פיתוח מודל חיזוי מבוסס GIS להערכת הסיכון להיפגעות בקטעי דרך בינעירוניים לפי פרמטרים תחבורתיים ומרחביים

דו"ח סופי

מוגש לקרן רן נאור

פרופ' אמציה פלד
קרן פולק, מ.א.

ל' כסלו תשע"ב **חיפה** **26.12.2011**

תוכן עניינים

1. מבוא 14
2. סקר ספרות 16
2.1 תאונות דרכים 16
2.1.1 כללי 16
2.1.2 התנהגות הנהג וקונספט הסכנה 17
2.1.3 משתנים מסבירים 19
2.1.4 מודלים ומחקרים נבחרים 26
2.1.5 איכות נתוני תאונות הדרכים והקושי בניתוחם 34
2.2 GIS ותחבורה 38
2.2.1 GIS-T 38
2.2.2 כריית מידע (Data Mining) 45
2.2.3 שיטות לייחוס גיאוגרפי 46
2.3 מתודות סטטיסטיות בשימוש חקר התחבורה 48
2.3.1 פיתוח מודלים סטטיסטיים 48
2.3.2 אגרגציה של מידע מרחבי וההשפעה על מאפייני הפיזור (פואסוני, נורמאלי, בינומי) 50
2.3.3 מתודות ניתוח סטטיסטיות נהוגות 51
2.3.4 תבניות מרחביות של נקודות (Spatial Point Patterns) ואישכול (Clustering). 73
3 מערך המחקר 92
3.1 מטרת המחקר 92
3.2 אזור המחקר 93
3.3 שיטות המחקר 96
3.4 נתונים 96
3.4.1 טבלאיים 96
3.4.2 שכבות מידע גיאוגרפי (GIS) 98
3.6.3 מודל גבהים סיפרתי (DEM) 98
4. מהלך המחקר 99
4.1 שלבי המחקר 99
4.2 בקרת איכות ועיבוד ראשוני (Pre-processing) של נתוני המקור 99
4.2.1 קבצי תאונות הדרכים של הלמ"ס 99
4.2.2 שכבת הק"מ בצמתים - חב' פלד מיפוי (בקרה ע"י קובץ PDF של הלמ"ס) 100
4.2.3 הצבת התאונות בכבישים בינעירוניים 102
4.2.4 שדה הכיווניות 103
4.2.5 קובץ הלמ"ס – נתוני תאורה וסטטוס יום/לילה 105
4.2.6 קובץ הלמ"ס - נתוני מסלול 106
4.2.7 קובץ הלמ"ס - רוחב השול 110
4.3 איכון תאונות הדרכים הבין עירונית, שאינן בצומת (קוד 4 של הלמ"ס) 112
4.3.1 כללי 112
4.3.2 הנתונים בהם נעשה שימוש עבור ההשמה במרחב 113
4.3.3 שלבים עיקריים בתהליך 114
4.4 חישוב פרמטרים עבור אירוע תאונת דרכים (נקודה) 120
4.4.1 שיוך "כיוון נסיעה" לתאונה. 120
4.4.2 חישוב זווית השמש עבור נקודה במישור אופקי 121
4.4.3 הוספת הערך "אזור טבעי" עבור כל תאונה בבסיס הנתונים 129
4.5 חישוב פרמטרים עבור מקטע דרך 131
4.5.1 כללי 131
4.5.2 מספר תאונות 131
4.5.3 חומרת תאונה 131
4.5.4 עקמומיות סגמנט 133
4.5.5 שיפוע סגמנט 136
4.5.6 נפח תנועה: AADT (Annual average daily traffic) 139
4.6 עריכה ושילוב הנתונים – השלב הסופי 147

4.6.1 התאמת השכבה הקווית לשכבת הגבהים (DEM) 147
4.6.2 סינון מקטעי דרך הנמצאים תחת השפעת צומת 148
4.6.3 החיבור הסופי בין שכבת הכבישים המחולקת לאורכים שונים לבין שכבת התאונות הנקודתית 150
4.6.4 דיוק המודל (איכון הנתונים) 151
5. ניתוח גיאו-מרחבי – נקודתי 153
5.1 Kriging 153
5.1.1 כללי 153
5.1.2 מהלך העבודה 154
5.1.3 ניתוח ותוצאות 155
5.2 איתור צבירים 158
5.2.1 כללי 158
5.2.2 מהלך העבודה 158
5.2.3 ניתוח ותוצאות 162
5.3 השוואת מתודת KRIGING למתודת איתור מקטעים מסוכנים לאורך ליניאמנטים 170
6. מערך ניסויים לבניית המודל 174
6.1 מהלך העבודה - בניית המודל הסטטיסטי 174
6.2 בחירת הפרמטרים המשמעותיים עבור המודל 175
6.3 שגרות מחשב 176
6.4 תוצרים וניתוחם (דוגמת 500 המטרים) 176
6.4.1 בחירת המודל 176
6.4.2 פרוצדורת GENMOD 176
6.4.3 CROSS VALIDATION של המודל 181
6.4.4 Empirical Bayes (להלן: EB) 187
7. ממצאים 190
7.1 מקטעי דרך באורך 500 מטרים 190
7.2 מקטעי דרך באורך 750 מטרים 190
7.2.1 בחירת המודל 190
7.2.2 פרוצדורת GENMOD 190
7.2.3 CROSS VALIDATION של המודל 193
7.2.4 Empirical Bayes (להלן: EB) – מקטעי דרך באורך 750מ'. 198
7.3 מקטעי דרך באורך 1000 מטרים 201
7.3.1 בחירת המשתנים עבור המודל 201
7.3.2 פרוצדורת GENMOD 201
7.4 מקטעי דרך באורך 1500 מטרים 211
7.4.1 בחירת המשתנים עבור המודל 211
7.4.2 פרוצדורת GENMOD 211
7.4.3 CROSS VALIDATION של המודל 213
7.5 ניתוחים משלימים 221
7.5.1 ניתוח אזורים (צפון, מרכז, דרום) 221
7.5.2 אינטרקציות 226
7.5.3 ניתוח תתי איזורים באזור הצפון 243
7.5.4 רגרסיה לוגיסטית 246
8. תוצרים וסיכום 251
9. סיכום מסקנות והמלצות 262
10. רשימת מקורות 266
נספחים 272

תקציר

מטרת המחקר היתה פיתוח מודל חיזוי מבוסס GIS להערכת הסיכון להיפגעות בקטעי דרך בינעירוניים לפי פרמטרים תחבורתיים ומרחביים. המחקר התמקד בחמש שאלות עיקריות:

1. בהינתן מידע בנוגע למאפיינים תחבורתיים ומרחביים עבור מקטע דרך מהירה בינעירונית - מהו שיעור תאונות הצפוי עבור אותו מקטע דרך, ומהי הסתברות להתרחשות של אפס תאונות או לפחות תאונה אחת?
2. בהינתן מידע בנוגע למאפיינים תחבורתיים ומרחביים עבור מקטע דרך מהירה בינעירונית, מהם המשתנים המשמעותיים ביותר, אשר להם השפעה על שיעור תאונות הדרכים?
3. האם מתקיימת אינטראקציה בין המשתנים: אזור ועקמומיות (האם העקמומיות הינה בעלת השפעה שונה על מספר התאונות באזורים שונים?), בין אזור ושיפוע (האם ניתן לאפיין מספר תאונות שונה עבור שיפועים באזורים שונים?) ובין נפח תנועה לבין אזור (האם השפעת נפח התנועה על שיעור תאונות הצפוי אינה אחידה, ותלויה באזור בו אנו נמצאים?).
4. האם מודלים סטטיסטיים דומים, המופעלים על מקטעי דרך באורכים שונים, יניבו תוצאות שונות בהיבט הפרמטרים המשמעותיים לקיום תאונות במקטע דרך?
5. מה היכולת לאתר צבירים, הן במרחב הליניאמנטים (קרי: לאורך הכבישים) והן במרחב הרסטרי (כלל שטח אזור המחקר) לרבות יתרונות וחסרונות של השיטות.

הפיזור המרחבי של אירועי תאונות דרכים (מודלי ספירה וחיזוי) נלמד בעניין רב במהלך העשורים האחרונים. מחקרים רבים בחנו גורמי חשיפה שונים וניסו לגבש משתנים מסבירים מתחומים שונים: צי רכב (פיזורו במרחב, חלוקה לקטגוריות שונות וגיל הרכב); נתיבים (כמות, מידות, הפרדה בין מסלולים) ; חשיפה (נסועה, נסועה לסוג כביש, נסועה למשתמשי דרך שונים, נסועה לקבוצות גיל ומגדר, כמות תנועה בסוגי דרך שונים); אוכלוסייה (גיל ומגדר של האוכלוסייה, מחזיקי רישיונות נהיגה לפי גיל ומגדר, צפיפות, אחוז נהגים צעירים) ; מזג אויר (טמפרטורה, רמות גשם ושלג, שמש, וקרח); בטיחות בדרכים (צריכת אלכוהול, שיעור חוגרי חגורות בטיחות, מהירות נסיעה, שיעור חובשי קסדות, שיעור כלי רכב המצוידים בכריות אוויר, הגבלות מהירות, מגבלת גיל לנהיגה, השקעות בבטיחות); מהירות; כלכלה (הכנסות למשק בית, צריכה למשק בית, שכר, תוצר לאומי גולמי, שיעור מובטלים, אוכלוסייה פעילה , ייצרו תעשייתי, מחירי גז, צריכת גז); בטיחות (השקעות בבטיחות הכבישים, הוצאות לאומיות עבור הנדסת הכבישים, השקעה לאומית בשיטור) ועוד [Page, 2001].

נפח התנועה (AADT) הינו גורם חשיפה, שנתפס לעיתים רבות כגורם בעל **ה**משמעות בגרימת אירועי תאונות דרכים. בהתאמה לכך, נחקר לעומק בעשרות השנים האחרונות תוך שימוש במגוון רחב של מתודות סטטיסטיות. זאת, החל מבחינת השפעות והקשרים ליניאריים תוך שימוש במודלים של רגרסיה ליניארית; דרך מודלים בהם נעשה שימוש בשיטת הריבועים הפחותים

(least squares); ולעיתים תוך שימוש בטרנספורמציות LOG או בפולינומים מדרגה שנייה ויותר. בשלבים מתקדמים יותר פותחו, כאלטרנטיבה לרגרסיה הליניארית, מודלי פואסון. זאת, עקב ההנחה של פיזור פואסוני לספירות אירועי תאונות דרכים, ומודלים בינומיים שליליים [Thomas, 1996]. מאוחר יותר אף פותחו מודלי Zero Inflated. אלה מאפשרים לכלול במודל גם מקטעי דרך המאופיינים בערך של אפס תאונות למקטע דרך [Qin et-al, 2005].

המחקר הנוכחי מחדש במספר היבטים. ראשית, בהיבט שילוב הפרמטרים המרחביים. במחקר שולבו פרמטרים מרחביים, בחלקם לא נמצא דיווח בהקשר לחקר הבטיחות בדרכים. כך לדוגמא, הוצגה היכולת לבצע חישוב פרמטר של זווית שמש למטרת בחינת תאונות שהתרחשו לכאורה כתוצאה מסינוור. בנוסף, חושבו פרמטרים של שיפוע ועקמומיות מקטעי הדרך. במחקר זה פרמטרים אלה נדונו לעומקם, תוך בחינת מספר נוסחאות על תוצאותיהן, ובחירת הנוסחה שסיפקה את המענה המרחבי הטוב ביותר. כמו כן, בוצע שיוך של כל מקטעי דרך לאזור גיאוגרפי. במקרה שלנו רק לשלושה: צפון, מרכז ודרום, באופן התואם את אזורי ה"פריפריה" (צפון ודרום) ואת אזור הליבה ("מטרופולין ת"א"). ניתוח "משתנה האזור" (Zone), כפי שנבחר באופן החלוקה והשילוב במחקר, העלה ממצאים מעניינים לגבי שיעור התאונות הגבוה הצפוי באזור הצפון ובמיוחד באזור "הגליל העליון המערבי". זה האחרון נמצא בניסוי בו נעשתה גם חלוקת משנה לאזור הצפון.

משתנים נוספים וייחודיים ששולבו במחקר זה הינם משתנים המייצגים אינטראקציות בין אזור לעקמומיות, בין אזור ושיפוע ובין נפח תנועה והאזור. אלה, אפשרו לבחון האם העקמומיות הינה בעלת השפעה שונה על מספר התאונות באזורים השונים; האם ניתן לאפיין מספר תאונות שונה עבור שיפועים באזורים השונים ; והאם השפעת נפח התנועה על שיעור תאונות הצפוי תלויה באזור בו נמצאים, בהתאמה.

ההיבט השני הינו ההיבט המתודולוגי. במהלך המחקר פותחה מתודולוגיה המאפשרת להצביע על מקטעי דרך קצרים מאוד, שהינם בעלי סיכון להתרחשות תאונת דרכים. להפניה מדויקת יש משמעות אופרטיבית של חיסכון במשאבים הן מבחינת התשומות והן מבחינת זמן הטיפול והשיפור. יתר על כן, חלוקת בסיס הנתונים למקטעי דרך באורכים שונים, אפשרה לנטרל את פרמטר "אורך הסגמנט", שבמחקרים רבים עלה כפרמטר משמעותי ברמת המודל. כמו כן, טוייב המודל על ידי שימוש בשיטת Empirical Bayes המעלה את דיוק ההערכה עקב שימוש במידע היסטורי ומתקנת את ההטייה עקב תופעת ה"רגרסייה אל הממוצע".

היבט מתודולוגי נוסף, נמצא בהתייחסות לכיוון התאונה בעת הצבת התאונות במרחב. על אף שברור כי נתון הכיווניות הינו בעייתי, נעשה בו שימוש, הן בפיתוח שאילתות SQL המאפשרות לפתח מתודולוגיה של חישוב זווית שמש למטרת איתור תאונות המרחשות כתוצאה מסינוור, והן באיכון התאונות במרחב. עוד, נכתב קוד SQL המאפשר לאתר צבירים במרחב הליניאמנטי המרחבי ולא במרחב הפוליגונאלי.

בסיכום המחקר נמצא כי מקטעי הדרך באורכים הקצרים יותר שנבחנו, 500 מטרים ו- 750 מטרים, הינם היציבים ביותר. המשמעות היא כי ניתן להמליץ על התייחסות והשקעה (הן בתשתית והן באכיפה) במקטעים ממוקדים, להבדיל מההתייחסות רחבה של שיפור מקטע ארוך יחסית של הכביש.

נוסחה a מציגה את שיעור תאונות הדרכים הצפוי (μ) עבור מקטעי דרך בינעירוניים באורך **500** מטרים:

[a]

$$\mu = e^{-7.091+0.714*AADT\_Log+(-0.678)*Ezor1+(0.569)*Ezor3+(1.594)*CurvePrm3}$$

נוסחה **b** מציגה את שיעור תאונות הדרכים הצפוי (μ) עבור מקטעי דרך בינעירוניים באורך של **750** מטרים:

[b]

$$\mu = e^{-5.9688+0.6420*AADT\_Log+(-0.7316)*Ezor1+(-0.5483)*Ezor3+(1.0741)*CurvePrm3}$$

כאשר:

aadt_log - ממוצע נפח תנועה יומי לימי חול, לאחר טרנספורמציה ללוגריתם טבעי.

ezor - משתנה האזור. **ezor1** = דרום, **ezor3** = מרכז ו- **ezor2** =צפון. המודל משווה ומנתח בין אזור הצפון לשני האזורים האחרים. כך, עבור מקטע המצוי באזור 2 יוצב הערך 0 במשתנים **ezor3** ו- **ezor1**.

CurvePrm3 – עקמומיות.

לאחר חישוב שיעורי התאונות הצפויים, בוצע הליך ה- EB לטיוב המודל, לפי נוסחה C המופיעה במאמרם של Sayed & Rodriguez [1999] ו- Ng et-al [2002], והמוצגת להלן:

[c] $$\left[\frac{\lambda}{\frac{1}{\alpha}+\lambda}\right]\times\left(\frac{1}{\alpha}+y\right)$$

כאשר:

α - פרמטר ה overdispersion , כפי שחושב עבור סט נתונים באורך מסויים

λ - שיעור תאונות בפועל

y- שיעור התאונות צפוי (μ בנוסחאות a ו- b).

בהתאם למודלים היציבים שנמצאו כמיטביים, **הפרמטרים המשמעותיים** והמשפיעים על שיעור התאונות הינם: **עקמומיות** מקטע הדרך, **האזור** בו נמצא מקטע הדרך **ונפח התנועה** עליו.
בהיבט **העקמומיות** הממצא תואם מחקרים אחרים שנערכו בתחום, דוגמת מחקרו של Miaou [1994]. אם כי במחקר זה, בשונה ממחקרו של Miaou [1994], שמצא קשר בין עקמומיות לבין שיעור התאונות רק עבור משאיות גדולות, נבנה המודל במחקר הנוכחי על כלל סוגי כלי-הרכב, ללא חלוקה לסוג רכבים מסויים, ובוצע תקנון עבור הנתונים תוך שימוש בשיטת EB. גם בהיבט פרמטר **נפח התנועה** שנמצא כמשמעותי במודל, נמצאה התאמה לספרות הקיימת, דוגמת מחקרם של: Oppe [1991], Levine et-al [1995] ו- Hauer [2001].
המשתנה השלישי שנמצא משמעותי, הינו כאמור, משתנה **האזור** – בישראל, כאשר החלוקה שנבחרה לשילוב במחקר (צפון, מרכז, דרום) תואמת בגדול את אזורי ה"פריפריה" (צפון ודרום) ואת אזור הליבה ("מטרופולין ת"א"). הממצא שעלה בעיקביות לאורך כלל הניתוחים, הינו שיעור התאונות **הגבוה הצפוי באזור הצפון** ובמיוחד באזור "**הגליל העליון המערבי**". נראה כי לממצא זה, המגלה כאמור עיקביות לאורך כלל הניתוחים הסטטיסטיים-מרחביים, משמעויות ונגזרות רבות. על מנת לספק הסבר מעמיק לממצא זה, יש לבוחנו באספקט רחב, וניכר כי יש לאסוף עוד מידע רב (לאו דווקא מתחומי התחבורה ה"טהורה"), שלא נאסף ושולב במחקר זה בצורה ייעודית. כוונה היא למידע דוגמת: סוג אוכלוסיה, מצב סוציואקונומי, השכלה, השקעה בתשתית ועוד ועוד. המשמעות האופרטיבית של ממצא זה ברורה. יש להסיט משאבים לאזורים אלה ולבצע פעולות מעשיות, החל מהכוונת אכיפה והשקעת כוח אדם, דרך שיפור תשתית, חינוך והסברה, וכמובן להמשיך במחקר המנסה לתת מענה ל"בולטות" של אזור ה"צפון" ותת-אזור ה"גליל העליון המערבי" בהיבט ריכוזי התאונות הגבוהים שבהם.

למטרת חישוב **ההסתברות להתרחשות של אפס תאונות או לפחות תאונה אחת במקטע דרך** הורץ מודל של רגרסיה לוגיסטית. תוצאות המודל לא איפשרו לנבא במידה מספקת, את ההסתברות לקיומן של אפס תאונות או לפחות תאונה אחת, היות ותוצאות המודל מצביעות על אקראיות כמעט מלאה. ממצא זה מצביע על כך, שקיימים גורמים נוספים אחרים שלא שולבו במודל. אולי חברתיים, אולי מאפייני דרך פיזיים נוספים דוגמת: רוחב שול, מספר נתיבים ואולי אחרים.

תוצאות ניתוחי האינטראקציות השונים הראו, כי לא מתקיימת אינטראקציה בין "אזור" ל"עקמומיות" ובין "אזור" ל"שיפוע", כאשר בבואנו לבחון את ה**אינטראקציה בין "נפח התנועה" לבין "אזור" נמצא כי אכן מתקיימת אינטראקציה**. קרי, השפעת נפח התנועה על שיעור תאונות הצפוי אינה אחידה, ותלויה באזור בו נמצאים.

בבחינת היכולת לאתר צבירים הן במרחב הליניאמנטים (קרי: לאורך הכבישים) והן במרחב הרסטרי (כלל שטח המחקר תוך שימוש במתודת – Krigging**),** נראה כי שתי המתודות מספקות מענה טוב לאיכון צבירים במרחב, כאשר על מנת לקבל תוצאה טובה וממוקדת יותר בעת השימוש ב- Krigging**,** מומלץ להקטין ככל האפשר את הרשת (Grid), ובכך למרכז את ההסתברויות לתאונה "על הכביש" ולא במרחב ההדמאה הכללי, בו ברור כי אין בפועל כבישים.

בהיבט הנתונים ה"גולמיים", ממצאי המחקר מצביעים על יישימות עקרונית לשימוש בנתוני מד"א למטרת טיוב נתוני המשטרה והלמ"ס באזורים עירוניים.

דומה כי עבודת מחקר זו, על מימצאיה, מהווה בסיס ופתח למחקרים מגוונים נוספים בתחום הבטיחות בדרכים. מחקר זה מאשש שוב את העוצמה והתועלת שבשימוש במערכת מידע גיאוגרפי (להלן: GIS – Geographic Information Systems) בתחום הבטיחות בדרכים (GIS-T). זו, מאפשרת את הצגת הנתונים ובחינתם במרחב ההתרחשות ה"אמיתי" (המיקום הפיזי של התרחשות האירוע), איטרטיביות וביצוע ניתוחים סטטיסטיים-מרחביים מורכבים' באופן פשוט יחסית.

רשימת טבלאות

טבלה 1 - מספר הרוגים בתאונות דרכים בישראל [מקור: אתר הרשות הלאומית לבטיחות בדרכים] 14
טבלה 2 - ערכי המעריך וערכי הקבועים שנמצאו אופטימאליים עבור כל מדינה [Oppe, 1991]. 30
טבלה 3 - תוצאות המודל של Page [2001] 31
טבלה 4 - גיל ומחלות לב כליליות (להלן: CHD) [Hosmer & Lemeshow, 2000] 70
טבלה 5 - תדירות קיום מחלת CHD בקרב קבוצות הגיל השונות [Hosmer & Lemeshow, 2000] 71
טבלה 6 - תדירות הפיזור של עצים באזורי הדגימה, ברדיוס של 6 מטרים [Cressie, 1993] 76
טבלה 7 - אינדקסים עבור בחינת פיזור [Cressie, 1993] 76
טבלה 8 - תאונות דרכים עם נפגעים - קובץ בקרה 100
טבלה 9 - ערכי הק"מ בצמתים בשכבת ה Roads אל מול קובץ הלמ"ס, בין צומת הכפר הירוק למחלף מורשה. 101
טבלה 10 - כמות אירועי תאונות הדרכים (למ"ס, 2005) בכבישים בינעירוניים, בחלוקה לחומרה וכיווניות 104
טבלה 11 - דרך חד מסלולית – התפלגות כמות אירועים וסוג 106
טבלה 12 - דרך רב מסלולית – התפלגות כמות אירועים וסוג 107
טבלה 13 -"רוחב הכביש" בפרסום הלמ"ס, והתפלגות מספר האירועים מכל סוג בכבישים הבינעירוניים 110
טבלה 14 - תאונות מסוג בין עירוני לא בצומת 113
טבלה 15 - עמודות שכבת הכבישים 118
טבלה 16 -ערכי קוד מינימום ומקסימום ל"כיווניות" תאונות דרכים בינעירוניות 121
טבלה 17 - נוסחאות עקמומיות שנבחנו עבור סגמנטים באורכים שונים 133
טבלה 18 - נוסחאות חישוב שיפוע מקטע דרך 136
טבלה 19 - סטטוס מספר הכבישים ונתוני AADT הקיימים עבורם 141
טבלה 20 - ערך AADT מחושב עבור מקטע דרך 5881 בשכבת הכבישים המחולקת לסגמנטים באורך 1000 מ' 145
טבלה 21 - תוצאות הבחינה הידנית עבור נפח תנועה מחושב 147
טבלה 22 - מקטעי הדרך המהווים את בסיס הנתונים הסופי לבניית המודל 150
טבלה 23 - תוצר AIC עבור מקטעי דרך באורך 500 מ' 176
טבלה 24 - Model Information - (מקטעי דרך באורך 500 מ') 177
טבלה 25 - טבלת Criteria For Assessing Goodness Of Fit (מקטעי דרך באורך 500 מ') 177
טבלה 26 - תוצר פרוצדורת GENMOD עבור מודל האימון של 2/3 הנתונים (ניתוח סגמנטים באורך 500 מ') 177
טבלה 27 - Contrast Estimate Results, מקטעי דרך באורך 500 מ'. 180
טבלה 28 - תוצר פרוצדורת GENMOD עבור נתוני קבוצת הבקרה של 1/3 הנתונים (ניתוח סגמנטים באורך 500 מ') 182
טבלה 29 - מבחן TOST – מקטעי דרך 500 מ' 184
טבלה 30 - ההסתברות להתרחשות N תאונות במקטעי הדרך באורך 500 מטרים 184
טבלה 31 - ממוצע ההסתברויות להתרחשות 0 תאונות ותאונה אחת במקטע דרך באורך 500 מ' 185
טבלה 32 - פרופורציות המקטעים האמיתיים בהם התרחשו N תאונות 185
טבלה 33 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 500 מ') – טבלת Analysis Of Maximum Likelihood Parameter Estimates 186
טבלה 34 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 500 מ') טבלת – Contrast Estimate Results 186
טבלה 35 - תוצר AIC עבור מקטעי דרך באורך 750 מ' 190
טבלה 36 - Analysis Of Maximum Likelihood Parameter Estimates, 750 מודל האימון 2/3 מהנתונים 190
טבלה 37 - Contrast Estimate Results, 750, מודל האימון (2/3 מהנתונים) 192
טבלה 38 - תוצר פרוצדורת GENMOD עבור נתוני קבוצת הביקורת של 1/3 הנתונים (ניתוח סגמנטים באורך 750 מ') 194
טבלה 39 - תוצאות מבחן TOST , סגמנטים באורך 750 מ' 195
טבלה 40 - ממוצע הסתברויות מחושב להתרחשות 0 תאונות ותאונה אחת במקטע דרך באורך 750 מ' 195
טבלה 41 - פרופורציות המקטעים האמיתיים בהם התרחשו N תאונות - 750 מ' 196
טבלה 42 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 750 מ') – טבלת Analysis Of Maximum Likelihood Parameter Estimates 197
טבלה 43 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 750 מ') טבלת – Contrast Estimate Results 197
טבלה 44 - תוצר AIC עבור מקטעי דרך באורך 1000 מ' 201
טבלה 45 - Analysis Of Maximum Likelihood Parameter Estimates, 1000 מודל האימון 2/3 מהנתונים 201
טבלה 46 - טבלת Contrast Estimate Results, מקטעי דרך 1000 מ'. 203
טבלה 47 - תוצר פרוצדורת GENMOD עבור נתוני קבוצת הביקורת של 1/3 הנתונים (ניתוח סגמנטים באורך 1000 מ') 205
טבלה 48 - מבחן TOST לשוויון – מקטעי דרך 1000 מ' 206
טבלה 49 - ממוצע הסתברויות מחושב להתרחשות 0 תאונות ותאונה אחת במקטע דרך באורך 1000 מ' 206
טבלה 50 - פרופורציות המקטעים האמיתיים בהם התרחשו N תאונות - 1000 מ' 207
טבלה 51 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 1000 מ') – טבלת Analysis Of Maximum Likelihood Parameter Estimates 208
טבלה 52 - תוצר AIC עבור מקטעי דרך באורך 1500 מ' 211
טבלה 53 - טבלת Analysis Of Maximum Likelihood Parameter Estimates , מקטעי דרך 1500 מ' 211
טבלה 54 - טבלת Contrast Estimate Results, מקטעי דרך 1500 מ' 213
טבלה 55 - תוצר פרוצדורת GENMOD עבור נתוני קבוצת הבקרה של 1/3 הנתונים (ניתוח סגמנטים באורך 1500 מ') 214

טבלה 56 - תוצאות מבחן TOST למקטעי דרך של 1500 מ'. 215
טבלה 57 - ממוצע הסתברויות מחושב להתרחשות 0 תאונות ותאונה אחת במקטע דרך באורך 1500 מ' 216
טבלה 58 - פרופורציות המקטעים האמיתיים בהם התרחשו N תאונות - 1500 מ' 217
טבלה 59 - טבלת Analysis Of Maximum Likelihood Parameter Estimates, עבור מודל מלא מ 1/3 נתוני הביקורת, מקטעי דרך של 1500 מ'. 218
טבלה 60 - השוואת אזור 1 (דרום) אל מול אזורים 2 (צפון) ו-3 (מרכז) עבור נפח בינוני-נמוך (5000 ל- 15000). מקטעי דרך של 500 מ' 221
טבלה 61 - השוואת אזור 2 (צפון) אל מול אזורים 1 (דרום) ו-3 (מרכז) עבור נפח בינוני-נמוך (5000 ל- 15000) מקטעי דרך של 500 מ' 222
טבלה 62 - השוואת אזור 1 (דרום) אל מול אזורים 2 (צפון) ו-3 (מרכז) עבור נפח תנועה בינוני-גבוה (15000 ל- 30000). מקטעי דרך של 500 מ' 222
טבלה 63 - השוואת אזור 2 (צפון) אל מול אזורים 1 (דרום) ו-3 (מרכז) עבור נפח תנועה בינוני-גבוה (15000 ל- 30000). מקטעי דרך של 500 מ' 223
טבלה 64 - השוואת אזור 1 (דרום) אל מול אזורים 2 (צפון) ו-3 (מרכז) עבור נפח בינוני-נמוך (5000 ל- 15000) 224
טבלה 65 - השוואת אזור 2 (צפון) אל מול אזורים 1 (דרום) ו-3 (מרכז) עבור נפח בינוני-נמוך (5000 ל- 15000) 224
טבלה 66 - השוואת אזור 1 (דרום) אל מול אזורים 2 (צפון) ו-3 (מרכז) עבור נפח תנועה בינוני-גבוה (15000 ל- 30000) 225
טבלה 67 - השוואת אזור 2 (צפון) אל מול אזורים 1 (דרום) ו-3 (מרכז) עבור נפח תנועה בינוני-גבוה (15000 ל- 30000) 225
טבלה 68 - השוואה בין ערכי ה AIC של מודל ה NULL לבין מודל בו מתקיימות אינטרציות בין המשתנים השונים (חלוקת נפח תנועה נמוך, בינוני וגבוה) 227
טבלה 69 - LR Statistics For Type 3 Analysis, מקטעי דרך של 500 מ' 227
טבלה 70 - מספר תאונות צפוי עבור מקטעי דרך בחלוקה לאזורים 229
טבלה 71 - תוצר מודל האינטרקציה. השוואת נפח תנועה נמוך (1) באזור 1 (דרום) אל מול נפחי 230
טבלה 72 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ' 231
טבלה 73 - תוצר מודל השוואת האינטרקציות - נפח תנועה גבוה (3) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'. 231
טבלה 74 - תוצר מודל השוואת האינטרקציות - נפח תנועה נמוך (1) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'. 231
טבלה 75 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'. 231
טבלה 76 - תוצר מודל השוואת האינטרקציות - נפח תנועה נמוך (1) באזור 3 (מרכז) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ' 232
טבלה 77 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 3 (מרכז) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ' 232
טבלה 78 - השוואה בין ערכי ה AIC של מודל ה NULL לבין מודל בו מתקיימות אינטרציות בין המשתנים השונים (חלוקת נפח תנועה נמוך, בינוני וגבוה) – 750 מ' 235
טבלה 79 - LR Statistics For Type 3 Analysis, מקטעי דרך של 750 מ' 235
טבלה 80 - מספר תאונות צפוי עבור מקטעי דרך בחלוקה לאזורים 236
טבלה 81 - תוצר מודל האינטרקציה. השוואת נפח תנועה נמוך (1) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ' 238
טבלה 82 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ' 238
טבלה 83 - תוצר מודל השוואת האינטרקציות - נפח תנועה גבוה (3) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ' 238
טבלה 84 - תוצר מודל השוואת האינטרקציות - נפח תנועה נמוך (1) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ' 239
טבלה85 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ' 239
טבלה 86 - תוצר מודל השוואת האינטרקציות - נפח תנועה גבוה (3) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'. 239
טבלה 87 - תוצר מודל השוואת האינטרקציות - נפח תנועה נמוך (1) באזור 3 (מרכז) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ' 239
טבלה 88 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 3 (מרכז) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'. 240
טבלה 89 - מספר תצפיות ששולבו במודל עבור טווח נפחי תנועה שונים 242
טבלה 90 - השוואה בין ערכי ה AIC של מודל ה NULL לבין מודל בו מתקיימות אינטרציות בין המשתנים השונים (חלוקת קבוצות נפח תנועה לפי התפלגות תחבורתית) 242
טבלה 91 - השוואה בין ערכי ה AIC של מודל ה NULL לבין מודל בו מתקיימות אינטרציות בין המשתנים השונים (חלוקת קבוצות נפח תנועה לפי התפלגות תחבורתית)- 750 מ' 243
טבלה 92 - השוואת תת אזור 21 (חיפה) לתת האיזורים האחרים בצפון 244

טבלה 93 - תוצאות ניתוח ההשוואת המרובות בין תתי האזורים בצפון 244
טבלה 94 - השוואת תת אזור 22 (עמקים) לתת האיזורים האחרים בצפון 245
טבלה 95 - השוואת תת אזור 24 (גליל תחתון מזרחי) לתת האיזורים האחרים בצפון 245
טבלה 96 - מקטעי הדרך בהם כלל לא התרחשו תאונות (726) אל מול מקטעי הדרך בהם התרחשה לפחות תאונה אחת (381)– נתוני אמת 246
טבלה 97 - ערכי AIC לפרמטרים שנבחנו עבור בניית מודל רגרסיה לוגיסטית 247
טבלה 98 - מקטעי הדרך בהם כלל לא התרחשו תאונות (361) אל מול מקטעי הדרך בהם התרחשה לפחות תאונה אחת (249)– נתוני אמת. מקטעי דרך של 750 מ' 248
טבלה 99 - ערכי AIC לפרמטרים שנבחנו עבור בניית מודל רגרסיה לוגיסטית. 249
טבלה 100 - תוצאות מסכמות נבחרות של מודלי הרגרסיה הבינומית השלילית, עבור ארבע סטי הנתונים שנבחנו במחקר 251
טבלה 101 - אינטרקציות – מספר תאונות צפוי עבור מקטעי דרך בחלוקה לאזורים 254
טבלה 102 - תוצאות בחינת השטח - בחיפה 278

רשימת איורים

איור מספר 1 - חישוב ממוצע עקמומיות משוקלל באורך הקטע [Road Research Laboratory ,1965].....25
איור מספר 2 -ניתוחים גיאומרחביים של תאונות דרכים במחקרם של Levine et-al [1995].....39
איור מספר 3 - יחסים מרחביים בין מקטעי דרך.....42
איור מספר 4 - פונקציה חשיפה ופונקציית בטיחות שאינה לינארית [Qin et-al , 2005].....44
איור מספר 5 - הליך סגמנטציה דינאמית [Cadkin, 2002].....47
איור מספר 6 - שיטת המיקום הדינאמי [Sutton & Wyman, 2000].....48
איור מספר 7 - היחס בין מחלות לב כליליות (CHD) לבין גיל [Hosmer & Lemeshow, 2000].....69
איור מספר 8 - שיעור נוכחות המחלה ביחס לגיל [Hosmer & Lemeshow, 2000].....71
איור מספר 9 - דגמי פיזור של נקודות.....74
איור מספר 10 - אזורי דגימה מעגליים המאופיינים ברדיוס של 6 מ'.....75
איור מספר 11 - פיזור מרחבי של נקודות הנמצאות במרחק שווה מאירוע נבחן [Nicholson, 1999].....77
איור מספר 12 - פיזור נקודות המאופיין בכיווניות דומה [Nicholson, 1999].....78
איור מספר 13 - אופני מדידת מרחק שונים [Nicholson, 1999].....78
איור מספר 14 - תבניות שונות של נקודות (מקובץ וסדור) [Nicholson, 1999].....82
איור מספר 15 - פונקצית הפיזור המצטברת [Nicholson, 1999].....83
איור מספר 16 - פיזור מרחבי של נקודות [Jones et-al, 1996].....85
איור מספר 17 - ערכי ההפרש K22 K11 עבור טווחי מרחקים נבחנים (עד 2 ק"מ).....87
איור מספר 18 - קונספט פונקציית K ופונקציית K עבור רשתות [Yamada & Thill,2004].....88
איור מספר 19 - סטטוס הכבישים לגביהם קיים נתון נפח תנועה (AADT).....94
איור מספר 20 - מקטעי הדרך של הכבישים הבינעירוניים ששולבו בבסיס נתוני המחקר (מקטעי דרך בחלוקה ל 1000 מ').....95
איור מספר 21 - תרשים זרימה - מהלך המחקר.....99
איור מספר 22 - כביש 5, דוגמא לפירוט נתון נפח תנועה.....101
איור מספר 23 - אורך מקטעים בין צומת הכפר הירוק למחלף מורשה.....102
איור מספר 24 - דוגמא לתקלה בנתון הכיווניות בסגמנט גירוגרפי.....105
איור מספר 25 - סוגי מסלולים (מתוך: http://www.teoria-yoav.com/dagesh/D1.asp).....106
איור מספר 26 - מספר נתיבים בכביש 70, 75 בין צומת יגור לג'למה –.....107
איור מספר 27 - מספר נתיבים בכביש 2 (החוף) בין מחלף חיפה דרום למחלף עתלית –.....109
איור מספר 28 - רוחב כביש עבור כל תאונה, בהתאם לקוד הלמ"ס.....111
איור מספר 29 - התאמת צמתים לפי שילובי קטעי דרך על בסיס מספר 'כביש מע"צ'.....115
איור מספר 30 - קבלת החלטות לשיוך צמתים בין שכבת "מפה" ו"פלד מיפוי".....117
איור מספר 31 - חלון "MAKE ROUTE EVENT LAYER".....120
איור מספר 32 - מישור הדקלינציה [מקור: http://cgate.co.il/navigation/declination.htm].....123
איור מספר 33 - חישוב כיוון השמש ביחס למישור אופקי.....125
איור מספר 34 - דוגמא לצורך בחישוב כיוון הכביש (אזימוט הרכב) ושיפועו ביחס לקטע בו נמצאת המכונית.....127
איור מספר 35 - דוגמא לצורך בחישוב כיוון הכביש (אזימוט הרכב) ושיפועו ביחס לקטע בו נמצאת המכונית.....127
איור מספר 36 - אזימוט כלי הרכב (כיווניות כלי הרכב ביחס לצפון).....128
איור מספר 37 - מפת חלוקה לאזורים גיאוגרפיים.....130
איור מספר 38 - חישוב עקמומיות - הסבר.....133
איור מספר 39 - עקמומיות מקטע דרך, נוסחאות 77-74, אורך סגמנט 1 ק"מ.....134
איור מספר 40 - עקמומיות מקטע דרך, נוסחאות 81-78, אורך סגמנט 1 ק"מ.....135
איור מספר 41 - כלי ה surface length בתוכבת ARCMAP.....136
איור מספר 42 - שיפוע מקטע דרך, נוסחאות 85-82, אורך סגמנט 1 ק"מ.....138
איור מספר 43 - שיפוע מקטע דרך, נוסחאות 87-86, אורך סגמנט 1 ק"מ.....139
איור מספר 44 - מקטעי דרך עבורם קיים מידע נפח תנועה (AADT) בחלוקה לקבוצות.....142
איור מספר 45 - השמת ערך AADT עבור מקטע דרך באורך L.....144
איור מספר 46 - בקרת איכות נתון נפח תנועה. מקרה לדוגמא, סגמנט 5881.....146
איור מספר 47 - שכבת הכבישים מוטלת על שכבת הגבהים.....148
איור מספר 48 - מתחמים עירוניים לאורך כבישי המחקר – דוגמא, אזור חיפה ועמקים.....149
איור מספר 49 - יצירת fishnet (רשת GRID).....154
איור מספר 50 - KRIGING - מפת הסתברות לתאונה , גודל תא שטח 1000X1000 מ'.....156
איור מספר 51 - KRIGING - מפת הסתברות לתאונה , גודל תא שטח 500X500 מ'.....157
איור מספר 52 - איתור צבירים - תרשים זרימה.....160
איור מספר 53 - תוצר הרצת קוד ה SQL לחילוץ צבירים.....161
איור מספר 54 - צבירי תאונות דרכים –מרחק של 200 מ' בין אירועי לאירוע (למעלה מ 20 אירועים בתקופת המחקר).....163
איור מספר 55 – המשך: צבירי תאונות דרכים – מרחק של 200 מ' בין אירועי לאירוע (למעלה מ 20 אירועים בתקופת המחקר) ...164
איור מספר 56 - צבירי תאונות דרכים –מרחק של 150 מ' בין אירועי לאירוע (למעלה מ 20 אירועים בתקופת המחקר).....165
איור מספר 57 - צבירי תאונות דרכים – מרחק של 100 מ' בין אירועי לאירוע (למעלה מ 15 אירועים בתקופת המחקר).....166

איור מספר 58 - צבירי תאונות דרכים שניתן לקושרם למזג האוויר– מרחק של 200 מ' בין אירועי לאירוע (3 אירועים ומעלה בתקופת המחקר) 167
איור מספר 59 - צבירי תאונות דרכים שניתן לקושרם למזג האוויר– מרחק של 150 מ' בין אירועי לאירוע (3 אירועים ומעלה בתקופת המחקר) 168
איור מספר 60 - צבירי תאונות דרכים שניתן לקושרם למזג האוויר– מרחק של 100 מ' בין אירועי לאירוע (3 אירועים ומעלה בתקופת המחקר) 169
איור מספר 61 - צבירי תאונות דרכים שניתן לקושרם למזג האוויר– מרחק של 100 מ' בין אירועי לאירוע (3 אירועים ומעלה בתקופת המחקר) 170
איור מספר 62 – אזורים בעלי סיכון לפי מתודת Kriging וקטעי דרך מסוכנים שאותרו לאורך ליניאמנטים (קוד SQL עבור מרחק של 100 מ') 172
איור מספר 63 - אזורים בעלי סיכון לפי מתודת Kriging וקטעי דרך מסוכנים שאותרו לאורך ליניאמנטים (קוד SQL עבור מרחק של 100 מ') 173
איור מספר 64 - מספר תאונות צפוי אל מול מספר תאונות אמיתי – מקטעי דרך באורך 500 מ' 181
איור מספר 65 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 500 מ' 188
איור מספר 66 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 500 מ' (ממספר תצפית 0 עד מספר תצפית 500) 188
איור מספר 67 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 500 מ' (ממספר תצפית 1500 עד מספר תצפית 2000) 189
איור מספר 68 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 500 מ' (ממספר תצפית 3000 עד מספר תצפית 3300) 189
איור מספר 69 - מספר תאונות צפוי אל מול מספר תאונות אמיתי – מקטעי דרך באורך 750 מ' 193
איור מספר 70 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 750 מ' (כלל התצפיות) 198
איור מספר 71 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 750 מ' (ממספר תצפית 0 עד מספר תצפית 500) 199
איור מספר72 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 750 מ' (ממספר תצפית 500 עד תצפית 1000) 199
איור מספר73 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 750 מ' (ממספר תצפית 1000 עד מספר תצפית 1800) 200
איור מספר74 - מספר תאונות צפוי אל מול מספר תאונות אמיתי – מקטעי דרך באורך 1000 מ' 204
איור מספר75 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1000 מ' (כלל התצפיות) 209
איור מספר 76 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1000 מ' (ממספר תצפית 0 עד מספר תצפית 400) 209
איור מספר 77 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1000 מ' (ממספר תצפית 400 עד מספר תצפית 800) 210
איור מספר 78 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1000 מ' (ממספר תצפית 800 עד מספר תצפית 1200) 210
איור מספר 79 - מספר תאונות צפוי אל מול מספר תאונות אמיתי – מקטעי דרך באורך 1500 214
איור מספר 80 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1500 מ' (כלל התצפיות) 219
איור מספר 81 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1500 מ' (ממספר תצפית 0 עד מספר תצפית 200) 219
איור מספר 82 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1500 מ' (ממספר תצפית 200 עד מספר תצפית 400) 220
איור מספר 83 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1500 מ' (ממספר תצפית 400 עד מספר תצפית 600) 220
איור מספר 84 - השפעת נפח תנועה קטגוריאלי, בחלוקה לאזורים, ... 228
איור מספר 85 - אינטקראציה בין נפח תנועה קטגוריאלי לבין אזור (מקטע דרך 750 מ') 236
איור מספר 86 - רגרסיה לוגיסטית, גרף ה ROC עבור מקטעי דרך באורך 500 מ' 248
איור מספר 87 - רגרסיה לוגיסטית, גרף ה ROC עבור מקטעי דרך באורך 750 מ' 249
איור מספר 88 - מספר תאונות צפוי, מקטעי דרך באורך 500 מטרים 258
איור מספר 89 - מספר תאונות צפוי, מקטעי דרך באורך 750 מטרים 259
איור מספר 90 - מספר תאונות צפוי, מקטעי דרך באורך 1000 מטרים 260
איור מספר 91 - מספר תאונות צפוי, מקטעי דרך באורך 1500 מטרים 261
איור מספר 92 - חילוץ מאפייני הכביש באמצעים של חישה מרחוק. 263
איור מספר 93 - דוגמא – השמת תאונות עירוניות. הסטת נקודת מרכז פוליגון לכביש הקרוב 276

1. מבוא

מידי שנה נהרגים בתאונות דרכים במדינת ישראל מאות בני אדם ואלפים נפצעים באורח קשה. למעשה, מספר הנפגעים בתאונות דרכים בישראל עולה על כלל מספר הנפגעים בכל מערכות ישראל [פלד, 1996]. טבלה מספר 1 מציגה את מספר ההרוגים בתאונות דרכים בישראל בין השנים 2003 ל - 2010. ניתן לראות כי עד שנת 2009 קיימת מגמת ירידה, כאשר מגמה זו נעצרה בשנת 2010.

טבלה 1 - מספר הרוגים בתאונות דרכים בישראל [מקור: אתר הרשות הלאומית לבטיחות בדרכים]

שנה	2003	2004	2005	2006	2007	2008	2009	2010
סה"כ הרוגים כולל מחוז ש"י	486	518	476	448	431	449	352	380

על אף מגמת הירידה, וחרף קיומם של מחקרים רבים ומגוונים בתחום תאונות הדרכים, עדיין קיים הצורך במחקרים משלימים (מכלל הסוגים: מתארים, מסבירים וחוזים), הבוחנים מגוון משתנים המקושרים לתאונות דרכים. זאת, היות ומספר ההרוגים העומד על מאות בשנה הינו כבד מנשוא, ויש להמשיך ולשאוף לצמצם ככל האפשר פגיעה בגוף ובנפש.

מחקר תאונות הדרכים החל להתבסס כתחום מחקרי בעל חשיבות החל משנות ה – 70 של המאה העשרים. בעשורים האחרונים נעשו מאמצים רבים בתחום חקר הבטיחות בדרכים בעולם ובישראל. מגמות המחקר בישראל מגוונות. חלק מהמחקרים עוסקים בתחום ה- ITS (Intelligent Transportation Systems) לשם קידום הבטיחות בתעבורה. במהלך מחקרים אלה נבחנת היכולת לשלב טכנולוגיות מתקדמות מתחומי החישה מרחוק (מכ"מ לייזר, אינפרה-אדום), האופטיקה ועיבוד התמונה, ניווט (GPS) ותקשורת נתונים (קווית ואלחוטית) בתחום התחבורה היבשתית. סוג נוסף של מחקרים עוסק בהערכת אפקטיביות בתחומים שונים, החל מהערכת אפקטיביות של מגרשי הדרכה לבטיחות בדרכים, דרך הערכת אפקטיביות "נוהל 6" (נוהל המחייב מעקב אחר תלונות המיוחסות לנהגים המועסקים על ידי מפעל או ארגון), עבור דרך אפקטיביות פעילות מערך משטרת התנועה והערכות אפקטיביות של קמפיינים הסברתיים שונים. פן נוסף, עוסק באפיון עמדות והתנהגות של משתמשי דרך בסקטורים שונים (דוגמת המגזר החרדי והערבי), כאשר תחום מחקר חשוב נוסף, עוסק במחקרים בהם נעשה ניסיון להתאמת משתנים מסבירים למגמות השונות של תאונות הדרכים, לרבות ניתוחים סטטיסטיים ובחינת מגמות רב שנתיות. אלה, מבוצעים על ידי גופים ממשלתיים וציבוריים דוגמת: הלמ"ס, משרד התחבורה, משטרת ישראל, אקדמיה, מכוני מחקר וגופי מחקר אחרים [הרשות הלאומית לבטיחות בדרכים].

על אף האמור לעיל, לא פותחו עד כה, אפליקציות התואמות את עומק היכולת של מערכות המידע הגיאוגרפי למטרת בטיחות בתחבורה. מערכות המידע הגיאוגרפי משמשות כיום בעיקר לזיהוי ולקביעת מיקומים בהם קיימת לכאורה סבירות גבוהה לתאונות דרכים. אם כי חשוב לציין, כי בעולם הולכת ומתפתחת מגמת מחקר, המשלבת ניתוחים מרחביים הנעשים תוך שימוש במערכות מידע גיאוגרפי עם מתודות סטטיסטיות הנפוצות בתחום חקר תאונות הדרכים. למרות הניסיונות בתחום זה, עדיין קיים הצורך בניתוח סטטיסטי משלים ומעמיק לקביעת הגורמים המשפיעים על איכון "מיקומים" אלה. בין גורמים אלה ניתן למנות: מזג אוויר, מאפיינים סוציו-אקונומיים של הנהג, מאפיניים סביבתיים ואחרים. הניסיון שהצטבר בעבודות קיימות, מראה כי בעת מימוש הניתוח במערכות המידע הגיאוגרפי, יש לשלב גורמים אלה ולא להסתפק רק בנתונים גיאומטריים של מקטעי הדרך או נתונים תחבורתיים דוגמת נפח תנועה. המגמה שתוארה לעיל, תואמת גם את המחקר בישראל, ולפיה מרבית המחקרים עד כה, לא התמקדו בהיבט הגיאו-מרחבי או הגיאו-סטטיסטי המרחבי.

לאור האמור, ניסה המחקר לתרום להשלמת הפער הקיים בתחום המיפוי המרחבי, ושם לו למטרה לשלב בין המתודות הסטטיסטיות הנהוגות בחקר הבטיחות בתחבורה, לבין הכלים המרחביים הקיימים במערכות המידע הגיאוגרפי. המחקר שואף לחדש הן ברמת "בסיס המידע" שימוזג ממספר מקורות (חלקם חדשניים), והן ברמת המשתנים והמידע המרחבי שישולבו במודל. על כן, הוגדרה מטרת המחקר כ: פיתוח מודל חיזוי מבוסס GIS להערכת הסיכון להיפגעות בקטעי דרך בינעירוניים לפי פרמטרים תחבורתיים ומרחביים.

המודל עתיד לתרום לאיכון אזורי תורפה ואיתור מיקומי תופעות נשנות. כך יתאפשר מיקוד באזורים בהם נדרשת השקעה בתשתית או אכיפה מוגברת. בנוסף, יאפשר המחקר להעמיק בלימוד וניתוח הגורמים לתאונות דרכים למטרת שיפור והקטנה של מספר אירועי התאונות. יישום המודל יאפשר להעריך טיבן של תוכניות קיימות בנוגע לדרכי ההתמודדות עם תאונות הדרכים, ואף יוכל לתרום לפיתוח תוכניות חדשות להתמודדות עם תאונות הדרכים הן ברמה המחקרית והן ברמה היישומית. השקעה במחקר המאפשרת גזירת תוכניות אופרטיביות, הקמת בסיס מידע אחוד, השקעת משאבים הן בתחום התשתית והן בתחומי האכיפה, ובמקביל, השקעת מאמץ בתחום ההסברתי והחינוכי הן על ידי גופים לאומיים-ממשלתיים (משרד החינוך, הרשות הלאומית לבטיחות בדרכים) והן על ידי גופים ציבוריים ופרטיים חיצוניים (דוגמת עמותת אור ירוק), אפשר שיובילו לשינוי המצופה, ואולי אף יפתחו פתח להפחתת מספר נפגעי תאונות הדרכים.

2. סקר ספרות

2.1 תאונות דרכים

בפרק זה תסקר ותוצג ספרות קיימת, הרלוונטית לנושא המחקר. הסקירה מחולקת מתודולוגית לשלושה חלקים. ראשית, תובא סקירה כללית בנוגע לתאונות דרכים, גורמים מקושרים ומודלים מסבירים. בחלק השני, יסקרו ההקשרים בין מערכות מידע גיאוגרפי לבין ניתוח מרחבי של תאונות. בחלק השלישי של סקירת הספרות, תוצגנה מתודות סטטיסטיות נפוצות הנמצאות בשימוש חקר התחבורה, כאשר לעיתים בגין מורכבות השיטה, הן תוסברנה תוך כדי הצגת מחקר לדוגמא.

2.1.1 כללי

תאונות דרכים גובות מחיר יקר הן בגוף, הן בנפש והן מבחינה כלכלית. ארגוני בריאות שונים בעולם מעריכים, כי בתאונות דרכים נהרגים מדי שנה כ- 600,000 איש וכ- 15 מיליון נפגעים [Page, 2001]. למרות העליה בבטיחות בשלושת העשורים האחרונים, תאונות דרכים עדיין מהוות את גורם המוות המוביל גם בארה"ב [Song et-al, 2006]. במדינת ישראל נהרגים מדי שנה בתאונות דרכים מאות בני אדם ואלפים נפצעים באורח קשה [הרשות הלאומית לבטיחות בדרכים, 2005]. מספר הנפגעים בתאונות דרכים בישראל עולה על מספר הנפגעים בכל מערכות ישראל [פלד, 1996]. אומדן עלות הנזק הכלכלי למשק הלאומי (במחירי יולי 2004) עומד על סכום של כ - 12.6 מיליארד ש"ח בשנה, שהם כ - 2.5% מהתל"ג. שיעור זה דומה לזה שנאמד בארצות מערב אירופה ובארה"ב, למרות הבדלים בשיטות ההערכה ובפרמטרים הכלכליים שבהם נעשה שימוש [הרשות הלאומית לבטיחות בדרכים, 2005]. על אף מגמת הירידה, שנצפתה בשנים האחרונות וחרף קיומם של מחקרים רבים ומגוונים בתחום תאונות הדרכים, עדיין קיים הצורך במחקרים משלימים, הבוחנים מגוון משתנים המקושרים לתאונות דרכים.

על אף שמספר תאונות הדרכים בעולם ומספר כלי הרכב עולה בהתמדה, במונחי בטיחות בתנועה דווקא קיים שיפור במדינות רבות, בעיקר במדינות המתועשות [Peled et-al, 1996; Page, 2001]. בשינויים במגמות התאונות, הן בטווח הארוך והן בטווח הקצר, מקובל לראות השתקפות של המדיניות והמאמץ הכולל של הרשויות להגברת הבטיחות בדרכים במדינה והשיפור בבטיחות הכלי רכב. אולם, כידוע, מספר התאונות וההרוגים שנצפו בשנה מסוימת אינו תואם תמיד את המאמץ שהושקע בבטיחות בשנה זו. זאת, מכיוון שמעבר למאמץ ממוקד של שיפור הבטיחות, על תופעת תאונות הדרכים משפיעים גם גורמים אחרים דוגמת: המצב הכלכלי, היקף נסועה, שינויים דמוגרפיים ועוד [גיטלמן וחובריה, 2004]. מכאן, שבכדי לנתח את נושא תאונות הדרכים, אין להסתפק במידע הנוגע לתאונות הדרכים בלבד. על מנת לקבל את הערך

המוסף של המידע, יש להשוותו לתקופות אחרות, להשוות בין אזורים ולקשרו למשתנים נוספים [Peled et-al, 1996].

הניסיון להסביר את הגורמים לתאונות דרכים, בייחוד ברמה הלאומית, מעסיק את החוקרים בתחום הבטיחות בדרכים כבר למעלה משלושים שנה. עם החמרת מצב הבטיחות בשנות השישים של המאה העשרים, והשינויים במדיניות התחבורה בעקבות משבר האנרגיה בתחילת שנות השבעים של המאה העשרים, שהובילו לירידה משמעותית במספר ההרוגים ברוב המדינות המפותחות, נולד הצורך בהבנת הסיבות לירידה זו. הצורך לבחון את מנגנוני ההשפעה על התאונות גבר. מודלים פשטניים לא הציגו תשובה מספקת. לכן, הצטרפו למאמץ, במהלך הזמן, מומחים מתחומי ידע מגוונים לפיתוח המודלים השונים [Hakim et-al, 1991].

2.1.2 התנהגות הנהג וקונספט הסכנה

בחלוקה לגורמי אנוש, רכב ודרך, מקובל לטעון כי גורמי אנוש מעורבים בכ- 95 אחוזים מהתאונות, גורמים הקשורים בדרך מעורבים בכ- 28 אחוזים מהתאונות וגורם הרכב ב- 8 אחוזים בלבד. קיימת כמובן חפיפה של גורמים, כך שבפחות מאחוז אחד מהתאונות ניתן למצוא את כל שלושת התחומים [הרשות הלאומית לבטיחות בדרכים, 2005].
Kim & Levine [1996], הגדירו יחס שונה ולפיו, בשבעים עד שמונים אחוזים מתאונות הדרכים מוגדר הגורם האנושי כגורם התאונה העיקרי. במחקרם הם מציינים כי נהוג להתייחס לגורם האנושי בשני מישורים: המישור ה"ביצועי" - מה מסוגל הנהג לעשות בהיבט הכישורים המוטוריים והתפיסתיים; והמישור "התנהגותי", קרי: מה מבצע לבסוף הנהג ברגע האמת בכביש. להגדרתם, מרבית המחקר, הניתוח והמידול בתחום הבטיחות בתחבורה, מתמקד במאפיינים, ביכולות ההתנהגותיות ובפעולות שננקטו על ידי הנהג הבודד. זאת, במקום לנסות ולצמצם את הפער בין התנהגות הנהג הבודד (רמת מיקרו) לבין ניתוח כוללני (מקרו) המבצע אגרגציה של מדוע, למה והיכן מרוכזות מרבית התאונות, וכיצד ניתן לצמצמן. כ- 15 שנה לאחר מחקרם של Kim & Levine [1996], ניתן לומר בוודאות גדולה, כי מגמה זו, שעיקרה בבחינת פעולות הנהג הבודד, השתנתה.

הדיון המנסה להסביר את השינויים החלים בתאונות דרכים, בהקשרם להתנהגות הנהג ולקונספט הסכנה, החל כבר בראשית שנות השבעים של המאה העשרים. ניסיונות מוקדמים להתמודד עם תופעת תאונות הדרכים, התמקדו בתחילה בשיפור ביצועי הנהג. זאת, תוך התעלמות מכך, שבקרב הנהגים קיימות נטיות ומניעים, אשר אינם עולים בקנה אחד עם נושא הבטיחות. כך לדוגמא, נטען כי קצב הנסיעה הנבחר על ידי הנהג, הינו קריטי יותר מאשר מקסום רמת המיומנות שלו [Summala, 1996].

Wilde [1982], הציג בתחילת שנות השמונים של המאה העשרים, את תיאוריית "יציבות הסיכון" (Risk Homeostatis). לפיה, התנהגות אנושית בה נלקחים סיכונים הינה תוצאה של ניסיון למצוא את האיזון בין סיכון רצוי/מבוקש לבין סכנה נתפסת/מובנת. משמע, אנשים נוטים להתאים את התנהגותם בתגובה לשינויים בעוצמת רמת הסיכון הנתפסת. כך, לפי תיאוריה זו, המדדים היחידים להם אפשרות השפעה קבועה על תאונות הדרכים, הינם אלו היכולים לגרום לשינוי גישה או תפיסת הנהג בנוגע לנטילת סיכונים [Hakim et-al, 1991]. גישה זו הינה פסימית, היות ולפיה לשיפורים בטיחותיים אין השפעה ארוכת טווח. גישה זו הניבה דיונים רבים בספרות, אולם לא נתמכה אמפירית [גיטלמן וחובריה, 2004].

גישה נוספת עושה שימוש בידיעה לפיה הנהג נוטה להגיב לשינויים במערכת התנועה (במידה ואלו מתרחשים ברכב, בסביבת הכביש או בהלך הרוח של הנהג עצמו), ואף מתאים את התנהגותו לתנאים ולתקנות לעיתים תוך שינוי ברמת הסיכון שהוא רגיל לקחת. עקרון זה של "הסתגלות התנהגותית", ידוע גם בתור תורת "פיצוי סיכון" (Risk Compensation) הינו נושא מרכזי בחקר התנהגות הנהג. מודל זה מבוסס על ההנחה, לפיה במצב דינאמי של נהיגה הנהג שולט בגבול הבטיחות. רק כשהסכנה והפחד חוצים גבול זה או מתקרבים בצורה משמעותית לגבול זה, מתרחשת השפעה על ההתנהגות. כך, לאחר התנסות במצבים חוזרים ונשנים, משתנה גם הגדרת הסף הסובייקטיבית של הנהג. במצב כזה הנהג עלול התייחס למצבים שבעבר היה מגדיר כמסוכנים, כאל מצבים שאינם מסוכנים [Summala, 1996 ; גיטלמן וחובריה, 2004].

לתיאוריה זו משקל רב, כאשר באים לדון בשילוב אלמנטים טכנולוגיים לשיפור הבטיחות (מצלמות, חיישנים וכיוצא בזאת). זאת, תחת ההנחה כי לרוב נוהג הנהג ברמת סיכון קבועה, כאשר אפשר, ובמידה וישולבו שיפורים טכנולוגיים, ישנה הנהג את תפיסת הסיכון שלו, יסתמך על הטכנולוגיה ויבחר בסיכונים גדולים יותר מאלה שהיה נוהג לקחת בעבר. דוגמה לכך, ניתן למצוא בשהתרחש בשנת 1967 בשבדיה. בשנה זו הוחלט לעבור לנסוע בצד ימין. בעקבות כך עלה החשש כי תתרחשנה תאונות רבות בעלות תוצאות בטיחותיות קשות מאוד. במציאות, מספר התאונות הקטלניות ירד בכעשרים אחוזים בשנה בה הוחל המעבר לצד ימין [פרוטוקול מספר 51, ישיבת ועדת המדע והטכנולוגיה, 2007].

<u>מורכבות המחקר האנושי</u> – במחקר בטיחות בדרכים נהוג לבחון תאונות ביחס לרמת החשיפה. כך לדוגמא, שיעורי תאונות מחושבים על בסיס מספר התאונות לנפש באוכלוסיה, מספר תאונות לכלי רכב רשומים, או תאונות לקילומטרים שנתיים לרכב (נסועה). למדדים אלה ניתן להתייחס כאל מדד סכנה לתאונות ומדדים יחסיים. ניתן לעשות שימוש במדדים אלה גם למטרות אחרות. כך לדוגמא, מדד תמותה לנפש, מהווה גם מדד עבור העלות הכוללת לחברה, בעוד ששיעור התמותה ביחס לנסועה, מתאר באופן כללי גם את המצב הבטיחותי של רשת הכבישים.

על אף חשיבותם של נתונים אלה, אין הם מספקים מענה על המכניזם הבסיסי של התנהגות הנהג. לצורך כך יש לפזר את המידע ולנתחו לפי גורם חשיפה ספציפי, דוגמת: סוג הכביש, רוחבו, נראות, זמן התרחשות התאונה במהלך היום, היום בשבוע בו התרחשה התאונה וכמובן קבוצת משנה של סוג הנהגים. עם זאת, הסקת מסקנות על המכניזם ההתנהגותי אינו כה פשוט. לדוגמא, במחקרים שנערכו בארה"ב בסוף שנות השמונים ותחילת שנות התשעים של המאה העשרים, נבדק הקשר בין גיל הנהג לתמותה. נמצא, כי בקרב נהגים מבוגרים קיים סיכוי הגדול פי חמש לתמותה ככל שהנסועה גבוהה יותר, מאשר בגילאי הביניים. עם זאת, בהערכות אלה התעלמו מהעובדה שנהגים מבוגרים נוטים שלא לנהוג עם רדת החשיכה, בשעות בהן דווקא צעירים וגילאי הביניים נוטים להיות מעורבים בתאונות חמורות. לפיכך, לנהגים מבוגרים, שיעור מעורבות גבוה בתאונות, על אף שאינם נוטים לנהוג כאשר קיימים תנאי דרך חריגים [Summala, 1996].

2.1.3 משתנים מסבירים

2.1.3.1 כללי

בשנות השבעים והשמונים של המאה העשרים, פותחו במדינות רבות מודלים מסבירים לתופעת תאונות הדרכים. מודלים אלה חיפשו קשרים סטטיסטים לתיאור המגמות במספר התאונות ובמספר הנפגעים וההרוגים בתאונות דרכים [Hakim et-al, 1991]. במודלים אלה ניתן להבחין במגוון רחב של גישות. מחד גיסא, ראיה גלובלית הבוחנת טווח ארוך (עשרות שנים) המתאימה נוסחה או מודל לעקומת ההרוגים (בהתבסס על משתני זמן ומספר מצומצם של משתנים כללים, דוגמת כמות כלי הרכב). מאידך גיסא, ניתן להבחין בגישה של "מידול" סדרות חודשיות של תאונות, במטרה לבודד גורם משפיע, דוגמת היקף ההסברה או כמות האכיפה.

בין "המשתנים המסבירים" במודלים שונים, ניתן לראות מדדים כללים של מערכת התחבורה, דוגמת: מספר כלי רכב מנועיים, מספר נהגים מורשים, אורך הכבישים והיקף נסועה. לעיתים, נמצא בין "המשתנים המסבירים" מדדים כלכליים, כגון: תל"ג, רמת אבטלה, או מדדים כלכליים תחבורתיים אחרים (צריכת דלק, השקעה בתשתית הכבישים ועוד). בנוסף, יש המשלבים במודל גם משתנים להם השפעה פוטנציאלית על רמת הבטיחות בדרך, דוגמת: כמות מכירת אלכוהול במדינה, שעות אכיפה בשטח או מספר דו"חות התנועה המופקים [גיטלמן וחובריה, 2004].

Page [2001], קיבץ במאמרו, בחלוקה לקבוצות, מגוון משתנים מסבירים, ששימשו בפיתוח מודלים שונים:

- <u>צי רכב</u> - צי רכב, פיזורו במרחב, חלוקה לקטגוריות שונות וגיל הרכב.

- נתיבים - אורך הכבישים במדינה, סוג הכביש.
- חשיפה - קילומטרים (נסועה), קילומטרים לסוג כביש, קילומטרים למשתמשי דרך שונים, קילומטרים לגיל ומגדר שונים, כמות תנועה בסוגי דרך שונים.
- אוכלוסיה - גיל ומגדר של האוכלוסייה, מחזיקי רישיונות נהיגה לפי גיל ומגדר, צפיפות, אחוז נהגים צעירים.
- מטאורולוגיה - טמפרטורה, רמות גשם ושלג, שמש, וקרח.
- בטיחות בדרכים - צריכת אלכוהול, שיעור חוגרי חגורות בטיחות, מהירות נסיעה, שיעור חובשי קסדות, שיעור כלי רכב המצוידים בכריות אוויר, הגבלות מהירות, מגבלת גיל לנהיגה, השקעות בבטיחות (בהתאם למדיניות בטיחות בדרכים).
- כלכלה - הכנסות למשק בית, צריכה למשק בית, שכר, תוצר לאומי גולמי, שיעור מובטלים, אוכלוסייה פעילה , ייצרו תעשייתי, מחירי גז, צריכת גז.
- כלכלה ובטיחות - השקעות בבטיחות הכבישים, הוצאות לאומיות בהנדסת הכבישים, השקעה לאומית בשיטור.
- אחרים - עלות התאונות, רמת חינוך, רמת פשיעה, שביתות, התאבדויות.

בניתוח להלן, יוצגו מספר משתנים נבחרים להם קשר אמפירי מוכח לתאונות דרכים.

2.1.3.2 נסועה

נמצא קשר ישיר בין עליה בנסועה[1] לבין הסיכוי למעורבות בתאונת דרכים. ככל שרכב יגמא קילומטרים רבים יותר, כך תעלה הסבירות למעורבותו בתאונה. בנוסף, הסיכוי למעורבות בתאונה ליחידת נסועה בקרב קבוצות סיכון מסוימות, דוגמת: נהגים צעירים או נהגים הנוסעים נסיעות שאינן קשורות לעבודה, גבוה יותר מאשר נהגים ותיקים וקבוצת המיוממים לעבודה[Hakim et-al, 1991]. עם זאת, Page [2001], מציין כי התלות בין מספר הרוגים לנסועה אינו ליניארי.

2.1.3.3 נפח תנועה

Winslott [2004] מציינת כי מחקרים העוסקים ביחס בין תדירות תאונות לנפח התנועה מצביעים על שונות גדולה בתוצאות. בחלק מהמחקרים לא אותר גידול מובהק בשיעור התאונות עם הגידול בזרם התנועה, כאשר לעיתים אף נמצא, כי כאשר מספר הכלי רכב עלה, שיעור התאונות דווקא ירד. בנוסף, היא מציינת שמחקרים אמפירים בודדים, עסקו בניתוח השפעות זרם התנועה בנפרד עבור כבישים שונים ובחלוקה לקבוצות מדגם

1 קִילוֹמֶטֶרז' - מספר הקילומטרים שנסעו בתקופה מסויימת

שונות. מעת פרסום מחקרה, ועד היום נראה כי מחקרים נוספים רבים שילבו ובחנו את פרמטר נפח התנועה, כאשר פרמטר זה נמצא ועולה כמובהק וכבעל משמעות גדולה במודלים השונים [Levine et-al,1995 ; Oppe , 1991]. כן, קיימת הסכמה בקרב החוקרים כי היחס בין מספר התאונות ונפח תנועה יומי, איננו ליניארי [Qin et-al , 2005].

2.1.3.4 גיל וותק הנהיגה

לגיל הנהג והוותק שלו בנהיגה, השפעה על מידת המעורבות בתאונות קטלניות וקשות. קבוצת הנהגים הצעירים נחשבת לקבוצה בעלת הסיכון הגבוה ביותר. שיעור המעורבות של נהגים בתאונות קטלניות וקשות הולך ופוחת ככל שגדל הוותק בנהיגה. נהגים בעלי וותק של יותר משנה ופחות משנתיים מעורבים בתאונות קטלניות וקשות בשיעור הגבוה כמעט ב- 60 אחוזים, מאשר נהגים בעלי ותק של עשר "שנות רישיון" ויותר-Hakim et al 1991 ; הרשות הלאומית לבטיחות בדרכים, 2005].

בנוסף, בקרב קבוצת הגיל הצעירה, שכיחות יותר עבירות תנועה המקושרות לתאונות הקטלניות. אלה, נובעות לרוב ממהירות מופרזת וסטייה מנתיב. עוד נמצא כי צעירים מעורבים בתאונות פי שלוש ממבוגרים [הרשות הלאומית לבטיחות בדרכים, 2005 ; Page, 2001].

תאונות בקרב נהגים צעירים, מקושרות להרגלי צריכת האלכוהול, כאשר בהתאמה לכך התבצעו במהלך השנים מחקרים רבים העוסקים בבחינת הקשר בין הגבלת הגיל המינימאלי לצריכת אלכוהול, לבין תאונות. מחקרים אלה מצביעים על כך שכ- 20 אחוזים מכלל התאונות הקשורות לצריכת אלכוהול בקרב צעירים, יכולות היו להימנע, במידה והגיל המינימאלי לצריכת אלכוהול היה גבוה יותר [Hakim et-al, 1991].

2.1.3.5 מהירות

בחינת נושא המהירות מורכבת ביותר. מחקרים רבים העוסקים במהירות, בחנו את משתנה המהירות ביחס למספר התאונות, ללא כל התייחסות למאפייני סביבת הכביש. גישה מעין זו עלולה להוביל לתוצאות מוטעות. בכדי לגשת לנושא באופן יעיל יש לדעת מהי סביבת הכביש, מהו סוג הכביש, אילו משתמשי דרך מצויים בו ומהי האינטראקציה הקיימת בין משתמשי הדרך השונים.

במהלך חמישים השנים האחרונות עברה סביבת התנועה שינויים בכל מדינות המערב. בהתאמה גם צפו ועלו מגוון בעיות בבטיחות בתנועה. בפרספקטיבה היסטורית, שינויים אלה השפיעו על מהלך מחקר תאונות הדרכים, היות ושיפור סטנדרט הכבישים הוביל

להגברת המהירות, בעוד שבמקביל שופרו המכוניות ובטיחותן וגובשו תקני בטיחות [Baruya, 1997].

למעט מחקרו של Solomon [1964], עד למשבר הנפט בראשית שנות ה-70 של המאה העשרים, לא בוצעו באופן מעמיק מחקרים הקשורים למהירות. זאת, למעט בשוודיה, שבה התקיימו מחקרים בעקבות המעבר מנהיגה בצד שמאל של הדרך לצד ימין. רק בסמיכות לסיום משבר הנפט, כאשר החל להיאסף מידע בנוגע לתוצאות הגבלת המהירות, החלו החוקרים לחפש הסבר רציונאלי לתצפית לפיה, כמות תאונות הדרכים הצטמצמה משמעותית עם הגבלת המהירות. מרבית המחקרים בנושא המהירות הניבו תוצאות חיוביות. משמע, נמצא כי קיים קשר בין המהירות לתאונות. יחד עם זאת, הערכת האפקט הייתה שונה ממדינה למדינה. חלק מהחוקרים הציעו כי השונות במהירות היא המשמעותית ולא המהירות הממוצעת. מספר חוקרים אף מצאו קשר שלילי.

מחקרים מוקדמים הבוחנים את הקשר בין מהירות לתאונות, התבצעו בעיקר על כבישים הממוקמים באזורים כפריים (לרבות אוטוסטראדות), בהן כל כלי רכב מהווה סיכון פוטנציאלי עבור משתמשי הדרך הסובבים אותו. בכבישים מסוג זה, מצאו מרבית המחקרים כי הפחתת המהירות הובילה להפחתה באירועי תאונות הדרכים. לחילופין, באזורים אורבאניים המצב מורכב יותר. זאת, היות וקיימים גורמים רבים, המשפיעים הן על מהירות התנועה ועל זרם התנועה. הולכי רגל, משאיות ואוטובוסים עלולים להאט את התנועה, ולהשפיע על מעבר חופשי של כלי רכב [Baruya, 1997].

Aarts & Van Schagen [2006], שסקרו מספר מחקרים העוסקים במהירות ותאונות דרכים, מדגישים אף הם, כי מהירות הנהיגה הינה מרכיב משמעותי בבטיחות בדרכים. המהירות לא רק שמשפיעה על חומרת התאונה, אלא גם מקושרת לסיכון להיות מעורב בתאונה מלכתחילה. מרבית המחקרים העוסקים בנושא התאונות והמהירות, בחנו את המהירות האבסולוטית ברמת רכב בודד או במקטע של כביש. בהתאמה נמצא, כי את היחס בין שיעור התאונות לבין המהירות ניתן לתאר על ידי פונקציה מעריכית (אסקפוננציאלית) או פונקציה ריבועית. נמצא כי שיעור התאונות עולה מהר יותר עם הגברת המהירות בכבישים שאינם ראשיים, לעומת שיעור התאונות המקושר למהירות בכבישים ראשיים. רוחב הנתיב, צפיפות הצמתים וזרם התנועה נמצאו כמשפיעים על היחס שבין מהירות ושיעור התאונות. מחקרים אחרים בחנו את פיזור ושונות המהירות, ומצאו כי גורם זה הינו משמעותי בגרימת תאונות דרכים. הבדלים גדולים ביחס המהירות בין מכוניות, נמצאו מקושרים לשיעור תאונות גבוה. רכב הנע במהירות

הגבוהה מאוד משאר הכלי רכב הסובבים אותו בדרך, נמצא כבעל סיכוי גבוה להיות מעורב בתאונה. התוצאות לגבי רכב הנע במהירות נמוכה באופן משמעותי מהסובבים אותו בדרך, אינן חד משמעיות.

בשנת 1991, נהרגו כ- 50,000 איש כתוצאה מתאונות דרכים במדינות האיחוד האירופי. מועצת הבטיחות בדרכים האירופאית (European Transport Safety Council) העריכה בראשית שנות ה- 90 של המאה העשרים, כי הפחתת המהירות הממוצעת בחמישה קמ"ש, יכולה הייתה לחסוך למעלה מ- 11,000 הרוגים וכ- 180,000 פצועים באירופה. יודגש, כי הוועדה המליצה למדינות החברות באיגוד, לנקוט צעדים נגד נהיגה במהירות מוגזמת ("excess"). משמע, נהיגה במהירות העולה על המהירות המותרת. בנוסף, הומלץ לנקוט צעדים נגד נהיגה שאינה הולמת ("inappropriate"), משמע נהיגה המהירה מדי ביחס לתנאי הדרך [Baruya, 1997].

2.1.3.6 תדירות תאונות והומוגניות קבוצות מדגם

Winslott [2004], פיתחה מודל לתיאור היחס בין מספר המכוניות לשעה לק"מ כביש, ותדירות תאונות הדרכים באזור הכפרי של שוודיה, בחלוקה למדגם הומוגני (סוג כלי רכב אחד) ושאינו הומוגני (המערב שני סוגי כלי רכב). המודל פותח תחת ההנחה כי נפח התנועה הינו הגורם היחידי המשפיע על מספר התאונות. במקרה הראשון נבחן, שטף הומוגני של מכוניות לשעה (q), המכיל רק מכוניות. במקרה השני נבחן זרם התנועה עבור מכוניות פרטיות (q_c) ועבור משאיות (q_l). תדירות התאונות (Z) חושבה, כמספר הצפוי של תאונות ליחידת זמן לקילומטר, כאשר תאונות שהתרחשו בצמתים הוסרו מהמדגם. המידע חולק לארבעה סוגים של כבישים בהתאם למהירות המותרת ולרוחב הכביש.

תוצאות מחקרה הראו כי שיעור התאונות יורד, כאשר בוחנים זרימת תנועה הומוגנית (משמע לא מערבבים סוגי כלי רכב). עם זאת, נמצא כי שיעור התאונות קבוע או עולה, כאשר בוחנים מדגם הומוגני של מכוניות, לעומת שיעור תאונות שדווקא יורד, כאשר בוחנים מדגם הומוגני של משאיות.

2.1.3.7 גורמים כלכליים

שיעור האבטלה הינו גורם שכיח לפיו נהוג לתאר ולאפיין מצב כלכלי. מאפיין זה נמצא כבעל קשר שלילי לתאונות עם נפגעים. עוד נמצא, כי לרמת ההכנסה השפעה על היכולת לממן את הנסיעה. כך, מושפעות בעיקרן נסיעות הקשורות לבילוי ופנאי, בניגוד לאלה אותן ניתן לקשור לנסיעות במסגרת עבודה. [Hakim et-al, 1991]. Peltzman [1975], מציע הסבר נוסף לתופעה. לגישתו, עליה ברמת ההכנסה מובילה לדרישה ויכולת

לרכישת מכוניות בטוחות יותר, ובהתאמה לכך, לירידה בשיעור תאונות עם נפגעים והרוגים. בנוסף, בעת צמיחה כלכלית קיימת השקעה גדולה יותר בתשתיות הכבישים. נציין כי, מאפיינים כלכליים רבים כגון: תל"ג, שיעורי ריבית, צריכה ממוצעת לאדם ועוד, יכולים להשתלב במודל. עם זאת, היות ורובם משקפים את אותה התופעה או לפחות את המגמה, לרוב מספיק לכלול משתנה אחד במודל [Hakim et-al, 1991].

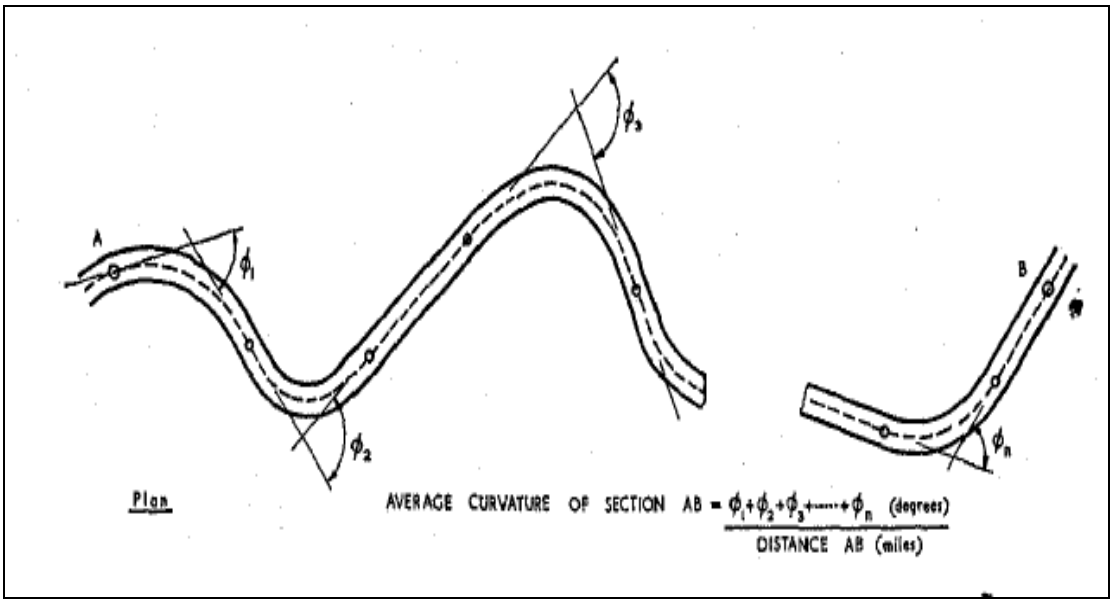
עוד מציינים Hakim et-al [1991], כי היחס בין מחירי הדלק לבין תאונות דרכים אינו ישיר. היות ומחיר הגז משפיע על צריכתו, צריך המחיר להופיע במשוואה המסבירה את הצורך בנסיעה. לפיכך, רכיב זה בא לידי ביטוי בשילוב משתנה ה"נסועה" כמשתנה במחקר.

במחקר אשר עסק בניתוח תאונות בהן היו מעורבים נהגים צעירים, נמצא קשר בין תאונות קטלניות לבין המצב הכלכלי של בית האב. בתקופות של מיתון כלכלי מספר התאונות בהן מעורבים נהגים צעירים (מתחת לגיל 24) נמוך בצורה משמעותית. זאת, על אף שלא נמצא שינוי משמעותי בנסועה הכללית. הסיבה לכך טמונה במקורות הכספיים העומדים לשם מימוש נסיעות פרטיות. החלטה זו מצמצמת את הנהיגה בכבישים שאינם עירוניים, במיוחד בשעות הלילה [Hakim et-al, 1991].

2.1.3.8 אלמנטים הנדסיים

Page [2001], מציין כי אחת הבעיות הקשות ביותר של המידול הסטטיסטי הינה החוסר במשתנים המקושרים ישירות לתשתית הכבישים (למעט ההשקעה הכספית בתשתיות). שיפורי בטיחות נכללים לרוב במודלים השונים ללא משתנה מפורש. זאת, בעוד שידוע כי לא ניתן להתעלם מתכנון ותחזוקת הכביש בבואנו לבחון את תחום תאונות הדרכים. כך לדוגמא, נמצא כי מעגלי תנועה מפחיתים באופן משמעותי את הסבירות לתאונה בצומת. [Page, 2001].

את החשיבות לתכנון הכביש בהיבט העקמומיות, ושילוב נתון זה במחקר, ניתן לראות כבר באמצע שנות ה- 60 של המאה העשרים, בפרסום של Road Research Laboratory [1965], שאיור ממנו, ניתן לראות באיור מספר 1.



איור מספר 1 - חישוב ממוצע עקמומיות משוקלל באורך הקטע [Road Research Laboratory ,1965]

הפרסומים של Miaou & Lum [1994] ו- Miaou [1993] היו בין הראשונים, בהם נבחן הקשר בין סוג כלי רכב מסויים לבין המאפיינים הפיסיים של הכביש. זאת, תוך שימוש במודלים ליניאריים ובמודלים מתקדמים (שאינם ליניאריים), דוגמת: בינומי שלילי, פואסוני ופואסוני Zero-Inflated. במחקרם ניסו לתאר את הקשר בין המשתנים הפיזיים של הכביש (אורך הסגמנט, עקמומיות, רוחב שול, שיפוע), משתנים תחבורתיים (נפח תנועה, מספר מיילים שעברה משאית) לבין מספר התאונות בהן מעורבות משאיות במקטע דרך מסויים. המחקר נועד לספק מענה לשאלות הבאות: בהינתן מידע בנוגע למאפיינים עבור מקטע דרך מהיר בינעירוני - לאיזה שיעור תאונות והסתברויות לתאונה ניתן לצפות; בהינתן סט משתנים עבור המאפיינים הגיאומטרים של הכביש, מהם המשתנים הרלוונטים בהיבט בטיחות הכביש ; לאיזה אחוז ירידה ניתן לצפות לאחר ביצוע שיפורי בטיחות נרחבים במקטע דרך.

מעניין לציין, כי במודלים הסטטיסטיים שבנו, נכללו גם השינויים הגיאומטריים שעברו מקטעי הדרך במהלך תקופת המחקר. זאת, על ידי כך שכל מקטע הדרך, שולב במודל חמש פעמים (אחת עבור כל שנה במחקר). מעניין לציין כי במחקרם, התגלה עבור המודלים הליניאריים שנבחנו, כי שילוב מקטעי דרך קצרים במיוחד במחקר (הקצרים מ-80 מ'), פגמה בדיוק התוצאות של המודל. לאחר שהוסרו אלה מבסיס הנתונים ונבנה מודל חדש, התקבלו תוצאות אמינות יותר. במודלים הפואסוניים התגלה כי לאורך המקטע אין השפעה בהיבט דיוק המודל. הפרמטרים שנמצאו כבעלי קשר חיובי לתאונות הדרכים היו: נפח תנועה, עקמומיות ושיפוע.

מחקר שנערך בשנים מאוחרות יותר, בוצע על ידי Aguero & Jovanis [2008]. במחקרם שולב פרמטר העקמומיות כמשתנה במחקר אם כי במאמר לא פורט כיצד חולץ פרמטר זה.

2.1.4 מודלים ומחקרים נבחרים

Smeed [1949], ביצע את אחד המחקרים המוקדמים בתחום הבטיחות בתחבורה ותאונות הדרכים. במחקרו מידל את הקשר בין מספר ההרוגים בתאונות דרכים (D), מספר כלי הרכב (N) וגודל האוכלוסייה (P), עבור 20 מדינות לאורך 16 שנה. בשלב הראשון של מחקרו הכין דיאגרמת פיזור (תוך שימוש ב LOG ליניארי) המראה את היחס בין מספר ההרוגים למספר כלי הרכב (D/N) , מספר ההרוגים לגודל האוכלוסיה (D/P) ומספר כלי הרכב לגודל האוכלוסייה (N/P). הגרפים הראו כי קיים קשר חזק בין שלושת המשתנים. בהמשך לכך, ניסה למצוא נוסחה לפיה ניתן יהיה לחזות את הקשר בין המשתנים במדויק. הקשר הוצע על ידי נוסחה מספר 1 :

$$[1] \quad D = AN^{\alpha}P^{\beta}$$

כאשר הקבועים A , α ו- β, חושבו על ידי שימוש בשיטת הריבועים המופחתים Least (squares), ונמצא כי הערכים 0.00030, 1.3 ו- 2.3 בהתאמה, מאפשרים לחזות את מספר ההרוגים על פי נוסחה מספר 1.

צדר וליבנה [1977], בחנו את הקשר בין צפיפות ושיעור התאונות לבין עוצמת התנועה היומית. במחקרם ניסו להבין את ההקשר בין שיעור תאונות הדרכים לבין מאפייני זרימת התנועה לאורך קטעי דרך המבודדים מצמתים. זאת, בכדי להעריך את התרומה הבטיחותית של שיפורים הנדסיים במקטעי דרך. כמו כן, ניסו לקשור בין סוג התאונה לבין תנאי זרימת תנועה שונים. מחקרם בחן את הנתונים בשתי רמות, רמת המיקרו ורמת המאקרו.

ברמת המאקרו נבחנו שני היבטים. ההיבט הראשון (מאקרו I), בו לא בוצעה הפרדה בין סוגי התאונות, כאשר קשרי הגומלין נחקרו בין נפח התנועה ובין סך כלל תאונות הדרכים. ההיבט השני (מאקרו II), ובו סוגי התאונות שוייכו לקבוצות, ונחקרו קשרי הגומלין בין נפח תנועה לתאונות מסוג: התנגשויות, רכב בודד ופגיעה בהולכי רגל.

ברמת המיקרו, נבחנו גם שני היבטים. ההיבט הראשון (מיקרו I), בו נבדקו קשרי הגומלין בין נפח תנועה ממוצע לשעה (q = כלי רכב/שעה) לבין שיעור תאונות דרכים מהסוגים: התנגשות, רכב בודד והתנגשות+רכב בודד. בהיבט השני (מיקרו II), נחקרו קשרי הגומלין בין נפח תנועה ממוצע לשעה (q) לבין מדד הסתברות לתאונה (המבוסס על שקלול מספר התאונות בהתאם לנפחי תנועה שונים) המתייחס לתאונות מסוג: התנגשות, רכב בודד והתנגשות ורכב בודד.

במחקרם נבחנו 32 מקטעי דרך בישראל, עבורם ידוע: ה ADT (נפח התנועה היומי) אורך הקטע, רוחב מיסעה[2], רוחב השולים ומספר הנתיבים. יודגש כי נבחרו למחקר דווקא מקטעי דרך "נקיים" מאלמנטיים גיאומטריים (שיפועים, עקמומיות), היות ובתפיסתם, גורמים אלה משפיעים על זרימת התנועה. נבחרו מקטעי דרך, בהם קיימם נפח תנועה מקסימאלי יומי לשעה, וכן כאלה בהם קיים פרופיל תנועה הקבוע יחסית במהלך מספר שנים. כמו כן, שולבו במחקר קטעי דרך באזורים שונים. היישום החשוב והעיקרי של מחקרם, הינו היכולת לעשות שימוש בשיטות שפיתחו, בכדי להעריך את התרומה הבטיחותית הנובעת משיפורים הנדסיים ואחרים.

Oppe [1991], פיתח מודל המקשר בין מספר ההרוגים בתאונות (F) לבין הנסועה (V). בבסיס המודל עומדות ההנחות לפיהן עליה או ירידה בנפח התנועה בפרק זמן t, הינה פונקציה של התפקוד ההיסטורי של המערכת טרם הרגע בה נבחנת המערכת, ובנוסף, שבטיחות בתחבורה משקפת את איכות מערכת התחבורה. לאור ההנחות שצויינן, לא ניתן לבחון את ההתפתחות בבטיחות התחבורה באופן עצמאי, אלא רק כחלק ממכלול ההתפתחות של מערכת התחבורה עצמה. S

מודלים תלויי זמן נפוצים לתיאור הקשר בין נפח התנועה (הבא לידי ביטוי על ידי משתנה הנסועה - V_t) לבין שיעור התמותה R_t. הקשר מתואר על ידי הנוסחה מספר 2 :

$$R_t = F_t / V_t \quad [2]$$

כאשר:

R_t - שיעור התמותה

F_t - מספר ההרוגים

V_t - הנסועה (סך הקילומטרים לשנה בכל מדינה)

לתפיסתו, ניתן לראות את מערכת התחבורה כמערכת יצרנית, בה מספר הקילומטרים לרכב הינו "יחידת הייצור", ושיעורי התמותה הינם ההסתברות המוערכת לכישלון עבור כל "יחידת ייצור". בהתאמה, הנסועה בשנה t (V_t), תחשב ל"תוצר המערכת", כאשר את ה"הפסד" הכולל על בטיחות, עבור כל שנה, ניתן לתאר בנוסחה מספר 3:

$$F_t = V_t(F_t / V_t) \quad [3]$$

שיפור הבטיחות של "תהליך היצור", משמעו הפחתה בהסתברות לתאונה בכל "יחידת ייצור". האופן הטוב ביותר לתאר תהליך יצרני, בו בתחילה קיימת דרישה גבוהה למוצר, ולאחר מכן איזון בביקוש, הוא על ידי מודל של פונקציה לוגיסטית[3]. כך, ניתן להגדיר את לוגריתם היחס בין V_t לבין $V_m - V_t$ (כאשר V_m - רמת "רווית" הנסועה או "הפוטנציאל המקסימאלי" של הנסועה)

[2] כביש - החלק עליו נוסעות המכוניות, כמו ש"מדרכה" היא החלק עליו דורכים הולכי הרגל

[3] הסבר לגבי הקשר שבין פונקציה ליניארית ללוגיסטית, ניתן למצוא בפרק 2.3.3.4

כפונקציה ליניארית של הזמן. דרך נוספת להגדיר את היחס, הינה לומר כי שיעור צמיחת התנועה, הינו פרופורציונאלי ליחס בין כמות התנועה הידועה לבין הפוטנציאל הנותר שלה, כמתואר בנוסחה מספר 4:

$$[4] \qquad \frac{d(S_t)}{dt} = a \cdot S_t$$

כאשר: $S_t = V_t / (V_m - V_t)$

ניתן לתאר את S_t ממשוואה [4] גם באופן הבא: $\ln(S_t) = at + b$, כאשר a ו - b הינם מקדמי המודל המותאמים עבור כל מדינה על סמך הניתוח הסטטיסטי של הנתונים שלה. עיבוד נוסחה זו (התואמת, כאמור, תפיסה של מודל ליניארי), כפונקציה לוגיסטית, מוצג בנוסחה מספר 5:

$$[5] \qquad V_t = \frac{V_m}{1 + e^{-(at+b)}}$$

כאשר:

V_t – הנסועה בשנה

V_m - "הפוטנציאל המקסימאלי" של הנסועה

חשוב לציין כי תיאור הנסועה בשנה תלוי בזמן, מאחר וניתן להניח כי מערכת התחבורה מתפתחת במהלך הזמן.

תנועה ממונעת קשורה לתופעות חברתיות, כאשר תהליך האימוץ וההסתגלות לסביבת הנסיעה אורך זמן. ניתן לתאר את תהליך ה"אימוץ וההסתגלות החברתי" בנוסחה מתמטית. זאת, אם מניחים, שבמידה ואדם חייב ללמוד להגיב לגירוי כלשהו בתגובה מסוימת (לדוגמא: אדם חייב להגיב למילה מסוימת המוצגת בפניו בביטויים מסוימים), מספר התשובות הנכונות במסלול למידה כלשהו, יהיה פרופורציונאלי לכמות התשובות שאינן נכונות. הנחה זו מתאימה למגוון מצבי למידה. עבור בעיית הבטיחות בדרכים יש לנסחה באופן הבא: שיפור הבטיחות על ידי כלל האנושות בזמן t, הינו יחסי למספר המצבים שעדין נותר לשפר. כך, ניתן לתאר את שיעור התמותה (R_t) בדומה לנוסחה מספר 4, כנוסחה מספר 6:

$$[6] \qquad \frac{d(R_t)}{dt} = \alpha \cdot R_t$$

את נוסחה מספר 6, ניתן לרשום גם באופן הבא $\ln(R_t) = \alpha t + \beta$, או בצורה מעריכית, כמוצג בנוסחה מספר 7:

$$[7] \qquad R_t = e^{\alpha t + \beta}$$

כאשר:

R_t – שיעור התמותה בזמן t
α , β - מקדמי המודל המותאמים עבור כל מדינה, על סמך הנתונים הסטטיסטיים הייחודיים עבורה.

Oppe [1991], הפעיל את נוסחאות [5] ו- [7] על נתוני תאונות הדרכים של שש מדינות: ארה"ב, בריטניה, גרמניה, יפן, ישראל והולנד. התוצאות הראו כי קיים יחס מיוחד, הניתן להשוואה, בין נגזרת נפח התנועה (המיוצגת כאמור על ידי הנסועה, בנוסחא 5) לבין מספר התמותות (נוסחא 7). מיחס זה ניתן היה להסיק כי שיעור התמותה, הינו פונקציה של נגזרת נפח התנועה ביחס לזמן. כך, נמצא כי F_t (ערך התמותה הכוללת או מספר ההרוגים) שווה לקבוע פעמים השורש הריבועי של נגזרת Vt (הנסועה) ביחס לזמן.

לאור שהוסק וצוין לעיל, בכדי לחזות את מספר ההרוגים חושבה נגזרת עבור נוסחה [5], ובהתאמה מוצגת נוסחה מספר [8], שחשיבותה ביכולתה לחזות את מספר ההרוגים (F_t = התמותה), ללא שימוש בנוסחה מספר [7].

$$E(F_t) = c \cdot d(V_t)/dt \qquad [8]$$

כאשר:

$E(F_t)$ - מספר ההרוגים הצפוי
c - קבוע הפונקציה

בעת הרצת נוסחה [8], התקבל ערך אחוז השונות המוסברת קרוב מאוד לערך השונות שחושב במודל המקורי.
בשלב הבא, היות ונמצא קשר בין מספר התמותות לבין השורש הריבועי של נגזרת הנסועה (V), התבצע ניתוח נוסף. זאת, בכדי להגדיר את הפרמטרים של הפונקציה באופן המיטבי. נמצא כי נוסחה [9] מאפשרת חיזוי אופטימאלי של מספר ההרוגים:

$$F = c \cdot (dV/dt)^q \qquad [9]$$

כאשר:

c - קבוע הפונקציה
q - מעריך
V - נסועה

טבלה מספר 2 מציגה את ערכי המעריך q בו נעשה שימוש, וערך הקבוע c, עבור כל מדינה שנבחנה.

טבלה 2 - ערכי המעריך וערכי הקבועים שנמצאו אופטימאליים עבור כל מדינה [Oppe, 1991].

Country	Exponent	Constant
USA	0.70	7.88
GBR	0.87	7.18
FRG	0.57	8.28
JAP	0.34	8.48
ISR	0.68	6.61
NET	0.64	7.09

בהתאם לטבלה מספר 1, חושב אחוז השונות המוסברת עבור כל מדינה. משמעות אחוז השונות המוסברת הינה, כי ניתן להסביר X אחוזים מתמותה במדינות הנבחנות, על ידי השינויים בנפח התנועה. נציין כי בששת המדינות שנבחנו, אחוז השונות המוסברת גבוהה – למעלה מ 77%.
חוזקו של המודל המוצע, ביכולתו לחזות באופן מיידי את ההשפעה של שינויים בנפח התנועה על התמותה. עם זאת, תיתכן תזוזה בזמן (קדימה או אחורה), בהתאמה לפרק הזמן ולמאמצים המושקעים בשיפורי הבטיחות. דוגמא לתזוזה שכזו נצפתה בהולנד, בה על פי המודל, ניתן להסביר 95% מהשונות בתמותה, על ידי גידול בכמות הכלי רכב בלבד. עם זאת, מגמה זו נצפתה כבעלת תזוזה בזמן של כ- 10 שנים [Oppe, 1991].

Page [2001], פיתח מודל סטטיסטי המשווה את רמת הבטיחות במדינות OECD[4], בין השנים 1980 ל- 1994. היות ושיעור מקרי המוות אינו מהווה מדד השוואתי מספק, הותאם מודל סטטיסטי המשלב בתוכו את ממוצעי המשתנים הבאים:

- אוכלוסיה (מיליוני תושבים), המבטאת את רמת החשיפה להיפגעות.
- מספר כלי רכב לנפש, מבטא את רמת מינוע.
- אחוז האוכלוסייה העירונית, מאפיין את ריכוז הנסיעות העירוניות, משמע הסתברות לתאונות ברמת חומרה פחותה.
- אחוז גילאים צעירים באוכלוסייה (בני 15 עד 24).
- אחוז אוכלוסיה פעילה ומועסקת, מאפיין את המצב הכלכלי.
- צריכת אלכוהול לנפש (בליטרים).
- אחוז האוטובוסים בצי הרכב.

במחקר זה, מדגיש Page [2001] כי ניתן היה לבחור גם במשתנים אחרים. במשתנים רבים, שסבר כי הינם משמעותיים: נסועה, שיעור אורך הכבישים המהירים מכלל הכבישים במדינה, אורך רשת הכבישים, אורך הכבישים המהירים, צפיפות אוכלוסייה, תל"ג, שיעור נשים עובדות

[4] Organization for Economic Co-operation and Development - המדינות החברות: גרמניה אוסטרליה, אוסטריה, בלגיה, קנדה, דנמרק, ספרד, ארה"ב, פינלנד, צרפת, יוון, אירלנד, איטליה, יפן, לוקסמבורג, נורווגיה, הולנד, פורטוגל, בריטניה, שוודיה ושוויץ (כאשר רק 21 מדינות מתוך ה - 25 שולבו במודל מסיבות שונות).

וצריכת דלק, לא ניתן היה לעשות שימוש בגין חוסר זמינות או בשל קולינאריות (Collinearity)[5] הנתונים. המודל הלוגריתמי שהוצע, מוצג בנוסחה מספר 10, כאשר בכדי להעריך את המקדמים, נעשה שימוש בתוכנת SAS:

$$[10] \quad \ln(F_{it}) = \beta_0 + \sum \beta_j \ln(X_{jit}) + v_i + e_i + \varepsilon_{it}$$

כאשר:

F – מספר ההרוגים בתאונות
i – אינדקס המדינה
t – אינדקס הזמן
Xj- משתנים "מסבירים"
β0, βj - מקדמי המודל.
$v_i + e_i + \varepsilon_{it}$ - שאריות שונות

יתרונו של המודל הלוגריתמי בכך, שעבור כל משתנה במודל, ידועה הגמישות, היות ושווה למקדם βj. טבלה מספר 3 מציגה את תוצאות מחקרו:

טבלה 3 - תוצאות המודל של Page [2001]

	Estimate of β_j
Intercept	−5.53
Population	+0.96
Percent youngsters	+0.83
Vehicle fleet per capita	+0.28
Percent buses and coaches in fleet	−0.16
Percent urban population	−0.41
Alcohol consumption	+0.39
Percent of employed population	+0.39

היות וכאמור, הגמישות שווה למקדם βj, כך עבור משתנה גודל האוכלוסייה (המאופיין בגמישות של 0.96). על כל גידול של 10% באוכלוסייה, תגדל התמותה ב 9.6%. ה- Intercept, מייצג את המוצע המשתנים החסרים (שלא שולבו) במודל הרגרסיה. כך, ניתן להשוות ולדרג את המדינות הנבחנות ביחס לרמת הקדמה בבטיחות תחבורתית.

בינסטוק וחובריו [1998] עסקו בבחינת הקשר שבין אכיפה (כפי שבאה לידי ביטוי במספר דו"חות התנועה) לבין תאונות הדרכים בישראל. המחקר התבצע בשני שלבים. הראשון, שלב מאקרו - ניסיון ללמוד את הקשר, על ידי ניתוח סדרות נתונים שנתיות של מספר תאונות הדרכים בישראל, אכיפה ומשתנים מצרפים נוספים. זאת, על בסיס נתונים שנאסף בין השנים 1960 ל-1995 (ללא ייחוס למקטע כביש ספציפי, היות וייתכן כי השפעות "נגררות" ממקטע אחד

[5] סדרות שנתיות של משתנים דמוגרפיים, כלכליים ותחבורתיים ככלל מדגימים קולינאריות גבוהה, דהיינו ערכי המשתנים עולים או יורדים במקביל, לאורך זמן. לדוגמא, עם השנים נצפית עליה בתוצר לאומי גולמי, הכנסה פרטית, נסועה שנתית, מספר כלי רכב רשומים ועוד. כתוצאה, עשוי להיות קושי בבידוד האפקט של אחד מהמשתנים על התאונות. כך לדוגמא, יהיה קשה להבדיל את אפקט השפעת השינוי הכלכלי על התאונות מהשפעתו על נסועה (גיטלמן וחובריה, 2004).

למשנהו). השני, בחינת הקשרים על בסיס נתונים חודשיים של תאונות ודו"חות שהתקבלו בשנים שבין 1993 ל- 1995, בקטעי כביש בין עירוניים. ברמת המיקרו, ניסו החוקרים לבחון האם קיימת תלות בין המגמה השלילית במספר התאונות לנסועה על פני הזמן בישראל (ירידה במהלך השנים 1960 -1995) לבין האכיפה המשטרתית.

היות ובשיעור האכיפה ובשיעור התאונות קיימת מגמה שלילית לאורך הזמן, הקפידו להימנע מ"רגרסיה מסולפת" (Spurious Regression)[6]. מאחר ובמחקר נבחן שיעור התאונות במספר מדינות בעולם, כאשר בכל המדינות מגמה שלילית דומה, חשבו כי לא סביר יהיה להניח, שהמגמה המשותפת הינה פרי יוצא של שיקולים מקומיים המשתנים ממדינה למדינה (דוגמת אכיפה). לפיכך, שיערו כי המכנה המשותף שהוביל למגמת הירידה, נעוץ בהתפתחות הטכנולוגית, הן מבחינת כלי הרכב והן מבחינת מבנה הכבישים. במחקרם, התאימו מודל מאקרו המסביר את המגמה בשיעור התאונות בישראל על ידי שני משתנים בלבד: שיעור תאונות בחו"ל ואחוז כלי רכב מיושנים (מעל 10 שנים). שיעור התאונות בחו"ל משקף ומגלם בתוכו את השיפורים הטכנולוגיים, בעוד המשתנה השני, מבטא את "פיגור" ישראל ביחס לחזית הטכנולוגית הבינלאומית. הניסיון לצרף למודל את משתנה שיעור האכיפה, כשל. המשמעות היא כי מדיניות האכיפה לא הסבירה באופן מהותי ומשמעותי את המגמה השלילית בשיעור תאונות הדרכים בישראל.

יצוין, כי החוקרים בנו מודל מאקרו נוסף בו לא שולב משתנה שיעור התאונות בחו"ל. אזי נמצא, כי שעור האכיפה הינו בעל השפעה שלילית על שיעור התאונות. במודל זה שולבו מלבד ממשתנה "שיעור האכיפה" שני משתנים נוספים: גיל ציי הרכב, וההשקעה ביחידת כביש (מלאי הון כביש חושב בבנק ישראל, לפי הנוסחה של השקעה צבורה בכבישים חלקי אורך הכבישים בישראל). המשתנה האחרון, מבטא את איכות הכביש. על פי תוצאות מודל המאקרו, עליה בשיעור האכיפה באחוז אחד, צפויה לצמצם את שיעור התאונות ב 0.28 אחוזים. מטעמים סטטיסטים שונים, מציינים החוקרים כי מודל המקרו הראשון שתואר הוא המיטבי לתיאור הקשר בין תאונות לאכיפה. ולפיו, הירידה המתמשכת בשיעור התאונות אינה משקפת את מדיניות האכיפה.

ברמת המיקרו החוקרים ניסו לבחון מספר היבטים. ההיבט הראשון: האם קיים קשר בין כמות הדו"חות בקטע כביש בחודש נתון לבין מספר התאונות באותו הקטע באותו החודש. ההיבט שני: האם מספר הדו"חות בחודש יכול להשפיע על מספר התאונות באותו קטע כביש בחודשים הבאים (מעבר להשפעה באותו החודש). היבט שלישי: האם מספר הדו"חות הכולל יכול להשפיע על מספר התאונות בקטע כביש נתון (מעבר להשפעה של מספר הדו"חות באותו קטע

[6] רגרסיה הנוטה לדחות/לבטל קשר אמיתי (TYPE I Spurious REGRESSION) או לדחות/לבטל קשר שאינו אמיתי (TYPE II Spurious REGRESSION) [Chiarella & Shenhuai ,2002].

כביש). היבט רביעי: האם סוג הדו"ח משפיע על מספר התאונות. ההיבט האחרון: האם מספר הדו"חות משפיע בצורה שונה על תאונות מדרגת חומרה שונה.

לבניית מודל המיקרו מוזגו נתונים משלושה מקורות: קובץ תאונות הדרכים של משטרת ישראל, קובץ הדוחות של משטרת ישראל וקובץ נפחי תנועה של הלמ"ס. הניתוח התבצע על 135 מקטעי כביש, כאשר לכל מקטע נבנה קובץ חודשי של: דו"חות, תאונות לפי חומרה, ונפחי תנועה ממוצעים לאורך הקטע.

ניתוח היחס בין מספר התאונות לאכיפה במרבית המקטעים הנבחנים הראה כי לא ניתן לאתר קשר מובהק.

<u>שיטת המחקר</u> – בניית מודל רגרסיה, המביא בחשבון השפעות של גורמים לתאונות: צפיפות, אכיפה (כפי שבאה לידי ביטוי בדו"חות) ומגמות זמן כלליות. כל קטע אופיין עבור כל חודש במודל המוצג בנוסחה מספר 11:

$$[11]\quad Y_{it} = \alpha_y + \beta_m + \gamma_i + \delta_0 \cdot N_{it} + \delta_1 \cdot N_{it-1} + \delta_2 \cdot N_{it-2} + \theta \cdot A_{it} + \mu \cdot Y_{it-1} + U_{it}$$

כאשר:

Y_{it} - מספר תאונות בקטע (או בכביש) i בחודש t

α_y - אפקט קבוע לשנה, β_m - אפקט קבוע לחודש. אפקטים אלו אומדים מגמות זמן כלליות המשקפות עונתיות, שינויים ברמת הבטיחות של כלי הרכב במהלך הזמן, משתני מדיניות שאינם קשורים לאכיפה ועוד.

γ_i - אפקט קבוע לכביש או קטע כביש i. אפקט זה אומד את ההשפעה של גורמים נוספים בלתי נצפים (תכונות גיאומטריות של הכבשים, או אחרים שאינם ידועים), המשפיעים על רמת התאונות בקטע, מעבר להשפעה של נפח תנועה, דו"חות ומגמות הזמן.

N_{it} - מספר דו"חות שנרשמו בקטע (או בכביש) i בחודש t

N_{it-1} - מספר דו"חות שנרשמו בקטע (או בכביש) i בחודש הקודם (t-1)

N_{it-2} - מספר דו"חות שנרשמו בקטע (או בכביש) i בחודשים הקודמים (t-2)

A_{it} - נפח התנועה בקטע (או בכביש) i בחודש t

Y_{it-1} - מספר התאונות בקטע (או כביש) i בחודש הקודם (t-1) – משתנה זה מתאר קשרים דינאמיים בין תאונות דרכים על פני הזמן, האם ריבוי תאונות בתקופה הקודמת מתואם עם התאונות בתקופה הנוכחית.

U_{it}- סטייה מקרית בקטע (או בכביש) i בחודש t

אומדני המקדמים $\delta_0, \delta_1, \delta_2$, הינם אומדני השפעת הדו"חות (בניכוי השפעות של הגורמים האחרים, כולל מגמות זמן). בחינת הקשר בין מספר הדו"חות לתאונות, התבצע באמצעות בדיקת השערות לגבי המובהקות הסטטיסטית של אומדנים אלו.

המודל הורץ מספר פעמים, עבור כלל התאונות ועבור תאונות בחלוקה לסוגים (קטלניות, קשות או קלות). עיקרי ממצאי המחקר:

- נמצאו רק קשרים שאינם ליניאריים, המתארים את הקשר בין מספר דו"חות לכמות התאונות. נמצא כי השפעת הדו"חות על התאונות הולכת וגדלה ככל שמספר הדו"חות גדל. משמעות ממצא זה ברמה המעשית-אופרטיבית היא כי רק אכיפה מאסיבית תוביל לשינויים משמעותיים.
- לא נמצא סוג מסוים של דו"חות המשפיע על מספר התאונות, יותר מסוגים אחרים.
- **נמצא כי קיים אפקט קבוע מובהק למקטע כביש המשפיע על רמת התאונה בקטע. אפקט זה משקף את ההבדלים בין הקטעים מבחינת התכונות הגיאומטריות, עומסי התנועה ועוד.**
- נמצא כי בחודשים ינואר, מרץ, יוני, יולי ואוגוסט, מספר התאונות גבוה יותר מזה שבשאר החודשים.
- בכל המודלים נמצאה השפעה חיובית ומובהקת של מספר התאונות בחודש הקודם על מספר התאונות בחודש הנוכחי. לפיכך, ניתן להסיק כי קיימת דינאמיקה של תאונות על פני הזמן.
- כמות הדו"חות בחודש הנוכחי אינה משפיעה על כמות התאונות בחודש ובחודשיים שלאחר מכן.
- לא נמצא קשר בין מספר הדו"חות לבין מספר התאונות הקטלניות.
- הגדלת סך האכיפה בכלל המקטעים, מקטינה את הסיכוי לתאונה במקטע בודד, גם אם האכיפה בקטע הנבדק לא גדלה באופן ספציפי.

2.1.5 איכות נתוני תאונות הדרכים והקושי בניתוחם

2.1.5.1 כללי

הקושי באסוף נתוני תאונות דרכים ובהשוואתם, הועלה כבר בראשית שנות החמישים של המאה העשרים על ידי Smeed [1949]. במחקרו ציין Smeed, כי במדינות שונות קיים קושי השוואתי בין נתונים, הן בגין שינויים בטרמינולוגיה והן בשל אופן איסוף הנתונים. כך לדוגמא, מדינות שונות אינן נוטות לקטלג ולכלול תאונות בין כלי רכב לרכבות כתאונות דרכים, בעוד שמדינות אחרות אינן מדווחת על תאונות של "ילידים" (ראה מקרה ניו זילנד והמאורים).

בישראל, הנתונים המדווחים בפרסומי הלשכה המרכזית לסטטיסטיקה (להלן: הלמ"ס) לניתוח, הינם נתוני תיקי תאונות דרכים עם נפגעים (תיקי ת"ד), כפי שמועברים ללמ"ס על ידי משטרת ישראל. קיימים מקורות מידע נוספים על תאונות דרכים: חברות הביטוח, אשפוזים בבתי חולים ופינויים ע"י מד"א. אולם, הקשיים בשילוב מקורות מידע אלו, הנובעים ממגבלות חיסיון המידע הרפואי, עבור דרך אי תאימות בהגדרות הבסיס של בסיס הנתונים, מקשים על בניית מאגר מידע אחיד [הרשות הלאומית לבטיחות בדרכים, 2005].

בתיקי ת"ד של משטרת ישראל אין מידע מקודד על מכלול הגורמים והנסיבות אשר הביאו לתאונה. אלה, מפורטים לעיתים בחלק המילולי המתאר את התאונה, כפי שמצא לנכון למלא בוחן התאונה. בנוסף, היות והמידע נאסף על ידי משטרת ישראל למטרות שאינן מחקר תופעת תאונות הדרכים, נוצר מצב בו המידע אינו שלם ואינו מייצג סטטיסטית. מצב זה מקשה על הסקת מסקנות נכונות מכלל נתוני תאונות הדרכים. כך לדוגמא, מספר תאונות הדרכים עם נפגעים המדווחות למשטרת ישראל הגיע בשנת 2005 ל-100 אלף, מהן כ-84 אלף תאונות עם נפגעים. המשטרה פתחה תיקי ת"ד רק ב-17 אלף מהן. בכל המקרים שבהם לא נפתח תיק ת"ד, נכללו נתוני התאונה, הנפגעים וכלי הרכב המעורבים בהן, בקובץ "כללי עם נפגעים" או בקובץ "נזק בלבד". רק החל משנת 2004 מעבירה המשטרה גם את קבצי "כללי עם נפגעים" ללמ"ס, אך הנתונים הנכללים בקובץ זה הם חלקיים ובלתי מדויקים, ולפיכך אינם ניתנים לשילוב מלא עם נתוני תיקי ת"ד.

קושי נוסף הינו הגדרת "פצוע קשה". משטרת ישראל, ובעקבותיה הלמ"ס מגדירה פצועים קשה כמי שאושפזו בבית חולים לפרק זמן העולה על 24 שעות, שלא למטרות השגחה. על פי נתוני משטרת ישראל מספר הפצועים קשה (כאמור - בתיקי התאונות בלבד) הוא 2,500 – 3,000 בשנה. לעומת זאת, על פי נתוני מערכת המידע של בתי החולים, מספר הפצועים שאושפזו בבתי חולים לפרק זמן העולה על 24 שעות קרוב יותר ל-6,000 בשנה.

היבט אחר של הגדרת פצוע קשה, הינו חומרת הפגיעה, כאשר תקופת האשפוז אינה מהווה בהכרח אינדיקציה לחומרת הפציעה. העוסקים ברפואה מגדירים חומרת פציעה ע"י מדד ISS – Injury Severity Score. מקובל לחלק את הסולם של ISS ל- 4 קבוצות חומרה: 1 – 8 פציעה קלה, 9 – 14 פציעה בינונית, 16 – 24 פציעה קשה ומעל 25 פציעה אנושה וקטלנית. על פי הגדרות רפואיות לפציעה קשה (מדד ISS גבוה מערך של 16), נפגעים בממוצע בשנה כ- 800 איש. עם זאת, חשוב לציין כי גם פציעות הנחשבות "קלות" ו"בינוניות" היות ואין בהן סיכון לחיי הנפגע, עלולות להיות קשות

מבחינת השפעתן על אורח חיי הנפגע ומשפחתו עקב תקופת אשפוז ושיקום ארוכה, או נכות קבועה [הרשות הלאומית לבטיחות בדרכים, 2005].

המצב שתואר לעיל נכון עבור מדינות רבות. מספר מדינות, ובהן הונלולו-הוואי, כדוגמא, החלו לטפל בפערים אלה. הונולולו, שהיתה אחת הערים החלוצות שעשו שימוש ביכולות מערכות המידע הגיאוגרפי למטרות ניתוח אירועי תאונות דרכים, החלה בראשית שנות התשעים של המאה ה- 20, לאסוף מידע על תאונות הדרכים מכלל הרשויות והוקם מאגר מידע מקיף, המבוסס על דיווחי משטרה, אמבולנסים, בתי חולים, חברות הביטוח, והמשלב מידע דמוגרפי מקבצי רישומי הרכב ורישיון הנהיגה ומהלשכה המרכזית לסטטיסטיקה האמריקאית. בנוסף, שולבו נתונים מרחביים [Kim & Levine, 1996].

2.1.5.2 הקושי הסטטיסטי בניתוח תאונות דרכים

תאונה היא אירוע אקראי. עקב כך, ניתוח התופעה מורכב ביותר. מורכבות זו מתוארת במחקרו של בר גרא [2007], שעסק בבניית מודל סטיסטי של תאונות דרכים בצמתים בינעירוניים. בר גרא מתאר את האתגר המרכזי בניתוח תאונות דרכים, ומסביר את הקושי בחישוב רמת סיכון לתאונה בצומת מסויים על ידי ניסוי שהוא מכנה "ניסוי הכובעים", שיפורט מיד.

נניח כי נתונים מספר כובעים ופתקים שעל חלק קטן מהם רשום "תאונה" ועל האחרים רשום "אין תאונה". בשלב הראשון, מארגן הניסוי מחלק את הפתקים בין הכובעים בשיטה אקראית כלשהיא, כך שיתכן כי בכל כובע יהיו בסופו של דבר מספר שונה של פתקים עליהם רשום "תאונה", ואפילו מספר שונה של פתקים בסה"כ. מארגן הניסוי אינו מאפשר לנו לדעת כמה פתקים יש בכל כובע, וכמובן שלא על כמה מהם רשום "תאונה". בכל יום אנו בוחרים באופן אקראי פתק אחד מכל כובע, פותחים אותו, רושמים האם הוא מציין "תאונה" או לא, ומחזירים בחזרה לכובע. לאחר מספר שנים אנו בודקים את מספר פתקי ה"תאונה" שהתקבלו בכל כובע. בניסוי זה מספר התאונות עבור כובע אחד, מפולג בינומית (או בקירוב פואסונית), אבל רמת הסיכון, כלומר הסיכוי לקבל פתק "תאונה", שונה מכובע לכובע.

במחקרו מדגיש הילל, כי רמת סיכון של כובע מסויים היא תוחלת[7] מספר התאונות לאותו הכובע במשך תקופת הניסוי. כלומר, הסיכוי לקבל פתק "תאונה" ביום אחד,

---

[7] **תוחלת** = ממוצע ערכי המשתנה משוקללים בהתאם להסתברות שלהם. לדוגמא, בעת זריקת קוביה, ניתן בכל זריקה לקבל אחת מתוך שש ספרות או שש תוצאות אפשריות (1,…,6). לפיכך התוחלת של X (=תוצאת הטלת הקוביה) תחושב לפי הנוסחה הבאה:

$$EX = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

המשמעות התוצאה 3.5 הינה: במידה ונזרוק מספר רב של פעמים את הקוביה, התוצאה הממוצעת תהיה 3.5 בקירוב [לויתן ורביב, 1997]

מוכפל בימי הניסוי. הדרך היחידה לדעת מהי רמת הסיכון של כובע מסויים, היא לבדוק כמה פתקים יש בכובע וכמה מתוכם הם פתקי "תאונה". במסגרת הניסוי שתואר, זהו מידע שאין באפשרותנו לקבל, לא לפני הניסוי ולא בתום הניסוי. כיוון שכך אין דרך לדעת מהי רמת הסיכון האמיתית של כל כובע, גם לא בתום הניסוי. מספר ה"תאונות", שהתקבל בכובע מסויים, בהיותו משתנה מקרי, אינו מייצג בצורה מדוייקת את "רמת הסיכון" של אותו הכובע.

מצב זה קורה בניתוח תאונות דרכים בצמתים. לכל צומת i רמת סיכון המאפיינת אותו μ_i, אשר מייצגת את תוחלת מספר התאונות באותו הצומת לאורך תקופת המחקר. לנו, אין שום דרך לגלות מהי אותה רמת הסיכון, אפילו לא לאחר שיש בידנו את כל המידע על התאונות שהתרחשו במשך תקופת המחקר. הסיבה היא כי מספר התאונות בצומת בפועל בצומת i לאורך תקופת המחקר הוא משתנה מקרי בהתפלגות פואסון עם תוחלת μ_i, ולכן מספר זה אינו משקף בהכרח את רמת הסיכון באותו הצומת.

כדי להמחיש את ההבדל בין "רמת הסיכון" של צומת לבין "מספר התאונות" באותו הצומת, מציג בן-גרא תאור של סימולציה, בה הוחלט לקבע את רמת הסיכון עבור צומת מסויים ל- 2.5 תאונות בשנה, ונקבע כי ערך זה ישאר קבוע במשך 10 שנים. באמצעות סימולצית מחשב נדגמו 10 ערכים מקריים מהתפלגות פואסון עם תוחלת של 2.5, כך שכל ערך מייצג את מספר התאונות בצומת בשנה מסויימת. מספר התאונות בשנה הרביעית והשישית היה 0, מספר התאונות בפועל בשנה הראשונה היה 1, בשנה הראשונה והשלישית היה 1; בשנים השניה החמישית והתשיעית היה 2, בשנה העשירית מספר התאונות היה 10 ; בשניה השביעית מספר התאונות היה 6 ; בשנה השמינית מספר התאונות היה 5. ברור כי ערכים אלה אינם מייצגים בצורה טובה את כלל רמת הסיכון עבור אותו הצומת, שנשארה 2.5 לאורך כל השנים. שינויים כאלה במספר תאונות דרכים משנה לשנה הם דבר צפוי כאשר לא נעשה שום שינוי בפועל.

בר-גרא מציין עוד, כי סימולציה זו ממחישה היטב גם את בעיית החזרה לממוצע. נניח כי בסוף השניה השמינית, לאחר שנתיים בעיתיות במיוחד (שש תאונות בשנה השביעית, וחמש תאונות בשנה השמינית), יתכנס צוות מומחים לדיון בשיפור הבטיחות בצומת. הצוות ימליץ על דרכים לשיפור הבטיחות וההמלצות ייושמו. אם בשנה התשיעית תחול ירידה במספר התאונות, המסקנה המתבקשת היא כי המלצות צוות המומחים הועילו לבטיחות. כפי שמראה הסימולציה, גם במידה ולא היה חל שום שינוי, וגם אם המלצות הצוות היו חסרות תועלת ולא שינו את רמת הסיכון, מספר התאונות בשנה התשיעית, שנדגם מהתפלגות פואסונית עם תוחלת של 2.5 צפוי להיות נמוך ממספר התאונות בשנה השמינית, שהיה גבוה באופן חריג. למעשה בכדי לדעת האם המלצות הצוות אכן

הועילו, צריך לבחון את השינוי ברמת הסיכון של הצומת (כאמור, מספר התאונות אינו מייצג את רמת הסיכון של הצומת).
בעיית החזרה אל הממוצע מחריפה, אם תהליך קבלת ההחלטות מבוסס באופן שיטתי על בחירת צמתים בהם קיים ריבוי של תאונות לצורך טיפול.

2.2 GIS ותחבורה

2.2.1 GIS-T

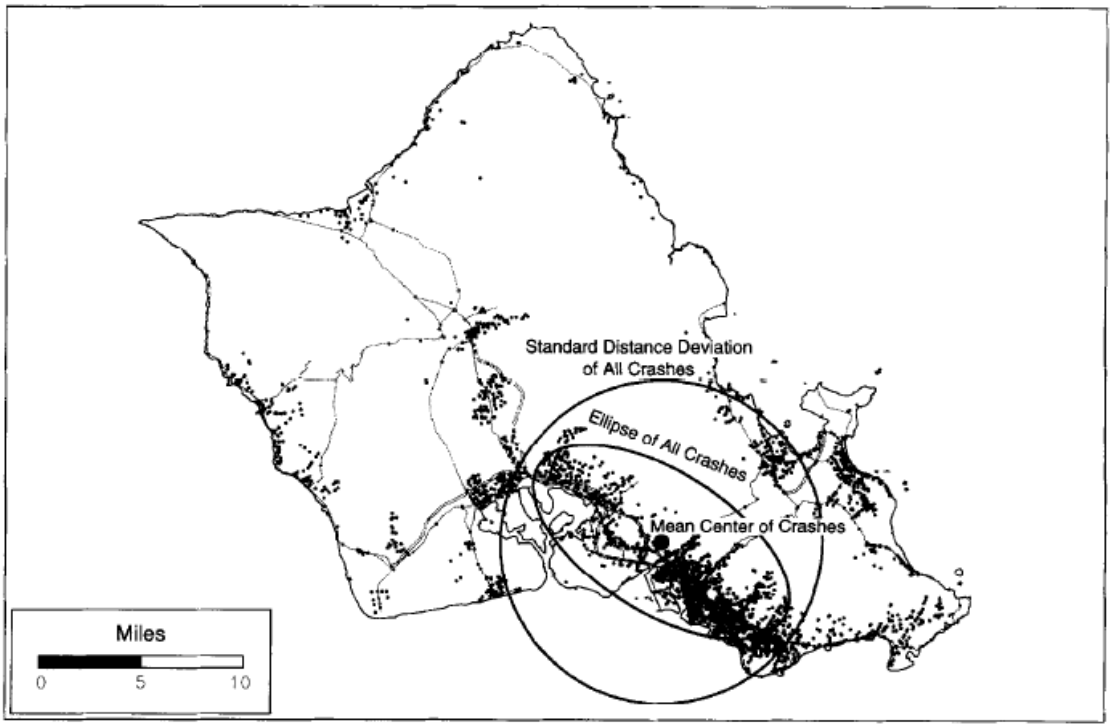
בראשית שנות התשעים של המאה העשרים הלך והתפתח שיח, שדן בטכנולוגיית מערכות המידע הגיאוגרפי והשימושים בטכנולוגיה זו למטרות חקר התחבורה (הן בדגש על תחום המחקר ההנדסי והן בתחום התכנוני). דומה היה כי לנושא הבטיחות בתחבורה (Traffic Safety) הקשרים פשוטים ולוגיים למערכות אלו [Kim & Levine, 1996].
עם זאת, בשלבים הראשונים שימשו מערכות המידע הגיאוגרפי בעיקר לזיהוי וקביעת מיקומים במרחב בהם קיימת סבירות גבוהה לתאונות דרכים. בין מחקרים ראשוניים אלה, ניתן לציין את עבודותיהם של Peled & Hakkert [1993], Peled et-al [1996] ו- Kim et-al [1995].
בנוסף, מחקרים רבים שעסקו בפיזור המרחבי של תופעות, התבססו בעיקר על בחינה ויזואלית. מתודות סטטיסטיות בהן נעשה שימוש למדידת אוטו-קורלציה בתבניות המרחביות, הניבו תוצאות המתארות את רמת האישכול של סט נקודות (המייצגות מאפיינים שונים באזור המחקר), אולם לא עסקו במרחב הגיאוגרפי ובקנה המידה בו מתרחשים האירועים [Jones et-al, 1996].
במהלך השנים גבר הצורך בניתוח סטטיסטי משלים ומעמיק לניתוח המרחבי. זאת, בכדי לקבוע את ההשפעות והגורמים השונים למיקומים המרחביים של תאונות הדרכים (דוגמת: מזג אוויר, מאפיינים סוציו-אקונומיים של הנהג, מאפיינים סביבתיים ואחרים) ושילובם בעת מימוש הניתוח במערכות המידע הגיאוגרפי [Kim & Levine, 1996]. הצורך בעבודות מסוג זה, במשולב עם שיפור היכולת הניתוחית הסטטיסטית בתוכנות מערכות המידע הגיאוגרפי הוביל למחקרים המשלבים ניתוחים סטטיסטיים וניתוחים מרחביים של תאונות דרכים. דוגמא לכך ניתן למצוא בעבודותיהם של: Jones et-al [1996], Ng et-al [2002], Schneider et-al [2004] ו Sabel et-al [2005].

לרוב, שילוב מידע תאונות דרכים במערכות מידע גיאוגרפי נתקל בקשיים רבים ואינו מיטבי. זאת, היות וכאמור, מידע בנוגע לתאונות נאסף באופן שונה על ידי רשויות שונות, המנהלות את המידע במגוון מערכות שאינו אחיד. קושי משמעותי ועיקרי נוסף הינו חוסר הדיוק בתיאור האירוע ובמיקומו במרחב. כך לעיתים, המיקום המרחבי של תאונות דרכים אינו מתואר ע"י מערכת קואורדינאטות קרטזית, אלה במונחים של: "מרחק מהצומת", "אבני דרך", "הק"מ ה..." או נקודות התייחסות אחרות, שמות רחובות וכיוצא בזאת. היעדר סטנדרטים גיאוגרפיים אחידים

לתיעוד המידע בעת דיווח על תאונות דרכים, נוסף על בעיות הקשורות למיפוי (דוגמת: מיפוי שאינו אחיד, המתבצע בקנה מידה שונים ובהיטלים שונים), מקשה על יצירת בסיס מידע עדכני ושלם, שיאפשר מחקר אמיתי ושלם של תאונות הדרכים [Kim & Levine, 1996].

אחד המחקרים הראשונים והחדשניים בתקופתו, בו נעשה ניתוח תאונות המשלב רכיב מרחבי, פורסם על ידי Levine et-al [1995] . במחקרם מציינים Levine et-al [1995] מפורשות כי בראייתם, מבחינת ההיבט המרחבי, תאונות דרכים אינן סתם אירועים אקראיים. בראייתם, קיימת דוגמא מרחבית המאפיינת את מיקומי התאונות,שהינה יותר מאשר "סתם" פיזור מרחבי חריג. במחקר נבחנו מספר כלים לתיאור הפיזור המרחבי של תאונות דרכים, כאשר שימוש בכלים האמורים, הראה כי ניתן לבצע אבחנה בין סוגים שונים של תאונות דרכים.
בראשית המחקר, עיסוק נרחב באופן איכון התאונות למקום התרחשותן במרחב. לאחר הצבת התאונות במרחב הגיאוגרפי, בוצעו מספר ניתוחים סטטיסטיים מרחביים (ראה איור מספר 2) :

- mean center - מרכז כובד התאונות
- Standard distance deviation - סטיית המרחק בין כל תאונה ("greay circle") לבין מרכז כובד התאונות.
- Standard deviational ellipse - מחשבת שתי סטיות שונות. האחת, לאורך ציר הריכוזים המקסימאלי של תאונות הדרכים, והשני בניצב לציר זה.
- השכן הקרוב ביותר - המודד את המרחק הממוצע מכל נקודה לנקודה הקרובה השכנה שלה ולאחר מכן משווה את התוצאה לפיזור הצפוי.



איור מספר 2 -ניתוחים גיאומרחביים של תאונות דרכים במחקרם של Levine et-al [1995]

הכלי שפיתחו מציג דוגמאות מרחביות אולם אינו מספק הסברים התנהגותיים, כאשר ביכולת התבניות המרחביות להצביע על הסברים אפשריים. כך לדוגמא, במחקר עלה כי תאונות רבות יותר מתרחשות בסמוך לאזורי תעסוקה, כאשר בנוסף, תאונות מקושרות לנפח התנועה (יותר תנועה = יותר תאונות). לאור האמור, ניתן לומר, כי תעסוקה יכולה להוות מנבא לריכוז תאונות. היחס המרחבי בין אוכלוסיה ואלכוהול מרמז כי השפעת האלכוהול באזורי אוכלוסיה, חמורה יותר מאשר באזורים בהם קיים ריכוז של מרכזי תעסוקה. כך, בפרוורים ובאזורים כפריים ניתן לקשור בין חומרת תאונות (מספר גבוה של הרוגים או פגועים קשה) לנהיגה בלילה תחת השפעת אלכוהול.

מאמר נוסף של אותם החוקרים (Levine et-al [1995] - חלק ב') ממשיך בפיתוח מודל סטטיסטי - אוטו קורלטיבי - מרחבי. במחקרם ניסו להסביר את הקשר בין אזורים שונים לבין אוכלוסיה, תעסוקה ומאפייני הכביש. המודל נבחן במתארים שונים: סוגי תאונות, שעות שונות בימי אמצע השבוע ובסופי השבוע. בהיבט מאפייני הכביש ששולבו במחקר נבחנו: נוכחות כביש מהיר החוצה צביר תאונות, אורכו של הכביש החוצה את צביר התאונות, אורך הכבישים מסדר נמוך וסדר גבוה החוצים את אזור מחקר הצבירים. בתוצאות המחקר מציינים החוקרים כי יתרונו של המודל שפיתחו הוא ביכולתו למקם את התאונות הצפויות בכל אזור המטרופולין ולא רק עבור מקטע כביש בודד או צומת.

Aguero & Jovanis [2008] מבקרים את מחקרם של Levine ומציינים, כי חולשתו בהנחה של חוקריו כי פיזור ספירות תאונות דרכים מתפלג באופן נורמאלי, ולא הניחו פיזור NB או פואסוני, המקובל כיום.

ניתוח מתקדם יותר בהיבט המידול המרחבי, בוצע על ידי Song et-al [2006]. שמצאו לנכון לציין, כי דווקא פיתוח מערכות המידע הגיאוגרפיות המתקדמות (GIS) דירבנה את הסטטיסטיקאים לפתח מודלים סטטיסטיים דוגמת מודלי ה- CAR (Conditional AutoRegressive), המספקים מענה גם לתחום המרחבי. הקושי במודלים הסטטיסטיים הנהוגים בספרות: הפואסוניות, בינומיות שלילית, לרבות ההתמודדות עם תופעת הרגרסיה אל הממוצע, המטופלת בשיטות של EMPIRICAL BAYES, ולרבות השימוש ברגרסיות לוגיסטיות, מתעלמות, לרוב, מהתלויות המרחביות המתקיימות בתאונות דרכים.

במחקרם התאימו את מודלי ה- CAR לניתוח רב משתני, למטרת הערכת שיעור תאונות דרכים מסוגים שונים. מידע התאונות חולק למספר קבוצות: תאונות המתרחשות בצומת, תאונות המתרחשות באזור ההשפעה של הצומת, ותאונות שאינן מתרחשות בצמתים. בנוסף שולבו שלושה תת משתנים נוספים (לגביהם אין במאמר פירוט כיצד חולצו): רטיבות, המייצגת את

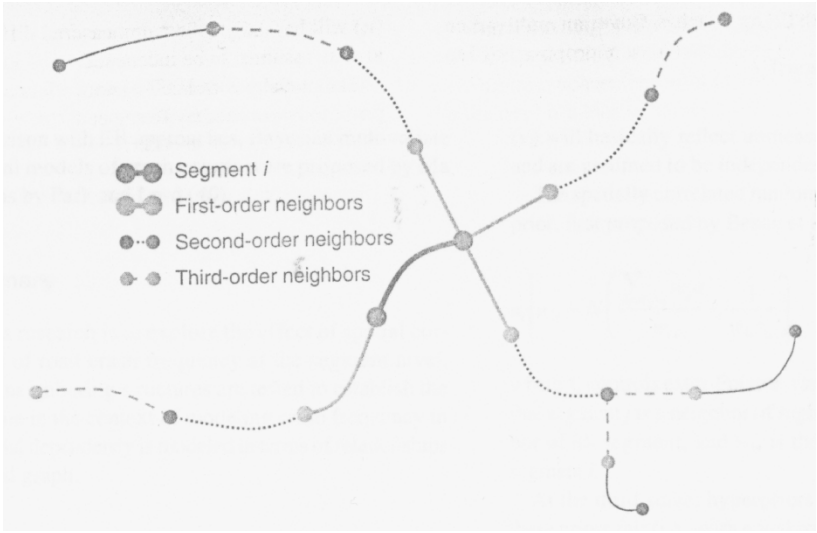
תנאי מזג האוויר ; שונות מרחבית באחוזים עבור עקומות אופקיות חדות ומשתנה המייצג את השונות המרחבית בתנאי צידי הדרך. בתוצאות המחקרם מוצגות מפות חיזוי מרחביות בחלוקה לסוגי התאונות השונים. יודגש כי המפות מציגות את תוצאות החיזוי בחלוקה לאזורים פוליגונאליים (ולא ליניאמנטיים). מחקרם העלה כי משתנה המייצג את תנאי מזג האוויר איננו תמיד מובהק בסוגי תאונות שונים ובעקמומיות שונה. מחקר הדומה באופייו בוצע על ידי [2003] Miaou et-al אלה, מדגישים במחקרם את היתרונות והצורך בהצגת הניתוחים המתקדמים של תאונות הדרכים באופן ויזואלי.

שיטה לניתוח מרחבי נוסף, מוצגת במחקרם של Aguero & Jovanis [2008]. אלה, עסקו בקורלציה מרחבית ברשתות ליניאריות. בראשית מחקרם מציינים כי במחקר קיימים מודלים סטטיסטיים העושים שימוש ברגרסיות שונות המנסות לקשור בין נתוני ספירות תנועה לנפח התנועה. אולם מדגישים כי ההתפתחות הנוכחית בתחום ה- HSM Highway Safety (Manual) מעוררת אף עניין גדול יותר במידול תאונות הדרכים. החוקרים שמו להם למטרה למדל גורמים מרחביים (דוגמת מאפייני הכביש) לרבות מידול הקורלציה בין גורמים אלה בקטעי דרך שונים. מבנים שונים של סגמנטים על שכניהם, נבחנו למטרת ביסוס מודל תדירות תאונות ברשתות ליניאריות.

ההנחה היא, כי ניתוח מקטע מסויים בו התרחשה תאונה, יושפע, לרוב, גם ממקטעים שכנים. לפיכך, השימוש בקורלציה המרחבית משפרת, לרוב, גם את טיב הערכת המודל. בנסיבות קיימת בהן קיימת השתנות אקראית רבה במידע, כמו במקרה של ניתוח תאונות דרכים, השימוש בקורלציה מרחבית, משמעותי אף יותר. בספרות הסטטיסטית מתיחסים לנושא זה, כבעיית: "הערכת האיזור הקטן" (small area estimation). ההנחה היא, כי כמות האפסים הרבה המאפיינת מקטעי דרך שונים (מספר תאונות במקטע שווה לאפס), במחקרים שונים, הינה תוצר של בחירת סקאלות זמן-מרחב שאינן מתאימות, חשיפה נמוכה או שתיהן. סיבה אפשרית נוספת לכך, הינה כמות הדיווחים בפועל על תאונות דרכים, דוגמת תאונות בהן התרחש נזק בלבד, שאינן מדווחות וכיוצא בזאת.

תלות מרחבית למעשה משלימה את מקומם של משתנים מסבירים, בלתי תלויים, שאינם ידועים, אינם מדידים, אבל בעלי רלוונטיות והשפעה, דוגמת: שימושי קרקע, אוכלוסיית הנהגים והשפעות מזג אוויר. בנוסף, הדיון בתלות המרחבית, מעלה מספר שאלות בעלותעניין: כמה קרובים צריכים להיות צומת או מקטע אחד לשני, על מנת שנוכל להחשיבם או להגדירם כקורלטיביים?, האם הקורלציה המרחבית שונה בכבישים מסדר שונה? מהי מערכת היחסים המרחבית בין סגמנטים הממוקמים בתוך אותו הצומת, האם סגמנטים שזוהו כבעלי קורלציה מרחבית משמעותית בינהם, ניתנים לקיבוץ למטרות ניתוח עתידי?

במחקרם Aguero & Jovanis [2008], שמו להם כמטרה לבחון השפעת יחסים מרחבים בין מקטעים עבור שכנים מסדר שונה על תאונות הדרכים. הבחינה התקיימה החל מהמבנה הפשוט ביותר (שכנים מסדר ראשון) ועד למבנה מורכב יותר, כמוצג באיור מספר 3:



איור מספר 3 - יחסים מרחביים בין מקטעי דרך
[מקור: Aguero & Jovanis, 2008]

שכנים מסדר ראשון הינם כלל המקטעים המחוברים לסגמנט באופן ישיר. שכנים מסדר שני, הינם אלה המחוברים ישירות לסגמנטים מסדר ראשון, וכן הלאה. הגדרת הסמיכות ההיררכית שתוארה, הינה טופולוגית, ולפיכך מתעלמת מהמרחק בין הסגמנטים. שכנים שונים ניתן לאפיין גם במשקלים שונים. ההיפותיזה מאחורי אנליזה מסוג זה, הינה: הכל קשור, אבל דברים "קרובים" קשורים יותר מאשר "מרוחקים", כאשר המשקל ניתן בהתאמה, קרוב מושקל באופן משמעותי הרבה יותר.

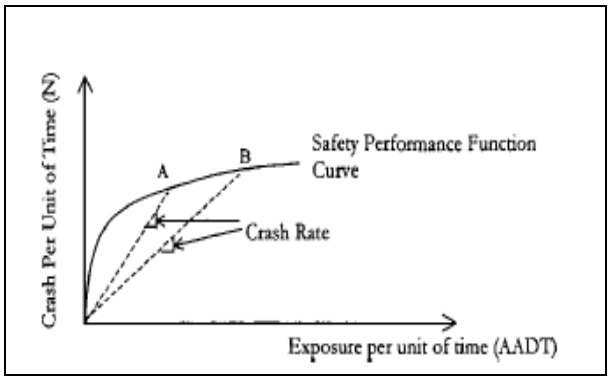
הפרמטרים ששולבו במחקר היו: קוד המייצג את המדינה, מספר כביש, נפח תנועה ממוצע יומי, סדר הכביש, רוחב שוליים, מהירות נסיעה מותרת, רוחב הנתיבים וקוד עבור אזור אורבני או כפרי ופרמטר עבור השכנים מסדר גודל כזה או אחר. אורך הסגמנט נורמל, ושולב כמשתנה OFFSET, קרי: תדירות התאונות הינה פרופורציונאליות לאורך הסגמנט. במחקרם עשו שימוש בשיטה הבייאסינית[8], היות ושיטה זו מיטבית עבור ניתוח מודלים מרחביים וקורלציות מרחביות ממספר היבטים. בראש ובראשונה, בגלל שילוב המידע הקודם, ההסטורי. שנית, השיטה לוקחת בחשבון את אי הוודאות המקושרת בהערכת פרמטרים, ומספקת הערכה מדוייקת של אי הוודאות עבור הפיזור האפוסטופריורי, של פרמטרים אלה.

מחקרם העלה כי המודלים בהם שולבו נתוני הקורלציה המרחבית, הראו התאמה טובה יותר למידע (לפי DIC[9]). מודלים בעלי קורלציה מרחבית מורכבת יותר, המכילה מקטעים מסדר שני ושלישי עלו כטובים יותר מאשר אלו, שמומשו עם בסיס נתונים חסר קורלציה מרחבית או

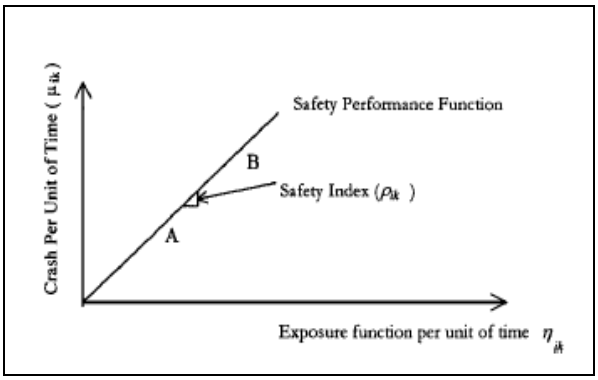
[8] מפורט בהרחבה בסעיף 2.3.3.4
[9] שהינו אקוולנטי לקריטריון AIC – כפי שמוסבר בהרחבה בפרק 7.2

שהקורלציה היתה רק מסדר ראשון. השימוש בקורלציה מרחבית, מקטין את ההטיה הנגרמת בשל שימוש בפרמטרים שמדידתם מוטה מלכתחילה או אינה מדוייקת עבור כלל תקופת המחקר, דוגמת AADT.

Qin et-al [2005] ביצעו מחקר בכבישים מהירים (בעלי שני נתיבים) באזורים כפריים, תוך שימוש בהערכה בייאסנית היררכית, ומתודת שרשרת מרקוב ומונטה קרלו (MCMC). במחקר מציינים כי הקשר בין מספר תאונות הדרכים לנפח התנועה אינו ליניארי. כתוצאה מכך, שיעור תאונות פשוט, המחושב כיחס בין מספר תאונות לנפח תנועה (N/AADT) , אינו מתאים כאשר מנסים להשוות בטיחות באתרים בעלי נפח תנועה שונה. על מנת לפתור את הבעיה, הציעו גישה חדשה לקישור בין נפח תנועה למספר תאונות דרכים. בבסיסה, יצירת פונקציית חשיפה חדשה, כאשר ההנחה היא כי פונקצית החשיפה היא אכן ליניארית ותותאם עבור סוגים שונים של תאונות. זאת, היות ועבור סוגי תאונות שונים קיימת רמת סיכון שונה. כך לדוגמא, עבור רכב בודד, הסכנה קשורה בעיקר להסתברות לפיה רכב ירד או יוסט מהכביש בשילוב האם קיימים אובייקטים המצויים בצידי הדרך. עבור מספר כלי רכב, ההסתברות לתאונה מותנית, בנוסף, בתכיפות שבה חולפים כלי רכב זה על פני זה וזה מול פני זה וכן הלאה.
איור מספר 4a מדגים את האמור לעיל, כאשר מוצגת בו פונקציית בטיחות שאינה ליניארית עבור קטעי כביש מהירים. עבור כל נקודה בפונקציה, שיעור התאונות (N/AADT ˙) מוגדר כשיפוע. היות ולא מדובר בקו ישר, שיעור התאונות נע בהתאם לזמן החשיפה. כך לדוגמא, מספר התאונות בנקודה B גבוה מזה שבנקודה A, אולם שיעור התאונות בנקודה B נמוך מזה שבנקודה A (השיפוע פחות תלול). מנקודת מבט של המהנדסים, השינוי בשיעור התאונות חל בגין השינוי בחשיפה, ולא כתוצאה משפור בבטיחות באתר זה. לאור האמור, עוזר לעיתים להגדיר פונקציית חשיפה חדשה, ולבצע טרנספורמציה לנתון "נפח התנועה" ההופך לפרמטר או מדד חשיפה $f(V)$, תוך שמירה על קשר ליניארי בין התאונות לבין אותו מדד חשיפה (איור 4). לאחר שקיים אותו השיפוע, שהינו אינדקס הבטיחות (המייצג שיעור תאונות מנורמל), ניתן להשוות בין ישויות שונות במקומות שונים, כמוצג באיור 4b.



a



b

איור מספר 4 - פונקציה חשיפה ופונקציית בטיחות שאינה לינארית [Qin et-al , 2005]

לפיכך, בסיס הנתונים במחקר חולק לארבע קבוצות: רכב בודד; מספר כלי רכב באותו הכיוון; מספר הכלי רכב בכיוון הנגדי; מספר כלי רכב בכיוונים שונים. במחקרם עשו שימוש בגישה בייאסינית היררכית ובמודלי רגרסיה מסוג Z-INFLATED פואסון, למטרת קבלת חיזוי מספר תאונות הדרכים כפונקצייה של הפרמטרים המרחביים הבאים: נפח תנועה יומי, אורך סגמנט, מהירות מותרת ורוחב שוליים, תוך שימוש במתודת MCMC.

במודל הסטטיסטי שולב נתון ה AADT , כאינדקס בטיחות, לאחר שבוצעה עליו הטרנספורמציה שצויינה לעיל.

נוסחת מודל החיזוי שנבנה, מוצגת כנוסחה מספר 12 :

$$\ln(\mu) = \text{int}\,ercept + \beta_\gamma D_\gamma + \beta_V \ln(V) + \beta_L \ln(L) + \beta_W W + \beta_S S \quad [12]$$

כאשר:

β – מקדם הרגרסיה

D_γ - וקטור המייצג את השפעת השנה

V – אינקדס הבטיחות (AADT לאחר שעבר טרנספורמציה)

L – אורך הסגמנט

W – רוחב השולים

S – הגבלת מהירות.

תוצאות מחקרם הראו כי היחס בין מספר התאונות ונפח תנועה יומי, אכן איננו ליניארי. לא נמצא שוני מובהק בין פונקציות החשיפה בארבע הערים שנבחנו. היחסים הלא ליניאריים בין תדירות התאונות לאורך הסגמנט מרמזים כי מהמודל הושמטו גורמים מסבירים חשובים, שכנראה היו קורלטיביים לאורך הסגמנט. פרמטרים כרוחב השוליים והמהירות המותרת, הראו שונות בהתאם למדינות השונות, כאשר חלק הראו יחסים חיוביים ביחס למספר התאונות וחלק הציגו מגמה הפוכה.

2.2.2 כריית מידע (Data Mining)

לאחרונה החלו גיאוגרפים לעשות שימוש בפוטנציאל הגלום בכריית מידע (data mining). כריית מידע מוגדרת כאיתור תבניות (patterns), שאינן טריוויאליות (אך בעלות פוטנציאל שימושי) מבסיסי מידע גדולים. כריית המידע הקלאסית משמשת לביצוע מחקרי גישוש, מידול מחקרי, מחקרי חיזוי, זיהוי תבניות וחוקים למטרת שליפת מידע. שילוב מידע מרחבי עושה שימוש באותן הטכניקות למטרת זיהוי מגמות ותבניות מרחביות. בהקשר שבין כריית המידע המרחבי לבין מערכות המידע הגיאוגרפי, כריית המידע נוטה להתעמק במחקר המרחבי, וניתן להגדירה כאוסף הטכניקות המאפשרות לתאר ולצפות במאפיינים מרחביים, לזהות מיקומים שאינם טיפוסיים או לאתר אירוע במדגם סטטיסטי המרוחק מהמקבץ העיקרי של האירועים באותו מדגם, לגלות תבניות והקשרים מרחביים, ולזהות מקבצי נתונים [Sabel et-al, 2005] .

Sabel et-al [2005] הציעו אף הם גישה לאיתור אוטומטי של מקבצים מרחביים סטטיסטים (תבניות מרחביות). במחקרם נעשה שימוש בשיטת Kernel Estimation (KE), לפיה אירועים המוגדרים כ"נקודות וקטוריות" עוברים טרנספורמציה למידע רסטרי.

פרמטר מוכר, שנמצא במחקרים רבים כבעל קשר הדוק לתאונות, הינו נפח התנועה. היכולת לקשור בין נתון "הכמות הממוצעת היומית של כלי רכב" לבין "כמות התאונות" במקטע הדרך, מאפשרת בניית מדד לנפח תנועה עבור תאונה ביום בו התרחשה. כך, על ידי פעולת קיבוץ (group by) של ערכי זרם התנועה (ערכים הקרובים ל- 100 במחקרם), ובמקביל, סכימת מספר התאונות שהתרחשו, נבנה גרף של תדירות תאונות "צפויה" עבור נפח תנועה כלשהו. מגמת "הסבירות לתאונה" נמצאה דומה עבור כבישים מהירים בהם כמות גבוהה של כלי רכב ועבור כבישים צדדיים בהם כמות נמוכה, יותר מאשר בכבישים עירוניים בהם זרם התנועה אינו קבוע.

בהנחה כי הסיכון לתאונה שווה לאורך כל מקטע הכביש (על אף כי במציאות קיימים גורמים נוספים המשפיעים על מספר תאונות הדרכים באזור מסוים), ניתן להשתמש ביחס שהוגדר, לסווג מחדש כל מקטע כביש, ולתת לו ערך המגדיר את רמת הסיכון שלו. היות והסיכון לתאונת דרכים אינו שווה בכלל מקטעי הדרך (66 אחוזים מכלל התאונות מתרחשות בצמתים), עולה הצורך לשלב במודל גם מידע לגבי צמתים. לפיכך, חושב המרחק של כל תאונה מהצומת הקרוב ביותר. נתון זה הושלם עבור כלל התאונות. כך, ניתן היה לאפיין ברמת סיכון שונה, כל מיקום לאורכו של כביש. שילוב נתון זה עם הנתון הקיים לגבי "זרימת המידע", איפשר להעריך את רמת הסיכון לכל מקטע כביש.

לאחר שאותרו התבניות המרחביות, בוצעה המתודולוגיה שהוצעה על ידי Kelsall & Diggel [1995] בנוגע לפונקצית הסיכון היחסי (Relative Risk Function). זו, עושה שימוש ביחס בין שני אומדנים של צפיפות, למטרת גיבוש מידע סינטטי חדש. לאחר מכן, שולבה סימולציית מונטה קרלו על התוצר החדש. כך, התאפשר לחלץ את המיקומים והאזורים המשמעותיים ביותר

מבחינה סטטיסטית. שילוב התוצרים של שתי השיטות איפשר למפות את האזורים בעלי הסיכון הרב ביותר לתאונות.

מחקר נוסף בתחום זה בוצע על ידי Geurts et-al [2005]. במחקרה ניסתה להבין מדוע תאונות דרכים נוטות להתקבץ (כצביר) דווקא במקטעי דרך מסויימים. החוקרים ניסו להבין את מאפייני התאונות הנוטות להתקבץ באזורים שחורים (להלן: Black Zones) בהשוואה לאלו המפוזרים לאורכו של הכביש. תוצאות מחקרם הראו כי תאונות הנוטות להתקבץ ב Black Zones, הופיעו בעיקר במצבים בהם התקיימו פניות שמאלה בצמתים שאינם מרומזרים, בעת התנגשויות עם הולכי רגל, איבוד שליטה עצמי על כלי הרכב ותנאי מזג אוויר. לעומת זאת, תאונות המתרחשות מחוץ ל Black Zones (מפוזרות במרחב) נטו להופיע בעיקר בעת פניה שמאלה בצמתים מרומזרים, בתאונות חזיתיות ובקרב נהגים הנוהגים תחת השפעת אלכוהול.

2.2.3 שיטות לייחוס גיאוגרפי

קיימים מצבים יומיומיים בהם נדרש למצוא מיקום על רשת התחבורה. דוגמא לכך הינה בשעת דיווח על אירוע חירום. תאור המיקום אינו כה פשוט כפי שמצטייר לעיתים. מערכות הקואורדינאטות אינן תמיד תואמות, כתובות הרחובות אינן תמיד מעודכנות וסימני דרך מוסכמים אחרים, אינם תמיד מובנים [Noronha & Church, 2002]. לפיכך, קיימת חשיבות בקביעת שיטת הייחוס ומידול העצמים ברשת. להלן תובא סקירה של מספר שיטות מקובלות.

2.2.3.1 שיטת הייחוס הקווי (linear referencing)

שיטת הייחוס הקווי משמשת כמתודת ליבה בתחום התחבורה, היות ומרבית העצמים בהם עוסקים בתחום התחבורה הינם קווים [Sutton & Wyman, 2000]. שיטת הייחוס הקווי מאפשרת לקשור מספר מאפיינים לאותו אלמנט, ועוסקת באחסון מידע גיאוגרפי תוך שימוש במיקום יחסי לאורך עצם קווי קיים. שיטה זו מאפשרת לזהות ולמפות באופן ייחודי, מיקום לאורך עצם קווי ללא שימוש מפורש בקואורדינאטות קרטזיות למשל של X ו- Y. המיקום מתואר במונחים של אלמנט קווי ידוע ומיקום או מדידה לאורכו. כך, לדוגמא הביטויים כביש 40 , הק"מ ה- 100, ימפו מיקום גיאוגרפי ללא שימוש במונחי רשת הקואורדינאטות [Noronha & Church, 2002].

2.2.3.2 הסגמנטציה הסטאטית (Static Segmentation)

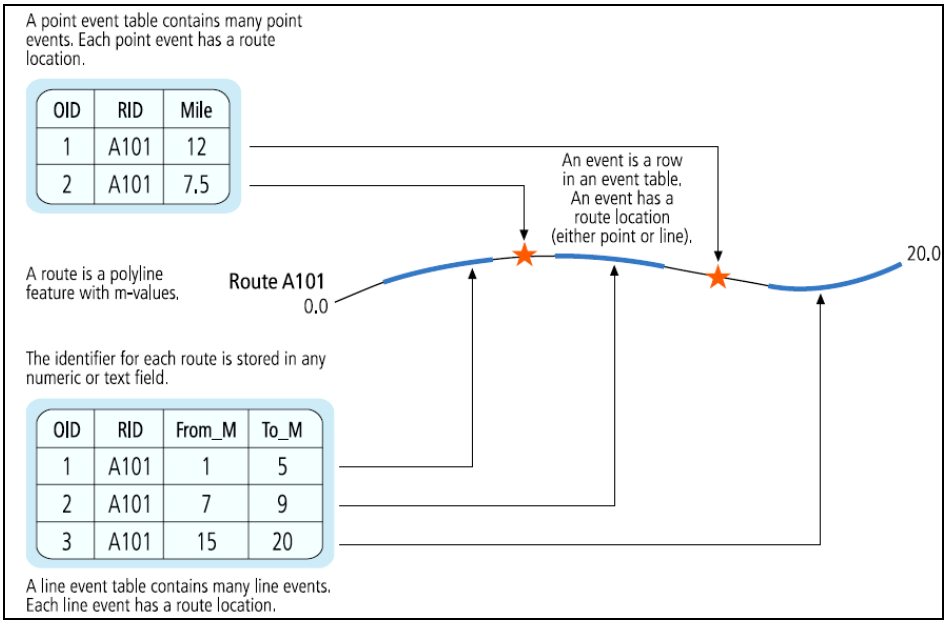
השיטה המסורתית למידול עצמים ברשת של סגמנטים, ולפיה מספר עצמים ה"קשורים" לאותו הסגמנט, מקושרים ישירות (על ידי מפתח כלשהו) לגיאומטריה של אותו הסגמנט.

הבעיה העיקרית בשימוש במתודה זו הינה, שהיא גורמת לסגמנטציה מוגזמת ולא מתואמת, הפוגמת בניתוח יעיל של רשת וביכולת לשיתוף מידע [Guo & Kurt, 2004].

2.2.3.3 שיטת הסגמנטציה הדינאמית (Dynamic Segmentation)

טכניקת הסגמנטציה הדינאמית, מתעלמת מהצורך לייצוג מפורש ואחסון הנתונים הגיאומטריים של עצם כלשהו כשכבה נפרדת בבסיסי המידע. כך, ניתן לקשור מספר מאפיינים לאותו מקטע באלמנט קווי כלשהו, כאשר אלמנט קווי (Route), יכול להיות כל עצם אליו ניתן לקשור אירועים, דוגמת: נהרות, רחובות, כבישים, צינורות וכיוצא בזאת.

תהליך הסגמנטציה הדינאמית מתואר כתהליך בו הופכים מידע טבלאי של אירועים (Events), בעלי ייחוס לאלמנטים קוויים כלשהם (Routes), למידע שניתן להציגו על מפה (ראה איור מספר 5). בכדי לבצע זאת, יש תחילה להגדיר את הפרמטרים של הייחוס בין הטבלה המאחסנת את האירועים לבין המקטעים אליהם עתידים להיות מקושרים. בנוסף, יש לוודא כי לכל Event בטבלת האירועים יהיה מזהה ייחודי ואת מיקומו לאורך האלמנט הקווי, כאשר לכל אלמנט קווי (Route) חייב להיות מזהה ייחודי [Cadkin, 2002 ; Cadkin & Brennan, 2002].

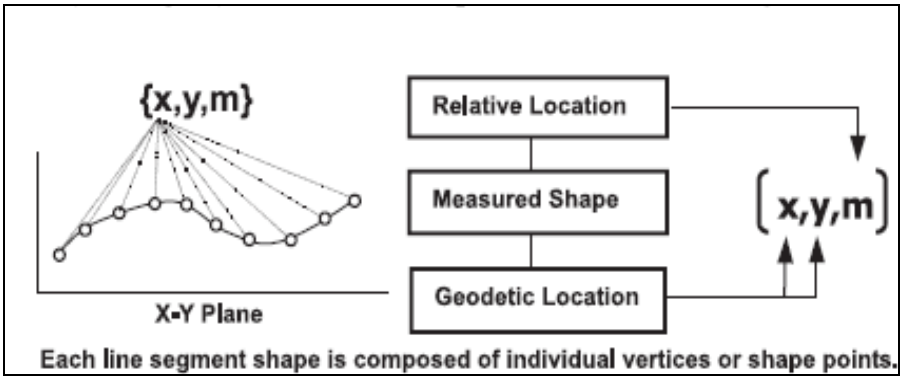


איור מספר 5 - הליך סגמנטציה דינאמית [Cadkin, 2002]

2.2.3.4 שיטת המיקום הדינאמי (Dynamic Location)

תהליך הייחוס הקווי, הממומש ללא קושי בסביבה טבלאית (דוגמת SQL), מעלה קשיים בעת שימוש בסביבת GIS עקב הצורך להתייחס לאלמנט המרחבי. Sutton & Wyman [2000], מציינים כי בכדי לספק מענה למגבלות ה GIS, פותח מודל המיקום הדינאמי.

המודל מתאר בצורה גרפית מפגשים או הצטלבויות מרחביות, ולפיו אובייקטים בעולם מתוארים בהתאם למיקום הקרטזי שלהם. אובייקטים לא ידועים: תחנת אוטובוס, שלטים, מיקום תאונות דרכים, יתוארו ויצוינו באותו האופן, על ידי ציון שתי קואורדינאטות קרטזיות ללא שימוש בטופולוגיה[10] . מסלולי נסיעה של אוטובוסים יתוארו אף הם כאירועים (Events) ויתוארו באופן קרטזי, ללא כל התייחסות למערכת של ייחוס קווי (ראה איור מספר 6).



איור מספר 6 - שיטת המיקום הדינאמי [Sutton & Wyman, 2000].

למעשה מדובר על שיטה הפוכה לשיטת הסגמנטציה הדינאמית, כאשר בניגוד לשיטת הסגמנטציה הדינאמית, שיטת המיקום הדינאמי מאחסנת את הגיאומטריה כאובייקט בתוך בסיס מידע יחיד. שיטה זו מונעת עיסוק נרחב בטבלאות רבות. עצמים ליניאריים מאוחסנים כמערך של קואורדינאטות (X,Y) ללא שום ייחוס לנתיב או מסלול. להגדרת מסלולים במערכת נבנים מערכים של X,Y ו - M (המגדירים מיקום יחסי או ערכי מדידות אחרות). כך מקבלים ערך מספרי לחישובים מתמטיים עתידיים. ניתן להתייחס לכל אובייקט לו צמד קואורדינאטות X, Y על פי המיקום שלו ברשת הקרטזית או בהתאם למספר השורה הייחודי שקיבל, או לפי כל הגדרה אחרת של צמד פרמטרים ובלבד שמוגדר כך מיקום ייחודי [Sutton & Wyman, 2000].

2.3 מתודות סטטיסטיות בשימוש חקר התחבורה

2.3.1 פיתוח מודלים סטטיסטיים

<u>גישות בבניית מודל</u>

קיימות שתי גישות עיקריות לבניית מודל: ניתוח רוחבי (Cross Sectional) וניתוח סדרות עיתיות (time-series). קיימים גם מודלים המשלבים בין שתי הגישות. הניתוח הרוחבי מתבסס על השוואה בין מאפייני תאונות ברשויות שונות ובאזורים שונים. לפיכך, ישים יותר בארצות

[10] סט של חוקים (דפוס הקשרים) המתארים יחסים מרחביים בין עצמים, דוגמת הקשרים המחברים נקודות וצמתים ברחבי הרשת.

גדולות דוגמת ארה"ב, אוסטרליה או אוסף מדינות דוגמת האיחוד האירופי. זאת, היות ובארצות מסוג זה, קיימים לרוב בסיסי נתונים המאפשרים השוואה "פנימית" בין המדינות. בגישה זו, ניתן לבחון פרק זמן קצר יחסית או מספר נקודות בזמן, כאשר קיים צורך להתחשב, דווקא במשתנים השונים בין המדינות. זאת, מאחר והשונות שלא נמדדה עשויה ליצוק קשר מוטעה בין מספר התאונות והמשתנים המסבירים שנכללו במודל. כך, עלול להיווצר פירוש לקוי של תוצאות המודל.

ניתוח הסדרה העיתי, הבוחן נתונים שנתיים, דורש סדרה ארוכה בכדי להבטיח תוצאה מובהקת של המודל. בדיקת סדרות שנתיות של תאונות מתייחסת בדרך כלל לתקופה של עשרות שנים, ולכן יש להביא בחשבון בעת הניתוח, גם את השינויים המהותיים שעברו על תוכן הרכב, את הרגלי הנהיגה, שימושי הקרקע, אופן רישום התאונות, ועוד. בחינת נתונים חודשיים של תאונות לאורך תקופה של 5-10 שנים, מבטיחה יותר נקודות תצפית לניתוח. יצוין, כי שיטה זו מוגבלת לעיתים, בגין חוסר היכולת לקבל נתונים חודשיים של משתנים רבים [Hakim et-al, 1991].

<u>קולינאריות (collinearity)</u>

סדרות שנתיות של משתנים דמוגראפיים, כלכליים ותחבורתיים, מדגימות (לרוב) קולינאריות גבוהה. דהיינו, ערכי המשתנה עולים במקביל או יורדים במקביל, במהלך הזמן. כך, במהלך השנים, נוטים לצמוח משתנים דוגמת: תל"ג, הכנסה פרטית, מספר הכלי רכב ונסועה. כתוצאה מכך עשוי להתעורר קושי בבידוד ההשפעה של המשתנה הקולינארי על התאונות. מגמת הזמן עלולה אף היא להיות קורלטיבית למשתנים התלויים. כך, עלול להיווצר מצב בו לא ניתן יהיה להבחין בהשפעת גורמי חשיפה שונים (או גורמים כלכליים על התאונות), עקב השיפור הרב-שנתי. בין אלה ניתן למצוא שיפורים במערכת הרפואה הדחופה. לשיפורים אלה השפעה אפשרית על תוצאות התאונה. מכאן, שקיימת חשיבות בבחינת המשתנים המשולבים במודל [Hakim et-al, 1991].

<u>המשתנים התלויים</u>

בעת בחירת המשתנים המסבירים, יש להעדיף את אלו המאופיינים בקורלציה נמוכה ככל האפשר. זאת, מכיוון שקורלציה נמוכה מחזקת את מקדמי המודל ומצמצמת את סטיית התקן (למעשה מצמצמת את השונות במודל). מחקרים מסוימים עושים שימוש בערכים מתוקננים, דוגמת: שיעור תמותה לנפש, שיעור תמותה לנסועה. שילוב "שיעור התאונות" כמשתנה תלוי במודל, מניח יחס ליניארי בין מספר התאונות או התמותה לבין הפרמטר המתוקנן. מצב זה מקשה על ביצוע בחינה סטטיסטית למודל. כך, עדיף לעשות שימוש במודל במספרי התאונות והנפגעים, ללא תקנון כלשהו [Hakim et-al, 1991 ; Page, 2001]. כמו כן, על מספר המשתנים במודל להיות לא גבוה מדי ולא נמוך מדי, בכדי לקבל תוצאות סטטיסטיות טובות [Page, 2001].

2.3.2 אגרגציה של מידע מרחבי וההשפעה על מאפייני הפיזור (פואסוני, נורמאלי, בינומי)

המרכיב המרחבי של תאונות הדרכים החל למשוך תשומת ליבם של חוקרים מתחום הבטיחות, בראשית שנות ה- 80 של המאה העשרים. עם זאת, תשומת לב מינורית ניתנה לבעיית קנה המידה בפיתוח מודלים של תאונות דרכים. מרבית המחקרים לא טיפלו בגודל וצורת היחידות המרחביות. סוגי מידע שונים שולבו יחדיו, ולרוב אף לא צוין כלל מהו אורך הסגמנט עליו בוצעה האגרגציה. לעיתים, חולקה רשת הכבישים לסגמנטים המאופיינים באורך קבוע, ולעיתים בסגמנטים בעלי אורך שונה. בנוסף, בניסיונות רבים למידול תופעת תאונות הדרכים התבצעה אגרגציה של תאונות דרכים על קווים (סגמנטים) או על משטח. אגרגציה זו מובילה לבעיה מאחר והגודל והצורה של יחידת המדידה המרחבית גורמת להסטה סטטיסטית של התוצאות. על אף חשיבות הנושא, ההשפעה העיקרית שנבחנה בהיבט קנה המידה במהלך השנים, התרכזה בשונות ובקורלציה, כאשר בעת הקטנת קנה המידה נוטה הקורלציה לעלות והשונות נוטה לרדת [Thomas, 1996].

Thomas [1996] עסקה בבחינת ההשפעה של אורך הסגמנט על שני משתנים הנמצאים בשימוש שכיח בחקר תאונות הדרכים: מספר תאונות למקטע כביש בפרק זמן נתון, וצפיפות התאונות (מספר התאונות ליחידת אורך). יודגש כי מחקרה ניסה לצמצם, ככל האפשר, השפעות של גורמים העלולים להטות מבחינה סטטיסטית את התוצאות. כך, הוצאו מהמדגם כלל תאונות הדרכים שמיקומם הוגדר כ "0". משמע, תאונות שהתרחשו בנקודות הקצה של מקטע הכביש או תאונות שקודדו כך, היות ומיקום האירוע לא צויין במדויק על ידי הבוחן שהכין את הדו"ח. בנוסף, עשתה שימוש בנתוני A.A.D.T. (Annual Average Daily Traffic). מאחר ונתוני מדד זה אינם כוללים מידע בנוגע לכלי הרכב העוברים בדרך בין השעות 22:00 ל- 6:00, הוסרו מבסיס המידע המקורי התאונות שהתרחשו בין השעות שצויינו. גם בעת קבלת ההחלטה על גודל (אורך) יחידת המחקר, נעשה ניסיון לצמצם מקורות, שיתכן ויהוו גורם לשונות. מכיוון שבכבישים המהירים בבלגיה, קיימת (לרוב) יציאה או כניסה כל 5 קילומטרים, כאשר אלה עלולות להוות מקור לשונות בתבניות ובכמות התנועה, נקבע כי גודלו של הסגמנט הגדול ביותר לא יעלה על 4.9 קילומטרים.

יחידת המחקר הראשית (Section) הוגדרה כחלק מכביש מהיר הנמצא בין שני מחלפים בכביש. לאחר הסינון נותרו בבסיס המידע 65 יחידות (Sections). יחידת המחקר המשנית, סגמנט (Segment), היוו חלק מיחידת מחקר באורך L. כך, מספר סגמנטים מרכיבים יחידו יחידת

מחקר (Section). גודלו של הסגמנט הגדול ביותר לא יעלה על 4.9 ק"מ (49 הקטומטר). גודלו הקטן ביותר של סגמנט הינו 100 מ'.

בשלב הבא עברו הסגמנטים המקוריים אגרגציה לסגמנטים בעלי אורך / הקטומטר. התקיימו שלוש אפשרויות של אגרגציה לסגמנטים: A,B, C (לכל סוג הוגדרו מספר סטים של נתונים).

A– עבור כל שני סגמנטים בעלי אורך / הקטומטר בוצעה אגרגציה, כאשר / קבוע עבור כל סט של נתונים. במצב זה , n (מספר הסגמנטים שנכללו בסט הנתונים) נע בין 3185 (כאשר 1 =/ הקטומטר) לבין 65 (כאשר 49 =/ הקטומטר). עבור / השווה ל - 1 הקטומטר, ניתן לומר כי N (המספר הכולל של התאונות שנצפו על n סגמנטים) של 3185 סגמנטים, שווה ל 411, כאשר אורך הסגמנט שונה מ 1 הקטומטר, נע ערך N בין 290 ל- 411 (היות ולא תמיד יתחלק L ללא שארית (במחקרה בחרה רק את השלמים).

B- בכל אחד מ- 65 יחידות המחקר (Sections), נבחר הסגמנט המרכזי. זאת, בכדי למזער ככל האפשר השפעות חיצוניות. כך n=65, ו N גדל ביחס ישר ובהתאמה לאורך /, ונע בין 11 (כאשר /=1 הקטומטר) לבין 411 (כאשר / =49 הקטומטר).

C– סגמנט שווה לsection, כאשר / הוא המרחק בין שתי יציאות מהכביש המהיר, ונע בין 50 ל 108 הקטומטר. כן n שווה ל- 65 ו- N שווה ל- 577.

תוצאות המחקר הראו כי ניתן לאפיין התנהגות סטטיסטית שונה בהתאם לשלוש הקבוצות של אורכי סגמנטים: סגמנטים קטנים מאוד (עד 1 הקטומטר) אופיינו בקירוב להתפלגות פואסון; סגמנטים גדולים (גדולים מ - 20 הקטומטר) אופיינו בהתפלגות נורמאלית ; וסגמנטים בינוניים (בטווח של 3-20 הקטומטר), בהם התאונות אופיינו בפיזור סטטיסטי הנע בין פילוג פואסוני לבין פילוג נורמאלי. המחקר מדגיש, כי מסקנות מניתוחים שהתבצעו על קנה מידה אחד, אפשרי ולא תהיינה נכונות עבור שימוש בקנה מידה שונה. לא נמצא קנה מידה אחד שניתן להגדירו כמיטבי. לפיכך, קיימת חשיבות רבה לתכנון תהליך האגרגציה והתאמת קנה המידה טרם ביצוע המחקר וביצוע מבחנים סטטיסטים מתוחכמים [Thomas, 1996].

2.3.3 מתודות ניתוח סטטיסטיות נהוגות

נושא תאונות הדרכים, נחקר במהלך עשרות השנים האחרונות, תוך שימוש במגוון רחב של מתודות סטטיסטיות ותוך בחינת משתנים מסבירים רבים. מודלים העוסקים בספירת תאונות וחיזוי מספר תאונות עתידי, החל מבחינת השפעות והקשרים ליניאריים, תוך שימוש במודלי רגרסיות ליניאריות, דרך מודלים בהם עשו שימוש בשיטת הריבועים הפחותים (least squares), לעיתים תוך שימוש בטרנספורמציות log-log או במשוואות ריבועיות. בכדי להשיג התאמה טובה יותר, פותחו בשלבים מאוחרים יותר, כאלטרנטיבה לרגרסיה הליניארית, מודלי הרגרסיה של פואסון ובינומי שלילי. זאת, עקב ההנחה של פיזור פואסוני של ספירות אירועי

תאונות דרכים [Thomas, 1996]. בהמשך החלו לשלב גם מודלים מסוג של Zero-Inflated. אלה, ממדלים גם מקטעי דרך המאופיינים בערך של 0 למקטע דרך [Qin et-al, 2005].
פרק זה יציג וירחיב את היריעה עבור מספר מאמרים העוסקים בחקר תאונות הדרכים (לאו דווקא במיקוד על מחקרים שעיקרם בספירת תאונות)

2.3.3.1 חי בריבוע Chi-Squared

מבחן חי (x^2) בריבוע לטיב התאמה, בוחן האם השכיחות הצפויה במשתנה אחד מתאימה לשכיחות הנצפית. זאת, על מנת לבדוק האם התפלגות השכיחויות במדגם שווה להתפלגות השכיחויות באוכלוסיה.

המבחן שימש את Schneider et-al [2004], אשר שילבו במחקרם הן ניתוחים מרחביים והן מודלי רגרסיה שונים, על מנת להבין את מכלול הגורמים המקושרים לסיכון התרחשות תאונת דרכים בה מעורב הולך רגל. מטרת מחקרם הייתה לבחון, האם קיומו של מידע תפיסתי (דוגמת אזור מסויים הנחשב מסוכן) יכול לשמש כפְּרוֹאַקְטִיבִי. משמע, קיום המידע יאפשר מניעת בעיות צפויות כדוגמת אירועי תאונות בהן מעורבים הולכי רגל. מטרתם תושג, במידה ויצליחו להוכיח בשלב ראשון, כי קיים שוני בין אזורים בהם מוערך סיכון גבוה לתאונות, לבין אזורים בהן מתרחשות התאונות בפועל. במידה ותבניות פיזור "התאונות בפועל" תמצא שונה מתבנית הפיזור של ה"תאונות הצפויות/מוערכות", ניתן יהיה להניח כי ביכולתו של המידע התפיסתי להוסיף מידע חשוב בנוגע לגורמים הסביבתיים המקושרים לתאונות הולכי רגל (מאפיני הכביש, חשיפה ועוד). בשלב השני, בוצע ניתוח הגורמים המקושרים לתאונות הולכי רגל בשני הפיזורים השונים דרך מודלי רגרסיה.
אזור המחקר אוכן באוניברסיטת צפון קרוליינה בארה"ב. המחקר התבצע במספר שלבים. בשלב הראשון, חולק המידע בנוגע ל 57 תאונות בהם מעורבים הולכי רגל (תאונות שדווחו על ידי <u>המשטרה</u> בתקופת המחקר שהוגדרה לחמש שנים), לארבע קבוצות עקריות תוך שימוש בתוכנת CRIME STAT. בשלב השני, חולקו שאלונים לשתי קבוצות עיקריות. הקבוצה הראשונה התמקדה במועמדים ועוברי אורח אקראיים בתחום הקמפוס (סטודנטים, עובדי האוניברסיטה, מרצים ואחרים). הקבוצה השנייה, הורכבה מבעלי כלי רכב להם היה היתר חנייה בתחומי האוניברסיטה (נוהגים בתחום הקמפוס). כל אחד התבקש, מנקודת מבטו, לסמן על גבי מפה שלושה אזורים המועדים לפורענות, בראייתו, בהיבט תאונות דרכים בקמפוס. עוברי האורח התבקשו לבצע זאת מנקודת מבט של הולכי הרגל, ובעלי היתר החניה מנקודת מבט של נהג רכב. כמו כן, מולאו מפות נפרדות עבור אזורי סיכון ביום ובלילה.

בחינת השאלונים והמפות הראתה כי סומנו 1835 נקודות בעלות סיכון לתאונה (קבוצת "תפיסת הסכנה") לעומת 57 התאונות ה"אמיתיות" שהתרחשו בין הולכי רגל לכלי רכב. עקב שינויים מועטים בין המפות והמיקומים שצויינו עבור יום ולילה, עובדו כלל 1835 מיקומי התאונות, שצויינו על ידי המשתמשים לכדי מפה אחת. זו, ייצגה את "תפיסת אזורי הסכנה" לתאונות בין כלי רכב להולכי רגל בקמפוס. כאמור, במידה ולא היה נמצא שוני בין הנתונים האמיתיים של מיקום תאונות הדרכים על פי נתוני המשטרה, לבין המפה ה"תפיסתית", לא היה למחקר ערך מוסף רב. אולם, בחינה וויזואלית פשוטה הראתה שוני רב, שיש לתומכו סטטיסטית. על מנת לבחון האם השוני בין פיזורי הנקודות הינו משמעותי מבחינה סטטיסטית, נעשה שימוש במתודות הסטטיסטיות הכמותיות (חי בריבוע) Chi-Squared ו- Nearest Neighbor Cluster (השכן הקרוב ביותר), כאשר היפותיזת ה- "0", היתה שהפיזור המרחבי של המיקומים הנתפסים כמסוכנים, אינם שונים מבחינה סטטיסטית, מהפיזור המרחבי של התאונות שהתרחשו בפועל.

במחקר עשו שימוש בשיטת Kernel density (להלן: KE) ליצירת משטח הסתברותי לתאונות הדרכים, כאשר גם שיטת KE וגם שיטת השכן הקרוב ביותר מניחות כי קיימת אקראיות מרחבית מלאה. הכוונה לכך היא, שיש להניח כי קיים פוטנציאל להתרחשות של תאונה בכלל שטח המפה (גם באזורי בניינים ושטחים פתוחים, שאינם מצויים בסמיכות לעורקי תחבורה). עם זאת, היות וידוע כי עיקר התאונות מתרחשות במגרשי החניה, לאורך הכבישים ובצמתים הרי שניתן לעשות שימוש בכלל נתוני המפה בעת ביצוע מחקר העוסק בתאונות דרכים.

כאמור, מבחן Chi-Squared בוחן במחקר את היחסים בין הנקודות במדגם. במקרה הנדון נבדקו היחסים בין מיקומי התאונות בכל פיזור נתון. המבחן מראה, האם כלל תאונות הדרכים הן אלה שדווחו על ידי המשטרה, והן אלה ה"נתפסות" כבעלות סיכון והנמצאות על הסגמנטים (הכבישים) והצמתים בקמפוס, יכולות להיות מסווגות לאזורים דומים, ולפיכך הפיזור דומה. יודגש, כי המבחן אינו לוקח בחשבון את המיקום במרחב של אותו סגמנט או צומת, אבל מאפשר לבצע השוואה ראשונית בין פיזור הנקודות בשתי הקבוצות, לאורך רשת הכבישים.

כל אחד מתשעים וארבעה הכבישים והצמתים במתחם הקמפוס, סווג בהתאם לספר התאונות שדווח עליו (0, 1, 2 או 3 ומעלה), לאחר מכן הושווה הסיווג למספר התאונות הצפוי על כל סגמנט או צומת. זאת, בהנחה כי התאונות תפוזרנה בתבנית מרחבית התואמת את המיקומים שנצפו. הסיווג המקביל שבוצע איפשר לעשות שימוש במבחן Chi-Squared, עבור הדיווחים האמיתיים שהתקבלו מהמשטרה, כמפורט בנוסחה מספר 13 :

$$X^2 = \sum_{i=1}^{n} \frac{(O_{i_r} - O_{i_e})^2}{O_{i_e}} \quad [13]$$

כאשר:

i - מייצג כל אחד מארבע הקטגוריות 0, 1, 2 או 3 ומעלה תאונות שדווחו על ידי המשטרה.

O_{i_γ} - מספר הסגמנטים/צמתים הנופלים בקטגוריה *i*, בהתאם למספר הדיווחים האמיתיים (משטרתיים).

O_{i_ε} - המספר הצפוי של סגמנטים והצמתים העתידים "ליפול" בקטגוריה *i*, במידה ודיווחי המשטרה יפוזרו באותה תבנית מרחבית לאורך רשת הכבישים באותו האופן ש"סקר התפיסה" הניב.

המבחן בוצע גם עבור המידע התפיסתי, תוך שימוש בנוסחה מספר 14, להלן:

$$X^2 = \sum_{i=1}^{n} \frac{(O_{i_P} - O_{i_e})^2}{O_{i_e}} \quad [14]$$

כאשר :

i - מייצג כל אחד מחמש הקטגוריות 0-29, 30-59, 60-89, 90-119 או 120 או יותר תאונות שנצפו.

O_{i_p} - מספר הסגמנטים/צמתים הנופלים בקטגוריה *i*, בהתאם למספר הדיווחים שדווח בסקר התפיסה.

O_{i_ε} - המספר הצפוי של סגמנטים והצמתים העתידים "ליפול" בקטגוריה *i*, במידה ודיווחי "סקר התפיסה", יפוזרו באותה תבנית מרחבית לאורך רשת הכבישים באופן בו דיווחי המשטרה התפזרו.

תוצאות המבחנים הראו כי מספר התאונות על כל סגמנט או צומת הן בקבוצה של התאונות ה"אמיתיות" המתועדות על ידי המשטרה, והן בקבוצה של המיקומים הנתפסים כבעלי סיכון, חולקו לקטגוריות נפרדות. משמע, קיים שוני באופן בו האירועים מפוזרים לאורך כבישי וצמתי הקמפוס בין שתי הקבוצות.

בשלב השני, נעשה שימוש בסוג השני של השוואה מרחבית - שיטת "השכן הקרוב ביותר". זאת, בכדי לזהות אשכולות של נקודות במרחב. הכוונה היא כי לאחר שתזוהנה

קבוצות, ניתן יהיה להשוות אשכול של קבוצה אחת לאשכול של הקבוצה השנייה. השיטה עושה שימוש בחישוב המרחק האוקלידי. כך, ניתן לזהות סט של נקודות המקובצות באופן משמעותי יותר מהצפוי בפיזור אקראי רגיל. נקודות תיחשבנה כאשכול, כאשר ממוצע המרחק האקראי ביניהן הינו פחות ממינימום המרחק המבוסס על שגיאה תיקנית של פיזור אקראי, כמצוין בנוסחה מספר 15, להלן:

[15] Minimum Distance = $0.5\sqrt{\frac{A}{N}} - t\left[\frac{0.26136}{\sqrt{N^2/A}}\right]$

כאשר:

A – שטח אזור המחקר

N - מספר מיקומי אירועי תאונות הדרכים (57 - מיקומי התאונות על פי דיווחי המשטרה, או לחלופין 1835, במידה ומדובר בסט הנקודות ה"תפיסתי").

t- רמת ההסתברות בפיזור

$[0.26136/\sqrt{N^2/A}]$ - השגיאה הסטנדרטית של מרחק בפיזור אקראי.

במילים אחרות, עבור הסתברות חד כיוונית P, פחות מ- P אחוזים מהנקודות, יאופיינו במרחק משכניהם, שיהיה קטן מגבול תחתון זה, במידה הנקודות יאופיינו בפיזור שהינו לגמרי אקראי.

במחקרם עשו Schneider et-al [2004], שימוש בערך של P =0.01, בכדי שיוכלו להיות בטוחים ב- 99 אחוזים שאשכולות דיווחי המשטרה והדיווחים ה"תפיסתיים" אינם "מתגבשים" באופן אקראי.

מספר הנקודות המינימאלי ליצירת אשכול הוגדר כ- 3 בסט הנתונים ה"אמיתי" שדווח (משטרה) ו- 40 בסט הנקודות ה"תפיסתי". אליפסת סטיית התקן מסביב לכל אשכול ומספר הנקודות שהאשכול מכיל דווחה אף היא. אליפסת סטיית התקן מודדת את הפיזור והכיוון של הנקודות מסביב למרכז הממוצע (אורך ממוצע ורוחב ממוצע של האשכול). הקבלה בין שני הפיזורים המרחביים, תתקיים בעת חפיפה של שתי האליפסות.

תוצאות ההשוואה עבור וודאות של 99%- דיווחי המשטרה על פי שיטת השכן הקרוב, קובצו לשלושה אשכולות. זאת, לעומת שישה אשכולות שקובצו בסט הנתונים התפיסתי. היות ומרכזי הכובד ותחומי האליפסות לא חפפו האחד את השני , הסיקו החוקרים כי בין שני הסטים של הנתונים קיימים מאפיינים מרחביים שונים. לפיכך קבעו כי מידע תפיסתי אכן מוסיף פן חשוב, שיאפשר למנוע מראש תאונות דרכים.

כאמור, במספר מקרים, מיקומי תאונות הדרכים האמיתיים לא חפפו את מיקומי האזורים ה"תפיסתיים". לפיכך, בכדי לבחון השפעה והקשרים בין גורמים שונים, גיאוגרפים ואחרים (דוגמת: שימושי הקרקע, מרחק מהספריה, קיום מדרכה, מספר אוטובוסים העוצרים באותו הסגמנט ומאפייני הדרך אחרים) על תאונות הדרכים, נעשה שימוש במודלי רגרסיה של פואסון (Poisson) ובינומי שלילי (Negative Binominal). מודלי הרגרסיה מדגימים את הקשר בין המרחב וניתוח תאונות רב משתני.

רשת הכבישים חולקה ל- 38 צמתים (15 מ' ממפגש הצירים של שניים או יותר כבישים) ו- 56 סגמנטים. לכל אחד מהמקטעים שוייכו מאפיינים כגון: חשיפה, רמת סיכון, שימוש קרקע, סוג הכביש, מרחקים ממקומות מרכזיים ואחרים. תוצאות המחקר הראו, כי ככל שהסגמנטים ארוכים יותר, וככל שכמות הולכי הרגל גבוהה יותר, כך גדלו דיווחי התאונות של המשטרה באותם המיקומים. עליה של אחוז אחד בכמות הולכי הרגל, העלתה את כמות דיווחי המשטרה בשיעור של 0.672%. ככל שבמקטע הנבחן מוקמו יותר מעברי חצייה, כך עלה מספר הדיווחים על תאונות הדרכים.עדין נדרש מחקר נוסף בנושא, היות ובהיבט המרחבי, המחקר לא בחן את הקשר בין תאונות לקרבה למעברי החצייה.

בהיבט המרחבי, נמצא קשר בין קרבה לבניינים אקדמיים ואצטדיונים לבין סיכוי גבוה יותר לתאונה. אזורים "קרובים לספריה" נמצאו כמאופיינים בסיכון נמוך לתאונה. מעניין לציין, כי דווקא אזורים אלו נתפסו בעיני משתמשי הסקר כאזורים בעלי סיכון. בחינת מספר תחנות האוטובוס הקיימות העלה כי מבחינה סטטיסטית, מספר גבוה יותר של תחנות אינו מקושר-סכנה לתאונה בה מעורבים הולכי רגל. בנוסף, מרחק ממעונות, בתי חולים מגרשי חניה וחדרי אוכל, נמצא כבעל קשר סטטיסטי נמוך לתאונות. עם זאת, הם מציינים כי היחס בין קרבה לשימושי קרקע מסוימים במרחב דורש מחקר נוסף.

2.3.3.2 מודל רגרסיה - פיזור פואסון

השימוש בהתפלגות פואסון עולה לעיתים קרובות, כאשר קיים צורך בשימושים מעשיים. המשותף לרוב השימושים הוא קיום זרם של אירועים לאורך זמן, באופן שהתפלגות מספר האירועים בקטע זמן כלשהו, תלויה בעיקר באורך הקטע. הנחה נוספת הינה, כי מספר האירועים בשני קטעי זמן זרים הם בלתי תלויים זה בזה. קרי: האירועים אקראיים.

נוסחת ההתפלגות הפואסונית מוצגת בנוסחה מספר 16:

$$P\left[x = k\ \right] = \frac{e^{-\lambda}\lambda^{k}}{k\ !} \qquad [16]$$

λ- קצב המופע של האירועים (ממוצע אירועים ליחידת זמן או שטח)

2.718 = e

K=0,1,2…

התפלגות "פואסון" מאופיינת בכך שהממוצע שווה לשונות. תכונה זו מונעת את הצורך לאמוד את סטיית התקן של ההתפלגות מתוך נתוני המדגם, אשר לעיתים יהיה קטן מדי ולא יאפשר קבלת אומדן סביר של סטיית התקן [לויתן ורביב , 1997 ; הרשות הלאומית לבטיחות בדרכים] .

Schneider et-al [2004], שמחקרם פורט בהרחבה בסעיף 2.3.3.1, עשו אף הם שימוש בהתפלגות פואסון כמפורט בנוסחה [17]. כאמור, מתוך הנחה כי הינה מתאימה לחישוב ההסתברות להתרחשות תאונה על כל סגמנט:

$$P\left[Y_i = y_i\right] = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \qquad [17]$$

כאשר:

Y_i - מספר התאונות **בפועל** על סגמנט/צומת i בזמן המחקר (i נע מ- 1 ועד N)

λ_i - התדירות **הצפויה** לתאונה בסגמנט/צומת i (i נע מ- 1 ועד N)

y_i - מספר התאונות שהתרחש בעבר על כל סגמנט/צומת i (y_i =0,1,2…). כאשר למעשה, על בסיס מספר התאונות הקודם בכל סגמנט ניתן להניח מהי ההסתברות לתאונה נוספת שתתרחש.

2.718= e

i = i נע מ- 1 ועד N, כאשר N=94

בכדי לשלב גורמי מסבירים (Xi) הוגדר הפרמטר λ_i באופן המוצג בנוסחה 18:

$$\lambda_i = \exp(\beta^i \chi_i) \qquad [18]$$

β^i - הוקטור של הפרמטר המוערך

χ_i - המשתנה המסביר (לדוגמא: הולך רגל, נוכחות מעבר חציה וכדומה)

כאמור, המודל מתאים עבור מדגמים בהם הממוצע והשונות שווים. לפיכך, במקרה שבחן Schneider , דווקא המודל הבינומי השלילי תאם יותר.

קלרמן [1986], מציין כי לאחר חישוב התפלגות פואסון, יש לבחון את ההשערה לפיה, מערך הנתונים הנבחן, אכן מאופיין בתהליך כזה. הוא הציע שני סוגים של מבחנים: אל-פרמטרים המשווים חישובים נצפים אל מול מיוחלים (דוגמת מבחן חי בריבוע, ומבחן קולומוגורוב-סמירנוב); ומבחן פרמטרי בשם יחס שונות/ממוצע (variance mean ratio test), המספקים רמז להשערות חלופיות (דגם סדיר או מקובץ). היות והממוצע והשונות של התפלגות פואסון שווים, ניתן לחשב את היחס על ידי שימוש במבחן T.

2.3.3.3 המודל הבינומי השלילי

ניתן לראות את התפלגות פואסון בתור התפלגות בינומית, כאשר מספר הניסויים המבוצע (n) הינו גדול מאוד (שואף לאינסוף), וההסתברות להצלחה בכל ניסוי (P) קרובה או שואפת לאפס [Schneider et-al, 2004].
במחקרים קודמים [Karlaftis & Tarko, 1998] נמצא כי שימוש ברגרסיות ליניאריות רבות משתנים אינו מתאים לחקר תאונות הדרכים. זאת, בשל ההנחות הסטטיסטיות הנדרשות בעת קיום רגרסיות ליניאריות (דוגמת משתנים בעלי שונות זהה), שאינן מתקיימות בעת ניתוח תאונות דרכים. שימוש במתודות אלה, תוך הפרת ההנחות הסטטיסטיות, עלול לגרום לסטייה ברווח הסמך של האומדים. כך, בכדי להעריך אירועי תאונות דרכים, מתוך ההנחה כי אירוע תאונה הינו אקראי, מתאימה רגרסיית פואסון. שימוש ברגרסיה זו עלול להניב תוצאות שגויות במידה והממוצע אינו שווה לשונות. במידה והשונות גדולה מהממוצע, נוטה הפונקציה לפיזור יתר. ניתן להתגבר על קושי זה על ידי שימוש בהנחה נוספת בנוגע לממוצע הצפוי - הנחת פיזור גמא (gamma). כלומר, הנחה לפיה מספר תאונות הדרכים יאופיין בפיזור פואסוני, עם ממוצע צפוי המאופיין בפיזור גמא. כך, למעשה ניתן להתייחס למספר התאונות כמאופיין בפיזור בינומי שלילי [Ng et-al, 2002]

Ng et-al [2002], עסקו בפיתוח אלגוריתם שיאפשר להעריך, על בסיס תאונות שהתרחשו בעבר, את מספר התאונות והסיכון לתאונה העתידיים. האלגוריתם משלב טכניקת מיפוי בממ"ג יחד עם מתודות סטטיסטיות (איכון אשכולות וניתוח רגרסיות). שימוש בממ"ג מאפשר למפות את התאונות ולבחון את הפיזור ויזואלית. האלגוריתם שפותח בנוי מארבעה מודולים: בניית בסיס מידע אחוד (מידע והגורמים בנוגע לתאונות); אישכול (Clustering) המידע; כיול המודל ; ואיתור "black sites". כאשר עבור כל אשכול נבחנו הגורמים להם ההשפעה הגדולה ביותר על התאונות.

אישכול – טכניקה סטטיסטית שמטרתה קיבוץ אירועים על בסיס הדומה או השונה בין האירועים הנבחנים והתצפיות. במחקר נעשה שימוש באלגוריתם המרחק המינימאלי של Ward [1963]. האלגוריתם עושה שימוש במרחק אוקלידי מינימאלי בין תצפיות. התבנית המרחבית של מיקום התאונות הוצגה במקור בתוכנת ה GIS: Map Info. כמו כן, נעשה שימוש בפונקציית thematic map בכדי להדגים את תוצאות האישכול על גבי המפה.

מודל הכיול – הורצו מספר מודלים של רגרסיה בינומית שלילית, על מנת להסביר את מספר התאונות במונחים של גורמים פוטנציאלים יומיומיים. מספר תאונות הדרכים הוגדר כמשתנה התלוי, כאשר הגורמים היומיומיים שימשו כמשתנים הבלתי תלויים. ההנחה היא כי מספר התאונות מתפלג בהתפלגות בינומית שלילית המוצגת בנוסחה מספר 19:

$$[19] \quad P(Y=y)=\frac{\Gamma(y+\alpha^{-1})}{y!\Gamma(\alpha^{-1})}\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{y}\left(\frac{1}{1+\alpha\lambda}\right)^{\alpha^{-1}}$$

כאשר :

Γ - פונקצית גמא.

α - פרמטר overdispersion בעל ערך חיובי (מצב זה מניח כי השונות ה"אמיתית" גבוהה יותר מאשר זו של המודל ה"תיאורטי").

λ - ממוצע התאונות הצפוי

y - מספר התאונות

בהתפלגות בינומית שלילית ניתן להציג את הממוצע והשונות של מספר התאונות , כ- λ וכ- $\lambda+\alpha\lambda^2$, בהתאמה.

את λ ניתן להציג כפי שמופיע בנוסחה מספר 20:

$$[20] \quad \lambda=\exp(x^t\beta)$$

נוסחה השוות ערך לנוסחה מספר 21:

$$[21] \quad \ln\lambda=x^t\beta$$

כאשר:

x - הווקטור המייצג את עמודת המשתנה המשמש כהסבר

x^t -x לאחר הטרנספורמציה

β - עמודת מקדם הרגרסיה

בכדי לחשב את מקדם הרגרסיה β ואת הפרמטר α, נעשה שימוש במתודת Maximum likelihood Estimation (להלן: MLE). המתודה שואפת להשיג את הפרמטרים של α ו - β על ידי הבאה למקסימום של פונקצית הסבירות (likelihood function). זאת, כאשר מקדם הרגרסיה β יוערך על ידי השימוש במתודת Raphson-Newton עבור ערכי α נבחרים. התהליך ישנה על עצמו עד שפונקצית הסבירות תגיע למקסימום ביחס למקדם הרגרסיה β וביחס ל α. במחקר נעשה שימוש בתוכנת MATHCAD בכדי לחשב את הפרמטרים בשיטת MLE.

בסטיית התקן עשו Ng et-al [2002], שימוש בכדי לבחון כיצד המודל הסטטיסטי שנבנה תואם את סט התצפיות. במידה והמודל מייצג את המידע באופן טוב יותר מהמציאות, ערך הסטייה יהיה קרוב יותר לדרגת החופש של מודל הרגרסיה. רמת המובהקות של הגורמים הנבחנים ניתנת למדידה על ידי ערך ה - t, המוגדר בהתאמה למקדמי הרגרסיה. ככל שערכו גבוה יותר, כך הינו משמעותי יותר כגורם להתרחשות התאונה.

בכדי לחשב את הסיכון לתאונה עתידית באזור מסוים הופעלה שיטת Empirical Bayes (להלן: EB). שיטה זו עושה שימוש במספר הנצפה ובמספר המוערך של תאונות בכדי להניב אומדן "טהור". הנוסחה בה נעשה שימוש לחישוב ה"סיכון" מופיעה בנוסחה מספר 22.

$$EB = \left[\frac{\lambda}{\frac{1}{\alpha} + \lambda} \right] * \left(\frac{1}{\alpha} + y \right) \qquad [22]$$

כאשר:

α - פרמטר overdispersion בעל ערך חיובי (מצב זה מניח כי השונות ה"אמיתית" גבוהה יותר מאשר זו של המודל ה"תיאורטי").

λ - ממוצע התאונות הצפוי

y - מספר התאונות ה"אמיתי"

סוגי התאונות חולקו לשלוש קבוצות בהתאם לחומרת התאונות: תאונות קטלניות (לפחות אדם אחד נהרג במקום או במהלך 30 יום מהתרחשות התאונה), חמורות (אשפוז מעל 12 שעות בבית החולים) וקלות (הנפגע שהה פחות מ - 12 שעות בבית החולים). במחקרם בחרו Ng et-al [2002], כמשתנים התלויים, את הגורמים הבאים: מספר התאונות הכולל; מספר התאונות הקטלניות ומספר התאונות בהם מעורבים הולכי רגל. הגורמים הבלתי תלויים הינם שימושי הקרקע, להם השפעה על ההתנהגות

האנושית. דוגמא לכך היא, כי הולכי רגל נוטים לנוע מהר יותר באזורים מסחריים מאשר באזורי המגורים שלהם. בנוסף, נמצא כי לפעולות אנושיות, דוגמת: התנהגות, פיזור אוכלוסיה, רמת התעסוקה ותבנית התנועה, השפעה על רמת התרחשות התאונות. עקב כך, לשימושי הקרקע השפעה על דרגת התאונה, וקיימת חשיבות בהגדרת גורמי המפתח, להם ההשפעה הגדולה ביותר.

האלגוריתם שיפר את אומדן ההערכה לסיכון להתרחשות תאונה בהשוואה להערכת סיכון המבוססת על מידע היסטורי בלבד. תוצאות המחקר הראו כי, האלגוריתם נמצא יעיל יותר בעת ניתוח מקרים של תמותה והולכי רגל. בנוסף, הצביעו התוצאות על קיומם של שני אשכולות (שני אזורים). האשכול הראשון "צפוני", בעל שיעור תאונות גבוה. האשכול השני, בעל שיעור תאונות נמוך. מודל הרגרסיה יוכל להסביר חלק גדול מהשונות באירועי התאונות, עבור ערכי השונות הקרובים לדרגות החופש של מודל הרגרסיה. כך, ככל שערכי ה- t של המשתנים יהיו גדולים מהערך הקריטי, תעלה רמת חשיבותם בהתרחשות התאונה. כמו כן נערך מבחן "חי בריבוע" אותו הישוו לתוצאות אמיתיות, ול- EB. התוצאות הראו כי שיטת ה- EB השיגה תוצאות הקרובות יותר למציאות. נמצא, כי שימושי הקרקע שאותרו כבעלי השפעה על התאונות באזור אחד, אינם תואמים את האזור השני שנבחן. באזור אחד נמצא כי: שימושי קרקע מסחריים, מספר המושבים בבתי הקולנוע, מספר המיטות בבתי החולים, מספר השווקים והאזורים הנמצאים בקרבת תחנות רכבת היו בעלי קשר חיובי ומובהק למספר התאונות הכולל. לעומת זאת, באזור השני נמצא כי שימושי קרקע מעורבים של מסחר ומגורים, אזורי תעשייה מיוחדים ואזורי בתי ספר, נמצאו כבעלי קשר שלילי מובהק לתאונות. עוד נמצא באזור השני, כי: שימושי קרקע של אזורי אחסון, אזורי פיתוח, שטחים ירוקים, אזורי מוסדות ממשלתיים, גני ילדים ומשרדים הינם בעלי קשר חיובי ומובהק לתאונות דרכים. אזורים של בתי ספר על יסודיים ואזורי תעשייה מיוחדים נמצאו כבעלי קשר שלילי מובהק לתאונות.

2.3.3.4 Safety Performance Functions (SPF) ושיטת Empirical Bayes

הבטיחות של ישויות דוגמת: מקטע כביש, צומת, נהג, וכיוצא בזאת, מוגדרת על ידי מספר התאונות הצפויות להתרחש באותה ישות בתקופה נבחרת (בהתאם לחומרה ולסוג התאונה). מאחר וה"צפוי" אינו ידוע, ניתן רק להעריך את הבטיחות של הישות, כאשר דיוק ההערכה נעשה בדרך כלל על ידי שימוש בסטיית תקן. בטיחות של ישויות עליהן מתרחשות תאונות רבות בפרק זמן קצר, ניתן ואף נהוג במקרים רבים, להעריך

באופן מדוייק יחסית על ידי ספירת התאונות באותה ישות. עם זאת, כאשר פרק הזמן בו בוחנים את מספר התאונות בישות הינו ארוך, ההערכה הופכת ללא מדוייקת.

חסרון נוסף של הערכת בטיחות המתבססת על ספירת תאונות, היא שכיחות ההטייה. לרוב, ישויות מושכות עניין בגין ריכוז תאונות עליהן או אינן מושכות עניין בגין מספר תאונות מועט. כך לרוב, הבחירה ביישות לבחינה וניתוח, אינה אקראית. בשני המקרים, הערכה העושה שימוש רק בספירת התאונות תהייה מוטה, בשל תופעת רגרסייה לכיוון הממוצע (regression to mean). מהות התופעה: כאשר דגימה ראשונה "מייצרת" ממוצע גבוה באופן חריג, בדרך כלל תהיה הדגימה הבאה נמוכה יותר וקרובה יותר אל הממוצע, משום שחלק מהגורמים ל"ביצועים" הגבוהים אינם נשמרים. כלל זה מתקיים גם בכוון ההפוך. בהתאמה, צפויה עלייה בביצועים לאחר ביצועים נמוכים. כך לדוגמא, כאשר מתגלה מספר תאונות בצומת בשנה מסוימת הגבוה מהממוצע, סביר שבשנה הבאה אחריה יהיה מספר התאונות קרוב יותר לממוצע, גם אם לא יעשה כל טיפול בצומת. כאשר משתמשים בסטטיסטיקה של מספר גדול של תאונות בקטע דרך כהצדקה להקצאת משאבים לטיפול באותו קטע (בין אם מדובר בשיפורי תשתית או באמצעי אכיפה), חשוב לנטרל את ההשפעה האפשרית של אפקט הנסיגה אל הממוצע. זאת, כדי שלא נייחס לטיפול השפעה שאינה קיימת ונגיע למסקנה כי התיקון שבוצע הוא שגרם לשינוי ולא הרגרסיה הטבעית לכיוון הממוצע.

שיטת EB מעלה את דיוק ההערכה עקב שימוש במידע היסטורי של שנתיים שלוש. בנוסף, השיטה מתקנת את ההטיה של תופעת ה"נסיגה/רגרסייה אל הממוצע", מכיוון שהיא יוצאת מתוך ההנחה שספירת התאונות בישות כלשהי, אינה הרמז היחידי. רמז נוסף הינו הידע הקיים בנוגע לישויות בעלי מאפיינים דומים. לדוגמא, נניח כי במדינה מסויימת נחשב אדם X לנהג חדש, שלא היה מעורב כלל בתאונת דרכים בשנת נהיגתו הראשונה. בנוסף, ידוע לנו, שממוצע תאונות הדרכים של נהגים צעירים באותה המדינה עומד על 0.08 תאונות בשנה. יהיה זה שגוי לטעון, כי אותו אדם X צפוי להיות מעורב ב-0 תאונות בשנה בהתבסס על רקורד התאונות האישי שלו. הנחה הגיונית חייבת לשלב את שני סוגי המידע. בדומה לכך, בכדי להעריך את הבטיחות על מקטע דרך, הנמצא לדוגמא בכביש כפרי בעל שני נתיבים, חשוב לעשות שימוש לא רק בספירת התאונות על אותו הסגמנט, אלא לעשות שימוש גם בתדירות התאונות המאפיינת כבישים מסוג זה [Hauer et-al , 2002].

שיטת ה EB עושה שימוש ברקורד התאונות בישות, במשולב עם תדירות התאונות הצפויה על אותו מקטע. תדירות התאונות הצפויה בישויות דומות נקבעת על ידי Safety Performance Function (SPF), כאשר שילוב שני סוגי המידע מיושם בדרך של

חישוב ממוצע משוקלל, כמפורט בנוסחה מספר 23, שבה נוסחת החישוב עבור מספר התאונות הצפוי עבור ישות:

[23]

$$Estimate\_of\_Expected\_Accidents\_for\_Entity =$$

$$Weight * Accidents\_Expected\_on\_Similar\_Entities + (1 - Weight) * Count\_of\_Accident\_on\_this\_Entity$$

$$where\_0 \le Weight \le 1$$

תוצאת הנוסחה נקבעת על ידי ה"משקל" ("Weight") הניתן לכל תאונה הצפויה להתרחש על ישות דומה. יתרונה של שיטת EB טמון בשימוש ב"משקל" המבוסס על נתונים אמיתיים. המשקל כפי שיוצג בהמשך, מחושב על בסיס עוצמת רקורד התאונות הצפוי ומהימנות ה SPF, בה מחושב בכמה שונה הבטיחות בישות מסויימת מהממוצע של SPF.

הליך ה- EB עושה שימוש גם בטווח נתונים קצר (abridged), המכיל מספר תאונות שהתרחשו בטווח זמן של שנתיים שלוש וממוצע נפח התנועה באותה התקופה. חשיבות היכולת לעשות שימוש במידע "עדכני" הינה היכולת לקשור התאונות למצב כביש עכשווי או למעשה עדכני.

תדירות התאונות הממוצעת של ישויות או אתרים דומים והשונות סביב אותו הממוצע, משולבת בשיטת EB בפונקצית ה- Safety Performance . SPF הינה משוואה המספקת הערכה של μ (ממוצע תאונות ביחידות של ק"מ/בשנה) כפונקציה של ערך מאפיין (דוגמת: average daily traffic ADT, רוחב הנתיבים וכו'). SPF מכויילת מהנתונים תוך שימוש בטכניקות סטטיסטיות. בבסיס הכיול, עומדת ההנחה המקובלת, לפיה תאונות מאופיינות בפיזור בינומי שלילי. אחד המאפיינים של הפיזור הינו פרמטר over dispersion המסומן בנוסחה על ידי האות φ. עבור סגמנט של כביש, הפרמטר מוערך עבור יחידת אורך [1/MILE] או [1/KM].

אם L הינו האורך של הסגמנט, ו- η הינו מספר התאונות הצפוי על אותו הסגמנט, אזי השונות של ספירת תאונות על סגמנט מסוג זה מוצגת בנוסחה מספר 24:

[24] $$\eta\left[1+\eta/\varphi L\right]$$

יש לשים לב ליחידות. אם במהלך הכיול φ חושב בק"מ גם L יוערך ביחידות ק"מ. עבור צמתים L נחשב כשווה 1 [Hauer et-al , 2002].

השלב הראשון בהליך ה- EB, הינו קביעת פונקציית ביצועי הבטיחות (Safety Performance Function ; להלן: SPF). ה- SPF הינה מודל מתמטי החוזה את מספר התאונות העתידיות עבור מקטע כביש i. לדברי Hauer et-al [2002] מידול התרחשות תאונות הינו מיטבי, כאשר עושים שימוש במודל מרובה משתנים. המודל הינו משוואה או מערך משוואות המקשרות בין תדירות התאונות הצפוייה על סגמנט, לבין מאפייני הסגמנט, דוגמת: AADT, אורך הכביש, רוחבו, רוחב השוליים, מספר נתיבים וכיוצא בזאת. לכל סוג דרך, עירונית או בינעירונית, תתקיים SPF שונה לקביעת מספר התאונות הצפוי.

Powers & Carson [2004], הדגימו כיצד לעשות שימוש בשיטת EB להשוואת בטיחות אזורים "לפני" שנעשתה בהם פעולה לשיפור הבטיחות על מנת להפחית אירועי תאונות דרכים ו "לאחר" שטופלו, במונטנה. במאמרם, מידלו בשלב הראשון, את SPF תוך שימוש ברגרסיה ליניארית רבת משתנים[11], תוך שימוש בנתוני תאונות דרכים של שלוש שנים. נוסחה מספר 25 מציגה את המידול:

$$[25]\quad SPF_i = \beta_0 + \beta_1\chi_{i1} + \beta_2\chi_{i2} + ... + +\beta_{p-1}\chi_{i,p-1} + \varepsilon_i$$

כאשר:

SPF_i - המשתנה התלוי (מספר תאונות בסגמנט i לפני שנערך כל שינוי בסגמנט)

$\chi_{i1} \to \chi_{i,p-1}$ - המשתנים המסבירים, הבלתי תלויים (דוגמת AADT, המהירות המותרת, רוחב הנתיב, מספר נתיבים וכיוצא בזאת)

$\beta_0 \to \beta_{p-1}$ - מקדמי הפונקציה

ε_i - הטעות האקראית, הבלתי מוסברת במודל

במחקרם מצאו כי נפח התנועה (AADT) ואורך הכביש הינם המשתנים המסבירים המשמעותיים ביותר מבחינה סטטיסטית, בהשפעתם על תאונות הדרכים, ובהתאמה, נוסחת ה SPF מוצגת בנוסחה 26.

$$[26]\quad SPF_i = \beta_0 + \beta_1(L_i) + \beta_2(AADT_i)$$

כאשר :

L_i מייצג את אורך קטע i

$AADT_i$ - **ממוצע שנתי** של נפח התנועה היומי התקופת המחקר.

[11] כאמור, מודל של רגרסיה ליניארית, אינו מקובל כיום למידול תאונות דרכים.

משתנים אחרים דוגמת: רוחב השול, מספר נתיבים והגבלת מהירות, לא נמצאו כמשמעותיים במודל, עקב היותם קבועים. כך לדוגמא, משתנה "רוחב השול" קבוע לאורך מקטעי הכביש השונים, גם במקטעים בהם התרחשו תאונות רבות וגם עבור מקטעים בהם התרחשו מעט תאונות. לדבריהם, עם שני משתנים מסבירים משמעותיים, טיב המודל (Goodness of Fit) הינו נמוך, אולם הם ממשיכים ומתארים את השיטה. כך, בשלב שני, לאחר בניית הפונקציה המתאימה, נדרשים להערכת פרמטר ה-overdispersion (ϕ).

כחלק מהערכת ה- SPF, נדרש לבחון מהי ההתפלגות המאפיינת את ההסתברות הבסיסית להתרחשות תאונה בתדירות מסויימת (במקטע דרך מסויים). היסטורית, תדירות תאונות דרכים, אופיינה בפיזור פואסון. זו מניחה כי הממוצע והשונות בהתפלגות, דומים. מחקרים הראו כי ההבדלים בין תדירות תאונה לבין מודל מנבא, אשר נעשה בו שימוש במודלים המבוססים בפזור פואסוני, אינם עקביים. המחקרים העדכניים מניחים כי פיזור בינומי שלילי משקף בצורה טובה יותר את פיזור תדירות התאונות.

אחד הפרמטרים המשמשים על מנת לוודא כי הפיזור שנבחר מתאים, הינו פרמטר ה-**overdispersion** [להלן: ϕ]. מידע יחשב כבעל פיזור יתר ("overdispered") במידה ושונות המשתנה התלוי גדולה יותר מהממוצע.

במחקרם מציגים Powers & Carson [2004], שתי שיטות להערכת פרמטר ה-overdispersion. **הראשונה**, מתחשבת באורך הסגמנט ובוחנת את סוג הפיזור המאפיין את ההסתברות להתרחשות תאונה בתדירות מסויימת. הם מדגישים כי במידה ועסקינן בקטעי כביש המאופיינים בפרמטרים זהים (דוגמת: אורך ומאפיינים גיאומטריים נוספים זהים), ניתן היה ליישם פרמטר overdispersion כללי עבור כלל הסגמנטים. אולם, מאחר ולרוב קטעי דרך שונים באורכם ובמאפיניהם, נדרש לחשב פרמטר overdispersion ייחודי עבור כל סגמנט. זאת, כאשר בכדי להעריך בצורה טובה יותר במודל, את מספר התאונות הצפוי במקטע בודד, ניתן להתאים את פרמטר ה-overdispersion בהתבסס על על אורך הסגמנט לפי משוואה מספר 27.

$$\phi_i = \phi \cdot L_i^{\beta} \qquad [27]$$

כאשר:

ϕ_i - פרמטר overdispersion מותאם לסגמנט i

ϕ - פרמטר overdispersion כללי , עבור כלל הסגמנטים.

L_i - אורך קטע הכביש

β – קבוע הנע בין 0 ל- 1. פרמטר זה מביא בחשבון את ההבדלים הנובעים ממאפיינים גיאומטרים ומאפייני התנועה בין כל סגמנט. במידה ובין מקטע כביש אחד לחברו, לא היה כלל דמיון (אף אחד ממאפיני הכביש לא היה דומה לשני), אזי β=0, והסגמנט היה מיוצג במשוואה על ידי פרמטר ה- overdispersion הכללי. ולהיפך, כאשר כלל קטעי הכביש מאופיינים בדיוק באותם מאפיינים β =1. Powers & Carson [2004] הניחו במחקר כי β =1 , בכדי לספק את ההערכה השמרנית ביותר להתרחשות תאונות עתידיות.

השיטה **השניה**, מניחה פיזור גמא עבור כל מקטע כביש, כאשר פרמטר ה- overdispersion האינדיוודואלי בשימוש במתודה חלופית זו ניתן לחישוב כמופיע בנוסחה 28, המוצגת להלן:

$$[28] \quad \phi_i = \phi \cdot SPF_i^{\gamma}$$

כאשר:

γ - קבוע הנע בין 0 ל- 1. במידה ו- $\gamma = 0$, אזי עסקינן במודל בינומי שלילי. אם γ גדול יותר מ- 0, אזי השונות של פיזור גמא יורדת ככל שה- SPF_i עולה.

Powers & Carson [2004] מציינים כי במספר מחקרים נעשה שימוש דווקא בשיטה השניה, כאשר שיטה זו, הניבה תוצאות מדוייקות יותר מהשיטה הראשונה שהוצגה, אם כי קביעת ערך ה γ הינה בעייתית. במחקרם בחרו לעשות שימוש בערך $\gamma = 1$.

השלב השלישי, הינו קביעת המשקל היחסי (α), שנעשה תוך שימוש בנוסחה 29:

$$[29] \quad \alpha_i = \frac{1}{1 + SPF_i / \phi_i}$$

לאחר מכן, בשלב הרביעי, נעשית קביעת מספר התאונות הצפוי (π), תוך שימוש בנוסחה מספר 30:

$$[30] \quad \pi_i = (\alpha_l) \cdot (SPF_l) + (1 - \alpha_l) \cdot (\lambda_i)$$

כאשר:

π_i - מספר התאונות הצפוי פר שלוש שנים על סגמנט i

λ_i - מספר התאונות שהתרחש בפועל בשלוש שנים על סגמנט i

השלב האחרון הינו קביעת אפקטיבית הטיפול/שיפור (θ). את אפקטיביות הטיפול, ניתן לבטא כהפרש היחסי בין מספר התאונות הצפוי לבין מספר התאונות שהתרחש בפועל. אולם אז, לא ניתן מענה הולם לחוסר וודאות הנובע מסיבות שונות, דוגמת: גודל דגימה קטן יחסית ; טיב מודל נמוך יחסית ; הנחות הנלקחות בעת בחירת פרמטר ה-Overdispersion והמשקל היחסי. לכן, במקום זאת, נלקח בחשבון אינדקס האפקטיביות θ_l. כאשר את השונות ניתן לחשב כמוצג בנוסחה מספר 31:

$$[31] \quad \sigma_i^2 = (1-\alpha_i)\pi_i = SPF_i\left[1+\frac{SPF_i}{\phi_i \cdot L_i}\right]$$

אזי, אינדקס האפקטיביות מוצג בנוסחה מספר 32:

$$[32] \quad \theta_i = \frac{\lambda_i / \pi_i}{1+(\sigma_i^2/\pi_i^2)}$$

El-Basyouny & Sayed [2010] מציינים כי אחד המשתנים המנבאים הנפוצים והמשמעותיים ביותר בבניית SPF הינו נפח התנועה, אם כי מראש ידוע כי קיימת אי וודאות בעת מדידתו. ידוע כי אי דיוק זה (Measurement Errors להלן: ME), מגדיל את ה - dispersion. במאמרם מציעים גישה חדשה, גישת MENB, המשלבת בין מודל המכונה ME (Measurement Errors) לבין מודל SPF מסוג בינומי שלילי. במחקרם הישוו בין תוצרי ניתוח של שתי השיטות בעת בחינת שני בסיסי נתוני מקטעי דרך בקולומביה הבריטית. תוצאות מחקרם מראה כי השימוש בשיטת MENB מוצדק כאשר השונות בנפחי התנועה בין השנים שנבחנו הינו גדול. אחרת, ניתן לעשות שימוש במודל מסוג בינומי שלילי רגיל.

כאמור, פרמטרים המשולבים במודל, שבמדידתם חוסר וודאות מסויים, מגדילים את פרמטר ה- dispersion. נפח התנועה, הינו דוגמא לפרמטר כזה. באופן אידיאלי, יש למדוד אותו עבור כל תקופת המחקר. בפועל, נפח התנועה נדגם לרוב רק עבור יום או יומיים, כאשר לאחר מכן מבוצע מיצוע של הנתונים. לעיתים אף מוסיפים עוד תצפיות ומעדכנים את המיצוע. פעולה זו יכולה לשפר את ההערכה, אולם לא לפתור לחלוטין את בעיית אי הדיוק. כך, במקום להשתמש בנוסחה הרגילה של המודל הבינומי השלילי, נעשה שימוש בנוסחה מספר 33:

[33] $$\ln(\mu_i) = \ln(\alpha_0) + \alpha_1 \ln(L_i) + \alpha_2 \ln(\overline{V}_i) + \sum_{j=1}^{m} \beta_j X_{ji} \Rightarrow i = 1,2,n$$

כאשר:

μ_i - מספר תאונות צפוי

L_i - אורך הסגמנט

$\overline{V}_i$ - נפח תנועה (AADT) מוערך עבור סגמנט בתקופת המחקר.

X_{ji} - משתנים המייצגים פרמטרים גיאומטריים שונים של הכביש

β ו α הפרמטרים של המודל.

בהמשך, נעשה שימוש במשוואות הבאות: מענה (34), חשיפה (35) וטעות המדידה (36), כמוצג להלן:

[34] $$\ln(\mu_i) = \ln(\alpha_0) + \alpha_1 \ln(L_i) + \alpha_2 \ln(F_i) + \sum_{j=1}^{m} \beta_j X_{ji} \Rightarrow i = 1,2,....n$$

[35] $$\ln(F_i) \sim N(\mu_F, \sigma_F^2) \Rightarrow i = 1,2,...n$$

[36] $$\ln(V_{it}) \sim N(F_i + \delta_t, \sigma_\varepsilon^2) \Rightarrow i = 1,2,...n, t = 1,2...., T$$

נוסחה 34 הינה נוסחת ה SPF הלוגריתמית רגילה, שמטרתה חישוב מספר התאונות הצפוי במקטע דרך (μ_i). נוסחה 35 עושה שימוש ב Fi, הממוצע (הלא ידוע) של ספירות נפח התנועה במהלך תקופת המחקר, עבור כביש i, כמשתנה לטנטי (משתנה סמוי)[12]. משתנה לטנטי הינו מושג תיאורטי, המייצג משתנה שאינו נמדד באופן ישיר, אשר אנו מנסים להסיק לגביו מתוך המשתנים הנצפים (אינדיקטורים). היות ומשתנה לטנטי אינו מדיד אלא מוסק על סמך מספר אינדיקטורים/מסבירים, הוא נמדד למעשה "ללא טעות".

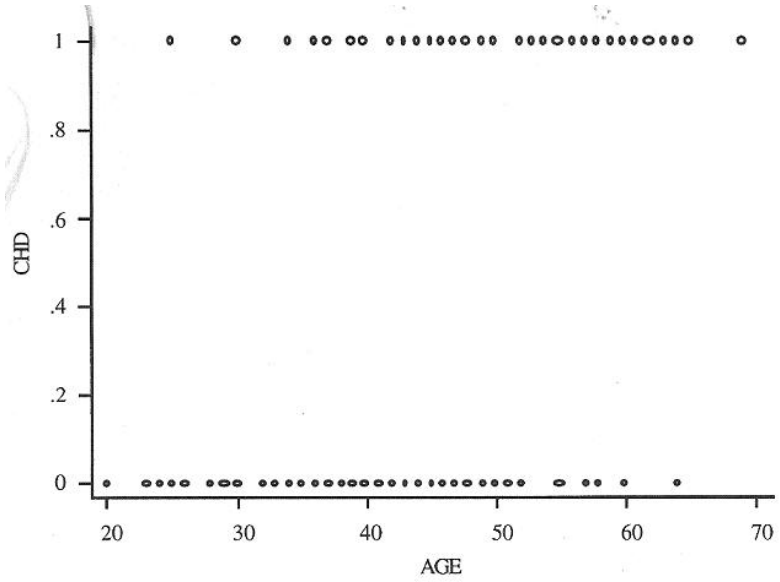
נוסחה 35 מתארת את הפיזור של גורם הסכנה הלא ידוע Fi באוכלוסיה כללית (עיר, איזור וכן הלאה). הפרמטר μ_F מייצג את כל ערכי ה- $\ln(F_i)$ הצפויים להתקבל בבסיס נתונים הומוגני, כאשר σ_F^2 מסמל את השונות מסגמנט לסגמנט. נוסחה 36, עושה שימוש במודל בנפח התנועה V_{it} עבור הכביש ה i בשנה מסויימת. δ_t - תיקון זמן ו- σ_ε^2 - השונות של טעות המדידה.

[12] *משתנה לטנטי מקור: http://www.bicohen.tau.ac.il/uploads/files/.pdf

2.3.3.5 רגרסיה לוגיסטית

הרגרסיה הלוגיסטית הינה למעשה מודל המתאר את היחס בין התוצאה (המשתנה התלוי) לבין סט שלם של משתנים מסבירים. רגרסיה זו, מתאימה עבור מידע המאופיין במשתנים בעלי מענה בינארי (קרי, המשתנה התלוי הינו דיכוטומי), כאשר במקום לנסות ולחזות ערכי 0 או 1 (משמע תוצאה של "כן/"לא"), מנסים להעריך מהי ההסתברות לכך שיתקבל אחד הערכים, לאחר שהוספנו מידע על היחסים בין המשתנים.

נניח כי אנו מעוניינים לחקור בקרב אוכלוסיית מדגם כלשהי את היחס שבין גיל ומחלות לב כליליות (להלן: CHD). המשתנה התלוי הינו CHD, כאשר קיום המחלה יקודד כ- 1 והעדרה יקודד כ- 0. כך, בהינתן משתנה דיכוטומי, יוצג היחס בין המשתנים בגרף הפיזור, המוצג באיור מספר 7:



איור מספר 7 - היחס בין מחלות לב כליליות (CHD) לבין גיל [Hosmer & Lemeshow, 2000]

ניתן לראות כי כל התצפיות מסתדרות בשני קווים מקבילים, כאשר Y=0 המשמעות היא כי היעדרות המחלה וכאשר Y=1 המשמעות היא כי המחלה קיימת. בנוסף, ניתן להסיק מהגרף, כי קיימת נטייה בקרב בני הגיל המבוגרים להראות תסמינים למחלה. אולם, על אף שהגרף מתאר את התוצאה הדיכוטומית של המשתנה התלוי, הוא אינו מספק תמונה ברורה מספיק על טבעו או מהותו של היחס בין מחלת ה CHD לבין הגיל. מתודה המאפשרת לקבל תמונה ברורה יותר, הינה יצירת אינטרוול עבור המשתנים הבלתי תלויים ("המסבירים"), תוך חישוב הממוצע עבור המשתנה התלוי ("התוצאה") בכל קבוצה. טבלה מספר 4 מציגה את משתנה הגיל המקורי (AGE) ובמקביל את עמודת AGRP, המחלקת לקטגוריות את משתנה קבוצת הגיל.

טבלה 4 - גיל ומחלות לב כליליות (להלן: CHD) [Hosmer & Lemeshow, 2000]

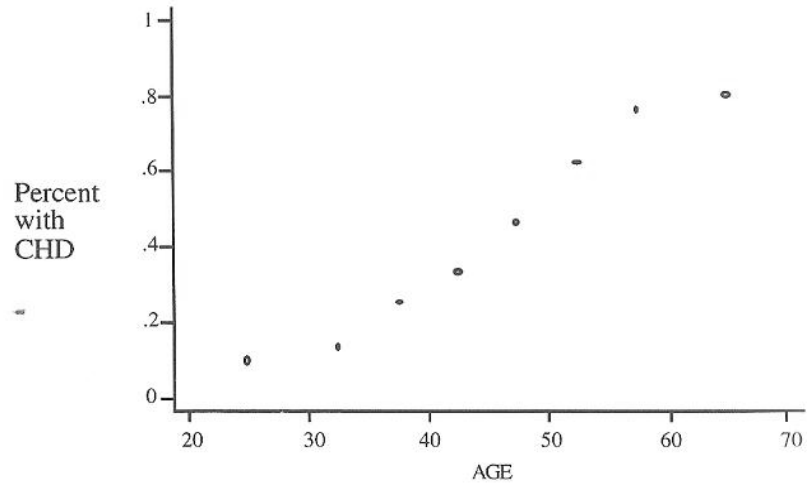
ID	AGE	AGRP	CHD	ID	AGE	AGRP	CHD
1	20	1	0	51	44	4	1
2	23	1	0	52	44	4	1
3	24	1	0	53	45	5	0
4	25	1	0	54	45	5	1
5	25	1	1	55	46	5	0
6	26	1	0	56	46	5	1
7	26	1	0	57	47	5	0
8	28	1	0	58	47	5	0
9	28	1	0	59	47	5	1
10	29	1	0	60	48	5	0
11	30	2	0	61	48	5	1
12	30	2	0	62	48	5	1
13	30	2	0	63	49	5	0
14	30	2	0	64	49	5	0
15	30	2	0	65	49	5	1
16	30	2	1	66	50	6	0
17	32	2	0	67	50	6	1
18	32	2	0	68	51	6	0
19	33	2	0	69	52	6	0
20	33	2	0	70	52	6	1
21	34	2	0	71	53	6	1
22	34	2	0	72	53	6	1
23	34	2	1	73	54	6	1
24	34	2	0	74	55	7	0
25	34	2	0	75	55	7	1
26	35	3	0	76	55	7	1
27	35	3	0	77	56	7	1
28	36	3	0	78	56	7	1
29	36	3	1	79	56	7	1
30	36	3	0	80	57	7	0
31	37	3	0	81	57	7	0
32	37	3	1	82	57	7	1
33	37	3	0	83	57	7	1
34	38	3	0	84	57	7	1
35	38	3	0	85	57	7	1
36	39	3	0	86	58	7	0
37	39	3	1	87	58	7	1
38	40	4	0	88	58	7	1
39	40	4	1	89	59	7	1
40	41	4	0	90	59	7	1
41	41	4	0	91	60	8	0
42	42	4	0	92	60	8	1
43	42	4	0	93	61	8	1
44	42	4	0	94	62	8	1
45	42	4	1	95	62	8	1
46	43	4	0	96	63	8	1
47	43	4	0	97	64	8	0
48	43	4	1	98	64	8	1
49	44	4	0	99	65	8	1
50	44	4	0	100	69	8	1

טבלה מספר 5 מציגה עבור כל קבוצת גיל את תדירות קיום המחלה. זאת, בנוסף לממוצע (שיעור נוכחות ה- CHD בקרב אותה קבוצה):

טבלה 5 - תדירות קיום מחלת CHD בקרב קבוצות הגיל השונות [Hosmer & Lemeshow, 2000]

Age Group	n	CHD		Mean (Proportion)
		Absent	Present	
20 – 29	10	9	1	0.10
30 – 34	15	13	2	0.13
35 – 39	12	9	3	0.25
40 – 44	15	10	5	0.33
45 – 49	13	7	6	0.46
50 – 54	8	3	5	0.63
55 – 59	17	4	13	0.76
60 – 69	10	2	8	0.80
Total	100	57	43	0.43

מהטבלה ניתן להסיק כי ככל שעולה הגיל, שיעור המראים תסמינים למחלת ה CHD גדל. ניתן לראות זאת באופן ברור על ידי הצגת דיאגראמת הפיזור של טבלה מספר 5, בגרף המוצג באיור מספר 8:



איור מספר 8 - שיעור נוכחות המחלה ביחס לגיל [Hosmer & Lemeshow, 2000]

תוצר הגרף שהוצג באיור מספר 8, אפשרי כי יהיה דומה לזה שיתקבל במידה ואותו תהליך של "קיבוץ" ו"מיצוע" יבוצע ברגרסיה ליניארית. זאת, כאשר ההבדל העיקרי הינו טבע היחסים בין המשתנה התלוי למשתנים הבלתי תלויים.

Hosmer & Lemeshow [2000], מציינים כי בכל בעיית רגרסיה מפתח המידה הינו הערך הממוצע של המשתנה הנחקר (התלוי) בהינתן הערכים של המשתנים המסבירים (הבלתי תלויים). מפתח זה מכונה ממוצע מותנה (conditional mean), ונהוג לסמנו $E(Y \mid x)$ כאשר, Y הערך הצפוי להתקבל בהינתן משתנה מנבא x . ברגרסיה ליניארית אנו מניחים כי ניתן להציג את הקשר כפי שמופיע בנוסחה מספר 37 :

$$[37] \qquad E(Y \mid x) = \beta_0 + \beta_1 x$$

כאשר:

β_0 קבוע המשוואה

β_1 המקדם של המשתנה המנבא (או המשתנים המנבאים).

אולם על פי המודל הליניארי קיימת האפשרות לקבלת ערכי הסתברות מעל 0 ומתחת ל-1 (עבור הפונקציה), בעוד שהסתברות חייבת "ליפול" בטווח שבין 0 ל-1. כך עולה הצורך לתאר את את היחס בין Y ל- X תוך שימוש בפונקציה שאינה קו ישר אלא בפונקציה בה הקו עקום (למעשה פונקצית Log, המאופיינת בצורת S).

לאור האמור, נסמן $E(Y \mid x) = \pi(x)$ בכדי להציג את הממוצע המותנה של Y בהינתן משתנה מנבא x , המודל הלוגיסטי מוצג בנוסחה מספר 38 (עבור משתנה מנבא אחד):

$$[38] \qquad \pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 X}}$$

ניתן לבצע טרנספורמציה של $\pi(x)$, כך שיעשה שימוש ב- log הטבעי (ln). הטרנספורמציה מכונה logistic transformation או בקיצור logit. נוסחה 38 שוות ערך לנוסחה מספר 39:

$$[39] \quad \ln\left(\frac{\pi(\mathrm{x})}{1-\pi(\mathrm{x})}\right) = \log it[\pi(\mathrm{x})] = \beta_0 + \beta_1 x$$

הבדל נוסף, בין הרגרסיה הלוגיסטית לליניארית, עוסק בתנאי הפיזור של המשתנה התלוי. ברגרסיה ליניארית אנו מניחים כי את תוצאות המשתנה התלוי ניתן יהיה להביע באמצעות נוסחה 40 :

$$[40] \quad y = E(Y \mid x) + \varepsilon$$

כאשר ε הינו "מרכיב השגיאה" , המייצג את הסטייה או הפער שבין הערך הנצפה של Y לבין הערך הממוצע הצפוי (הממוצע) לפי תוצאות הרגרסיה. ההנחה היא כי מרכיב השגיאה הינו בעל התפלגות נורמאלית ושונות קבועה עבור כלל המשתנים הבלתי תלויים (משמע, פיזור שווה של Y סביב כל ערך של X).

הנחות אלה <u>אינן תקפות</u> עבור מודל של רגרסיה לוגיסטית, בה המשתנים התלויים הינם דיכוטומיים. במצב זה יוגדר ערך המשתנה התלוי בהינתן x כמופיע בנוסחה 41:

[41] $y = \pi(x) + \varepsilon$

כאשר השונות של ε מוצגת בנוסחה מספר 42 :

[42] $\pi(x)[1-\pi(x)]$

כך לדוגמא, בהנחה כי π משמעו הסתברות להצלחה, והיחס $\pi(x)/1-\pi(x)$ שווה לסיכויים להצלחה. כך לדוגמא, במידה וסיכוי היחס להצלחה של מובטלים למצוא עבודה יהיה שווה ל 0.20, לאחר שנציב בנוסחה (p(x)/(1- p(x)) , נקבל את הערך 0.25, שמשמעו כי הסיכוי להצלחה (מציאת עבודה בדוגמא שלנו) גדול פי ארבע מהסיכוי להיכשל.

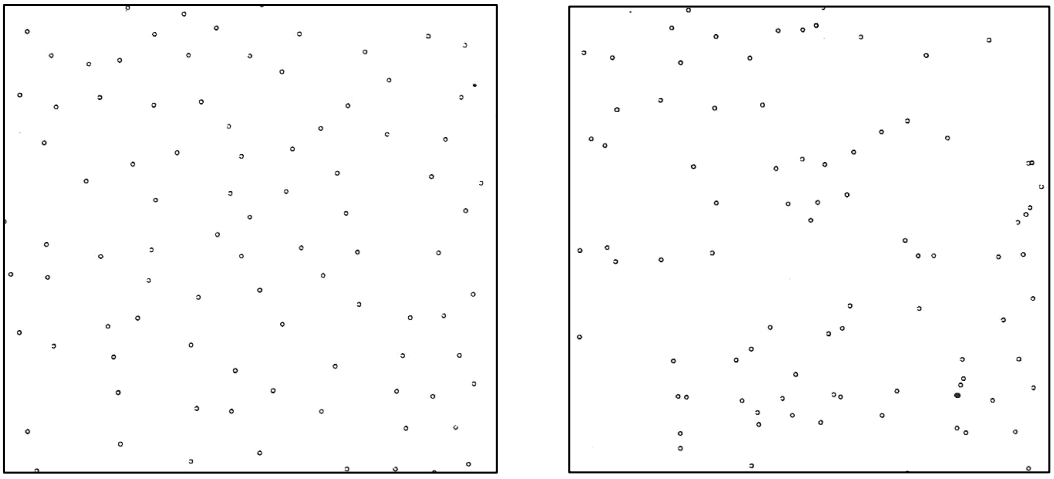
בבחינת מודל ליניארי ניתן להעריך את הפרמטרים β ו- α על ידי שימוש בשיטת הריבועים הפחותים, המאפשרת לקבוע מהו הקו הישר המתאים ביותר עבור סט הנתונים. בבסיס השיטה מונח הרעיון להביא למינימום את סכום ריבועי המרחק (האנכי) בין כל התצפיות Y לבין קו הרגרסיה. כלומר, נמצא ערכים של הפרמטרים אשר יגדירו קו מגמה ליניארי שהוא הקרוב ביותר, סימולטנית, לכל התצפיות Y. עם זאת, בבואנו לבחון משתנה תלוי דיכוטומי, יש להעריך את המקדמים β ו- α על ידי שימוש בפונקצית הנראות המקסימאלית (Maximum Likelihood Method) [Hosmer & Lemeshow, 2000]. בפונקציה זו, המקדמים להם הסבירות הגבוהה ביותר להיות הקרובים ל"מציאות" או לתוצאה האמיתית, יבחרו. את יחס ה"חשיבות" בין המשתנים ניתן להעריך על ידי שימוש במבחן Wald [Chatterjee et-al, 2000].

2.3.4 תבניות מרחביות של נקודות (Spatial Point Patterns) ואישכול (Clustering).

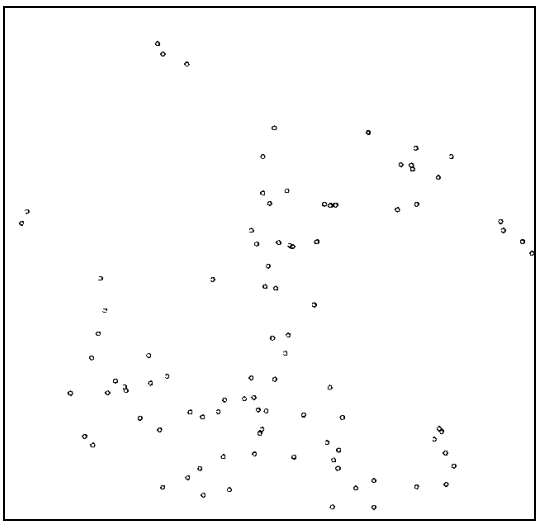
2.3.4.1 כללי

תאונת דרכים מושפעת מגורמים רבים, ויכולה להתרחש בהסתברות מסוימת בכל מקום ובכל עת. לרוב, ניתוח הפיזור המרחבי של תאונות דרכים, מתבסס כמעט בכללותו על בחינה ויזואלית. תהליך ההערכה והניתוח של זיהוי התבניות המרחביות, הינו תהליך סובייקטיבי. מכאן שצופים שונים עלולים שלא להסכים על קיומן של תבניות בפיזור המאורעות. כך, בכדי לזהות את היקף האירועים החורגים מאקראיות מרחבית מלאה, נדרש לעשות שימוש בטכניקות סטטיסטיות.[Nicholson, 1999 ; Cressie, 1993].

לעיתים תבניות המאופיינות באקראיות מרחבית מלאה, עשויות להידמות כאשכולות ("צבירים") לעין בלתי מיומנת. איור מספר 9 מציג שלוש דוגמאות שונות לתבניות מרחביות:



A אקראי B - פזור



C – מקובץ ("מאושכל")

איור מספר 9 - דגמי פיזור של נקודות

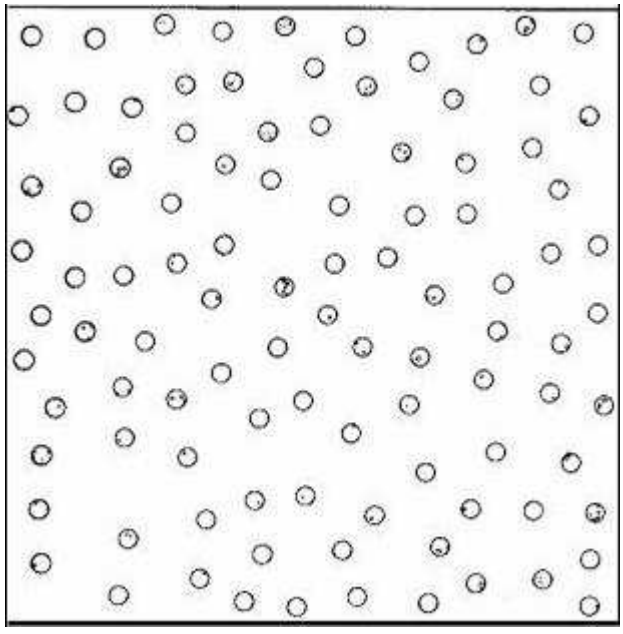
על מנת לגלות מהו "הסוג הנכון" של הפיזור המרחבי, קיימות שתי גישות עיקריות. הראשונה הינה גישת ה- Quadrats והשנייה מבוססת על בחינת המרחק בין נקודות במרחב. גישות אלה תפורטנה בסעיפים הבאים.

2.3.4.2 שיטת ה-"Quadrats" (שיטת "הרביעים")

בבסיס השיטה, ספירות "אירועים" בשטח מסוים. שמה ניתן לה, היות ולרוב נהוג לחלק את אזור הדגימה לריבועים, על אף שניתן עקרונית לחלק את שטח המחקר לכל צורה

בה נחפוץ. בשלב הראשון, על פי שיטה זו, תחת ההנחה כי האזור המאופיין באירועי תאונות חולק ל"רשת ריבועית", נסכם מספר תאונות הדרכים בכל ריבוע בגריד. במידה והתהליך המרחבי אקראי לחלוטין, נצפה כי סכום התאונות בכל "משבצת", יאופיין בהתפלגות פואסון (ראה פרק 2.3.3.2). בשלב הבא נבחן האם קיימת אקראיות בפיזור המרחבי של תאונות הדרכים בכל אחד מריבועי הרשת. זאת, ניתן יהיה לבחון על ידי שימוש במבחן סטטיסטי דוגמת קולמוגורוב-סמירנוף. חולשתה של שיטה זו, טמונה בכך שהתוצאות תלויות לעיתים בקבלת החלטות שרירותיות הקשורות לגודל הגריד, צורתו ומיקומו [Nicholson, 1999 ; ; Diggle, 1983 ; Cressie, 1993]. יש לציין כי אם ניתן לזהות את קנה המידה שבו דגם מסויים סוטה באופן מרבי מאקראיות, אזי ניתן להשתמש בידע זה בכדי להגיע להבנה מעמיקה יותר של מבנה דגם ההתפלגות. מומלץ לנסות ולעשות שימוש במספר קני מידה טרם ביצוע האינטרפולציה [קלרמן, 1986].

בכדי לבחון האם סט הנתונים "מקובץ", יש לבחון תחילה האם אינו מאופיין באקראיות מרחבית מלאה. היות ולמחקרים העוסקים בתבניות מרחביות של נקודות, היסטוריה רבה בתחומי האקולוגיה (בייחוד ייעור), הגיאוגרפיה, הקוסמולוגיה והארכיאולוגיה, נדגים את השיטה על סט נקודות של עצים. איור מספר 10 מציג 100 אזורי דגימה מעגליים המאופיינים ברדיוס של 6 מ'. מיקום אזורי הדגימה, נבחר כך שלא תתקיים חפיפה בין אזור אחד למשנהו.



איור מספר 10 - אזורי דגימה מעגליים המאופיינים ברדיוס של 6 מ'

מספר העצים בכל אזור דגימה נסכם. טבלה מספר 6 מציגה את תדירות פיזור העצים באזורי הדגימה, ברדיוס של 6 מטרים.

טבלה 6 - תדירות הפיזור של עצים באזורי הדגימה, ברדיוס של 6 מטרים [Cressie, 1993]

Trees per Quadrat	Observed Frequency	Expected Frequency
0	34	23.93
1	33	34.22
2	17	24.47
3	7	11.66
4	3	4.17
5	1	1.19
6	1	0.28
7	2 } 6	0.06 } 1.54
8	1	0.01
9	0	0.00
10	1	0.00

[a]Observed frequencies are compared to expected Poisson (mean = 1.43) frequencies.

כאמור, תחת אקראיות מרחבית מלאה, מספר העצים באזור הדגימה, יאופיין בהתפלגות פואסונית. התדירויות האמיתיות שנצפו, הושוו לתוצאות שאמורות היו להתקבל במידה והייתה מתקיימת התפלגות פואסון. כך, בכדי לבחון האם מדובר באקראיות מרחבית מלאה בוצע מבחן Chi-Squared (ראה פרק 2.3.31), ובעקבות התוצאות שהתקבלו, ניתן היה לומר בוודאות כי העצים אינם מאופיינים באקראיות מרחבית מלאה, אלא מקובצים לאשכולות.

לאחר שהנחת האקראיות המלאה לגבי סט הנקודות, נדחתה, ניתן לבחון האם הנתונים מקובצים. למטרה זו קיימים בספרות מספר אינדקסים המוצגים בטבלה מספר 7, להלן (נסביר את השנים הראשונים):

טבלה 7 - אינדקסים עבור בחינת פיזור [Cressie, 1993]

Index	Estimator	Realization	Reference
I	$\frac{S^2}{\bar{X}}$	2.335	Fisher et al. (1922)
ICS	$\frac{S^2}{\bar{X}}-1$	1.335	David and Moore (1954)
ICF	$\frac{\bar{X}^2}{S^2-\bar{X}}$	1.072	Douglas (1975)
$\dot{X}$	$\bar{X}+\frac{S^2}{\bar{X}}-1$	2.765	Lloyd (1967)
IP	$\frac{\dot{X}}{\bar{X}}$	1.933	Lloyd (1967)
I_δ	$\frac{n\sum_{i=1}^{n}X_i(X_i-1)}{n\bar{X}(n\bar{X}-1)}$	1.930	Morisita (1959)

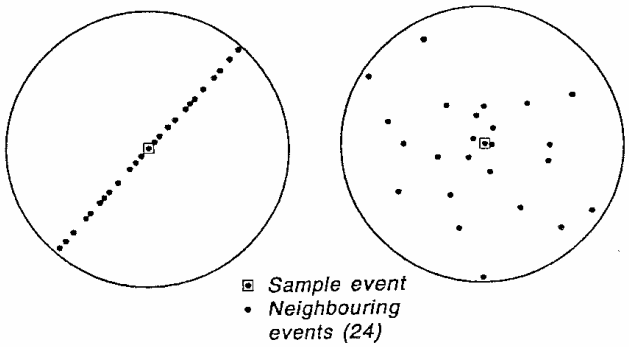
[a]X_i is the number of trees in the ith quadrat, $\bar{X}$ is the sample mean of the quadrat counts, an S^2 is the sample variance.

עבור האינדקס הבסיסי I (המייצג את השונות היחסית) ככל שערכו גדול יותר מ – 1, המשמעות הינה כי סט הנקודות הנבחן מקובץ לאשכול. לגבי אינדקס ICS, במידה וערכו

גדול מ – 0 מרמז הדבר כי סט הנקודות הנבחן מקובץ לאשכול, כאשר במידה וערך האינדקס קטן מ - 0, המשמעות הינה, שמדובר בפיזור רגיל. אינדקס זה מושפע מאוד מגודל אזור המדגם. במידה ואזורי הדגימה (Quadrats) גדולים ביחס לאזור בו מתרכזים כלל האירועים ("האזור המקובץ"), ובמידה ומספר האירועים באשכול (cluster) מאופיינים בפיזור פואסון עם ממוצע μ, אזי אינדקס ICS שווה בקירוב ל μ [Cressie, 1993].

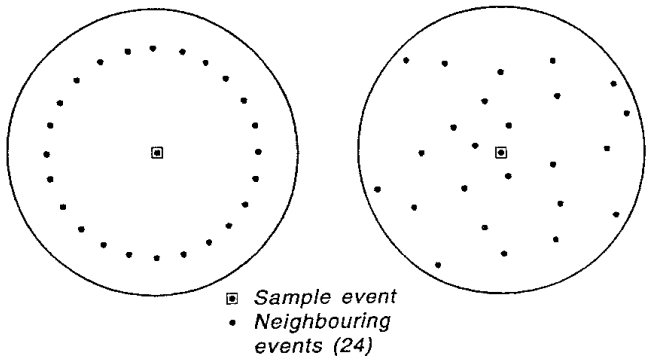
2.3.4.3 מתודת מרחק - השכן הקרוב ביותר (Nearest –neighbor)

השיטה עושה שימוש במידע מדוייק בנוגע למיקומי תאונות, ומתעלמת ממידע בנוגע לגודל וצורה. השיטה בוחנת "מרחק" בלבד ומאתרת את ה"שכן" (האירוע) הקרוב ביותר לפיו. כך, בבואנו לבחון רק את ה"מרחק" בין שני מדגמים מרחביים שונים, אחד המאופיין בדוגמא קווית והשני בפיזור שונה (ראה איור 11), אנו עלולים לקבל תוצאות דומות. אולם, אם בנוסף נבחן גם את הכיווניות, נקבל שתי תוצאות שונות לחלוטין.



איור מספר 11 - פיזור מרחבי של נקודות הנמצאות במרחק שווה מאירוע נבחן [Nicholson, 1999].

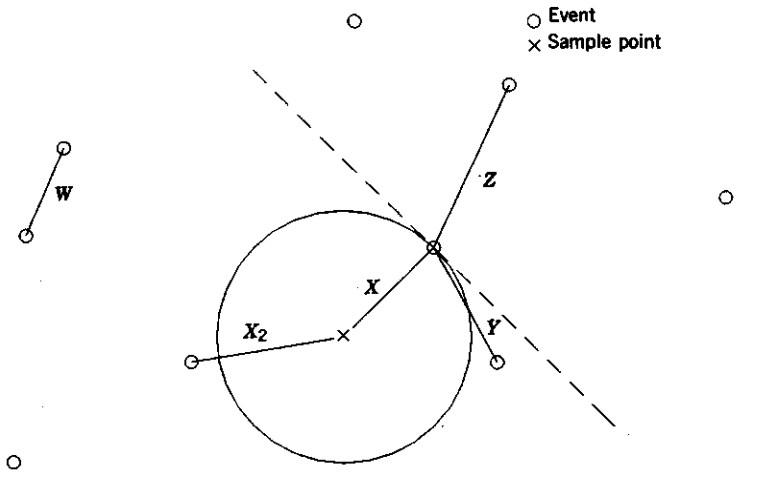
במקרים מסוימים, למשל בפיזור בצורת מעגל, במידה ונבחן רק את הכיווניות, ללא שימת דגש על המרחק, נוכל לקבל תוצאות, שאינן משקפות בהכרח בצורה המיטבית את המציאות (איור מספר 12).



איור מספר 12 - פיזור נקודות המאופיין בכיווניות דומה [Nicholson, 1999].

לכן, בהתאם לאמור, יש לבחון הן את הכיוון והן את המרחק [Nicholson, 1999].

<u>בהיבט ניתוח המרחק</u> - ניתן למדוד מרחק בין אירוע לאירוע (event to event) או מנקודה מסוימת לאירוע (point to event), בצורות שונות, כמוצג באיור מספר 13 [Cressie , 1993]:



איור מספר 13 - אופני מדידת מרחק שונים [Nicholson, 1999].

- בין אירוע לבין האירוע הקרוב ביותר, הנמצא בשכנות אליו (**W**)
- בין נקודת דגימה נבחרת, לבין האירוע הקרוב ביותר, הנמצא בשכנות לנקודת הדגימה (**X**). נקודות הדגימה בדרך כלל ממוקמות אחראית באזור המחקר.
- בין נקודת דגימה נבחרת לבין האירוע השני הקרוב ביותר, הנמצא בשכנות לנקודת הדגימה (**X_2**).
- בין האירוע הקרוב ביותר לנקודת דגימה מסוימת לבין השכן הקרוב ביותר של אותו אירוע מסוים (**Y**). המרחק Y אינו בלתי תלוי במרחק הרדיוס X. זאת, היות ושום

אירוע לא יכול להתרחש בתחום רדיוס המעגל הסובב את נקודת הדגימה. שהרי מדובר במרחק אל האירוע הקרוב ביותר, לכן הוצעה שיטה נוספת T- Square.

- שיטת ה- T- Square - בבסיסה חלוקת האזור לשני חצאי מישורים. בשיטה זו נמדד מרחק **Z** בין האירוע הקרוב ביותר לנקודת דגימה מסוימת לבין האירוע הקרוב ביותר. זאת, ללא תלות במרחק X.

קלרמן [1986], מציין כי מתודת המרחק מהשכן הקרוב ביותר הינה טכניקה המאפשרת מדידת פיזור במרחב, כאשר ממוצע המרחקים מכל נקודה אל הנקודה הקרובה ביותר, מושווה לערכים מיוחלים המייצגים שלוש אפשרויות של פיזור נקודות: אקראי, מקובץ או סדיר. המרחק מהשכן הקרוב ביותר R מוגדר בנוסחה מספר 43:

$$R = \frac{\bar{r}_0}{\bar{r}_e} \qquad [43]$$

כאשר:

$\bar{r}_0$ - הממוצע (הנצפה) של מרחקי כל נקודה מהנקודה הקרובה אליה ביותר.

$\bar{r}_e$ - הערך המיוחל (חישובו יוסבר בהמשך).

המרחק הממוצע ניתן לחישוב על ידי שימוש בנוסחה 44:

$$\bar{r}_0 = \frac{\sum_{i=1}^{n} d_{ij}}{N} \qquad [44]$$

כאשר:

d_{ij} - במרחק הישר בין נקודה כלשהי i , לבין הנקודה הקרובה ביותר אליה. הערך ניתן לחישוב על ידי שימוש במשפט פיתגורס. הערך המיוחל מהשכן, מוגדר בנוסחה מספר 45:

$$\bar{r}_E = \frac{1}{2\sqrt{\lambda}}, \qquad \lambda = \frac{N}{A} \qquad [45]$$

כאשר :

N- מספר הנקודות

A - שטח

ערכו של R עשוי לנוע בין הערך 0 לבין הערך 2.1491. ערך של R=0 מציין התלכדות מוחלטת של כלל הנקודות בנקודה אחת ; ערך R=1 מציין תפרוסת אקראית; בעוד שהערך R=2.1491 מציין פיזור מרבי [קלרמן, 1986].

Nicholson [1999], מציין כי כאשר קיים מדגם אינטנסיבי המתמקד באזור מצומצם יחסית מכלל אזור המחקר (הכוונה לכך שמספר הנקודות הנדגמות באזור מסוים, גדול יותר מ - 10% ממספר האירועים הכולל במדגם), אזי מדידות שנבחנות על ידי שיטת ה"שכן הקרוב ביותר" אינן בלתי תלויות. כל הניתוחים הסטטיסטיים שתוארו על ידי Cressie [1993], מחשבים למעשה את המרחק בין ה"שכן הקרוב ביותר" הראשון והשני, כאשר בפועל בכדי להבדיל בין תבניות במרחב, קיים צורך לחשב מרחק לשכנים מסדר גבוה יותר. כך מסיק Cressie, כי אף על פי שהניתוחים הסטטיסטיים מספקים מידע שימושי, אין באלה בכדי להספיק לבחינת מידע ממופה. במקרה זה מוצעת שיטה אלטרנטיבית של שימוש בפונקצית K (ראה פרוט בסעיף 2.3.4.4)

<u>בהיבט ניתוח הכיוון</u> - טווח הכיוון לשכן הקרוב ביותר נע בין 0º ל 360º, ולפיכך נדרשת טכניקה מיוחדת לתיאור וניתוח המידע המעגלי. כל תצפית בכיוון Qi תחשב שוות ערך, לווקטור בכיוון Qi, כאשר גודלה של הזוית יחושב ביחס לזווית ה - 0º. כלומר, האזימוט ביחס לכיוון הצפון. לפיכך, במידה ובמדגם קיימים N כיוונים שונים $Q_1,Q_2 \ldots Q_N$, הווקטור האקוויוולנטי של האורך (R) יחושב על פי נוסחה מספר 46:

$$[46] \qquad R= \left\{\left[\left(\sum_i \cos Qi\right)^2 + \left(\sum_i \sin Qi\right)^2\right]\right\}^{0.5}$$

R שווה ל – 0, כאשר קיים מידע המבטל את עצמו לחלוטין, ושווה ל- N, כאשר כלל המידע מקביל.

חישוב R הינו בעל משמעות במצב בו קיים ריכוז תצפיות אונימודלי (uni-modal). לדוגמא, במידה וקיימים שני אזורים המאופיינים בתדירות גבוהה של תאונות (mode) והנמצאים במרחק של ד האחד מהשני, הרי שערך R יהיה נמוך. זאת, מאחר והוקטורים יבטלו האחד את השני. עם זאת, במידה וקיימת סברה, ולפיה עסקינן בתצפיות בהן קיימים K ערכים בעלי תדירות או שכיחות גבוהה, המפוזרים באופן שווה במרחב (multi-modal), יש ל"עבד" את המידע כך שהווקטורים לא יבטלו האחד את השני, ולפיכך יחושב הכיוון על ידי נוסחה מספר 47:

[47] $$\overline{R_k} = \left\{\left[\left(\sum_i \cos kQi\right)^2 + \left(\sum_i \sin kQi\right)^2\right]\right\}^{0.5} / n$$

כאשר:

n - גודל הדגימה.

קיימים מספר מבחנים סטטיסטיים המתבססים על פונקציות של פיזור מצטבר, שאינם מתאימים לניתוח מידע מעגלי. זאת, לרבות מבחן קולמוגורוב-סמירנוף, בו אפשר והכיוון יקבל את הערך אפס. בכדי לנתח מידע מסוג זה מוצעים מספר מבחנים נוספים, דוגמת מבחן Rayleigh ו- Kuiper. מבחן Rayleigh, בוחן האם ערך R גדול דיו בכדי להצדיק דחיית ההיפותזה בנוגע לאחדות. מבחן Kuiper, משלב השוואה בין הפיזור המצטבר של פונקציה Qi, עם פונקצית הפיזור האחוד, ומהווה גרסה משופרת למבחן קולמוגורוב-סמירנוף [Nicholson, 1999].

2.3.4.4 פונקציית K

טכניקות מוקדמות הקשורות לניתוח תבניות מרחביות של נקודות לא פותחו תחת מדע הסטטיסטיקה, אלא דווקא עבור שימוש בתחומי מדע אחרים דוגמת: אקולוגיה (בייחוד בתחום הייעור), גיאוגרפיה, קוסמולוגיה וארכיאולוגיה. נקודת מפנה, מהווה מאמרו של RIPLEY [1977], המתאר את פיתוחה של פונקצית K שהוצעה במקור על ידי BARTLETT בשנת 1964 [Cressie, 1993]. יתרון השימוש בפונקציית K, הוא היכולת לבחון תבניות של נקודות המאופיינות בקני מידה שונים, תוך בחינת כל אירוע ללא צורך בחלוקה וביצוע אגרגציה לאזורים. הפונקציה נמצאת בשימוש נרחב בתחומי מחקר שונים, דוגמת: פיזור אוכלוסיה, התנהגות בעלי חיים, פיזור חולי סרטן, מודלי תמותה ועוד [Yamada & Thill, 2004].

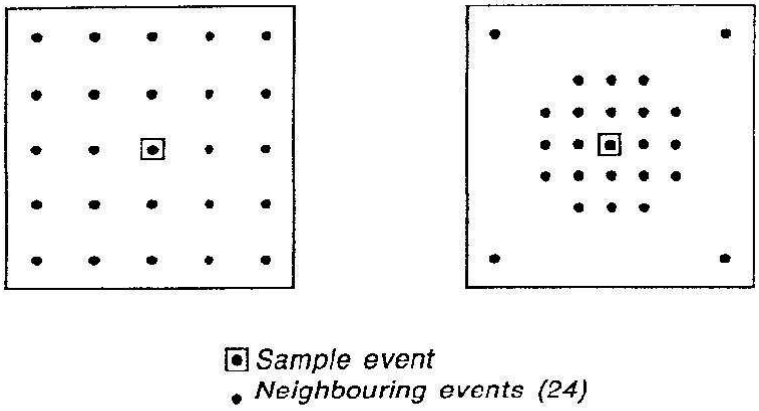
פונקציית K מבוססת על בחינת מרחקים בין אירועים, או בין נקודות המפוזרות באופן אקראי, ומשמשת ככלי בתהליך ניתוח מידע נקודתי "ממופה" [Dixon, 2002] , כאשר נוסחת הפונקציה מוצגת בנוסחה מספר 48:

[48] K(h) = E [number of extra events within distance h of a randomly selected event]/λ

כאשר λ מסמלת את צפיפות (אינטנסיביות)התהליך המרחבי (ביטוי למספר אירועים ליחידת שטח). עבור אקראיות מרחבית מלאה, ערך הפונקציה K(h) הינו πh^2. עבור אירועים המפוזרים באופן שווה הפונקציה נוטה לקבל ערך הנמוך מ πh^2, ועבור אירועים

המאופיינים בפיזור אשכולות (clustered) ערך פונקצית K נוטה להיות גדול יותר מ- πh^2 [Nicholson, 1999].

כאמור, פונקצית K מאפשרת לעשות שימוש בשיטת "השכן הקרוב ביותר" ולהבחין בין תבניות שונות במרחב, במגוון קנה מידה. בשלב בו נצפה פיזור במרחק מסוים מאירוע נבחן, ואותרו N שכנים קרובים ביותר מאירוע נבחן כלשהו (בהתאם למרחק שהוגדר), קיימת חשיבות לבחינת השאלה: האם הפיזור שונה באופן משמעותי מהפיזור שהיינו מצפים לקבל, במידה והאירוע היה מאופיין בפיזור אקראי מרחבי מלא. בכדי לקבוע זאת, ניתן להריץ מבחן דוגמת קולמוגורוב -סמירנוף (Kolmogorov –Smirnov). חוזקו של מבחן קולמוגורוב -סמירנוף ביכולתו לקבוע שני סוגים של סטיות מאקראיות מרחבית מלאה (קיבוץ עודף לאשכולות או פיזור "רגיל") ובכך לקבוע האם שני סטים של מידע שונים באופן משמעותי האחד מהשני. יתרון נוסף של המבחן, הוא שאין צורך להניח מראש הנחות כלשהן לגבי פיזור המידע. איור מספר 14 מציג שני סטים שונים של פיזור מרחבי המתארים תאונת דרכים, כאשר סביבה 24 תאונות נוספות. האחד מאופיין בפיזור סדור והשני בריכוז באשכול.



איור מספר 14 - תבניות שונות של נקודות (מקובץ וסדור) [Nicholson, 1999].

הפיזור המצטבר במצב של אקראיות מרחבית, מוגדר בנוסחה מספר 49:

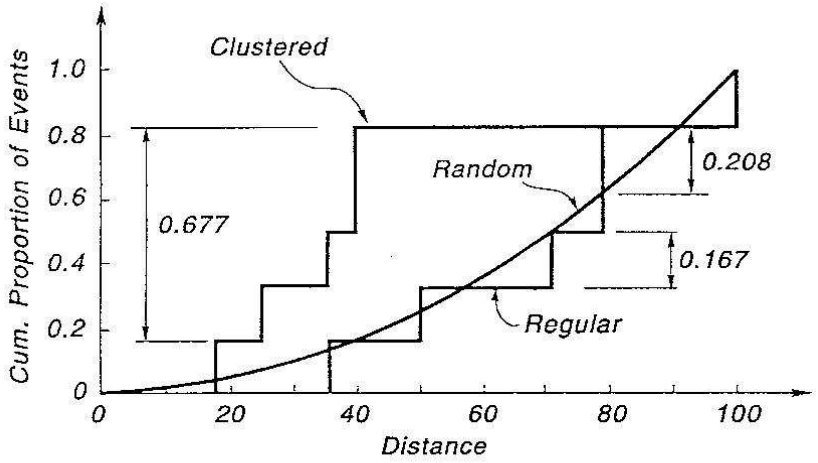
[49] $P(h_i) = (h_i / h_N)^2$

כאשר :

h_i - המרחק לשכן הקרוב ביותר של אירוע i , כאשר (i = 1,2,3,…N)

N - מספר השכנים קרובים ביותר לאירוע נבחן.

תוצר פונקצית הפיזור המצטבר של המרחק (המציגה את היחס הצפוי של N אירועים במרחק h_i) עבור תבנית סדורה, תבנית מאושכלת ותבנית אקראית מרחבית מלאה, מוצגת באיור מספר 15:



איור מספר 15 - פונקצית הפיזור המצטברת [Nicholson, 1999].

בדוגמא המוצגת באיור מספר 15, ניתן לראות כי תוצאות מבחן קולמוגורוב-סמירנוף , עבור תבנית מאושכלת ועבור פיזור סדור הינו 0.677 ו- 0.208 בהתאמה. תוצאה המצביעה על כך, שמבחינה סטטיסטית, האשכול משמעותי יותר [Nicholson, 1999].

Jones et-al [1996] העוסקים בגיאוגרפיה של הרפואה, שילבו גם הם במחקרם פונקצית K. זאת, היות והמתודה מאפשרת ניתוח המידע וחלוקתו לאשכולות (clustering) בסקאלות גיאוגרפיות נרחבות, תוך צמצום ההנחות התיאורטיות לגבי פיזור המידע. במחקרם עשו שימוש בנתוני תאונות דרכים שהתרחשו באזור הכפרי של Norfolk באנגליה בתקופה שבין השנים 1987 ל- 1991. אזור המחקר נבחר, מאחר ושיעור תאונות הדרכים באזור זה היה נמוך יותר מאשר הממוצע בכלל אנגליה, ושיעור התמותה היה גבוה הרבה יותר מהממוצע באנגליה. מטרת המחקר הייתה בחינת הגורמים השונים לשיעור התמותה הגבוה, תוך מתן דגש למשתנה "זמן הגעת האמבולנס" (חושב תוך שימוש ב GIS), בכדי לקבוע את הסבירות להישרדות בכל תאונת דרכים.

בשלב הראשון של המחקר הורצה רגרסיה לוגיסטית. המשתנה הבינארי התלוי הוגדר בכדי לאפשר מתן אינדיקציה בנוגע למשתנים המשפיעים על כל נפגע (חמור או קטלני). על פי תוצאות המודל, המשתנים המסבירים שנמצאו משמעותיים סטטיסטית לתמותה או לפציעה חמורה הם: גילאים הנעים בין 16-59 ; גיל 60+ ; כבישים בהם המהירות המותרת היא 40-50 מייל לשעה ; רוכבי אופניים ; רוכבי אופנוע ; תאונות שאופיינו בשני נפגעים ; תאונות שאופיינו בשלושה נפגעים ; תאונות שאופיינו בארבעה נפגעים ועוד.

במחקרם נמצא, כי קיימת סבירות גבוהה לתמותה בקרב: נפגעים מעל גיל 60, נפגעים המעורבים בתאונות שהתרחשו בכבישים בהם המהירות המותרת הינה לפחות 60 מייל לשעה, ובתאונות בהן היו מעורבים שלושה או יותר נפגעים.

באירועים שנמצאו חריגים, בהם חזה המודל מקרים קטלניים כאשר תוצאת האמת הייתה פגיעה חמורה, ובמקרים בהם חזה המודל פגיעה חמורה, כאשר תוצאת האמת הייתה פגיעה קטלנית, נבחן הפיזור הגיאוגרפי, תוך שימוש בפונקצית K, כמוסבר להלן.
נתייחס לסט הנקודות המפוזרות במרחב כאל N ואל אזור המחקר כאל A. כזכור, פונקצית K מאפשרת לקבוע את רמת ה"אישכול" (clustering) בעת השוואת סט נתונים אחד למשנהו. לפיכך, נדרשים שני סטים של נתונים: סט "אירועים" וסט "בקרה". במחקרם, סט ה"אירועים" הוגדר עבור תאונות בהם הנפגעים נפטרו על אף שעל פי מודל הרגרסיה צפוי היה כי ישרדו (קבוצה n_1), כאשר סט ה"בקרה" הוגדר עבור אירועים בהם נפגעים שרדו, בשעה שבהתאם לתוצאות המודל צפוי היה כי לא ישרדו (קבוצה n_2). כך, על ידי חישוב פונקציית K עבור שתי הקבוצות (n_1 ו - n_2) והשוואת התוצאות, ניתן היה לקבוע לאיזו קבוצה שייך אירוע מסוים.
לאחר שאירועי תאונות הדרכים מופו (בהתאם לקבוצה אליה השתייכו) ניתן היה לראות ויזואלית כי קיים ריכוז של תאונות של קבוצת n_1 בשלושה אזורים אורבאניים מרכזיים. תופעה זו מרמזת כי קיים גורם משפיע נוסף שלא שולב במחקר ובמודל הרגרסיה הלוגיסטית, ולפיכך מודל הרגרסיה אינו שלם. דוגמה אפשרית לכך – מבנה מערך הכבישים באזורים אורבאניים.

Jones et-al [1996] גם הציעו תיקון לפונקציית K, בכדי להופכה למדוייקת יותר ב"קצוות". במחקרם ציינו כי ניתן לתאר סט נקודות מרחבי, על ידי פונקציות מסדר ראשון (λ) ושני, כאשר פונקציה מסדר ראשון, עוסקת בצפיפות ובאינטנסיביות התהליך (מספר הנקודות חלקי השטח הנחקר), כמוצג בנוסחה 50, ופונקציה מסדר שני, עוסקת בפיזור המרחבי של אירועי תאונות הדרכים, כמוצג בנוסחה 51.

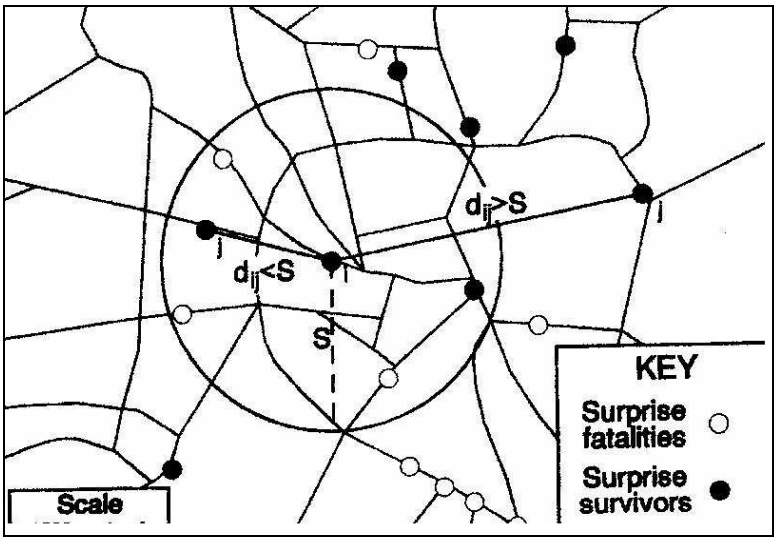
[50] $$\hat{\lambda}_j = \frac{n_j}{|A|}$$

j –סך אירועי תאונות הדרכים הסובבים אירוע נבחן

A - השטח הנחקר

[51] $$\lambda_1 \hat{K}_{11}(S) = \sum_{i=1}^{n_1} \left\{ \sum_{j \neq i} I(d_{ij} \leq S) \right\}$$

בכדי להבין את הפונקציה נתבונן באיור מספר 16:



איור מספר 16 - פיזור מרחבי של נקודות [Jones et-al, 1996].

כאשר:

I – אירוע תאונה נבחן

j –אירועי תאונות הדרכים הסובבים את האירוע הנבחן.

d_{ij} - המרחק האוקלידי בין האירוע הנבחן (*i*) לבין אירועים אחרים הסובבים אותו (*j*)

S – רדיוס איזור החיפוש הנקבע בין האירוע הנבחן לבין האירועים הסובבים.

I – אינדיקטור הפונקציה

משמעות הביטוי בסוגריים: במידה ויתקיים $d_{ij} \le S$, משמע האירוע "נופל" בתוך "אזור החיפוש" שהוגדר על ידי היחס בין גודל הרדיוס ומיקומו של I, אזי הפונקציה תחזיר את הערך 1 (I=1). במידה והתנאי אינו מתקיים תחזיר הפונקציה ערך השווה ל- 0. ה- Σ המסומן מחוץ לסוגריים, משמעו כי יש לחזור על נוסחה זו עבור כלל הנקודות בשני הסטים הנבחנים.

החישוב הבסיסי של פונקצית K כפי שהוצגה, נוטה לסבול מאי דיוקים בבחינת האירועים המתרחשים בגבול השטח הנבחן (Edge effect). זאת מאחר, שבמידה ונבחן אירוע הממוקם בקצה אזור המחקר, הרדיוס הנבחן יחרוג מגבול שטח המחקר, ולפיכך לא בהכרח ימצא מידע בטווח זה. קיימות מספר מתודות לתיקון אי הדיוקים בקצוות. אחת המתודות הנפוצות הוגדרה על ידי Ripeye. מתודה זו עושה שימוש במשקל (w_{ij}) שהינו פרופורציונאלי להיקף המעגל המוגדר על ידי רדיוס המרחק d_{ij} המוכל בתוך שטח

המחקר A. אם היחס בין היקף המעגל (המוגדר על ידי רדיוס המרחק d_{ij}) לבין השטח A גדול מאוד, ערך המשקל w_{ij} יהיה בקרוב 1.

נוסחה 52 מציגה את משוואת K לרבות תיקון הקצוות:

$$\lambda_1 \hat{K}_{11}(S) = \sum_{i=1}^{n_1} \left\{ \sum_{j \neq i} w_{ij} I(d_{ij} \leq S) \right\} \quad [52]$$

בהמשך יש לבצע תיקון נוסף, בגין העובדה כי דרגת חופש אחת, אבדה במהלך חישובי המרחק. זאת, היות וכאשר חושב המרחק בין אירוע תאונה בודד l, לבין שאר אירועי התאונות j, l לא נכלל בהערכה. בכדי לתקן את האמור, יש לחלק את תוצאות נוסחה 52 על ידי הנוסחה 53 :

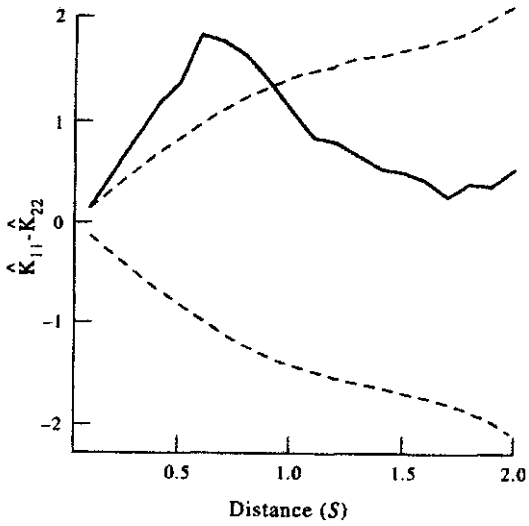
$$\frac{(n_1 - 1)}{|A|} \quad [53]$$

ולפיכך הנוסחאות הסופיות בהן עשו Jones et-al [1996] שימוש במהלך מחקרם 54, ו 55 מוצגות להלן:

$$\lambda_1 \hat{K}_{11}(S) = |A| \{n_1(n_1 - 1)\}^{-1} \sum_{i=1}^{n_1} \left\{ \sum_{j=1} w_{ij} I(d_{ij} \leq S) \right\} \quad [54]$$

$$\lambda_1 \hat{K}_{12}(S) = |A| \{n_2(n_2 - 1)\}^{-1} \sum_{i=n_j+1}^{n_1} \left\{ \sum_{j=n_1+1} w_{ij} I(d_{ij} \leq S) \right\} \quad [55]$$

ההערכה של פונקצית K צריכה להתבצע על מגוון טווחי S. במצב רגיל, הערכים מחושבים עבור 1/3 מהיקף אזור המחקר. עבור S גדול יותר, ההערכה אינה יעילה. לאחר הרצת התהליך עבור שתי הקבוצות, מתקבלים שני תוצרים. ההפרש בין שני התוצרים, יאפשר הצגת תוצר הפונקציה באופן גרפי, עבור מגוון ערכי ה – S הנבחנים. בכך יתאפשר לקבוע ולהשוות בין ה"אישכול" של שתי הקבוצות. לבסוף בכדי לבחון את המובהקות, יש לחשב את השגיאה הסטנדרטית ורווח הסמך (באמצעות סימולציית מונטה קרלו או בקומבינטוריקה). במחקרם בחרו לעשות שימוש בתוכנת SPLANCS, העושה שימוש בקומבינטוריקה. בשלב בו רווח הסמך חושב ניתן להוסיפו לגרף. ראה איור מספר 17.



איור מספר 17 - ערכי ההפרש K22 K11 עבור טווחי מרחקים נבחנים (עד 2 ק"מ).
רווח הסמך מסומך כקו מקווקו
[Jones et-al, 1996].

במידה ובערכי S מסוימים, ההפרש סוטה מהערך 0, ביותר מ ±2 סטיות תקן, ניתן לומר כי אחת מתוך שתי התבניות של הנקודות מקובצת לאשכול בהשוואה לקבוצה השנייה באותו הערך. בכדי שהקיבוץ עבור ערך S מסוים יהיה מובהק ברמת ביטחון של 95 אחוזים, ערך ההפרש (K11-K12) בין הפונקציות חייב להתקבל מחוץ לקו רווח הסמך (המקווקו באיור). אחרת,לא תתקיים וודאות סטטיסטית להשוואה המרחבית של פיזור המקרים.

תוצאות המחקר מראים כי ערכי S המקורבים למרחק של 0.9 ק"מ, מצויים מעל לקו המקווקו (רווח הסמך). המשמעות הינה כי עבור S האשכול של קבוצת "התמותה המפתיעה" משמעותי יותר מאשר עבור קבוצת ה"שורדים המפתיעים", וכאשר הערכים נמצאים מתחת לקו, המשמעות הינה הפוכה [Jones et-al, 1996].

2.3.4.5 פונקצית רשת K (Network K- Function)

הפסקה הקודמת עסקה ביתרון השימוש בפונקציית K. עם זאת, כאשר עסקינן בתופעות המתרחשות ברשת מרחבית, הניתוח המסורתי של פונקציית K העושה שימוש במרחק האוקלידי (קו ישר בין שתי נקודות במישור) מעלה מספר בעיות. ראשית, הנחת ה"מרחב" הינה בעיתית. פונקצית K מבוססת על הנחה של מרחב מישורי מתמשך, כאשר בפועל מרבית האירועים מתרכזים רק באזור מסוים, בחלק מהמרחב. באם עסקינן בתאונות דרכים, הרי שהן תתרחשנה על רשת הכבישים ולא במרחב כולו. בעיה נוספת הינה נושא חישוב המרחק בין האירועים. המרחק יוצג בצורה מדויקת יותר, באם יחושב בהתאם לדרך הקצרה ביותר על גבי רשת הכבישים ולא לפי המרחק האוקלידי, כפי שמחושב במקור בפונקציית K [Yamada & Thill, 2004]. לפיכך, הציעו Okabe

and Yamada [2001] פונקציית K עבור רשת, המהווה הרחבה לפונקציה המקורית. איור מספר 18 מציג וויזואלית את ההבדל התפיסתי בין שתי הפונקציות ה- K המדוברות.

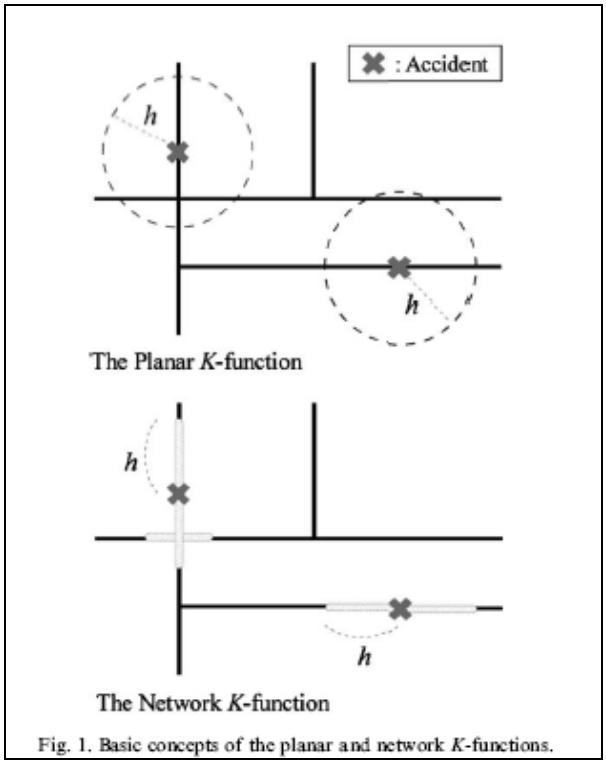


Fig. 1. Basic concepts of the planar and network K-functions.

איור מספר 18 - קונספט פונקציית K ופונקציית K עבור רשתות [Yamada & Thill, 2004]

פונקציית K הוגדרה עבור סט נתונים $p = \{p_1, ..., p_n\}$, המתפזר לאורך רשת סופית L_T באמצעות נוסחה מספר 56:

$$K^{net}(h) = \frac{1}{\rho} E\left(\text{מספר הנקודות בסט נתונים P, הנמצאות במרחק רשת של h מנקודה P על } L_T \right)$$

h≥ 0

נניח כי $L_p(h)$ הינו תת קבוצה של L_T , באופן בו מרחק הרשת בין p לבין כל נקודה על $L_p(h)$ קטן או שווה ל h, אזי נוסחה מספר (X) תהפוך לנוסחה מספר 57:

[57] $$K^{net}(h) = \frac{1}{|L_T|} \int_{p \in L_T} |L_P(h)| dp$$

מקדם הפונקציה ניתן על ידי נוסחה 58:

[58] $$\hat{K}^{net}(h) = \frac{L_T}{n(n-1)} \sum_{i \neq j} \sum I_h(s_{ij})$$

כאשר:

s_{ij} - מסמן את המרחק בין נקודה P_i לבין P_j.

$I_h(s_{ij})$ - האינדיקטור של הפונקציה המוגדר בנוסחה מספר 59:

[59] $$I_h(s_{ij}) = \begin{cases} 1 & if\ d_{ij} \leq h, \\ 0 & otherwise \end{cases}$$

היות ומודל הרשת מניח כי מדובר במרחב מוגדר של רשת, לא קיים בו אפקט הגבול, היות ועל פי ההנחה, מידע לא קיים מחוץ לרשת, שהיא למעשה אזור מחקר. בנוסף, ציינו כי קיים מודל כלאיים של פונקצית K עם אילוצי רשת.

Yamada & Thill [2004] בחנו את שני סוגי פונקציות ה- K על רשת הכבישים המהירים של ניו-יורק. היות ופונקצית K מושפעת מהטופולוגיה של הרשת וצפיפות התאונות, נבחרו שלושה אזורי מחקר, כאשר בכל אחד מאזורי המחקר התרחשו 450 תאונות. בתחילה נבחן הפיזור של כלל התאונות בכל אחד מהאזורים. במחקר המשך, נותחו תאונות שהתרחשו בסופי השבוע באזורים (1 ו- 3). זאת מאחר, ובימי סוף השבוע נמצא כי תאונות נוטות להיות מפוזרות יותר מאשר תאונות המתרחשות במהלך ימי חול באזורים אלה.

2.3.4.6 שיטת Kernel Density Estimation (KDE) – מפות צפיפות

שיטת ניתוח נוספת, לניתוח מידע ממופה של תבניות מרחביות. מניחה כי אזור המחקר A חולק לגריד ריבועי שווה, כאשר גודלו של כל אזור דגימה a x a. כך, ניתן לקבל באמצעות שיטה זו מושג כללי לגבי ריכוז אירועים ליחידת שטח [Cressie, 1993].

Tessa [2009] מציג במחקרו גישה, המשלבת GIS ו- Kernel Density Estimation (להלן: KDE) לזיהוי אזורים בהם קיים ריכוז של תאונות (להלן: Hotspots). במחקר זה, נעשה שימוש במתודות סטטיסטיות בכדי לבצע אישכול של ה- Hotspots בהתאם למידע מאפיין, לרבות מידע או פרמטרים ידועים לגבי האזור הסובב את מיקום התאונה. במחקרו מציין Tessa [2009] כי בספרות המקצועית העוסקת בתאונות דרכים, לא קיימת הגדרה כללית מקובלת לאיתור Hotspots. זאת מאחר וניתוח ה- Hotspots התרכז באופן מסורתי במקטעי כביש או צמתים ספציפיים, כאשר העיסוק בניתוח מרחבי/אזורי ופיזור הסיכון הנוצר מההתנגשויות, נזנח. חלק מהחוקרים מאתרים ריכוזי

תאונות תוך שהם מתייחסים לכמות התאונות בעוד אחרים בוחנים תדירויות (מספר תאונות בק"מ כבישׁ).

באופן מסורתי, מ- 1970 ואילך, נעשה שימוש במודלים סטטיסטיים. המודלים הראשוניים הניחו כי תאונות מתפלגות בצורה נורמאלית, כאשר בשלב הבא נעשה שימוש ברגרסיית פואסון. כיום נהוג לעשות שימוש גם במודל בינומי שלילי. בכלל העבודות נעשה שימוש במספר התאונות המדווח בתקופה נבחנת, כאשר מאפיינים מרחביים מודלו כערכים קבועים. בשונה ממחקרים אלה, מחקרו של Tessa [2009] מבוסס על ההנחה כי תאונות דרכים המתרחשות באזור מסויים תלויות לא רק בצומת או ברשת הכבישים, אלא גם במאפיינים מרחביים.

סיווג Hotspots מבוסס לרוב, על המידע המצורף לתאונה (סוג רכב, סוג הנפגע, שעת התרחשות התאונה וכיוצא באלה). עובדה זו מגבילה לעיתים את הבנת מורכבותו של אירוע תאונת דרכים. זאת, היות ומתעלמים מגורמים סביבתיים, חברתיים וכלכליים שגם להם השפעה על התרחשותה של תאונה, אולם לגביהם לא נאסף מידע. החידוש במחקר הינו, בשלב ראשון זיהוי Hotspots, ולאחר מכן אישכולם ויצירת טופולוגיה המבוססת על אינדיקטורים מרחביים.

לשיטת Kernel Density Estimation (להלן: KDE) יתרונות רבים. היתרון העיקרי הינו בקביעת "פיזור הסיכון" של התאונה. הכוונה ב"פיזור הסיכון" הינה להגדרת אזור הסובב אשכול מסויים, כאשר באזור זה קיימת סבירות הולכת וגדלה להתרחשות תאונה בהתבסס על תלות מרחבית. בבסיס השיטה, הנחת משטח סימטרי עבור כל נקודה, ולאחר מכן הערכת המרחק מהנקודה המסויימת למיקום מוגדר (למעשה לשאר הנקודות). המשמעות של הצבת ה- KERNEL, על נקודה, הינה יצירת משטח רציף. המתודה נקראת KDE היות ועבור כל נקודה שבה קיים אינדיקטור או מאפיין כלשהו, נוצר אזור מעגלי המוגדר בהתאם לגודל ה- KERNEL או רוחב הפס, שהינו למעשה רדיוס המוגדר על ידי החוקר. ערך האינדיקטור באותה נקודה "מפוזר" במרחב, בהתאם להגדרת פונקציה מסויימת. סכימת כל הערכים באותה הנקודה או אותו תא שטח, לרבות נקודות בהן לא קיימים אירועים מאותו האינדיקטור (או בעלי אותם מאפיינים) מספקת משטח צפיפות מוערך. הצפיפות יכולה להימדד בשתי שיטות. האחת Simple והשניה Kernel.

בשיטת Simple, כלל איזור המחקר מחולק לריבועים, כאשר עבור כל תא ברשת הריבועים מחושב ערך צפיפות, בהתאם ל"לרדיוס חיפוש" מוגדר. ערך הצפיפות הינו היחס בין מספר העצמים הנופלים ב"אזור החיפוש" (הנוצר בהתאם להגדרת גודל רדיוס החיפוש) לבין גודל אזור המחקר. ככל שהרדיוס גדל, גדלה הסבירות, לכך שהשכנים הסובבים יכילו נקודות רבות יותר, מה שיוביל ליצירת משטח צפיפות חלק יותר.

בשיטת ה- "Kernel" כלל איזור המחקר מחולק לריבועים, אולם, במקום להניח שכנות מעגלית סביב לכל תא, מוגדר מעגל (ברדיוס מסויים) סביב כל עצם נקודתי (במקרה שלנו - תאונה), ולאחר מכן פונקציה מתמטית מחשבת ומציבה ערכים הנעים בין 1 ל- 0, כאשר 1 הינו ערך המרחק הקרוב ביותר לעצם הנבדק (תאונה), ו 0 הינו בגבול התאים השכנים. הגורמים המשפיעים ביותר על תוצר ה KDE, הינם אורך הרדיוס המוגדר מסביב לכל תא בגריד או כל נקודה (תאונה) וגודל התא בגריד.

סיכום פרק הספרות

בפרק זה נסקרו לעומקם מחקרים שנערכו בשנים האחרונות בתחום תאונות הדרכים, לרבות הרחבה והתמקדות במתודות הסטטיסטיות הנהוגות. בהרחבה, נסקרו והוצגו מחקרים ומתודות שעיקרן בניתוח ספירת כמות אירועי תאונות דרכים למקטע דרך, ובהתאמה הותאמו מודלי חיזוי. לאור סקר הספרות, בעת בניית המודל יעשה שימוש במחקר הנוכחי במודלים המקובלים כיום, המניחים פיזור נתונים פואסוני ובינומי שלילי. לאור ריבוי מקטעי הדרך בהם קיים מספר תאונות השווה ל 0, קרי: מקטעי דרך בהם לא התרחשו כלל אירועי תאונות דרכים בתקופת המחקר, יבחנו גם מודלי Zero-Inflated, שיודעים למדל גם בסיסי נתונים המאופיינים בערכי 0 רבים.

בהיבט הפרמטרים המרחביים - ישולבו, יבחנו וידונו לעומקם, פרמטרים דוגמת שיפוע ועקמומיות, ששולבו במספר מצומצם של מחקרים עד כה, על אף היכולת לחילוצם בפשטות יחסית באמצעות שימוש ב GIS. עוד תוצג היכולת לעשות שימוש בחישוב זווית השמש למטרת איתור תאונות שהתרחשו תחת השפעת סינוור.

בנוסף יעשה ניסיון לאיתור צבירי אירועי תאונות דרכים (Hotspots) בתוכנת ArcGIS, במרחב המישור הכולל של אזור המחקר, תוך שימוש במתודת Kriging (בדומה למתודת ה KDE שתוארה לעיל). במקביל, ובהשראת הפונקציונאליות שתוארה עבור פונקציית K עבור רשתות ליניאריות, ננסה באמצעות נוסחה מפושטת יותר לאתר צבירים במרחב הליניאמנטי. האמור יבוצע תוך בחינת המרחק (בסיפים שונים) בין תאונות דרכים המתרחשות לאורך כביש ספציפי, כאשר על מנת לפשט את תהליכי המשקול של הנתונים, יחולק בסיס הנתונים, מראש, לקבוצות שונות (נוסף על ניתוח כלל אירועי הדרכים לאורך כביש מסויים)

3 מערך המחקר

3.1מטרת המחקר

מטרת המחקר הינה פיתוח מודל חיזוי מבוסס GIS להערכת הסיכון להיפגעות בקטעי דרך בינעירוניים לפי פרמטרים תחבורתיים ומרחביים.

המחקר בחן מספר היבטים ושאלות:

(1) בהינתן מידע בנוגע למאפיינים תחבורתיים ומרחביים עבור מקטע דרך מהירה בינעירונית - מהו שיעור תאונות הצפוי עבור אותו מקטע דרך, ומהי ההסתברות להתרחשות של אפס תאונות או לפחות תאונה אחת?

(2) בהינתן מידע בנוגע למאפיינים תחבורתיים ומרחביים עבור מקטע דרך מהירה בינעירונית - מהם המשתנים המשמעותיים ביותר, אשר להם השפעה על שיעור תאונות הדרכים?

(3) האם מתקיימת אינטראקציה בין המשתנים: אזור ועקמומיות (האם העקמומיות הינה בעלת השפעה שונה על מספר התאונות באזורים שונים?); בין אזור ושיפוע (האם ניתן לאפיין מספר תאונות שונה עבור שיפועים באזורים שונים?); ובין נפח תנועה לבין אזור (האם השפעת נפח התנועה על שיעור תאונות הצפוי אינה אחידה, ותלויה באזור בו אנו נמצאים?).

(4) האם מודלים סטטיסטיים דומים, המופעלים על מקטעי דרך באורכים שונים, יניבו תוצאות שונות בהיבט הפרמטרים המשמעותיים לקיום תאונות במקטע דרך?

(5) מה היכולת לאתר צבירים, הן במרחב הליניאמנטים (קרי: לאורך הכבישים) והן במרחב הרסטרי (כלל שטח אזור המחקר) לרבות יתרונות וחסרונות של השיטות.

בבסיס המחקר, נאספו נתונים ייעודיים עבור כל מקטע דרך, ולאחר מכן, בוצעו ניתוחים מסוגים שונים:

מרחביים - איתור צבירים לאורך ליניאמנטים, גזירת מאפיינים מרחביים דוגמת שיפוע ועקמומיות.

סטטיסטיים – בחינת ארבע סוגים של מודלים מתקדמים: פואסון, בינומי שלילי, פואסון Zero-Inflated , בינומי-שלילי Zero-Inflated. מבחנים לטיב מודל, מבחני שיוויון והערכת הסתברויות. אלה בוצעו בתוכנת SAS.

גיאו-סטטיסטיים – יצירת מפות חיזוי (prediction) לחישוב הסתברות לאירוע של תאונות דרכים במרחב הפיזי-רסטרי. אלה בוצעו בתוכנת ה- GIS (Geographic Informaotion) ArcGIS System, שיתרונה, בביצוע ניתוחים מרחביים מתקדמים תוך שילוב ניתוחים סטטיסטיים. יתרה מזאת, סביבת העבודה הגיאוגרפית הינה איטרטיבית, וניתן לבצע בה חזרה נשנית על הניסויים, למטרת טיוב המודל.

המודל מציע שילוב פרמטרים ייחודי, מרחבי בעיקרו, דוגמת: עקמומיות, שיפועים וזויות השמש (איתור סינוור), לשימוש בעת ניתוח תאונות הדרכים. בנוסף, בוחן לראשונה בישראל, את נתוני

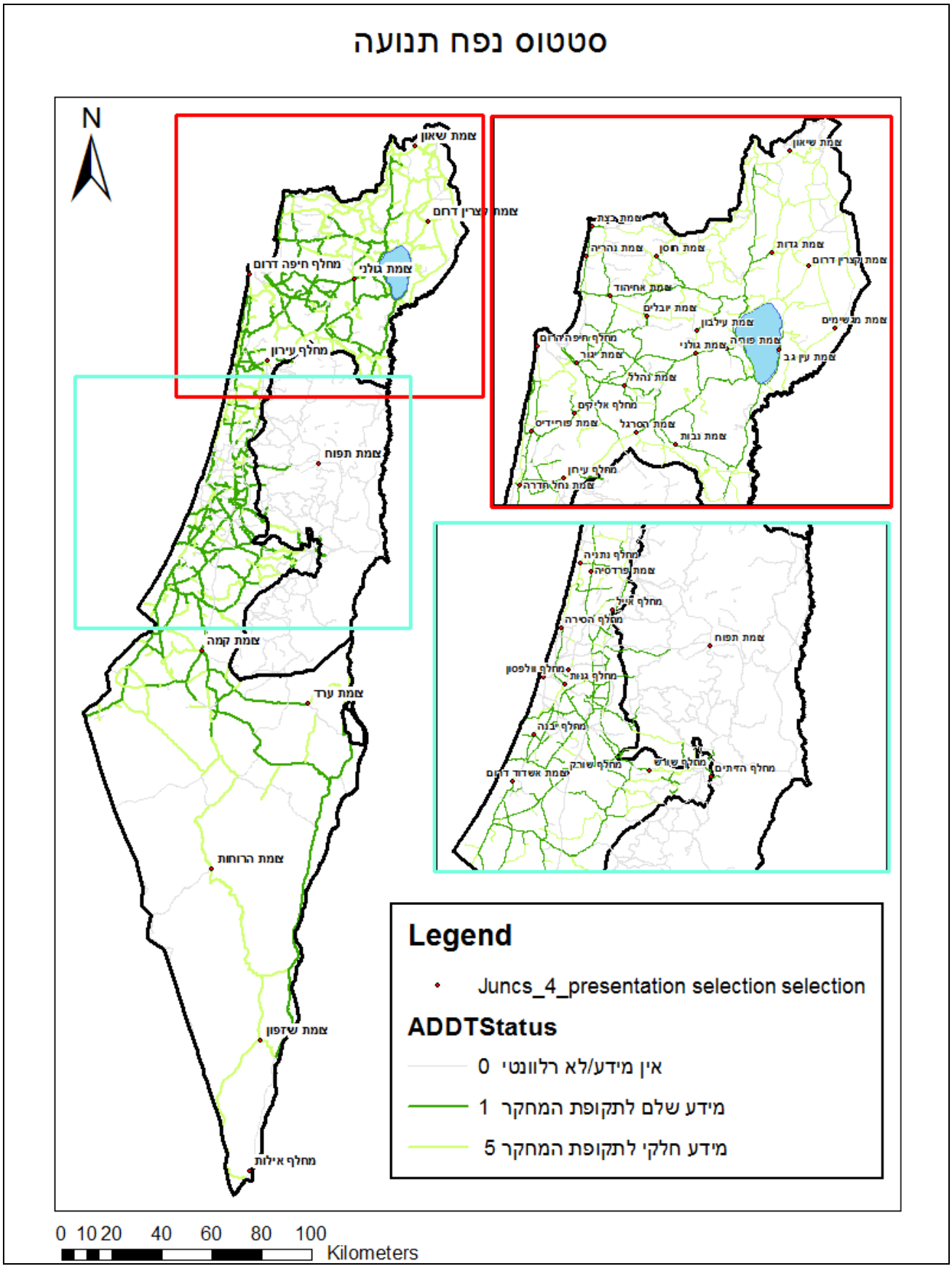
מד"א והיכולת לעשות בהם שימוש בחקר תאונות הדרכים. המודל נבנה על בסיס קבצי "תאונות דרכים עם נפגעים" של הלמ"ס בין השנים 2005-2007.

יתרונו של המודל, באיכון אזורי תורפה של תופעות הנשנות במרחב. כך, ניתן יהיה להתמקד באזורים בהם נדרשת השקעה תשתית או אכיפה מוגברת. כן, תומך המודל, בלימוד וניתוח הגורמים לתאונות הדרכים לצורך שיפור והקטנה של מספר אירועי התאונות. יישום המודל יאפשר להעריך טיבן של תוכניות קיימות בנוגע לדרכי ההתמודדות עם תאונות הדרכים, ואף אפשר כי יתרום לפיתוח תוכניות חדשות להתמודדות עם נושא תאונות הדרכים הן ברמה המחקרית והן ברמה היישומית.

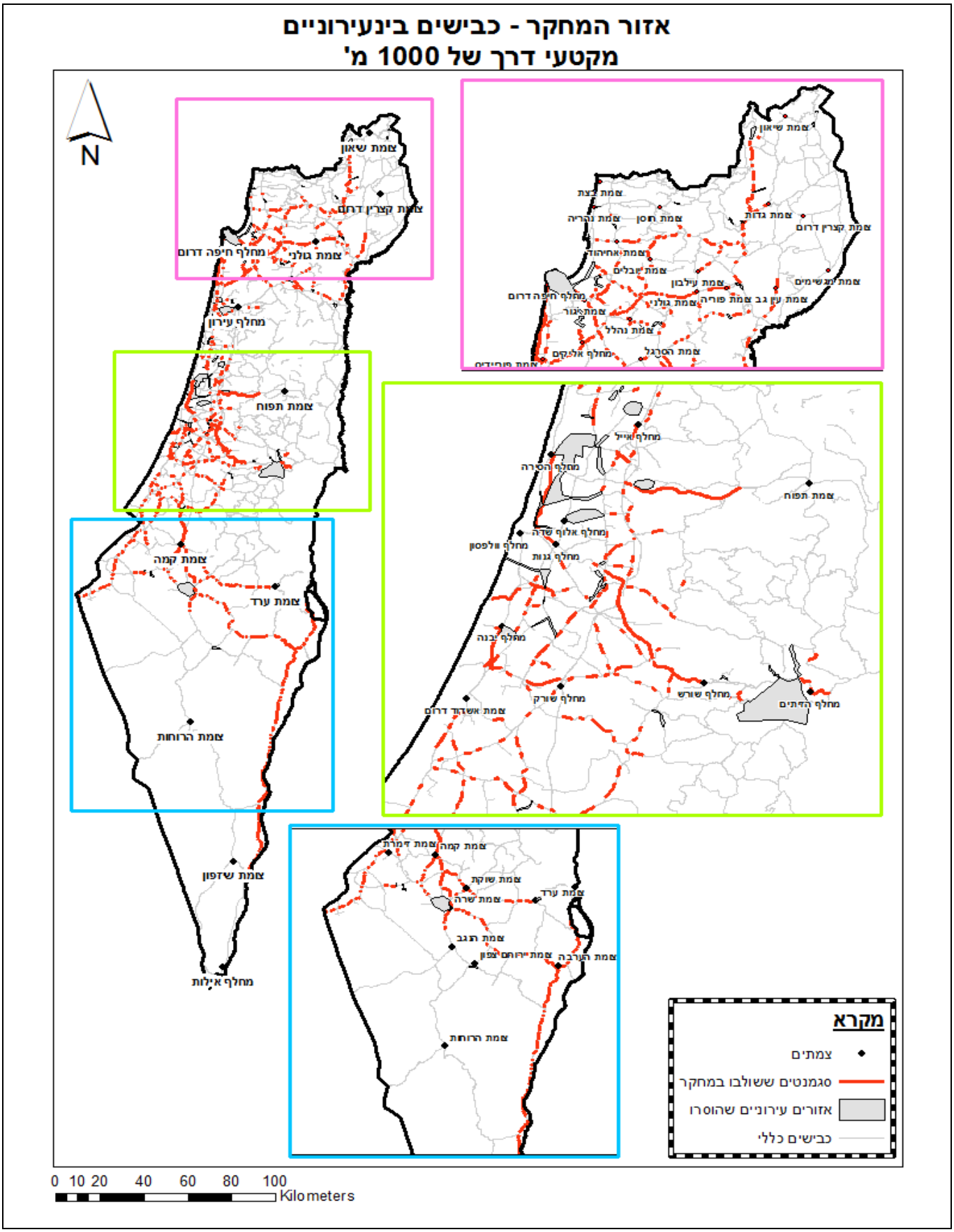
3.2 אזור המחקר

ביסודו שם לו המחקר למטרה לבחון את תאונות הדרכים ולאפיין מודל עבור מקטעי דרך בינעירוניים בישראל. איור מספר 19 מציג את כבישי ישראל העיקריים, לרבות הסטטוס בנוגע לקיומו של נתון נפח תנועה (AADT). בירוק כהה סומנו מקטעי הדרך עבורם קיים מידע שלם בתקופת המחקר (2005-2007); בירוק בהיר סומנו קטעי דרך עבורם קיים מידע חלקי (נפח תנועה עבור שנה או שתיים מתוך שלוש שנות המחקר); ובאפור סומנו קטעי הדרך לגביהם לא קיים כלל מידע בתקופת המחקר. היות ובמחקרים רבים עולה, כי נתון זה הינו משמעותי ביותר לפיתוח מודלים הבוחנים ומנתחים מספר אירועי תאונות דרכים באזור מסויים, הוחלט לשלב בבסיס נתוני המחקר, רק את מקטעי הדרך עבורם קיים מידע מלא לכל תקופת המחקר. כך צומצם בסיס הנתונים, באופן די ניכר, כבר בתחילת המחקר.

המקטעים הסופיים ששולבו במודל מופיעים במפה באיור 20, שבה מוצג בסיס הנתונים בחלוקה לקטעי דרך באורך של 1000 מטרים. בחירת המקטעים הסופיים הושפעה בעיקרה מארבעה פרמטרים. הראשון, מעבר באזור בעל השפעה של עיר או יישוב – מקטעי דרך באזורים אלה הוסרו. פרמטר נוסף הינו טווח ההשפעה של צומת – כל מקטע שנחתך עם צומת, הוסר גם הוא. פרמטר משמעותי נוסף הינו נפח תנועה (AADT), והאחרון הינו חיתוך עם נתוני DEM – במקומות בהם לא היו נתוני גובה "הוסרו" הסגמנטים מבסיס הנתונים.



איור מספר 19 - סטטוס הכבישים לגביהם קיים נתון נפח תנועה (AADT)



איור מספר 20 - מקטעי הדרך של הכבישים הבינעירוניים ששולבו בבסיס נתוני המחקר (מקטעי דרך בחלוקה ל 1000 מ')

3.3 שיטות המחקר

בקרת איכות -

בקרת איכות מרחבית - הבחינה המרחבית בוצעה בשני אופנים שונים. הראשון, בחינה וויזואלית - מיפוי ובחינת נכונות האזור הגיאוגרפי של קבצי המידע שהתקבלו. השני, בחינה לאיתור שגיאות במיפוי הגיאוגרפי (מפערים, Over/Under shoot).

בקרת איכות סטטיסטית – בוצעה בתחילה על הנתונים הטבלאים והפרמטרים הקיימים. במהלך בקרה זו, נבחנו כדוגמא: אחוז הנתונים הרלוונטיים החסרים בבסיס הנתונים (לכמה אירועים חסר נתון המייצג מיקום מרחבי או נתון בנוגע לגורם התאונה וכיוצא באלה) והתפלגות הנתונים. בשלבים מתקדמים יותר של המחקר, בעת בניית המודל הסטטיסטי, בוצעו בקרות איכות משולבות על תוצרי המודלים השונים. זאת, לרבות בקרת איכות בסיסית על המודל בדרך של בניית מודל אימון (בקרה) על בסיס 2/3 הנתונים ומודל ביקורת שנבנה על סמך 1/3 הנתונים הנותרים.

עיבוד ראשוני (Pre-processing) – בוצע תהליך Zero-Check, אשר במהלכו, טופלו ותוקנו הנתונים בכדי להביאם למצב התחלתי בו מתקיים מספר שגיאות מינימאלי (בשאיפה ל- 0), הן במיפוי והן בנתוני המקור.

ניתוחים סטטיסטיים ובניית המודל - בחינת המודל הסטטיסטי המתאים (פואסוני, בינומי שלילי), מבחנים סטטיסטים, איכון צבירים פרמטריים וצבירי מאפיינים במרחב הפרמטרי.

ניתוחים מרחביים - מיפוי אירועים, איכון צבירים ("אשכולות") מרחביים במרחב הפיזי ובניית מפות הסתברות (Kriging)

ניתוחים משולבים - איכון צבירים במרחב הפרמטרי (במקום לתאר תופעה במרחב הפיזי, דוגמת X ו-Y, יתוארו צירי המרחב כפרמטריים).

3.4 נתונים

3.4.1 טבלאיים

3.4.1.1 תאונות דרכים למ"ס

הלמ"ס מפרסמת נתונים על תאונות דרכים עם נפגעים, אשר נחקרו על ידי משטרת ישראל ונרשמו על ידה. אלה, מכילים שלושה קבצים עיקריים: תאונות, מעורבים ורכב. בקבצים, מפורט מידע דוגמת: מועד התרחשות התאונה, סוג התאונה, סוג דרך, מיקום

התאונה (ק"מ ומספר כביש במידה ועסקינן בתאונות בין עירוניות, ומספר בית ורחוב במידה ועסקינן בתאונות עירוניות), מספר הנפגעים בתאונה, כלי הרכב המעורבים בכל תאונה, מידע לגבי הנהגים שנהגו באותם כלי הרכב.
מסד הנתונים של הלמ"ס מבוסס על קובץ מנהלי המתקבל מדי חודש ממשטרת ישראל, המכיל את התאונות שדווחו למשטרה, ושנפתח עבורן תיק תאונה במשטרת ישראל. הנתונים כוללים רק תאונות שדווחו למשטרה, כאשר משטרת ישראל אינה רושמת תאונת בהן מעורב כלי רכב צבאי. קיימת הפרדה בין תאונה שהוגדרה במשטרה כתאונה הנחקרת על ידה (תיק כ"ד) לבין תאונה שאינה נחקרת ("תיק כללי עם נפגעים").

3.4.1.2 תאונות דרכים מד"א

מד"א מתחזקת מערכת מידע גיאוגרפי מתקדמת, המאפשרת למוקד השליטה לבחון את פריסת כלל האמבולנסים והנהגים בזמן אמת. בעת הגעה לארוע מדווח רכב החירום למוקד. על אף שקיימת היכולת ברמה הטכנית, לא מתבצע איכון מרחבי על ידי מי מהמעורבים (המוקד/רכב החירום). זאת, על אף שכאמור קיימות מערכות GPS בכלי הרכב (אמבולנסים).
המטרה שהוגדרה עבור השימוש בקבצי מד"א היתה, מראשיתה, בחינת יכולת טיוב נתוני תאונות הדרכים. לאור האמור, ולאור כפילויות רבות בבסיס הנתונים, עלה כי ניתן לעשות שימוש בנתוני מד"א, ברמת סמך גבוהה, רק למטרת טיוב נתוני הלמ"ס באזורים עירוניים בהם קיימת כתובת מדוייקת אותה ניתן לקשור לפי התאריך, ולא במקטעים בינעירוניים. הנתונים שהתקבלו ממד"א הגיעו בקבצי XLS, ששולבו בשלב מאוחר יותר בבסיס נתונים MDB.

3.4.1.3 שכבת שמות רחובות טבלאית (annotation.address)

שכבה טבלאית המכילה את שמות הרחובות העירוניים, בערים הנבחנות.

3.4.1.4 AADT

נתוני ספירות התנועה בכבישים בינעירוניים מופיעים בפרסום הלמ"ס: "תאונות דרכים עם נפגעים 2007 חלק ב: תאונות בדרכים לא-עירוניות", בפורמט PDF.
הפרסום שימש לעדכון הק"מ בצמתים בשכבת ROADS שהתקבלה מחברת "פלד מיפוי ממ"ג בע"מ" (להלן: "פלד מיפוי"). מידע זה נדרש על מנת לאכן את התאונות במרחב הגיאוגרפי הבינעירוני. במקומות בהם היה חסר ערך הק"מ בצומת בקובץ של "פלד מיפוי", הושלם הנתון על פי קובץ זה. בנוסף, במקומות בהם לא התקיימה הלימה בין שכבת "פלד מיפוי" וקובץ הלמ"ס, טוייב ועודכן הק"מ בצומת לפי פרסום הלמ"ס.

3.4.2 שכבות מידע גיאוגרפי (GIS)

3.6.2.1 שכבת מבנים

של הממ"ג הלאומי (שכבה פוליגונאלית Layer 7)

3.6.2.2 שכבת כבישים

<u>עירוניים</u> - של הממ"ג הלאומי באזורי **חיפה וירושלים** (שכבה קווית Layer 1)

<u>בינעירוניים</u> – חברת מפה. עדכנית לרבעון שני של 2008.

3.6.2.3 שכבת Roads – "פלד מיפוי".

מכילה כבישים בינעירוניים וכבישים עירוניים ראשיים לרבות טופולוגיה של צמתים (בפורמט Coverage) . שכבה זו נקלטה ע"י "פלד מיפוי" ממדידות בשטח וממפות מע"צ. טבלת המאפיינים של קטעי הכביש כוללת, בין היתר, את המאפיינים הבאים: מספר הכביש של מע"צ, ק"מ בתחילת הקטע; ק"מ בסוף הקטע ; "מצומת" ו"אל צומת". טבלת המאפיינים של הצמתים כוללת, בין היתר, את שמות הצמתים ומספרם. לפני תחילת העבודה הוסרו מהשכבה כל קטעי הכביש שאין להם מספר מע"צ.

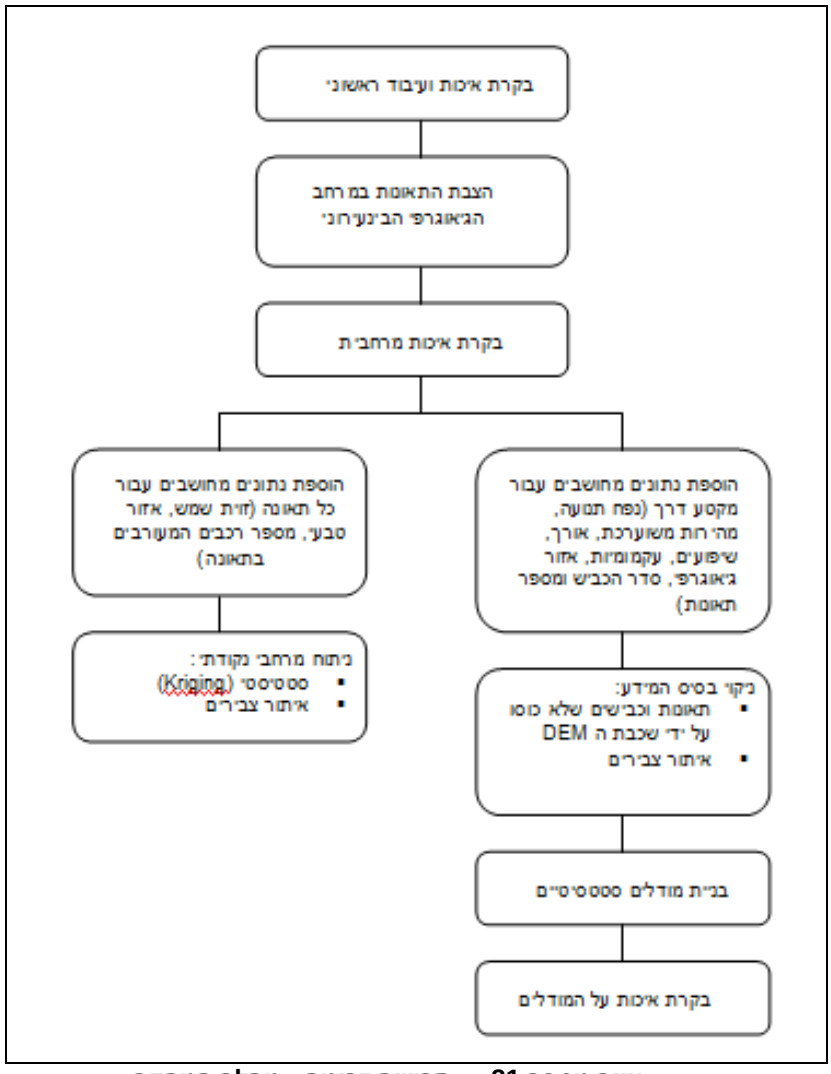
3.6.3 מודל גבהים סיפרתי (DEM)

שכבת רסטר המייצגת גבהים. רזולוציה מוערכת של כ- 33X33 מ' עבור כל פיקסל. שכבה זו שימשה לחילוץ השיפועים עבור מקטעי הדרך. בנוסף ניתן היה לעשות בה שימוש בעת עיסוק בנושא הסינוור (איכון השיפוע של כלי הרכב).

4. מהלך המחקר

4.1 שלבי המחקר

איור מספר 21 מציג תרשים כללי של מהלך המחקר



איור מספר 21 - תרשים זרימה - מהלך המחקר

4.2 בקרת איכות ועיבוד ראשוני (Pre-processing) של נתוני המקור

תת פרק זה מתאר את מהלך בקרת האיכות והעיבוד הראשוני של הנתונים, טרם הצבתם במרחב.

4.2.1 קבצי תאונות הדרכים של הלמ"ס

ראשית, יובאו כלל נתוני תאונות הדרכים לתוכנת SPSS. לאחר היבוא התבצעה בדיקה אל מול קובץ הבקרה המצורף לנתוני הלמ"ס (טבלה מספר 8). קובץ זה מכיל את ממוצעי עמודות הנתונים בטבלאות המידע של תאונות הדרכים, ומטרתו בקרת איכות לאחר ייבוא הנתונים.

ממוצעי השדות, כפי שחושבו לאחר ההעברה, הושוו לנתונים המופיעים בקובץ הבקרה של הלמ"ס.

טבלה 8 - תאונות דרכים עם נפגעים - קובץ בקרה

תאונות דרכים עם נפגעים-רכב

שם משתנה	ממוצע
PK_TEUNA_FIKT	2005043756.22
MISPAR_REHEV_fikt	1.56
SHNAT_TEUNA	2005.00
HODESH_TEUNA	6.50
NEFAH	6.31
SHNAT_YITZUR	1907.76
KIVUNE_NESIA	1.10
MATZAV_REHEV	0.04
SHIYUH_REHEV_LMS	1.11
SUG_REHEV_LMS	3.22
MEKOMOT_YESHIVA_LMS	61.26
MISHKAL_KOLEL_LMS	1.06
SUG_TIK	1.00

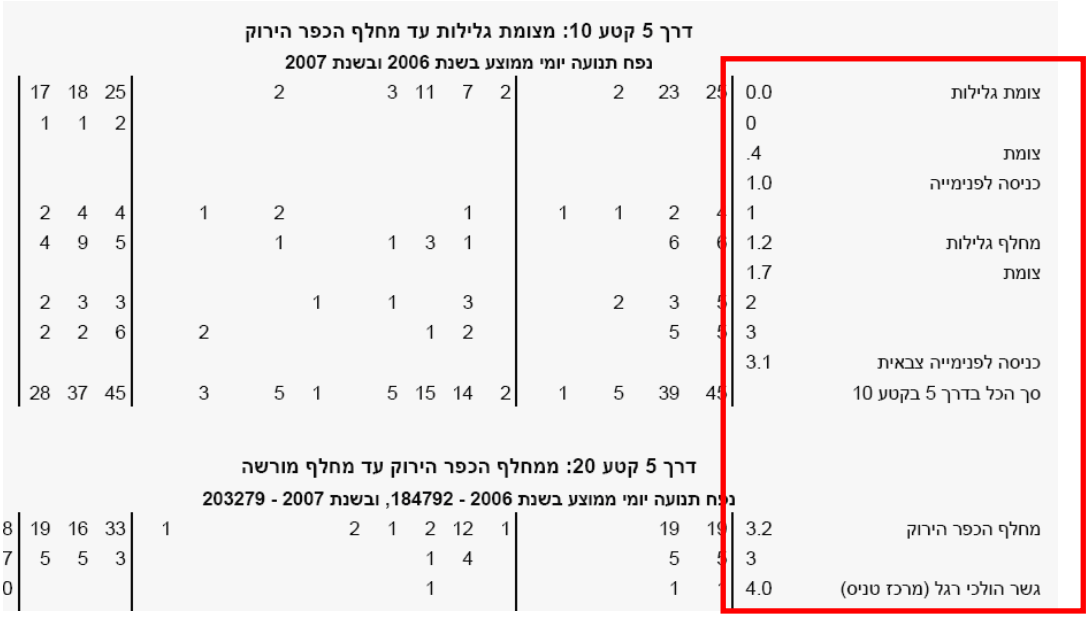
בשלב השני בוצעה בדיקת ערכי קצה עבור השדות הרלוונטיים. במסגרת זו נבדק: אי קיום ערכים שליליים, ערכי מינימום וערכי מקסימום.

4.2.2 שכבת הק"מ בצמתים - חב' פלד מיפוי (בקרה ע"י קובץ PDF של הלמ"ס)

על מנת להציב את אירועי תאונות הדרכים על רשת הכבישים, תוך שימוש בכלי: Make Route Events , נדרש בשלב הראשון לעדכן ק"מ בצמתים. היות ולא ניתן היה לקבל שכבת ק"מ בצמתים ממע"צ, בפורמט SHP (עבור תוכנת ArcMap) או בפורמט טבלאי (שהיינו מייבאים לתוך תוכנת ArcMap), נעשה שימוש בשכבה הגיאוגרפית Roads של "פלד מיפוי", המכילה את הק"מ בצמתים. כאמור, שכבה זו נקלטה ע"י "פלד מיפוי" ממדידות בשטח וממפות מע"צ.

בקרה האיכות על שכבה זו התבצעה אל מול קובץ PDF של הלמ"ס. קובץ זה של הלמ"ס והרשות הלאומית לבטיחות בדרכים:" *תאונות דרכים עם נפגעים, 2007 חלק ב :תאונות בדרכים לא-עירוניות*", מפרט עבור כל מקטע כביש את ערך הק"מ הרץ בצמתים. באיור מספר 22 מוצג לדוגמא קטע כביש 5, עבורו הצומת הראשון, המקבל את הערך 0.0 הינו צומת גלילות[13]. הצומת המשמעותי הבא אחריו הינו מחלף גלילות[14], המקבל ערך 1.2 ק"מ , וכן הלאה.

[13] כיום מחלף גלילות מערב
[14] כיום מחלף גלילות מזרח



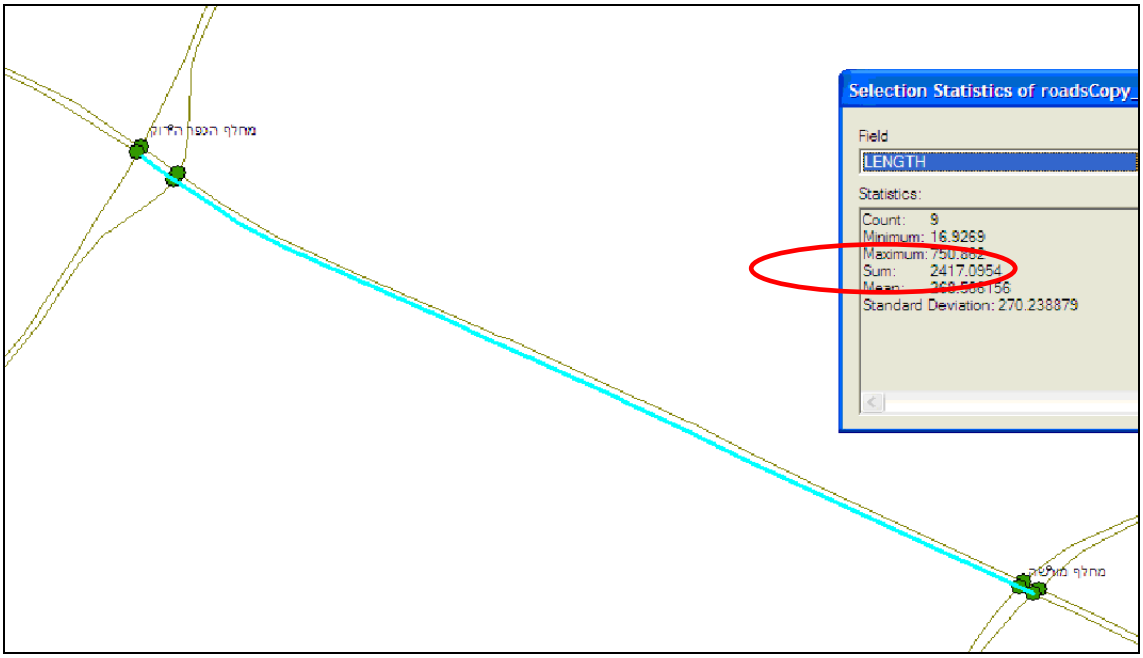
איור מספר 22 - כביש 5, דוגמא לפירוט נתון נפח תנועה
[מקור: *תאונות דרכים עם נפגעים, 2007 חלק ב :תאונות בדרכים לא-עירוניות*]

כבר בתחילה ניתן היה להבחין, כי ערכי הק"מ עבור הצמתים בשכבת Roads של "פלד מיפוי" וערכי הק"מ בצומת בקובץ הלמ"ס, אינם תואמים. בכדי לוודא ממה נובע הפער, נבחנו מדגמית מספר מקטעים. כך לדוגמא, טבלה מספר 9 מציגה את ערכי הק"מ בצמתים בשכבת ה Roads אל מול קובץ הלמ"ס, בין צומת הכפר הירוק למחלף מורשה.

טבלה 9 - ערכי הק"מ בצמתים בשכבת ה Roads אל מול קובץ הלמ"ס, בין צומת הכפר הירוק למחלף מורשה.

צומת	מחלף הכפר הירוק	מחלף מורשה	הפרש
פלד מיפוי	3.17	5.57	2.4 ק"מ
למ"ס	3.2	5.8	2.6 ק"מ

במהלך הבדיקה, סומנו בשכבה הגיאוגרפית של הכבישים, אורכי המקטעים הנכללים בטווח שבין שני הצמתים (ראה איור מספר 23, בו המקטעים מסומנים בתכלת). המקטעים שנבחרו, מייצגים את המרחק המקסימאלי בין שני הצמתים. לאחר מכן נסכמו אורכי המקטעים באופן אוטומאטי (כאמור, עבור המרחק המקסימאלי), והתקבל כי סכום מרחק המקטעים המקסימאלי הינו 2.41 ק"מ. כפי שניתן לראות בטבלה מספר 9, המרחק ששולב בקובץ הלמ"ס עומד על 2.6, שהינו ערך הגבוה יותר (ב 200 מטרים) מהמרחק המקסימאלי הקיים בפועל. מדוגמא זו, כמו בדוגמאות אחרות שנבחנו, עולה כי נתוני שכבת ה- Roads, המבוססים על נתוני מע"צ ומדידות עצמאיות של חברת "פלד מיפוי", מדוייקים יותר.



איור מספר 23 - אורך מקטעים בין צומת הכפר הירוק למחלף מורשה.

כאשר היה חסר הק"מ בצומת, בשכבת roads, הושלם הנתון בהתאם לפרסום הלמ"ס.

4.2.3 הצבת התאונות בכבישים בינעירוניים

השימוש בכלי ה- GIS: Make Route Events מספק קובץ נתוני שגיאה עבור תאונות שלא הושמו במרחב. הקובץ אינו מספק את הסיבה לטעות, אלא מכוון לנקודות שיש לבחון אותן למטרת טיובם. הבעיות העיקריות בגין לא הוצבו הנקודות הינן:

- ק"מ חסר (לא עודכן ידנית) עבור צומת בראשית/בסיום קטע הדרך.
- קיטוע/קיטום פיזי בשכבת הכבישים.
- כיווניות שאינה נכונה (מפריע לבניית המסלול)
- לעיתים, לא תמיד הק"מ עבור הצומת הראשון הנמצא בתחילת הכביש עומד על ערך של 0 ק"מ. כך, אירע לעיתים, שהק"מ שעודכן עבור תאונה מסויימת (ע"י משטרת ישראל), עומד על ערך הקטן, מערך הק"מ הקיים בתחילת הכביש. כך לדוגמא: בכביש מספר 437, ערך המרחק הרץ עבור הצומת הראשון בכביש הינו: 6 ק"מ. כך, לא ניתן היה לאכן 19 תאונות, אשר עבורן היה ערך בבסיס הנתונים, ערך של ק"מ הקטן מ 6 ק"מ. דוגמא נוספת ניתן לראות בכביש 4. עבור צומת "חבל עזה", שהינו הראשון בכביש, מתקיים מרחק רץ של 51 ק"מ. על כן, לא ניתן היה להציב מרחבית, על המפה, 17 תאונות עבורן נתון היה ערך של ק"מ הקטן מק"מ 51.

פערים אלה, יכולים לנבוע מטעות רישום (במקרה זה אין כל דרך לאתרה) או מטעמים היסטוריים הנובעים משינויים בערך הק"מ בכביש, במידה ולדוגמא, ושונה נתיב או שהכביש הוארך, והחלו למדוד את ראשיתו מצומת מרוחק יותר.

4.2.4 שדה הכיווניות

הכוונה במונח כיווניות, הינה להגדרת כיוון התנועה ביחס לאבני הק"מ המסומנות בכבישים. בשימוש בביטוי: "עם אבני הק"מ" הכוונה הינה לתנועה הזורמת בכיוון הנע בין הק"מ הראשון של הכביש לקילומטר השני, בין הק"מ השני לשלישי וכן הלאה. בדומה, בביטוי "נגד כיוון אבני הק"מ", הכוונה הינה לנסיעה בכיוון ההפוך (מהקילומטר השלישי לקילומטר השני, מהשני לראשון וכן הלאה).

לפרמטר ה"כיוון", השפעה רבה על ניתוח תאונות דרכים. זאת, הן עבור איכון והשמת התאונות במרחב הגיאוגרפי והן בהיבט חישוב פרמטרים נוספים, דוגמת חישוב זווית השמש. כך, בוצע אימות על נתון זה הן בקובץ הכבישים של חברת מפה (אל מול קובץ הלמ"ס ופרסומי הרשות הלאומית לבטיחות בדרכים), והן בקובץ הלמ"ס עצמו.

להלן תוצגנה שתי דוגמאות לאימות. הראשונה בוצעה על "הכיווניות" בנתוני תאונות הדרכים הבינעירוניות בשנת 2005, שהתקבלו מהלמ"ס, השניה בוצעה על נתוני המקור הגיאוגרפיים של חברת מפה.

<u>דוגמא א – בחינת נתוני הכיווניות של נתוני הלמ"ס, בדרכים בינעירוניות בשנת 2005</u>

טבלה מספר 10 מציגה את כמות אירועי תאונות הדרכים בכבישים בינעירוניים, בחלוקה לחומרה וכיווניות, בקובץ תאונות הדרכים של הלמ"ס בשנת 2005.

<u>כיווניות</u>:

0 - לא ידוע

1 - עם כיוון הספירה (הכיוון ה"עולה" של ערכי הק"מ בצמתים)

2 - כיוון הספירה ההפוך (הכיוון ה"יורד" של ערכי הק"מ בצמתים)

3 – רכב חונה

<u>חומרה</u>: 1 – הרוג ; 2 – פצוע קשה ; 3 – פצוע קל

טבלה 10 - כמות אירועי תאונות הדרכים (למ"ס, 2005) בכבישים בינעירוניים, בחלוקה לחומרה וכיווניות

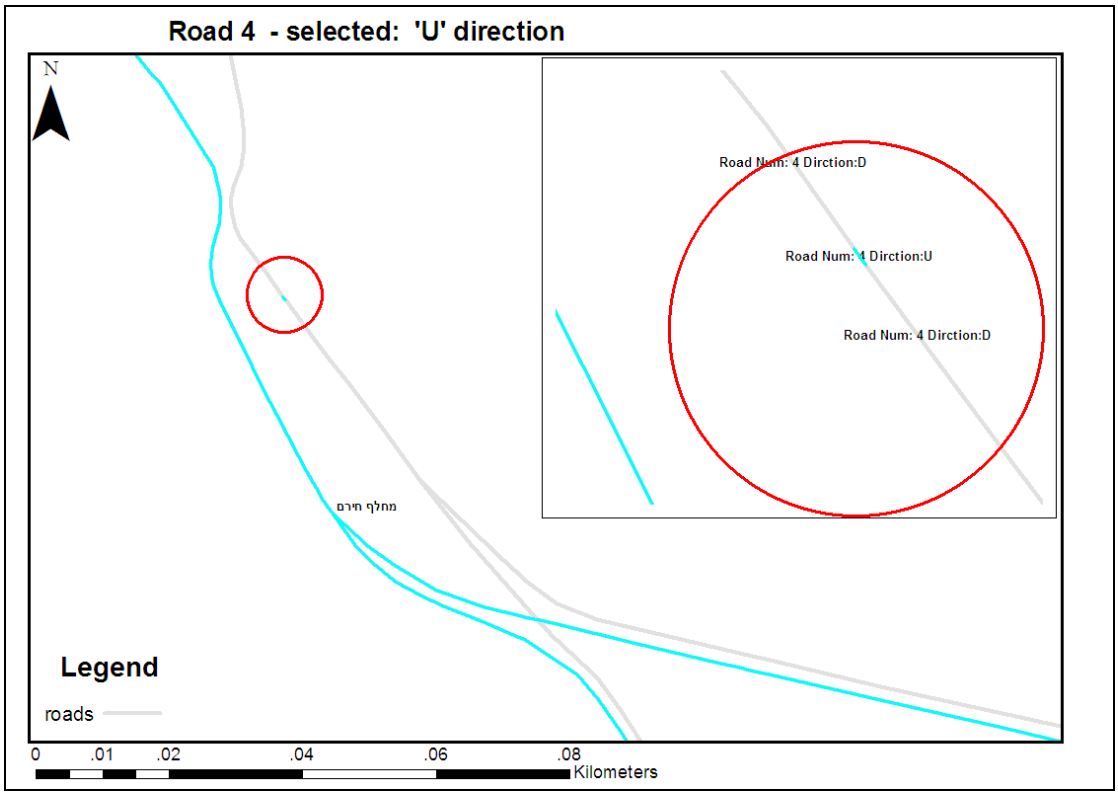
KIVUNE_NESIA	HUMRAT_PGIA	CountOfKIVUNE_NESIA
0	0	45
1	0	1170
2	0	240
3	0	47
0	1	5
1	1	163
2	1	57
3	1	6
0	2	5
1	2	544
2	2	147
3	2	17
0	3	48
1	3	2246
2	3	474
3	3	52

מתוצאות הסכימה עולה, כי החלק הארי של התאונות מסומן בבסיס הנתונים, כמתרחש עם כיוון הספירה. אומנם, היות ותאונות הינן אירועים שאינם צפויים, לא נצפה להתפלגות נורמאלית, ובכל זאת עולה החשש כי "נוח" לסמן, כברירת מחדל, את הערך "1" (בעיקר בתאונות בהן לא יוצא בוחן למקום התאונה). בכדי לבדוק מגמה זו נבחן פרמטר נוסף - "חומרת הפגיעה". זאת, מתוך הנחה כי בתאונות קטלניות בדרך כלל מגיע בוחן פיזית למקום האירוע, והסבירות לטעות במילוי שדה זה בטופס נמוכה. בדיקת האימות הראתה, כי במגמה זו לא חל שינוי. כלומר, גם לאחר שנבחן רכיב "חומרת הפגיעה", עדין עולה כי יותר תאונות מתרחשות בכיוון הספירה של אבני הק"מ בצמתים. לפיכך, הוחלט לעשות שימוש בשדה הכיווניות, על אף שברור כי הנתון הינו נתון בעייתי.

דוגמה ב' - בחינת נתוני הכיווניות של נתוני המקור הגיאוגרפיים מחברת מפה

הבקרה על שדה "הכיווניות" בוצעה על קובץ המקור של חברת מפה. בשכבת הכבישים הבינעירוניים של חברת מפה קיים שדה בשם DirectionM. שדה זה מצביע על כיווניות ביחס לאבני הק"מ. ערך U משמעו עם כיוון הק"מ, ערך D משמעו כיוון הק"מ היורד, ו- B - הינו עבור מקטעי דרך בהם מותר לנסוע לשני הכיוונים, כאלה שנקלטו כאלמנט אחד.

כפי שעולה מאיור מספר 24 באזור מחלף חירם (מרחב חיפה) נקלט קטע כביש שאמור להיות בערך D , כערך U. תקלות מסוג זה, תוקנו ידנית בבסיס הנתונים הגיאוגרפי.



איור מספר 24 - דוגמא לתקלה בנתון הכיווניות בסגמנט גירוגרפי

4.2.5 קובץ הלמ"ס – נתוני תאורה וסטטוס יום/לילה

למטרת איתור צבירי תאונות דרכים שהתרחשו בלילה, בוצע אימות של נתוני התאורה בקובץ נתוני תאונות הדרכים של הלמ"ס (עבור כבישים בינעירוניים). הבדיקה התבצעה תוך שימוש ב-SQL.

נבדק האם מתקיים מצב בו עבור תאונה כלשהי בבסיס הנתונים, קיים קוד תאורה 1 (יום) ו"גם" סטטוס לילה (קוד: 3 – לילה, פעלה תאורה ;4 – לילה, קיימת תאורה, בלתי תקינה/ לא פעלה ;5 – לילה, לא קיימת תאורה ;6 – לילה, לא ידוע). התוצאה שהתקבלה הינה 0. משמע הנתון תקין ומצב זה אינו מתקיים בבסיס הנתונים.

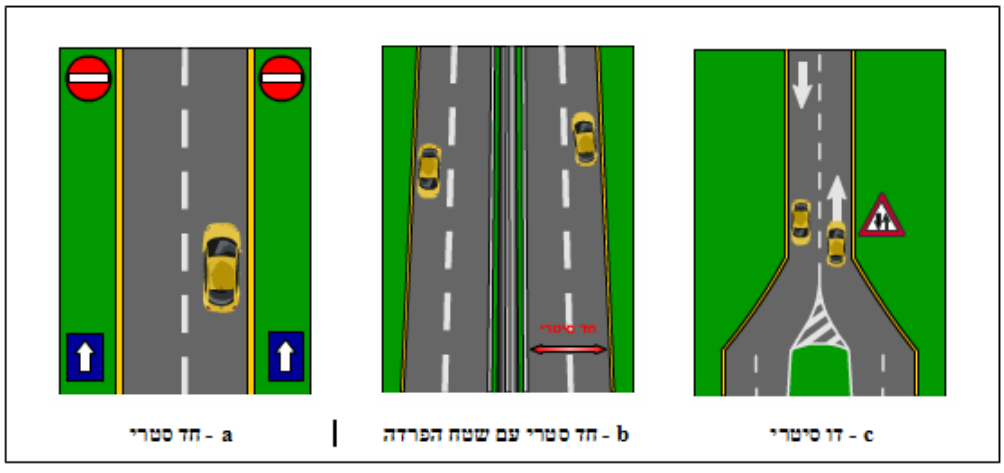
בדיקה נוספת בחנה את טווח השעות. נבחרו תאונות שהוגדר עבורן בבסיס הנתונים קוד 5 (לילה) בשילוב הגדרת סטטוס "לילה" (כאמור קודים: 3,4,5,6). עבור תאונות אלה בחנו כי טווח הזמנים בעמודת הזמן ("TIME_") אכן משתייכת לשעות הלילה. גם בבדיקה זו לא עלו ממצאים חריגים.

4.2.6 קובץ הלמ"ס - נתוני מסלול

אחד הפרמטרים שנבחן לשילוב בבסיס הנתונים כמשתנה, הינו מספר הנתיבים. משתנה זה נחקר בעבר במספר מחקרים. לצורך שילוב נתון זה, בוצעה בקרת איכות משולבת על הנתונים. הראשונה, בוצעה ברמת בסיסי הנתונים ושלמותם. קרי, עבור כמה תאונות מעודכנים ערכים בבסיסי הנתונים. טבלאות מספר 11 ו- 12 מציגות התפלגות כמות אירועי תאונות דרכים וסוג לפי קוד הלמ"ס.

טבלה 11 - דרך חד מסלולית – התפלגות כמות אירועים וסוג

קוד	כמות אירועים	הגדרה	הערות
1	449	חד סטרי	איור 25a במקרה המצוייר גם דו נתיבי.
2	1848	דו-סטרי+קו הפרדה רצוף	
3	1893	דו-סטרי+אין קו הפרדה רצוף	איור 25c (החלק העליון באיור, בו מקטע הכביש הוא חד מסלולי - דו כיווני).
4	81	אחר	
9	161	לא ידוע מספר מסלולים	



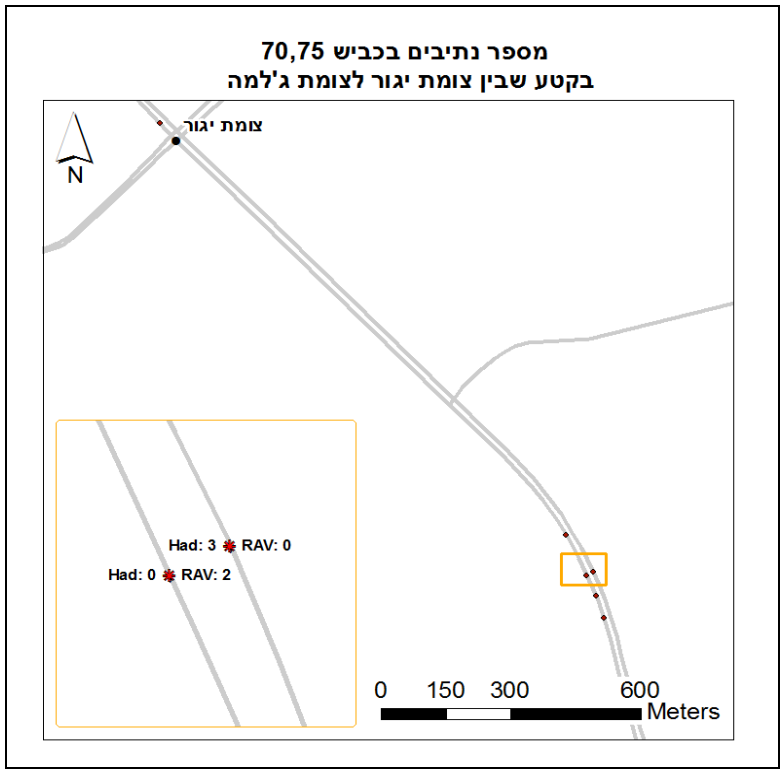
איור מספר 25 - סוגי מסלולים (מתוך: http://www.teoria-yoav.com/dagesh/D1.asp)

טבלה 12 - דרך רב מסלולית – התפלגות כמות אירועים וסוג

קוד	כמות אירועים	הגדרה
1	147	מפרדה מסומנת בצבע
2	1833	מפרדה עם גדר בטיחות
3	991	מפרדה בנוייה, ללא גדר בטיחות
4	35	מפרדה לא בנוייה (עפר/כורכר)
5	21	אחר

בדרך חד מסלולית קיימים בבסיס הנתונים 4432 אירועים. מתוכם, קודים 4 ו- 9 ("אחר" ו"לא ידוע", בהתאמה), מהווים כ 6% אחוזים. בדרך רב מסלולית קיימים בבסיס הנתונים 3027 אירועים. מתוכם, קוד 5 – אחר, מהווה פחות מאחוז אחד. המשמעות היא, כי מידע בתחום זה, לרוב, ממולא כראוי בדו"ח תאונות הדרכים.

בשלב האחרון, נותר לבחון את דיוקם ואמינותם של הנתונים בצורה ויזואלית-מרחבית. בדיקה זו היא המשמעותית ביותר, ובמהלכה נבחנו מספר כבישים מוכרים. כך לדוגמא, באיור מספר 26 ניתן לראות את כביש 75 בקטע דרך הסמוך לצומת יגור.



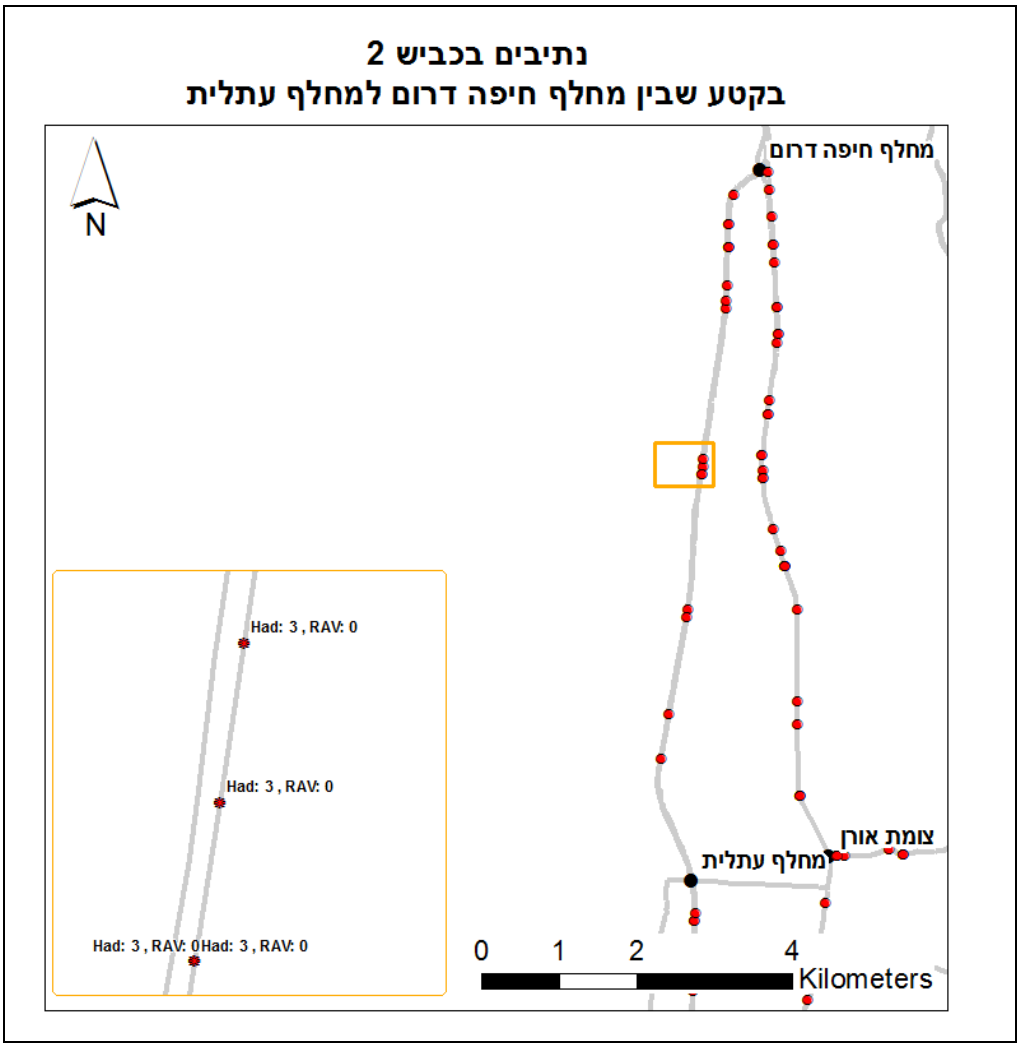
איור מספר 26 - מספר נתיבים בכביש 70, 75 בין צומת יגור לג'למה – דוגמא לאפיון רב מסלולי וחד מסלולי, כפי שמופיע בקובץ הלמ"ס.

במציאות, באזור זה של הדרך קיים מתאר דומה של הכביש, הן לכיוון צפון-מערב (מג'למה ליגור) והן לכיוון דרום-מזרח (מיגור לג'למה): חד סיטרי, עם שני נתיבים בכל כיוון, שבינהם קו הפרדה מקווקו[15] .

איור מספר 26 מראה בברור את חוסר האחידות בבסיס הנתונים עבור פרמטר זה. כאמור, מתאר הכביש דומה בשני הכיוונים. עם זאת, התאונות קודדו באופן הבא:
<u>האחת</u>, כרב מסלולית, ***קוד 2 = מיפרדה עם גדר בטיחות*** – מה שאכן נכון במקטע הכביש המדובר, <u>והשניה</u> כחד מסלולית, **קוד *3 =דו סיטרי אין קו הפרדה רצוף*** .
היות וברור כי לא מדובר במקטע כביש שהינו חד מסלולי (קיימים בוודאות 2 נתיבים), עולה כי התאונה קודדה באופן שגוי או שאולי כוונת ה"משורר" היתה כי מדובר בכביש רב נתיבי – חד סיטרי.

דוגמא נוספת לכך, הינה כביש החוף (כביש 2) במקרים רבים הכביש מקודד כ: HAD=0 ; RAV=2, כלומר:"**דרך רב מסלולית, מפרדה עם גדר בטיחות**". מאידך, מהיכרותנו עם כביש מספר 2, ברור כי המפרדה נמצאת בין הנתיבים הנוסעים בכיוון אחד, לבין הנתיבים בכיוון השני (כאשר לרוב קיימים שני מסלולים בכל כיוון). נדמה כי הגדרה של "חד סיטרי – רב מסלולי" מתארת את המציאות בצורה מוגדרת יותר, וחסרה גם במקרה זה. דוגמא לבעיה נוספת מוצגת באיור 27 כביש 2 בקטע הדרך הנע בין מחלף חיפה דרום, למחלף עתלית, ניתן למצוא תאונות המוגדרות באופן הבא: HAD=3דו-סטרי+אין קו הפרדה רצוף. שוב ברור כי ההגדרה שגויה.

[15] הנתוני חברת מפה, מאופן דיגיטציית (קליטת) הנתונים, ברורה מאוד, האבחנה בין כבישים בהם קיימת הפרדה פיזית, היות ונקלטים כשני קווים נפרדים (אחד בכל כיוון, ללא קשר למספר הנתיבים) לבין כבישים בהם לא קיימת הפרדה – נקלטים כ"קן אחד".



איור מספר 27 - מספר נתיבים בכביש 2 (החוף) בין מחלף חיפה דרום למחלף עתלית – דוגמא לאפיון רב מסלולי וחד מסלולי, כפי שמופיע בקובץ הלמ"ס.

המסקנה העיקרית העולה מניתוח פרמטר זה, הינה כי נדרש לבצע מחשב מחודש על קטגוריה זו. מוצע לפשט את ההגדרות, ובתחילה להגדיר קוד המפריד בין חד סיטרי ודו-סיטרי, ובעמודה נוספת לציין את סוג המפרדה הקיימת בדו סיטרי (מקווקו, מלא, מפרדה נמוכה – שאינה בנויה, וכיוצא בזאת). בנוסף, יש להשלים בעמודה נוספת או קוד אחר פרמטרים בנוגע ל: מספר הנתיבים, רוחב המפרדה והחומר ממנו בנויה המפרדה (בטון, מתכת, ניוגרזים וכיוצא בזאת). כך לדוגמא, ניתן יהיה למשקל בצורה אחרת, תאונה שהתרחשה במקום בו קיימת תנועה דו סיטרית עם מפרדה מצויירת בלבד, לבין תאונה שהתרחשה בדו סיטרי עם מפרדה בנוייה.

היות ונראה כי לפרמטר זה קשר משמעותי לתאונות דרכים (בהיבט השתלבות התנועה במסלולים והסתברות לתאונות חזיתיות), בייחוד במודלי ספירה של תאונות, והיות ובמקביל עולה התחושה כי נתון זה אינו ברור מספיק, אינו מדוייק ונתון לפרשנות רבה של בוחני התנועה, ולפיכך רחוק מלשקף את המציאות, הוחלט שלא לעשות בו שימוש. יתר על כן חסר מידע הנוגע למספר הנתיבים בקידוד של רב מסלולי.

4.2.7 קובץ הלמ"ס - רוחב השול

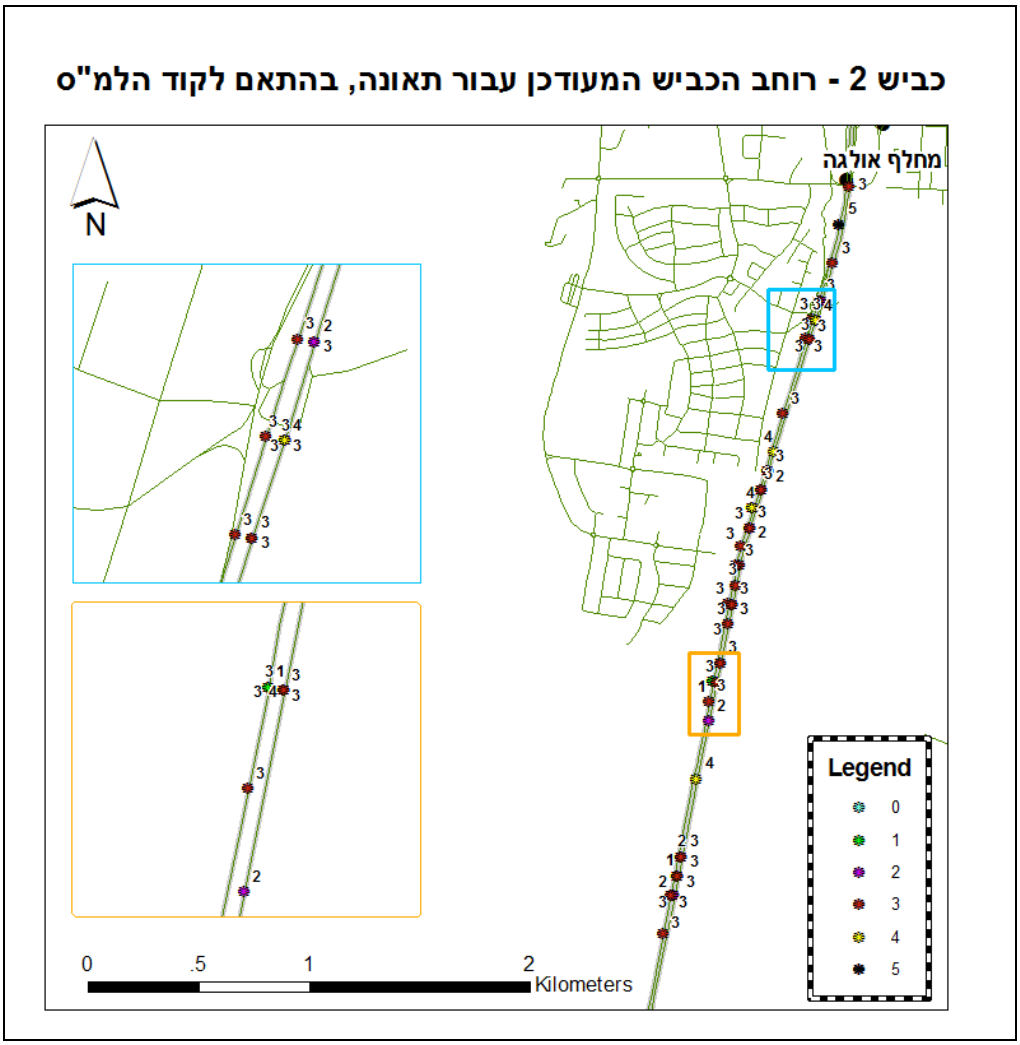
בחיי היום יום יוצא לכולנו להתקל בנהגים שאינם נוהגים בהתאמה לאופי הדרך, בעקיפות מסוכנות, בבלימה פתאומית. אלה ואחרים, גורמים לנו לסטות לשולי הדרך בכדי למנוע תאונה והיפגעות. במקרים מסוג זה, קיומו של שול רחב מונע פגיעה, היות ויכול לשמש כנתיב נוסף ל"מילוט". היות וכך, חשוב היה לקשור גם בין רוחב השוליים לכמות התאונות וחומרתן.

בנתוני תאונות הדרכים בקובץ הלמ"ס מופיעים הנתונים העוסקים ברוחב השול, כעמודה קטגוריאלית. טבלה מספר 13 מציגה את קטגוריות רוחב הכביש ואת מספר התאונות בהן מעודכנת קטגורית רוחב כביש זו, עבור המקטעים הבינעירוניים.

טבלה 13 -"רוחב הכביש" בפרסום הלמ"ס, והתפלגות מספר האירועים מכל סוג בכבישים הבינעירוניים

רוחב הכביש (מטרים)	קוד הלמ"ס	מספר תאונות בהם מעודכנת קטגורית רוחב כביש מסויימת
לא ידוע	0	161
עד 5	1	279
5 עד 7	2	1817
7 עד 10.5	3	4016
10.5 עד 14	4	796
יותר מ 14	5	390

במהלך בקרת האיכות, נבחנו הכבישים המתאפינים בקוד המרחק המקסימאלי. כבר מבדיקה זו, נראה כי פרמטר זה בעייתי. היות ונמצא כי עבור כביש 6 מעודכן בבסיס הנתונים הערך המקסימאלי, כאשר אחת הביקורות העיקריות בנוגע לכביש זה, הינה כי נסלל עם שוליים צרים במיוחד. על כן, נבחן ויזואלית גם פרמטר זה (ראה איור מספר 28 המציג את כביש 2 במקטע הסמוך לחדרה).



איור מספר 28 - רוחב כביש עבור כל תאונה, בהתאם לקוד הלמ"ס

איור מספר 28 מציג את הקושי בנושא. ניתן לראות כי תאונות שהתרחשו באותו מקטע דרך, עודכנו לעיתים כקוד 3 ולעיתים 4, במקרים אחרים, עודכנו התאונות כקוד 1 ובמקביל כקוד 3. מקרים אלה חוזרים ונשנים בבסיס הנתונים.

לאחר שהוברר כי נתוני הלמ"ס בעייתיים בהיבט האמור, נבחן האם ניתן יהיה להשלים את הנתונים ממקורות אחרים. לרשות המחקר עמדו, כאמור, נתוני חברת "מפה", נתוני הממ"ג הלאומי, ונתוני חברת "פלד מיפוי".

בנתוני חברת מפה לא מופיע מידע בנוגע לרוחב השול, כאשר בנתוני הממ"ג הלאומי מופיעים נתונים ששולבו ואינם מדוייקים. נראה כי נתוני הממ"ג הלאומי אוכנו בהתאמה לסדר הכביש, כלומר, כביש בעל סיפרה אחת קיבל ערך מסויים, כביש בעל שני ספרות, קיבל ערך אחר, וכן הלאה. נתוני חברת "פלד מיפוי" היו המדויקים ביותר (בוצעו מדידות בפועל), אולם נתונים אלה לא עודכנו עבור כלל כבישי הארץ (בוצעה מדידה חלקית).

עקב ההבנה כי נתון זה הינו בעל פוטנציאל רב, לשמש כנתון בעל קשר או חשיבות בהיבט כמות וחומרת התאונות הוחלט שלא לעשות בו שימוש. המסקנה המתבקשת הינה שיש להחזיק

פרמטר עדכני בבסיס הנתונים עבור רוחב השול במקום התאונה. אפשר ופרמטר זה **ימדד בפועל** על ידי בוחן התנועה (במדוייק) או שהגופים האמונים על הכנת בסיסי נתונים גיאוגרפיים, ישלבו בשכבות המידע המרחביות, **ערך מדוד** עבור כלל כבישי ישראל. זאת, על מנת שמחקרים עתידים יוכלו לעשות בו שימוש ולנתחו.

4.3 איכון תאונות הדרכים הבין עירונית, שאינן בצומת (קוד 4 של הלמ"ס)

4.3.1 כללי

ניתן לאפיין את ריכוז תאונות הדרכים המתרחשות בצמתים, בצורה השונה מהותית, מתאונות המתרחשות במקטעי כביש (להלן: סגמנטים). צמתים, עקב היותם מפגש של שני כבישים, מועדים הרבה יותר לתאונות דרכים.

המחקר שם לו למטרה לבנות מודל עבור מקטעי דרך, שאינם חלק מצומת (קוד תאונה בקובץ הלמ"ס: לא עירוני לא בצומת , Sug_Dereh=4).

בתחילתו של המחקר, עלתה המחשבה להתמקד גם בסוג נוסף של תאונות: עירוני לא בצומת (קוד למ"ס Sug_Dereh=2). אולם, כבר בשלב הראשוני של המחקר נראה היה כי עבור תאונות מסוג זה נדרש להתאים מודל ייחודי, שאינו עולה בקנה אחד עם המודל שיפותח עבור הכבישים הבינעירוניים. הפער בין המודלים הינו בראש ובראשונה בהיבט הפרמטרים המשפיעים השונים. אחד הפערים העיקרים בין הנתונים הקיימים עבור כבישים בינעירוניים לעומת כבישים עירוניים, הינו בנתון ה- AADT (המייצג, כזכור, נפח תנועה או ספירות תנועה במקטע דרך). בערים שנבדקו: ירושלים וחיפה, נפח התנועה אינו נתון זמין עבור צמתים מרכזיים בתקופת המחקר. לפיכך, לא ניתן לבנות מודל שלם בעיר. יודגש, כי מכלל הספרות שנבחנה בנושא, עולה כי פרמטר נפח התנועה, הינו לרוב, אחד הפרמטרים המשמעותיים המשולבים במודל. בנוסף, באזורים עירוניים קיימת חשיבות גם לנושא שימושי הקרקע, ולפרמטרים אחרים שאינם משפיעים על התנועה בדרך בינעירונית. נספח מספר 1 – מציג את הליך השמת התאונות במרחב העירוני, לרבות האימות שבוצע עם נתוני מד"א, לתאונות הדרכים של הלמ"ס.

השלב הראשון, הינו איכון התאונות במרחב הגיאוגרפי, כאשר בסופו של הליך איכון הנתונים, יתקבלו שני סטים של נתונים שינותחו. **האחד**, **בסיס נתונים נקודתי, בו** כל "נקודה" מייצגת תאונת דרכים, על כל הפרמטרים הקיימים במקור עבורה (בקובץ הלמ"ס). כן נוספו פרמטרים שחושבו עבורה במסגרת מחקר זה , דוגמת זווית שמש שחושבה למטרת בחינת סינוור. **הסט השני, הינו קווי**. בסיסו רשת הכבישים הראשיים של ישראל, שחולקה למקטעי כביש (להלן: סגמנטים) באורכים שונים: חצי קילומטר, שבע מאות וחמישים מטרים, קילומטר וקילומטר וחצי. עבור כל מקטע חושבו פרמטרים דוגמת: עקמומיות, שיפוע ממוצע ועוד, כפי שיפורט בהמשך.

בנוסף, לכל מקטע דרך שוייכו התאונות שהתקיימו בתחומו, ויבוצעו ניתוחים סטטיסטיים הקשורים לספירת מספר התאונות במקטע.

על מנת להציב את התאונות הבין עירוניות על רשת הכבישים, נבנה מודל אוטומטי להשמת התאונות. השיטה האוטומטית פותחה על בסיס שכבת Roads, שהתקבלה מ"פלד מיפוי", המכילה את הק"מ הרץ בצמתים. נדרש לעדכן את השכבה, היות וכבישים דוגמת כביש 6 לא הופיעו בשכבה זו.

ברמת העיקרון, בכל מקום בו נמצאה אי התאמה בק"מ המעודכן עבור צומת מסויים בין שכבת Roads לבין קובץ הלמ"ס המפרט את הק"מ בצמתים, עודכן הק"מ בצומת לפי קובץ הלמ"ס. טבלה מספר 14, מציגה את התפלגות נתוני תאונות הדרכים מסוג 4: בין עירוני – לא בצומת.

טבלה 14 -תאונות מסוג בין עירוני לא בצומת

שנת נתונים	מספר תאונות בבסיס הנתונים	מספר תאונות Sug_Dereh = 4	מתוכן KM=NULL	לא הוצב במרחב הג"ג**	סך תאונות שהוצבו במרחב *
2005	**17509**	2841	168	81	2592
2006	**17687**	2806	174	74	2558
2007	**16576**	2708	255	80	2373

** ההפרש הקיים בין סך התאונות ה"חוקיות" בבסיס הנתונים לבין התאונות שהוצבו בפועל במרחב, מתקיים בשל מספר סיבות. הראשונה והמשמעותית ביותר, הינה חוסר הלימה בין הק"מ בתחילת הכביש לבין הק"מ שעודכן עבור התאונה בבסיס הנתונים של המשטרה/הלמ"ס. לעיתים רבות, מסיבות הסטוריות, הצומת הראשון בכביש אינו מתחיל בערך ק"מ השווה ל- 0. כך, לדוגמא עבור כביש 4 בשנת 2005, 17 תאונות לא הוצבו היות והתרחשו בטווח של ק"מ הנע בין 0 ל- 22.9, כאשר בפועל (לפי קובץ הלמ"ס המציג את הק"מ בצמתים), הצומת הראשון של כביש 4 הינו צומת חבל עזה, שערכו ק"מ 51. סיבה נוספת הינה מחסור בנתוני הק"מ בין שני צמתים. מקרה זה מאפיין כבישים בעלי 4 ספרות.

4.3.2 הנתונים בהם נעשה שימוש עבור ההשמה במרחב

(א) Is_Main – שכבת הכבישים המרכזיים הבינעירוניים - חברת מפה. המרה למערכת קואורדינאטות גיאוגרפית, ולאחר מכן המרה להיטל ISRAEL TM.

(ב) ROADS – שכבת כבישים בינעירוניים בפורמט Coverage עם טופולוגיה של צמתים - פלד מיפוי. טבלת המאפיינים של קטעי הכביש כוללת, בין היתר, את המאפיינים הבאים:

- ROAD-NO1 – מספר הכביש של מע"צ

- KM-START – ק"מ בתחילת הקטע
- KM-END – ק"מ בסוף הקטע
- #fNODE - מספר הצומת הראשון בקטע
- #TNODE - מספר הצומת האחרון בקטע

(ג) שכבת ק"מ בצמתים - קובץ נקודות, שהתקבל מ"פלד מיפוי", מסוג COVERAGE המכיל עבור כל כביש את הק"מ המשוייך לצומת (עבור כל כביש). טבלת המאפיינים של הצמתים כוללת, בין היתר, את המאפיינים הבאים:

- JUNC.NAME – שם הצומת
- #ROADS - מספר הצומת

שכבה זו נקלטה, על ידי "פלד מיפוי" וכוללת כבישים בינעירוניים וגם כבישים עירוניים ראשיים. לפני תחילת העבודה נמחקו מהשכבה כל קטעי הכביש שאין להם מספר מע"צ. בנוסף, היו קטעי כביש עם ק"מ שלילי, הק"מ לקטעים אלה אופס. אם הערכים של שני המאפיינים של הק"מ שווים לאפס, אזי המשמעות היא שהק"מ איננו ידוע.

4.3.3 שלבים עיקריים בתהליך

(א) עדכון השדות JUN NAME , Y ,X

על בסיס שכבת Roads נבנתה טבלת צמתים (Junctions.dbf) המכילה את כלל הצמתים המופיעים בשכבה. המאפיינים העיקריים שמכילה הטבלה:

- JUN NAME - שם הצומת
- NumRoads – מספר הכבישים (בעלי מספר מע"צ) הנפגשים בצומת
- RoadNo1 – מספר הכביש (של מע"צ) של הקטע הראשון המחובר לצומת
- RoadKm1 - הק"מ של קטע הכביש הראשון המחובר לצומת
- RoadNo2 – מספר הכביש (של מע"צ) של הקטע השני המחובר לצומת
- RoadKm2 - הק"מ של קטע הכביש השני המחובר לצומת
- RoadNo3 – מספר הכביש (של מע"צ) של הקטע השלישי המחובר לצומת
- RoadKm3 - הק"מ של קטע הכביש השלישי המחובר לצומת

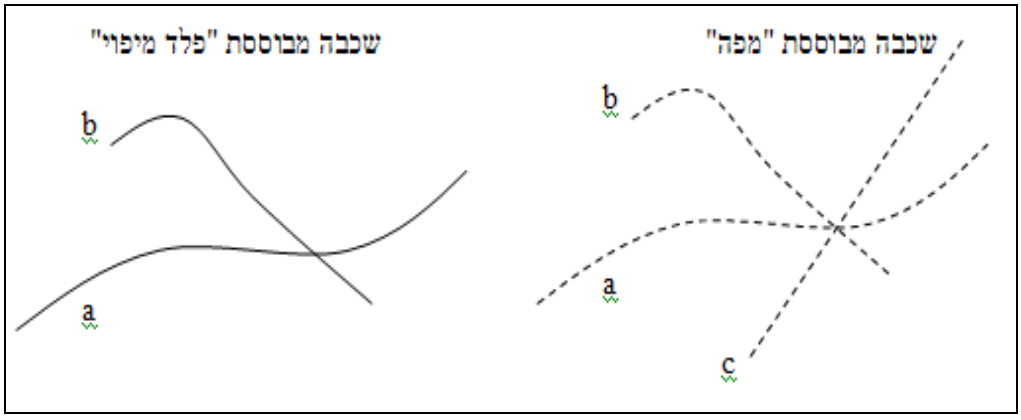
(ב) עדכון כלל הכבישים המצטלבים באותה הנקודה (אותו הצומת).

(ג) יצירת שכבה חדשה, נקודתית, בה כל נקודה מייצגת צומת (Juncs.dbf). שכבה זו מבוססת על נתוני הכבישים של חברת מפה. במידה ושדה FJUN מופיע יותר מפעם אחת ניתן לומר שמדובר בצומת בו נפגשים שני כבישים או יותר.

המאפיינים העיקריים שמכילה הטבלה:

- Name – שם הצומת (לפי JUNC.NAME)
- NumRoads – מספר קטעי הכביש הנפגשים בצומת (הכוונה רק לקטעים בעלי מספר מע"צ ועם ק"מ) .
- "1" – משמע מדובר בצומת פסדאו כביש "מצטלב" עם עצמו.
- 2 או 3 - צומת בו מצטלבים שני כבישים או שלושה כבישים.
- RoadNo1 ; RoadKm1 ; RoadNo2 ; RoadKm2 ; RoadNo3 ; RoadKm3 - כמפורט בסעיף א' בפסקה זו.
- סטטוס
- NumNodes – מספר ה"התאמות" של הצמתים בין השכבה המבוססת "מפה", לבין השכבה המבוססת "פלד מיפוי". האלגוריתם מחפש את כלל קומבינציות הכבישים הקיימות בשכבת "מפה" ומשווה אותן לאלו הקיימות ב"פלד מיפוי".

במצב המופיע באיור מספר 29 תזוהה קומבינציית a,b,c בשכבה של "מפה" לעומת a ו b בשכבה של "פלד מיפוי". היות וקיימת התאמה אחת (של a ו b) בין 2 השכבות, יעודכן ערך "1" בשדה NumNodes.



איור מספר 29 - התאמת צמתים לפי שילובי קטעי דרך על בסיס מספר 'כביש מע"צ'

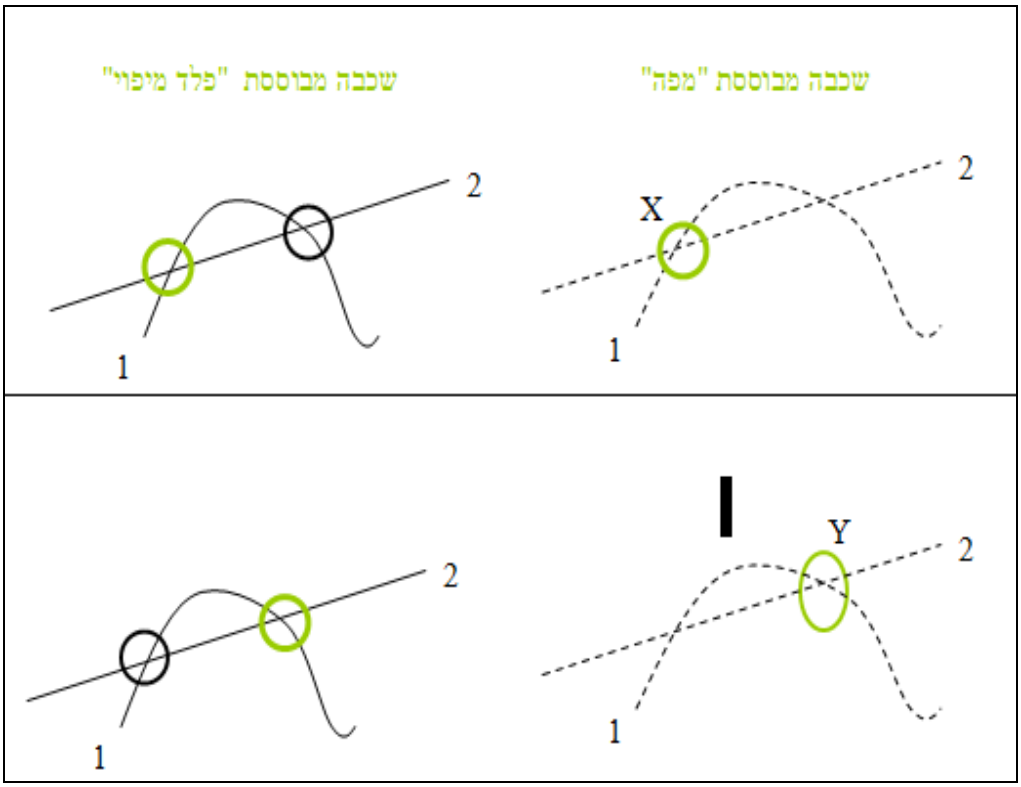
- distance - חישוב ההפרש, בין הנקודה שזוהתה בשכבת "מפה" לבין הנקודה שזוהתה בשכבה של "פלד מיפוי". שוב, שדה זה משמש למטרת בקרת איכות למקומות בעייתיים.
- בכל צומת מעודכנים מספרי מע"צ של הכבישים הנפגשים בו (עמודת RoadNoX) .בשלב זה נבדק האם הצומת מופיע בבסיס הנתונים. במידה ולא, מיוצר צומת חדש. כאמור, בכדי לדעת כמה כבישים מצטלבים בצומת מסויים (לעדכון שדהNumRoads) ניתן לסכום את שדה FJUN. כמו כן

מתעדכנים שמות הצמתים. וערכי הק"מ מאותחלים בערך "1-", על מנת לציין כי בשלב זה עדיין לא נקבע הק"מ עבור הצומת.

(ד) סינכרון השכבה שנוצרה על בסיסי נתוני חברת "מפה" (בשלב ג') עם השכבה שהתקבלה מ"פלד מיפוי", המכילה את הק"מ בצומת (נוצרה בשלבים א' ו ב') ועדכון השכבה Juncs.

- בכל אחת מהשכבות, עבור כל צומת, משווים את הערך המופיע בשדה NumRoads.
- במידה והערך הינו "1" (משמע מדובר בצומת פסאודו) משאירים את ערך "1-" בשדה הק"מ (RoadKm1) – מצב זה לא נבדק. במידה והערך שווה ל-2 או ל-3 (כאמור, באותו הצומת מצטלבים שניים או שלושה כבישים), נבדקת מהי "קומבינצית הכבישים" שזוהתה בנתוני חברת "מפה" אל מול "קומבינצית הכבישים" שזוהתה בשכבה של "פלד מיפוי". במידה ונמצאה קומבינציה מתאימה ("אחד לאחד"), לדוגמא: כביש 75 מצטלב עם כביש 77 גם בנתונים של חברת"פלד מיפוי" וגם בנתוני חברת "מפה**". כך, יעודכן ערך הק"מ לפי מה שמופיע בקובץ הנתונים של "פלד מיפוי" (שכבת Junctions).** כן יעודכן שדה NumNode, המייצג את מספר הצמתים שהתאימו אצל "פלד" וגם אצל "מפה".

- קיימים מקרים בהם קיימת התאמה של "אחד לרבים". ראה איור מספר 30 בחלקו העליון, בו כביש 1 מצטלב פעמיים עם כביש 2. במצב זה, האלגוריתם מחפש לעדכן ק"מ עבור צומת X (בשכבה המבוססת "מפה"), ו"מוצא" כי קיימים שני צמתים "מתאימים" בשכבה של "פלד מיפוי". אלו הן תוצאת החיתוך של כביש 1 וכביש 2. במצב מעין זה בוצע עדכון לפי המרחק הקרוב יותר. כך, עבור צומת X יעודכן הק"מ המתאים לו המופיע ב"פלד מיפוי" בצומת המוקף בירוק. כך גם יתקיים הדבר עבור צומת Y, המוצג בחלקו התחתון של האיור.



איור מספר 30 - קבלת החלטות לשיוך צמתים בין שכבת "מפה" ו"פלד מיפוי"

- בשלב זה מתבצע גם עדכון עמודת המרחק בין הצמתים (Distance) בין הצומת, שזוהה בשכבה שנבנתה על ידי "פלד מיפוי", לבין הצומת שזוהה בשכבה שנבנתה על בסיס "מפה". צמתים שימצאו במרחק הגדול מסף מסוים, יהיה צורך לבדוק מדוע, ואף לטפל אינטראקטיבית, במידת הצורך.
- בשלב זה מתקיים גם עדכון שדה הסטטוס (status):

1 – לכל הכבישים הנפגשים בצומת נקבע ק"מ (מצב תקין)

2 – לחלק מהכבישים בצומת נקבע ק"מ ולחלק לא. לדוגמא: בשכבה של "מפה" אותרו שלושה כבישים ואילו בשכבה המבוסס על "פלד מיפוי" אותרו רק שני כבישים עם ק"מ.

11 - צמתים וירטואליים (Pseudo): כביש אחד שלמעשה "מצטלב" עם עצמו (מתקיים לדוגמא ב"קצוות" כמו כביש 4 המסתיים בגבול הלבנון, או צמתים בהם "הוסרו" כבישים מסדר נמוך, שאין להם מספר מע"צ). צמתים אלו לא טופלו ונשאר בהם ערך של 1- בעמודת הק"מ.

12- לא נמצאה כלל התאמה. כל הק"מ עודכן ב "1-".

(ה) תיקון שכבת הצמתים בצורה ידנית - יש לקחת בחשבון כי האלגוריתם של הצבת ערכי הק"מ פועל על רצף של קטעי כביש השייכים לאותו מספר כביש המתחיל ומסתיים בצמתים בעלי ק"מ. במידה והרצף מתחיל בצומת עם ק"מ

ומסתיים בצומת ללא ק"מ, אזי לא מתבצעת הצבה של ערכי הק"מ. על כן, בשלב זה (4א), יש לקבוע ערכי ק"מ בצמתים בקצוות הכבישים ובהסתעפויות. התיקון וההשלמה של שכבת הצמתים בצורה אינטראקטיבית, בוצע על ידי איסוף חומר ממקורות מידע אחרים, כגון: מפות מע"צ או פרסומי הלמ"ס.

(ו) יצירת שכבת קטעי כביש חדשה ("Roads"), על בסיס שכבת הקטעים של חברת "מפה". בשכבה זו שוכפלו קטעי הכביש השייכים לשני מספרי כביש שונים (לדוגמא: כביש 1 וכביש 6 חולקים סגמנטים משותפים). לשכבה הוספו מאפיינים (טבלה 15), שחלקם נוספו למטרות אימות (Factor, Status Difference) כפי שיורחב בהמשך וחלקם למטרות ביצוע ה routing והצבת התאונות במרחב (דוגמת RoadNo, שהוא שדה ייחודי המורכב מצירוף של כיוון עם מספר הכביש 2-U וכיוצא בזאת).

טבלה 14 - עמודות שכבת הכבישים

FILED NAME	TYPE
FID	Object ID
Shape	Geometry
FJunciton	Double
TJunction	Double
Length	Double
oneway	Text
Roadnamemz	Text
DirectionM	Text
RoadNo	Long Integer
RoadID	Text
StartKm	Double
LastKm	Double
Direction	Text
Factor	Double
Status	Short Integer
Difference	Double

(ז) שכפול כל קטעי הכביש הדו-סטריים ("B="DirectionM) בשכבת roads, לשני קטעים חד סטריים.

(ח) עדכון השדות: Status , Factor , Difference בטבלת המאפיינים שנוצרה בשלבים הקודמים

Status - 0 – לא שוייך ק"מ (0=LastKm & 0=StartKm)

1 – שוייך ק"מ עבור הצומת

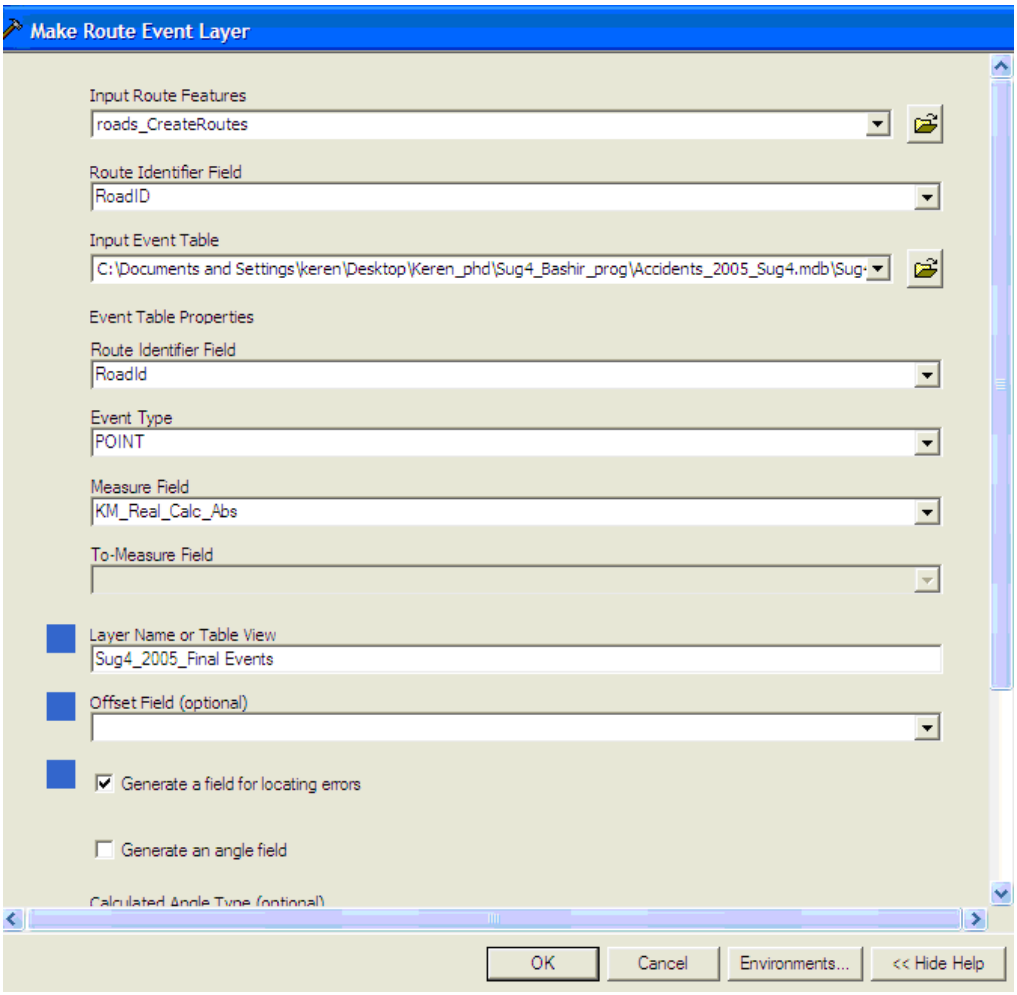
Factor – **היחס** בין האורך לפי הק"מ המוכרז עבור כל צומת לבין האורך לפי קואורדינטות של כלל הכביש (במציאות תוואי הכביש בגיאוגרפי). פקטור גדול מ 2 או 1.5 מרמז על בעייה אפשרית.

Difference – הפרש האורך לפי הק"מ המוכרז עבור כל צומת לבין האורך לפי הקואורדינטות של כלל הכביש.

(ט) הצבת הנקודות ב ArcMap

- בתום טיוב שכבת הק"מ בצמתים, בוצע הפיזור וההצבה של ה"נקודות" (תאונות הדרכים) על רשת הכבישים. זאת, תוך שימוש בכלים: Create Routes ו- Make Route Events, בתוכנת Arc Map.
- **יצירת מסלולים (Create Routes) מ"אלמנטיים קווים"** - בשלב זה חושב ערך M, המייצג מרחק עבור כל VERTEX (נקודת מפנה) לאורכו של האלמנט הלינארי. בכדי ליצור מסלול ולחשב את ערכי ה- vertices לאורכו, יש לוודא כי קיים בקובץ ה- input שדה identifier, המשמש לזיהוי האלמנט הליניארי. ה- Identifier מופיע בשכבת Roads בשדה RoadId ומורכב ממספר הכביש +כיוון.
- שימוש בכלי ה- **Make Route Events - להשמת התאונות לאורך הכביש**

הערה לגבי שדה Measure filed: יש לוודא כי היחידות הינן בקילומטרים. יש לשים לב כי יחידות עמודת **הק"מ** בקובץ הלמ"ס המקורי הינן במאות מטרים (ערך של 20 בשדה זה משמעו 2.0 ק"מ, ערך שלילי בעמודה זו, משמעו שערך המרחק המדוייק אינו ידוע, לדוגמא: '30-' משמעו כי התאונה התרחשה בין ק"מ 3 לבין ק"מ 3.1) . לפיכך נוספה עמודה המציגה את הערכים בק"מ (בערכים מוחלטים). זאת, על מנת שבשלב מאוחר יותר, בעת חישוב המסלולים והצבת התאונות תוך שימוש ב Make Route Events (איור מספר 31), ניתן יהיה לעשות שימוש בתוכנת ARCMAP ולחבר בין היחידות.



איור מספר 31 - חלון "MAKE ROUTE EVENT LAYER"

4.4 חישוב פרמטרים עבור אירוע תאונת דרכים (נקודה)

4.4.1 שיוך "כיוון נסיעה" לתאונה.

המחקר שם לעצמו מטרה לזהות ולהגדיר רמת סיכון בקטעי דרך שונים. בהתאמה לכך קיימת חשיבות להצבת הנקודות בכיוון ה"נכון" או ה"אמיתי ככל האפשר" בו התרחשה התאונה. פרוט בעניין הכיוון בכביש בו התרחשה התאונה, מפורט בקובץ הכלי רכב של הלמ"ס, כאשר מופיעות בו הקטגוריות הבאות: 0 = לא ידוע ; 1 = עם כיוון הספירה (הק"מ ה"עולה") ; 2 = נגד כיוון הספירה (הק"מ ה"יורד") ; 3 = רכב חונה. בנתונים הגיאוגרפיים, של חברת מפה, בעמודת "DirectionM", מפורט כיוון הכביש באופן הבא: U = עם הק"מ העולה ; D = עם הק"מ היורד ו B = קטע כביש המשמש לנסיעה לשני הכיוונים[16].

[16] כדי להתאים ל"מפה"3 – עדכון: הערך 1 ב – U ; הערך 2 ב D ; הערכים 0 ו 3 ב B .

עבור כל תאונה בהיבט כיוון הנסיעה נבחנו כלל הצירופים. כך, אותרו מקרים בהם עבור אותה תאונה (במידה וקיימים מספר כלי רכב), מעודכן ערך כיווניות של 1 ובנוסף ערך כיווניות של 2. הסבר אפשרי לכך הינו שכלי הרכב "עף" לכיוון ההפוך. הסבר נוסף אפשרי, הינו טעות רישום של בוחן התנועה או טעות כלשהי בקידוד הנתונים ב"מעלה הדרך". בנוסף, נדגמו גם צירופים שונים, דוגמת: 0, 1,3 ועוד.

היות והכוונה היא לייצג כל תאונה במרחב הגיאוגרפי, כנקודה אחת, חושבו מינימום ומקסימים עבור כל תאונה, כאשר עבור כל צירוף של מינימום ומקסימום נבחן איזה ערך כיווניות יש להגדיר. יודגש, כי ברור היה מהתבוננות ראשונית בנתונים, כי בקרב בוחני התנועה או נציגי המשטרה הממלאים את טפסי התאונה, קיימת נטייה למלא את הערך 1.

על אף שדיוקו של נתון הכיווניות, מוטל בספק, הוחלט לעשות בו שימוש. זאת, היות ואלו הנתונים הקיימים בפועל כיום בבסיס הנתונים של הלמ"ס. טבלה מספר 16 מציגה את ערך ה"כיווניות" (מינימום ומקסימום) המעודכן עבור כל תאונה בבסיס הנתונים. זאת, תחת ההנחה שמעורבים מספר כלי רכב באותה התאונה.

טבלה 16 - ערכי קוד מינימום ומקסימום ל"כיווניות" תאונות דרכים בינעירוניות

מינימום	מקסימום	החלטה
0	0	1
0	1	MAX
0	2	MAX
0	3	1
1	1	1
2	2	2
1	2	1
1	3	MIN
2	3	MIN
3	3	1

4.4.2 חישוב זווית השמש עבור נקודה במישור אופקי

4.4.2.1 כללי

אחד הפרמטרים הייחודיים שנבחן במחקר, הינו פרמטר סינוור השמש. זאת, היות ולעיתים רבות בעת נהיגה בדרך, נתקלים נהגים בשדה ראיה הנחסם עקב סינוור השמש. זאת, הן במצב בו השמש מוקמה ישירות אל מול כיוון הנסיעה, והן במצבים בהם השמש מצוייה בצידו של הרכב, כך שלא ניתן היה כלל לעשות שימוש במראות הצד של הרכב. בהינתן תאריך התרחשות התאונה, מיקומה (קואורדינאטות אורך ורוחב) ושעת התאונה, ניתן לחשב את מיקום השמש (גובהה ביחס לאופק ואת אזימוט מיקומה). פרק זה יתאר בהרחבה את נוסחאות החישוב, שמומשו בהתאם למאמר שפורסם ב: http://answers.google.com/answers/threadview/id/782886.html.

בקרת האיכות בוצעה מול תוצאות ודוגמאות חישוב שבוצעו בשני אתרים. הראשון, אתר NOAA בכתובת *http://www.srrb.noaa.gov/highlights/sunrise/azel.html*, השני מול אתר *http://www.pveducation.org/pvcdrom/properties-of-sunlight/sun-position-calculator*.

שאילתות ה SQL שנכתבו עבור המימוש מצורפות כנספח מספר 2. יודגש, כי עד כה בוצעו החישובים במערכת UTM. לפני תחילת הרצת כלל השאילתות והנוסחאות, יש לחשב עבור כל תאונה בבסיס הנתונים, קואורדינאטות במעלות עשרוניות. בכדי לבצע זאת, יש לשנות את ה- DataFrame לדאטום הגיאודטי WGS84 לייצא ולשמור כשכבה חדשה. בהיטל זה ניתן לחשב באופן אוטומטי, תוך שימוש בתוכנת ARCGIS, את הקואורדינאטות במעלות עשרוניות. עבור חישובים אלה, הוספו שני שדות מסוג FLOAT עבור X ועבור Y.

4.4.2.2 המרת זמן מפורמט של יום, שנה ושעה לערך של מעלות (ערך G)

שלב זה מתבצע בשני תתי שלבים. הראשון, חישוב ערך N – מספר היום בשנה מסויימת, כמוצג בנוסחה מספר 60:

[60] **N = day number ; January 1 = day 1 ; January 2 = day 2, etc.**

כך לדוגמא, ה 15 בנובמבר 2006, הינו יום מספר 319. בשלב השני, יש לבטא את השעה בצורה עשרונית, בהתאם לנוסחא מספר 61, בה מוצג כדוגמא המרת ערך השעה 10:35 לחלקי השעה:

[61] **X(h)+Y(min)/60 ➔ example: 10h 35m = 10 + 35/60 = 10.58333**

נוסחה מספר 62 מציגה את נוסחת ההמרה משנה ושעה לערך מעלות.

[62] **g = (360/365.25)*(N + hour/24)**

כך לדוגמא, יומר ערך היום, השנה והשעה (G) עבור תאונה שהתרחשה ב 15 בנובמבר 2006 ל :

g = (360/365.25)*(319 + 10.58333/24) = 314.849

בתאונות שהתרחשו בזמן שמופעל "שעון קיץ", יש להוריד שעה אחת טרם תחילת החישוב. לפי חוק שעון הקיץ, מידי שנה יחל שעון הקיץ ביום שישי האחרון שלפני ה-2 באפריל בשעה 02:00 ויסתיים ביום ראשון האחרון שלפני י' בתשרי בשעה 02:00. כך, בתאונות שהתרחשו בין התאריכים (שצויינו לעיל, ויפורטו להלן) הופחתה שעה.

עבור התאונות שהתרחשו ב 2005 , נדרש להפחית שעה בין התאריכים **1.4.2005** לבין **9.10.2005** (היות והתאריך הלועזי של י' בתשרי תשס"ו הינו 13.10.2005, ויום ראשון שלפניו הוא התשיעי בחודש).

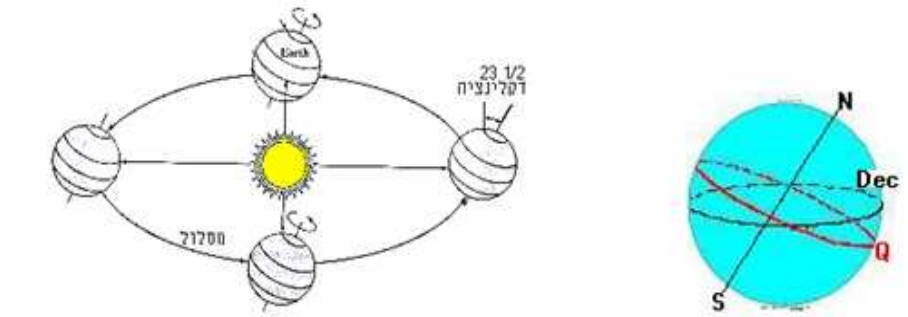
עבור התאונות שהתרחשו בשנת 2006 – נדרש להפחית שעה בין התאריכים **31.3.2006** – לבין **1.10.2006** (היות והתאריך הלועזי של י' בתשרי תשס"ז הינו 2.10.2006 , ויום ראשון שלפניו הוא הראשון בחודש).

בשנת 2007 - נדרש להפחית שעה בין התאריכים 30.3.2007 – לבין 16.09.2007 (היות והתאריך הלועזי של י' בתשרי התשס"ח הינו 22 בספטמבר 2007, ויום ראשון שלפניו הוא הראשון בחודש).

בנתוני תאונות הדרכים עבור כל תאונה מתקבל טווח ערכים (מסוג טקסט) עבור שעת התרחשות התאונה (ולא ערך בדיד). לפיכך, נבחר המייצג את הערך המרכזי (על ידי הוספת 7 דקות לערך הדקות המינימאלי).

4.4.2.3 חישוב הדקלינציה של השמש (declination)

כדור הארץ סובב סביב עצמו וסביב השמש. נטיתו של ציר סיבוב כדור הארץ סביב עצמו אינו מאונך למסלול הסיבוב סביב השמש אלא נטוי בזוית של כ- 23.5 מעלות. הדקלינציה היא הזווית שבין השמש לבין מישור קו המשווה, כאשר מישור הדקלינציה הוא המישור הנוצר מסיבוב כדוה"א סביב השמש (ראה איור 32).



איור מספר 32 - מישור הדקלינציה [מקור: http://cgate.co.il/navigation/declination.htm]

נוסחה מספר 63 מציגה את אופן חישוב זווית הדקלינציה:

[63] $$D = 0.396372-22.91327*\cos(g)+4.02543*\sin(g)-0.387205*\cos(2*g)+0.051967*\sin(2*g)-0.154527*\cos(3*g) + 0.084798*\sin(3*g)$$

בדוגמא D = -18.6162. נדגיש, כי על מנת לחשב שלב זה, יש להכניס לנוסחה, ערך של G ברדיאנים.

4.4.2.4 חישוב תיקון זמן עבור זווית השמש.

בכדי לבצע תיקון זמן עבור ההפרש בין זווית השמש לזוית כדוה"א הופעלה נוסחה מספר 64, וחושב ערך TC (Time Correction)

[64] TC = 0.004297+0.107029*cos(g)-1.837877*sin(g)-0.837378*cos(2*g)-2.340475*sin(2*g)

בדוגמא לעיל: TC = 3.72763. מודגש שוב, כי על מנת לחשב שלב זה, יש להכניס לנוסחה, ערך של G ברדיאנים, כאשר ערך תיקון הזמן (TC) מתקבל לבסוף במעלות.

4.4.2.5 חישוב Solar Hour Angle (SHA)

Solar Hour Angle (זווית השעה) הינה המרחק הזוויתי שכדוה"א עובר במהלך יממה. נוסחה מספר 65 מציגה את חישוב ה SHA:

[65] SHA = (hour-12)*15 + Longitude + TC

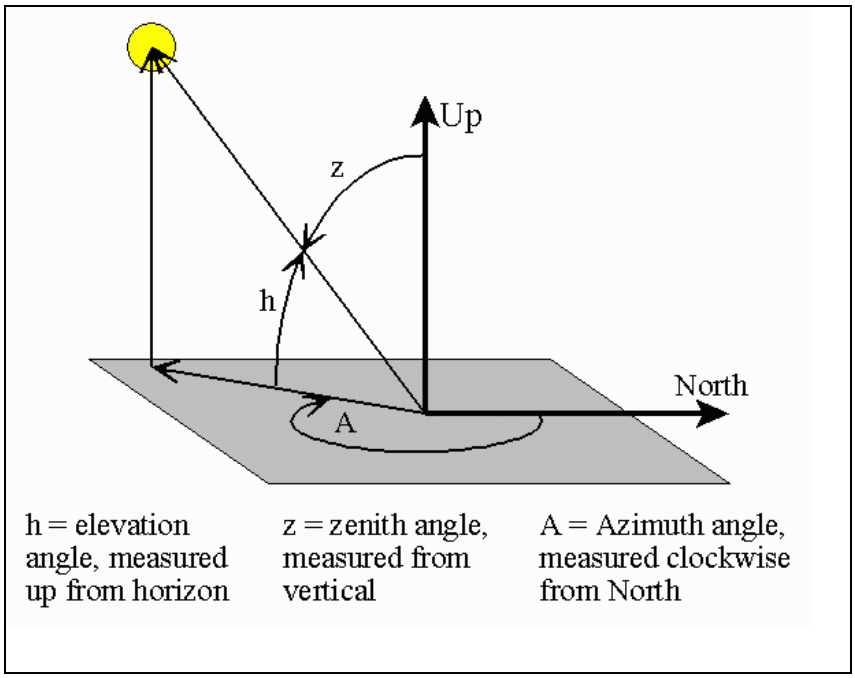
נציין, כי ה- Longitude הינו קו האורך במעלות (חיובי עבור אזורים מזרחית לגריניץ'), אותו יש להציב במעלות עשרוניות (DD). נוסחה מספר 66 מציגה כיצד ממירים מפורמט של מעלות דקות שניות למעלות עשרוניות.

[66] 02°17'23 E = 2+17/60+23/3600 = 2.28972

כך בדוגמא שלנו: SHA = (10.58333-12)*15 + 2.28972 + 3.72763 = -15.2327 יודגש, כי במידה ו- SHA גדול מ 180, יש לגרוע 360º מהתוצאה, ואם ה- SHA קטן מ- 180º-, יש להוסיף 360º לתוצאה.

4.4.2.6 חישוב SZA (Sun Zenith Angel)- זווית/זנית השמש ביחס לאופק (אנך לקרקע)

זנית הינה מצב בו השמש נמצאת בניצב (90 מעלות) לנקודה על פני כדוה"א. איור מספר 33 מציג את ה - SZA כלומר את זווית השמש ביחס לאנך מהקרקע (ה Z באיור, הנמדד מקו אנכי לנקודה בכדוה"א), כאשר נוסחה מספר 9 מציגה את אופן חישוב ה SZA.



איור מספר 33 - חישוב כיוון השמש ביחס למישור אופקי

[67] **cos(SZA) = sin(Latitude)*sin(D)+cos(Latitude)*cos(D)*cos(SHA)**

ניקח לדוגמא נקודה בקו רוחב 48°49'N, ונמיר קו הרוחב ממעלות ודקות למעלות עשרוניות (DD) כפי שמוצג בנוסחה מספר 68:

[68] **Latitude = 48°49'N = 48 + 49/60 = 48°.81667**

כך לדוגמא:

cos(SZA) = sin(48.81667)*sin(-18.6162)+cos(48.81667)*cos(-18.6162)*cos(-15.2327)
= 0.36184

נדגיש כי, כל הזווית בנוסחה 67, הינן ברדיאנים. לאחר שחושב COS בזווית, ניתן לחשב את ה SZA (Sun Zenith Angle) לפי arcos, כמוצג בנוסחה מספר 69. במחקר מומש החישוב באקסל, היות ופונקציית ARC COS אינה קיימת בגרסת ה - ACCESS בה נעשה שימוש.

[69] **SZA = arccos(0.36184) = 68.78676**

יש לשים לב כי הזווית עבורה מחשבים arccos (בדוגמא 0.36184) הינה ברדיאנים, כאשר התוצאה מתקבלת במעלות. על מנת לחשב זאת נכון באקסל הרצנו את נוסחה 70 :

[70] **DEGREES(ACOS(0.36184))**

4.4.2.7 חישוב SEA (Sun Elevation Angle)

זווית גובה השמש (H באיור מספר 33) הינה למעשה הזווית המשלימה ל SEA. זווית זו מאפשרת לבחון את הסינוור עבור נהג בכביש, ולשלב אותו כפרמט נוסף בבסיס הנתונים. מחושבת בנוסחה מספר 71:

[71] SEA = 90- SZA = 90-68.78676 = 21.213°(21°12 '48")

4.4.2.8 אזימוט השמש Azimuth Angle (AZ)

[72] cos(AZ) = (sin(D)-sin(Latitude)*cos(SZA))/(cos(Latitude)*sin(SZA)) = -0.96367

הערכים בנוסחה הינם ברדיאנים.

[73] AZ = arccos(-0.96367) = 164.50848 (164°30'30. 5")

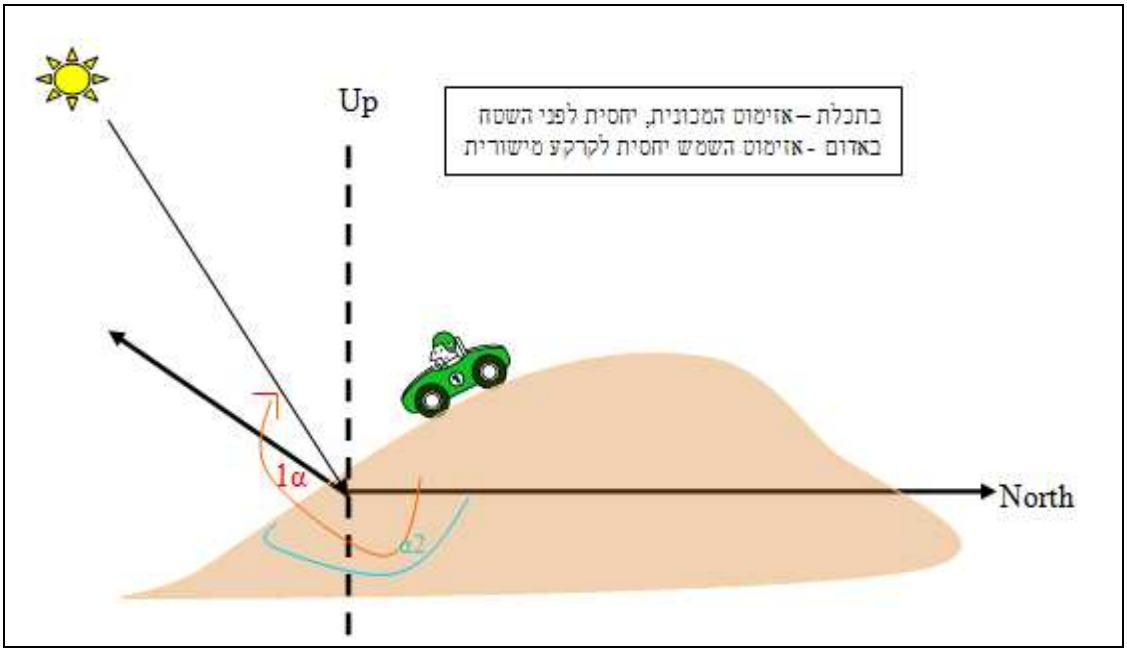
יש להקפיד ולבחון את נוסחאות חישוב האזימוט וזוויות גובה אל השמש, ובמיוחד, המעברים מהמערכות הסלסטיאליות[17] ושעות הזמן, אל מערכות האופק המקומי שבהן אנו מחשבים בפועל את האזימוט ואת זוויות הגובה אל הגופים השמימיים. במקרה שלנו, מדובר בהתמרה אל השמש. נוסחאות ההתמרה הסופיות, המופיעות במרבית המקורות בספרות, בנויות על חילוץ הזוית דרך ה Cos שלה, בחילוץ פשוט של משפט הקוסינוסים מהמשולש הספרי שנוצר. על כן, יש לשים לב, שהתוצאות המתקבלות, תהינה גבולות מ- 0 ועד 180. מכאן, שיש לבחון את תופעת הקוסינוס של זוית α, השווה לקוסינוס של מינוס α. הווה אומר, להתאים את הזווית בצורה סימטרית סביב המרידיאן לפי השעה 12:00. המשמעות המעשית הינה שכל אזימוט אשר מתקבל עבור מיקום בשעה שבין 12:00 ל 24:00, יש להטילו סימטרית סביב ציר הזמן של השעה 12:00 וסביב מרידיאן הדרום באזימוט 180 מעלות. כך, עבור תאונות שהתרחשו לאחר השעה 12:00 בצהריים, בכדי למצוא את אזימוט השמש הנכון, יש לבצע הפחתה באופן הבא: 360°-AZ .

4.4.2.9 חישוב זווית השמש עבור נקודה במישור משופע

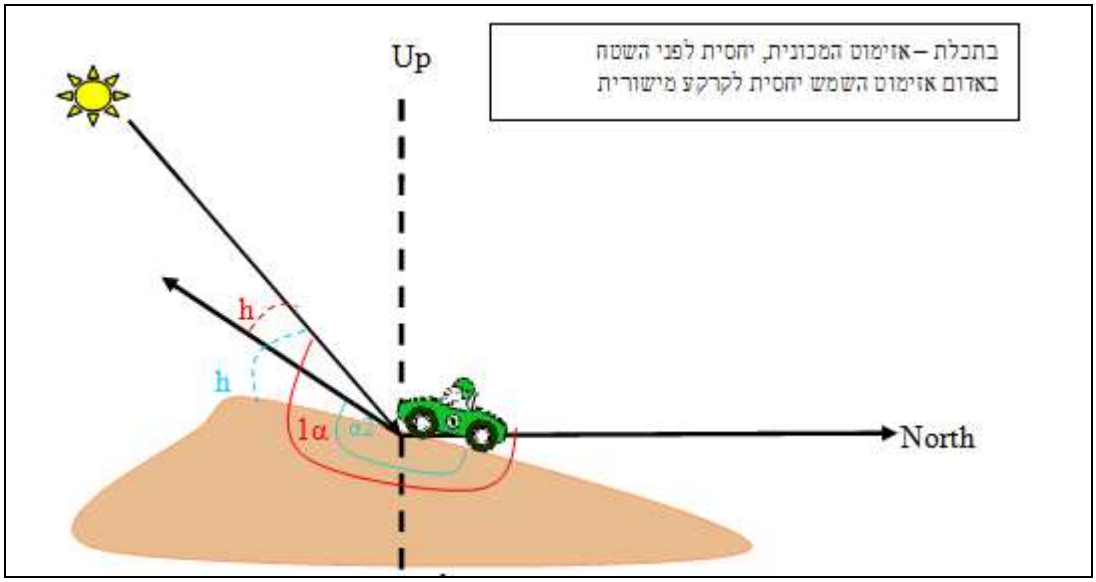
עד כה חושבה הזווית h, שהינה הזווית המתקבלת בין מיקום השמש בשמים לבין מישור אופקי (תוך הנחה של פני קרקע שטוחים. בשלב הבא, היות וכלי רכב, לרוב אינם נוסעים במישור אופקי, יש לחשב את כיוון הכביש לפי הקטע בו נמצאה המכונית

[17] שמימית

(האזימוט של הרכב) ואת השיפוע של הכביש (ממנו נחלץ את ה h עבורו). כמוצג באיורים מספר 34 ו- 35.



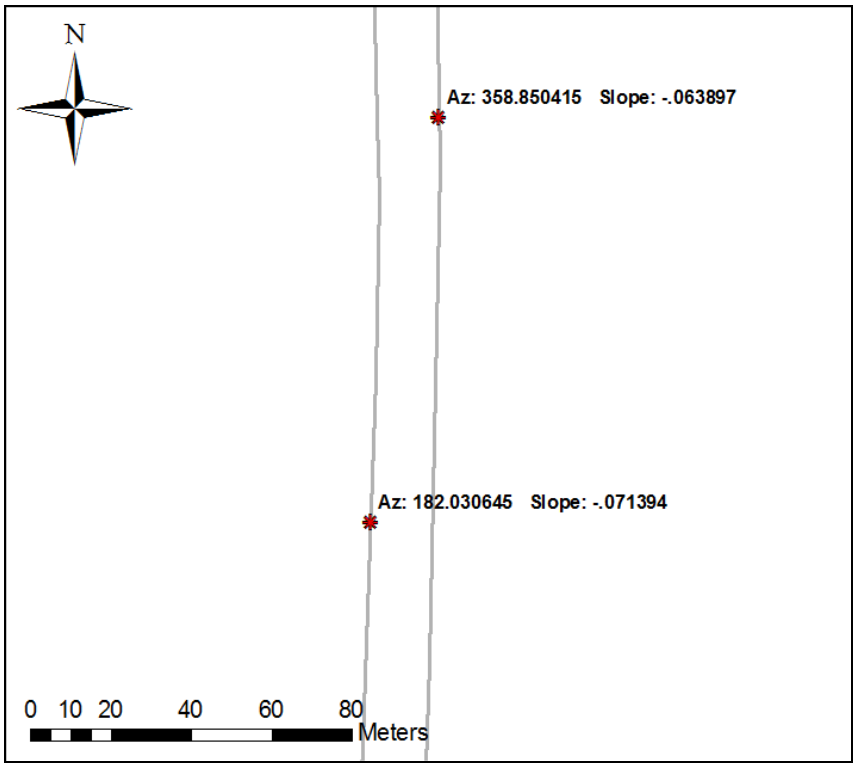
איור מספר 34 - דוגמא לצורך בחישוב כיוון הכביש (אזימוט הרכב) ושיפועו ביחס לקטע בו נמצאת המכונית



איור מספר 35 - דוגמא לצורך בחישוב כיוון הכביש (אזימוט הרכב) ושיפועו ביחס לקטע בו נמצאת המכונית

4.4.2.10 היחס בין זווית השמש בנקודה לבין זווית השמש בהתחשב בשיפוע ובכיוון הנסיעה.

על מנת לדעת האם סינוור משפיע על התאונה, יש להגדיר את היחס בין שני פרמטרים. הראשון: היחס בין אזימוט השמש לבין האזימוט בו נמצא כלי הרכב המעורב בתאונת הדרכים. השני, שיפוע הרכב (במעלות) לבין שיפוע/זווית השמש בשמים.
האזימוט חושב לפי נתון הכיווניות בתאונת הדרכים (על אף שברור היה כבר בשלבים הראשונים, כי רמת הדיוק הנתון זה נמוכה, מבחינה מתודולוגית, קיימת חשיבות בפיתוח היכולת). איור מספר 36 מציג את אזימוט כלי הרכב (קרי: כיווניות כלי הרכב ביחס לצפון).



איור מספר 36 - אזימוט כלי הרכב (כיווניות כלי הרכב ביחס לצפון)

כפי שניתן לראות מאיור מספר 36, בכביש צפון דרום (דוגמת כביש 4, 2, 90 וכו') מכונית הנוסעת בכיוון צפון, נוסעת בקירוב באזימוט הקרוב ל 0º/360º. מכונית הנוסעת בכיוון דרום, נוסעת באזימוט 180º.

על מנת לחשב את הסינוור באזימוט (במישור הסבסוב, קרי: ימינה שמאלה) – נניח כי רכב הנמצא בטווח אזימוט השמש +/- 15º, הינו רכב בעל פוטינצאל סינוור. היות, וכאמור חושב אזימוט כלי הרכב במישור, נחשב שתי עמודות נוספות (לדוגמא: SunAZMin & SunAZMax) המכילות ערך מחושב של אזימוט השמש פלוס

15º ואזימוט השמש מינוס 15º. בשלב הבא נבחן האם ערך אזימוט כלי הרכב נמצא בטווח אזימוט השמש +/- 15º.

לאחר מכן, נבחנה זווית גובה השמש ברקיע h או SEA (Sun Elevation Angle), הניתנת להקבלה למישור עלרוד (="אף המכונית"). גם כאן, ניתן להניח כי שיפוע הרכב הנע בטווח של +/- 15 מעלות ביחס לזווית השמש הינו בעל פוטנציאל סינוור . גם כאן, נוספו שתי עמודות המכילות ערך מחושב של זווית גובה השמש ברקיע, ועוד 15º , וזווית גובה השמש ברקיע פחות 15º (לדוגמא:CarSlpMin &CarSlpMax), לאחר מכן נבחן האם ערך שיפוע הרכב נמצא בטווח זווית גובה השמש ברקיע +/- 15º.

אירועי תאונות הדרכים, בהם מתקיימים שני התנאים שצויינו לעיל, גם יחד, ניתן להניח כי אפשר והתרחשו בגין סינוור. במחקר לא מומש הלכה למעשה שלב איתור הצבירים של תאונות המושפעות מסינוור, וזאת משני טעמים. הראשון, שכבר צויין, הינו האבחנה הברורה כי ערך ה"כיווניות" המעודכן עבור כל תאונה בבסיס הנתונים אינו מדוייק. השני, בגין האופן בו נרשמת שעת התאונה בבסיס נתוני הלמ"ס. כל תאונה בבסיס הנתונים מעודכנת כאילו התרחשה בטווח של רבע שעה. כך לדוגמא, תאונה שהתרחשה בשעה 12:08 דקות, תשוייך לקוד שעה המייצג את טווח זמנים הנע בין 12:00 ל 12:15, כאשר על מנת לחשב את מיקום השמש אנו נדרשים להגדיר שעה מדוייקת יותר. במחקר נבחר הערך המרכזי. כלומר, עבור תאונה שהתרחשה בין השעות 12:00 ל 12:15, עודכן כאילו התרחשה בשעה 12:07. עדכון השעה באופן הנ"ל הוביל לסטיה די משמעותית האזימוט הסופי. כך לדוגמא, עבור תאונה שהתרחשה בצהרי היום, כאשר השמש נמצאת בדרום האזימוט הינו 180º, כאשר האזימוט שחושב עבור תאונה בבסיס הנתונים שהתרחשה בשעה זו, מתקבל קרוב יותר לערך של 172º.

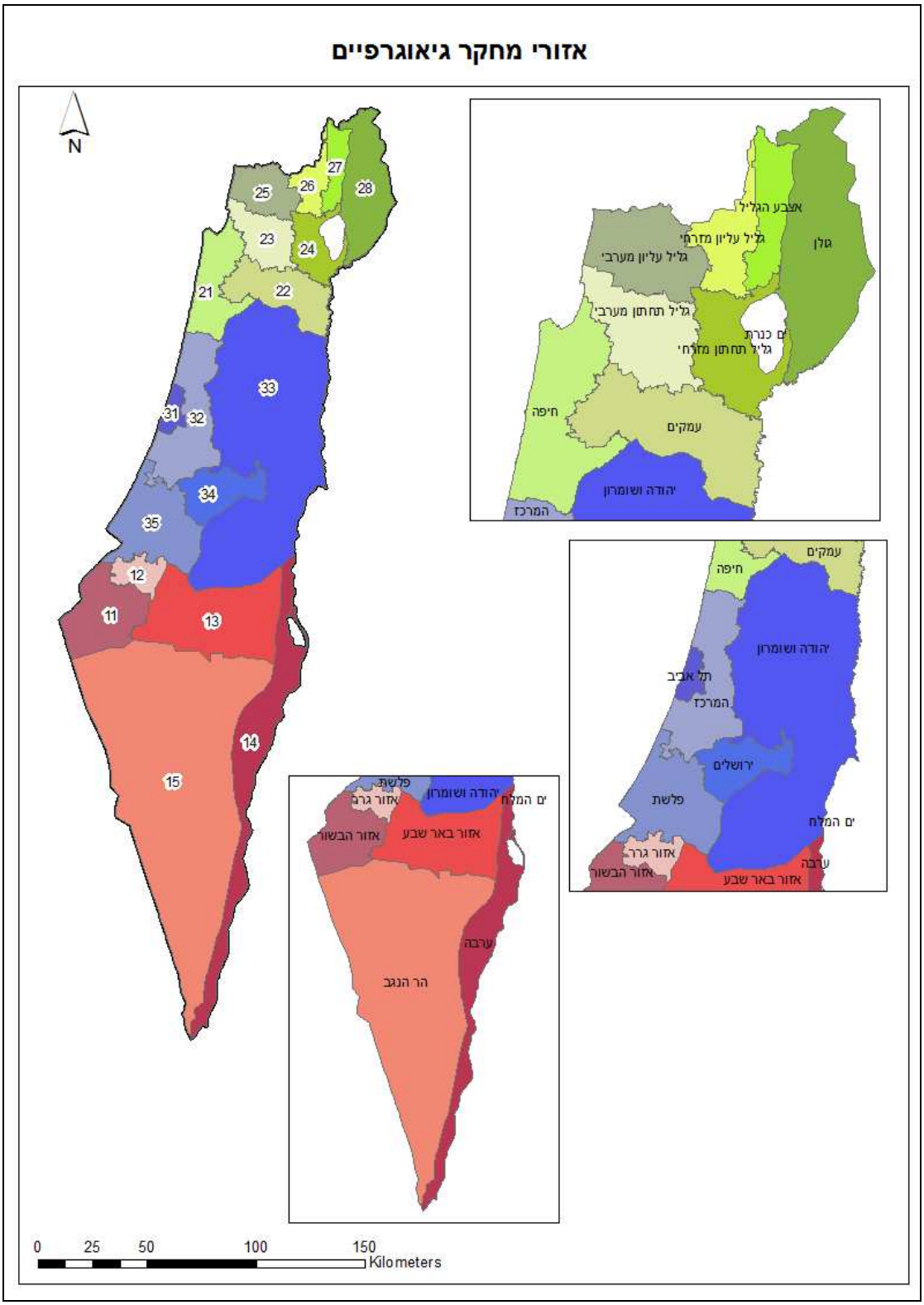
4.4.3 הוספת הערך "אזור טבעי" עבור כל תאונה בבסיס הנתונים

במהלך המחקר נבחן גם האם ניתן לאפיין או לקשור בין מאפייני תאונות דרכים לבין המיקום המרחבי-גיאוגרפי בו התקיימו. על מנת לבחון זאת, שייכנו לכל נקודה בבסיס הנתונים, פרמטר המייצג "אזור טבעי-גיאוגרפי". פרמטר זה, מורכב מערך המחולק לשניים:

ספרת העשרות - מצביעה על אזור גיאוגרפי רחב: צפון - 2 ; מרכז - 3 ; דרום - 1

ספרת האחדות – חלוקה פנימית לאזורי משנה בתוך האזור הג"ג.

איור מספר 37 מציג את מפת החלוקה לאזורים גיאוגרפיים.



איור מספר 37 - מפת חלוקה לאזורים גיאוגרפיים

נציין, כי שכבת הנתונים (בפורמט SHP) המקורית של מחוזות משרד הפנים. עברה שינויי באזור מחוז דרום. מחוז זה פוצל במסגרת המחקר, ונוצר "מלאכותית", מחוז נוסף שישמש במחקר

בשם: פלשת. מחוז זה מכיל את האזורים: אשדוד, מלאכי, לכיש ואשקלון, כפי שנלקחו ממפת האזורים הטיבעיים.

על מנת לספק ערך של "אזור גיאוגרפי" לכל אלמנט קווי בשכבת הכבישים לאחר חלוקתם לסגמנטים בעלי אורך דומה, בוצעה על שכבת התאונות פעולת Spatial Join עם השכבה הפוליגונאלית של המחוזות. לאחר ביצוע האיחוד, שוייך לכל תאונה בבסיס הנתונים המחוז שלה. בשלב ההחלטה כיצד לעדכן מקטעי כביש שחלקם עובר באזור X וחלקם באזור Y, נבחר לעדכן את הערך המקסימאלי.

4.5 חישוב פרמטרים עבור מקטע דרך

4.5.1 כללי

כאמור, עבור כל תאונה (בסיס הנתונים ה"נקודתי"), חושב בסיס נתונים קווי, עליו יבוצעו בהמשך ניתוחי ספירות אירועי תאונות הדרכים. בסיס הנתונים הקווי המקורי הינו שכבת הכבישים. שכבה זו חולקה לתתי מקטעים (סגמנטים) באורכים שונים. לאחר חלוקת רשת הכבישים למקטעים (שווים בגודלם) שוייכו לכל סגמנט התאונות הקשורות אליו (לרבות הנתונים שהופיעו בבסיס הנתונים עבור כל תאונה). נבחנו התפלגויות, חושבו ונבחנו אלמנטים מרחביים וסטטיסטיים עבור כל קטע כביש. הפרמטרים שחושבו עבור המקטעים הקווים יפורטו בסעיפים הבאים, להלן.

4.5.2 מספר תאונות

בכל סט נתונים על אורך הסגמטים השונים שבו, חושב מספר התאונות במקטע דרך (להלן: Acc_Cnt).

4.5.3 חומרת תאונה

נוסף על מספר התאונות במקטע דרך (AccCnt), חושב ערך חומרת תאונות ממושקל עבור אותו סגמנט (להלן: SumHumraMemushkelet). המשקול בוצע באופן הבא:
קטלנית (קוד 1) מושקל לערך 10; קשה (קוד 2) - מושקל לערך 5 ; קלה (קוד 3) – מושקל לערך 1

מהלך העבודה:

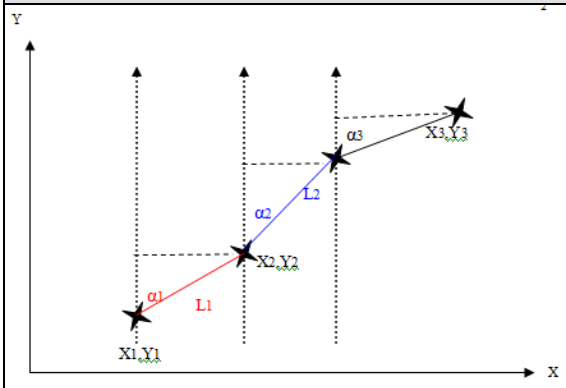
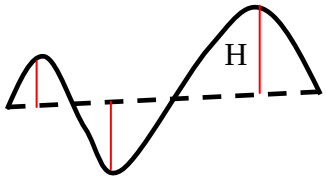
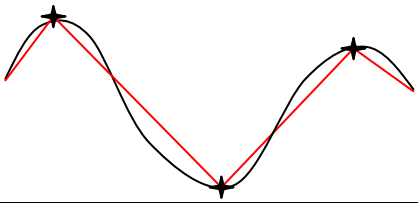
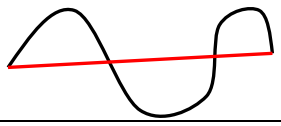
- בשכבת התאונות נוספה עמודת חומרה משוקללת (HumraMeshu), ובה עודכנה חומרת התאונה הממושקלת עבור כל תאונה בבסיס הנתונים (בהתאם למישקול שצויין לעיל)

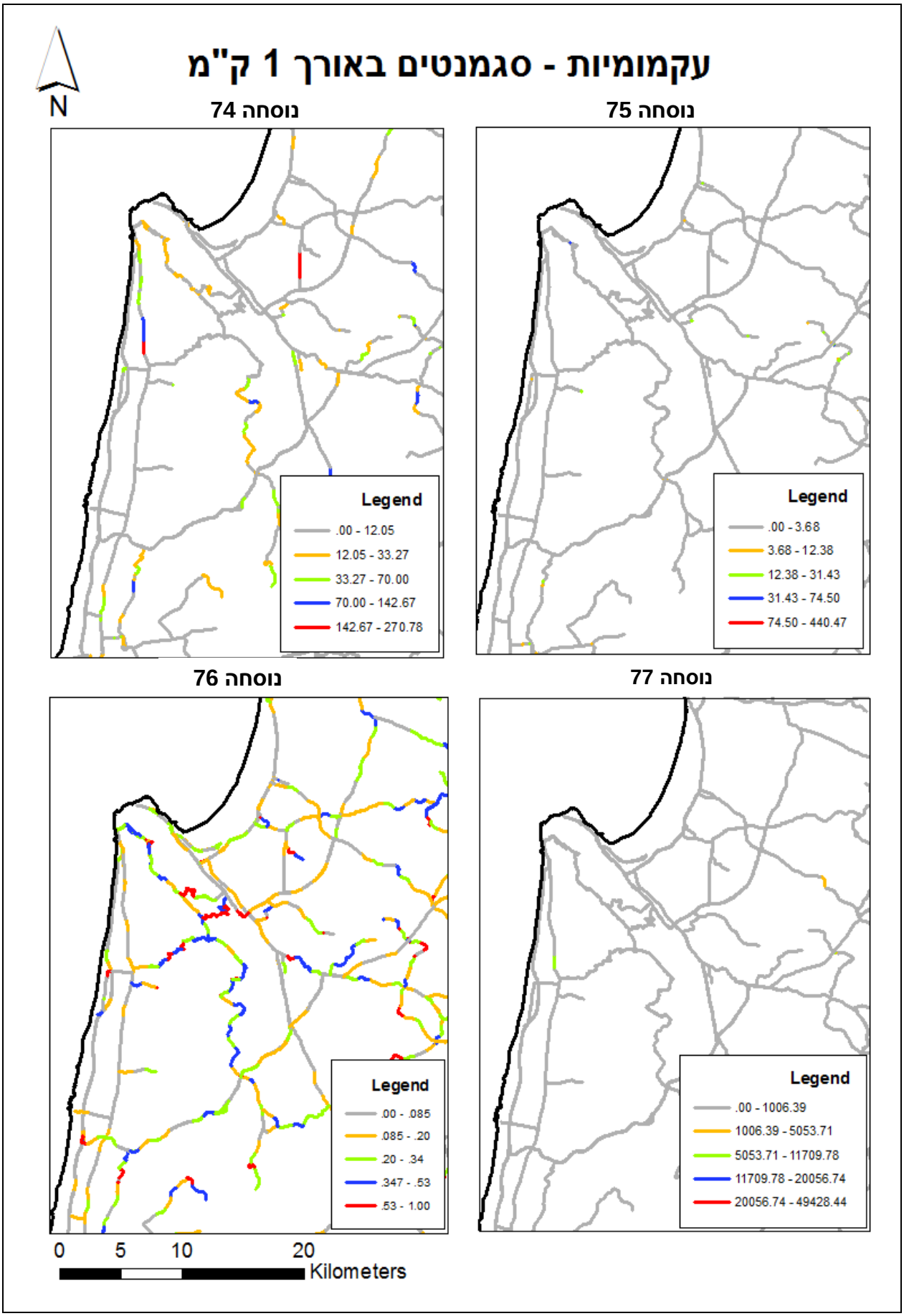
  התוצר: *All_DB_ITM_02_06_2011_NoJun_HumratTeona* יובא לתוך MDB ייחודי.
- לאחר מכן נבחרה שכבת הנקודות (*All_DB_ITM_02_06_2011_NoJun_HumratTeona*), ובוצע SPATIAL JOIN עם שכבת Split500_Final_Reduced270111 (השכבה הקווית שכבר אינה מכילה סגמנטים באזורים שהוגדרו אורביניים ובצמתים). בשלב זה נמצא כי התוצר למעשה שגוי (Acc02062011_SpatialJoinWith_Seg500FinalRed270111), היות וגם עבור תאונות, שאינן נחתכות בפועל עם כבישים, הוצבו ערכים.
- דרך חלופית היתה בחירה ייעודית של התאונות הנחתכות עם שכבת הכבישים (selection פשוט), וייצוא אותן תאונות לשכבה חדשה (Acc02062011_SpatialJoinWith_Seg500FinalRed270111_b). זו גם השכבה עליה נמשכה העבודה בתוך בסיס המידע.
- שכבת התוצר הנקודתית (כפי שתוארה בסעיף הקודם) והשכבה הקווית Split500_Final_Reduced270111 יובאו לתוך ה- MDB . שתי הטבלאות חוברו דרך שדה מקשר SplitID, כאשר עבור כל השדות נבחר בפונקציית Group By, ועבור עמודת HumraMeshu נבחר בפונקציית SUM. כך, בתום הרצת השאילתה, נוצרה עמודה חדשה SumHumraMemushkelet, המשקללת את חומרת התאונות עבור מקטע דרך. זו נשמרה לטבלה חדשה Final_Reduced_NewTbl_500.
- בטבלה הסופית, במקטעי הדרך, בהם לא התרחשו תאונות , עודכן בעמודת SumHumraMemushkelet הערך 0 בטבלה הסופית Final_Reduced_NewTbl_500, שמשמעו אין תאונות על המקטע בתקופת המחקר המדוברת.
- בשלב האחרון בוצע החיבור לשכבות הג"ג מחדש. לטבלת Split500_Final_Reduced270111 צורפה טבלת Final_Reduced_NewTbl_500 לפי שדה מקשר SPLITID. התוצר נשמר כשכבה חדשה: Split500_Final_Reduced03062011 ונוספה לו עמודה חדשה (AccSumHelkeyCnt) מסוג FLOAT. עמודה זו מכילה את תוצר סכום התאונות הממושקל (SumHumraMemushkelet) חלקי מספר התאונות במקטע (Acc_Cnt). עבור מקטעי דרך שלא התרחשו עליהם תאונות, עודכן ערך של 0 בעמודה זו (לא ניתן לחלק ב 0).
- בשלב האחרון נמחקו עמודות כפולות, שנוצרו כתוצאה מהאיחודים השונים.

4.5.4 עקמומיות סגמנט

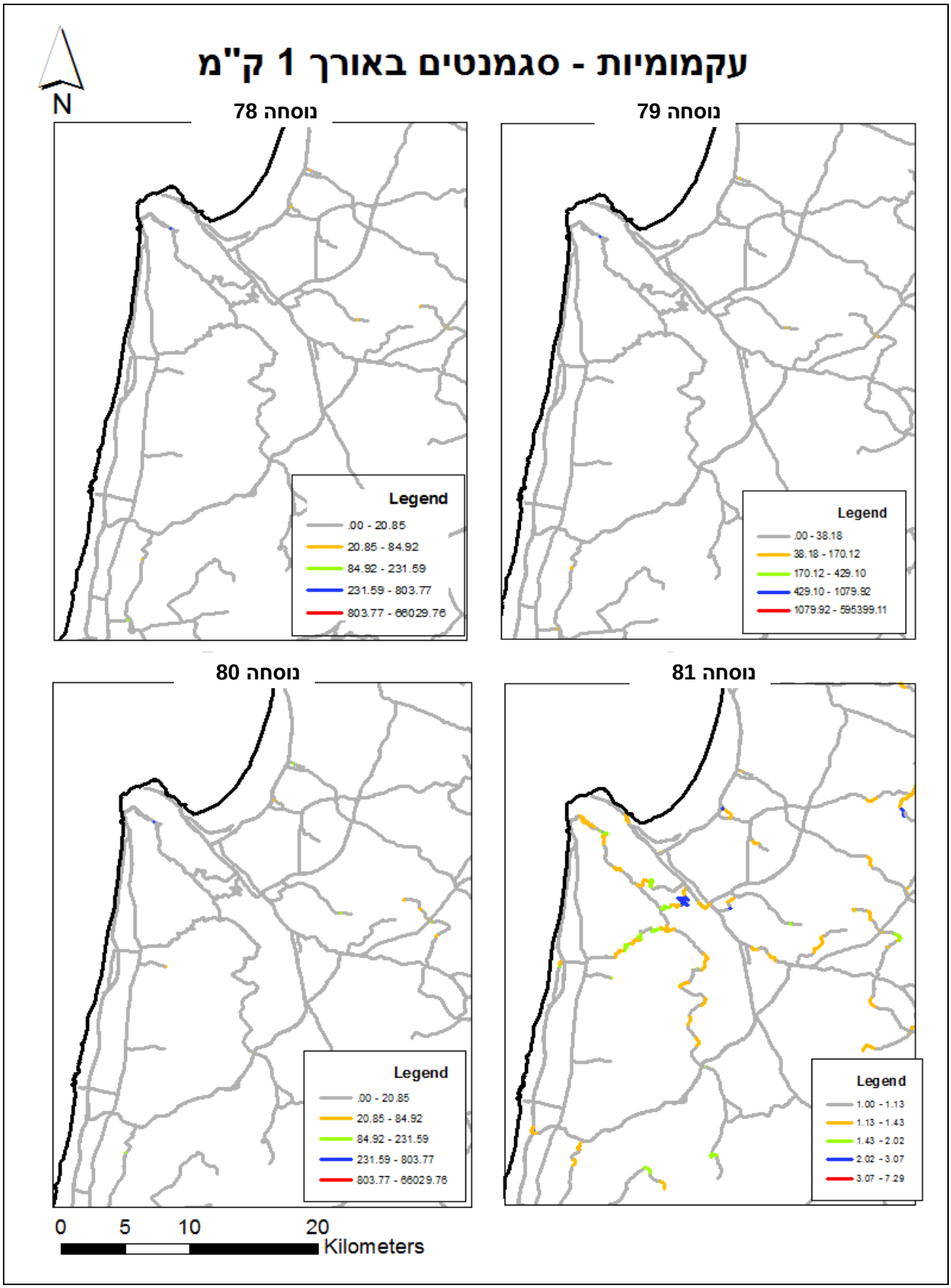
עבור כל סגמנט בבסיס הנתונים חושבה העקמומיות בפרמטרים שונים. טבלה מספר 17 מציגה את הנוסחאות השונות שנבחנו. נוסחאות 76 ו-81 (ראה איור מספר 39 ו- 40), הן אלה שהראו את הרגישות הגבוהה ביותר והנכונה עבור העקמומיות, והם אלה ששולבו במודל הסטטיסטי בשלבים הבאים.

טבלה15 - נוסחאות עקמומיות שנבחנו עבור סגמנטים באורכים שונים

מספר נוסחה	נוסחה	הסבר	איור מסביר
74	$Prm1 = \frac{\sum(\lvert Az_{i+1} - Az_i \rvert * l_i)}{\sum l_i}$	**ממוצע שינוי הכיוון (הפרש האזימוטים**), משוקלל באורך הקטע l_i (בין שני vertices).** ****הפרש האזימוטים – הפרש זויות האלפא הניתנות לחישוב באופן הבא:** ➔ $\alpha_i = Arc\tan(\frac{\Delta X}{\Delta Y})$ ➔ $\alpha_1 = Arc\tan(\frac{X_2 - X_1}{Y_2 - Y_1})$	 איור מספר38 - חישוב עקמומיות - הסבר
75	$Prm2 = \frac{\sum\lvert Az_{i+1} - Az_i \rvert}{Total\_Length}$	**חישוב ממוצע שינוי הכיוון משוקלל באורך L (הקטע כולו)**	ראה איור 38
76	$Prm3 = \frac{\sum\lvert H_{i+1} - H_i \rvert}{Total\_Length}$	**H – באדום, המרחק האנכי בין קצה הפרבולה לקו המחבר בין קווצות הקטע (המקווקו)** **סכום ההפרשים בין H אחד לשכנו, חלקי אורך הסגמנט כולו.**	
77	$Prm4 = \frac{\sum(\lvert PA_i \rvert * PW_i)}{\sum(PW_i)}$	**חלוקת הסגמנט לפרבולות, לפי שלוש נקודות עוקבות .** **PA מוקד הפרבולהA (כוכב)** **PW סכום שני המרחקים בין קצות הפרבולה (באדום).** **סכום , משוקלל באורך המרחקים בין נקודות הפרבולה**	
78	$Prm5 = \frac{\sum(\lvert Az_{i+1} - Az_i \rvert / (l_i * l_i))}{\sum l_i}$	**חישוב ממוצע שינוי הכיוון בריבוע משוקלל באורך הקטע**	ראה איור 38
79	$Prm6 = \frac{\sum(\lvert Az_{i+1} - Az_i \rvert / (l_i * l_i))}{\sum(1/(l_i * l_i))}$	**חישוב ממוצע שינוי הכיוון משוקלל באורך**	ראה איור 38
80	$Prm7 = \frac{\sum(\lvert Az_{i+1} - Az_i \rvert / (l_i))}{\sum(1/l_i)}$	**חישוב ממוצע שינוי הכיוון משוקלל באורך**	ראה איור 38
81	$Prm8 = \frac{\text{Total Curved Length}}{\text{Distance between end points}}$	**אורך הסגמנט (בשחור) חלקי אורך הישר העובר בין שני קצוותיו (באדום)**	



איור מספר 39 - עקמומיות מקטע דרך, נוסחאות 74-77, אורך סגמנט 1 ק"מ

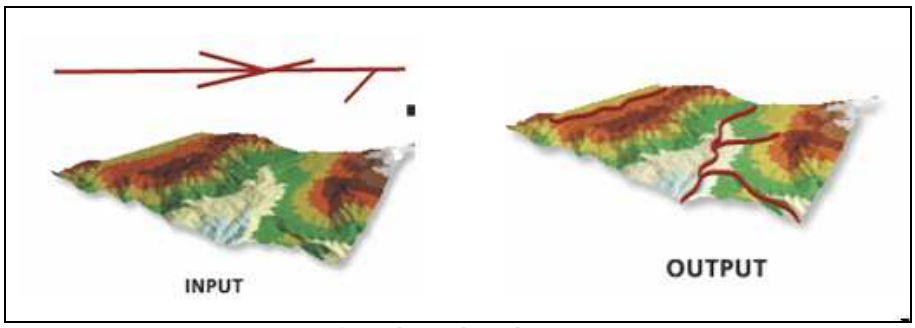


איור מספר 40 - עקמומיות מקטע דרך, נוסחאות 78-81, אורך סגמנט 1 ק"מ

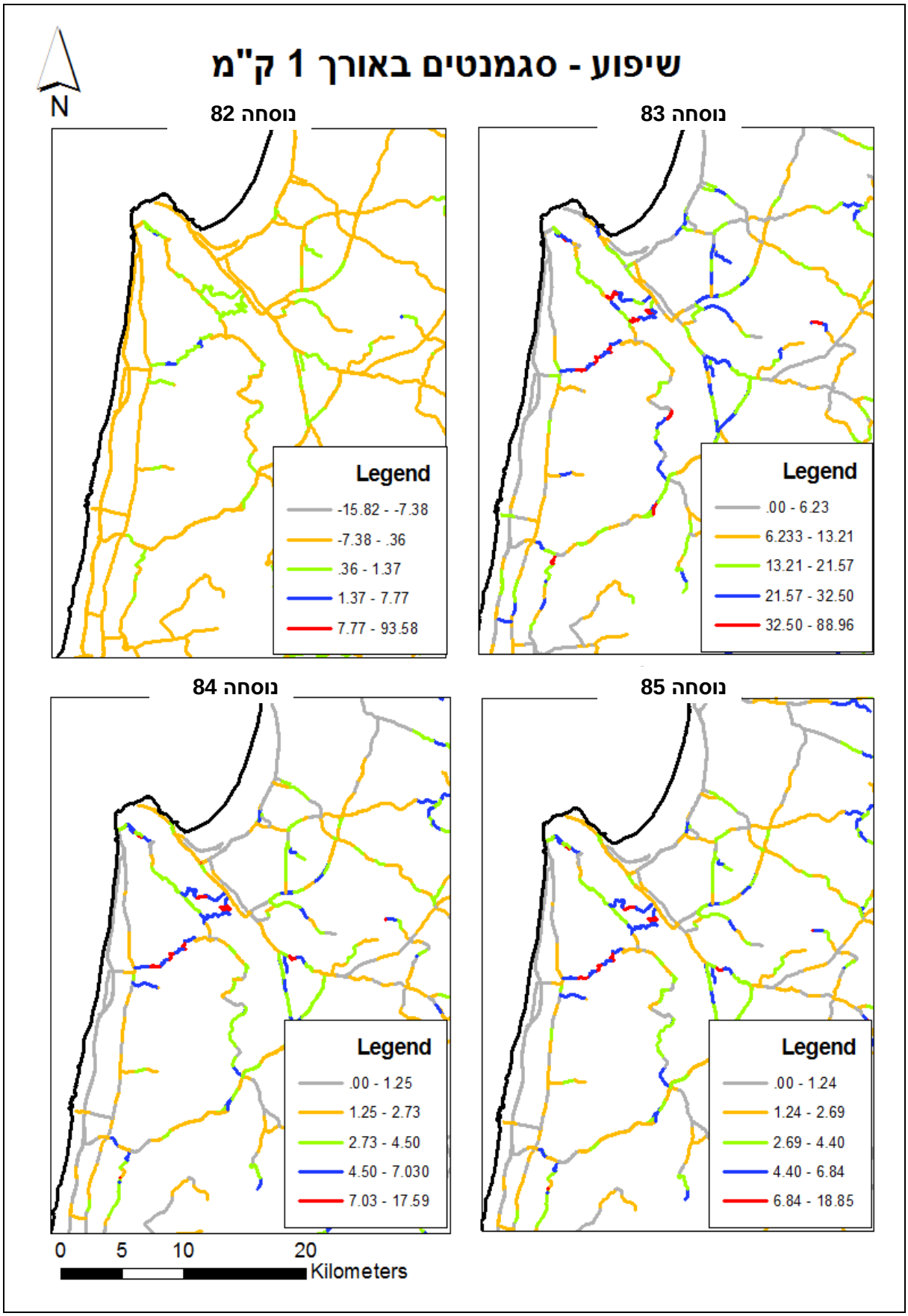
4.5.5 שיפוע סגמנט

היכולת לחשב שיפוע עבור מקטע כביש תלוייה ברזולוצית ה- DTM. ב- DTM שברשותנו מאופיין כל פיקסל בגריד בגודל של 33 מ' על 33 מ'. נבחנו השיפועים של מקטעי הדרך באורכים השונים. נוסחאות 85 ו- 87 הניבו את התוצאות הטובות ביותר, ולפיכך שולבו במודל הסטטיסטי. איורים מספר 42 ו- 43 מציגים את תוצרי הנוסחאות השונות, וטבלה מספר 18 מציגה את אופן חישוב השיפוע.

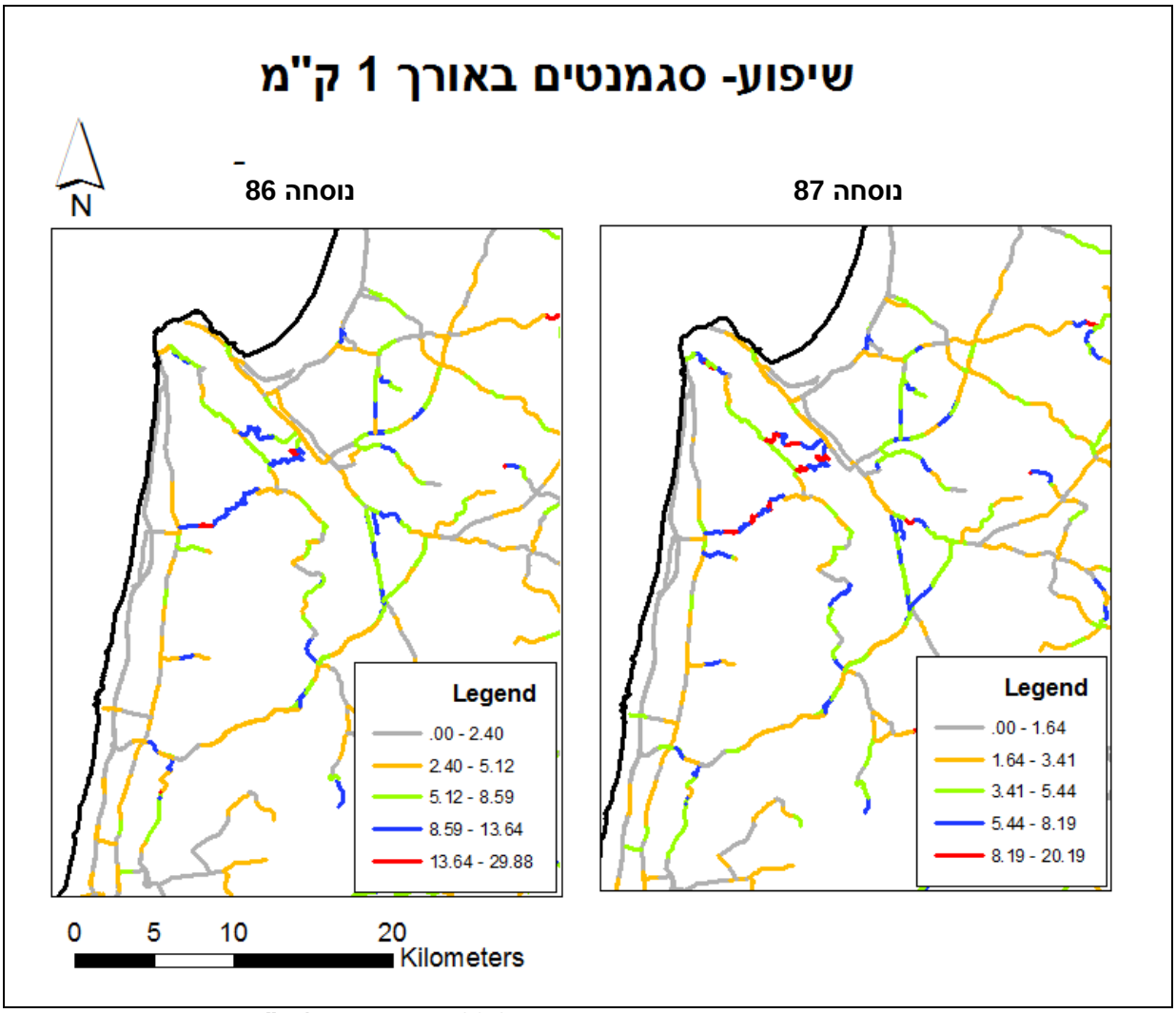
טבלה 16 - נוסחאות חישוב שיפוע מקטע דרך

מספר נוסחה	שם מאפיין	תיאור
82	sLenCalc	כלי ה surface length מחשב אורך תלת מימדי עבור אלמנטים קוויים המונחים על פני השטח. המרחק התלת מימדי גדול או שווה למרחק הדו מימדי של הסגמנט המחושב. הכלי מביא בחשבון שינויים בערכי ה Z (הגובה) כאשר הקו עובר על פני שטח גיבעיים, כמוצג באיור מספר 41:  **איור מספר 41 - כלי ה surface length בתוכנת ARCMAP** לאחר מכן שולב תוצר ה - surface length במודל תוך שילוב בנוסחה מספר Y: כאשר: $sLength_i$ - אורך סגמנט מחושב לאחר הרצת כלי ה surface length. l_i - אורך הסגמנט המקורי
83	dMinMaxSlp	כל סגמנט (קטע כביש) מורכב ממספר מקטעים (בהתאם ל vertices) . פרמטר זה מייצג את ההפרש בין השיפוע המינימאלי למקסימאלי במקטע הדרך, כמוצג בנוסחה מספר YY, כאשר MinSlp הינו השיפוע המינימאלי של הסגמנטים המרכיבים את קטע הכביש ו MaxSlp הינו השיפוע המקסימאלי. |MaxSlp-MinSlp|
84	AvrAbsSlp	ממוצע שיפועי הסגמנטים המרכיבים את קטע הכביש (בערכים מוחלטים), כפי שמוצג בנוסחה מספר BB: $\frac{\lvert S_1\rvert + \lvert S_2\rvert + .. + \lvert S_i\rvert}{n}$ כאשר:

		Si - שיפוע במקטע דרך n - מספר המקטעים המרכיבים את הסגמנט המקורי
85	WAvrAbsSlp	ממוצע ממושקל (באורך הסגמנט) של השיפועים המוחלטים של הסגמנטים המרכיבים את קטע הכביש, כמוצג בנוסחה מספר ה CC: $\frac{\lvert S_1 \rvert * L_1 + \lvert S_2 \rvert * L_2 + .. + \lvert S_i \rvert * L_i}{\sum L}$ כאשר: Si - שיפוע במקטע דרך Li – אורך הסגמנט (בין שני Vertices)
86	RngSlp	ההפרש בין הממוצע הממושקל (באורך הסגמנט) של הסגמנטים בעלי השיפוע החיובי (בדומה לנוסחה מספר CC) לבין הממוצע הממושקל (באורך הסגמנט) של כל הסגמנטים בעלי השיפוע השלילי (בדומה לנוסחה מספר CC). במידה ולא נמצאו שיפועים חיוביים אזי נחשב הערך הממוצע שלהם כאפס. בדומה לכך בשיפועים השליליים.
87	VarSlp	שונות השיפועים, מחושבת לפי שורש היחס בין הסכום הממושקל של ריבועי השיפועים לבין אורך קטע הכביש הכולל. המשקלים נקבעו לפי אורך כל סגמנט, כמוצג בנוסחה מספר VV: $\sqrt{\frac{S_1 * S_1 * L_1 + S_2 * S_2 * L_2 + .. + S_i * S_i * L_i}{\sum L}}$



איור מספר 42 - שיפוע מקטע דרך, נוסחאות 82-85, אורך סגמנט 1 ק"מ



איור מספר 43 - שיפוע מקטע דרך, נוסחאות 86-87, אורך סגמנט 1 ק"מ

4.5.6 נפח תנועה: AADT (Annual average daily traffic)

4.5.6.1 כללי

נתוני נפח התנועה (להלן: AADT) עבור כלל הכבישים (למעט כביש 6) נלקחו מפרסום הלמ"ס: *ספירות תנועה בדרכים לא-עירוניות 2001-2007, ירושלים, אדר ב' תשס"ח, מרס 2008*. עבור כביש 6, עודכנו נתונים לשנת 2009, כפי שהופיעו בפרסום הלמ"ס "נפח תנועה לפי דרך, קטע ושעת היממה (ממוצע לשעה לכל ימי השבוע), 2009-2004". זאת, היות ובקובץ פרסום הלמ"ס של ספירות תנועה בדרכים עירוניות 2001-2007, לא הופיעו נתונים עבור כביש זה ופנייה לחברת דרך ארץ לקבלת נתוני ספירות תנועה עבור שנה זו, נענתה בשלילה.

עבור כל סגמנט בשכבת הכבישים עודכנו ערכי AADT במידה והערך היה ידוע/קיים. הוחלט להציב ערך AADT **ממוצע יומי לימי חול (א-ה).** נתון זה נבחר, היות ולטעמנו נתון ה AADT בימי עבודה, משקף בצורה טובה את המציאות בהיבט הצפיפות בכביש. תחושתינו היא כי בתאונות המתרחשות בסופי שבוע ה AADT אינו הגורם המכריע. אלה, מושפעות יותר מפרמטרים אחרים, דוגמת: גיל הנהג, ניסיון מהירות ושכרות.

4.5.6.2 כיוון התנועה

בקובץ נתוני הלמ"ס קיים מידע כיווני עבור נפח התנועה. כן, קיימת אבחנה ברורה בין כביש חד סטרי לדו סיטרי . כך לדוגמא, עבור קטע 70 בדרך 40 מצומת ראם (מסמיה) עד צומת גדרה יופיעו הנתונים באופן הבא:

- 0 - התנועה בשני הכיוונים.
- 1 - כיוון ל ...התנועה לצומת ראם (מסמיה).
- 2- בכיוון מ ...התנועה מצומת ראם (מסמיה) לכיוון צומת גדרה.

לאור האמור, הוחלט כי עבור מקטעים הנמצאים <u>במעגלי תנועה</u> יעודכן מחצית ערך ה AADT. בנוסף, כאשר קיים נתון מסוג 0 (קרי, ספירות תנועה מייצגות את התנועה בשני הכיוונים), עבור כביש שבפועל (בשכבת הכבישים של חב' מפה) מפוצל לשני מסלולים (אחד לכל כיוון), יפוצל גם נתון ה- AADT לשניים.

4.5.6.3 כמות מידע קיימת והשלמת ערכי AADT למקטע דרך.

כבר מבחינה ראשונית עלה כי קיים פער בין מספר הכבישים עליהם קיים נתון נפח התנועה. בשכבת הכבישים הגיאוגרפית של חברת מפה, מעודכנים 400 כבישים בעלי מספר מע"ץ (כבישים ראשיים), לעומת 330 כבישים המופיעים בקובץ ספירות התנועה של הלמ"ס. לאור הפער שצויין, בעת עדכון הנתונים בשלב ההתחלתי, הוגדרו חמש אפשרויות, עבור כל מקטע כביש:

1 - עודכן ערך אמיתי עבור 3 שנות המחקר.

2 - לא עודכן, קיים ערך בקובץ הלמ"ס אולם לא ידענו כיצד לעדכן.

3 - אין נתונים בקובץ הלמ"ס

4 - <u>עודכן, אולם קיים קטע כביש עבורו חסר מידע</u> - מרבית הכביש עודכן והושלם לפי ערכים ממוצעים של מקטעים לפני ואחרי המקטע עליו חסר מידע. אולם קיים קטע בכביש, עבורו נקבע, כי במידה ויושלמו נתוני כמות כלי הרכב, תרד רמת הדיוק משמעותית, ולפיכך, הוחלט שלא להשלים מידע.

5 - הושלם ערך עבור קטע כביש מסויים – הושלם ערך מחושב על פי נתוני הסביבה, או שהוצבו עבור כלל הכביש הנתונים שהיו קיימים בקטע מהכביש. השלמה זו הינה סובייקטיבית, ומושפעת מהכרות עם הכבישים.

6 - קיים נתון AADT אולם לא לשנות המחקר (2005-2007)

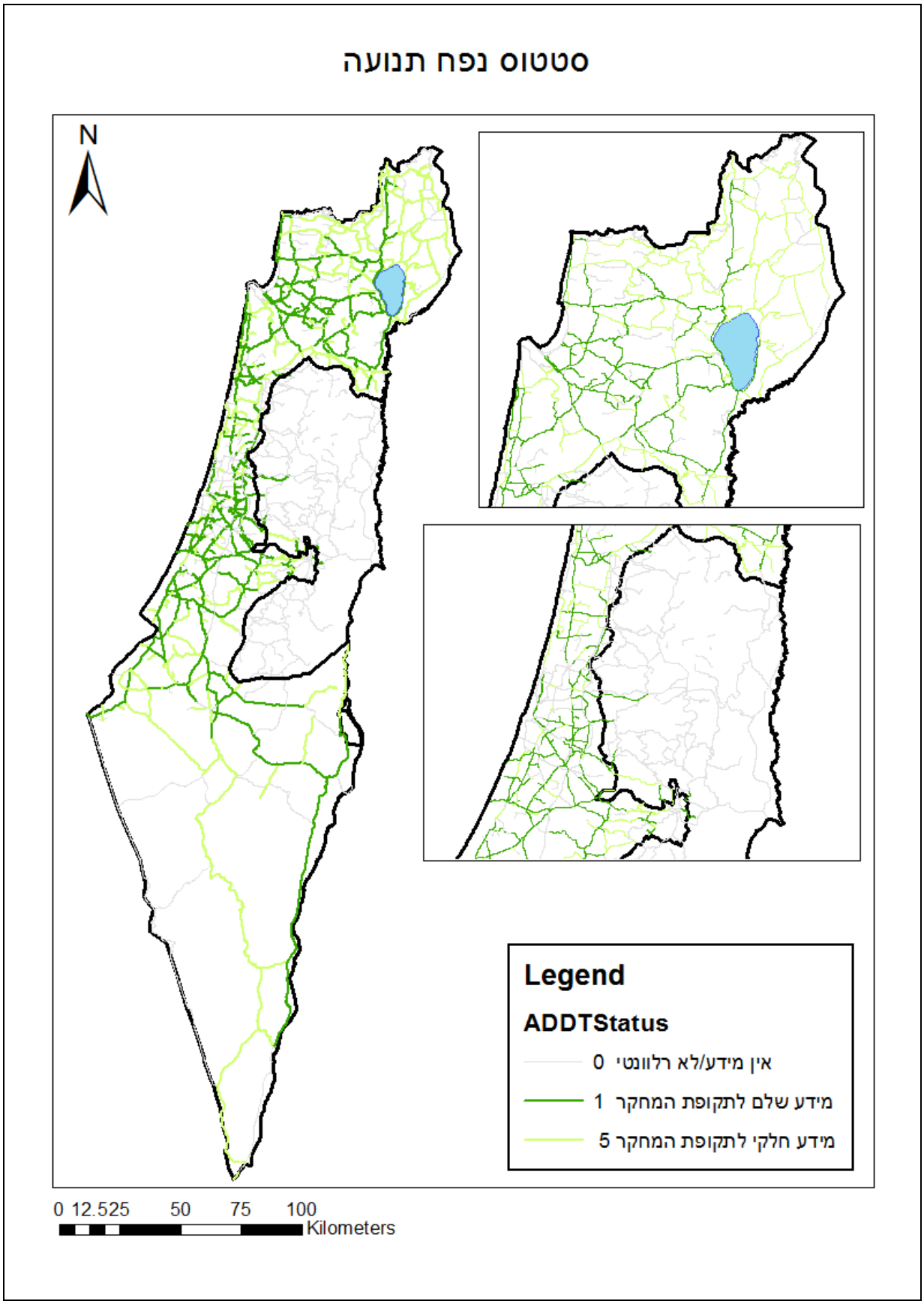
יודגש כי **בסופו של דבר**, לאחר גזירת המשמעויות הסטטיסטיות, שולבו במודל רק מקטעי דרך בעלי קוד 1, כלומר מקטעי דרך עבורם קיים היה ערך AADT מלא בשלוש שנות המחקר. איור מספר 44 (בעמוד הבא) מציג (בירוק כהה) את מקטעי הדרך ששולבו כחלק אינטגראלי במחקר (קוד שהוגדר=1) , טרם ש"נוקו" מהם מקטעי הדרך הנמצאים תחת השפעת הצמתים.

טבלה מספר 19 מציגה את כמות הכבישים שעודכנו ואת סטטוס עדכונם.

טבלה 17 - סטטוס מספר הכבישים ונתוני AADT הקיימים עבורם

מספר כבישים שנספרו	סטטוס
147	אין נתונים בקובץ הלמס
37	הושלם ערך עבור קטע כביש מסויים
2	לא עודכן
190	עודכן
24	עודכן אולם קיים קטע כביש עבורו חסר מידע
400	SUM

ניתן לראות כי למעלה מ 50% מהכבישים (190) עודכנו באופן ישיר מנתוני ספירות התנועה המפורסמות על ידי הלמ"ס, כאשר רק שני כבישים לא עודכנו כלל (מתוך חוסר בהירות בקובץ הלמ"ס כיצד לעדכנם). עם זאת, חסרים בקובץ הלמ"ס נתוני ספירות תנועה עבור 147 כבישים.



איור מספר 44 - מקטעי דרך עבורם קיים מידע נפח תנועה (AADT) בחלוקה לקבוצות

4.5.6.4 חישוב ערך AADT עבור מקטע כביש אגרגטיבי

המחקר שם לו למטרה להגדיר מקטעי דרך באורכים שונים, ולבחון איזה מהם הוא המומלץ ביותר או איזה אורך סגמנט יניב את התוצאות המשמעותיות ביותר בהיבט טיב המודל הסטטיסטי. כך, חולקה רשת הכבישים לסגמנטים באורכים שונים, החל ממקטעים באורך של 500 מ', עבור דרך 750מ', קילומטר וקילומטר וחצי.

לאור האמור, לאחר עדכון ערך AADT במקטע המקור בשכבת הכבישים הגיאוגרפית של חברת מפה[18], השלב הבא הינו חישוב והצבת ערך AADT עבור מקטע הדרך לאחר שעבר אגרגציה לסגמנט באורך מסויים. איור מספר 45 מדגים את "ההרכבה" או את החישוב של ערך AADT למקטע דרך חדש באורך מסויים (L).

יש לציין כי עבור כל אחד ממקטעי הדרך באורך L (שהינו כאמור, אורך המקטע לאחר שעבר אגרגציה של מספר סגמנטים "מקוריים" l_i) הוצב ערך AADT ממוצע ומשוקלל (באורך קטע המקור) , לפי נוסחה מספר 74:

$$\overline{AADT} = \frac{\sum_{i=1}^{N} AADT_{l_i} * l_i}{L} \qquad [74]$$

כאשר:

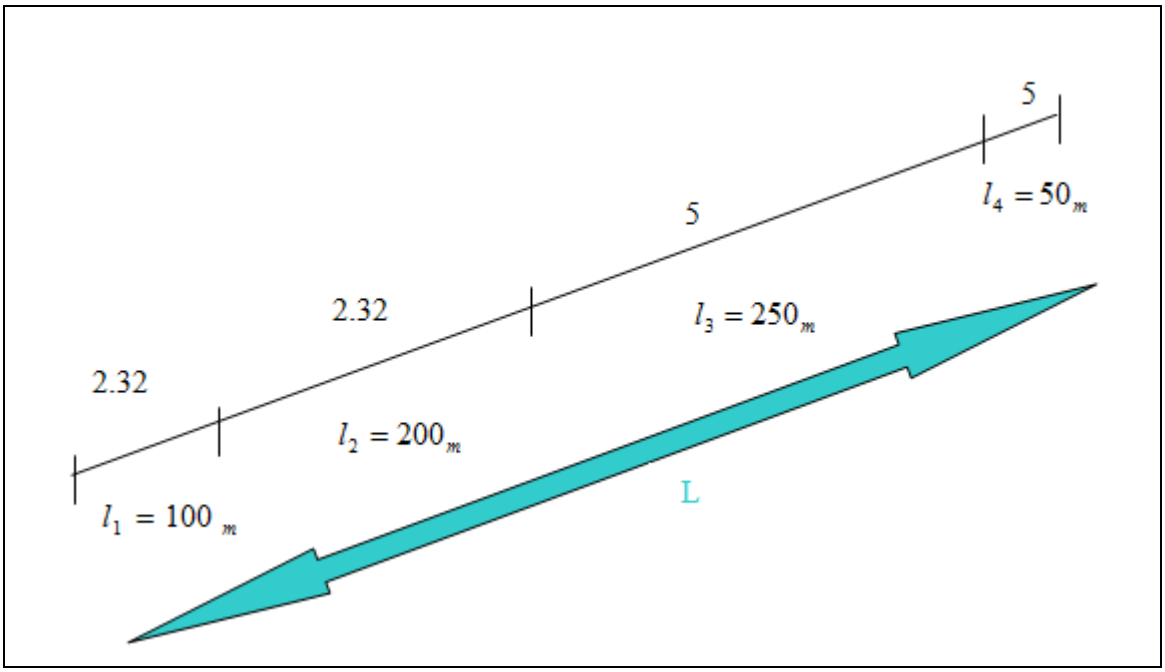
L - אורך מקטע ש"נחתך" מכביש המקור (בהתאם להגדרה ראשונית). בנוי ממספר סגמנטים קטנים יותר.

l_i - אורך סגמנט, קטע כביש מקורי (בשכבת הכבישים הבינעירוניים של חברת "מפה")

N – מספר המקטעים המרכיב את הסגמנט

AADT – נפח תנועה בסגמנט l

[18] אורך המקטעים המקוריים בשכבת מפה הינו תלוי דיגיטציה.



איור מספר 45 - השמת ערך AADT עבור מקטע דרך באורך L

מספר הערות לחישוב:

- במידה ובמקטעי ה l_i לא היה כלל מידע בנוגע ל AADT , הוצב גם עבור L, הערך NULL בשדה ה - AADT.
- במידה וחסר מידע AADT במקטע **שאורכו עולה על מחצית** האורך הכולל של הסגמנט L , אזי הוצב ערך NULL עבור כלל המקטע. כך לדוגמא, אם אורך מקטע L הינו 1000 מ', וחסרים היו נתונים על 500 מ' מתוכו, כלל הסגמנט הוצב ערך NULL.
- במידה וחסר מידע AADT במקטע **שאורכו קטן מאורך** מחצית הסגמנט L, הוצב עבורו הערך שחושב מהסגמנטים הסובבים אותו. כך לדוגמא, עבור אורך מקטע L השווה ל 1000 מ', במידה וחסר מידע עבור 100 מ' מתוכו, הוצב עבור L ערך ה AADT הממוצע, לפי נוסחה 74, **ללא התחשבות במקטע החסר.** כלומר, במידה ו L הורכב מארבעה סגמנטים "מקוריים" l_i, כאשר באחד אין מידע, הוצב ערך AADT של L, כאילו היה מורכב רק מ 3 סגמנטים ולא מארבעה.

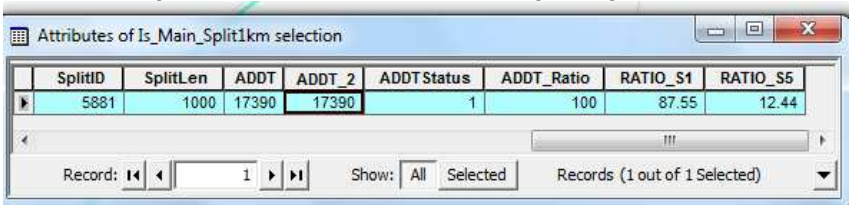
4.5.6.5 בקרת איכות - חישוב AADT ממוצע למקטעי כביש חתוכים

התקיימה בדיקה מדגמית של חישוב AADT ממוצע עבור מקטע דרך L. הבדיקה בוצעה על שכבת הכבישים שחולקה למקטעים באורך 1000 מ'. בדיקה זו בוצעה רק על בסיס נתונים אחד. זהו מדגם מייצג ומספק לביצוע פעם אחת, היות והקוד שפותח לביצוע חלוקת בסיס המידע המקורי לאורכים שונים, הינו אותו הקוד, בשינוי פרמטר האורך L .

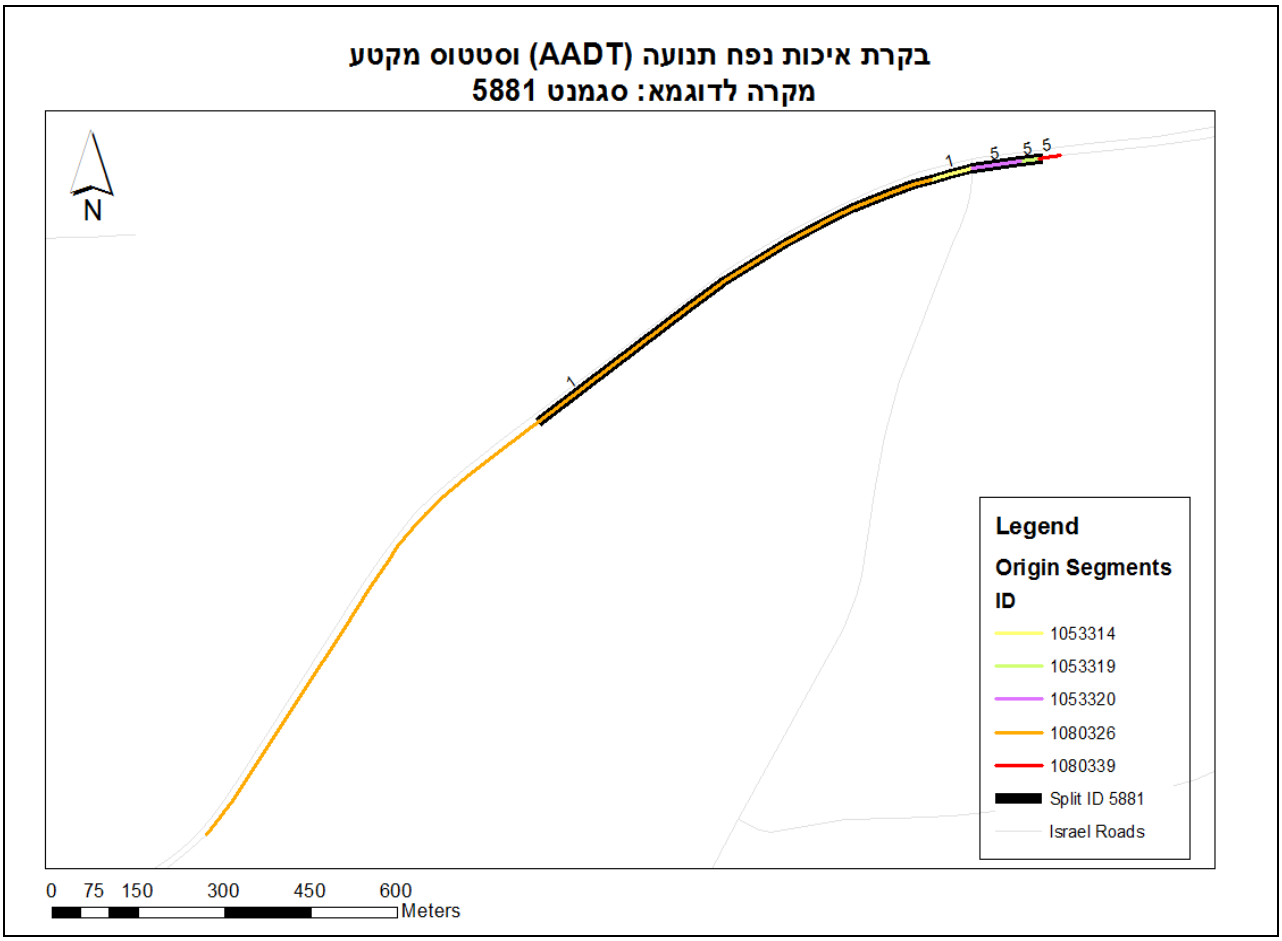
במהלך האימות, מהשכבה ה"מחולקת" Is_Main_Split1km=, שהינה שכבת תוצר החלוקה של רשת הכבישים למקטעים באורך 1000מ', נבחר סגמנט L. סגמנט זה נבחן אל מול מקטעי הדרך (l_i) המרכיבים אותו בשכבת המקור של חברת מפה (שכבת Is_Main_AADT).

לדוגמא, יוצג סגמנט SPLITID=5881 משכבת Is_Main_Split1km (מסומן בשחור באיור 46), עבורו הוצב ערך AADT השווה ל 17390 (ראה עמודת AADT_2 בטבלה מספר 20, המייצגת, את תוצאות חישוב האלגוריתם, לפי הקוד שפותח). נבחן האם הערך המחושב הנ"ל, חושב נכון.

טבלה 18 - ערך AADT מחושב עבור מקטע דרך 5881 בשכבת הכבישים המחולקת לסגמנטים באורך 1000 מ'



לשם כך, חולצו בשכבת המקור (Is_Main_AADT) המקטעים המרכיבים את סגמנט 5881 (המקטעים הצבעוניים באיור מספר 46)



איור מספר 46 - בקרת איכות נתון נפח תנועה. מקרה לדוגמא, סגמנט 5881

ניתן לראות בברור כי הסגמנט ה"כתום" (1080326) וה"אדום" (1080339) מוכלים חלקית בסגמנט 5881. מכאן, כי בעת שיקלולם בחישוב, יש לעשות שימוש באורכם היחסי (נתון זה מופיע בעמודת "**אורך מקטע המקור המוכל בסגמנט 5881**" בטבלה מספר 21).

טבלה 21, מציגה את תוצאות הבחינה הידנית אל מול הנוסחה מספר 74 המוצגת בסעיף 4.5.6.4 , כאשר העמודות המסומנות בצבע צהוב בהיר הן העמודות אליהן התייחסנו בעת מימוש הנוסחה (LEN = אורך הסגמנט ; AvgAADT_3Y = נפח תנועה ממוצע בשלוש שנות המחקר). ניתן לראות כי ה- AADT המשוקלל שחושב ידנית בקובץ האקסל (באדום בטבלה 14) שווה ערך לתוצאת ה- AADT המופיעה בעמודת "AADT_2" בטבלה מספר 20 (עמודה זו מייצגת, כאמור, את תוצאות חישוב האלגוריתם, לפי הקוד שפותח) . המשמעות היא כי האלגוריתם מחשב נכון את נתון AADT שאמור להיות מוצב בכל מקטע דרך מחולק.

טבלה 19 - תוצאות הבחינה הידנית עבור נפח תנועה מחושב

USERID	Road	Len	AADT_STATU	AvgAADT_3Y	אורך מקטע המקור המוכל בסגמנט 5881	AvgAADT_3Y * li
1053314	65	74.93	1	17000	74.93	1273771
1053319	65	32.45	5	20133	32.45	653361
1053320	65	85.06	5	20133	85.06	1712470
1080339	65	36.39	5	20133	6.86	138115
1080326	65	1712.39	1	17000	800.70	13611900
					1000.00	
					sum: (AADT * li)	**17389616**
					AADTמשוקלל	**17390**

בדיקה נוספת שהתקיימה היא ערך סטטוס ה- AADT שנבחר עבור הסגמנט. ניתן לראות שהסגמנט קיבל ערך סופי של 1 (מידע שלם של AADT עבור תקופת המחקר), למרות שקיימים במקטעים המקוריים גם נתונים שהסטטוס עבורם הינו 5 – מידע חלקי של AADT בתקופת המחקר) .

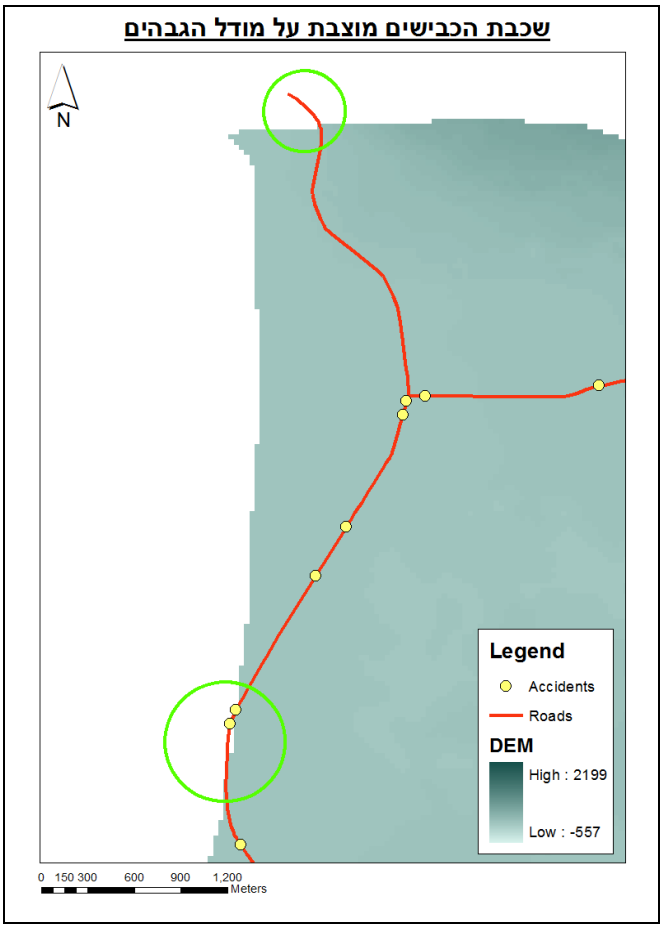
4.6 עריכה ושילוב הנתונים – השלב הסופי

לאחר שחושבו הפרמטרים עבור מקטע דרך, ופרמטרים נקודתיים עבור כל תאונה בבסיס הנתונים בוצעו מספר השלמות בהיבט עריכת הנתונים. אלה יפורטו בסעיפים הבאים.

4.6.1 התאמת השכבה הקווית לשכבת הגבהים (DEM)

במחקר נעשה שימוש במודל גבהים למטרת חישוב נתונים שונים, דוגמת שיפועים. במהלך השימוש נמצא כי מודל הגבהים אשר נרכש, אינו חופף באופן שלם את שכבת הכבישים (ראה איור מספר 47). מצב זה יוצר קושי בעת ביצוע חישובים וחיתוכים אשר עבורם נדרש מידע בנוגע לגובה.

לאור האמור, בקצוות (ראה לדוגמא המעגל העליון באיור מספר 47) "נחתכו" "קצות הכבישים" בשכבת הכבישים ההתחלתית (טרם שבוצעו עבורה החישובים השונים). במקרה המורכב יותר, בו קטע מידע ב- DEM לגבי מקטע מרכזי בכביש (ראה לדוגמא המעגל התחתון באיור מספר 47), הוחלט להשאיר את הנתונים כפי שהם. בקטעים אלה, בוצע הסינון בשלב מאוחר יותר, לאחר ששכבת הכבישים חולקה לסגמנטים קצרים יותר (500, 1000, 1500, 2000) בכדי לאבד כמה שפחות מידע.



איור מספר 47 - שכבת הכבישים מוטלת על שכבת הגבהים

בנוסף, נבדק כי אין תאונות שהתרחשו במקטע. במידה ולא היה קיים מידע AADT עבור הכביש, סולקו ממערך המחקר גם קטעים אמצעיים. בכביש 10 נמצא כי התרחשו מספר תאונות, אולם היות והכביש סגור כיום לתנועה, והיות ובחלקו הגדול נופל מחוץ לתחום שכבת ה- DEM, נמחק כלל הכביש מבסיס הנתונים.

4.6.2 סינון מקטעי דרך הנמצאים תחת השפעת צומת

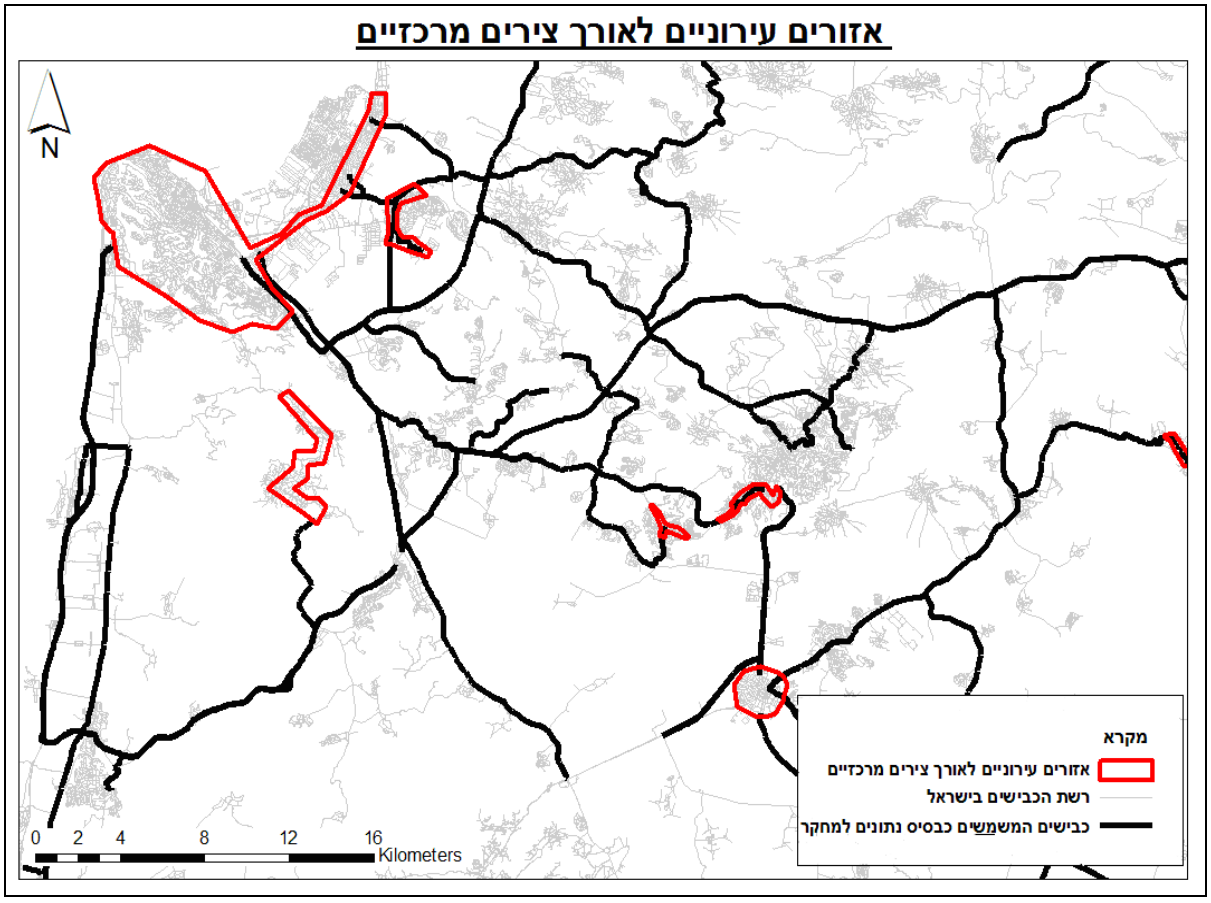
כזכור, כבר בראשית המחקר, היות וברור היה כי ההסתברות לתאונה בצומת שונה מאשר ההסתברות לתאונה במקטע דרך, התקבלה ההחלטה לפתח את המודל עבור תאונות במקטעי דרך בינעירוניים. עקב כך, מלכתחילה לא שולבו בבסיס הנתונים אירועי תאונות דרכים שהתרחשו בצומת. כפועל יוצא, על מנת שסגמנטים הנמצאים תחת השפעת צומת, לא ישולבו כקלט במודל ויטו את התוצאות (היות ומדובר בסגמנט עליו אפס תאונות, לא משום שבפועל לא התרחשו עליו תאונות, אלא משום שמראש המידע לגבי מספר התאונות בצומת לא שולב בבסיס המידע), סוננו מקטעי הדרך הנמצאים תחת השפעת צמתים.

הסינון בוצע במספר שלבים. בשלב **הראשון**, נוצרה משכבת הכבישים הקיימת, שכבת צמתים חדשה על ידי שימוש בכלי "*Feature vertices to point*" בתוכנת ARCMAP . כלי זה יוצר

צומת בכל מקום בו נחתכים שני קווים. לאחר מכן , בוצעה פעולת ניקוי אקטיבית ("ידנית") על הצמתים על פי הקריטריונים הבאים:

- צמתי פְּסֵאוּדוֹ (pseudo)– בנתוני חברת מפה קיימת קטגורית" *ROADFUNCTI*, המכילה בתוכה הסבר לגבי השימושיות של הכביש. כך ניתן היה, בקלות יחסית ל"נקות" צמתים שאינם אמיתיים. לדוגמא: חיתוך כביש : עם דרכי עפר, גשרי הולכי רגל, מעברים תת קרקעיים של הולכי רגל, גשרים, כניסות ויציאות מתחנות דלק, גשרים, דרכי 4x4, כניסה לחניות, תעלות, כניסה ויציאה מחניות.
- מחלפים - בנתוני הלמ"ס לא מופיע נתון המאפשר לדעת האם תאונה התרחשה ברמפת העליה למחלף או ברמפת הירידה ממנו. כך בהתאמה, ניתן היה, מבחינה מתודולוגית, למקם תאונה רק על הציר הראשי של הכביש, ולכן גם נקודות החיתוך שנוצרו במפגש שבין רמפות העליה והירידה (SLIPS) לכביש הראשי, הוסרו מבסיס הנתונים.
- מפגשי כביש שירות עם דרכים ראשיות – לרוב כבישי השירות הינם מקביליים לדרכים ראשיות. בדרך כלל מדובר באזורים עירוניים, שמראש לא שולבו בהם נתונים בנוגע לתאונות דרכים. כבישים אלה הוסרו לרוב בחיתוך עם גבולות האזורים העירוניים.

בשלב השני סוננו האזורים העירוניים. איור מספר 48 מציג באדום, כדוגמא, את האזורים שסוננו בגין היותם תחת השפעה של סביבה עירונית.



איור מספר 48 - מתחמים עירוניים לאורך כבישי המחקר – דוגמא, אזור חיפה ועמקים

טבלה מספר 22 מציגה את סך מקטעי הדרך ששימשו במחקר, כפי שנותרו לאחר הסינונים השונים (קוד 1 , מידע AADT מלא לכלל תקופת המחקר).

טבלה 20 - מקטעי הדרך המהווים את בסיס הנתונים הסופי לבניית המודל[19]

אורך	AADT Status	CountOf SplitLen	מספר סגמנטים	לאחר "ניקוי" צמתים ואזורים עירוניים
1500	0	1622	6030	595
	1	1525		572
	5	1126		428
1000	0	2552	8478	1191
	1	2389		1166
	5	1780		833
750	0	3509	10954	1870
	1	3260		1817
	5	2428		1287
500	0	5419	15893	3276
	1	5004		3259
	5	3713		2323

4.6.3 החיבור הסופי בין שכבת הכבישים המחולקת לאורכים שונים לבין שכבת התאונות הנקודתית

4.6.3.1 כללי

לאחר שחולקה רשת הכבישים, למקטעים בעלי אורכים שונים, נדרש היה לקשור את נתוני התאונות לכל מקטע. היות ומרבית המודלים הסטטיסטיים שנבנו במחקר, אמורים היו לנסות ולהסביר את המשתנה התלוי שהינו "מספר תאונות במקטע דרך" חשוב היה למנות כמה תאונות התרחשו על כל מקטע בתקופת המחקר.

היכולת לעשות שימוש באפשרות ה "SPATIAL JOIN" בתוכנת ARCMAP, לאיחוד שכבת הכבישים עם שכבת התאונות הנקודתית, אין בה בכדי לספק מענה שלם. על אף שהתוכנה מייצרת עמודה נוספת בתוצר המאוחד, המכילה את מספר התאונות בסגמנט (COUNT), בבסיס הנתונים יופיעו רק סגמנטים עליהם התרחשו תאונות. לצורך המחקר, נדרשים כלל הסגמנטים בבסיס המחקר, לרבות אלה שלא התרחשו עליהם אירועי תאונות. קושי נוסף העולה כאן, הוא שלא ניתן להוסיף לשכבה ה"קווית" (הכבישים המחולקים) נתונים אחרים המצויים בבסיס הנתונים הנקודתי של תאונות הדרכים, למשל: דרך רב/חד מסלולית, רוחב הכביש, שוליים ועוד. לאור האמור, היה צורך בדרך חלופית לביצוע החיבור בין שכבת הקווים (הכבישים) לשכבת הנקודות של התאונות.

[19] לאחר סינון מקטעי הדרך העירוניים ומקטעי הדרך תחת השפעת הצמתים

4.6.3.2 מהלך העבודה לשילוב המידע

- ביצוע SPATIAL JOIN על שכבת התאונות הנקודתית עם Is_Main_SplitXXX (שכבת הכבישים הסופית המחולקת למקטעים באורך 500 מ', 1000 מ' וכן הלאה). ההרצה בוצעה תוך בחירה באפשרות בה לכל תאונה הוצמדו כלל מאפייני הקו המשוייך לה.
- לאחר האיחוד, שוייך לכל תאונה בבסיס הנתונים סגמנט קווי באורך L כלשהו. תוצר האיחוד נקרא להלן: Acc_SpatialJoin_with_SegXXX.
- ביצוע "Join Attributes from a table" על שכבת הכבישים המחולקת הסופית (Is_Main_Split1k) עם תוצר ה- join XXX_join, לפי שדה מקשר SplitID – אינו מספק. זאת, היות ולכל מקטע קווי, תשוייך רק תאונה אחת, בעוד שבמקרה זה, נדרש קשר ONE to MANY. קשר זה בוצע ב- Access . על מנת לבצע את הקשר האמור, יובאו תוצר ה- JOIN ושכבת התאונות לתוך MDB, וחוברו על ידי השדה המקשר SPLITID, כאשר טבלת התוצר הינה: Seg_with_AccCnt.
- בסיום האיחודים כפי שתוארו לעיל, נסכם מספר התאונות המשוייכות לכל מקטע קווי.
- בתום הרצת השאילתות התקבל תוצר טבלאי המכיל את כלל הפרמטרים הרלוונטיים עבור מקטעי הדרך. על מנת לשייכו חזרה למרחב הגיאוגרפי, בוצע הליך של "JOIN ATTRIBUTE FROM A TABLE" בין שכבת הכבישים המקורית (Is_Main_SplitXXX) וטבלת התוצר Seg_with_AccCnt. גם כאן, לפי שדה מקשר SPLITID.
- הסרת כפילויות ממסד הנתונים.
- הוספת עמודת ROADORDER, המייצגת את סדר הכביש. הכוונה הינה לסימול הכבישים לפי מספר הספרות של חברת מע"צ. כביש מדרגה ראשונה, דהינו בעל סיפרה אחת במע"צ, יקבל את הערך 1. כביש מסדר שני, המאופיין בשתי ספרות במע"צ, יקבל את הערך 2, וכן הלאה.

4.6.4 דיוק המודל (איכון הנתונים)

בעת בחינת חומר המקור, השמת התאונות במרחב וביצוע הניתוחים נתגלו לעיתים מספר אי-דיוקים עימם נדרש להתמודד, וחשוב לציינם. **בראש ובראשונה**, קיימות בעיות בדיוק כבר טרם ההשמה במרחב הגיאוגרפי. אלה נובעות מהאופן שבו התאונות מעודכנות בבסיס הנתונים של הלמ"ס, ובו המידע ניתן בדיוק של **מאות מטרים**. כך, מלכתחילה, הסיכוי לתזוזה במיקום התאונה הינו במינימום כ - 100 מטרים.

שנית, קיימות סטיות הנובעת מנתון הק"מ בצמתים. הק"מ הרץ בצמתים, אמור להתחיל מהערך 0, כאשר אבני הק"מ אמורות להיות מוצבות באופן שווה, לאורך כל 1 ק"מ. לא כך המצב בפועל. בין אבני הק"מ קיימים מרחקים שונים. מצב זה מוביל לכך, שקיים פער (משמעותי לעיתים) בין הק"מ המעודכן בין שני צמתים (בקבצי הלמ"ס/מע"צ) לבין המרחק ה"אמיתי" הנמדד בין הצמתים.

5. ניתוח גיאו-מרחבי - נקודתי

Kriging 5.1

5.1.1 כללי

בפרקים הקודמים הוסבר כיצד נבנה בסיס הנתונים הנקודתי, בו כל רשומה (נקודה) מייצגת אירוע תאונת דרכים. על מנת לנסות ולאכן את מרכזי הכובד של התופעה (צבירים), ניתן לעשות שימוש בכלים המובנים של תוכנת Arc GIS. בעיקר מדובר בכלים המופיעים ב- ToolBox Spatial Statistics תחת Analyzing Patterns (דוגמת: Ripleys K Function, Morans I, Getis-Ord General G) וב-Measuring Geographic Distribution (דוגמת: Mean Center).

עם זאת, כלים אלה, עושים שימוש במרחב הכללי ולא במרחב הליניאמנטי (לאורך צירי הכבישים). במידה ונעשה שימוש בפונקציות לאיתור צבירים עבור תאונות דרכים, הרי באזורים בהם קיים ריכוז גדול של כבישים הקרובים זה לזה פיזית, תחוברנה תאונות שהתרחשו בכבישים שונים, לצביר אחד גדול. על אף האמור, הושקע מאמץ במציאת דרך, בה יתאפשר לעשות שימוש ביכולות המובנות של התוכנה. כך נבחר לעשות שימוש במתודת ה- Kriging.

Kriging הינו כלי סטטיסטי מתקדם המבצע אינטרפולציה, כך שמנתוני קלט נקודתיים בעלי ערך Z כלשהו (Z-Values), יתקבל תוצר רסטרי מוערך (משטח מוערך). הכלי, מאפשר חקירה אינטראקטיבית של ההתנהגות המרחבית טרם קבלת תוצר הפלט שהינו משטח רסטרי, והינו מבוסס על תאוריית המשתנה האזורי (regionalized variable), שבבסיסה ההנחה לפיה הדוגמא המרחבית יכולה להתפרס בכל מקום במשטח הנחקר. ברור מאופן ההצגה הנ"ל כי על מנת שנוכל לעשות שימוש במתודה זו, יש לערוך את נתוני הקלט הנקודתיים (תאונות), באופן שיענה על הנחה זו. לדוגמא, נציב במודל נקודות המייצגות מיצוע כלשהו עבור תא שטח מסויים.
בשיטת ה - Kriging ההנחה היא כי המרחק (או הכיוון) בין נקודות, משקף קורלציה מרחבית, בעזרתה ניתן להסביר את שונות התופעה בפני השטח. אלגוריתם ה - Kriging מתאים מודל מתמטי עבור מספר ספציפי של נקודות, או עבור כלל הנקודות הנכללות ברדיוס מסויים. בנוסף, הוא ממשקל את הערכים הסובבים בכדי לבצע ניבוי עבור המקום (הגיאוגרפי) בו לא התקבלה דגימה. נוסחה מספר 75 מציגה את נוסחת ה Kriging הבסיסית:

$$\hat{Z}(S_0) = \sum_{i-1}^{N} \lambda_i Z(S_i) \qquad [75]$$

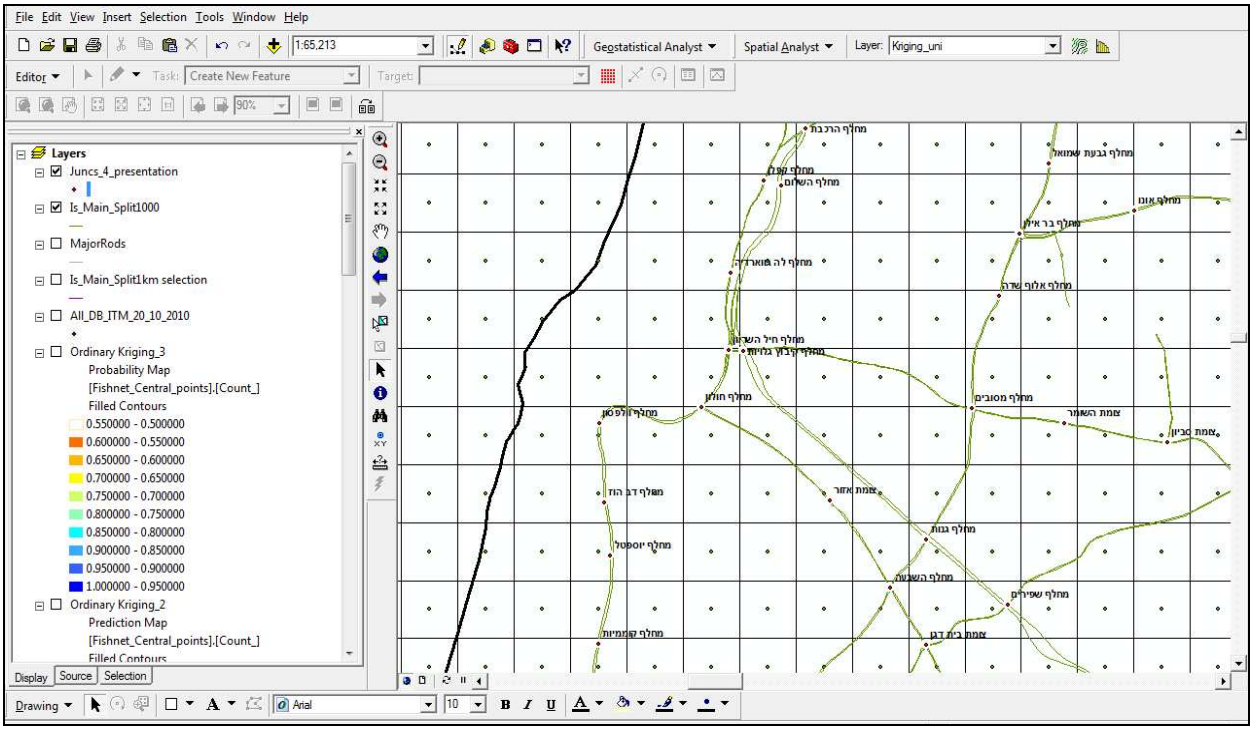
כאשר:
$Z(s_i)$ – הערך הנמדד במיקום ה i.
λ_i – המשקל (הלא ידוע) עבור הערך הנמדד במיקום ה i.
s_0 – המיקום הצפוי
N – מספר הערכים שנמדדו (מספר הנקודות בבסיס הנתונים)

בשונה מאינטרפולציית ה- Inverse Distance Weighted (IDW), שבה המשקל λ_i תלוי אך ורק במרחק (והמניחה כי אלמנטים הקרובים אחד לשני דומים יותר מאלה המרוחקים אחד מהשני), ב Kriging המשקול מבוסס לא רק על המרחק בין הנקודות הנמדדות והמיקומים הצפויים, אלא גם על הסידור המרחבי של כלל הנקודות. על מנת לעשות שימוש בתבנית המרחבית במישקול יש צורך להגדיר אוטוקורלציה מרחבית[20].

על מנת לבצע תחזית כזו, יש לבחון את תנאי התלות ורק אז לבצע את החיזוי. התוכנה יוצרת variograms (structural analysis) ופונקציית שונות משותפת בכדי להגדיר את האוטוקורלציה המרחבית. בשלב הבא הליך ה- Kriging חוזה את המיקום הלא ידוע. כך עושה ה Kriging שימוש במידע פעמים. פעם אחת על מנת להעריך את האוטוקורלציה ופעם נוספת, בכדי לבצע את החיזוי [מתוך: HELP תוכנת ArcGIS ; Cressie1990,].

5.1.2 מהלך העבודה

- יצירת fishnet[21] ריבועי, שגודל כל תא בו הינו 500X500 מטרים (350 עמודות על 900 שורות), כפי שניתן לראות באיור מספר 49. לאחר מכן נבנה (בנוסף), גריד של 1000X1000 מ' לכל תא .



איור מספר 49 - יצירת fishnet (רשת GRID)

20 המתאם בין ערך של משתנה באזור מסויים, לבין הערכים של אותו משתנה באזורים סמוכים Unwin & O'Sullivan[2003,]

GRID 21

- הרצת כלי ה - Feature To Poligon, ההופך את הרשת לפוליגונים ריבועיים.
- עבור כל פוליגון, חושבו קואורדינאטות X, Y מרכזיים (פורמט DOUBLE)
- ביצוע SPATIAL JOIN בין השכבה הפוליגונאלית של הריבועים, לבין שכבת תאונות הדרכים. כך עבור כל פוליגון מתקבל מספר התאונות שהתרחש בשטחו. ערך זה יהווה את המשקל במהלך ריצת ה- Kriging.
- חישוב קואורדינאטות X Y של מרכז הפוליגון וייצוא לפורמט אקסל.
- ייבוא קואורדינאטות X Y (מהשלב הקודם) וייצירת שכבת נקודות חדשה, המהווה את קובץ הקלט לkriging.

5.1.3 ניתוח ותוצאות

איורים מספר 50 ו- 51 מציגים את ההסתברות לתאונת דרכים במרחב, כפי שהתקבלה מהרצת ה kriging בגדלי תא שטח של 1000X1000 מ' ו- 500X500 מטרים בהתאמה.

באיור מספר 50, ניתן לראות כי ההסתברות תאונות מאפיינת גם שטחים נרחבים, בהם ברור כי אין נתיבים תחבורתיים. כזכור ההתייחסות לכלל המרחב, כפי שצויין לעיל, הינה אחת מהחולשות המוכרות של השימוש בכלי זה.

היות ותבנית הכבישים מוכרת, ניתן לראות כי ההסתברויות הגבוהות ביותר לתאונה (בכחול), ממוקמות לאורך צירי האורך המרכזיים: כביש 2 (כביש החוף) ; כביש 4 (הכביש הישן, בואכה נהריה); כביש 6, וכביש 70 בצפון.

בהיבט כבישי הרוחב (במרכז הארץ): כביש 44 וכביש 1 (תל אביב בואכה מחלף שורש).

באזור הצפון ניתן לראות הסתברויות גבוהות במיוחד וריכוז גדול ב: כביש 85 (עכו-כרמיאל) ; כביש 65 בחלקו הצפוני (צ. גולני – צ. קדרים) ; כביש 75 (בין צ. נהלל לצ. יגור) וכביש 79 (בין נצרת לאפק)

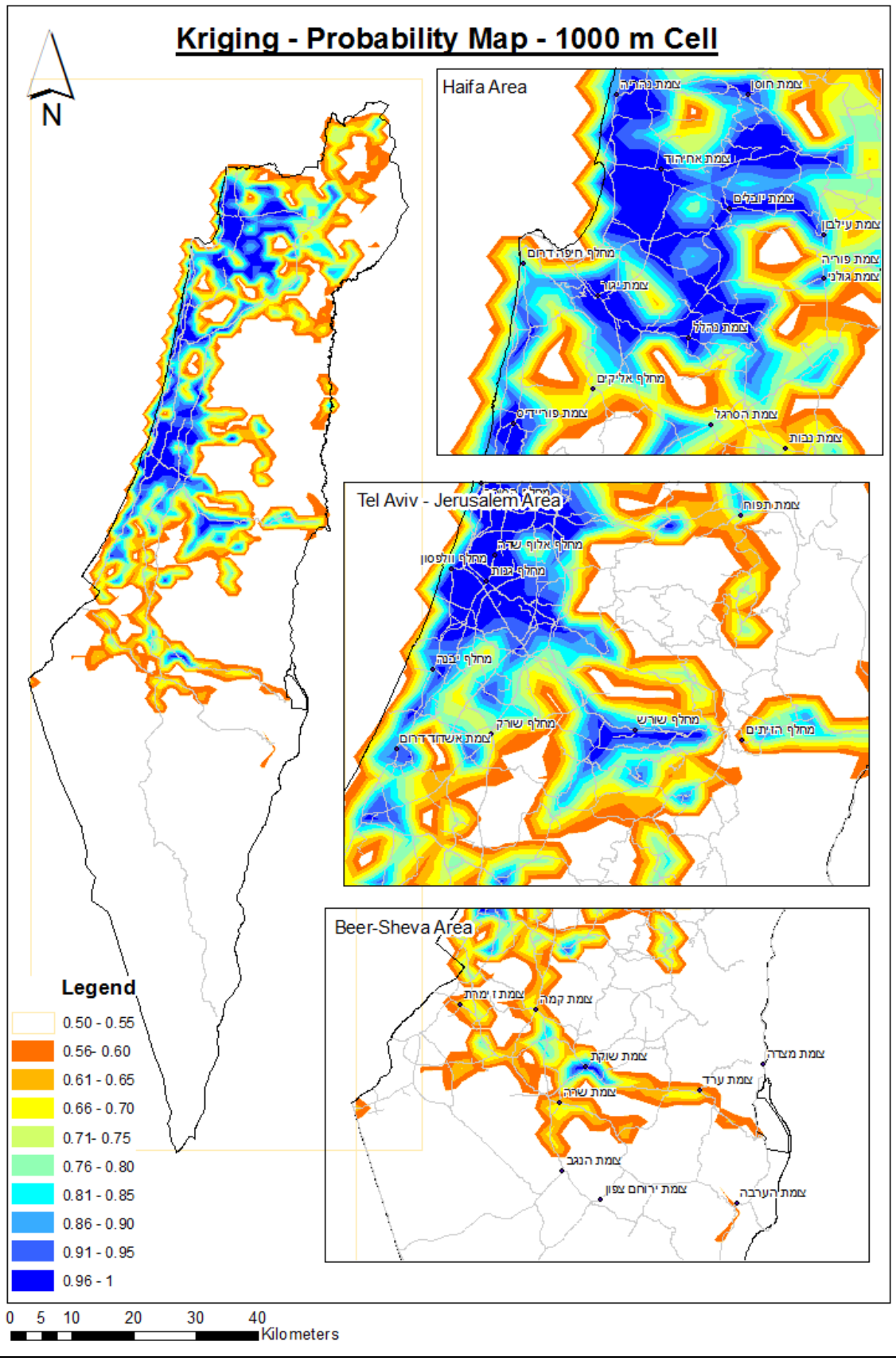
באזור הדרום, ניתן להבחין בהסתברות גבוהה, בכביש 31, באזור צומת שוקת.

באיור מספר 51, ניתן לראות את התוצאות שהתקבלו בניתוח דומה, עבור גריד מקור קטן יותר. כך, התוצר הסופי "נקי" יותר אם גם מצומצם יותר. ניתן לראות שמרכזי הכובד סמוכים יותר לכבישים (מעט פחות הסתברויות נמצאות באזורים בהם אין כבישים במרחב הגיאוגרפי), נמצאה מגמה דומה מאוד לזו של הניתוח הקודם.

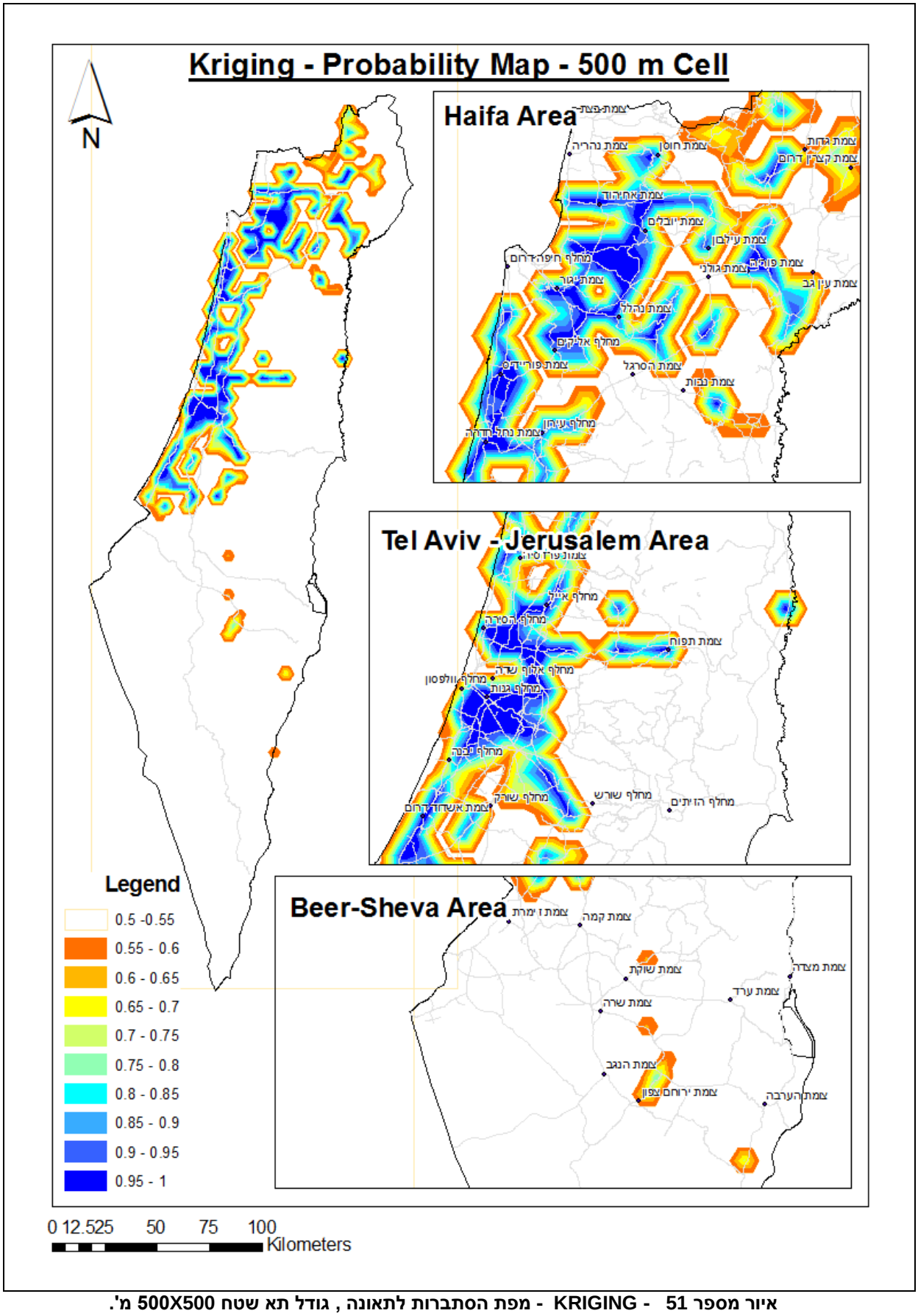
במרכז, ניתן לראות כי בשונה מהניתוח הקודם, כביש 5 (ובהמשך 505) בואכה אריאל מקבל משקל הסתברותי רב יותר. מצפון ליישוב, ארגמן, בבקעת הירדן, בכביש 90 , ניתן לראות הסתברות גבוהה נקודתית.

באזור באר שבע, הניתוח ממוקד מאוד, והסתברויות נמוכות יותר לתאונה, מאשר במרכז או בצפון.

לאור הניתוח שתואר לעיל, מומלץ להקטין ככל האפשר את הגריד, בבואנו לבצע ניתוח kriging. כך גם תימנע שגיאה של ריכוז אירועים בכבישים סמוכים לצביר אחד.



איור מספר 50 - KRIGING - מפת הסתברות לתאונה , גודל תא שטח 1000X1000 מ'.



איור מספר 51 - KRIGING - מפת הסתברות לתאונה , גודל תא שטח 500X500 מ'.

5.2 איתור צבירים

5.2.1 כללי

הפונקציות המאכנות צבירים בצורה אוטומאטית בתוכנת ArcGIS, עושות שימוש בכלל הנקודות הנמצאות במרחב בהתאם לרדיוס הניתן להגדרה על ידי המשתמש. לאחר בחינתן (דוגמה לכך הוצגה בסעיף 5.1 בתיאור הליך ה kriging), נמצא כי פונקציות אלה אינן מספקות מענה מספק לאיכון צבירים לאורך אלמנט קווי (להלן: לינאמנט). לאור האמור, ועל מנת לאתר צבירים במרחב הגיאוגרפי לאורך לינאמנטים, פותח קוד SQL ייעודי.

שוב יצויין, כי במהלך בחינת נתוני תאונות הדרכים, עלה חשד בנוגע לנכונות המידע בנושא "כיווניות" תאונות הדרכים (האם התאונה התרחשה בק"מ ה"עולה" או בק"מ ה"יורד" של הכביש, או במילים אחרות האם התאונה התרחשה בכיוון אחד של הכביש או בעברו השני). הוזכר כבר כי ניכר היה שהערך "1" מופיע הרבה יותר, מאשר הערך "2".

כזכור, על אף האמור, בעת הצבת התאונות במרחב הגיאוגרפי, הוחלט להציבן בכיוון התואם את המידע שהופיע בבסיס הנתונים של הלמ"ס (המכיל כאמור את המידע המתקבל ממשטרת ישראל). זאת, על אף שסביר היה כי נפלה טעות אנוש בקביעה זו.

בעת הבחינה של המקטעים המסוכנים, בשיטת ה- Region Growing , נאלצנו שוב להתמודד עם אותה הבעיה. ההחלטה בניתוח זה, המתמקד באיתור הצבירים, היתה כי **לא יעשה שימוש בכיווניות**, כך ניתן היה לאתר מצבורים של תאונות, לקבוע מקטעי סיכון, ולאחר מכן לנתחם.

5.2.2 מהלך העבודה

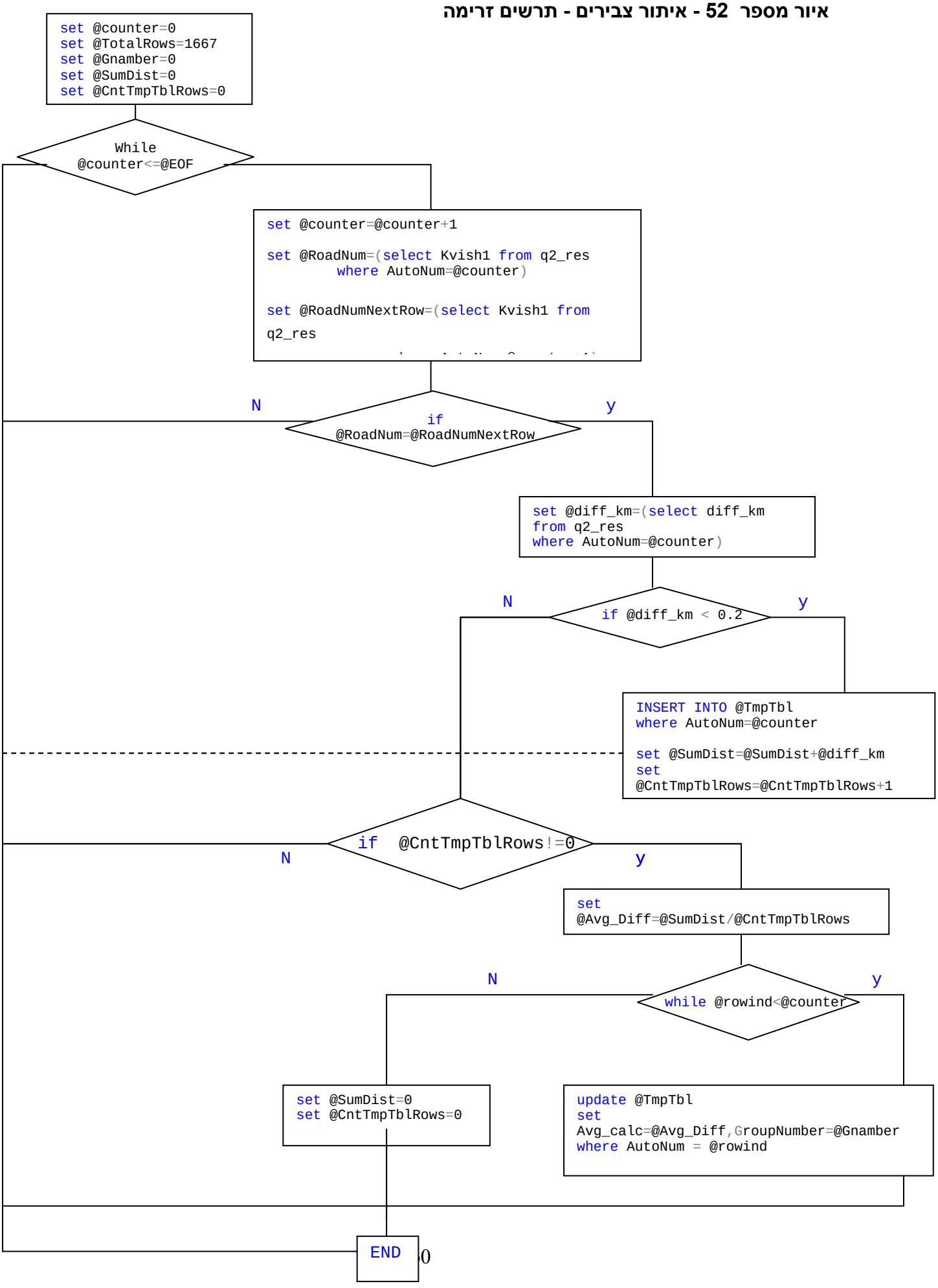
- בשלב הראשון, לאחר ניקוי בסיס הנתונים של תאונות הדרכים מתאונות שהתרחשו בצמתים (All_DB_ITM_17_12_2010_NoJun), פוצל בסיס הנתונים לקטגוריות הרצויות. לדוגמא: חזית צד ; self inflicted ; מזג אוויר - גשום ; מזג אוויר ערפילי[22].
- הורצה שגרה (Q1) בה בוצע סידור (ORDER BY) של נתוני התאונות (All_DB_ITM_17_12_2010_NoJun) במהלך שלוש שנות המחקר לפי מספר כביש וק"מ. סדר זה נשמר כטבלה חדשה (FullDB_Ordered).
- לאחר מכן נוספה לתוצר השלב הקודם (FullDB_Ordered) עמודת מספרים רצים AutoNum

[22] בוצע אימות לקבוצה אותה ניתן לקשור למזג האוויר. נבדק האם אותן תאונות בהן הכביש דווח כרטוב, תואמות לאלה שדווח עבורן מזג אוויר ערפילי או גשום. בסך הכל דווח עבור 652 אירועי תאונות מזג אוויר ערפילי וגשום, ועבור 712 דווח כביש רטוב ממים. בוצע חיבור בין שתי הטבלאות, כאשר כשדה מקשר שימש מספר התאונה (חד ערכי). חיבור זה הניב 615 רשומות תואמות. מכאן ניתן היה להניח כי קיימת התאמה בין מצב פני הכביש לבין מזג האוויר ולגזור תת קבוצה נוספת.

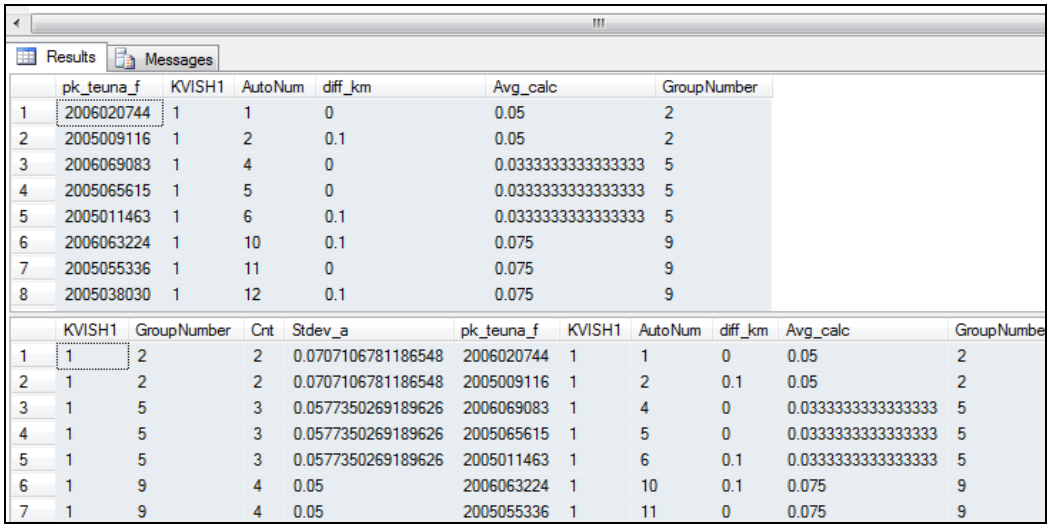
- הורצה שגרה (Q2) המחברת בין תאונה אחת לזו שלאחריה. כך ניתן לחשב מהו המרחק בין התאונות ולאתר צבירים.
- החיבור שתואר לעיל, יצר מצבים שלכאורה נדרש ל"נקותם". ניתן לזהותם על ידי המעבר בין מספרי הכביש. נבחר שלא ל"נקות" את התאונה האחרונה על מנת לא לפגוע ברצף ה- AUTO NUM. בכל מקרה, באופן גורף, המרחק בין התאונות בסיום הכביש גדול באופן משמעותי מ- 0.2 ק"מ, שהוגדר כמרחק החיפוש.

התהליך שתואר לעיל, בוצע בסביבית תוכנת ACCESS , כאשר התוצר הינו טבלת q2_res. טבלה זו יובאה לתוכנת sql server 2005 ושם הורצה השאילתה שמטרתה איתור הצבירים (השאילתה מופיעה בפרוט בנספח מספר 3), עץ המבנה מתואר באיור מספר 52:

איור מספר 52 - איתור צבירים - תרשים זרימה



לאחר שהורץ קוד ה SQL (מופיע בנספח מספר 3) , התקבל תוצר טבלאי , כמוצג באיור מספר 53.



Results | Messages

	pk_teuna_f	KVISH1	AutoNum	diff_km	Avg_calc	GroupNumber
1	2006020744	1	1	0	0.05	2
2	2005009116	1	2	0.1	0.05	2
3	2006069083	1	4	0	0.0333333333333333	5
4	2005065615	1	5	0	0.0333333333333333	5
5	2005011463	1	6	0.1	0.0333333333333333	5
6	2006063224	1	10	0.1	0.075	9
7	2005055336	1	11	0	0.075	9
8	2005038030	1	12	0.1	0.075	9

	KVISH1	GroupNumber	Cnt	Stdev_a	pk_teuna_f	KVISH1	AutoNum	diff_km	Avg_calc	GroupNumbe
1	1	2	2	0.0707106781186548	2006020744	1	1	0	0.05	2
2	1	2	2	0.0707106781186548	2005009116	1	2	0.1	0.05	2
3	1	5	3	0.0577350269189626	2006069083	1	4	0	0.0333333333333333	5
4	1	5	3	0.0577350269189626	2005065615	1	5	0	0.0333333333333333	5
5	1	5	3	0.0577350269189626	2005011463	1	6	0.1	0.0333333333333333	5
6	1	9	4	0.05	2006063224	1	10	0.1	0.075	9
7	1	9	4	0.05	2005055336	1	11	0	0.075	9

איור מספר 53 - תוצר הרצת קוד ה SQL לחילוץ צבירים

על מנת לאתר באופן חזותי את המקטעים בהם קיים ריכוז תאונות, בוצעה פעולת איחוד (JOIN) בין התאונות בחלוקה השונה לטבלת התוצר של הרצת ה SQL (שנשמרה בפורמט XLS) . כאמור, קוד ה- SQL מבצע חיבור בין תאונה אחת לזו העוקבת, ובתנאי כי מתקיים מרחק הקטן מ L כלשהו הקבוע מראש (במקרה שלנו מרחק הקטן מ- 100 ו- 200 מטרים), ולפיכך בהיבטי התצוגה על מנת להציג את מספר התאונות המעורבות באשכול יש להוסיף לעמודת ה CNT ערך 1, ולהוסיף לתצוגה את התאונה האחרונה.

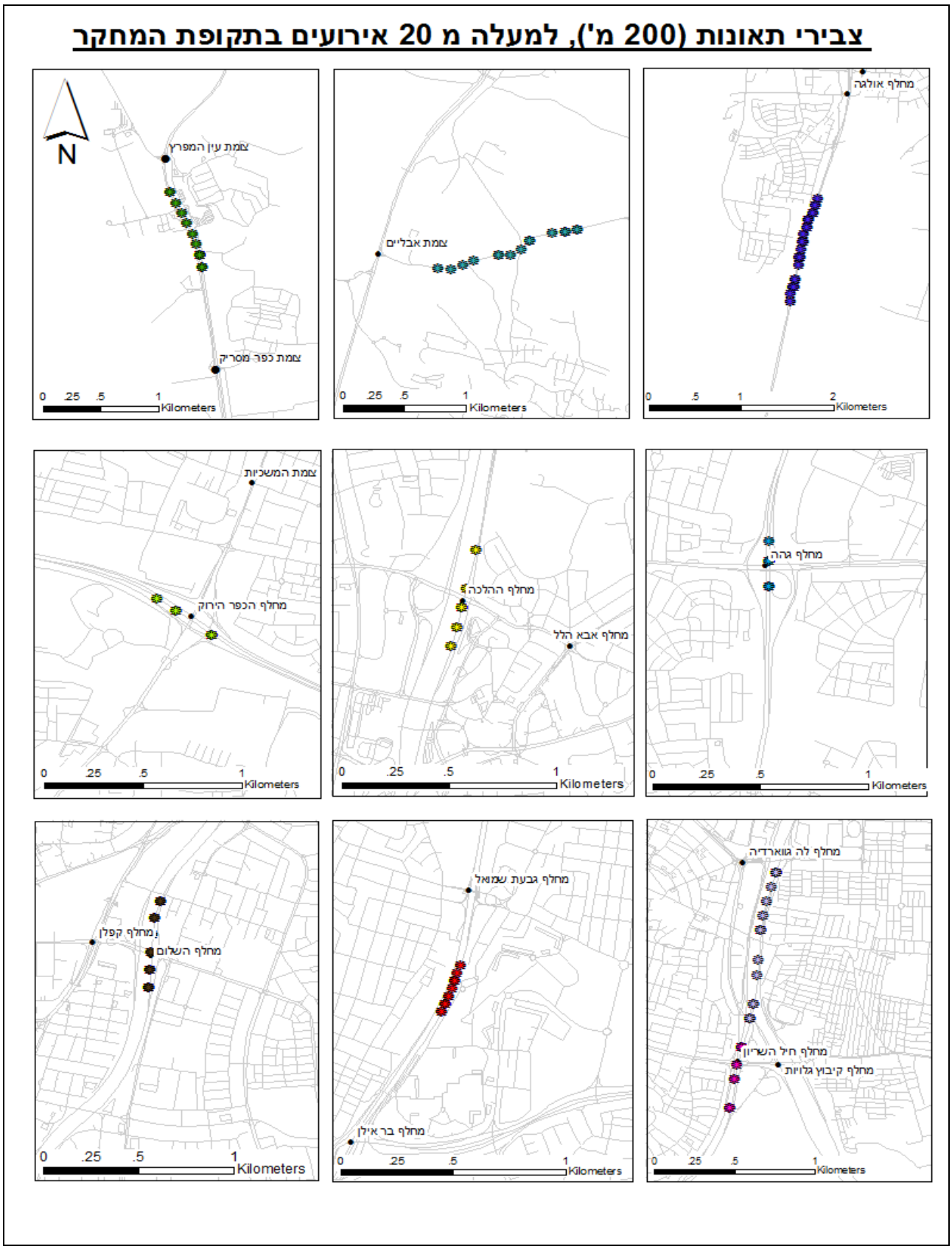
נבצע זאת באופן הבא:

- בוצע JOIN בין שכבת ה SHP הרלוונטית לבין תוצר קוד ה SQL (שנשמר בפורמט XLS, לדוגמא מזג אוויר - LONG SHEET). את התוצר נשמור ב MDB הרלוונטי (בשם XXXX_Clastered_Wirhout_LastAcc).
- הורצה שאילתה בתוך מסד נתוני ה ACCESS (MDB), המבצעת חיבור בין XXXX_Clastered_Wirhout_LastAcc לבין q2_res. שאילתה זו משלימה לכל "קבוצה" את התאונה האחרונה, על כל הפרמטרים שלה (מספר קבוצה, CNT וכיוצא בזאת).
- שוב, בוצע JOIN בין שכבת ה SHP הרלוונטית (לדוגמא מזג אוויר) לבין התוצר שהתקבל בשלב הקודם. נבצע ייצוא לפורמט SHP בתוך ה MDB הרלוונטי.
- הוספת עמודה נוספת (Cnt_UpDate) עבור ה CNT העדכני (יש להוסיף CNT+1·).
- ביצוע BUFFER +DISSOLVED לפי GROUP NUMBER (לקבוצת ה 100 מ' BUFFER של 55 מטרים ולקבוצת ה 200 מ' BUFFER של 110.

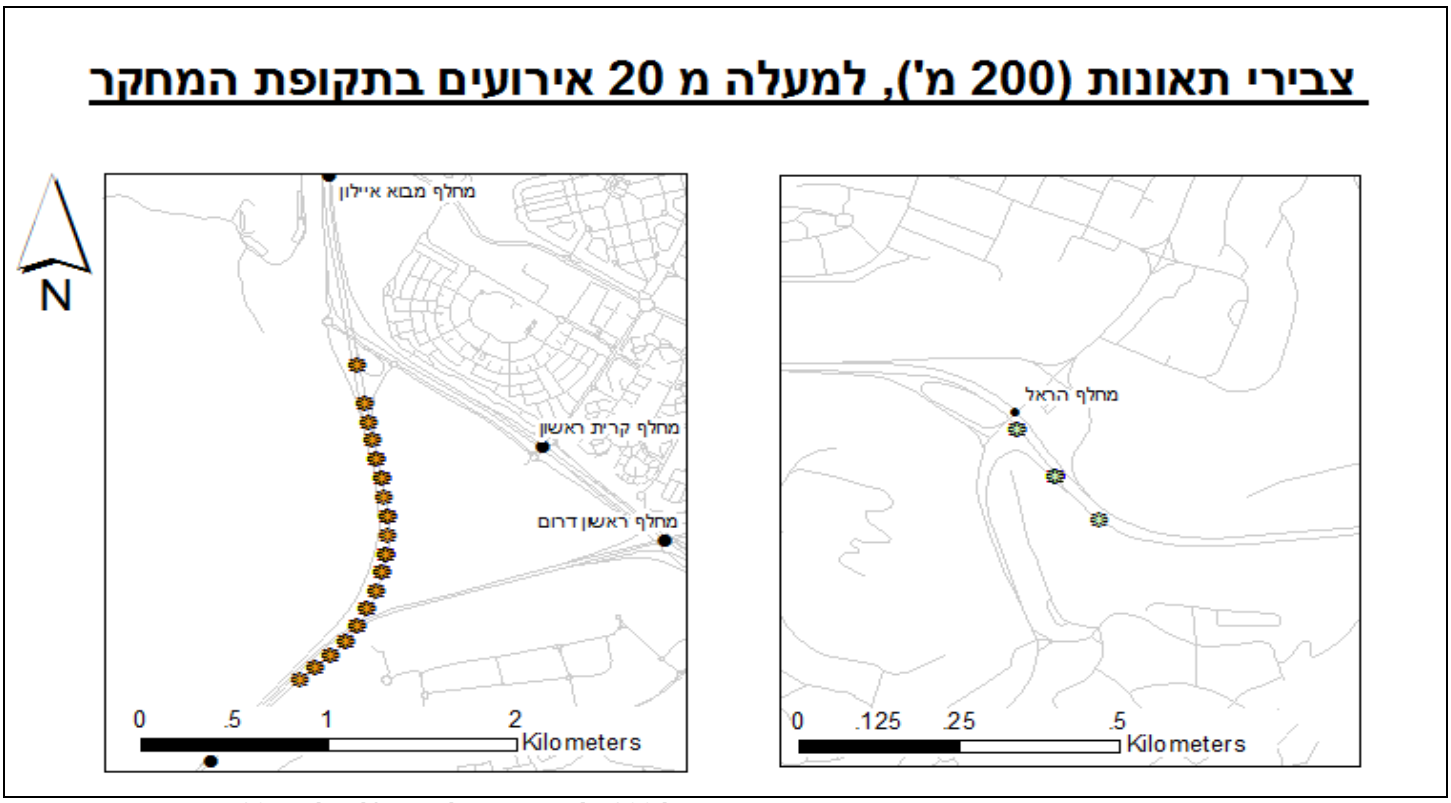
5.2.3 ניתוח ותוצאות

איורים 57-54 מציגים את תוצאות איתור הצבירים עבור כלל תאונות הדרכים בתקופת המחקר, במרחקים שונים (200 מ', 150 מ' ו - 100מ'). ניתן לראות באופן ברור, כי קיים ריכוז תאונות במחלפים, המאופיינים בנפח תנועה גבוה וריבוי נתיבים (דוגמת נתיבי איילון).

איורים 61-58 מציגים את תוצאות איתור צבירי תאונות, שניתן לקושרן להתרחשות בהשפעת מזג אוויר גשום. בתחילה הועלתה סברה, לפיה, ניתן יהיה להבחין בהתרחשות התאונות בעקומות חדות, אולם התברר כי הבחינה החזותית אינה מאתרת קשר שכזה.

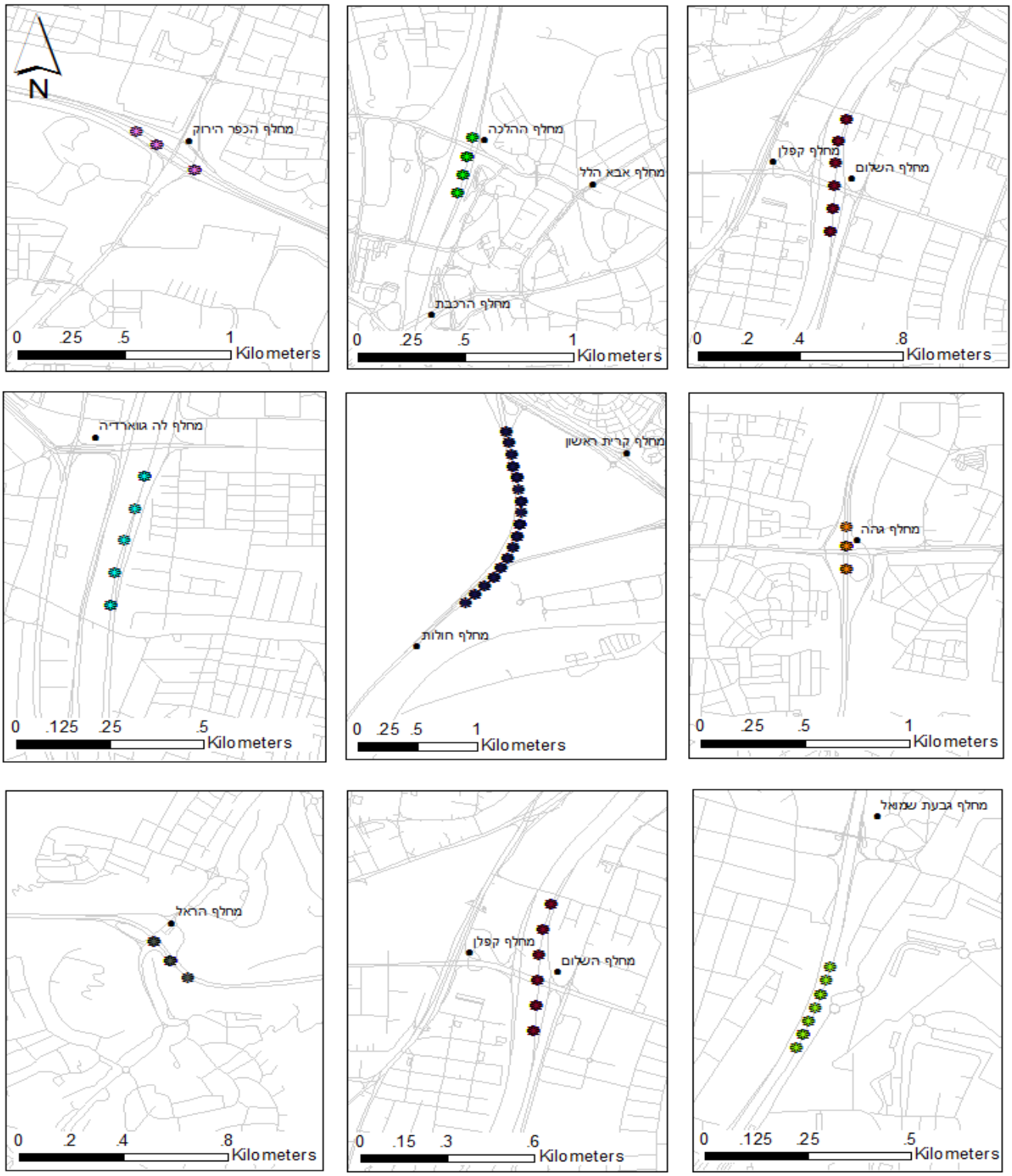


איור מספר 54 - צבירי תאונות דרכים –מרחק של 200 מ' בין אירועי לאירוע (למעלה מ 20 אירועים בתקופת המחקר)

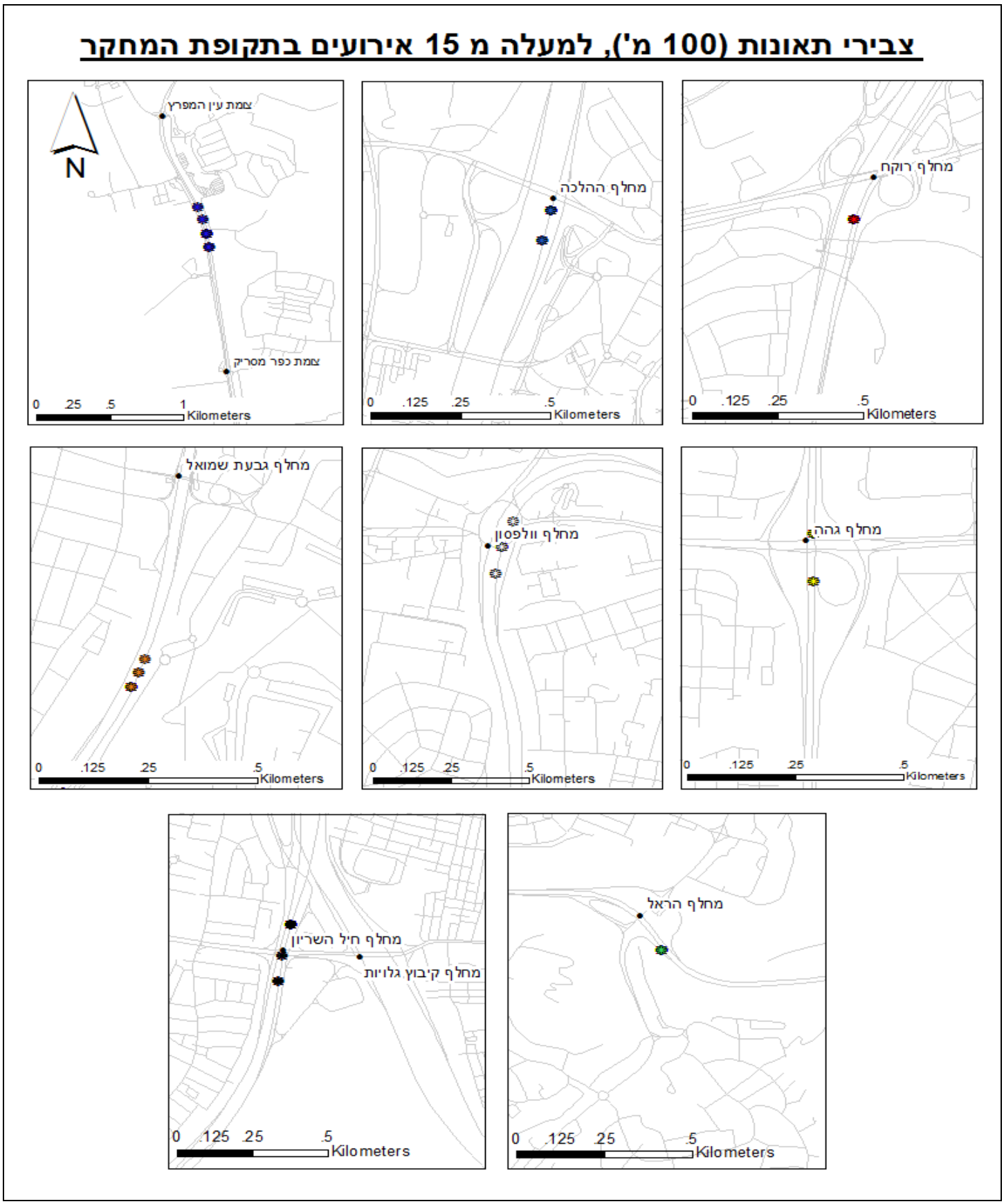


איור מספר 55 – המשך: צבירי תאונות דרכים – מרחק של 200 מ' בין אירועי לאירוע (למעלה מ 20 אירועים בתקופת המחקר)

צבירי תאונות (150מ') - למעלה מ 20 אירועים בתקופת המחקר

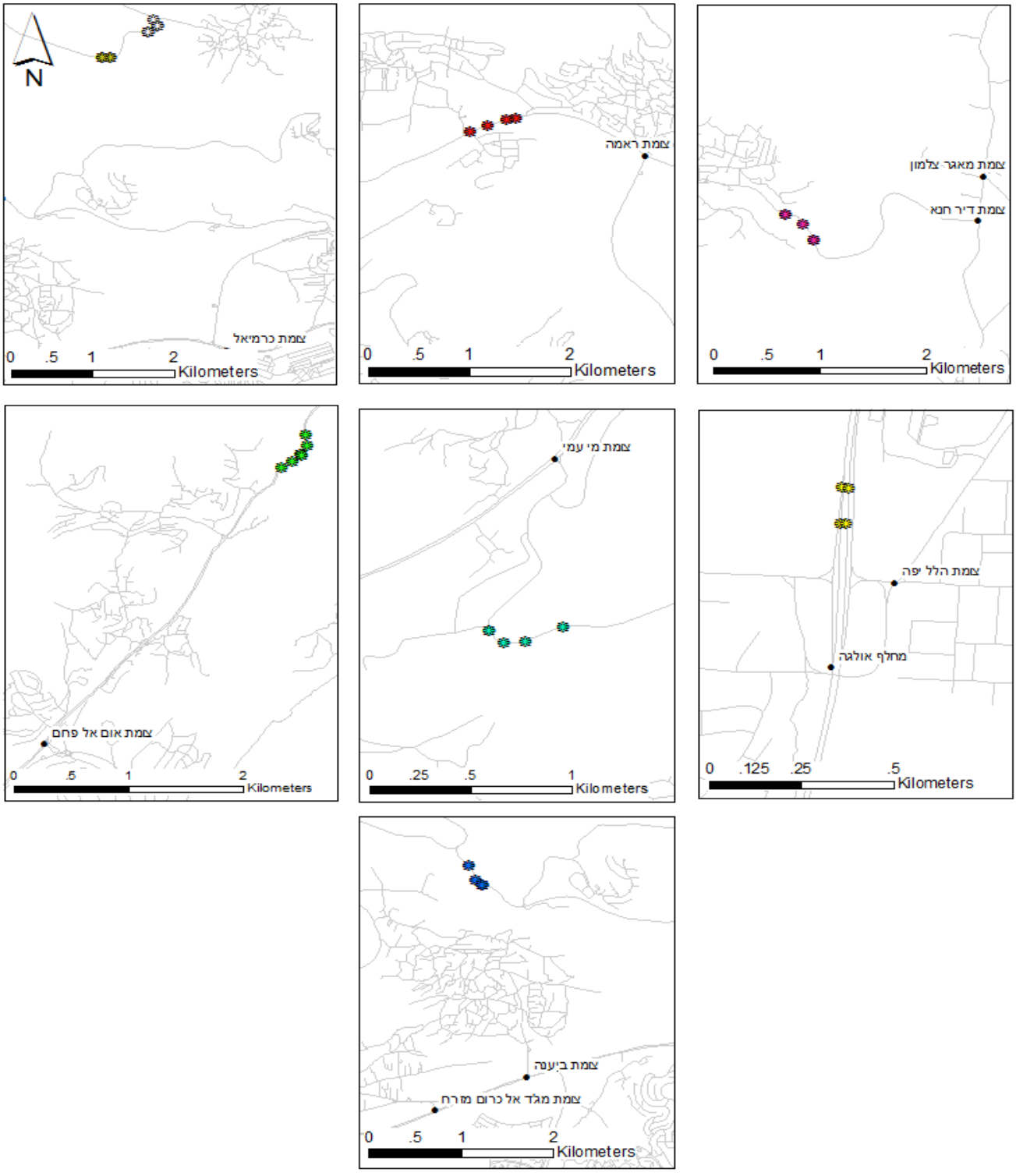


איור מספר 56 - צבירי תאונות דרכים –מרחק של 150 מ' בין אירועי לאירוע (למעלה מ 20 אירועים בתקופת המחקר)



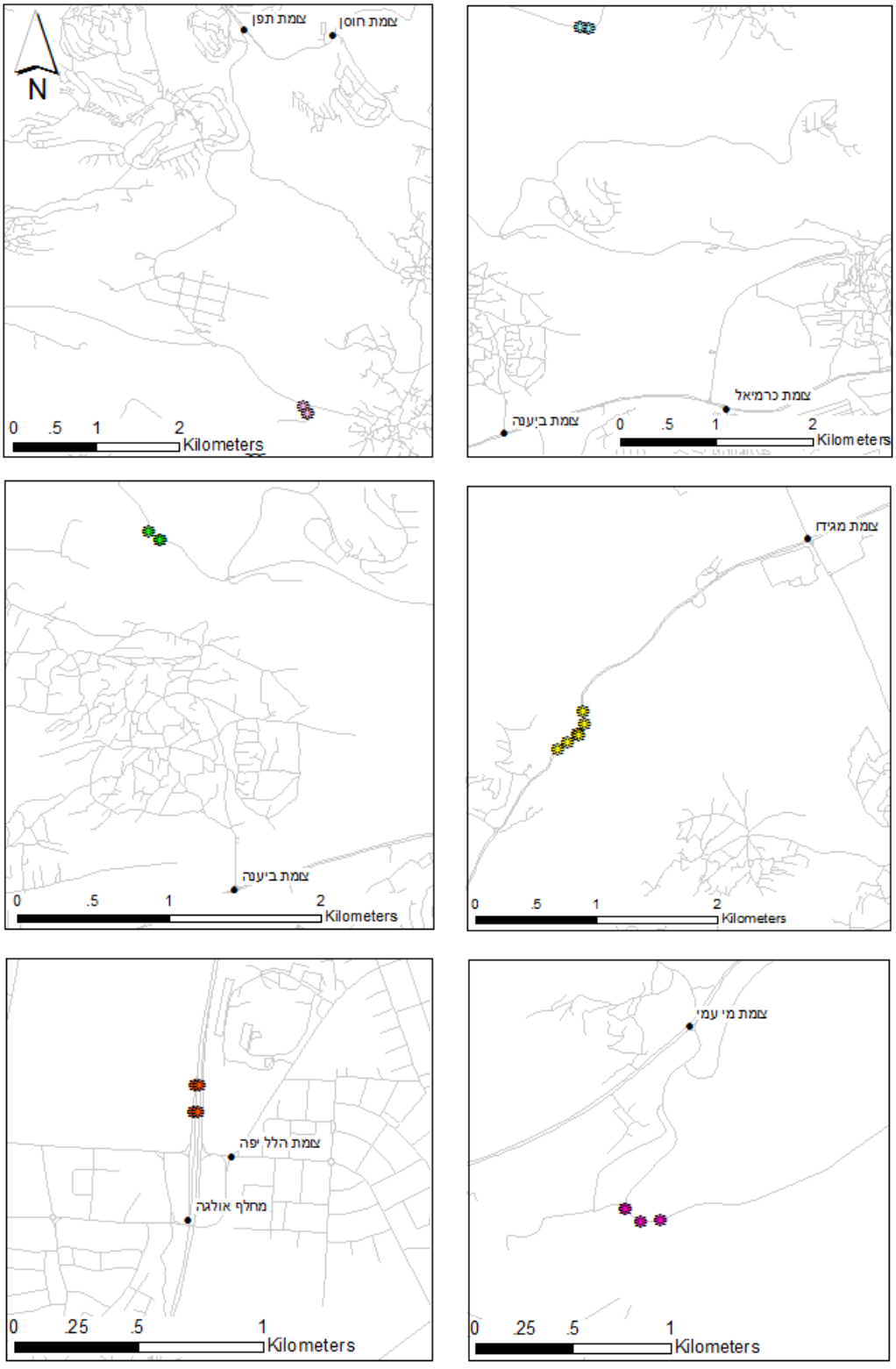
איור מספר 57 - צבירי תאונות דרכים – מרחק של 100 מ' בין אירועי לאירוע (למעלה מ 15 אירועים בתקופת המחקר)

צבירי תאונות תחת השפעת מזג אוויר (200 מ') - 3 אירועים ומעלה בתקופת המחקר

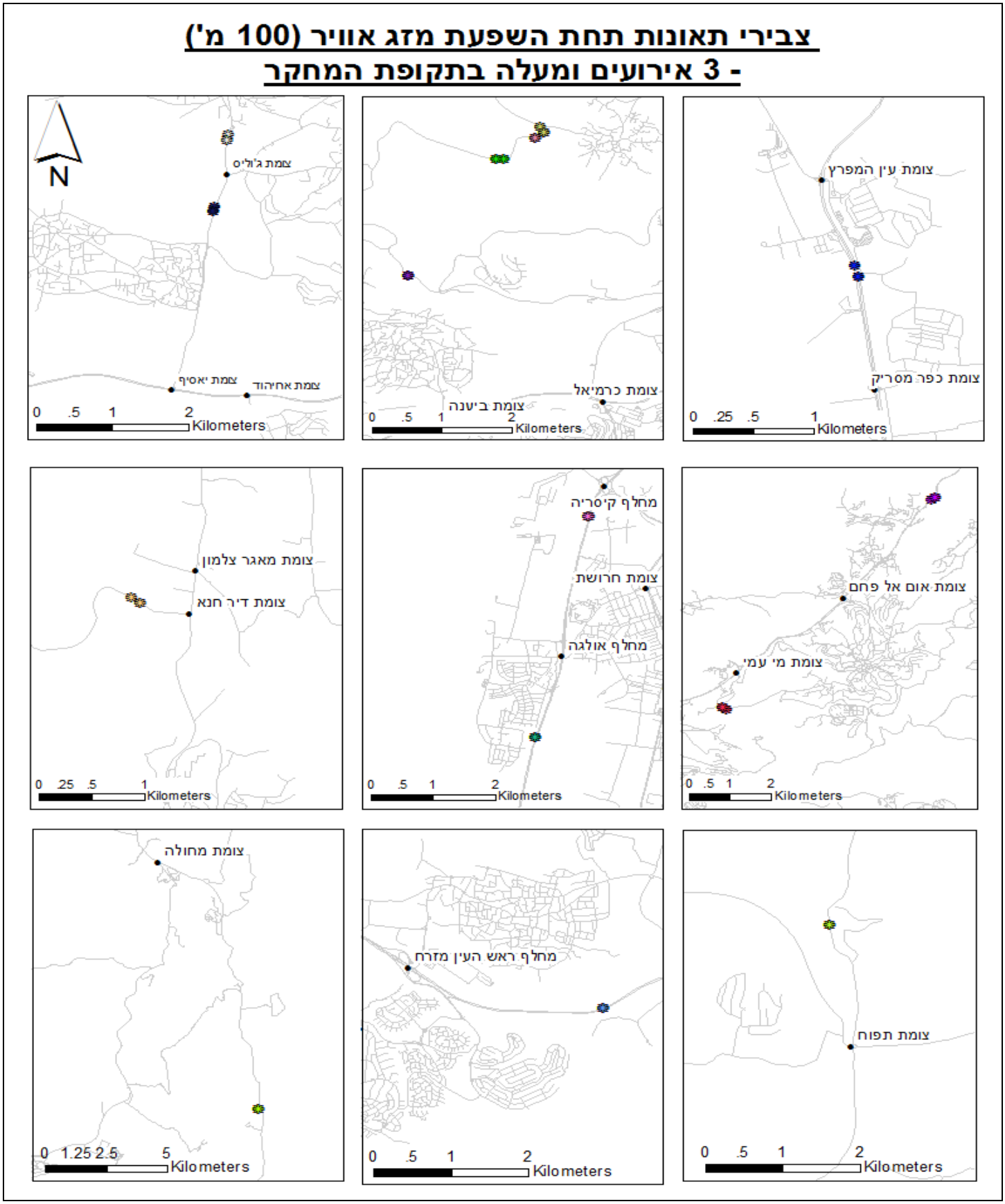


איור מספר 58 - צבירי תאונות דרכים שניתן לקושרם למזג האוויר– מרחק של 200 מ' בין אירועי לאירוע (3 אירועים ומעלה בתקופת המחקר)

צבירי תאונות תחת השפעת מזג אוויר (150מ') - 3 אירועים ומעלה בתקופת המחקר



איור מספר 59 - צבירי תאונות דרכים שניתן לקושרם למזג האוויר– מרחק של 150 מ' בין אירועי לאירוע (3 אירועים ומעלה בתקופת המחקר)



איור מספר 60 - צבירי תאונות דרכים שניתן לקושרם למזג האוויר– מרחק של 100 מ' בין אירועי לאירוע (3 אירועים ומעלה בתקופת המחקר)

איור מספר 61 - צבירי תאונות דרכים שניתן לקושרם למזג האוויר– מרחק של 100 מ' בין אירועי לאירוע (3 אירועים ומעלה בתקופת המחקר)

5.3 השוואת מתודת KRIGING למתודת איתור מקטעים מסוכנים לאורך ליניאמנטים

איורים מספר 62 ו- 63 מציגים את תוצאות השוואת ניתוח ה– kriging המובנית בתוכנת ה- ArcGIS, לניתוח מקטעי הדרך המסוכנים (צבירי אירועים), כפי שאותרו לפי קוד ה- SQL שפותח.

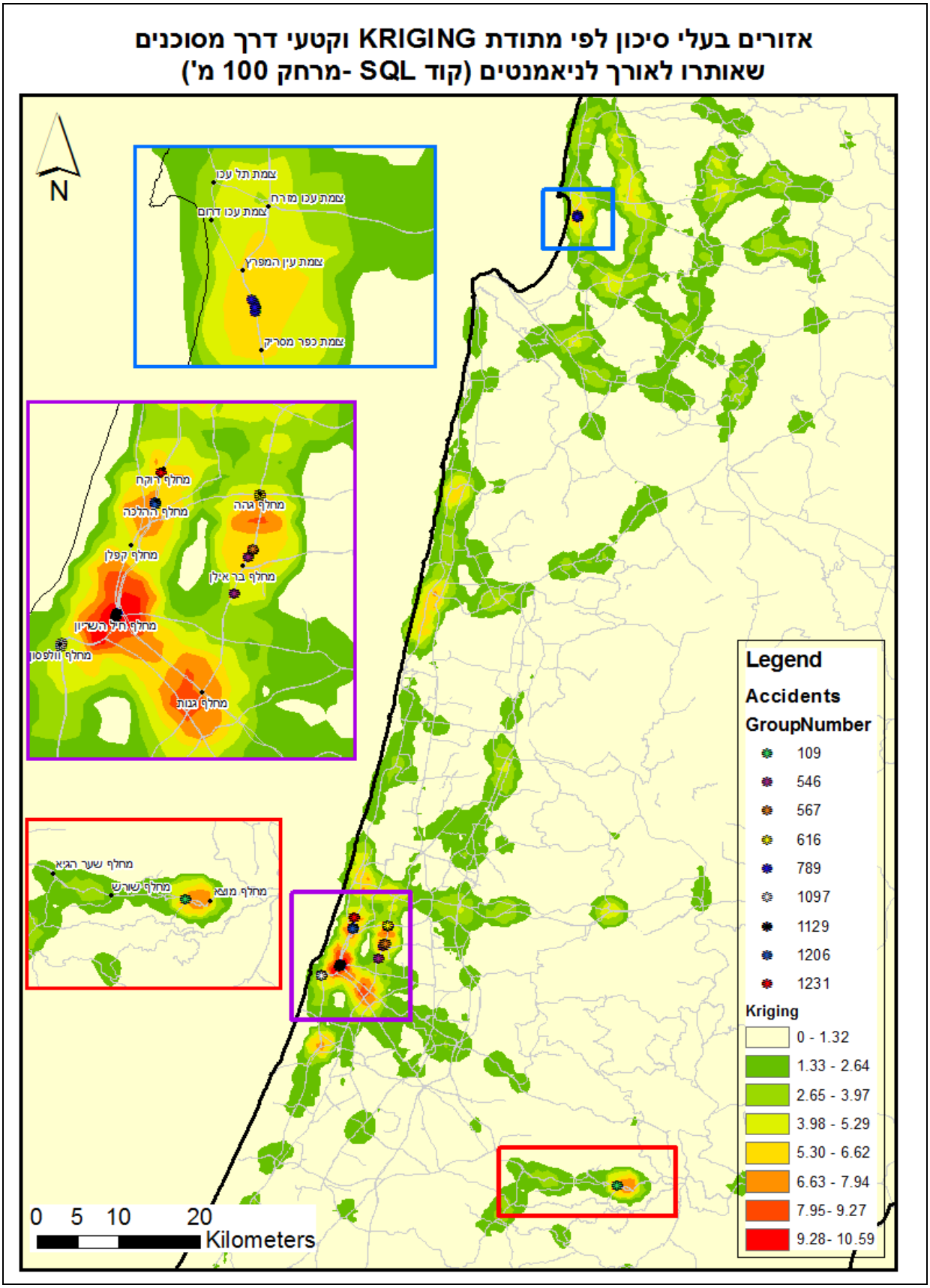
נזכיר, כי באיורים אלה, חושבה מתודת ה - KRIGING בהתאם לקלט נקודתי, המייצג את מספר אירועי תאונות דרכים עבור תא ברשת. למשל, כאשר במפת התוצר מופיע תא בצבע אדום משמעו כי צפוי שם מספר גבוה של תאונות. בנוסף, בקוד ה- SQL, חושבו הצבירים לפי שני מרחקי חיפוש של 100 מ' ו- 200 מ', בין תאונה לתאונה.

ניתן לראות כי, לרוב, קיים מתאם טוב בין הצבירים שאותרו תוך שימוש בקוד ה - sql לבין האזורים שאותרו בשיטת ה- Kriging (הפוליגונים המופיעים בגווני "אדום" ו"כתום" במפה). כן נמצא מתאם פחות טוב כאשר מבוצעת השוואה הפוכה.

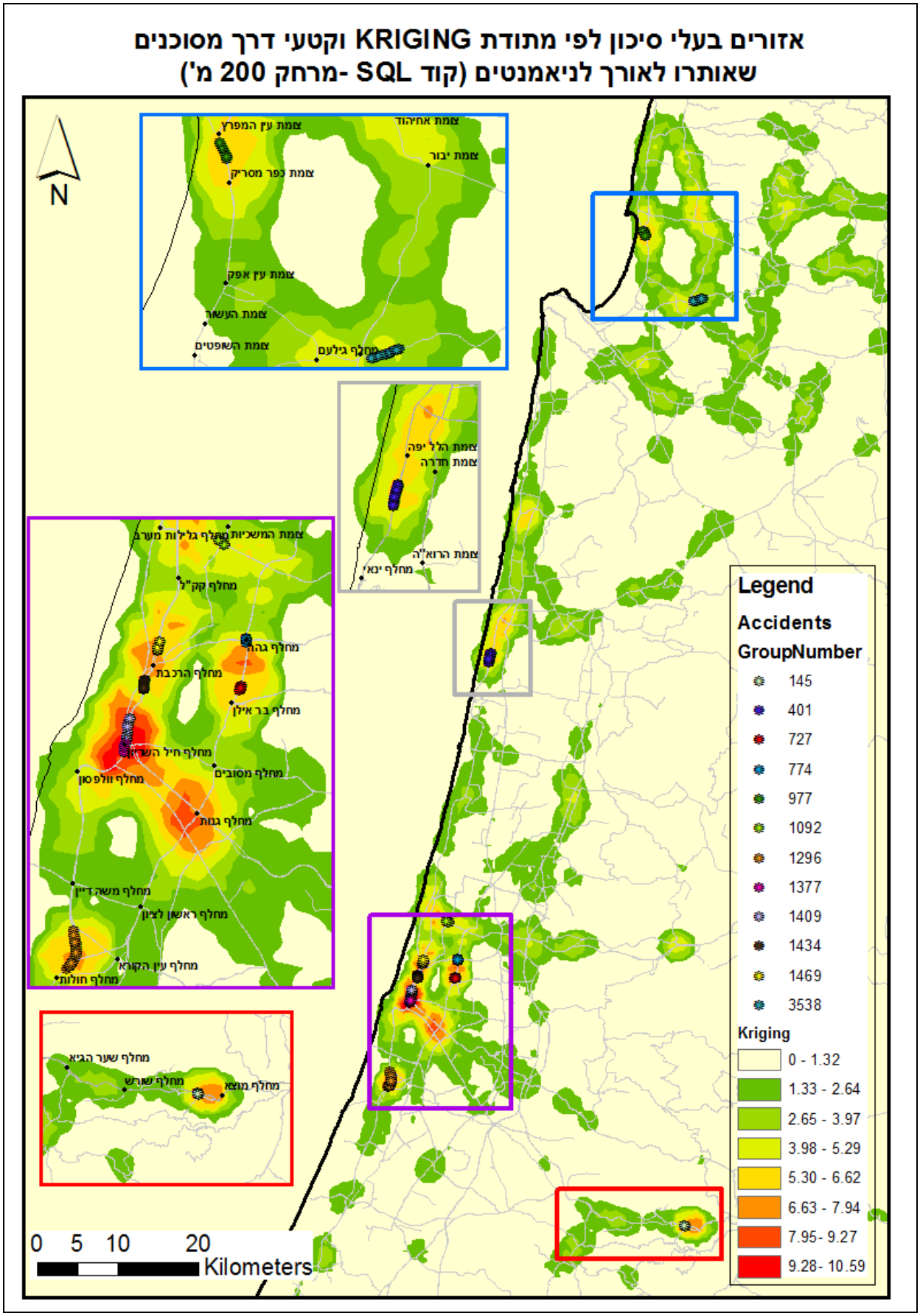
באופן ברור, ניתן לראות כי השימוש בשיטת ה - Kriging, אינו מיטבי עבור ניתוח מידע המצוי לאורך ליניאמנט. זאת, מאחר והאלגוריתם ממשקל את הערכים הסובבים בכדי לבצע ניבוי עבור מיקום גיאוגרפי עבורו לא התקבלה דגימה. המישקול, אגב, מבוסס לא רק על המרחק בין הנקודות הנמדדות והמיקומים הצפויים, אלא גם על הסידור המרחבי של כלל הנקודות. תחת הנחה זו, ניתן

לראות כי לרוב מאותרים אזורים בעלי מספר תאונות צפוי גבוה (אדום) וגם בינוני (ירקרק) באזורי צמתים, ובאזורים בהם קיימים כבישים הקרובים (פיזית, מבחינת מרחק) זה לזה. יודגש, כי ניתן להקטין השפעות אלה, במידה ויוגדר תא רשת קטן (רזולוציה מרחבית גבוהה).

עוד חשוב לציין, כי התוצאות שהתקבלו עבור הצבירים שהוגדרו על בסיס מרחק חיפוש בטווח של 200 מ', נמצאו מתאימות יותר למתודת ה- Kriging מאשר התוצאות שהתקבלו עבור הצבירים שנקבעו לפי מרחק חיפוש של 100 מ'. ממצא זה גם תואם את דיוק המודל. כזכור, מידע על מיקום תאונות הדרכים, מסופק מלכתחילה באי ודאות של עשרות עד למאות מטרים. משמע, הסיכוי ל"תזוזה" במיקום התאונה מוערך על ידינו במצב הנוכחי, בכ - 100 מטרים, לפחות.



איור מספר 62 – אזורים בעלי סיכון לפי מתודת Kriging וקטעי דרך מסוכנים שאותרו לאורך ליניאמנטים (קוד SQL עבור מרחק של 100 מ')



איור מספר 63 - אזורים בעלי סיכון לפי מתודת Kriging וקטעי דרך מסוכנים שאותרו לאורך ליניאמנטים (קוד SQL עבור מרחק של 100 מ')

6. מערך ניסויים לבניית המודל

6.1 מהלך העבודה - בניית המודל הסטטיסטי

- המודל הסטטיסטי חושב לחוד על בסיס מקטעי הדרך לאורכיהם השונים, שסופקו כ INPUT לתוכנת ה- SAS.
- עבור בניית המודלים, חולק בסיס המידע המקורי, באופן אקראי ל - **2/3, המהווים את "מודל האימון"**, ול - **1/3 המהווים את "קבוצת הביקורת".**
- **1/3 הנתונים היווה כאמור "מודל ביקורת"**. במודל זה שולבו אותם הפרמטרים שהתקבלו משמעותיים במודל **האימון**, על מנת לבחון את יציבות המודל. בנוסף, כחלק מבקרת האיכות למודל, נבנה מ- 1/3 הנתונים גם "**מודל ביקורת מלא**" (במודל זה הורצו כלל הפרמטרים הקיימים בבסיס הנתונים, ונבחן אלו פרמטרים משמעותיים עבור מודל 1/3 הנתונים. זאת, **ללא תלות במודל האימון**, בכדי לבחון פעם נוספת את יציבות המודל.
- עבור כל סט נתונים, נבחנו ארבעה מודלים סטטיסטיים המאופיינים בהנחות שונות בנוגע להתפלגות מספר תאונות הדרכים. זאת, מתוך ההנחה כי מספר התאונות במקטע דרך אינו מתפלג נורמאלית :

(א) פואסון (Poisson) .

(ב) בינומי שלילי (negative binomial).

(ג) Zero-Inflated Poisson. במציאות נתוני תופעות שונות, מאופיינים ב over - dispersion ובאפסים רבים. מודלי ספירה מסוג Zero-Inflated מספקים דרך למידול תופעות, כאשר בבסיס הנתונים קיימים אפסים רבים. בנוסף מאפשרים גם overdispersion. מודלים מסוג זה מניחים כי המידע המתקבל הינו תהליך בו מעורבים שני סוגי מידע שונים. הראשון הינו "אפסים", השני הינו מידע בעל פיזור מסוג פואסון או בינומי שלילי. תוצאת משפט ברנולי משמשת בכדי לקבוע איזה משני התהליכים הוא זה שיקבע את ערך התצפית הצפויה.

[http://www2.sas.com/proceedings/forum2008/322-2008.pdf]

מכאן, שנוסף על מידול הפרמטרים המשמעותיים לניבוי מספר תאונות, מודל זה מאפשר גם למדל את הפרמטרים המסבירים את ההבדל בין 0 (מקטעים בהם אין כלל תאונות) לבין 1 (מקטעים בהם התרחשה לפחות תאונה אחת) בתקופת המחקר. יצויין, כי בשימוש במודל זה, יכול בהחלט להיות שהפרמטרים המסבירים והמשמעותיים בחיזוי מספר התאונות, יתקבלו שונים לגמרי מאלו המסבירים את ההבדל בין 0 ל 1.

(ד) Zero-Inflated negative binomial, בדומה לשהוסבר בסעיף הקודם, אלא שהמודל הינו מסוג בינומי שלילי ולא פואסוני.

- אחת השאלות המרכזיות שעלתה במהלך סקר הספרות, ועבורה נדרשה מחשבה מיוחדת, היתה כיצד להתייחס למשתנים: "אורך הסגמנט" ונפח התנועה (" AADT"). האם לשלבם כמשתנים מסבירים במודל, או האם לנרמלם כדי לעשות בהם שימוש כמשתני offset. כך לדוגמא, במקום לשלב את נפח התנועה AADT כמשתנה מסביר, ניתן לשלבו במודל כמשתנה חדש, דוגמת מספר תאונות מתוקנן ב AADT. היות ובמחקרים רבים עלה כי ה-AADT הינו אחד מהמשתנים המסבירים המשמעותיים ביותר, הוחלט לשלבו כמשתנה מסביר במודל. ההתמודדות עם שאלת "אורך הסגמנט" קוימה, על ידי חלוקת בסיס הנתונים למקטעים הזהים באורכם. כך, למעשה "נוטרלה" השפעת פרמטר זה. משמע, הפרמטר לא שולב במודל לא כמשתנה offset ולא כמשתנה מסביר.
- היות והתפלגות פרמטר ה- AADT הינה א-סימטרית ימנית, בוצעה טרנספורמציה למשתנה על ידי שימוש ב LOG עם בסיס טבעי (AADT_LOG), כאשר זהו הנתון ששולב במודל. יודגש, כי טרם שבוצעה הטרנספורמציה, נבחן טיב המודל, והתגלה כי אכן טרנספורמציה זו משפרת את טיב ההתאמה של המודל.
- עבור כל מקטע דרך, חושב פרמטר המייצג את סכום חומרת התאונות הממושקל עליו (להלן: SumHumraMemushkelet). פרמטר זה, לא שולב בסופו של דבר כפרמטר אינטגרלי בניתוח הסטטיסטי השוטף. זאת, מכיוון שפרמטר זה מייצר קשר מלאכותי בין חומרת התאונות שהתרחשו על הסגמנט לבין מספר התאונות על אותו סגמנט, שהרי מקטע עליו לא התרחשו כלל תאונות יאופיין בערך חומרה של 0. והיות ובבסיס הנתונים הנבחן, למעלה מחמישים אחוזים ממקטעי הדרך מאופיינים בערך של 0 תאונות, עלול היה להיווצר מצב, לפיו, גם אם לא קיים קשר סטטיסטי מובהק בין חומרת התאונה למספרן, הוא בכל זאת היה נמצא ככזה.

6.2 בחירת הפרמטרים המשמעותיים עבור המודל.

לרוב, על מנת לבחור איזה פרמטר מסביר נשלב במודל, נחפש איזה פרמטר עולה כמובהק. במחקרים בהם בסיס הנתונים הינו גדול באופן יחסי (דוגמת מחקר זה), בחינת המובהקות איננה מספקת, מכיוון שככל שבסיס הנתונים גדול יותר, סביר כי השפעות מאוד נמוכות של הפרמטרים יהיו מובהקות.

אחד הפיתרונות לכך, הינו שימוש ב- AIC (Akaike Information Criterion), המשמש כקריטריון לבחינת טיב התאמת המודל. קריטריון זה עושה שימוש בפונקציית ה- likelihood. פונקציה זו בוחנת כיצד הפרמטר הממודל מתאים לתצפיות האמיתיות שלנו, או במילים אחרות, מהי ההתאמה הטובה ביותר ("המקסימאלית") בין ההסתברות הנאמדת (ה"צפויה") לבין ההסתברות ה"אמיתית" (זו שחושבה על בסיס כלל התצפיות במדגם).

ה- AIC עושה שימוש בפונקציית ה likelihood , כאשר עבור כל משתנה נוסף שנכלל במודל ההליך משלב מעין "עונש", היות ולמעשה, כל משתנה נוסף שישולב במודל תמיד ישפר אותו. המטרה היא לקבל AIC כמה שיותר קטן. ככול שערך ה - AIC קטן יותר, משמע טיב המודל טוב יותר.

6.3 שגרות מחשב

בתוכנת SAS 9.2 קיימות שתי פרוצדורות (שגרות) שונות בהן נעשה שימוש. GENMOD ו-COUNTREG. כל אחת מספקת הן נתונים חופפים והן נתונים משלימים לפרוצדורה האחרת. מודלים מסוג Zero-Inflated , מצויים רק בפרוצדורת ה- COUNTREG.

6.4 תוצרים וניתוחם (דוגמת 500 המטרים)

בתת פרק זה, נציג כדוגמא, את מהלך בניית המודל עבור סט הנתונים של מקטעי הדרך שחולקו לאורך של 500 מטרים.

6.4.1 בחירת המודל

טבלה מספר 23 מציגה את תוצר ה AIC עבור מקטעי דרך באורך של 500 מטרים. כאמור, הערך הנמוך ביותר הוא הטוב ביותר, מבחינת טיב המודל.

טבלה 21 - תוצר AIC עבור מקטעי דרך באורך 500 מ'

AIC	L=500
פואסוני	5494
בינומי-שלילי	**4738**
Zero-Inflated פואסון	4941
Zero-Inflated בינומי-שלילי	4740

לאור האמור, המודל שיצא הטוב ביותר עבור סט הנתונים של 500 מטרים – הינו Negative Binomial.

6.4.2 פרוצדורת GENMOD

6.4.2.1 ניתוח טבלת "Analysis Of Maximum Likelihood Parameter Estimates".

טבלת ה- **Model Information** (טבלה 24) - מספקת מידע על הנתונים שהוכנסו כקלט לפרוצדורה. כך לדוגמא, מוצג נתוח שבוצע על הסגמנטים באורך של 500 מטרים (דגימת

2/3 מהמידע). המשתנה התלוי (**Dependent Variable**) אותו ננסה להסביר באמצעות המשתנים המסבירים, הבלתי תלויים הינו מספר התאונות במקטע דרך.

טבלה 22 - (מקטעי דרך באורך 500 מ') Model Information

Model Information		
Data Set	K.TRAIN500	
Distribution	Negative Binomial	
Link Function	Log	
Dependent Variable	AccCNT	AccCNT

טבלה מספר 25, **Criteria For Assessing Goodness Of Fit** , מרכזת פרמטרים המייצגים את טיב המודל, לרבות פרמטר ה AIC שהורחב לגביו.

טבלה 23 - טבלת Criteria For Assessing Goodness Of Fit **(מקטעי דרך באורך 500 מ')**

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	2183	1719.9737	0.7879
Scaled Deviance	2183	1719.9737	0.7879
Pearson Chi-Square	2183	2398.1971	1.0986
Scaled Pearson X2	2183	2398.1971	1.0986
Log Likelihood		-1419.0795	
Full Log Likelihood		-2363.1286	
AIC (smaller is better)		4738.2571	
AICC (smaller is better)		4738.2956	
BIC (smaller is better)		4772.4016	

Deviance – ערך הקטן מ 1 מצביע על התאמה טובה מבחינת המודל

טבלה 24 - תוצר פרוצדורת GENMOD עבור מודל האימון של 2/3 הנתונים (ניתוח סגמנטים באורך 500 מ')

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Likelihood Ratio 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-7.0910	0.6395	-8.3444	-5.8376	**-8.3537**	**-5.8444**	122.95	<.0001
aadt_log	1	0.7144	0.0648	0.5875	0.8414	**0.5883**	**0.8423**	121.73	<.0001
ezor1	1	-0.6788	0.1210	-0.9160	-0.4415	**-0.9173**	**-0.4423**	31.45	<.0001
ezor3	1	-0.5694	0.0890	-0.7438	-0.3950	**-0.7444**	**-0.3952**	40.94	<.0001

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
CurvePrm3	1	1.5948	0.4107	0.7899	2.3997	**0.7953**	**2.4085**	15.08	0.0001
Dispersion**	1	1.7419	0.1351	1.4771	2.0067	**1.4925**	**2.0234**		[23]

Note: The negative binomial dispersion parameter was estimated by maximum likelihood.

טבלה מספר 26 מציגה את המשתנים שנמצאו כחשובים במודל:

- **aadt_log** - ממוצע נפח תנועה יומי לימי חול, לאחר שעבר טרנספורמציה ללוגריתם טבעי.
- **ezor1** (דרום), **ezor3** (מרכז) – מייצגים משתנה האזור, כאשר המודל משווה ומנתח בין אזור 2 (צפון) לשני האזורים האחרים.
- **CurvePrm3** – עקמומיות.

בנוסף קיים פרמטר ה- **Dispersion**. פרמטר זה, מגדיר האם עסקינן במודל בינומי שלילי או מודל פואסוני. ברגרסייה פואסונית, מניחים כי ערכו של פרמטר ה- **Dispersion** שווה ל 1 (משמע אין פיזור יתרה).

בעזרת מבחן יחס הנראות, ניתן לבחון האם פרמטר פיזור היתרה שונה באופן מובהק מ-1. בכדי לבצע זאת, יש לבדוק עד כמה מקדם המשתנה (Estimate) של פרמטר ה-Dispersion , שערכו במודל לדוגמא שלנו 1.7419, קרוב מספיק או רחוק מספיק מ-1. לשם כך נבחן את רווח הסמך המופיע בטבלת **Analysis Of Maximum Likelihood Parameter Estimates** בעמודת **"Likelihood Ratio 95% Confidence Limits"**, כאשר:

במידה והערך **1** נופל בטווח רווח הסמך- נקבל את H_0: **$\alpha = 1$; משמע אין** פיזור יתר (overdispersion) והמודל המתאים הוא פואסוני.

כאשר הערך **1** אינו נופל בטווח רווח הסמך נקבל את H_1: **$\alpha \neq 1$; משמע יש** פיזור יתר (overdispersion) ואז המודל המתאים הוא מסוג negative binomial ולא פואסון.

כפי שניתן לראות בטבלה מספר 28 הערך 1, אינו נופל ברווח הסמך (הנע בין 1.4925 ל 2.0234) ברמת מובהקות 95%. לאור האמור, ניתן להסיק כי קיים פיזור יתרה בהתפלגות של מספר תאונות הדרכים, ולפיכך **המודל המתאים יותר הינו מסוג בינומי שלילי**.

[23] ** שורת ה- Dispersion בטבלת **Analysis Of Maximum Likelihood Parameter Estimates** , למעשה מקבילה ל **Alpha_ המופיע בטבלת Parameter Estimates**, בפרוצדורת COUNTREG .

בנוסף, על פי טבלה מספר 21, ניתן לבנות עבור מודל האימון שלנו (2/3 הנתונים), את **הנוסחה לניבוי תוחלת תאונות הדרכים μ** (ממוצע תאונות למקטע דרך לתקופת המחקר העומדת על שלוש שנים). הערכים המנובאים מתקבלים דרך הקשר הלוגריתמי. בהתאם לנוסחה 77

[77] $$Log\mu = b_0 + a_1x_1 + a_2x_2 + ...$$

כאשר יוצבו ערכי המקדמים המופיעים בטבלה 27 בנוסחה מספר 77, תתקבל נוסחה 78 לפיה נחשב ערכים מנובאים מהמודל שנבנה על 2/3 הנתונים.

[78] $$Log\mu = -7.091 + 0.714 * AADT\_Log + (-0.678) * Ezor1 + (0.569) * Ezor3 + (1.594) * Curve\Pr m3$$

מעיבוד נוסחה 78 תתקבל המשוואה הסופית 79, עבור תוחלת התאונות התצפויה לפי המודל שנבנה מ 2/3 הנתונים:

[79]

$$\mu_{2/3} = e^{-7.091+0.714*AADT\_Log+(-0.678)*Ezor1+(0.569)*Ezor3+(1.594)*Curve\Pr m3}$$

בנוסף, מטבלת **Analysis Of Maximum Likelihood Parameter Estimates** (טבלה מספר 27), ניתן להסיק לגבי היחסים בין המשתנים. כך לדוגמא, נבחן את משתנה האזור: **ezor1** , **ezor3**. המינוס משמעו שבאזור 2 (צפון) = **ezor2**, אליו משווים את שני האזורים האחרים, צפויות להתרחש יותר תאונות מאשר באזור 1 (דרום) שערך המקדם שלו עומד על -0.6788. כמו כן, באזור 2 (צפון) גם מתרחשות יותר תאונות מאשר באזור 3 (מרכז) שערך המקדם שלו עומד על -0.5694. מסקנה אופרטיבית הינה כי נדרש שיפור תשתית הכבישים בצפון הארץ.

6.4.2.2 ניתוח טבלת Contrast Estimate Results

תוצר נוסף של הפרוצדורה, הינה טבלת ***Contrast Estimate Results*** (טבלה מספר 27). מטבלה זו ניתן לגזור את משמעויות המשתנים. בעמודת **"Mean Estimate"** מוצג e בגובה "המקדם" המופיע בטבלת ***Analysis Of Maximum Likelihood Parameter Estimates*** (טבלה מספר 26, בעמודת ה- ***Estimate***)

טבלה 25 - Contrast Estimate Results, **מקטעי דרך באורך 500 מ'.**

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
adj. rr: CurvePrm3	4.9275	2.2032	11.0203	1.5948	0.4107	0.05	0.7899	2.3997	15.08	0.0001
Exp(adj. rr: CurvePrm3)				4.9275	2.0236	0.05	2.2032	11.0203		
adj. rr: AADT_log	2.0431	1.7995	2.3195	0.7144	0.0648	0.05	0.5875	0.8414	121.73	<.0001
Exp(adj. rr: AADT_log)				2.0431	0.1323	0.05	1.7995	2.3195		
adj. rr: ezor1 vs 2	0.5072	0.4001	0.6430	-0.6788	0.1210	0.05	-0.9160	-0.4415	31.45	<.0001
Exp(adj. rr: ezor1 vs 2)				0.5072	0.0614	0.05	0.4001	0.6430		
adj. rr: ezor3 vs 2	0.5659	0.4753	0.6737	-0.5694	0.0890	0.05	-0.7438	-0.3950	40.94	<.0001
Exp(adj. rr: ezor3 vs 2)				0.5659	0.0504	0.05	0.4753	0.6737		

מטבלה 27 עולה כי באזור הצפון (2) קיים שיעור תאונות צפוי הגדול **פי** 1.97 משיעור התאונות הצפוי בדרום (אזור 1).
באזור הצפון (2) קיים שיעור תאונות צפוי הגדול פי 1.76 משיעור התאונות באזור המרכז (3).

AADT_log – AAAD מייצג כאמור, ממוצע נפח תנועה יומי לימי חול (מספר כלי הרכב שנעו בקטע דרך מסויים בימי חול). מטעמי טיב המודל, בוצעה טרנספורמצית LN(x) על כלל נתוני נפח התנועה במקטעי הדרך. המשמעות היא כי כל ערך מקורי ישולב וייוצג במודל לאחר הטרנספורמציה, על ידי הלוגריתם הטבעי (שערכו 2.7182) בחזקת מעריך כלשהו.

מבחינת המודל הסטטיסטי בהיבט נפח התנועה, כאשר עולים ביחידה אחת של המשתנה AADT_log (שמשמעה נפח הגדול פי 2.718 מהערך הקודם), שיעור התאונות הצפוי עתיד לגדול פי 2.04. כך לדוגמא, עבור מקטע דרך המאופיין בנפח תנועה של 1000 מכוניות ו 3 תאונות בתקופת המחקר, במידה ונפח התנועה יגדל ביחידה אחת (דהיינו פי 2.7) ויגיע לנפח של 2700 מכוניות, שיעור התאונות עתיד לגדול פי 2.04, וצפוי להיות: 6.12.

CurvePrm3 – ערכי יחידות העקמומיות נעות בין 0 ל- 1. ככל שהערך גבוה יותר, כך העקמומיות בקטע הדרך גבוהה יותר. בניתוח האמור, כאשר נעלה

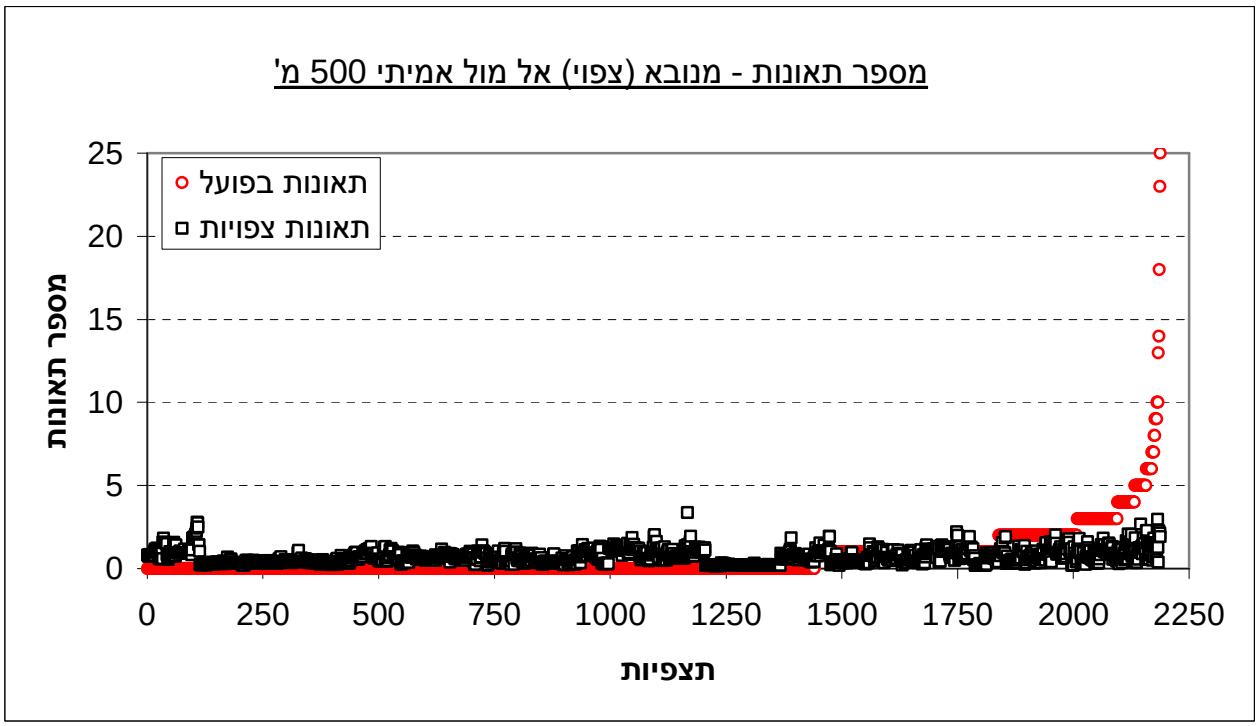
ביחידה אחת של עקמומיות, מספר התאונות במקטע דרך יעלה בכמעט פי 5 (4.9275) .

6.4.3 CROSS VALIDATION של המודל

בכדי להעריך את טיב המודל בוצעו מספר בדיקות, כפי שיפורטו בסעיפים שלהלן.

6.4.3.1 מספר תאונות צפוי אל מול גרף התאונות אמיתי

איור מספר 64 מציג את הגרף המשווה בין מספר התאונות הצפוי (EXPECTED) לפי המודל לבין מספר התאונות בפועל (OBSERVED) בקטעי הדרך באורך 500 מ'. באדום מוצג מספר תאונות הדרכים שהתרחשו בפועל, ובשחור מוצג מספר התאונות הצפוי.



איור מספר 64 - מספר תאונות צפוי אל מול מספר תאונות אמיתי – מקטעי דרך באורך 500 מ'

ניתן לראות כי המתאם אינו מספיק טוב, ולפיכך נדרש להליך ה- EB. הליך זה פורט, כזכור בהרחבה בסקר הספרות בסעיף 2.3.3.4, ובפרק זה בפסקה 6.4.4.

6.4.3.2 השוואת מודל האימון (2/3 הנתונים), למודל שנבנה מנתוני קבוצת הביקורת (1/3 הנתונים).

ההליך שתואר לעיל ובוצע עבור 2/3 מהנתונים, בוצע גם על מסד נתוני הביקורת, המכיל 1/3 מהנתונים. יודגש, כי מראש שולבו במודל זה הפרמטרים המשמעותיים, כפי שעלו

ממודל האימון (2/3 מהנתונים). טבלה מספר 28 מציגה את המקדמים עבור קבוצת הביקורת.

טבלה 26 - תוצר פרוצדורת GENMOD עבור נתוני קבוצת הבקרה של 1/3 הנתונים (ניתוח סגמנטים באורך 500 מ')

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.7780	0.9234	-9.5878	-5.9682	-9.6087	-5.9821	70.95	<.0001
AADT_LOG	1	0.8019	0.0937	0.6183	0.9856	0.6197	0.9878	73.23	<.0001
EZOR1	1	-0.7366	0.1629	-1.0559	-0.4174	-1.0584	-0.4186	20.45	<.0001
EZOR3	1	-0.8810	0.1249	-1.1258	-0.6362	-1.1274	-0.6369	49.75	<.0001
CURVEPRM3	1	0.9894	0.5845	-0.1563	2.1351	-0.1523	2.1489	2.86	0.0905
Dispersion	1	1.5877	0.1764	1.2420	1.9334	1.2701	1.9650		

גם פה ניתן לחשב ערכים מנובאים לפי נוסחה מספר 80:

[80]

$$\mu_{1/3} = e^{-7.778+0.801*AADT\_Log+(-0.736)*Ezor1+(-0.881)*Ezor3+(0.989)*Curve\Pr m3}$$

לאחר שהתקבלו אומדנים משני המודלים, נבחן, האם האומדנים שהתקבלו במודל קבוצת הביקורת (1/3 הנתונים) אינם שונים באופן מובהק ממקדמי מודל נתוני האימון (2/3). כלומר, נבדוק האם "אומדני ה- 1/3" נופלים ברווח הסמך של "מקדמי ה- 2/3", ברמת ביטחון של 95% . לשם כך, יושוו מקדמי המשתנים המסבירים במודל הביקורת המופיעים בעמודת Estimate בטבלת Analysis Of Maximum Likelihood Parameter Estimates (טבלה 28), אל נתוני טווח רווח הסמך של נתוני מודל האימון (2/3) המופיעים בעמודת Likelihood Ratio 95% Confidence Limits (טבלה 26).

כך לדוגמא, ערך מקדם המשתנה נפח תנועה (AADT_LOG) הינו **0.8019**. נבדוק האם מספר זה נופל בתחום רווח הסמך של פונקציית ה Likelihood של נתוני ה 2/3 (טבלה מספר 26). ואכן ניתן לראות, כי מספר זה אכן נכלל בטווח שבין **0.5883** ל **0.8423**. המשמעות היא כי המודל טוב מספיק עבור משתנה זה. כך מהשוואת טבלאות 26 ו- 28 ניתן לראות כי הדבר מתקיים עבור כל המשתנים במודל, **למעט מקדם** המשתנה **ezor3**, המתייחס להבדל בין אזור המרכז (**ezor3**) לאזור הצפון (**ezor2**), שערכו הינו **-0.8810** , ערך אשר אינו נופל בטווח רווח הסמך של נתוני 2/3 הנע בין **-0.3952** ל **-0.7444**. המשמעות הינה כי עבור פרמטר זה המודל אינו טוב מספיק. אפשר והסיבה לכך הינה, שבמודל הביקורת, שנבנה על סמך 1/3 הנתונים המשלימים לסט הנתונים הכולל, שולבו, מלכתחילה, הפרמטרים שיצאו כמשמעותיים במודל הראשוני (2/3 הנתונים). אפשר ובניית מודל מלא מ- 1/3

הנתונים, הבוחן ומשלב את כלל הפרמטרים המקוריים, יניב משתנים אחרים המשמעותיים עבור המודל, ויספק את התשובה לאי ההתאמה בפרמטר זה. בדיקה זו תושלם בהמשך. בנוסף ניתן לראות כי המשתנה **CURVEPRM3** המייצג עקמומיות אינו מובהק במודל הביקורת. נדגיש כי משתנה זה שולב במודל בגין ערך ה- AIC הנמוך שלו (כלומר נמצא בעל משמעות), היות ובמדגמים גדולים נהוג לעשות שימוש בפרמטר זה .

6.4.3.3 מבחן TOST לשוויון (TOST Equivalence Test)

כאמור ניתן להשוות בין הערכים הצפויים למספר תאונות במקטע דרך, לפי המודל האימון (2/3 הנתונים) ולפי מודל הביקורת (1/3 הנתונים). בפועל, בעולם אוטופי, נצפה כי הפרשי התוחלת יהיו 0, היות ושני המודלים מנבאים את אותו הדבר. אולם, היות וברור כי אין אנו חיים בעולם אוטופי, והבדלים יתקיימו, ניתן להניח כי הפרש תוחלת התאונות מצוי בטווח מוגדר כלשהו הנע סביב ה – 0. ננסה לבדוק האם ההפרשים נמצאים בתוך אותו הטווח המוגדר. לשם כך נעשה שימוש במבחן (Two one-sided tests) TOST המאפשר להוכיח שיוויון עבור טווח מוגדר כלשהו[24].

במקרה שלנו הוגדר הטווח כנע בין 0.25 תאונה ל 0.25- תאונות. במידה והפרש התוחלת μ יפול בטווח רווח הסמך להפרש, המשמעות היא כי המודלים בהיבט מספר התאונות הצפוי (EXPECTED) מספקים ניבוי דומה. נוסחאות מספר 81 ו - 82 מציגות את השערות המבחן.

[81] $H_0 = \mu < \theta_L \quad or \quad \mu > \theta_U$ ➔ $H_0 = \mu < \frac{1}{4} \quad or \quad \mu > \frac{1}{4}$

[82] $H_1 = \theta_L \le \mu \le \theta_U$ ➔ $H_1 = -\frac{1}{4} \le \mu \le \frac{1}{4}$

כאשר:

$\theta_L \ \theta_U$: הגבול העליון והתחתון בהתאמה (+/- רבע תאונה)

μ : הפרש התוחלת.

בטבלה מספר 29 מוצגות תוצאות המבחן. במידה והערכים המופיעים בעמודת "גבולות רווח הסמך" (**90% CL Mean**) נכללים בין הטווח העליון (**Upper Bound**) לטווח התחתון (**Lower Bound**) , אזי נקבל את H_0 ונדחה את H_1 .

[24] מבחן T לשתי אוכלוסיות תלויות אינו מספיק לבחינה מעין זו, מכיוון שבמבחן T, מניחים כי הנתונים מתפלגים נורמאלית (בניגוד לאירועים אקראיים, דוגמת תאונות דרכים), ובנוסף, מעוניינים להוכיח אי שיויון , כאשר פה, אנו מעוניינים להוכיח דווקא שיוויון.

טבלה 27 - מבחן TOST – מקטעי דרך 500 מ'

Mean	Lower Bound		90% CL Mean			Upper Bound	Assessment
-0.0548	-0.25	<	-0.0832	-0.0264	<	0.25	Equivalent

המשמעות היא כי התוצאות שוות דיין, כלומר מספר התאונות הצפוי, כפי שחושב משני סטי הנתונים (1/3 ו- 2/3) קרוב דיו.

6.4.3.4 השוואת פרופורציות ההסתברות לאפס תאונות

במודל שניבנה, ניתן לחשב ערכים מנובאים עבור ההסתברות הממוצעת לאפס תאונות, לתאונה אחת, שתי תאונות וכן הלאה. במידה והמודל שנבנה הינו מודל טוב, אזי, אם נשווה בין אחוז המקטעים בהם התקיימו "אפס תאונות" בפועל, לבין ההסתברות הממוצעת להתרחשות של "אפס תאונות" במודל המחושב, נצפה כי הפרופורציות תהינה דומות.

טבלה מספר 30 מציגה את ההסתברויות להתרחשות N תאונות במקטעי הדרך של 500 מטרים. כל שורה בטבלה מייצגת מקטע דרך ספציפי באורך מוגדר (כאמור, 500 מטרים בדוגמא שלעיל) כאשר בעמודות השונות מפורטות ההסתברות להתרחשות N תאונות באותו מקטע דרך. כך, בעמודת po0 ההסתברות להתרחשות של 0 תאונות, בעמודת po1 ההסתברות להתרחשות של תאונה אחת וכן הלאה.

טבלה 28 - ההסתברות להתרחשות N תאונות במקטעי הדרך באורך 500 מטרים

po0	po1	po2	po3	po4	po5	po6	po7	po8
0.6102773378	0.2021285764	0.0917807833	0.0454338563	0.0234212922	0.0123615103	0.0066255263	0.0035899153	0.0019608
0.602394973	0.2027937693	0.0935947405	0.0470925361	0.0246749411	0.013237001	0.0072112615	0.0039714387	0.0022048
0.6061459777	0.2024918782	0.0927388184	0.0463040849	0.0240757855	0.0128165475	0.0069286689	0.0037865486	0.0020861
0.622027764	0.2009173652	0.0889710565	0.0429520422	0.0215934632	0.0111145072	0.0058096008	0.0030698513	0.001635
0.6145998758	0.2017137307	0.090761575	0.0445217661	0.0227429164	0.0118945867	0.0063174341	0.0033919314	0.001835
0.6086037628	0.2022796466	0.0921707895	0.0457865823	0.0236857182	0.0125448166	0.0067473035	0.0036686911	0.0020109
0.5962728167	0.2032296575	0.0949626287	0.0483751326	0.0256622671	0.0139378986	0.0076875464	0.0042864047	0.0024093
0.6201240354	0.2011315691	0.0894345832	0.0433545348	0.0218860298	0.0113117258	0.0059371623	0.0031502422	0.0016850
0.6220480003	0.2009150508	0.0889661124	0.0429477634	0.021590361	0.0111124209	0.0058082544	0.0030690047	0.001634
0.6211552797	0.2010164018	0.089183883	0.0431365152	0.0217273695	0.0112046597	0.0058678413	0.0031065119	0.0016579
0.6095072185	0.2021987541	0.0919605723	0.0455962004	0.023542851	0.0124456872	0.0066813908	0.0036260171	0.0019837
0.610902529	0.2020707764	0.0916344264	0.0453020254	0.0233227622	0.0122933933	0.0065803916	0.0035607926	0.0019424

טבלה מספר 31 מציגה בעמודת ה MEAN את ממוצע עמודות po0 ו po1 (ההסתברות הממוצעת ל 0 תאונות,ההסתברות הממוצעת לתאונה אחת וכן הלאה).

טבלה 29 - ממוצע ההסתברויות להתרחשות 0 תאונות ותאונה אחת במקטע דרך באורך 500 מ'

Variable	N	Mean	Std Dev	Minimum	Maximum
po0	2188	**0.6582303**	0.1082179	0.3312983	0.8639541
po1	2188	**0.1837504**	0.0245488	0.1115330	0.2043939

כך לדוגמא, ממוצע הערכים (ההסתברויות להתרחשות תאונה אחת) בעמודת po1, עומד על 0.1837 (משמע כ- 18.37%) . ערך זה קרוב לערך 18.14% המופיע תחת עמודת AccCNT בטבלה מספר 32, המציגה את אחוז המקטעים האמיתיים, בהם התרחשו 0 תאונות, תאונה אחת, שתי תאונות, וכן הלאה.

טבלה 30 - פרופורציות המקטעים האמיתיים בהם התרחשו N תאונות

AccCNT				
AccCNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1441	65.86	1441	65.86
1	**397**	**18.14**	**1838**	**84.00**
2	170	7.77	2008	91.77
3	87	3.98	2095	95.75
4	37	1.69	2132	97.44
5	25	1.14	2157	98.58
6	12	0.55	2169	99.13
7	5	0.23	2174	99.36
8	2	0.09	2176	99.45
9	4	0.18	2180	99.63
10	3	0.14	2183	99.77
13	1	0.05	2184	99.82
14	1	0.05	2185	99.86
18	1	0.05	2186	99.91
23	1	0.05	2187	99.95
25	1	0.05	2188	100.00

דוגמא נוספת הינה ההסתברות להתרחשות 0 תאונות. ממוצע הערכים של ההסתברות להתרחשות תאונה כפי שעלה מהמודל (טבלה 31), עומד על 65.82%, כאשר במציאות (מטבלה 32) ניתן לראות כי הערך עומד על ערך קרוב מאוד – 65.86%.

6.4.3.5 בניית מודל חדש מלא מנתוני הביקורת (1/3 הנתונים), והשוואתו למודל האימון (2/3 הנתונים)

מ- 1/3 הנתונים (נתוני הביקורת) נבנה מודל חדש ומלא, שבמסגרתו נבחנו כלל הפרמטרים הקיימים בבסיס הנתונים. תוצרי המודל מוצגים בטבלאות 33 ו- 34.

בחינת מודל זה בהשוואה למודל שנבנה מהחלק הארי של הנתונים (2/3), מהווה בקרה נוספת על מודל האימון (2/3) בכמה מישורים. ראשית, אימות נוסף עבור סוג המודל, בהיבט פיזור הנתונים. שנית, בקרה על הפרמטרים שהתקבלו כמשמעותיים במודל, וכמו כן בקרה על האומדנים שהתקבלו במודל.

טבלה 31 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 500 מ') – טבלת Analysis Of Maximum Likelihood Parameter Estimates

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.7780	0.9234	-9.5878	-5.9682	-9.6087	-5.9821	70.95	<.0001
AADT_LOG	1	0.8019	0.0937	0.6183	0.9856	0.6197	0.9878	73.23	<.0001
EZOR1	1	-0.7366	0.1629	-1.0559	-0.4174	-1.0584	-0.4186	20.45	<.0001
EZOR3	1	-0.8810	0.1249	-1.1258	-0.6362	-1.1274	-0.6369	49.75	<.0001
CURVEPRM3	1	0.9894	0.5845	-0.1563	2.1351	-0.1523	2.1489	2.86	0.0905
Dispersion	1	1.5877	0.1764	1.2420	1.9334	1.2701	1.9650		

טבלה 32 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 500 מ') טבלת – Contrast Estimate Results

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
adj. rr: CurvePrm3	2.6896	0.8553	8.4577	0.9894	0.5845	0.05	-0.1563	2.1351	2.86	0.0905
Exp(adj. rr: CurvePrm3)				2.6896	1.5722	0.05	0.8553	8.4577		
adj. rr: AADT_log	2.2298	1.8557	2.6794	0.8019	0.0937	0.05	0.6183	0.9856	73.23	<.0001
Exp(adj. rr: AADT_log)				2.2298	0.2090	0.05	1.8557	2.6794		
adj. rr: ezor1 vs 2	0.4787	0.3479	0.6588	-0.7366	0.1629	0.05	-1.0559	-0.4174	20.45	<.0001
Exp(adj. rr: ezor1 vs 2)				0.4787	0.0780	0.05	0.3479	0.6588		
adj. rr: ezor3 vs 2	0.4144	0.3244	0.5293	-0.8810	0.1249	0.05	-1.1258	-0.6362	49.75	<.0001
Exp(adj. rr: ezor3 vs 2)				0.4144	0.0518	0.05	0.3244	0.5293		

ראשית נבחן את סוג הפיזור. כפי שניתן לראות בטבלה מספר 33 הערך 1, אינו נופל ברווח הסמך (הנע בין 1.2701 ל 1.9650) ברמת מובהקות 95%. לפיכך, ניתן להסיק כי קיים פיזור יתר בהתפלגות של מספר תאונות הדרכים, ולפיכך **גם פה, המודל המתאים**

יותר הינו מסוג בינומי שלילי. **ממצא זה תואם** את ממצאי שני המודלים הקודמים (מודל האימון של 2/3 הנתונים ומודל הביקורת שנבנה מ 1/3 הנתונים ובו שולבו הפרמטרים שנמצאו מתאימים עבור מודל האימון 2/3 הנתונים).

נבחן את **הפרמטרים** שיצאו כמשמעותיים במודל הביקורת המלא הנוכחי: עקמומיות - **CURVEPRM3**, אזור - **EZOR1** ונפח תנועה - **AADT_LOG**. פרמטרים אלה, **הינם אותם** הפרמטרים שנמצאו משמעותיים במודל האימון (2/3 הנתונים) ובמודל הביקורת הראשוני (1/3 הנתונים). לממצא זה חשיבות רבה בהיבט יציבות המודל. בבואנו לבחון את האומדים של מודל הביקורת המלא אל מול רווחי הסמך של מודל האימון (2/3 הנתונים) ברמת ביטחון של 95%, נמצא כי הפרמטרים **AADT_LOG , EZOR1 , CURVEPRM3** – נמצאים בטווח רווח הסמך של מודל האימון (2/3), כאשר המשתנה **EZOR3** הינו היחידי שלא נפל בטווח רווח הסמך. משתנה ה CURVEPRM3 המייצג עקמומיות אינו מובהק, ושולב במודל בגין ערך ה - AIC שהתקבל כאשר שולב במודל.

6.4.4 Empirical Bayes (להלן: EB)

כאמור, בתום ביצוע החיזוי במודלים השונים, בוצע הליך ה- EB. הליך זה, כפי שפורט בהרחבה בסעיף 2.3.3.4, משמש ל"תיקון" מספר התאונות הצפוי, תוך שהוא מביא בחשבון את מספר התאונות האמיתי שהתרחש באותו מקטע הדרך.

ההליך בוצע לפי הנוסחה המופיעה במאמרם של Sayed & Rodriguez [1999] ו- Ng et-al [2002], והמוצגת להלן:

$$\left[\frac{\lambda}{\frac{1}{\alpha}+\lambda}\right]\times\left(\frac{1}{\alpha}+y\right) \qquad [83]$$

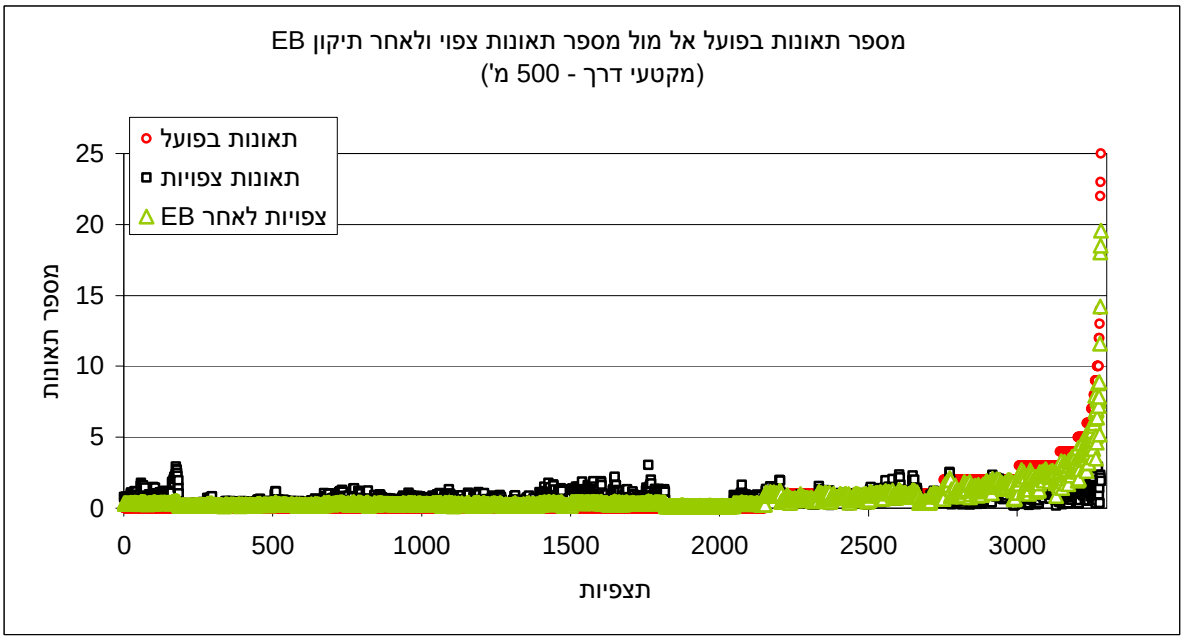
כאשר:

α - פרמטר ה overdispersion , כפי שחושב עבור סט נתונים באורך מסויים

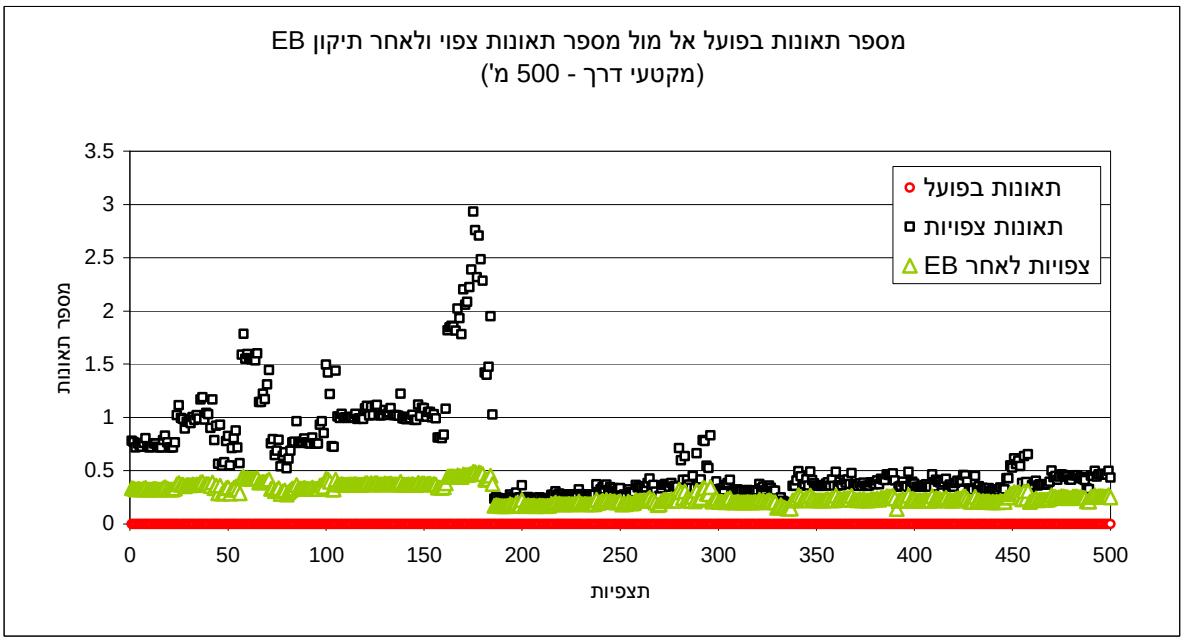
λ - מספר תאונות בפועל

y- מספר תאונות צפוי

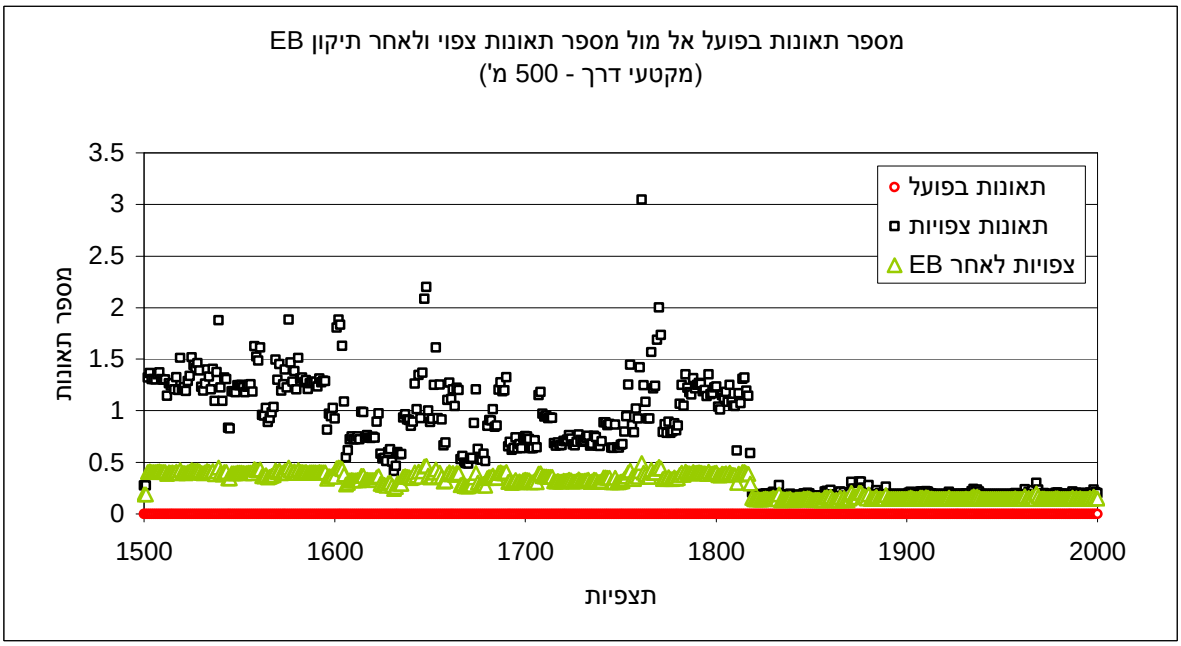
איור מספר 65 מציג את תוצאות הליך ה- EB עבור מקטעי דרך באורך של 500 . איורים מספר 66-68 מציגים בהרחבה את תוצאות המקור, בפילוח מספר תצפיות שונה (על מנת שנוכל להבחין טוב יותר במגמות) .



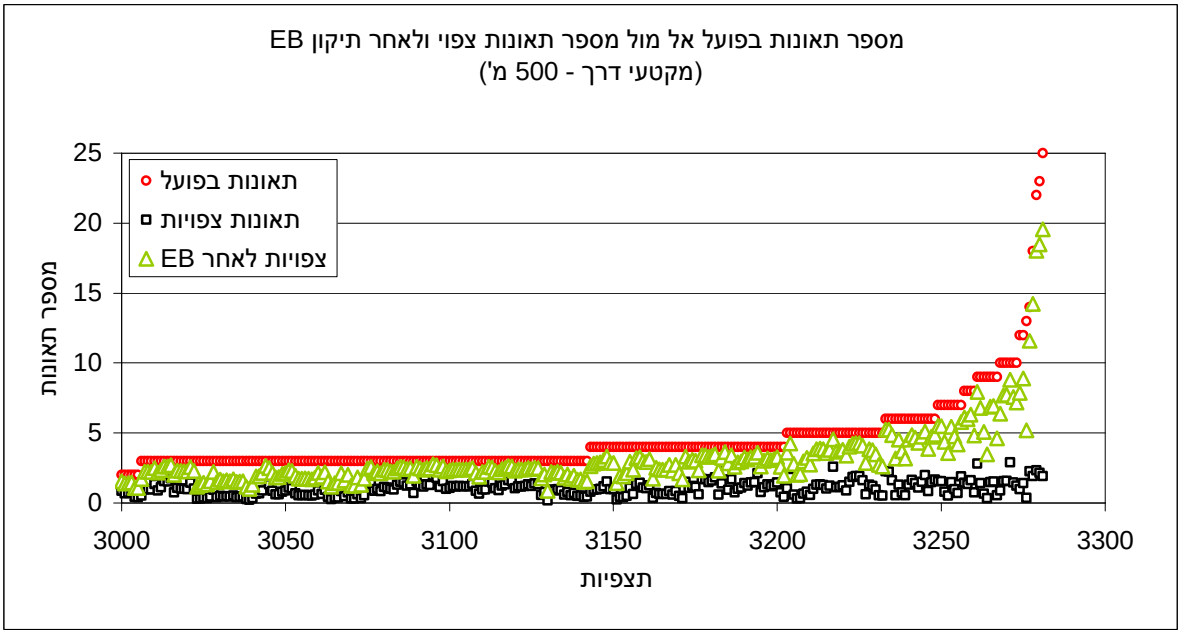
איור מספר 65 - **מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 500 מ'**



איור מספר 66 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 500 מ' (ממספר תצפית 0 עד מספר תצפית 500)



איור מספר 67 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 500 מ' (ממספר תצפית 1500 עד מספר תצפית 2000)



איור מספר 68 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 500 מ' (ממספר תצפית 3000 עד מספר תצפית 3300)

7. ממצאים

7.1 מקטעי דרך באורך 500 מטרים

הניתוח הסטטיסטי עבור מקטעי דרך באורך זה הופיע בפיסקה הקודמת, כהדגמה לתהליך העבודה.

7.2 מקטעי דרך באורך 750 מטרים

7.2.1 בחירת המודל

טבלה מספר 35 מציגה את תוצר ה AIC עבור מקטעי דרך באורך של 750 מטרים. המודל שיצא הטוב ביותר עבור סט הנתונים של 750 מטרים – הינו מסוג Negative Binomial.

טבלה 33 - תוצר AIC עבור מקטעי דרך באורך 750 מ'

AIC	750
פואסוני	3937
בינומי-שלילי	**3267**
Zero-Inflated פואסון	3470
Zero-Inflated בינומי-שלילי	3271

7.2.2 פרוצדורת GENMOD

7.2.2.1 ניתוח טבלת Analysis Of Maximum Likelihood Parameter Estimates

טבלה מספר 36, מציגה את תוצאות טבלת **Analysis Of Maximum Likelihood Parameter Estimates**.

טבלה 34 - Analysis Of Maximum Likelihood Parameter Estimates 750, מודל האימון 2/3 מהנתונים

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.9688	0.7698	-7.4776	-4.4600	-7.4918	-4.4692	60.12	<.0001
aadt_log	1	0.6420	0.0778	0.4895	0.7946	0.4905	0.7961	68.04	<.0001
ezor1	1	-0.7316	0.1386	-1.0034	-0.4599	-1.0045	-0.4602	27.85	<.0001
ezor3	1	-0.5483	0.1076	-0.7592	-0.3374	-0.7601	-0.3375	25.96	<.0001
CurvePrm3	1	1.0741	0.4636	0.1655	1.9827	0.1725	1.9944	5.37	0.0205
Dispersion	1	1.6071	0.1401	1.3324	1.8817	1.3511	1.9024		

הפרמטרים המשמעותיים במודל, דומים לאלה שהתקבלו ככאלה בניתוח מקטעי הדרך באורך של 500 מטרים.

- **aadt_log** - ממוצע נפח תנועה יומי לימי חול, לאחר שעבר טרנספורמציה ללוגריתם טבעי.
- **ezor1** (דרום), **ezor3** (מרכז) – מייצגים את משתנה האזור. כאשר המודל משווה ומנתח בין אזור 2 (צפון) לשני האזורים האחרים.
- **CurvePrm3** - עקמומיות

בעזרת מבחן יחס הנראות, נבחן האם פרמטר פיזור היתרה שונה באופן מובהק מ-1. כאמור, בכדי לבצע זאת, ייבדק עד כמה מקדם המשתנה (Estimate) של פרמטר ה Dispersion, שערכו במודל הנוכחי שלנו 1.6071, קרוב מספיק או רחוק מספיק מ- 1. לשם כך יש לבחון את רווח הסמך המופיע בטבלת **Analysis Of Maximum Likelihood Parameter Estimates** בעמודת **"Likelihood Ratio 95% Confidence Limits"**. כפי שניתן לראות בטבלה מספר 36, הערך 1 אינו נופל ברווח הסמך (הנע בין 1.3511 ל 1.9024) ברמת מובהקות 95%, לאור האמור ניתן להסיק כי קיים פיזור יתר בהתפלגות של מספר תאונות הדרכים, ולפיכך, גם בסט נתונים זה, **המודל המתאים יותר הינו מסוג בינומי שלילי**.

בנוסף, בהתאם לטבלה 36, נוכל לבנות עבור מודל האימון של "2/3 הנתונים", את **הנוסחה לניבוי תוחלת תאונות הדרכים μ** (=ממוצע תאונות למקטע דרך לתקופת המחקר העומדת על שלוש שנים). הערכים המנובאים התקבלו דרך הקשר הלוגריתמי. בהתאם לנוסחה מספר 84:

[84]

$$\mu_{2/3} = e^{-5.9688+0.6420*AADT\_Log+(-0.7316)*Ezor1+(-0.5483)*Ezor3+(1.0741)*CurvePrm3}$$

כמו כן, ניתן להבחין כי קיימת התאמה לממצאי הניסויים שבוצעו על מקטעי הדרך באורך של **500 מטרים** בנוגע לפרמטר ה"אזור". המינוס במשתנים **ezor1** , **ezor3**. משמעו, שבאזור 2 (צפון) מתרחשות יותר תאונות מאשר באזור 1 (דרום) ואזור 3 (מרכז).

7.2.2.2 ניתוח טבלת Contrast Estimate Results

טבלה מספר 37 הינה טבלת **Contrast Estimate Results** עבור מקטעי דרך באורך 750 מ'.

טבלה 35 - Contrast Estimate Results, מודל האימון (2/3 מהנתונים), 750

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
adj. rr: CurvePrm3	2.9273	1.1800	7.2622	1.0741	0.4636	0.05	0.1655	1.9827	5.37	0.0205
Exp(adj. rr: CurvePrm3)				2.9273	1.3570	0.05	1.1800	7.2622		
adj. rr: AADT_log	1.9003	1.6314	2.2135	0.6420	0.0778	0.05	0.4895	0.7946	68.04	<.0001
Exp(adj. rr: AADT_log)				1.9003	0.1479	0.05	1.6314	2.2135		
adj. rr: ezor1 vs 2	0.4811	0.3666	0.6313	-0.7316	0.1386	0.05	-1.0034	-0.4599	27.85	<.0001
Exp(adj. rr: ezor1 vs 2)				0.4811	0.0667	0.05	0.3666	0.6313		
adj. rr: ezor3 vs 2	0.5779	0.4680	0.7136	-0.5483	0.1076	0.05	-0.7592	-0.3374	25.96	<.0001
Exp(adj. rr: ezor3 vs 2)				0.5779	0.0622	0.05	0.4680	0.7136		

מטבלה מספר 37, עולה כי באזור הצפון (2) שיעור התאונות הצפוי גדול **פי** 2.07 משיעור התאונות הצפוי בדרום (אזור 1).
בנוסף, באזור הצפון (2) שיעור התאונות הצפוי גדול פי 1.73 משיעור התאונות הצפוי באזור המרכז (3).

AADT_log – AAAD מייצג כאמור, ממוצע נפח תנועה יומי לימי חול (מספר כלי הרכב שנעו בקטע דרך מסויים בימי חול). עבור עליה ביחידה אחת של המשתנה AADT_log (שמשמעה נפח הגדול פי 2.718 מהערך הקודם), שיעור התאונות עתיד לגדול פי 1.90.

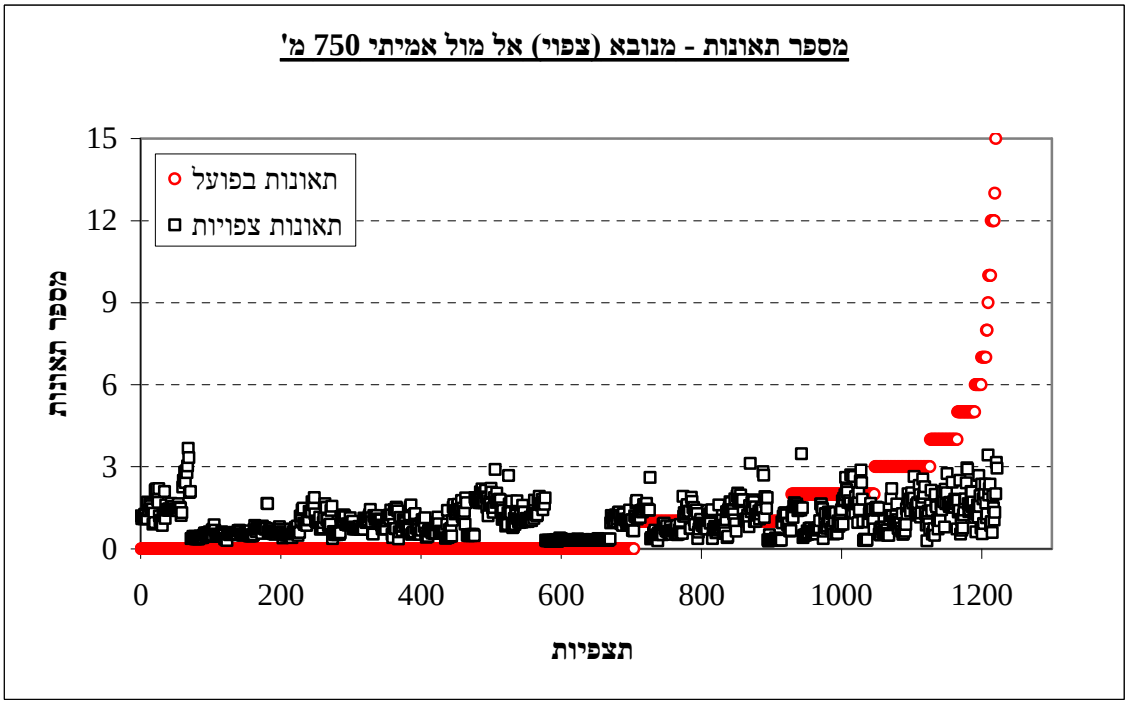
כך לדוגמא, עבור מקטע דרך שנפח התנועה עליו 1000 מכוניות, כאשר נפח התנועה יגדל ביחידה אחת (דהיינו פי 2.7) ויגיע לנפח של 2718 מכוניות, שיעור התאונות יגדל פי 1.9, כלומר בדוגמא שלנו מספר התאונות צפוי להיות : 5.7.

CurvePrm3 – יחידות העקמומיות נעות בין 0 ל- 1. ככל שהערך גבוה יותר, כך העקמומיות בקטע הדרך גבוהה יותר. בניתוח האמור, כאשר נעלה ביחידה אחת של עקמומיות, מספר התאונות הצפוי במקטע דרך יגדל כמעט פי 3 (2.92) .

7.2.3 CROSS VALIDATION של המודל

7.2.3.1 מספר תאונות צפוי אל מול גרף התאונות אמיתי

איור מספר 69 מציג את הגרף המשווה בין מספר התאונות הצפוי (EXPECTED) לפי המודל לבין מספר התאונות בפועל (OBSERVED) בקטעי הדרך באורך 750 מ'. באדום מוצג מספר תאונות האמיתי ובשחור מוצג מספר התאונות הצפוי.



איור מספר 69 - מספר תאונות צפוי אל מול מספר תאונות אמיתי – מקטעי דרך באורך 750 מ'

ניתן לראות כי המתאם אינו מספיק טוב, ולפיכך נדרש להליך ה EB. הליך זה פורט, כזכור בהרחבה בסקר הספרות בסעיף 2.3.3.4, ובפסקה 6.4.4.

7.2.3.2 השוואת מודל האימון (2/3 הנתונים), למודל שנבנה מנתוני קבוצת הביקורת (1/3 הנתונים).

אותו ההליך שתואר לעיל ובוצע עבור מודל האימון (2/3 מהנתונים), בוצע גם על מסד נתוני הביקורת (1/3 הנתונים) של מקטעי הדרך באורך 750 מ'. כאמור, מראש שולבו במודל זה הפרמטרים המשמעותיים, כפי שעלו ממודל האימון של מקטעי הדרך באורך 750 מ'. טבלה מספר 38 מציגה את המקדמים עבור מודל קבוצת הביקורת.

טבלה 36 - תוצר פרוצדורת GENMOD עבור נתוני קבוצת הביקורת של 1/3 הנתונים (ניתוח סגמנטים באורך 750 מ')

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-8.9175	1.0530	-10.9813	-6.8536	-11.0092	-6.8700	71.72	<.0001
aadt_log	1	0.9400	0.1063	0.7317	1.1483	**0.7335**	**1.1513**	78.22	<.0001
ezor1	1	-0.4848	0.1985	-0.8739	-0.0957	**-0.8759**	**-0.0954**	5.96	0.0146
ezor3	1	-0.7456	0.1592	-1.0576	-0.4336	**-1.0607**	**-0.4346**	21.93	<.0001
CurvePrm3	1	1.8142	0.5827	0.6722	2.9562	**0.6918**	**2.9871**	9.70	0.0018
Dispersion	1	1.5757	0.1912	1.2009	1.9504	**1.2365**	**1.9913**		

בהתאם לטבלה מספר 38 , נבנתה נוסחה מספר 85, לחישוב ערכי תאונות דרכים מנובאים:

$$\mu_{1/3} = e^{-8.917+0.940*AADT\_Log+(-0.484)*Ezor1+(-0.745)*Ezor3+(1.814)*Curve\,Prm3} \quad [85]$$

כאמור, לאחר שהתקבלו אומדנים משני המודלים, יש לבחון האם האומדנים שהתקבלו מהמודל שנבנה מנתוני הביקורת (1/3) אינם שונים באופן מובהק ממקדמי מודל האימון (2/3 הנתונים). זאת, על ידי בחינה הבודקת האם מקדמי אומדני מודל הביקורת נופלים בטווח רווח הסמך של מקדמי מודל האימון, ברמת ביטחון 95% .

כזכור, לשם כך, יש להשוות את מקדמי המשתנים המסבירים מנתוני מודל הביקורת המופיעים בעמודת **Estimate** בטבלת **Analysis Of Maximum Likelihood Parameter Estimates**(טבלה 38), אל נתוני טווח רווח הסמך של נתוני מודל האימון המופיעים בעמודת **Likelihood Ratio 95% Confidence Limits** (טבלה 36).כך לדוגמא, ערך מקדם המשתנה נפח תנועה (AADT_LOG) הינו 0.9400. יש לבדוק האם מספר זה נופל בתחום רווח הסמך של פונקציית ה Likelihood של נתוני מודל האימון (2/3). ניתן לראות כי מספר זה אינו נכלל בטווח שבין 0.4905 לבין 0.7961 (טבלה מספר 31). המשמעות היא כי המודל אינו טוב מספיק עבור משתנה זה, כאשר ההבדל בשיעור תאונות הדרכים הקשור בנפחי התנועה, גבוה יותר בסט נתוני הביקורת (1/3 הנתונים). מהמשך השוואת טבלאות 38 ו- 36 ניתן לראות כי עבור המשתנים האחרים ששולבו במודל, ערכי המקדמים אכן נמצאים בטווח רווח הסמך.

7.2.3.3 מבחן TOST לשוויון (TOST Equivalence Test)

בטבלה מספר 39 מוצגות תוצאות המבחן עבור מקטעי דרך של 750 מ'.

טבלה 37 - תוצאות מבחן TOST , סגמנטים באורך 750 מ'

Mean	Lower Bound		90% CL Mean			Upper Bound	Assessment
0.0520	-0.25	<	0.0335	0.0706	<	0.25	Equivalent

הערך המופיע בעמודת "גבולות רווח הסמך" (90% CL Mean), אכן נכלל בין הטווח העליון (Upper Bound) לטווח התחתון (Lower Bound). המשמעות היא כי נקבל את H_0 ונדחה את H_1. משמע התוצאות שוות דיין, ומספר התאונות הצפוי, כפי שחושב משני מערכי הנתונים ("1/3" ו- "2/3") קרוב דיו.

7.2.3.4 השוואת יחסי ההסתברות לאפס תאונות

כאמור, במודל שניבנה, ניתן לחשב ניבוי עבור ההסתברות הממוצעת לאפס תאונות, לתאונה אחת, שתי תאונות וכן הלאה. במידה והמודל שנבנה הינו מודל "טוב", אזי, אם נשווה בין אחוז המקטעים בהם התקיימו "אפס תאונות" בפועל, לבין ההסתברות הממוצעת להתרחשות של "אפס תאונות" במודל המחושב, נצפה כי הפרופורציות תהינה דומות.

טבלה מספר 40 מציגה בעמודת ה MEAN את ממוצע ההסתברויות המחושב להתרחשות אפס תאונות (po0) והסתברות להתרחשות תאונה אחת (po1) במקטעי הדרך של 750 מטרים.

טבלה 38 - ממוצע הסתברויות מחושב להתרחשות 0 תאונות ותאונה אחת במקטע דרך באורך 750 מ'

Variable	N	Mean	Std Dev	Minimum	Maximum
po0	1222	0.5724056	0.1161340	0.3006056	0.8020515
po1	1222	0.1982929	0.0151256	0.1489596	0.2112989

טבלה מספר 41 מציגה את אחוז המקטעים האמיתיים, בהם התרחשו 0 תאונות, תאונה אחת, שתי תאונות, וכן הלאה.

טבלה 39 - פרופורציות המקטעים האמיתיים בהם התרחשו N תאונות - 750 מ'

AccCNT				
AccCNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	704	57.61	704	57.61
1	224	18.33	928	75.94
2	119	9.74	1047	85.68
3	79	6.46	1126	92.14
4	39	3.19	1165	95.34
5	25	2.05	1190	97.38
6	9	0.74	1199	98.12
7	7	0.57	1206	98.69
8	2	0.16	1208	98.85
9	1	0.08	1209	98.94
10	4	0.33	1213	99.26
12	5	0.41	1218	99.67
13	1	0.08	1219	99.75
15	1	0.08	1220	99.84
24	1	0.08	1221	99.92
28	1	0.08	1222	100.00

בבחינת טבלה 40, המציגה כאמור את ממוצע ההסתברויות המחושב להתרחשות אפס תאונות, עולה כי עבור ההסתברויות להתרחשות תאונה אחת (במקטע דרך של 750מ') ממוצע הערכים בעמודת po1, עומד על **0.5724056** (משמע כ- 57.24%). ערך זה קרוב לערך 57.61% המופיע בטבלה מספר 41, המציגה כאמור את פרופורציית המקטעים האמיתיים בהם התרחשו אפס תאונות במקטע דרך. דוגמא נוספת היא 19.82% בערכים המחושבים עבור תאונה אחת (טבלה 40) לעומת 18.33% בפרופורציות המקטעים האמיתיים בהן התרחשה תאונה אחת (טבלה 41).

7.2.3.5 בניית מודל חדש מלא מנתוני הביקורת (עבור 1/3 הנתונים), והשוואתו אל מודל האימון (2/3 הנתונים)

מ- 1/3 הנתונים (נתוני הביקורת) נבנה מודל חדש ומלא, ובמסגרתו נבחנו כלל הפרמטרים הקיימים בבסיס הנתונים. תוצרי המודל מוצגים בטבלאות 42 ו- 43.

טבלה 40 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 750 מ') – טבלת Analysis Of Maximum Likelihood Parameter Estimates

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-8.9175	1.0530	-10.9813	-6.8536	-11.0092	-6.8700	71.72	<.0001
aadt_log	1	0.9400	0.1063	0.7317	1.1483	**0.7335**	**1.1513**	78.22	<.0001
ezor1	1	-0.4848	0.1985	-0.8739	-0.0957	**-0.8759**	**-0.0954**	5.96	0.0146
ezor3	1	-0.7456	0.1592	-1.0576	-0.4336	**-1.0607**	**-0.4346**	21.93	<.0001
CurvePrm3	1	1.8142	0.5827	0.6722	2.9562	**0.6918**	**2.9871**	9.70	0.0018
Dispersion	1	1.5757	0.1912	1.2009	1.9504	**1.2365**	**1.9913**		

טבלה 41 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 750 מ') טבלת – Contrast Estimate Results

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
adj. rr: CurvePrm3	6.1363	1.9586	19.2249	1.8142	0.5827	0.05	0.6722	2.9562	9.70	0.0018
Exp(adj. rr: CurvePrm3)				6.1363	3.5754	0.05	1.9586	19.2249		
adj. rr: AADT_log	2.5600	2.0786	3.1530	0.9400	0.1063	0.05	0.7317	1.1483	78.22	<.0001
Exp(adj. rr: AADT_log)				2.5600	0.2721	0.05	2.0786	3.1530		
adj. rr: ezor1 vs 2	0.6158	0.4173	0.9087	-0.4848	0.1985	0.05	-0.8739	-0.0957	5.96	0.0146
Exp(adj. rr: ezor1 vs 2)				0.6158	0.1222	0.05	0.4173	0.9087		
adj. rr: ezor3 vs 2	0.4744	0.3473	0.6482	-0.7456	0.1592	0.05	-1.0576	-0.4336	21.93	<.0001
Exp(adj. rr: ezor3 vs 2)				0.4744	0.0755	0.05	0.3473	0.6482		

כאמור, בחינת מודל זה בהשוואה למודל האימון שנבנה מהחלק הארי של הנתונים (2/3), מהווה בקרה נוספת על מודל האימון המקורי בשני מישורים. ראשית, יש בזה משום אימות נוסף עבור סוג המודל, בהיבט פיזור הנתונים. שנית, זו גם בקרה על הפרמטרים שיצאו משמעותיים במודל, וכמו כן בקרה על האומדנים שהתקבלו במודל.

סוג הפיזור, כפי שניתן לראות בטבלה מספר 42, הערך 1 אינו נופל ברווח הסמך (הנע בין **1.2365 ל- 1.9913**) ברמת מובהקות 95%. לפיכך, ניתן להסיק כי קיים פיזור יתרה בהתפלגות של מספר תאונות הדרכים, ולפיכך **המודל המתאים יותר הינו מסוג בינומי שלילי**. **ממצא זה תואם** את ממצאי שני המודלים הקודמים (המודל האימון, ומודל הביקורת).

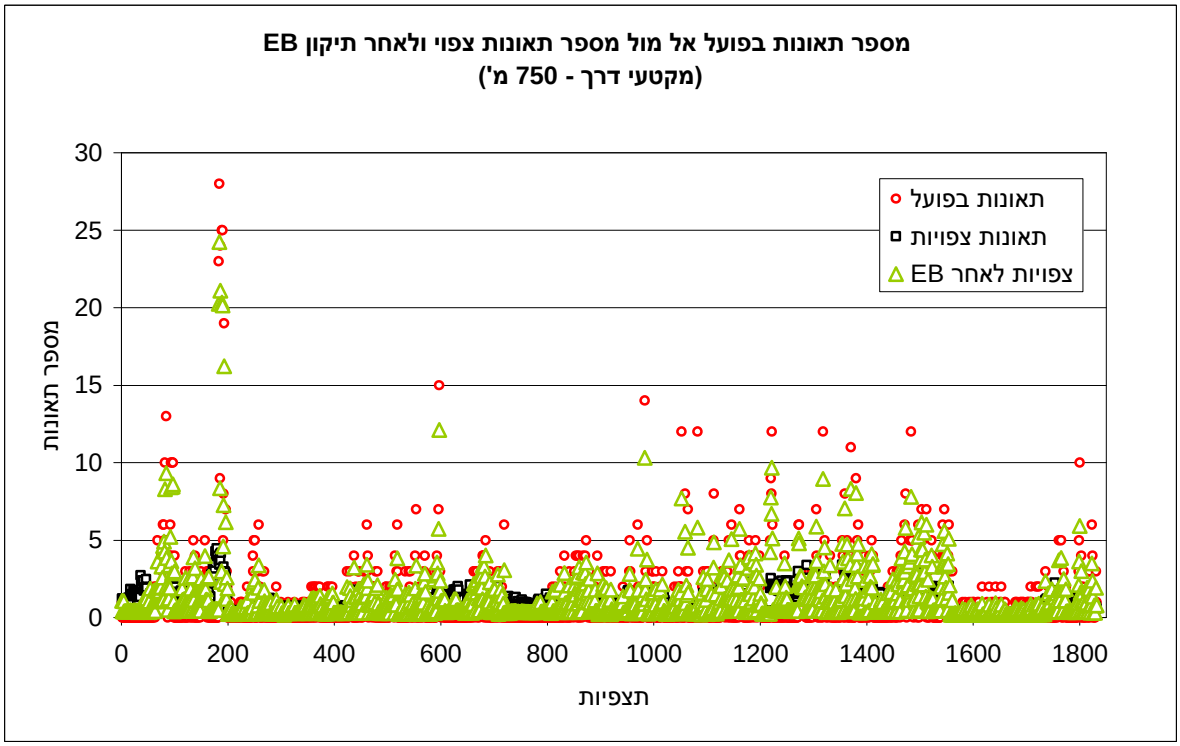
נשווה את **הפרמטרים** שנמצאו כמשמעותיים בשני מערכי הנתונים. הפרמטרים שהתקבלו משמעותיים עבור מודל זה, **הינם אותם** הפרמטרים שנמצאו משמעותיים במודל שנבנה עבור 2/3 הנתונים (עקמומיות - **CURVEPRM3** , אזור - **EZOR1** ונפח תנועה - **AADT_LOG**).

בדיקה נוספת הינה בדיקת האומדים. נבחן האם אומדני המודל החדש והמלא, שנבנה מ-1/3 הנתונים, נופלים בטווח רווח הסמך של מודל האימון (2/3), ברמת ביטחון 95%. הפרמטרים, **EZOR1** , **EZOR3** **CURVEPRM3** – אכן נפלו בטווח רווח הסמך של מודל האימון (2/3). המשתנה **AADT_LOG** הינו היחידי שלא נפל בטווח רווח הסמך.

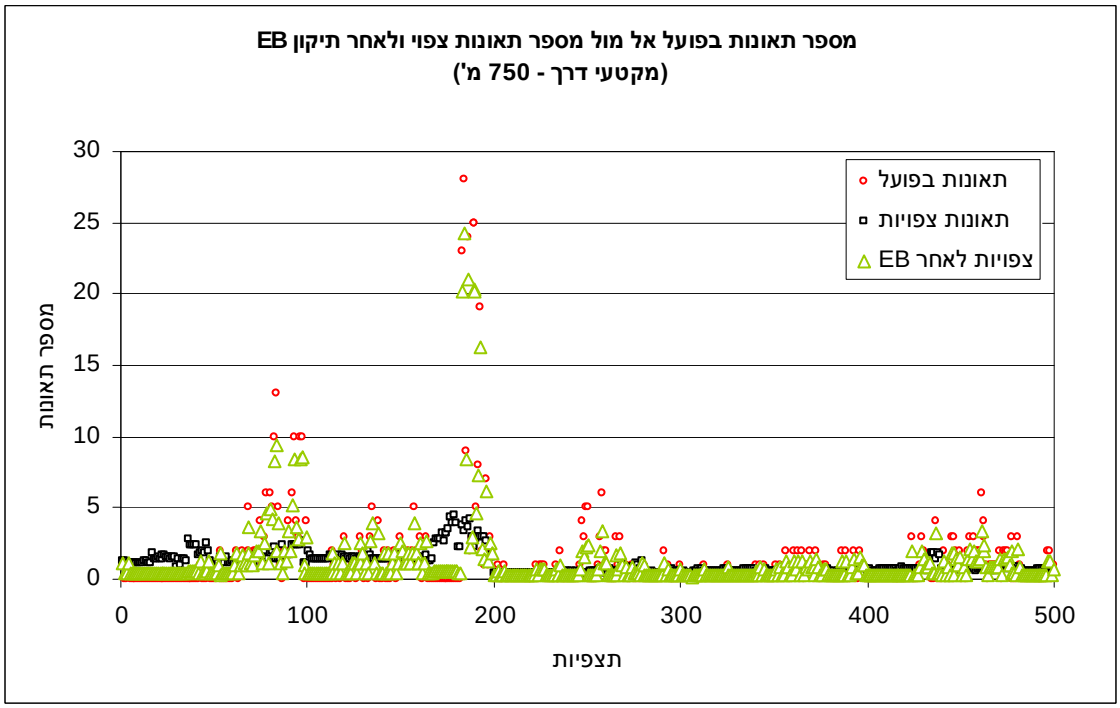
7.2.4 Empirical Bayes (להלן: EB) – מקטעי דרך באורך 750מ'.

כזכור, ההליך בוצע לפי הנוסחה המופיעה במאמרם של Sayed & Rodriguez [1999] ו- Ng et-al [2002], כפי שפורטה סעיף 6.4.4.

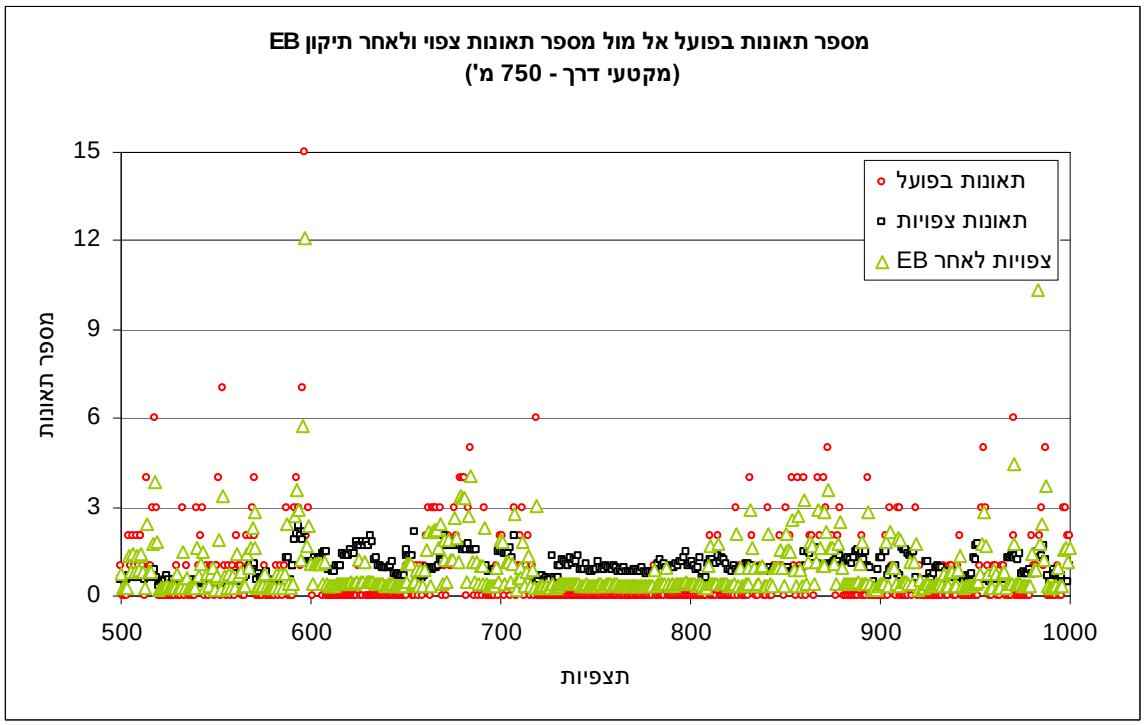
איור מספר 70 מציג את תוצאות הליך ה- EB עבור מקטעי דרך באורך של 750 . איורים 71-73 מציגים בהרחבה את תוצאות המקור, בפילוח מספר תצפיות שונה (על מנת שנוכל להבחין טוב יותר במגמות) .



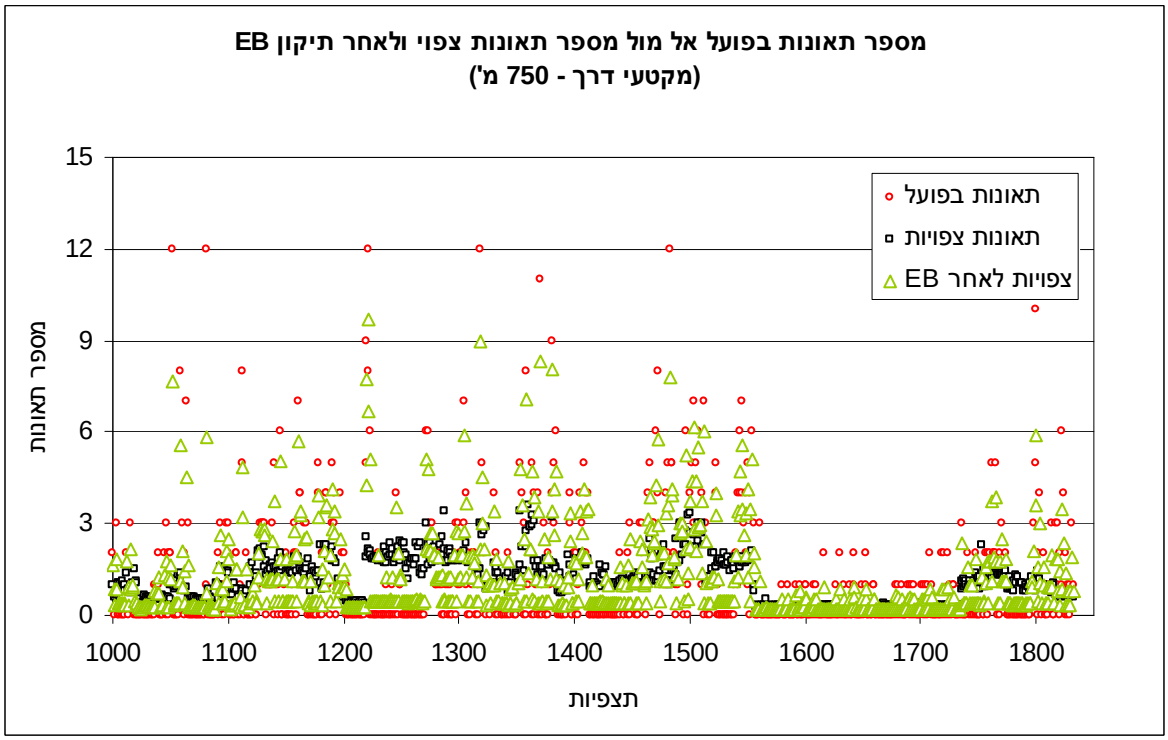
איור מספר 70 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 750 מ' (כלל התצפיות)



איור מספר 71 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 750 מ' (ממספר תצפית 0 עד מספר תצפית 500)



איור מספר 72 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 750 מ' (ממספר תצפית 500 עד תצפית 1000)



איור מספר 73 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 750 מ' (ממספר תצפית 1000 עד מספר תצפית 1800)

7.3 מקטעי דרך באורך 1000 מטרים

7.3.1 בחירת המשתנים עבור המודל

טבלה מספר 44 מציגה את תוצר ה- AIC עבור מקטעי דרך באורך של 1000 מטרים. המודל שהתקבל כטוב ביותר עבור סט הנתונים באורך זה – הינו מסוג Negative Binomial.

טבלה 42 - תוצר AIC עבור מקטעי דרך באורך 1000 מ'

AIC	1000
פואסוני	2964
בינומי-שלילי	**2430**
Zero-Inflated פואסון	2521
Zero-Inflated בינומי-שלילי	2431

7.3.2 פרוצדורת GENMOD

7.3.2.1 ניתוח טבלת Analysis Of Maximum Likelihood Parameter Estimates

טבלה מספר 45 מציגה את תוצר טבלת **Analysis Of Maximum Likelihood Parameter Estimates** עבור מקטעי דרך באורך 1000מ'.

טבלה 43 - Analysis Of Maximum Likelihood Parameter Estimates 1000, מודל האימון 2/3 מהנתונים

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.7103	0.8380	-7.3527	-4.0678	**-7.3695**	**-4.0773**	46.43	<.0001
aadt_log	1	0.6450	0.0848	0.4787	0.8112	**0.4798**	**0.8130**	57.82	<.0001
ezor1	1	-0.6049	0.1483	-0.8955	-0.3143	**-0.8965**	**-0.3141**	16.64	<.0001
ezor3	1	-0.6027	0.1237	-0.8452	-0.3601	**-0.8465**	**-0.3603**	23.72	<.0001
CurvePrm3	1	0.9208	0.4540	0.0309	1.8107	**0.0368**	**1.8227**	4.11	0.0426
Dispersion	1	1.3564	0.1348	1.0921	1.6206	**1.1126**	**1.6437**		

הפרמטרים המשמעותיים במודל, דומים לאלה שהתקבלו ככאלה, בניתוח מקטעי הדרך באורך של 500 ו- 750 מטרים:

- **aadt_log** - ממוצע נפח תנועה יומי לימי חול, לאחר שעבר טרנספורמציה ללוגריתם טבעי.
- **ezor1** (דרום), **ezor3** (מרכז) – מייצגים את משתנה האזור. כאשר המודל משווה ומנתח בין אזור 2 (צפון) לשני האזורים האחרים.
- **CurvePrm3 - עקמומיות**

בעזרת מבחן יחס הנראות, יש לבחון האם פרמטר פיזור היתרה שונה באופן מובהק מ-1. כאמור, בכדי לבצע זאת, יש לבדוק עד כמה מקדם המשתנה (Estimate) של פרמטר ה Dispersion, שערכו במודל הנוכחי שלנו 1.3564 קרוב מספיק או רחוק מספיק מ- 1. לשם כך ייבחן רווח הסמך המופיע בטבלת **Analysis Of Maximum Likelihood Parameter Estimates** בעמודת **"Likelihood Ratio 95% Confidence Limits"**. כפי שניתן לראות בטבלה מספר 45 הערך 1, אינו נופל ברווח הסמך (הנע בין 1.1126 ל **1.6437**) ברמת מובהקות 95%. מכאן, שיש להסיק כי קיים פיזור יתר בהתפלגות של מספר תאונות הדרכים. לפיכך, גם בסט נתונים זה של מקטעי דרך באורך של 1000 מ', **המודל המתאים יותר הינו מסוג בינומי שלילי**.

בנוסף, בהתאם לטבלה זו, ניתן לבנות עבור מודל האימון (2/3 הנתונים) את **הנוסחה לניבוי תוחלת תאונות הדרכים μ** (=ממוצע תאונות למקטע דרך לתקופת המחקר העומדת על שלוש שנים). הערכים המנובאים התקבלו דרך הקשר הלוגריתמי. בהתאם לנוסחה מספר 86:

[86]

$$\mu_{2/3} = e^{-5.7103+0.6450*AADT\_Log+(-0.6049)*Ezor1+(-0.6027)*Ezor3+(0.9208)*CurvePrm3}$$

בנוסף, קיימת התאמה לממצאי מקטעי הדרך באורכים של 500 מטרים ו- 750 מטרים בנוגע לפרמטר ה"אזור". המינוס במשתנים **ezor1** , **ezor3**. משמעו, שבאזור 2 (צפון) צפויות יותר תאונות מאשר באזור 1 (דרום). כמו כן, באזור 2 (צפון) גם צפויות גם יותר תאונות מאשר באזור 3 (מרכז).

7.3.2.2 ניתוח טבלת Contrast Estimate Results

טבלה מספר 46 מציגה את תוצר טבלת Contrast Estimate Results עבור מקטעי של של 1000 מ'.

טבלה 44 - טבלת Contrast Estimate Results**, מקטעי דרך 1000 מ'.**

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > Chi Sq
adj. rr: CurvePrm3	2.5113	1.0314	6.1145	0.9208	0.4540	0.05	0.0309	1.8107	4.11	0.0426
Exp(adj. rr: CurvePrm3)				2.5113	1.1402	0.05	1.0314	6.1145		
adj. rr: AADT_log	1.9059	1.6140	2.2506	0.6450	0.0848	0.05	0.4787	0.8112	57.82	<.0001
Exp(adj. rr: AADT_log)				1.9059	0.1617	0.05	1.6140	2.2506		
adj. rr: ezor1 vs 2	0.5461	0.4084	0.7303	-0.6049	0.1483	0.05	-0.8955	-0.3143	16.64	<.0001
Exp(adj. rr: ezor1 vs 2)				0.5461	0.0810	0.05	0.4084	0.7303		
adj. rr: ezor3 vs 2	0.5473	0.4295	0.6976	-0.6027	0.1237	0.05	-0.8452	-0.3601	23.72	<.0001
Exp(adj. rr: ezor3 vs 2)				0.5473	0.0677	0.05	0.4295	0.6976		

מטבלה 46 עולה כי באזור 1 (דרום) שיעור התאונות הצפוי גבוה פי 0.5461 משיעור התאונות הצפוי באזור 2 (צפון). שוב, במילים אחרות, ניתן לומר כי באזור הצפון (2) שיעור התאונות הצפוי גדול **פי** 1.83 משיעור התאונות הצפוי בדרום (אזור 1).

כמו כן, נמצא כי שיעור התאונות הצפוי באזור 3 (מרכז) גבוה פי 0.5473 משיעור התאונות הצפוי באזור 2 (צפון). גם במקרה זה ניתן לומר, כי באזור הצפון (2) שיעור התאונות הצפוי גדול פי 1.82 משיעור התאונות הצפוי באזור המרכז (3).

AADT_log – AAAD מייצג כאמור, ממוצע נפח תנועה יומי לימי חול (מספר כלי הרכב שנעו בקטע דרך מסויים בימי חול). עבור עליה ביחידה אחת של המשתנה AADT_log (שמשמעה נפח הגדול פי 2.7182 מהערך הקודם), שיעור התאונות יגדל כמעט פי 2 (1.90).

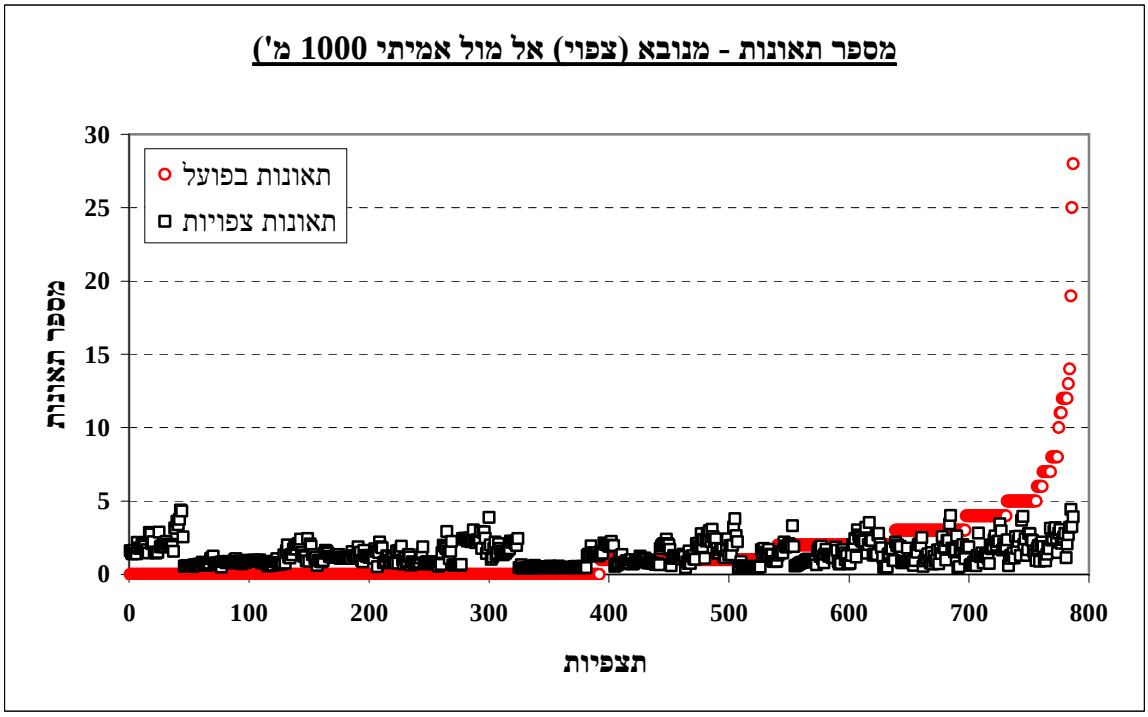
שוב, כדוגמא, עבור מקטע דרך המאופיין בנפח תנועה של 1000 מכוניות ו- 3 תאונות בתקופת המחקר, במידה ויגדל נפח התנועה ביחידה אחת, דהיינו פי 2.7, ויגיע לנפח של 2718 מכוניות, שיעור התאונות יגדל פי 1.9. בדוגמא שלנו מספר התאונות צפוי להיות כמעט כפול ולעמוד על ערך של 5.7.

CurvePrm3 – יחידות העקמומיות מתקבלות בערכים הנעים בין 0 ל- 1. ככל שהערך גבוה יותר, כך העקמומיות בקטע הדרך גבוהה יותר. בניתוח האמור, כאשר נעלה ביחידה אחת של עקמומיות, מספר התאונות במקטע דרך יעלה פי 2.5.

7.3.3 CROSS VALIDATION של המודל

7.3.3.1 מספר תאונות צפוי אל מול גרף התאונות אמיתי

איור מספר 74 מציג את הגרף המשווה בין מספר התאונות הצפוי (EXPECTED) לפי המודל לבין מספר התאונות בפועל (OBSERVED) בקטעי הדרך באורך 1000 מ'. באדום מוצג מספר התאונות האמיתי ובשחור מוצג מספר התאונות הצפוי.



איור מספר 74 - מספר תאונות צפוי אל מול מספר תאונות אמיתי – מקטעי דרך באורך 1000 מ'

ניתן לראות כי המתאם אינו מספיק טוב, ולפיכך נדרש להליך ה EB. הליך זה פורט כאמור, בהרחבה בסקר הספרות בסעיף 2.3.3.4, ובפסקה 6.4.4.

7.3.3.2 השוואת מודל האימון (2/3 הנתונים), למודל שנבנה מנתוני קבוצת הביקורת (1/3 הנתונים).

ההליך שתואר לעיל ובוצע עבור מודל האימון (2/3 מהנתונים), בוצע גם על מסד נתוני הביקורת, המכיל 1/3 מהנתונים. יודגש, כי מראש שולבו במודל זה הפרמטרים המשמעותיים, כפי שעלו ממודל האימון. טבלה מספר 47 מציגה את תוצאות מודל קבוצת הביקורת.

טבלה 45 - תוצר פרוצדורת GENMOD עבור נתוני קבוצת הביקורת של 1/3 הנתונים (ניתוח סגמנטים באורך 1000 מ')

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.8227	1.2955	-8.3618	-3.2837	-8.4068	-3.3026	20.20	<.0001
aadt_log	1	0.6524	0.1313	0.3950	0.9098	0.3973	0.9147	24.68	<.0001
ezor1	1	-0.7360	0.2585	-1.2427	-0.2293	-1.2453	-0.2269	8.11	0.0044
ezor3	1	-0.3423	0.1996	-0.7336	0.0490	-0.7381	0.0491	2.94	0.0864
CurvePrm3	1	0.6821	0.8787	-1.0401	2.4043	-1.0137	2.4535	0.60	0.4376
Dispersion	1	2.0795	0.2690	1.5524	2.6067	1.6095	2.6744		

גם כאן ניתן לחשב ערכים מנובאים לפי נוסחה מספר 89:

[89]

$$\mu_{1/3} = e^{-5.8227+0.652*AADT\_Log+(-0.736)*Ezor1+(-0.342)*Ezor3+(0.682)*CurvePrm3}$$

לאחר שהתקבלו אומדנים משני המודלים, נבחן, האם האומדנים שהתקבלו מהמודל שנבנה מנתוני הביקורת (1/3 הנתונים) אינם שונים באופן מובהק ממקדמי מודל האימון (2/3 הנתונים). לשם כך, יש לבדוק האם אומדני מודל הביקורת נופלים ברווח הסמך של מקדמי מודל האימון, ברמת ביטחון 95%.

לשם כך, יש להשוות את מקדמי המשתנים המסבירים מנתוני מודל הביקורת המופיעים בעמודת **Estimate** בטבלת **Analysis Of Maximum Likelihood Parameter Estimates** (טבלה מספר 47), אל נתוני טווח רווח הסמך של מודל האימון כפי שמופיעים בעמודת **Likelihood Ratio 95% Confidence Limits** (טבלה מספר 45).

כך, מטבלאות 45 ו- 47, ניתן לראות כי עבור החלק הארי של המשתנים במודל, ערך המקדם נופל בתחום רווח הסמך של פונקציית ה Likelihood של נתוני מודל האימון (2/3). **מקדם** המשתנה **ezor3**,המתייחס כאמור להבדל בין אזור המרכז (**ezor3**) לאזור הצפון (**ezor2**) , שערכו הינו -0.3423 , אינו נופל בטווח רווח הסמך של נתוני מודל האימון (2/3) הנע בין **-0.8465** ל **-0.3603** . המשמעות הינה כי עבור פרמטר זה המודל אינו טוב מספיק, וההבדל בשיעור תאונות הדרכים בין האזורים גדול יותר בסט נתוני ה 1/3. אפשר והסיבה לכך הינה, שבמודל האימות שנבנה על סמך 1/3 הנתונים המשלימים לסט הנתונים הכולל, שולבו <u>מלכתחילה</u>, הפרמטרים שיצאו כמשמעותיים במודל הראשוני (2/3 הנתונים). אפשר ובניית מודל עבור 1/3 הנתונים, הבוחן ומשלב את כלל הפרמטרים המקוריים, תניב משתנים אחרים המשמעותיים עבור המודל, ותיספק את התשובה לאי ההתאמה בפרמטר זה.

7.3.3.3 מבחן TOST לשוויון (TOST Equivalence Test)

בטבלה מספר 48 מוצגות תוצאות המבחן עבור מקטעי דרך של 1000 מ'.

טבלה 46 - מבחן TOST לשוויון – מקטעי דרך 1000 מ'

Mean	Lower Bound		90% CL Mean			Upper Bound	Assessment
0.0547	-0.25	<	0.0339	0.0756	<	0.25	Equivalent

הערך המופיע בעמודת "גבולות רווח הסמך" (90% CL Mean), אכן נכלל בין הטווח העליון (Upper Bound) לטווח התחתון (Lower Bound). המשמעות היא כי יש לקבל את H_0 ולדחות את H_1. משמע, התוצאות שוות דיין. כלומר מספר התאונות הצפוי, כפי שחושב משני מערכי הנתונים ("1/3" הנתונים ו- "2/3" הנתונים) קרוב דיו.

7.3.3.4 השוואת יחסי ההסתברות לאפס תאונות

כאמור, במודל שניבנה ניתן לחשב ניבוי עבור ההסתברות הממוצעת לאפס תאונות, לתאונה אחת, שתי תאונות וכן הלאה. במידה והמודל שנבנה הינו מודל "טוב", אזי, אם נשווה בין אחוז המקטעים בהם התקיימו "אפס תאונות" בפועל, לבין ההסתברות הממוצעת להתרחשות של "אפס תאונות" במודל המחושב, נצפה כי הפרופורציות תהינה דומות.

טבלה מספר 49 מציגה בעמודת ה- MEAN את ממוצע ההסתברויות המחושב להתרחשות אפס תאונות (po0) והסתברות להתרחשות תאונה אחת (po1) במקטעי הדרך של 1000 מטרים.

טבלה 47 - ממוצע הסתברויות מחושב להתרחשות 0 תאונות ותאונה אחת במקטע דרך באורך 1000 מ'

Variable	N	Mean	Std Dev	Minimum	Maximum
po0	787	0.4928599	0.1177859	0.2384680	0.7220342
po1	787	0.2114902	0.0143081	0.1506593	0.2255882

טבלה מספר 50 מציגה את אחוז המקטעים האמיתיים, בהם התרחשו 0 תאונות, תאונה אחת, שתי תאונות, וכן הלאה.

טבלה48 - פרופורציות המקטעים האמיתיים בהם התרחשו N תאונות - 1000 מ'

AccCNT				
AccCNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	392	49.81	392	49.81
1	148	18.81	540	68.61
2	98	12.45	638	81.07
3	59	7.50	697	88.56
4	34	4.32	731	92.88
5	25	3.18	756	96.06
6	5	0.64	761	96.70
7	7	0.89	768	97.59
8	6	0.76	774	98.35
10	1	0.13	775	98.48
11	2	0.25	777	98.73
12	5	0.64	782	99.36
13	1	0.13	783	99.49
14	1	0.13	784	99.62
19	1	0.13	785	99.75
25	1	0.13	786	99.87
28	1	0.13	787	100.00

מהשוואת שתי הטבלאות עולה כי ממוצע הערכים של ההסתברויות להתרחשות תאונה אחת במקטע דרך של 1000 מ', בעמודת po1, עומד על 49.28%. ערך זה קרוב לערך 49.81% המופיע בטבלה מספר 50, המציגה כאמור את פרופורצית המקטעים האמיתיים בהם התרחשו 0 תאונות במקטע דרך. דוגמא נוספת היא ערך של 21.14% בערכים המחושבים לתאונה אחת לעומת 18.81% בפרופורציות המקטעים האמיתיים בהן התרחשה תאונה אחת.

7.3.3.5 בניית מודל חדש מלא מנתוני הביקורת (עבור 1/3 הנתונים), והשוואתו אל מודל האימון (2/3 הנתונים)

טבלה מספר 51 מציגה את תוצאות מודל הביקורת המלא, עבור 1/3 הנתונים.

טבלה 49 - מודל מלא שנבנה מ 1/3 נתוני קבוצת הביקורת (מקטעי דרך 1000 מ') – טבלת Analysis Of Maximum Likelihood Parameter Estimates

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.6635	1.2763	-8.1651	-3.1619	-8.2075	-3.1782	19.69	<.0001
aadt_log	1	0.6452	0.1307	0.3890	0.9013	0.3911	0.9062	24.36	<.0001
ezor1	1	-0.7633	0.2563	-1.2656	-0.2610	-1.2683	-0.2589	8.87	0.0029
ezor3	1	-0.3531	0.1996	-0.7442	0.0380	-0.7488	0.0382	3.13	0.0768
Dispersion	1	2.0876	0.2696	1.5592	2.6160	1.6164	2.6838		

במודל זה, בשונה ממודל האימון, התקבל כי פרמטר **העקמומיות אינו משמעותי**, והוא לא שולב במודל.

פרמטר האזור נמצא כמשמעותי במודל, כאשר שיעור התאונות הצפוי בצפון גדול יותר משיעור התאונות הצפוי במרכז ובדרום. מעניין לציין כי שיעור התאונות באזור המרכז (3) אינו מובהק ושולב במודל בגין ערך ה- AIC.

גם פרמטר נפח התנועה AADT_log (המייצג כאמור, ממוצע נפח תנועה יומי לימי חול) עלה כבעל משמעות במודל. עבור עליה ביחידה אחת של המשתנה AADT_log (שמשמעה נפח הגדול פי 2.7182 מהערך הקודם), שיעור התאונות צפוי להיות גדול כמעט פי 2 (פי 1.90).

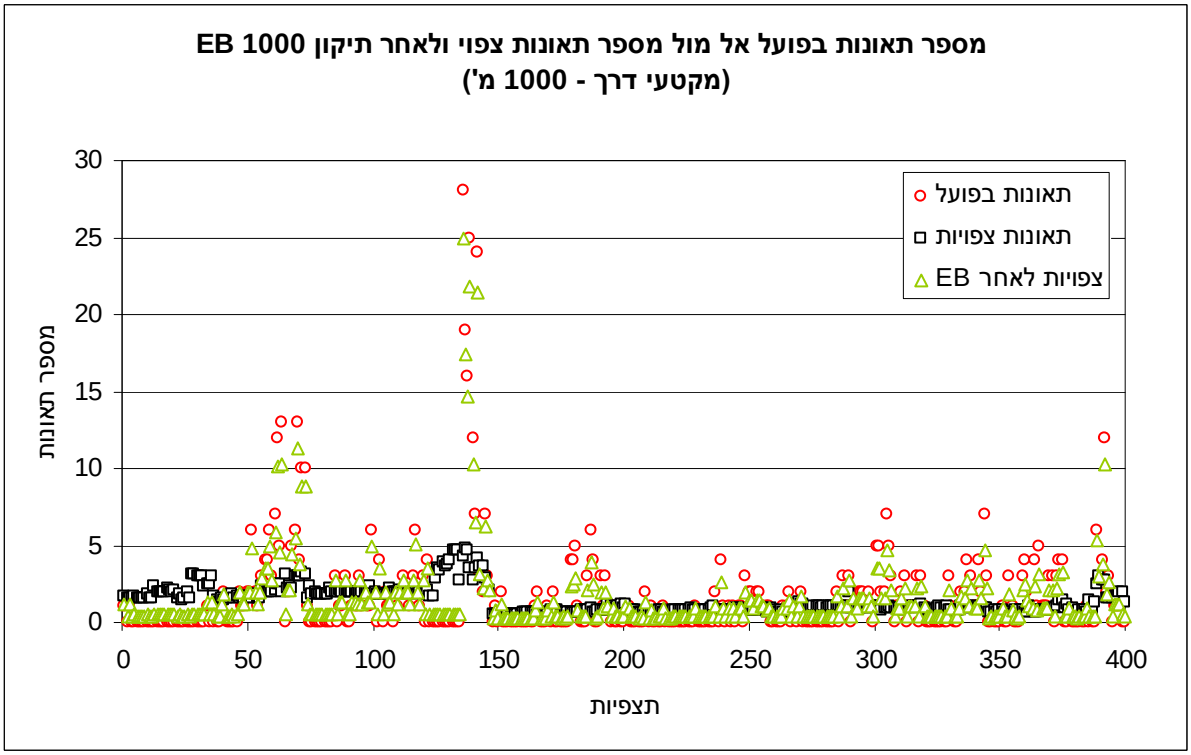
שוב, כדוגמא, עבור מקטע דרך המאופיין בנפח תנועה של 1000 מכוניות ו- 3 תאונות בתקופת המחקר, הרי שבמידה ויגדל נפח התנועה ביחידה אחת, דהיינו פי 2.7, ויגיע לנפח של 2718 מכוניות, שיעור התאונות צפוי לעלות פי 1.9. בדוגמא שלנו מספר התאונות צפוי להיות כמעט כפול ולעמוד על ערך של 5.7.

7.3.3.6 Empirical Bayes (להלן: EB) – מקטעי דרך באורך של 1000 מ'.

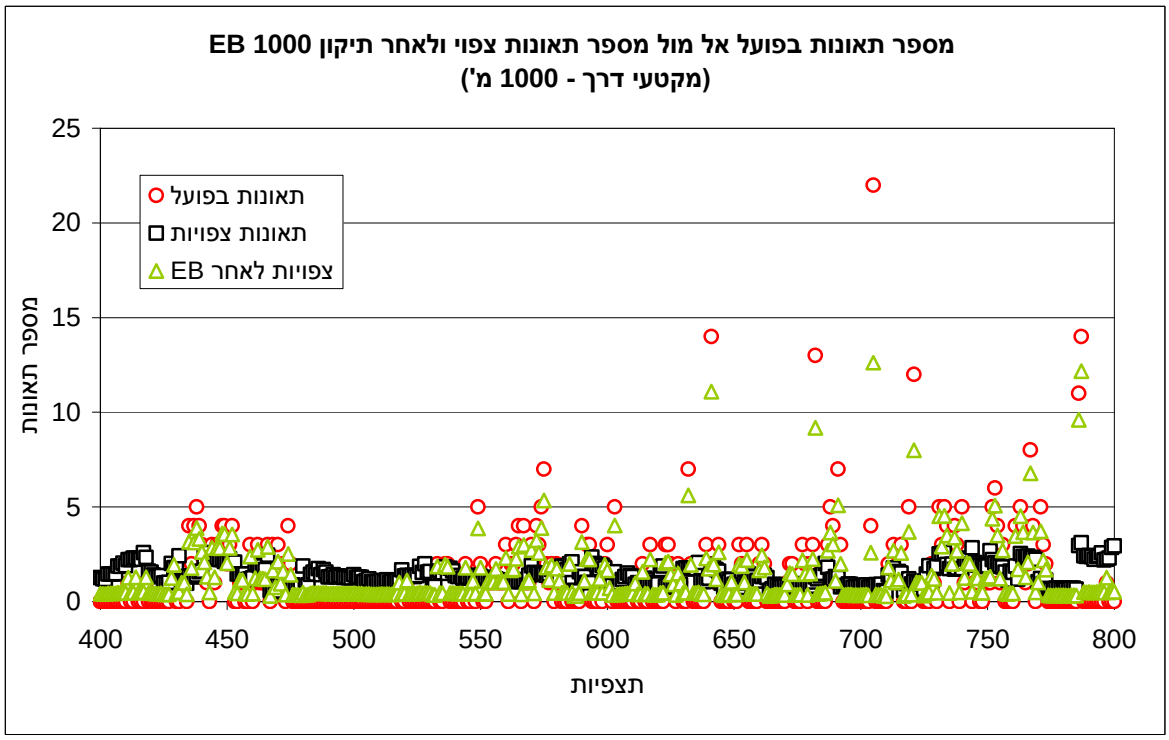
כזכור, ההליך בוצע לפי הנוסחה המופיעה במאמרם של Sayed & Rodriguez [1999] ו- Ng et-al [2002], כפי שפורטה סעיף 6.4.4.

איור מספר 75 מציג את תוצאות הליך ה EB עבור מקטעי דרך באורך של 1000 מ'. איורים מספר 76-78 מציגים בהרחבה את תוצאות המקור, בפילוח מספר תצפיות שונה (על מנת שנוכל להבחין טוב יותר במגמות) .

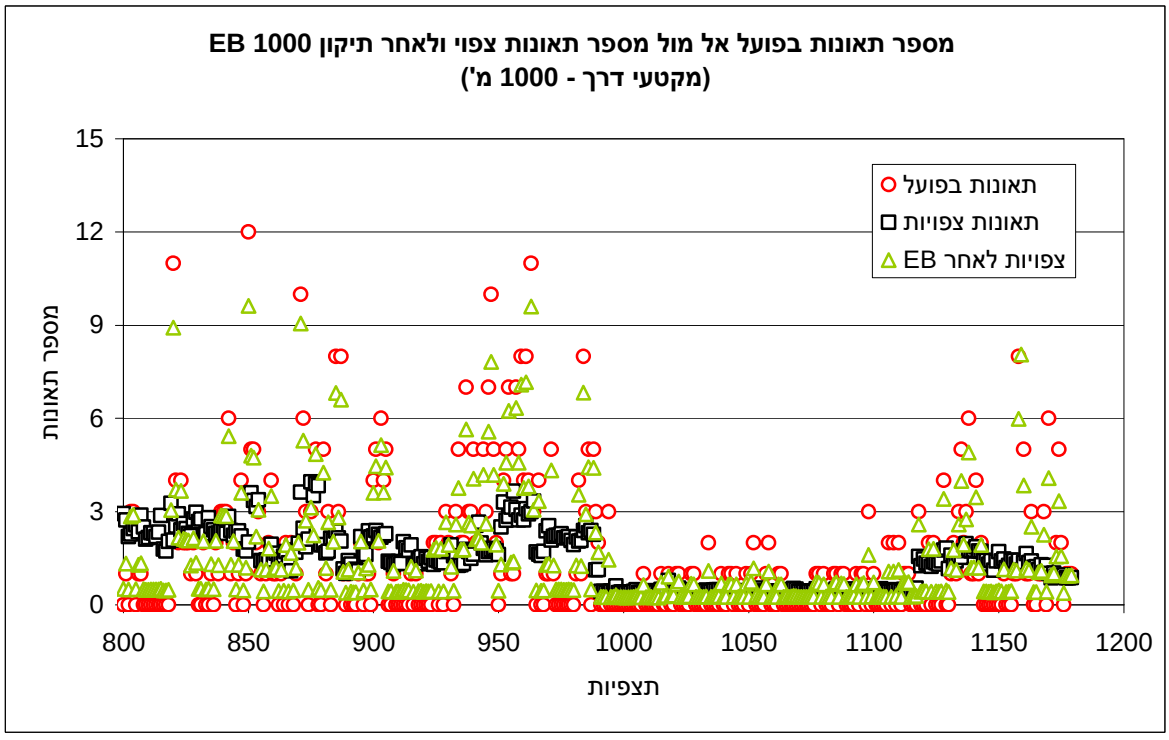
איור מספר 75 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1000 מ' (כלל התצפיות)



איור מספר 76 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1000 מ' (ממספר תצפית 0 עד מספר תצפית 400)



איור מספר 77 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1000 מ' (ממספר תצפית 400 עד מספר תצפית 800)



איור מספר 78 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1000 מ' (ממספר תצפית 800 עד מספר תצפית 1200)

7.4 מקטעי דרך באורך 1500 מטרים

7.4.1 בחירת המשתנים עבור המודל

טבלה מספר 52 מציגה את תוצר ה AIC עבור מקטעי דרך באורך של 1500 מטרים.

המודל הטוב ביותר עבור סט הנתונים באורך זה – הינו מסוג Negative Binomial.

טבלה 50 - תוצר AIC עבור מקטעי דרך באורך 1500 מ'

AIC	L=1500
פואסוני	1794
בינומי-שלילי	**1339**
Zero-Inflated פואסון	1473
Zero-Inflated בינומי-שלילי	1340

7.4.2 פרוצדורת GENMOD

7.4.2.1 ניתוח טבלת Analysis Of Maximum Likelihood Parameter Estimates

טבלה מספר 53 מציגה את תוצר טבלת Analysis Of Maximum Likelihood Parameter Estimates עבור מקטעי דרך באורך 1500 מ'.

טבלה 51 - טבלת Analysis Of Maximum Likelihood Parameter Estimates **, מקטעי דרך 1500 מ'**

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.5077	1.2367	-8.9315	-4.0839	**-8.9687**	**-4.0979**	27.69	<.0001
aadt_log	1	0.7452	0.1233	0.5035	0.9869	**0.5054**	**0.9910**	36.52	<.0001
ezor1	1	-0.6473	0.2190	-1.0766	-0.2180	**-1.0786**	**-0.2159**	8.73	0.0031
ezor3	1	-0.5933	0.1852	-0.9562	-0.2304	**-0.9600**	**-0.2303**	10.27	0.0014
CurvePrm3	1	1.5118	0.6182	0.3001	2.7234	**0.3192**	**2.7566**	5.98	0.0145
Dispersion	1	1.6370	0.2032	1.2387	2.0353	**1.2799**	**2.0837**		

הפרמטרים שנמצאו כמשמעותיים מתאימים למקטעי הדרך באורכים של 500, 750 ו- 1000 מטרים.

- **aadt_log** - ממוצע נפח תנועה יומי לימי חול, לאחר שעבר טרנספורמציה ללוגריתם טבעי.
- **ezor1** (דרום), **ezor3** (מרכז) – מייצגים את משתנה האזור. כאשר המודל מאפשר השוואה וניתוח בין אזור 2 (צפון) לשני האזורים האחרים.
- **CurvePrm3** - עקמומיות

בעזרת מבחן יחס הנראות, נבחן האם פרמטר פיזור היתר שונה באופן מובהק מ-1. כאמור, בכדי לבצע זאת, נבדוק עד כמה מקדם המשתנה (Estimate) של פרמטר ה Dispersion, שערכו במודל הנוכחי שלנו 1.6370 קרוב מספיק או רחוק מספיק מ- 1. לשם כך נבחן את רווח הסמך המופיע בטבלת **Analysis Of Maximum Likelihood Parameter Estimates** בעמודת **"Likelihood Ratio 95% Confidence Limits"** (טבלה 54). כפי שניתן לראות בטבלה מספר 53 הערך 1, אינו נופל ברווח הסמך (הנע בין **1.2799 ל 2.0837**) ברמת מובהקות 95%. מכאן, ניתן להסיק כי קיים פיזור יתר בהתפלגות של מספר תאונות הדרכים, ולפיכך, גם בסט נתונים זה, **המודל המתאים יותר הינו מסוג בינומי שלילי**.

בנוסף, בהתאם לטבלה 53, ניתן לבנות עבור מודל האימון (2/3 הנתונים) את **הנוסחה לניבוי תוחלת תאונות הדרכים μ** (=ממוצע תאונות למקטע דרך לתקופת המחקר העומדת על שלוש שנים). הערכים המנובאים התקבלו דרך הקשר הלוגריתמי. בהתאם לנוסחה מספר 87:

[87]

$$\mu_{2/3} = e^{-6.5077+0.7452*AADT\_Log+(-0.6473)*Ezor1+(-0.5933)*Ezor3+(1.5118)*CurvePrm3}$$

גם באורך זה, קיימת התאמה לממצאים של מקטעי הדרך של 500, 750 ו- 1000 מטרים בנוגע לפרמטר ה"אזור". המינוס במשתנים ezor1 , ezor3. משמעו, שבאזור 2 (צפון) צפויות להתרחש יותר תאונות מאשר באזור 1 (דרום). כמו כן, באזור 2 (צפון) גם צפויות להתרחש יותר תאונות מאשר באזור 3 (מרכז).

7.4.2.2 ניתוח טבלת Contrast Estimate Results

טבלה מספר 54 מציגה את טבלת **Contrast Estimate Results** עבור מקטעי הדרך באורך של 1500 מטרים.

טבלה 52 - טבלת Contrast Estimate Results**, מקטעי דרך 1500 מ'**

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
adj. rr: CurvePrm3	4.5348	1.3500	15.2326	1.5118	0.6182	0.05	0.3001	2.7234	5.98	0.0145
Exp(adj. rr: CurvePrm3)				4.5348	2.8034	0.05	1.3500	15.2326		
adj. rr: AADT_log	2.1068	1.6545	2.6828	0.7452	0.1233	0.05	0.5035	0.9869	36.52	<.0001
Exp(adj. rr: AADT_log)				2.1068	0.2598	0.05	1.6545	2.6828		
adj. rr: ezor1 vs 2	0.5234	0.3407	0.8041	-0.6473	0.2190	0.05	-1.0766	-0.2180	8.73	0.0031
Exp(adj. rr: ezor1 vs 2)				0.5234	0.1147	0.05	0.3407	0.8041		
adj. rr: ezor3 vs 2	0.5525	0.3843	0.7942	-0.5933	0.1852	0.05	-0.9562	-0.2304	10.27	0.0014
Exp(adj. rr: ezor3 vs 2)				0.5525	0.1023	0.05	0.3843	0.7942		

מטבלה 54 עולה כי באזור הצפון שיעור תאונות הצפוי גדול **פי 1.91** משיעור התאונות הצפוי בדרום (אזור 1).

בנוסף, באזור הצפון (2) שיעור תאונות הצפוי גדול פי 1.80 משיעור התאונות באזור המרכז (3).

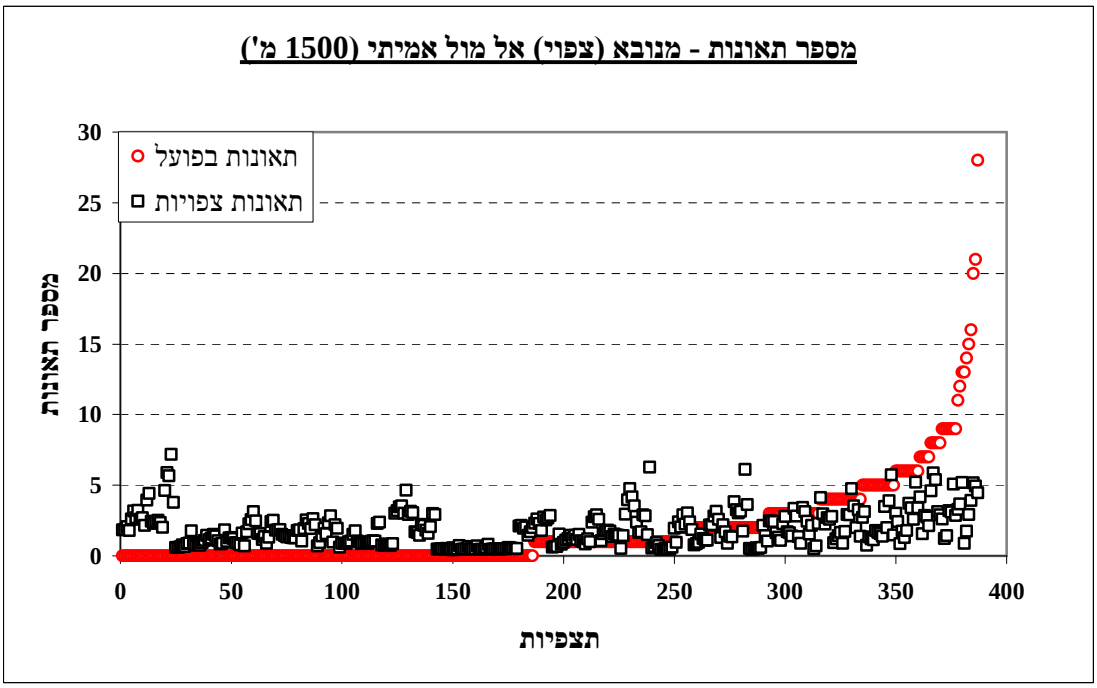
AADT_log – AAAD מייצג כאמור, ממוצע נפח תנועה יומי לימי חול (מספר כלי הרכב שנעו בקטע דרך מסויים בימי חול). עבור עליה ביחידה אחת של המשתנה AADT_log (שמשמעה נפח הגדול פי 2.71 מהערך הקודם), שיעור התאונות יגדל פי 2.1.
שוב, כדוגמא, במקטע דרך המאופיין בנפח תנועה של 1000 מכוניות ו 3 תאונות בתקופת המחקר, במידה ויגדל נפח התנועה ביחידה אחת, דהיינו פי 2.7, ויגיע לנפח של 2718 מכוניות, שיעור התאונות יגדל פי 2.1. בדוגמא שלנו מספר התאונות צפוי לעמוד על ערך של 6.3.

CurvePrm3 – יחידות העקמומיות התקבלו בערכים הנעים בין 0 ל- 1. ככל שהערך גבוה יותר, כך העקמומיות בקטע הדרך גבוהה יותר. בניתוח האמור, כאשר נעלה ביחידה אחת של עקמומיות, מספר התאונות במקטע דרך יעלה פי 4.5.

7.4.3 CROSS VALIDATION של המודל

7.4.3.1 מספר תאונות צפוי אל מול גרף התאונות אמיתי

איור מספר 79 מציג את הגרף המשווה בין מספר התאונות הצפוי (EXPECTED) לפי המודל לבין מספר התאונות בפועל (OBSERVED) בקטעי הדרך באורך 1500 מ'. באדום מוצג מספר התאונות האמיתי ובשחור מספר התאונות הצפוי.



איור מספר 79 - מספר תאונות צפוי אל מול מספר תאונות אמיתי – מקטעי דרך באורך 1500

ניתן לראות מהגרף, כי המתאם אינו טוב מספיק, ולפיכך נדרש הליך ה- EB. הליך זה פורט, כזכור בהרחבה בסקר הספרות בסעיף 2.3.3.4, ובפסקה 6.4.4.

7.4.3.2 השוואת מודל האימון (2/3 הנתונים), למודל שנבנה מנתוני קבוצת הביקורת (1/3 הנתונים).

ההליך שתואר לעיל ובוצע עבור 2/3 מהנתונים, בוצע גם על מסד נתוני הביקורת, המכיל 1/3 מהנתונים. יודגש, כי מראש שולבו במודל זה הפרמטרים המשמעותיים, כפי שעלו ממודל האימון (2/3 מהנתונים). טבלה מספר 55 מציגה את המקדמים עבור קבוצת הביקורת.

טבלה 53 - תוצר פרוצדורת GENMOD עבור נתוני קבוצת הבקרה של 1/3 הנתונים (ניתוח סגמנטים באורך 1500 מ')

Analysis Of Maximum Likelihood Parameter Estimates -									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.9050	1.3903	-7.6299	-2.1802	-7.6754	-2.1815	12.45	0.0004
aadt_log	1	0.6242	0.1433	0.3434	0.9050	0.3443	0.9107	18.98	<.0001
ezor1	1	-0.7598	0.2865	-1.3213	-0.1983	-1.3271	-0.1956	7.03	0.0080
ezor3	1	-0.4338	0.2609	-0.9452	0.0775	-0.9529	0.0795	2.76	0.0964
CurvePrm3	1	-0.1837	0.9052	-1.9578	1.5904	-1.9491	1.6414	0.04	0.8392
Dispersion	1	1.4825	0.2338	1.0242	1.9408	1.0851	2.0160		

גם כאן ניתן לחשב ערכים מנובאים לפי נוסחה מספר 88:

[88]

$$\mu_{1/3} = e^{-4.9050+0.6242*AADT\_Log+(-0.7598)*Ezor1+(-0.4338)*Ezor3+(-0.1837)*Curve\,Prm3}$$

לאחר שהתקבלו אומדנים משני המודלים, נבחן, האם האומדנים שהתקבלו מהמודל שנבנה מנתוני המבחן (1/3) אינם שונים באופן מובהק ממקדמי מודל נתוני האימון (2/3). כלומר, נבדוק האם אומדני ה- 1/3 נופלים ברווח הסמך של מקדמי 2/3, ברמת ביטחון 95%. לשם כך, נשווה את מקדמי המשתנים המסבירים מנתוני מודל הביקורת המופיעים בעמודת **Estimate** בטבלת **Analysis Of Maximum Likelihood Parameter Estimates** (טבלה 55), אל נתוני טווח רווח הסמך של נתוני מודל האימון (2/3) המופיעים בעמודת **Likelihood Ratio 95% Confidence Limits** (טבלה 53).

ערך מקדם המשתנה עקמומיות הינו -0.1837 . מספר זה אינו נופל בתחום רווח הסמך של פונקציית ה Likelihood של נתוני ה 2/3, הנע בין הערכים **0.3192** ל **2.7566**. המשמעות היא כי המודל איננו טוב מספיק עבור משתנה זה. בנוסף, ניתן לראות כי משתנה העקמומיות אינו מובהק – ושולב במודל בשל ערך ה AIC הנמוך שלו.

7.4.3.3 מבחן TOST לשוויון (TOST Equivalence Test)

בטבלה מספר 56 מוצגות תוצאות המבחן עבור מקטעי דרך של 1500 מ'.

טבלה 54 - תוצאות מבחן TOST למקטעי דרך של 1500 מ'.

Mean	Lower Bound		90% CL Mean			Upper Bound	Assessment
0.4102	-0.25	<	0.3466	0.4737	>	0.25	Not equivalent

הערך המופיע בעמודת "גבולות רווח הסמך" (**90% CL Mean**), **<u>אינו</u>** נכלל בין הטווח העליון (**Upper Bound**) לטווח התחתון (**Lower Bound**). משמע, יש לקבל את H_1 ולדחות את H_0. המשמעות היא כי התוצאות אינן שוות דיין. כלומר, מספר התאונות הצפוי כפי שחושב משני מערכי הנתונים ("1/3" ו- "2/3") **<u>אינו</u>** קרוב דיו, והמודלים שהושוו, בהיבט מספר התאונות הצפוי, אינם מספקים ניבוי דומה.

7.4.3.4 השוואת יחסי ההסתברות לאפס תאונות

כאמור, במודל שניבנה, ניתן לחשב ניבוי עבור ההסתברות הממוצעת לאפס תאונות, לתאונה אחת, שתי תאונות וכן הלאה. במידה והמודל שנבנה הינו מודל "טוב", אזי, אם נשווה בין אחוז המקטעים בהם התקיימו "אפס תאונות" בפועל, לבין ההסתברות הממוצעת להתרחשות של "אפס תאונות" במודל המחושב, נצפה כי היחסים יהיו דומים.

טבלה מספר 57 מציגה בעמודת ה MEAN את ממוצע ההסתברויות המחושב להתרחשות אפס תאונות והסתברות להתרחשות תאונה אחת במקטעי הדרך באורך של 1500 מטרים.

טבלה 55 - ממוצע הסתברויות מחושב להתרחשות 0 תאונות ותאונה אחת במקטע דרך באורך 1500 מ'

Variable	N	Mean	Std Dev	Minimum	Maximum
po0	387	0.4706385	0.1305741	0.2110189	0.7234223
po1	387	0.1897242	0.0190649	0.1188117	0.2097248

טבלה מספר 58 מציגה את אחוז המקטעים האמיתיים, בהם התרחשו 0 תאונות, תאונה אחת, שתי תאונות, וכן הלאה.

טבלה 56 - פרופורציות המקטעים האמיתיים בהם התרחשו N תאונות - 1500 מ'

AccCNT				
AccCNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	186	48.06	186	48.06
1	65	16.80	251	64.86
2	41	10.59	292	75.45
3	23	5.94	315	81.40
4	19	4.91	334	86.30
5	15	3.88	349	90.18
6	11	2.84	360	93.02
7	5	1.29	365	94.32
8	5	1.29	370	95.61
9	7	1.81	377	97.42
11	1	0.26	378	97.67
12	1	0.26	379	97.93
13	2	0.52	381	98.45
14	1	0.26	382	98.71
15	1	0.26	383	98.97
16	1	0.26	384	99.22
20	1	0.26	385	99.48
21	1	0.26	386	99.74
28	1	0.26	387	100.00

מהשוואת שתי הטבלאות עולה כי ממוצע הערכים (של ההסתברויות להתרחשות תאונה אחת במקטע דרך של 1500 מ') בעמודת po1, עומד על 47.06%. ערך זה קרוב לערך 48.06% המופיע בטבלה מספר 58, המציגה כאמור את פרופורצית המקטעים האמיתיים בהם התרחשו 0 תאונות במקטע דרך. דוגמא נוספת היא הערך של 18.97% בערכים המחושבים לתאונה אחת לעומת 16.80% בפרופורציות המקטעים האמיתיים בהן התרחשה תאונה אחת.

7.4.3.5 בניית מודל חדש מלא מנתוני הביקורת (עבור 1/3 הנתונים), והשוואתו אל מודל האימון (2/3 הנתונים)

טבלה מספר 59, מציגה את תוצאות טבלת **Analysis Of Maximum Likelihood Parameter Estimates** עבור מודל חדש ומלא מנתוני הביקורת.

טבלה 57 - טבלת Analysis Of Maximum Likelihood Parameter Estimates**, עבור מודל מלא מ 1/3 נתוני הביקורת, מקטעי דרך של 1500 מ'.**

Analysis Of Maximum Likelihood Parameter Estimates -									
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.9586	1.3648	-7.6335	-2.2837	-7.6808	-2.2874	13.20	0.0003
aadt_log	1	0.6268	0.1426	0.3472	0.9064	0.3483	0.9122	19.31	<.0001
ezor1	1	-0.7515	0.2835	-1.3071	-0.1959	-1.3125	-0.1930	7.03	0.0080
ezor3	1	-0.4348	0.2607	-0.9458	0.0761	-0.9536	0.0780	2.78	0.0953
Dispersion	1	1.4825	0.2339	1.0241	1.9408	1.0850	2.0160		

מטבלה 59, ניתן לראות כי פרמטר העקמומיות במקטעי הדרך באורך זה נמצא כלא משמעותי, ולפיכך לא שולב במודל זה. בנוסף, פרמטר האזור נמצא כמשמעותי במודל, כאשר שיעור התאונות הצפוי בצפון גדול יותר משיעור התאונות הצפוי במרכז ובדרום. מעניין לציין כי שיעור התאונות באזור המרכז (3) אינו מובהק. גם פה שולב במודל בגין ערך ה- AIC.

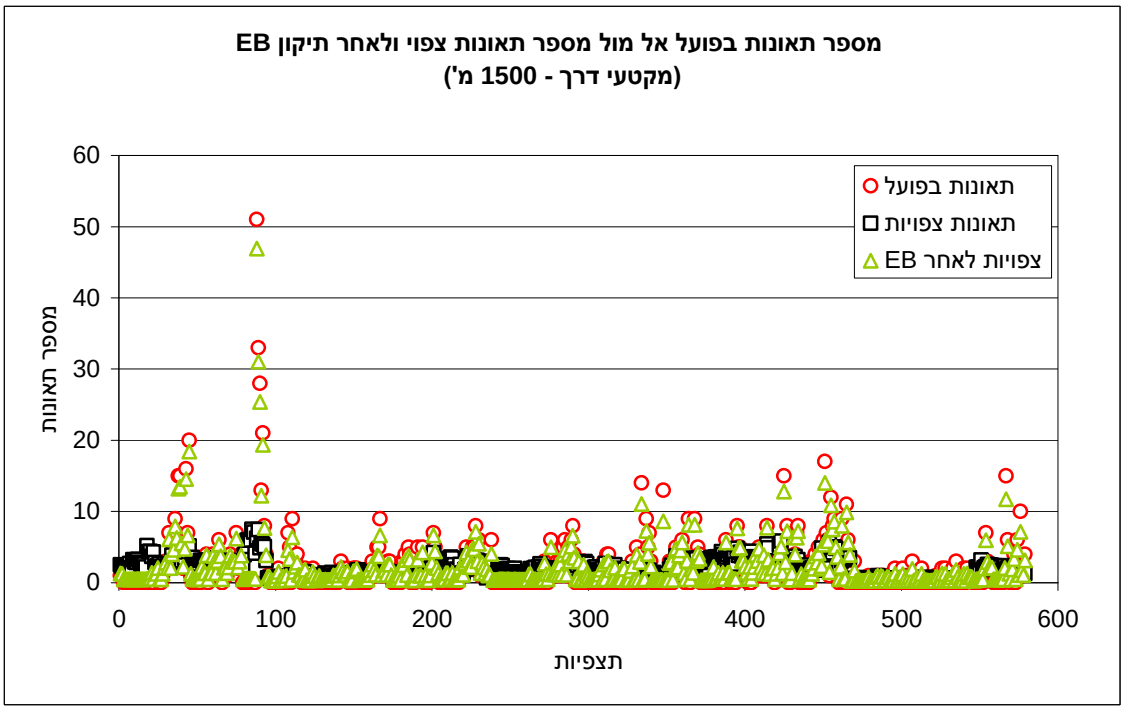
גם פרמטר נפח התנועה - AADT_log (המייצג כאמור, ממוצע נפח תנועה יומי לימי חול) נמצא כבעל משמעות במודל. עבור עליה ביחידה אחת של המשתנה AADT_log, שיעור התאונות הצפוי עתיד לגדול פי 1.87.

שוב, כדוגמא, עבור מקטע דרך המאופיין בנפח תנועה של 1000 מכוניות ו- 3 תאונות בתקופת המחקר, במידה ויגדל נפח התנועה ביחידה אחת, דהיינו פי 2.7, ויגיע לנפח של 2718 מכוניות, שיעור התאונות יגדל פי 1.87, וצפוי לעמוד על ערך של 5.61.

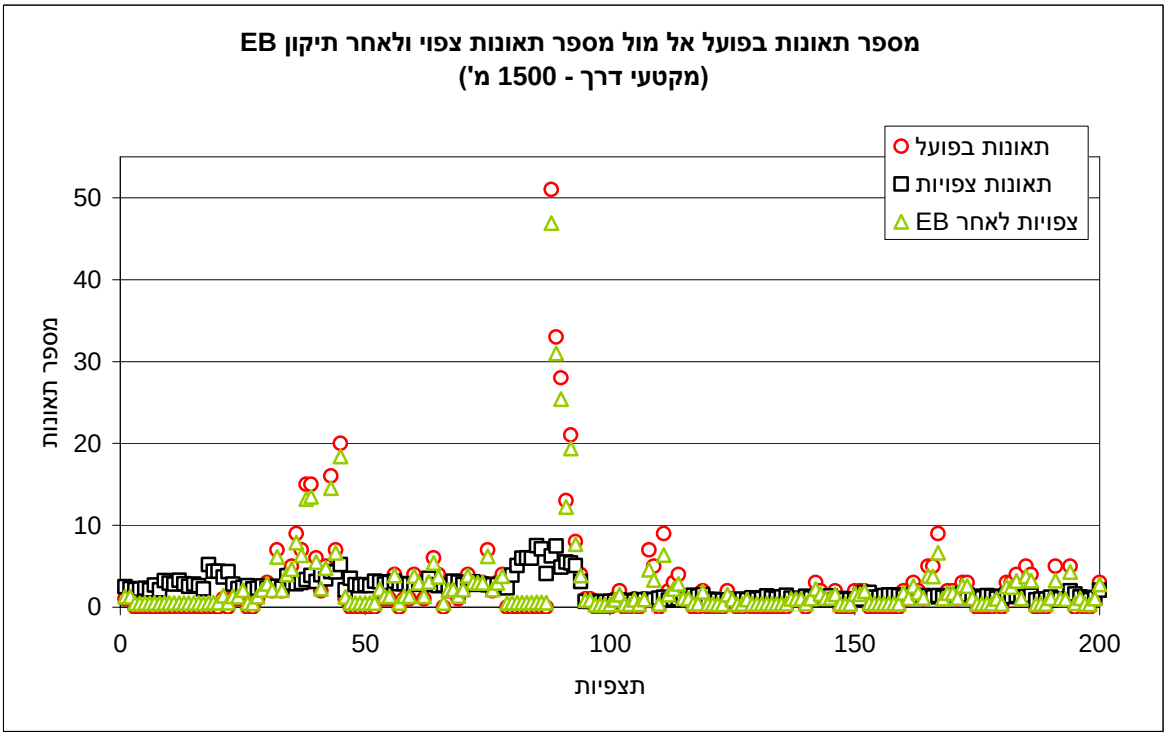
7.4.3.6 Empirical Bayes (להלן: EB) – מקטעי דרך באורך של 1500 מ'.

כזכור, ההליך בוצע לפי הנוסחה המופיעה במאמרם של Sayed & Rodriguez [1999] ו- Ng et-al [2002], כפי שפורטה סעיף 6.4.4.

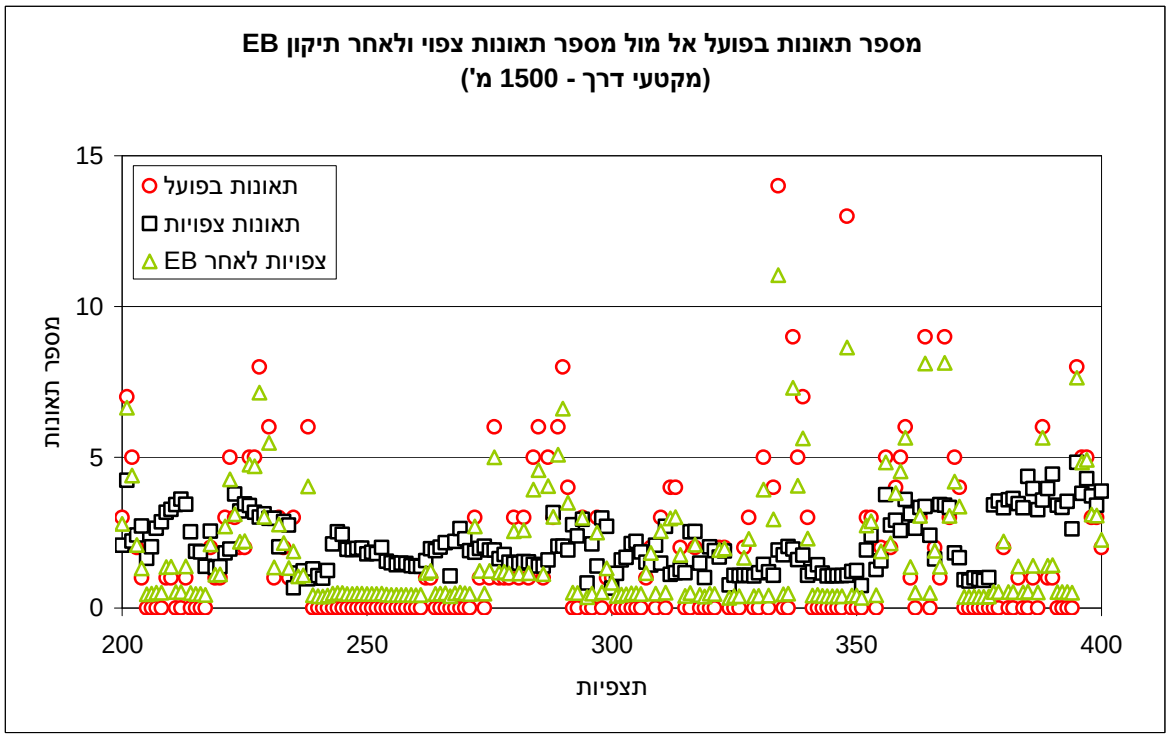
איור מספר 80 מציג את תוצאות הליך ה- EB עבור מקטעי דרך באורך של 1500 מ'. איורים מספר 81-83 מציגים בהרחבה את תוצאות המקור, בפילוח מספר תצפיות שונה (על מנת שנוכל להבחין טוב יותר במגמות) .



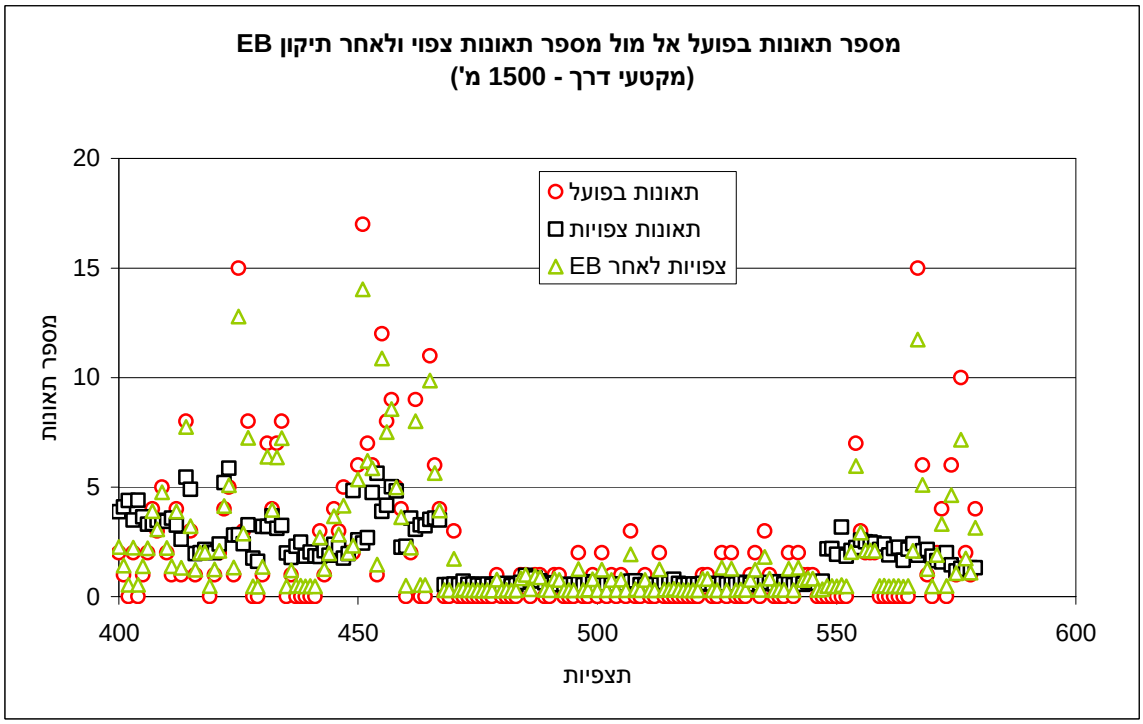
איור מספר 80 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1500 מ' (כלל התצפיות)



איור מספר 81 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1500 מ' (ממספר תצפית 0 עד מספר תצפית 200)



איור מספר 82 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1500 מ' (ממספר תצפית 200 עד מספר תצפית 400)



איור מספר 83 - מספר תאונות בפועל אל מול מספר תאונות צפוי, ומספר תאונות לאחר ביצוע תיקון EB, מקטעי דרך באורך 1500 מ' (ממספר תצפית 400 עד מספר תצפית 600)

7.5 ניתוחים משלימים

7.5.1 ניתוח אזורים (צפון, מרכז, דרום)

7.5.1.1 כללי

היות ובמודלים שתוארו בסעיפים הקודמים עלה בצורה עקבית, כי אזור הצפון הינו בעל שיעור תאונות צפוי הגבוה מהאזורים האחרים, בוצעו ניתוחים נוספים על משתנה האזור.
בניתוח זה נבחנו הבדלים סטטיסטיים בין האזורים צפון, מרכז ודרום. ניתוח זה, בשונה מניתוח האינטראקציות (המפורט בסעיף 7.5.3), מספק הסבר ה"נקי" מהשפעת משתנה נפח התנועה. זאת משני טעמים. הראשון, בסיס הנתונים חולק (בחלוקה תחבורתית) לנפחי תנועה בטווח בינוני, נמוך וגבוה (כפי שיפורט בהמשך). השני, נפח התנועה שולב אף הוא כמשתנה במודל. כך, בסיכומו של ניתוח זה, ניתן יהיה לומר כי באזור מסויים מתקיימות יותר או פחות תאונות דרכים, ללא תלות בנפח התנועה.

בחלוקה **התחבורתית** בסיס הנתונים חולק לארבע תתי-קבוצות בהתאם לנפח תנועה המאפיין טווחים בכבישים מסדר ראשון, שני,שלישי ורביעי, קרי: גדול מ- 30,000 ; בין 30,000 ל- 15,000; בין 5000 ל- 15,000 ; קטן מ- 5000 , בהתאמה.
יודגש, כי עקב הצורך לשמור על קבוצות מאוזנות עבור הניתוח הסטטיסטי, נותחו לבסוף שתי קבוצות בהן היה מדגם נתונים מספק. קבוצה: 5000 ל- 15,000 להלן תכונה: "בינוני-נמוך", והקבוצה הנעה בטווח נפח תנועה בין 15,000 ל- 30,000 להלן תכונה: "בינוני-גבוה".

7.5.1.2 חלוקה תחבורתית – מקטעי דרך באורך של 500 מ'.

טבלאות 60 ו-61 מציגות את תוצאות הניתוח עבור נפחי התנועה הבינוניים-נמוכים במקטעי הדרך באורך 500 מטרים.

טבלה 58 - השוואת אזור 1 (דרום) אל מול אזורים 2 (צפון) ו-3 (מרכז) עבור נפח בינוני-נמוך (5000 ל- 15000). מקטעי דרך של 500 מ'

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-5.7974	1.4795	-8.6972	-2.8977	15.36	<.0001
AADT_log		1	0.5062	0.1623	0.1882	0.8243	9.73	0.0018
Ezor	**2**	**1**	**0.8198**	**0.1264**	**0.5720**	**1.0676**	**42.05**	**<.0001**
Ezor	**3**	**1**	**0.2809**	**0.1283**	**0.0294**	**0.5324**	**4.79**	**0.0286**
Dispersion		1	1.8729	0.2014	1.4781	2.2677		

טבלה 59 - השוואת אזור 2 (צפון) אל מול אזורים 1 (דרום) ו-3 (מרכז) עבור נפח בינוני-נמוך (5000 ל- 15000) מקטעי דרך של 500 מ'

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-4.9776	1.4969	-7.9116	-2.0437	11.06	0.0009
AADT_log		1	0.5062	0.1623	0.1882	0.8243	9.73	0.0018
Ezor	1	1	-0.8198	0.1264	-1.0676	-0.5720	42.05	<.0001
Ezor	**3**	**1**	**-0.5389**	**0.1201**	**-0.7743**	**-0.3036**	**20.14**	**<.0001**
Dispersion		1	1.8729	0.2014	1.4781	2.2677		

מטבלה מספר 60, עולה כי **בנפחי התנועה הבינוניים-נמוכים**:

- שיעור התאונות הצפוי באזור 2 (צפון) גדול פי 2.27 משיעור התאונות הצפוי באזור 1 (דרום)
- שיעור התאונות הצפוי באזור 3 (מרכז) גדול פי 1.32 משיעור התאונות הצפוי באזור 1 (דרום).

טבלה מספר 61, משלימה את ההשוואה בין אזורים 2 (צפון) ו-3 (מרכז) עבור **נפחי התנועה הבינוניים-נמוכים**, ולפיה:

- שיעור התאונות הצפוי באזור 2 (צפון) גדול פי 1.71 משיעור התאונות הצפוי באזור 3 (מרכז).

טבלאות 62 ו - 63 מציגות את תוצאות הניתוח של נפחי התנועה **הבינוניים-גבוהים**.

טבלה 60 - השוואת אזור 1 (דרום) אל מול אזורים 2 (צפון) ו-3 (מרכז) עבור נפח תנועה בינוני-גבוה (15000 ל- 30000). מקטעי דרך של 500 מ'

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-8.6569	2.6923	-13.9337	-3.3802	10.34	0.0013
AADT_log		1	0.8314	0.2727	0.2969	1.3658	9.30	0.0023
Ezor	**2**	**1**	**0.5038**	**0.1806**	**0.1498**	**0.8578**	**7.78**	**0.0053**
Ezor	**3**	**1**	**-0.3297**	**0.1920**	**-0.7061**	**0.0467**	**2.95**	**0.0860**
Dispersion		1	1.5378	0.1519	1.2401	1.8354		

מטבלה מספר 62, עולה כי **בנפחי התנועה הבינוניים-גבוהים**:

- שיעור התאונות הצפוי באזור 2 (צפון) גדול פי 1.65, משיעור התאונות הצפוי באזור 1 (דרום).
- שיעור התאונות הצפוי באזור 1 (דרום) גדול פי 1.39, משיעור התאונות הצפוי באזור 3 (מרכז).

טבלה 61 - השוואת אזור 2 (צפון) אל מול אזורים 1 (דרום) ו-3 (מרכז) עבור נפח תנועה בינוני-גבוה (15000 ל-30000). מקטעי דרך של 500 מ'

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-8.1532	2.7167	-13.4777	-2.8286	9.01	0.0027
AADT_log		1	0.8314	0.2727	0.2969	1.3658	9.30	0.0023
Ezor	1	1	-0.5038	0.1806	-0.8578	-0.1498	7.78	0.0053
Ezor	**3**	**1**	**-0.8335**	**0.1063**	**-1.0419**	**-0.6252**	**61.47**	**<.0001**
Dispersion		1	1.5378	0.1519	1.2401	1.8354		

טבלה מספר 63, משלימה את ההשוואה בין אזורים 2 (צפון) ו- 3 (מרכז) עבור **נפחי התנועה הבינוניים-גבוהים**, ולפיה:

- שיעור התאונות הצפוי באזור 2 (צפון) גדול פי 2.30 משיעור התאונות הצפוי באזור 3 (מרכז).

<u>לסיכום ניתוח מקטעי הדרך באורך 500 מ'</u>

עבור נפחי התנועה הבינוניים-נמוכים והבינוניים גבוהים נמצאה מגמה דומה בהשוואה בין האזורים 2 (צפון) ל- 1 (דרום), ובין האזורים 2 (צפון) ל- 3 (מרכז), כאשר:

- שיעור התאונות הצפוי באזור 2 (צפון) גדול משיעור התאונות הצפוי באזור 1 (דרום)
- שיעור התאונות הצפוי באזור 2 (צפון) גדול משיעור התאונות הצפוי באזור 3 (מרכז).

בהשוואת שיעור התאונות הצפוי עבור אזורים 3 (מרכז) ו- 1 (דרום), עבור **נפחי התנועה הבינוניים-נמוכים** ו**הבינוניים- גבוהים** נמצאה מגמה הפוכה.

בנפחי התנועה **הבינוניים-נמוכים** נמצא כי שיעור התאונות הצפוי באזור 3 (מרכז) גדול משיעור התאונות הצפוי באזור 1 (דרום), כאשר בנפחי התנועה **הבינוניים-גבוהים** נמצא כי שיעור התאונות הצפוי באזור 1 (דרום) גדול משיעור התאונות הצפוי באזור 3 (מרכז).

7.5.1.3 חלוקה תחבורתית – מקטעי דרך באורך של 750 מ'.

טבלאות 64 ו- 65 מציגות את תוצאות הניתוח עבור נפחי התנועה הבינוניים-נמוכים במקטעי הדרך באורך 750 מטרים.

טבלה 62 - השוואת אזור 1 (דרום) אל מול אזורים 2 (צפון) ו-3 (מרכז) עבור נפח בינוני-נמוך (5000 ל- 15000)

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-5.7367	1.6899	-9.0488	-2.4246	11.52	0.0007
AADT_log		1	0.5439	0.1855	0.1802	0.9076	8.59	0.0034
Ezor	**2**	**1**	**0.8391**	**0.1459**	**0.5532**	**1.1251**	**33.08**	**<.0001**
Ezor	**3**	**1**	**0.3209**	**0.1459**	**0.0349**	**0.6070**	**4.84**	**0.0278**
Dispersion		1	1.5593	0.1894	1.1880	1.9305		

טבלה 63 - השוואת אזור 2 (צפון) אל מול אזורים 1 (דרום) ו-3 (מרכז) עבור נפח בינוני-נמוך (5000 ל- 15000)

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-4.8976	1.7142	-8.2574	-1.5378	8.16	0.0043
AADT_log		1	0.5439	0.1855	0.1802	0.9076	8.59	0.0034
Ezor	1	1	-0.8391	0.1459	-1.1251	-0.5532	33.08	<.0001
Ezor	**3**	**1**	**-0.5182**	**0.1435**	**-0.7994**	**-0.2369**	**13.04**	**0.0003**
Dispersion		1	1.5593	0.1894	1.1880	1.9305		

מטבלאות 64 ו-65, ניתן לומר כי עבור נפחי **התנועה בינוניים-נמוכים**, **במקטעי הדרך באורך של 750 מ'**:

- שיעור התאונות הצפוי באזור 2 (צפון) גדול פי 2.31 משיעור התאונות הצפוי באזור 1 (דרום).
- שיעור התאונות הצפוי באזור 3 (מרכז) גדול פי 1.37 משיעור התאונות הצפוי באזור 1 (דרום).
- שיעור התאונות הצפוי באזור 2 (צפון) גדול פי 1.67 משיעור התאונות הצפוי באזור 3 (מרכז).

טבלאות 66 ו-67, מציגות את תוצאות הניתוח עבור נפחי **התנועה בינוניים-גבוהים**, **במקטעי הדרך באורך של 750 מ'**:

טבלה 64 - השוואת אזור 1 (דרום) אל מול אזורים 2 (צפון) ו-3 (מרכז) עבור נפח תנועה בינוני-גבוה (15000 ל-30000)

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-5.4058	3.3602	-11.9917	1.1800	2.59	0.1077
AADT_log		1	0.5503	0.3406	-0.1173	1.2179	2.61	0.1062
Ezor	2	1	**0.4275**	**0.2052**	**0.0254**	**0.8296**	**4.34**	**0.0372**
Ezor	3	1	**-0.4674**	**0.2211**	**-0.9006**	**-0.0341**	**4.47**	**0.0345**
Dispersion		1	1.3572	0.1575	1.0486	1.6658		

טבלה 65 - השוואת אזור 2 (צפון) אל מול אזורים 1 (דרום) ו-3 (מרכז) עבור נפח תנועה בינוני-גבוה (15000 ל-30000)

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-4.9783	3.3918	-11.6261	1.6694	2.15	0.1422
AADT_log		1	0.5503	0.3406	-0.1173	1.2179	2.61	0.1062
Ezor	1	1	-0.4275	0.2052	-0.8296	-0.0254	4.34	0.0372
Ezor	3	1	**-0.8949**	**0.1304**	**-1.1504**	**-0.6394**	**47.12**	**<.0001**
Dispersion		1	1.3572	0.1575	1.0486	1.6658		

מטבלאות 66 ו-67, ניתן לומר כי עבור נפחי **התנועה הנמוכים-בינוניים**, **במקטעי הדרך באורך של 750 מ'**:

- שיעור התאונות הצפוי באזור 2 (צפון) גדול פי 1. 53 משיעור התאונות הצפוי באזור 1 (דרום).
- שיעור התאונות הצפוי באזור 1 (דרום) גדול פי 1.59 משיעור התאונות הצפוי באזור 3 (מרכז).
- שיעור התאונות הצפוי באזור 2 (צפון) גדול פי 2.44 משיעור התאונות הצפוי באזור 3 (מרכז).

בסיכום ניתוח האזורים עבור מקטעי הדרך באורך של 750 מ', נמצאה מגמה דומה לתוצאות שהתקבלו בניתוח מקטעי הדרך באורך 500 מטרים, ולפיה, **עבור נפחי התנועה הבינוניים-נמוכים והבינוניים גבוהים נמצאה מגמה דומה** בהשוואה בין האזורים 2 (צפון) ל- 1 (דרום), ובין האזורים 2 (צפון) ל- 3 (מרכז), כאשר:

- שיעור התאונות הצפוי באזור 2 (צפון) גדול משיעור התאונות הצפוי באזור 1 (דרום)

- שיעור התאונות הצפוי באזור 2 (צפון) גדול משיעור התאונות הצפוי באזור 3 (מרכז).

בהשוואת שיעור התאונות הצפוי עבור אזורים 3 (מרכז) ו- 1 (דרום), עבור **נפחי התנועה הבינוניים-נמוכים** ו**הבינוניים- גבוהים** נמצאה מגמה הפוכה.
בנפחי התנועה **הבינוניים-נמוכים** נמצא כי שיעור התאונות הצפוי באזור 3 (מרכז) גדול משיעור התאונות הצפוי באזור 1 (דרום), כאשר בנפחי התנועה **הבינוניים-גבוהים** נמצא כי שיעור התאונות הצפוי באזור 1 (דרום) גדול משיעור התאונות הצפוי באזור 3 (מרכז).

7.5.2 אינטרקציות

אפשר ופרמטר מסויים, דוגמת עקמומיות, אינו משפיע לבדו על התרחשות אירוע של תאונת דרכים, אולם בשילוב עם משתנה אחר (דוגמת אזור), הוא עתיד להימצא כבעל משמעות במודל.
במחקר נבחנו בתחילה אינטרקציות בין אזור לעקמומיות. במקרה זה, נבחן האם העקמומיות הינה בעלת השפעה שונה על מספר התאונות באזורים שונים. כן נבחנה האינטראקציה בין אזור ושיפוע. במקרה זה נבדק, האם ניתן לאפיין מספר תאונות שונה עבור שיפועים באזורים שונים, עבור מערכי הנתונים באורכים של 500 ו- 750 מ'. אלה, לא יצאו כמובהקים ומשפיעים במודל, ולפיכך לא נבחנו שני מערכי הנתונים באורכים האחרים. בנוסף, נבחנו אינטרקציות בין נפח התנועה (AADT) לבין אזור. ניתוחים אלה מוצגים ומוסברים בפירוט בפסקאות 7.5.2.1-7.5.2.4.

7.5.2.1 ניתוח קטעי דרך בקבוצות AADT נמוך, גבוה ובינוני. משתנה תלוי AccCNT (500 מ').

אחת ההנחות שהתגבשו במהלך סקר הספרות היא כי בעבור AADT שונה, אפשר ויתקבלו תוצאות אחרות בהיבט ניתוח מספר אירועים למקטע דרך. היות וכך, נבחן שילוב נפח התנועה הקטגוריאלי, כאינטרקציה עם פרמטרים אחרים במודל (עקמומיות, אזור ונפח תנועה). במידה ותמצא אינטראקציה בין נפח התנועה לבין משתנה מסויים, אזי **המשמעות היא כי השפעת נפח התנועה על שיעור תאונות הדרכים הצפוי אינה אחידה, ותלויה במשתנה שנמצא כבעל אינטרקציה במודל.**

ניתוח זה בוצע על בסיס נתוני מקטעי הדרך באורך 500 מ', כאשר המשתנה התלוי הינו AccCNT, המייצג את מספר תאונות במקטע דרך. הניתוח בוצע על כלל החומר ללא חלוקה לקבוצת אימון וקבוצת ביקורת. זאת, היות וחלוקה לקבוצת אימון וביקורת ובנוסף חלוקה לקבוצות לפי נפח תנועה מוגדר, מקטינה באופן משמעותי כמות הנתונים בכל קבוצה ופוגמת בניתוח הסטטיסטי. בסיס הנתונים חולק לשלושה תתי-קבוצות בטווחי AADT שונים: 1/3 ערכי ה - AADT הנמוכים, 1/3 הבינוניים ו- 1/3 הערכים הגבוהים באופן הבא[25]:

[25] החלוקה בוצעה על עמודת נפח התנועה לאחר שעברה טרנספורמציה ללוג טבעי, כך ש:
9.28421→ 10766.6642
9.90515 →20033.2766

```
if
aadt_log<=9.28421 then aadt_cat=1 ;
else
if 9.90515=>aadt_log>9.28421 then aadt_cat=2 ;
else
if aadt_log>9.90515 then aadt_cat=3;
```

טבלה מספר 68 מציגה את ההשוואה בין ערכי ה- AIC של מודל ה- NULL (מודל המכיל בתוכו את המשתנים: עקמומיות, אזור ונפח תנועה, ללא משתנה המייצג אינטראקציה), אל מול שלושה מודלים בהם שולב גם משתנה המייצג אינטראקציה בין הפרמטר הנבחן לבין נפח תנועה קטגוריאלי. כך לדוגמא, ערך ה- AIC הרשום בטבלה 68 תחת משתנה העקמומיות (CurvePrm3), הינו תוצר של מודל, בו שולבו המשתנים המשמעותיים: עקמומיות, אזור ונפח תנועה, ובנוסף משתנה המייצג אינטראקציה בין עקמומיות לנפח תנועה קטגוריאלי. במידה וערך ה- AIC שיתקבל עבור מודל זה, יהיה קטן יותר מערך מודל ה- NULL, המשמעות היא כי האינטרקציה משפרת את המודל. כפי שניתן לראות בטבלה 68, ערך ה- AIC שהתקבל עבור המודל בו שולב המשתנה הקטגוריאלי המייצג את האינטראקציה בין עקמומיות (CurvePrm3) לנפח תנועה קטגוריאלי, עומד על 7199 לערך. תוצאה זו הינה גבוהה מערך ה- AIC של מודל ה- NULL העומד על 7195. המשמעות היא כי אינטרקציה זו אינה משפרת את המודל.

טבלה 66 - השוואה בין ערכי ה AIC של מודל ה NULL לבין מודל בו מתקיימות אינטרציות בין המשתנים השונים (חלוקת נפח תנועה נמוך, בינוני וגבוה)

AIC	Aadt_log	ezor	CurvePrm3	Null model
Aadt_cat	7195.9376	7171.5060	7199.2169	7195.2649

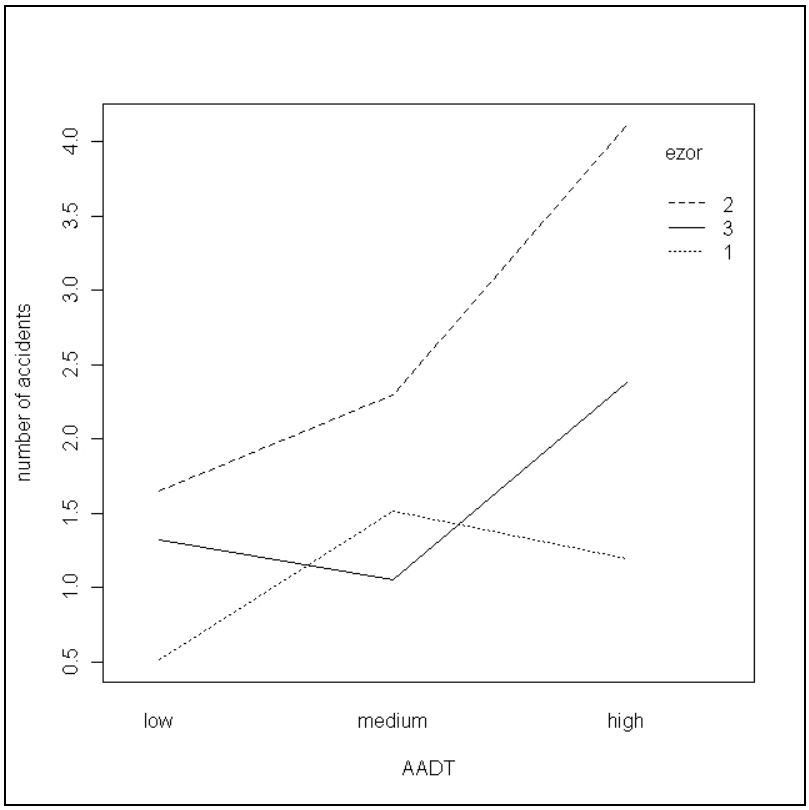
מטבלה מספר 68, עולה כי האינטרקציה בין משתנה ה"האזור" לבין נפח התנועה הקטגוריאלי, משפרת את המודל. מטבלה מספר 69, ניתן גם לראות כי האינטקרציה בין האזור לבין נפח התנועה הקטגוריאלי, היא מובהקת. **המשמעות היא כי השפעת נפח התנועה על שיעור תאונות הדרכים הצפוי אינה אחידה, ותלויה באזור בו אנו נמצאים.**

טבלה 67 - LR Statistics For Type 3 Analysis, מקטעי דרך של 500 מ'

LR Statistics For Type 3 Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
AADT_log	1	111.85	<.0001
aadt_cat	2	31.05	<.0001
Ezor	2	79.52	<.0001

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
CurvePrm3	1	15.17	<.0001
Ezor*aadt_cat	4	31.76	<.0001

הגרף המציג את האינטראקציה בין נפח התנועה הקטגוריאלי (נמוך, בינוני, גבוה) לבין שיעור התאונות באזורים השונים, מופיע באיור מספר 84. במידה ולא היתה מתקיימת כלל אינטראקציה בין נפח התנועה לאזור, היו מתקבלים שלושה קווים מקבילים.



איור מספר 84 - השפעת נפח תנועה קטגוריאלי, בחלוקה לאזורים, על מספר התאונות המשוער (מקטעי דרך באורך 500 מ')

ניתן לראות, עבור אזור **2 (אזור הצפון) ככל שנפח התנועה גדול יותר, אזי שיעור התאונות הצפוי גדול יותר.**

בהשוואת שיעור התאונות בין אזור 1 (דרום) לאזור 3 (מרכז) נמצא כי המגמות הפוכות. באזור 3 (מרכז), כאשר נפח התנועה נמוך שיעור התאונות הצפוי הוא בינוני, כאשר נפח התנועה בינוני מתקבל שיעור התאונות הצפוי הנמוך ביותר, וכאשר נפח התנועה גבוה מתקבל שיעור התאונות הצפוי הגבוה ביותר. באזור 1 (דרום), כאשר נפח התנועה נמוך מתקבל שיעור תאונות הצפוי הנמוך, בנפח התנועה בינוני מתקבל שיעור התאונות הצפוי הגבוה ביותר וכאשר נפח התנועה גבוה מתקבל שיעור תאונות צפוי בינוני.

בטווח נפח התנועה הנמוך, התוצאה לפיה באזור **הדרום צפויות פחות תאונות מאזור המרכז סבירה**. זאת, היות ועסקינן בטווח (כזכור, האמירה "נמוך" מייצגת טווח הקטן מ

10,000 כלי רכב). הווה אומר , כי בנפח תנועה "נמוך" באזור הדרום, אזי המשמעות היא כי כמעט ואין מכוניות. זאת, לעומת נפח תנועה "נמוך" במרכז הארץ (גם ביחס לכבישים אחרים במרכז), אשר בכל זאת מתנהלת בו תנועה. **בטווח נפח התנועה הגבוה באזור הדרום, ניכר כי קיים קושי** כפי שיוסבר בהמשך, לאחר שתוצג טבלה 69 המציגה את מספר התאונות הצפוי עבור מקטעי דרך בחלוקה לאזורים. יצויין, כי הסבר זה, בנוגע להשפעת טווח הנתונים, מתאים גם עבור נפח התנועה ה"גבוה". בבואנו להשוות בין אזור המרכז לדרום, **התוצאה לפיה מספר התאונות הצפוי במרכז, גבוה יותר מזה מהצפוי בדרום, סבירה. זאת, על רקע העובדה כי נפחי התנועה באזור המרכז גבוהים יותר מאשר באזור בדרום.**

<u>בהתייחס לאיור מספר 84 ולחישובים ידניים שבוצעו, טבלה מספר 70 מציגה את מספר התאונות המשוער עבור מקטעי דרך בחלוקה לאזורים: צפון (2), מרכז (3) ודרום (1):</u>

טבלה 68 - מספר תאונות צפוי עבור מקטעי דרך בחלוקה לאזורים (מקטעי דרך באורך 500 מ')

נפח תנועה	אזור	מספר תאונות משוער למקטע דרך
נמוך	דרום	0.51
בינוני	דרום	1.51
גבוה	דרום	1.19
נמוך	צפון	1.64
בינוני	צפון	2.29
גבוה	צפון	4.10
נמוך	מרכז	1.32
בינוני	מרכז	1.05
גבוה	מרכז	2.38

מטבלה 70 (נוסף על הקושי שצויין בניתוח הגרף המוצג באיור מספר 82), מסתמנת בעיתיות באזור 1 (דרום) בהשוואת נפחי התנועה ה"גבוהים" וה"בינוניים", כאשר עבור נפח תנועה "גבוה" מספר התאונות הממוצע המשוער נמוך יותר מזה הצפוי עבור נפח התנועה הבינוני (1.19 לעומת 1.51 למקטע דרך, בהתאמה). בדיקה של ערכי נפח התנועה המקסימאלים בכל קבוצה העלתה כי הניתוח שבוצע מוטה. נפח התנועה המקסימאלי באזור הדרום עומד על כ- 23000, כלומר הקבוצה ה"גבוהה" ששולבה בניתוח, כמעט ואינה מתקיימת בפועל באזור הדרום (72 רשומות בלבד), מה שמטה את תוצאות הניתוח.

קושי נוסף מסתמן באזור המרכז בנפחי התנועה הבינוניים והנמוכים. לכאורה, בהשואה בין מספר התאונות הצפוי בנפח תנועה "נמוך" לבין נפח תנועה "בינוני", היינו מצפים לקבל מספר תאונות צפוי גבוה יותר בנפחי התנועה ה"בינוניים". לפיכך נבדקו ערכי נפח התנועה הקיימים בכבישי אזור המרכז בקבוצות נפח התנועה ה"בינוני" ו"נמוך", ונמצא כי קבוצת המדגם תקינות. לאור האמור, נראה כי הדבר נגרם בשל טווח החלוקה לקבוצות. באזור המרכז, קיימים לא מעט מקטעי דרך בבסיס הנתונים (כשישית, 106 רשומות מתוך 596)

השייכים לקבוצת נפח התנועה הנמוך והמאופיינים בערך AADT הגדול מ- 10000. אי לכך, אפשר והקירבה לגבול שבין הקבוצות, יוצרת ההטייה בתוצאות.

על מנת שנוכל להמשיך ולעשות שימוש בניתוח של הפילוג שתואר להלן, יש להמנע מלנתח את תוצאות הקבוצה המוטה. קרי, לא ינותחו תוצאות ערכי נפח התנועה ה"גבוהים" באזור הדרום (1). תחת ההנחות וההסתיגויות שצויינו לעיל, המשיך הניתוח.

בשלב הבא, על מנת להשוות בין האינטרקציות השונות באזורים השונים, נבחן תחילה האם תוצרי המודל המשווה בין האינטרקציות השונות בנפחי התנועה והאזורים השונים, אכן מובהקים. במידה ויעלו כמובהקים, יבחן פי כמה גדול או קטן שיעור התאונות. כך לדוגמא, טבלה מספר 71 מציגה את תוצר המודל המשווה בין האינטרקציה של נפח תנועה נמוך ואזור 1 (דרום) לבין האינטרקציות של נפחי התנועה השונים באזורים השונים, עבור קטעי דרך באורך 500 מ' (המספר המופיע בצבע סגול - מסמל את ה"אזור", והמספר המופיע בצבע הירוק מסמל את "נפח התנועה הקטגוריאלי"[26]).

טבלה 69 - תוצר מודל האינטרקציה. השוואת נפח תנועה נמוך (1) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	2	2	1	-0.6397	0.2168	-1.0645	-0.2148	8.71	0.0032
Ezor*aadt_cat	2	3	1	0.0467	0.3002	-0.5417	0.6350	0.02	0.8764
Ezor*aadt_cat	3	2	1	-1.0540	0.2302	-1.5052	-0.6027	20.96	<.0001
Ezor*aadt_cat	3	3	1	-0.6039	0.2955	-1.1831	-0.0247	4.18	0.0410

ניתן לראות כי השוואת האינטרקציה בין נפח התנועה הנמוך (1) באזור הדרום (1) לבין נפח התנועה הגבוה (3) באזור הצפון (2) אינה מובהקת. כאשר השוואת האינטראקציה בין נפח התנועה הנמוך (1) באזור הדרום (1) לבין נפח תנועה בינוני (2) באזור הצפון (2) מובהקת ולפיכך ניתנת להשוואה. גם השוואת האינטראקציה בין נפח התנועה הנמוך (1) באזור הדרום (1) לבין האינטרקציה בין אזור המרכז (3) ונפח תנועה בינוני (2) מובהקת וניתנת להשוואה.

טבלאות 72-77 משלימות את השוואות מודלי האינטרקציה השונים לפי הפירוט בכותרת.

[26] כזכור, אזור 1=דרום אזור 2=צפון ואזור 3=מרכז. כאשר במודל המוצג, נפח תנועה נמוך=1, בינוני=2 וגבוה=3.

טבלה 70 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	2	1	1	0.6397	0.2168	0.2148 1.0645	8.71	0.0032
Ezor*aadt_cat	2	3	1	0.6864	0.2858	0.1262 1.2465	5.77	0.0163
Ezor*aadt_cat	3	1	1	1.0540	0.2302	0.6027 1.5052	20.96	<.0001
Ezor*aadt_cat	3	3	1	0.4500	0.2973	-0.1327 1.0328	2.29	0.1301

טבלה 71 - תוצר מודל השוואת האינטרקציות - נפח תנועה גבוה (3) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	2	1	1	-0.0467	0.3002	-0.6350 0.5417	0.02	0.8764
Ezor*aadt_cat	2	2	1	-0.6864	0.2858	-1.2465 -0.1262	5.77	0.0163
Ezor*aadt_cat	3	1	1	0.6039	0.2955	0.0247 1.1831	4.18	0.0410
Ezor*aadt_cat	3	2	1	-0.4500	0.2973	-1.0328 0.1327	2.29	0.1301

טבלה 72 - תוצר מודל השוואת האינטרקציות - נפח תנועה נמוך (1) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	2	1	0.6397	0.2168	0.2148 1.0645	8.71	0.0032
Ezor*aadt_cat	1	3	1	-0.0467	0.3002	-0.6350 0.5417	0.02	0.8764
Ezor*aadt_cat	3	2	1	-0.4143	0.2021	-0.8103 -0.0182	4.20	0.0403
Ezor*aadt_cat	3	3	1	-0.6506	0.1857	-1.0146 -0.2867	12.28	0.0005

טבלה 73 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	1	1	-0.6397	0.2168	-1.0645 -0.2148	8.71	0.0032
Ezor*aadt_cat	1	3	1	-0.6864	0.2858	-1.2465 -0.1262	5.77	0.0163
Ezor*aadt_cat	3	1	1	0.4143	0.2021	0.0182 0.8103	4.20	0.0403
Ezor*aadt_cat	3	3	1	-0.2363	0.1744	-0.5782 0.1055	1.84	0.1754

טבלה 74 - תוצר מודל השוואת האינטרקציות - נפח תנועה נמוך (1) באזור 3 (מרכז) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	2	1	1.0540	0.2302	0.6027	1.5052	20.96	<.0001
Ezor*aadt_cat	1	3	1	0.6039	0.2955	0.0247	1.1831	4.18	0.0410
Ezor*aadt_cat	2	2	1	0.4143	0.2021	0.0182	0.8103	4.20	0.0403
Ezor*aadt_cat	2	3	1	0.6506	0.1857	0.2867	1.0146	12.28	0.0005

טבלה 75 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 3 (מרכז) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 500 מ'

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	1	1	-1.0540	0.2302	-1.5052	-0.6027	20.96	<.0001
Ezor*aadt_cat	1	3	1	-0.4500	0.2973	-1.0328	0.1327	2.29	0.1301
Ezor*aadt_cat	2	1	1	-0.4143	0.2021	-0.8103	-0.0182	4.20	0.0403
Ezor*aadt_cat	2	3	1	0.2363	0.1744	-0.1055	0.5782	1.84	0.1754

מטבלאות 72-77, ניתן לראות כי תוצרי האינטרקציות הבאות נמצאו כלא מובהקים, ולפיכך לא ניתן לבצע השוואה בינהם:

דרום נפח תנועה נמוך אל מול צפון גבוה ; דרום נפח תנועה בינוני אל מול מרכז נפח תנועה גבוה ; דרום נפח תנועה גבוה אל מול צפון נפח תנועה נמוך ; דרום נפח תנועה גבוה אל מול מרכז נפח תנועה בינוני ; צפון נפח תנועה בינוני אל מול מרכז נפח תנועה גבוה ; מרכז נפח תנועה בינוני אל מול דרום נפח תנועה גבוה ; מרכז נפח תנועה בינוני אל מול צפון נפח תנועה גבוה.

כך ניתן להשוות בין האינטראקציות באותו האזור, כמפורט להלן:

באזור הצפון (2)

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "גבוה", גדול פי 2.49 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "נמוך".
- שיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "גבוה" גדול פי 1.78 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "בינוני".

באזור המרכז (3):

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "גבוה", גדול פי 1.8 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "נמוך".

- שיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "גבוה" גדול פי 2.25 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "בינוני".

<u>באזור הדרום (1):</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "בינוני", גדול פי 2.95 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "נמוך".

נשווה ונבחן את שיעור התאונות הצפוי עבור אותה קטגוריית נפח תנועה באזורים שונים:

<u>בקטגוריית נפח תנועה "גבוה":</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "גבוה" באזור הצפון (2), גדול פי 1.72 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "גבוה" באזור המרכז (3).

<u>בקטגוריית נפח תנועה "בינוני":</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "בינוני" באזור הצפון (2), גדול פי 2.17 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "בינוני" באזור המרכז (3).
- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "בינוני" באזור הצפון (2), גדול פי 1.51 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "בינוני" באזור הדרום (1).
- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "בינוני" באזור הדרום (1), גדול פי 1.43 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "בינוני" באזור מרכז (3).

<u>בקטגוריית נפח התנועה הנמוך:</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "נמוך" באזור הצפון (2), גדול פי 1.24 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "נמוך" באזור המרכז (3),
- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "נמוך" באזור הצפון (2), גדול פי 3.21 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "נמוך" באזור הדרום (1),

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "נמוך" באזור מרכז (3), גדול פי 2.57 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "נמוך" באזור הדרום (1),

7.5.2.2 ניתוח קטעי דרך לפי קבוצות AADT נמוך, גבוה ובינוני, משתנה תלוי AccCNT - 750 מ'.

פסקה זו תציג את תוצאות ניתוח האינטרקציה עבור מקטעי הדרך באורך 750 מ'. גם ניתוח זה בוצע על כלל החומר ללא חלוקה לקבוצת אימון וקבוצת ביקורת, זאת היות וחלוקה לקבוצת אימון וביקורת ובנוסף חלוקה לקבוצות לפי נפח תנועה מוגדר, מקטינה באופן משמעותי כמות הנתונים בכל קבוצה ופוגמת בניתוח הסטטיסטי.

בדומה לניתוח שבוצע על מקטעי הדרך באורך 500 מטרים, גם בניתוח זה, חולק בסיס הנתונים לשלוש תתי-קבוצות בטווחי AADT שונים: 1/3 ערכי ה - AADT הנמוכים, 1/3 הבינוניים ו 1/3 הערכים הגבוהים[27] באופן הבא:

```
if
aadt_log<=9.28421 then aadt_cat=1 ;
else
if 9.90515=>aadt_log>9.28421 then aadt_cat=2 ;
else
if aadt_log>9.90515 then aadt_cat=3;
```

טבלה מספר 78 מציגה את ההשוואה בין ערכי ה- AIC של מודל ה- NULL (מודל המכיל בתוכו את המשתנים: עקמומיות, אזור ונפח תנועה, <u>ללא</u> משתנה המייצג אינטראקציה), אל מול שלושה מודלים בהם שולב גם משתנה המייצג אינטראקציה בין הפרמטר הנבחן לבין נפח תנועה קטגוריאלי. כך לדוגמא, ערך ה- AIC הרשום בטבלה 78 תחת משתנה העקמומיות (CurvePrm3), הינו תוצר המודל בו שולבו המשתנים המשמעותיים: עקמומיות, אזור ונפח תנועה, ובנוסף משתנה המייצג אינטראקציה בין עקמומיות לנפח תנועה קטגוריאלי. במידה וערך ה- AIC שהתקבל במודל זה יהיה קטן יותר מערך מודל ה- NULL, ניתן יהיה לומר כי האינטרקציה משפרת את המודל. כמו שניתן לראות בטבלה 78, ערך ה- AIC שהתקבל עבור המודל בו שולב גם המשתנה הקטגוריאלי המייצג את האינטראקציה בין עקמומיות לנפח תנועה קטגוריאלי, עומד על 4890 לערך. תוצאה זו הינה גבוהה מערך ה- AIC של מודל ה NULL, קרי, האינטרקציה הזו, אינה משפרת את המודל.

[27] החלוקה בוצעה על עמודת נפח התנועה לאחר שעברה טרנספורמציה ללוג טבעי, כך ש:
9.28421→ 10766.6642
9.90515 →20033.2766

טבלה 76 - השוואה בין ערכי ה AIC של מודל ה NULL לבין מודל בו מתקיימות אינטרציות בין המשתנים השונים (חלוקת נפח תנועה נמוך, בינוני וגבוה) – 750 מ'

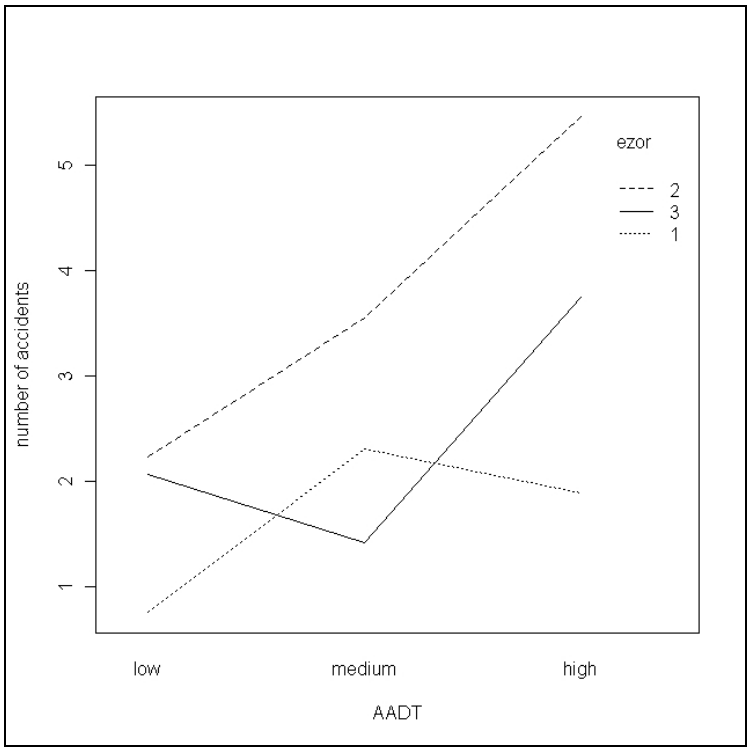
AIC	Aadt_log	ezor	CurvePrm3	Null model
Aadt_cat	488	4869.8530	4890.4683	4887.6420

מטבלה מספר 78, ניתן לראות כי האינטרקציה בין משתנה ה"האזור" לבין נפח התנועה הקטגוריאלי, משפרת את המודל. מטבלה מספר 79 ניתן לראות, כי האינטקרציה בין האזור לבין נפח התנועה הקטגוריאלי, היא מובהקת . **המשמעות היא, כי עבור מקטעי דרך של 750 מ' (בדומה לתוצאות שהתקבלו עבור מקטעי הדרך באורך 500 מ'), השפעת נפח התנועה על שיעור תאונות הדרכים, אינה אחידה ותלויה באזור בו אנו נמצאים.**

טבלה 77 - LR Statistics For Type 3 Analysis**, מקטעי דרך של 750 מ'**

LR Statistics For Type 3 Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
Ezor	2	16.31	0.0003
CurvePrm3	1	8.34	0.0039
aadt_cat	2	40.15	<.0001
Ezor*aadt_cat	4	31.56	<.0001

הגרף המציג את האינטראקציה בין נפח התנועה הקטגוריאלי (נמוך, בינוני, גבוה) לבין שיעור התאונות באזורים השונים עבור מקטעי דרך באורך של 750 מ', מופיע באיור מספר 85. במידה ולא היתה מתקיימת כלל אינטראקציה בין נפח התנועה לאזור, היינו מקבלים שלושה קווים מקבילים.



איור מספר 85 - אינטקראציה בין נפח תנועה קטגוריאלי לבין אזור (מקטע דרך 750 מ')

<u>בהתייחס לאיור מספר 85 ולחישובים ידניים שבוצעו, טבלה מספר 80 מציגה את מספר התאונות המשוער עבור מקטעי דרך בחלוקה לאזורים: צפון (2), מרכז (3) ודרום (1):</u>

טבלה 78 - מספר תאונות צפוי עבור מקטעי דרך בחלוקה לאזורים (מקטעי דרך באורך של 750 מ')

נפח תנועה	אזור	מספר תאונות משוער למקטע דרך
נמוך	דרום	0.74
בינוני	דרום	2.3
גבוה	דרום	1.89
נמוך	צפון	2.22
בינוני	צפון	3.55
גבוה	צפון	5.45
נמוך	מרכז	2.06
בינוני	מרכז	1.41
גבוה	מרכז	3.74

כפי שניתן לראות מאיור מספר 83 וטבלה 80, המגמה דומה מאוד לזו שהתקבלה עבור מקטעי הדרך באורך 500 מ' (איור מספר 82 וטבלה מספר 70), אם כי מוקצנת יותר. גם בניתוח זה **באזור 2 (אזור הצפון) ככל שנפח התנועה גדול יותר, אזי שיעור התאונות הצפוי גדול יותר**.

בטווח נפח התנועה הנמוך, התוצאה לפיה באזור **הדרום צפויות פחות תאונות מאזור המרכז נראית סבירה**. שוב, היות ועסקינן בטווח ("נמוך" מייצג טווח הקטן מ 10,000 כלי

רכב). הווה אומר, כי כאשר מגדירים נפח תנועה "נמוך" באזור הדרום, אזי המשמעות היא כי בפועל על הכביש כמעט ואין מכוניות. זאת, לעומת נפח תנועה "נמוך" במרכז הארץ (גם ביחס לכבישים אחרים במרכז), אשר בכל זאת מתנהלת בהם תנועה. שוב, יצויין כי הסבר זה, בנוגע להשפעת טווח הנתונים, מתאים גם עבור נפח התנועה ה"גבוה". בבואנו להשוות בין אזור המרכז לדרום, **התוצאה לפיה מספר התאונות הצפוי במרכז, גבוה יותר מזה מהצפוי בדרום, סבירה. זאת, על רקע העובדה כי טווח נפחי התנועה באזור המרכז גבוהים יותר מאשר באזור בדרום.**

באזור 1 (דרום) מספר התאונות הממוצע המשוער בנפח התנועה הגבוה, נמוך יותר מזה הצפוי עבור נפח התנועה הבינוני (1.89 לעומת 2.3 למקטע דרך, בהתאמה). בדיקה של ערכי נפח התנועה המקסימאלים בכל קבוצה העלתה כי הניתוח שבוצע מוטה. נפח התנועה המקסימאלי באזור הדרום עומד על כ- 23000, כלומר הקבוצה ה"גבוהה" ששולבה בניתוח, שוב כמעט ואינה מתקיימת בפועל באזור הדרום (45 רשומות בלבד), מה שכנראה מטה את תוצאות הניתוח.

קושי נוסף מסתמן באזור <u>המרכז</u> בנפחי התנועה ה"בינוניים" וה"נמוכים". לכאורה, בהשואה בין מספר התאונות הצפוי בנפח תנועה "נמוך" לבין נפח תנועה "בינוני", היינו מצפים לקבל מספר תאונות צפוי גבוה יותר בנפחי התנועה ה"בינוניים". לפיכך, נבדקו ערכי נפח התנועה הקיימים בכבישי אזור המרכז בקבוצות נפח התנועה הבינוני ונמוך, ונמצא כי קבוצת המדגם תקינות. לאור האמור, אפשר וגם כאן, ההסבר טמון בגין טווח החלוקה לקבוצות. והעובדה, לפיה קיימים מקטעי דרך הקרובים לגבול הטווחים שבין הקבוצות, משפיעה על התוצאות.

על מנת שנוכל להמשיך ולעשות שימוש בניתוח של הפילוג שתואר להלן, יש להמנע מלנתח את תוצאות הקבוצה המוטה. <u>קרי, לא ינותחו תוצאות ערכי נפח התנועה הגבוהים באזור הדרום</u>. תחת ההנחות וההסתיגויות שצויינו לעיל, המשיך הניתוח.

בשלב הבא, על מנת להשוות בין האינטרקציות השונות באזורים השונים, נבחן תחילה האם תוצרי המודל המשווה בין האינטרקציות השונות בנפחי התנועה והאזורים השונים – אכן מובהקים. במידה ויעלו כמובהקים, נבחן פי כמה גדול או קטן שיעור התאונות. כך לדוגמא, טבלה מספר 81 מציגה את תוצר המודל המשווה בין האינטרקציה של נפח תנועה נמוך (1) ואזור 1 (דרום) לבין האינטרקציות של נפחי התנועה השונים באזורים השונים, עבור קטעי

דרך באורך 750 מ' (המספר המופיע בצבע סגול - מסמל את ה"אזור", והמספר המופיע בצבע הירוק מסמל את "נפח התנועה הקטגוריאלי"[28]).

טבלה 79 - תוצר מודל האינטרקציה. השוואת נפח תנועה נמוך (1) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	2	2	1	-0.5402	0.2528	-1.0357 -0.0447	4.57	0.0326
Ezor*aadt_cat	2	3	1	-0.0347	0.3380	-0.6972 0.6277	0.01	0.9181
Ezor*aadt_cat	3	2	1	-1.2268	0.2693	-1.7545 -0.6991	20.76	<.0001
Ezor*aadt_cat	3	3	1	-0.7323	0.3288	-1.3767 -0.0880	4.96	0.0259

ניתן לראות כי השוואת האינטרקציה בין נפח התנועה הנמוך (1) באזור הדרום (1) לבין נפח התנועה הגבוה (3) באזור הצפון (2) אינה מובהקת. כאשר השוואת האינטראקציה בין נפח התנועה הנמוך (1) באזור הדרום (1) לבין נפח תנועה בינוני (2) באזור הצפון (2) מובהקת ולפיכך ניתנת להשוואה. גם השוואת האינטראקציה בין נפח התנועה הנמוך (1) באזור הדרום (1) לבין האינטרקציה בין אזור המרכז (3) ונפח תנועה בינוני (2) מובהקת וניתנת להשוואה.

טבלאות 82-85 משלימות את השוואות מודלי האינטרקציה השונים לפי הפירוט בכותרת.

טבלה 80 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	2	1	1	0.5402	0.2528	0.0447 1.0357	4.57	0.0326
Ezor*aadt_cat	2	3	1	0.5055	0.3212	-0.1241 1.1351	2.48	0.1156
Ezor*aadt_cat	3	1	1	1.2268	0.2693	0.6991 1.7545	20.76	<.0001
Ezor*aadt_cat	3	3	1	0.4945	0.3381	-0.1682 1.1571	2.14	0.1436

טבלה 81 - תוצר מודל השוואת האינטרקציות - נפח תנועה גבוה (3) באזור 1 (דרום) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	2	1	1	0.0347	0.3380	-0.6277 0.6972	0.01	0.9181
Ezor*aadt_cat	2	2	1	-0.5055	0.3212	-1.1351 0.1241	2.48	0.1156
Ezor*aadt_cat	3	1	1	0.7323	0.3288	0.0880 1.3767	4.96	0.0259
Ezor*aadt_cat	3	2	1	-0.4945	0.3381	-1.1571 0.1682	2.14	0.1436

[28] כזכור, אזור 1=דרום אזור 2=צפון ואזור 3=מרכז. כאשר במודל המוצג, נפח תנועה נמוך=1, בינוני=2 וגבוה=3.

טבלה 82 - תוצר מודל השוואת האינטרקציות - נפח תנועה נמוך (1) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	2	1	0.5402	0.2528	0.0447	1.0357	4.57	0.0326
Ezor*aadt_cat	1	3	1	0.0347	0.3380	-0.6277	0.6972	0.01	0.9181
Ezor*aadt_cat	3	2	1	-0.6866	0.2491	-1.1749	-0.1983	7.59	0.0059
Ezor*aadt_cat	3	3	1	-0.6976	0.2285	-1.1455	-0.2497	9.32	0.0023

טבלה 83 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	1	1	-0.5402	0.2528	-1.0357	-0.0447	4.57	0.0326
Ezor*aadt_cat	1	3	1	-0.5055	0.3212	-1.1351	0.1241	2.48	0.1156
Ezor*aadt_cat	3	1	1	0.6866	0.2491	0.1983	1.1749	7.59	0.0059
Ezor*aadt_cat	3	3	1	-0.0110	0.2169	-0.4361	0.4141	0.00	0.9595

טבלה 84 - תוצר מודל השוואת האינטרקציות - נפח תנועה גבוה (3) באזור 2 (צפון) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'.

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	1	1	-0.0347	0.3380	-0.6972	0.6277	0.01	0.9181
Ezor*aadt_cat	1	2	1	0.5055	0.3212	-0.1241	1.1351	2.48	0.1156
Ezor*aadt_cat	3	1	1	0.6976	0.2285	0.2497	1.1455	9.32	0.0023
Ezor*aadt_cat	3	2	1	0.0110	0.2169	-0.4141	0.4361	0.00	0.9595

טבלה 85 - תוצר מודל השוואת האינטרקציות - נפח תנועה נמוך (1) באזור 3 (מרכז) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	2	1	1.2268	0.2693	0.6991	1.7545	20.76	<.0001
Ezor*aadt_cat	1	3	1	0.7323	0.3288	0.0880	1.3767	4.96	0.0259
Ezor*aadt_cat	2	2	1	0.6866	0.2491	0.1983	1.1749	7.59	0.0059
Ezor*aadt_cat	2	3	1	0.6976	0.2285	0.2497	1.1455	9.32	0.0023

טבלה 86 - תוצר מודל השוואת האינטרקציות - נפח תנועה בינוני (2) באזור 3 (מרכז) אל מול נפחי התנועה השונים באזורים השונים – מקטעי דרך באורך 750 מ'.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Ezor*aadt_cat	1	1	1	-1.2268	0.2693	-1.7545	-0.6991	20.76	<.0001
Ezor*aadt_cat	1	3	1	-0.4945	0.3381	-1.1571	0.1682	2.14	0.1436
Ezor*aadt_cat	2	1	1	-0.6866	0.2491	-1.1749	-0.1983	7.59	0.0059
Ezor*aadt_cat	2	3	1	0.0110	0.2169	-0.4141	0.4361	0.00	0.9595

מטבלאות 82-88, ניתן לראות כי תוצרי האינטרקציות הבאות מצאו כלא מובהקים, ולפיכך <u>לא ניתן</u> לבצע השוואה בינהם:

אזור הדרום - נפח תנועה נמוך ואזור הצפון – נפח תנועה גבוה

אזור הדרום - נפח תנועה בינוני ואזור הצפון – נפח תנועה גבוה

אזור הדרום - נפח תנועה בינוני ואזור מרכז – נפח תנועה גבוה

אזור הדרום - נפח תנועה גבוה ואזור צפון – נפח תנועה נמוך

אזור הדרום - נפח תנועה גבוה ואזור צפון – נפח תנועה בינוני

אזור הדרום - נפח תנועה גבוה ואזור מרכז – נפח תנועה בינוני

אזור צפון - נפח תנועה נמוך ואזור דרום – נפח תנועה גבוה

אזור צפון - נפח תנועה בינוני ואזור דרום – נפח תנועה גבוה

אזור צפון - נפח תנועה בינוני ואזור מרכז – נפח תנועה גבוה

כך ניתן להשוות בין האינטראקציות באותו האזור, כמפורט להלן:

<u>באזור הצפון (2)</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה **"גבוה"**, גדול פי 2.45 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה **"נמוך"**.
- שיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "**גבוה"** גדול פי 1.53 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה **"בינוני".**

<u>באזור המרכז (3):</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה **"גבוה"**, גדול פי 1.81 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה **"נמוך"**.
- שיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה **"גבוה"** גדול פי 2.64 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה **"בינוני"**.

<u>באזור הדרום (1):</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "בינוני", גדול פי 3.08 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "נמוך".

נשווה ונבחן את שיעור התאונות הצפוי עבור אותה קטגוריית נפח תנועה באזורים שונים:

<u>בקטגוריית נפח תנועה "גבוה" (נשווה כאמור רק את אזור הצפון והמרכז):</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים **בנפח תנועה "גבוה" באזור הצפון (2)**, גדול פי 1.45 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם **נפח תנועה "גבוה" באזור המרכז (3)**.

<u>בקטגוריית נפח תנועה "בינוני":</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים **בנפח תנועה "בינוני" באזור הצפון (2)**, גדול פי 2.50 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם **נפח תנועה "בינוני" באזור המרכז (3)**.
- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים **בנפח תנועה "בינוני" באזור הצפון (2)**, גדול פי 1.53 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם **נפח תנועה "בינוני" באזור הדרום (1)**.
- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים **בנפח תנועה "בינוני" באזור הדרום (1)**, גדול פי 1.62 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם **נפח תנועה "בינוני" באזור מרכז (3)**.

<u>בקטגוריית נפח התנועה ה"נמוך":</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים **בנפח תנועה "נמוך" באזור הצפון (2)**, גדול פי 1.07 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם **נפח תנועה "נמוך" באזור המרכז (3)**.
- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים **בנפח תנועה "נמוך" באזור הצפון (2)**, גדול פי 2.97 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם **נפח תנועה "נמוך" באזור הדרום (1)**.
- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים **בנפח תנועה "נמוך" באזור מרכז (3)**, גדול פי 2.76 משיעור התאונות הצפוי להתרחש במקטעי דרך בהם **נפח תנועה "נמוך" באזור הדרום (1)**.

7.5.2.3 ניתוח קטעי דרך לפי לפי קבוצות AADT בהגדרה תחבורתית – 500 מ'

כאמור, אחת ההנחות שהתגבשו במהלך סקר הספרות היא כי בעבור AADT שונה, אפשר ויתקבלו תוצאות אחרות בהיבט ניתוח מספר אירועים למקטע דרך. היות וכך, נבחן שילוב נפח התנועה הקטגוריאלי, כאינטרקציה במודל. באופן המוצע, נוכל לנתח השפעת פרמטרים שונים על מספר התאונות, בעבור כל קבוצת AADT.

ההבדל בין הניתוח הנוכחי, לזה שבוצע בפיסקה הקודמת, הינו בחלוקה לפי לוגיקה תחבורתית (כבישים מסדר 1, 2 3 ו - 4). בסיס הנתונים חולק לארבע תתי-קבוצות לפי נפח תנועה המתאים לטווחים התחבורתיים כמפורט בטבלה מספר 89. הניתוח בוצע על כלל החומר, בכדי להביא לכך שתהינה כמה שיותר תצפיות בכל קבוצה.

טבלה 87 - מספר תצפיות ששולבו במודל עבור טווח נפחי תנועה שונים

טווח AADT	מספר תצפיות
קטן מ 5000	77
5000-15000	1641
15000-30000	1216
גדול 30000	388

טבלה מספר 90 מציגה את ההשוואה בין ערכי ה- AIC של מודל ה NULL (המודל ששולבו בו המשתנים המשמעותיים: אזור, עקמומיות ונפח תנועה, **ללא** איטרקציות), אל מול המודל ששולבו בו המשתנים המשמעותיים ובנוסף בכל פעם משתנה (אחר) המייצג אינטראקציה בין אותו משתנה לבין נפח התנועה הקטגוריאלי. במידה וערך ה AIC שהתקבל יהיה קטן יותר במודל, הרי שניתן לומר כי האינטרקציה משפרת את המודל.

טבלה 88 - השוואה בין ערכי ה AIC של מודל ה NULL לבין מודל בו מתקיימות אינטרציות בין המשתנים השונים (חלוקת קבוצות נפח תנועה לפי התפלגות תחבורתית)

AIC	Aadt_log	ezor	CurvePrm3	Null model
Aadt	7203.5068	7203.8751	7207.9921	7203.2828

כפי שניתן לראות מטבלה 90 אף אחת מהאינטראקציות שנבחנו לא טייבה את המודל. המשמעות היא, שלפי חלוקה זו, לא ניתן לאפיין ולומר כי קיימת השפעה שונה של פרמטר נפח התנועה על שיעור התאונות באזורים השונים, או לומר כי קיימת השפעה שונה של פרמטר נפח התנועה על שיעור התאונות כתלות בעקמומיות.

7.5.2.4 ניתוח קטעי דרך לפי לפי קבוצות AADT בהגדרה תחבורתית – 750 מ'

גם בניתוח זה, חולק בסיס הנתונים לארבע תתי-קבוצות לפי נפח תנועה המתאים לטווחים התחבורתיים שפורטו בסעיף הקודם, כאשר הניתוח בוצע על כלל החומר, בכדי להביא לכך שתהינה כמה שיותר תצפיות בכל קבוצה.

טבלה מספר 91 מציגה את ההשוואה בין ערכי ה AIC של מודל ה NULL (המודל ששולבו בו המשתנים המשמעותיים: אזור, עקמומיות ונפח תנועה, **ללא** איטרקציות), אל מול המודל ששולבו בו המשתנים המשמעותיים ובנוסף בכל פעם משתנה (אחר) המייצג אינטראקציה בין אותו משתנה לבין נפח התנועה הקטגוריאלי. במידה וערך ה- AIC שהתקבל יהיה קטן יותר במודל, הרי שניתן לומר כי האינטרקציה משפרת את המודל.

טבלה 89 - השוואה בין ערכי ה AIC של מודל ה NULL לבין מודל בו מתקיימות אינטרציות בין המשתנים השונים (חלוקת קבוצות נפח תנועה לפי התפלגות תחבורתית)- 750 מ'

AIC	Aadt_log	ezor	CurvePrm3	Null model
AADT	4892.9045	4893.6708	4896.1859	4892.1044

שוב בדומה לתוצאה של מקטעי הדרך באורך 750 מ', וכפי שניתן לראות מטבלה 91, אף אחת מהאינטראקציות שנבחנו לא טייבה את המודל. המשמעות היא, שלפי חלוקה זו, לא ניתן לאפיין ולומר כי קיימת השפעה שונה של פרמטר נפח התנועה על שיעור התאונות באזורים השונים, או לומר כי קיימת השפעה שונה של פרמטר נפח התנועה על שיעור התאונות כתלות בעקמומיות.

7.5.3 ניתוח תתי איזורים באזור הצפון

כאמור, היות ובמודלים שתוארו בסעיפים הקודמים עלה בצורה עקבית, כי אזור הצפון הינו בעל שיעור תאונות צפוי הגבוה מהאזורים האחרים, בוצע ניתוח ייעודי עבור אזור זה, המשווה את שיעור התאונות בין תתי האזורים. כזכור (בהתאם למוצג באיור מספר 37), האזורים בצפון הינם: 21 = חיפה ; 22 = עמקים ; 23 = גליל תחתון מערבי ; 24 = גליל תחתון מזרחי; 25 = גליל עליון מערבי ; 26 = גליל עליון מזרחי; 27 = אצבע הגליל; 28 = רמת הגולן - לא שולב בניתוח בגין מיעוט תאונות

הניתוח בוצע על בסיס המידע המלא (ללא חלוקה ל- 2/3 ול- 1/3) על מנת שניתן יהיה להתבסס על כמה שיותר דגימות. יודגש כי אזור 28 (רמת הגולן) הושמט מהניתוח, היות ומספר אירועי תאונות הדרכים שבו היה נמוך בצורה יוצאת דופן (2 אירועים בלבד).

טבלה מספר 92, מציגה כדוגמא, את תוצאות השוואת תת איזור 21 לשאר האזורים בצפון. המינוס בעמודת ה Estimate (לדוגמא בטבלה מספר 92 – בשורה של TAT_Ezor 22),

משמעו, ששיעור התאונות באזור המושווה (אזור 21, חיפה, במקרה זה) גבוה יותר משיעור התאונות באזור 22 (עמקים). בשאר השורות בטבלה זו, עמודת ה- Estimate חיובית. לפיכך ניתן לומר, כי שיעור התאונות באזורים אלה, גבוה יותר משיעור התאונות באזור חיפה. בכדי לדעת פי כמה, יש להעלות את e בחזקת המקדם המופיע תחת עמודת Estimate.

קיימת כמובן גם חשיבות למובהקות התוצאה. בדוגמא שהוצגה לעיל ערך ה Pr > ChiSq עבור שורת TAT_Ezor מספר 22 מראה כי התוצאה איננה מובהקת. מכאן שבסיכומו של דבר, לא ניתן לומר ששיעור התאונות באזור 21, חיפה, גבוה יותר משיעור התאונות באזור 22 העמקים).

טבלה 90 - השוואת תת אזור 21 (חיפה) לתת האיזורים האחרים בצפון

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-7.1348	1.0951	-9.2812	-4.9883	42.44	<.0001
AADT_log		1	0.7110	0.1097	0.4960	0.9260	42.01	<.0001
TAT_Ezor	**22**	**1**	**-0.1589**	**0.1363**	**-0.4261**	**0.1082**	**1.36**	**0.2436**
TAT_Ezor	23	1	0.3083	0.1221	0.0690	0.5476	6.38	0.0116
TAT_Ezor	24	1	0.0682	0.1620	-0.2493	0.3858	0.18	0.6737
TAT_Ezor	25	1	0.6306	0.1564	0.3241	0.9372	16.26	<.0001
TAT_Ezor	26	1	0.0448	0.2508	-0.4467	0.5364	0.03	0.8581
TAT_Ezor	27	1	0.3436	0.1921	-0.0329	0.7201	3.20	0.0736
CurvePrm3		1	1.2703	0.4241	0.4391	2.1015	8.97	0.0027
Dispersion		1	1.0395	0.1077	0.8285	1.2505		

היות ובוצעו השוואות מרובות על התוצרים, נעשה שימוש בשיטת linear step-up procedure (להלן: שיטת BH) לניתוח השוואות מרובות (FALSE DISCOVERY RATE להלן: FDR). ניתוח זה מאפשר לאתר מצבי false positives. במצבים מסוג זה, מתקבלות תוצאות מובהקות, למרות שבפועל התוצאות אינן מובהקות. טבלה מספר 93 מציגה את תוצאות ניתוח ה - FDR, בו מתקיים אחוז false positives שאינו עולה על 5% .

טבלה 91 - תוצאות ניתוח ההשוואת המרובות בין תתי האזורים בצפון

TAT_EZOR_1	TAT_EZOR_2	Pr > ChiSq	criteria		conclusion
21	25	0.0001	0.002381	1	significant
22	25	0.0001	0.004762	2	significant
22	23	0.0018	0.007143	3	significant
24	25	0.0045	0.009524	4	significant
21	23	0.0116	0.011905	5	significant
22	27	0.015	0.014286	6	not significant
25	26	0.0341	0.016667	7	not significant
23	25	0.0545	0.019048	8	not significant
21	27	0.0736	0.021429	9	not significant

23	24	0.161	0.02381	10	not significant
24	27	0.1865	0.02619	11	not significant
25	27	0.203	0.028571	12	not significant
22	24	0.2032	0.030952	13	not significant
21	22	0.2436	0.033333	14	not significant
26	27	0.2984	0.035714	15	not significant
23	26	0.307	0.038095	16	not significant
22	26	0.4387	0.040476	17	not significant
21	24	0.6737	0.042857	18	not significant
21	26	0.8581	0.045238	19	not significant
23	27	0.8609	0.047619	20	not significant
24	26	0.9305	0.05	21	not significant

בטבלה 93 ניתן לראות, כי קיימת שונות אמיתית בין האזורים שסומנו כמשמעותיים (significant). כך, ניתן לומר כי קיים שוני משמעותי ואמיתי (סטטיסטית), בין אזור 21 (חיפה) לבין האזורים 25 (גליל עליון מערבי) ו- 23 (גליל תחתון מערבי). בנוסף, ניתן לראות כי קיימת שונות אמיתית בין אזור 22 (עמקים) לאזורים 25 (גליל עליון מערבי) ו- 23 (גליל תחתון מערבי). כמו כן , קיימת שונות בין אזור 24 (גליל תחתון מזרחי) לאזור 25 (גליל עליון מערבי).

על כן תוצגנה גם טבלאות 94 ו- 95, המכילות את תוצאות השוואת תת איזור 22 (עמקים) לשאר האזורים בצפון, ותת איזור 24 (גליל תחתון מזרחי) לשאר האזורים בצפון, בהתאמה.

טבלה 92 - השוואת תת אזור 22 (עמקים) לתת האיזורים האחרים בצפון

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-7.2937	1.0845	-9.4192	-5.1682	45.23	<.0001
AADT_log		1	0.7110	0.1097	0.4960	0.9260	42.01	<.0001
TAT_Ezor	21	1	0.1589	0.1363	-0.1082	0.4261	1.36	0.2436
TAT_Ezor	23	1	0.4672	0.1496	0.1740	0.7605	9.75	0.0018
TAT_Ezor	24	1	0.2272	0.1785	-0.1227	0.5771	1.62	0.2032
TAT_Ezor	25	1	0.7896	0.1789	0.4389	1.1402	19.48	<.0001
TAT_Ezor	26	1	0.2038	0.2632	-0.3120	0.7196	0.60	0.4387
TAT_Ezor	27	1	0.5026	0.2067	0.0975	0.9076	5.91	0.0150
CurvePrm3		1	1.2703	0.4241	0.4391	2.1015	8.97	0.0027
Dispersion		1	1.0395	0.1077	0.8285	1.2505		

טבלה 93 - השוואת תת אזור 24 (גליל תחתון מזרחי) לתת האיזורים האחרים בצפון

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-7.0665	1.0349	-9.0950	-5.0381	46.62	<.0001
AADT_log		1	0.7110	0.1097	0.4960	0.9260	42.01	<.0001
TAT_Ezor	21	1	-0.0682	0.1620	-0.3858	0.2493	0.18	0.6737
TAT_Ezor	22	1	-0.2272	0.1785	-0.5771	0.1227	1.62	0.2032
TAT_Ezor	23	1	0.2401	0.1713	-0.0956	0.5757	1.97	0.1610

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
TAT_Ezor	25	1	0.5624	0.1978	0.1747	0.9501	8.09	0.0045
TAT_Ezor	26	1	-0.0234	0.2684	-0.5494	0.5026	0.01	0.9305
TAT_Ezor	27	1	0.2754	0.2085	-0.1332	0.6840	1.75	0.1865
CurvePrm3		1	1.2703	0.4241	0.4391	2.1015	8.97	0.0027
Dispersion		1	1.0395	0.1077	0.8285	1.2505		

מבחינת מהות השוני בטבלאות 61, 63-64, ניתן לגזור את המשמעויות, כמפורט:

- שיעור התאונות באזור 25 (גליל עליון מערבי) גבוה פי 1.36 משיעור התאונות באזור 21 (חיפה)
- שיעור התאונות באזור 23 (גליל תחתון מערבי), גבוה פי 1.87 משיעור התאונות באזור 21 (חיפה).
- שיעור התאונות באזור 23 (גליל תחתון מערבי) גבוה פי 1.59 משיעור התאונות באזור 22 (עמקים).
- שיעור התאונות באזור 25 (גליל עליון מערבי) גבוה פי 2.2 משיעור התאונות באזור 22 (עמקים)
- שיעור התאונות באזור 25 (גליל עליון מערבי) גבוה פי 1.75 משיעור התאונות באזור 24 (גליל תחתון מזרחי).

7.5.4 רגרסיה לוגיסטית

7.5.4.1 מקטעי דרך באורך 500 מ'

בניתוח זה נעשה ניסיון למדל את ההסתברות להתרחשות אפס תאונות או לפחות תאונה אחת במקטע דרך. הניתוח בוצע על מודל אימון (2/3 הנתונים), מתוך מחשבה כי במידה והמודל יניב תוצאות טובות, יבנה מודל ביקורת עבור 1/3 הנתונים. טבלה מספר 96 מציגה את נתוני האמת, כלומר מספר מקטעי הדרך בהם כלל לא התרחשו תאונות (726) אל מול מקטעי הדרך בהם התרחשה לפחות תאונה אחת (381), עבור סט הנתונים באורך 500 מ'.

טבלה 94 - מקטעי הדרך בהם כלל לא התרחשו תאונות (726) אל מול מקטעי הדרך בהם התרחשה לפחות תאונה אחת (381)– נתוני אמת

Response Profile		
Ordered Value	acccnt_dih	Total Frequency
1	0	726
2	1	381

בתחילה נבנה מודל ראשוני ללא משתנים מסבירים. עבור מודל זה חושב ערך AIC (להלן: ערך ה - AIC המוגדר כ NULL). לאחר מכן, נבנה מודל ייחודי עבור כל אחד מהמשתנים המסבירים, ונבחנה תוספת המשתנה למודל. על מנת שמשתנה מסביר יהווה גורם בעל משמעות במודל לאחר ששולב בו, על ערך ה AIC של המודל, להיות נמוך יותר מערך ה- AIC של מודל ה- NULL הראשוני.

טבלה מספר 97 מציגה את הפרמטרים ששולבו במודל ואת השפעתם (לפי ערך ה AIC): עקמומיות (CurvePrm3), אזור (ezor), נפח תנועה (AADT_log), שיפוע (VarSlp) ומהירות משוערכת (SPEED). כפי שניתן לראות, המשתנים המסבירים ששולבו במודל, והניבו ערך AIC, הנמוך יותר ממודל ה- NULL הראשוני הינם "נפח תנועה" (AADT_log) ו"אזור" (ezor).

טבלה 95 - ערכי AIC לפרמטרים שנבחנו עבור בניית מודל רגרסיה לוגיסטית

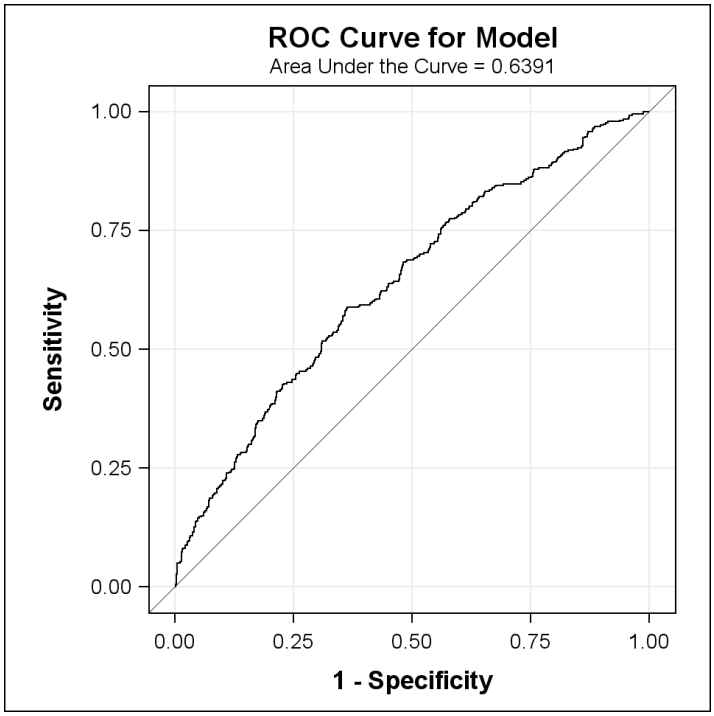
AIC	500 m
AADT_log	1405.299
CurvePrm3	1427.354
ezor	1393.266
VarSlp	1429.255
SPEED	1427.45
AADT_log , ezor	1373.233
NULL	**1427.296**

כאמור, בניתוח זה נעשה נסיון למדל את ההסתברות להתרחשות אפס תאונות או לפחות תאונה אחת במקטע דרך. איור מספר 79 מציג אחד מתוצרי הניתוח - עקומת ROC **(receiver operating characteristic), המציגה רגישות אל מול אחד פחות סגוליות. נסביר את המונחים.**

רגישות Sensitivity – מייצגת את ההסתברות ל"לפחות תאונה אחת", במקטעים בהם התרחשה לפחות תאונה אחת

סגוליות Specificity - מייצגת את ההסתברות ל"אפס תאונות" במקטעים בהם לא התרחשה שם אף תאונה.

המטרה הינה למקסם את שני הפרמטרים הנ"ל, כאשר הקו הישר בגרף (ב - 45°) מייצג ניבוי אקראי לגמרי (של 50 אחוזים). ככל שעקומת הגרף תעבור כמה שיותר קרוב לפינה השמאלית עליונה, אזי מתקיים הניבוי הטוב ביותר, היות ואז ה- "Sensitivity" וה- "1-Specificity" הינם בעלי ערך מקסימאלי. כלל השטח שמעל לגרף מסתכם בערך 1, כאשר ככל שהשטח מתחת לעקומה גדול יותר, כך המודל מנבא טוב יותר. במודל טוב השטח מתחת לעקומה צריך להיות קרוב ל-1.



איור מספר 86 - רגרסיה לוגיסטית, גרף ה ROC עבור מקטעי דרך באורך 500 מ'

כפי שניתן לראות באיור מספר 86, הגרף רחוק מהפינה השמאלית העליונה וקרוב מאוד לקו 45° (המסמן אקראיות), כאשר השטח מתחת לגרף עומד על 0.63 בלבד. כזכור, שטח של 0.5 מסמן כי הניבוי מהמודל הוא אקראי. המשמעות היא כי מודל הרגרסיה הלוגיסטית, אינו מנבא טוב מספיק את ההסתברות לקיומם של 0 תאונות או לפחות תאונה אחת, במקטעי הדרך באורך 500 מטרים. התוצאה מצביעה על מודל כמעט אקראי לגמרי.

7.5.4.2 מקטעי דרך באורך 750 מ'

כאמור, בניתוח זה ניסינו לבחון את האפשרות למדל את ההסתברות להתרחשות אפס תאונות או לפחות תאונה אחת עבור מקטעי דרך באורך 750מ'. הכוונה היתה לבחון האם תתקבל הסתברות שאינה אקראית כפי שעלה במודל שנבנה עבור מקטעי דרך באורך של 500 מטרים.

הניתוח בוצע על מודל אימון (2/3 הנתונים), כאשר טבלה מספר 98 מציגה את נתוני האמת, קרי, מספר מקטעי הדרך בהם כלל לא התרחשו תאונות, אל מול מקטעי הדרך בהם התרחשה לפחות תאונה אחת, עבור מקטעי הדרך באורך 750 מ'.

טבלה 96 - מקטעי הדרך בהם כלל לא התרחשו תאונות (361) אל מול מקטעי הדרך בהם התרחשה לפחות תאונה אחת (249)– נתוני אמת. מקטעי דרך של 750 מ'

Response Profile		
Ordered Value	acccnt_dih	Total Frequency
1	0	361
2	1	249

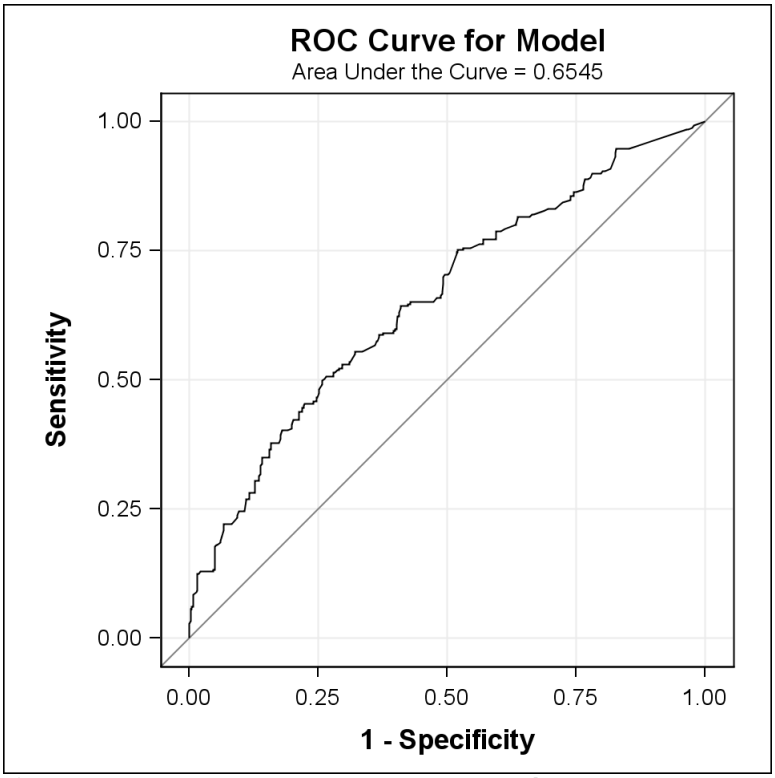
כאמור, בתחילה נבנה מודל ראשוני ללא משתנים מסבירים. עבור מודל זה חושב ערך AIC (להלן: ערך ה - AIC המוגדר כ- NULL). לאחר מכן, נבנה מודל ייחודי עבור כל אחד מהמשתנים המסבירים, ונבחנה תוספת המשתנה למודל. על מנת שמשתנה מסביר יהווה גורם בעל משמעות במודל לאחר ששולב בו, על ערך ה- AIC של המודל, להיות נמוך יותר מערך ה AIC של מודל ה NULL הראשוני.

טבלה מספר 99 מציגה את הפרמטרים ששולבו במודל ואת השפעתם (לפי ערך ה- AIC): עקמומיות (CurvePrm3), אזור (ezor), נפח תנועה (AADT_log), שיפוע (VarSlp) ומהירות משוערכת (SPEED). כפי שניתן לראות, המשתנים מסבירים ששולבו במודל, והניבו ערך AIC, הנמוך יותר ממודל ה- NULL הראשוני הינם נפח תנועה ואזור.

טבלה 97 - ערכי AIC לפרמטרים שנבחנו עבור בניית מודל רגרסיה לוגיסטית.
מקטעי דרך 750 מ'

AIC	500 m
AADT_log	818.494
CurvePrm3	827.968
ezor	793.439
VarSlp	827.689
SPEED	828.958
AADT_log , ezor	789.190
NULL	826.958

איור מספר 87, מציג את תוצאות עקומת ROC.



איור מספר 87 - רגרסיה לוגיסטית, גרף ה ROC עבור מקטעי דרך באורך 750 מ'

כפי שניתן לראות באיור מספר 85 (בדומה לממצא של מקטעי הדרך באורך 500 מ'), הגרף רחוק מהפינה השמאלית העליונה וקרוב מאוד לקו 45 מעלות (המסמן אקראיות), כאשר השטח מתחת לגרף עומד על 0.65 בלבד. כזכור, שטח של 0.5 מסמן כי הניבוי מהמודל הוא אקראי. המשמעות היא כי מודל הרגרסיה הלוגיסטית, אינו מנבא טוב מספיק את ההסתברות לקיומם של 0 תאונות או לפחות תאונה אחת, במקטעי הדרך באורך 750 מטרים. גם כאן, התוצאה מצביעה על מודל כמעט אקראי לגמרי.

היות ובבחינת הרגרסיה הלוגיסטית, לא נמצאה ישימות עבור מקטעי הדרך באורכים של 500 ו- 750 מטרים, הרי שעבור מקטעי הדרך באורכים 1000 ו 1500 לא נבנו מודלים מסוג זה.

8. תוצרים וסיכום

פרק זה מטרתו לרכז את עיקרי הממצאים, לאור ריבוי הניתוחים במהלך העבודה.

כאמור, במהלך המחקר נבחנו ארבעה מערכי נתונים, הנבדלים מבחינת אורך המקטע. בראש ובראשונה, מתוך רצון לבחון האם ניתן להצביע על "מקטעים אדומים" בעלי אורך "פיזי" קצר במקום להצביע על "כבישים אדומים". כך, ניתן יהיה לטפל ולהשקיע ביחידות קטנות לקבלת תפוקות גדולות יחסית לתשומות. בנוסף, במחקר זה, על גישתו המרחבית, בשימוש ב- GIS, נעשה נסיון לגרום לכך שפרמטר האורך לא יעמוד כפרמטר בפני עצמו, לאור העובדה כי "אורך הקטע" עלה במחקרים רבים כמשתנה בעל משמעות במודל.

טבלה מספר 100 מציגה את **סיכום תוצאות המודלים** שנבנו עבור ארבעת מערכי הנתונים (מקטעי דרך על אורכיהם השונים: 500 מ', 750 מ', 100 מ' ו- 1500 מ'). ארבעת המודלים שנבנו, יושוו בשני מישורים. הראשון, יציבות המודל. קרי איזה הוא המודל שכלל בדיקות הביקורת וה cross validation עבורו, היו העיקביים ביותר. השני, ברמת התוצאות והתובנות מכל מודל.

טבלה 98 - תוצאות מסכמות נבחרות של מודלי הרגרסיה הבינומית השלילית, עבור ארבע סטי הנתונים שנבחנו במחקר

אורך סגמנט	מודל פיזור מתאים	מודל אימון (2/3 הנתונים): פרמטרים משמעותיים	מודל ביקורת (1/3 הנתונים):[29] בטווח רווח סמך מודל האימון+הערות	TOST	פרמטרים משמעותיים במודל ביקורת מלא (1/3 הנתונים)	הערות ותובנות
500	בינומי שלילי	נפח תנועה	כן	שווה	זהים למודל האימון. אותו מודל עקומיות לא מובהקת (לפי AIC). בטווח רווח הסמך אזור 3 מובהק. לא בטווח רווח הסמך	כאשר נפח התנועה יגדל ביחידה אחת (דהיינו פי 2.7) ,שיעור התאונות יגדל פי 2.04.
		אזור (1)	כן			▪ בצפון שיעור התאונות הצפוי, הגדול פי 1.97 משיעור התאונות הצפוי בדרום. ▪ שיעור התאונות הצפוי בצפון, גדול פי 1.76 משיעור התאונות הצפוי באזור המרכז.
		אזור (3)	לא			
		עקומיות	כן, אבל לא מובהקת			עליה ביחידת עקומיות, משמעה כי שיעור התאונות יעלה בכמעט פי 5 (4.9275).
750	בינומי שלילי	נפח תנועה	לא, אבל מובהק	שווה	זהים למודל האימון. אותו מודל AADT לא בטווח רווח הסמך	כאשר נפח התנועה יגדל ביחידה אחת (דהיינו פי 2.7) , שיעור התאונות יגדל פי 1.9.
		אזור (1)	כן			▪ באזור הצפון שיעור התאונות הצפוי, גדול פי 2.07 משיעור התאונות הצפוי בדרום. ▪ בצפון שיעור התאונות הצפוי גדול פי 1.73 משיעור התאונות בצפוי במרכז.
		אזור (3)	כן			
		עקומיות	כן			עליה ביחידת עקומיות, משמעה כי שיעור התאונות יעלה פי 2.92
1000	בינומי שלילי	נפח תנועה	כן	שווה	נפח תנועה אזור (1) אזור (3)-לא מובהק. ולא בטווח רווח הסמך של מודל האימון אין עקומיות	במידה ויגדל נפח התנועה ביחידה אחת, דהיינו פי 2.7, שיעור התאונות יגדל פי 1.9.
		אזור (1)	כן			▪ בצפון קיים שיעור תאונות צפוי הגדול פי 1.83 משיעור התאונות הצפוי בדרום. ▪ שיעור תאונות הצפוי בצפון גדול פי 1.82 משיעור התאונות הצפוי באזור המרכז.
		אזור (3)	לא, ולא מובהק			
		עקומיות	כן, אבל לא מובהק			עליה ביחידת עקומיות, משמע כי שיעור התאונות במקטע דרך יעלה פי 2.5.
1500	בינומי	נפח תנועה	כן	לא	נפח תנועה	במידה ויגדל נפח התנועה ביחידה אחת,

[29] כאמור, מראש, שולבו במודל הביקורת החלקי אותם הפרמטרים שנמצאו כמשמעותיים במודל האימון (שנבנה על 2/3 מהנתונים). ולפיכך, נציין בטבלה זו, האם הליך ה CROSS VALIDATION , בקרת האיכות, עלה כתקין כן או לא.

	שלילי			**שווה**	**אזור (1)**	**דהיינו פי 2.7, , שיעור התאונות יגדל פי 2.1**
		אזור (1)	**כן**		**אזור (3) – לא מובהק אין עקמומיות**	▪ **באזור הצפון שיעור התאונות הצפוי גדול פי 1.91 משיעור התאונות הצפוי בדרום.** ▪ **שיעור תאונות הצפוי בצפון גדול פי 1.80 משיעור התאונות הצפוי באזור המרכז.**
		אזור (3)	**כן, אבל לא מובהק**			
		עקמומיות	**לא. לא מובהק**			**כאשר נעלה ביחידה אחת של עקמומיות, מספר התאונות במקטע דרך יעלה פי 4.5.**

שני המודלים שהתגלו כיציבים ביותר, בבחינת כלל הפרמטרים ובקרות האיכות, הינם המודלים של מקטעי הדרך באורך 500 מ' ו- 750 מ'. מבין שני המודלים הללו, המודל שנבנה עבור מקטעי הדרך באורך 500 המטרים, הראה את הרגישות הגבוהה ביותר לשינוי **בעקמומיות**, כאשר עליה ביחידת עקמומיות הגדילה באופן הגבוה ביותר את מספר התאונות הצפוי (ממצא התואם גם את ה"שכל הישר", שהרי ככל שמקטע כביש קצר יותר מאופיין בערך עקמומיות גבוה יותר, סביר כי העקומות תהינה חדות יותר). בנוסף, גילה מודל זה את הרגישות הגבוהה ביותר לגידול **בנפח התנועה**. כאשר גידול של יחידה אחת (דהיינו נפח תנועה הגדול פי 2.7, מנפח התנועה הראשוני שנבחן) צפוי כי יוביל לשיעור תאונות הגדול פי 2.04 משיעור התאונות הראשוני. כך לדוגמא, במקטע דרך בו נפח תנועה ממוצע של 1000 מכוניות ו- 3 תאונות בתקופת המחקר, במצב בו נפח התנועה יגדל ביחידה אחת (דהיינו פי 2.7) ויגיע ל- 2700 מכוניות, שיעור התאונות יגדל פי 2.04, וצפוי להגיע ל- 6.12 תאונות.

המודל שלאורך כלל הבדיקות הראה את היציבות **הנמוכה ביותר** הינו זה שנבנה עבור מקטעי הדרך באורך 1500 מ'. אפשר שמודל זה, הינו הכי פחות יציב, בגין העובדה, שמספר התצפיות ששולבו במודל נמוך ממספר התצפיות במודלים האחרים. אפשר ומספר התצפיות הנמוך, מסביר גם את העובדה שהעקמומיות במקטעי הדרך באורך 1500 אינה מובהקת במודל הביקורת. מודל שהיה מעט יותר יציב מקודמו (אולם עדין בעל יציבות נמוכה) הינו מודל מקטעי הדרך באורך 1000 מ'. בשני המודלים הללו, במודל הביקורת המלא (שנבנה כזכור על 1/3 מהנתונים) "נשמט" פרמטר העקמומיות ולא נמצא כבעל משמעות במודל.

הממצא, לפיו מקטעי הדרך באורכים בקצרים יותר שנבחנו (500 מ' ו 750 מ') התגלו כיציבים ביותר, משמעו כי ניתן, להמליץ על התייחסות והשקעה (הן בתשתית והן באכיפה) במקטעים ממוקדים. זאת להבדיל מההתייחסות הרווחת של שיפור מקטע ארוך יחסית של הכביש.

נוסחה מספר 89, לניבוי תוחלת תאונות הדרכים μ עבור מקטעי הדרך באורך **500** מטרים מוצגת להלן:

[89]

$$\mu = e^{-7.091+0.714*AADT\_Log+(-0.678)*Ezor1+(0.569)*Ezor3+(1.594)*Curve\Pr m3}$$

נוסחה מספר 90 מציגה את נוסחת ניבוי תוחלת תאונות הדרכים עבור מקטעי דרך באורך של **750** מ', להלן:

[90]

$$\mu = e^{-5.9688+0.6420*AADT\_Log+(-0.7316)*Ezor1+(-0.5483)*Ezor3+(1.0741)*CurvePrm3}$$

כאשר:

aadt_log - ממוצע נפח תנועה יומי לימי חול, לאחר שעבר טרנספורמציה ללוגריתם טבעי.

ezor - משתנה האזור. **ezor1** = דרום, **ezor3** = מרכז ו- **ezor2** =צפון. המודל משווה ומנתח בין אזור הצפון לשני האזורים האחרים. כך, עבור מקטע המצוי באזור 2 יוצב הערך 0 במשתנים **ezor3** ו- **ezor1**.

CurvePrm3 – עקמומיות, בהתאם לחישוב המוצג בנוסחה מספר 76.

בהיבט **ניתוח משתנה ה"אזור"**, בוצעו ניתוחים עבור מספר היבטים. הניתוח **הראשון** התקיים במסגרת בניית מודל האימון. פרמטר זה עלה כמשמעותי בארבעת המודלים, כפי שנבנו עבור מקטעי הדרך באורכים השונים. **בכל ארבעת המודלים נמצא כי שיעור התאונות בצפון צפוי להיות גבוה יותר (פי 2 לערך) משיעור התאונות במרכז ובדרום** (הפרוט המדוייק מופיע בטבלה מספר 100). יודגש, כי ניתוח זה הינו הכללי מבין ניתוחי האזורים האחרים, כפי שיפורטו בהמשך.

בניתוח השני נבחנו הבדלים הסטטיסטיים בין האזורים: צפון, מרכז ודרום, ללא תלות בנפח התנועה. ניתוח זה, בשונה מניתוח האינטראקציות (השלישי, שנדון בו בהמשך) מספק הסבר ה"נקי" מהשפעת משתנה נפח התנועה, וזאת משני טעמים. הראשון, בשל העובדה כי בסיס הנתונים חולק (בחלוקה תחבורתית) לנפחי תנועה בטווחים:"נמוך", "בינוני-נמוך", "בינוני-גבוה" ו"גבוה". השני, נפח התנועה שולב אף הוא כמשתנה במודל. כך, **בסיכומו של ניתוח זה, ניתן היה לגבש אמירה, ולפיה: באזור מסויים צפויות להתקיים יותר (או פחות) תאונות דרכים, <u>ללא תלות בנפח התנועה</u>**. כזכור עקב הצורך לשמור על קבוצות מאוזנות עבור הניתוח הסטטיסטי, נותחו הקבוצות בהן היה מדגם נתונים מספק - קבוצת **ה"בינוני-נמוך" (5000 ל- 15,000)** וקבוצת **ה"בינוני-גבוה" (15,000 ל- 30,000).**

בניתוח זה הסתמנו מגמות דומות עבור מקטעי הדרך באורך 500 ו- 750 מטרים, ולפיהן, בנפחי התנועה ה"**בינוניים-נמוכים" וה"בינוניים-גבוהים"** נמצא כי:

- **שיעור התאונות הצפוי באזור 2 (צפון) גדול משיעור התאונות הצפוי באזור 1 (דרום)**
- **שיעור התאונות הצפוי באזור 2 (צפון) גדול משיעור התאונות הצפוי באזור 3 (מרכז).**

מגמה זו תואמת גם את התוצאות שהתקבלו בניתוח האזורים הראשון, שבוצע כאמור בבניית מודל האימון. ולפיו שיעור התאונות הצפוי באזור הצפון גבוה מהצפוי באזור המרכז והדרום.

ממצא מעניין נוסף, מתקבל בעת השוואת שיעור התאונות הצפוי עבור אזורים 3 (מרכז) ו- 1 (דרום). זאת, מכיוון שעבור **נפחי התנועה הבינוניים-נמוכים** ו**הבינוניים- גבוהים** נמצאה מגמה הפוכה, ולפיה:

- בנפחי התנועה **הבינוניים-נמוכים** נמצא כי **שיעור התאונות הצפוי באזור 3 (מרכז) גדול משיעור התאונות הצפוי באזור 1 (דרום).** נראה כי ההסבר לכך, טמון בעובדה לפיה באזור הדרום כאשר עסקינן בשיעור תאונות "בינוני-נמוך", הרי שמדובר בעיקרו בנפחי התנועה הנמוכים יותר, קרי מדובר בכבישים בהם התנועה דלילה מאוד (כאשר הלוגיקה לפיה, היכן שאין רכבים הרי שההסתברות לתאונה יורדת בצורה דרסטית, מתקיימת). זאת, בהשוואה לאזור המרכז, כאשר בשטחו מדובר בנפחי התנועה הבינוניים יותר מתוך קבוצת ה"בינוני-נמוך", כך שמתקיימת תנועה.
- בנפחי התנועה **הבינוניים-גבוהים** נמצא כי **שיעור התאונות הצפוי באזור 1 (דרום) גדול משיעור התאונות הצפוי באזור 3 (מרכז).** ההסבר המשוער לכך, הינו שעם הגידול בנפח התנועה באה לידי ביטוי איכות התשתית (לרבות כניסות רבות שאינן מוסדרות, המבוצעות ישירות לכבישים המהירים), כאשר באזור הדרום איכות התשתית ירודה יותר מאשר באזור המרכז.

ניתוח שלישי בהיבט האזורים היה **ניתוח אינטרקציות**, שבוצע על מקטעי דרך באורך 500 מ' ו- 750 מ', שחולקו לשלוש קבוצות: טווח "נמוך" בו ערכי נפח התנועה מאופיינים בערכים הנעים בין 0 ל- 10766 ; טווח "בינוני" – ערכי נפח התנועה נעים בין 10767 לבין 20033; וטווח "גבוה" - ערכי נפח התנועה הגבוהים מ- 20033.

בסיום הבדיקה, נמצאה כי קיימת **אינטרקציה בין "אזור" לבין "נפח התנועה"**. עבור שני מערכי הנתונים שנבחנו. **המשמעות היא כי השפעת נפח התנועה על שיעור תאונות הדרכים הצפוי, אינה אחידה, ותלויה באזור בו אנו נמצאים.**

מגמות הניתוח עבור מקטעי הדרך באורך 500 מ' ו- 750 מ' התקבלו דומות (אם כי הערכים, כמובן שונים). טבלה מספר 101 מרכזת את תוצאות הניתוח עבור מקטעי הדרך באורכים השונים.

טבלה 99 - אינטרקציות – מספר תאונות צפוי עבור מקטעי דרך בחלוקה לאזורים (מקטעי דרך באורך 500 מ' ו-750מ')

סט נתונים	מספר תאונות צפוי למקטע דרך	אזור	נפח תנועה
500 מ'	0.51	דרום	נמוך
	1.51	דרום	בינוני
	1.19	דרום	גבוה
	1.64	צפון	נמוך
	2.29	צפון	בינוני
	4.10	צפון	גבוה
	1.32	מרכז	נמוך
	1.05	מרכז	בינוני
	2.38	מרכז	גבוה
750 מ'	0.74	דרום	נמוך
	2.3	דרום	בינוני

	1.89	דרום	גבוה
	2.22	צפון	נמוך
	3.55	צפון	בינוני
	5.45	צפון	גבוה
	2.06	מרכז	נמוך
	1.41	מרכז	בינוני
	3.74	מרכז	גבוה

<u>מגמות ניתוח בחלוקה לאזורים וקטגוריות "נפח תנועה", עבור 500 ו- 750 מ', יצאו אף הן דומות, ומפורטות להלן:</u>

<u>באזור הצפון</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "גבוה", גדול משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "נמוך".
- שיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "גבוה" גדול משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "בינוני".

<u>באזור המרכז</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה **"גבוה"**, גדול משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה **"נמוך"**.
- שיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה **"גבוה"** גדול משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה **"בינוני"**.

<u>באזור הדרום</u>

- שיעור התאונות הצפוי להתרחש במקטעי דרך המאופיינים בנפח תנועה "**בינוני**", גדול משיעור התאונות הצפוי להתרחש במקטעי דרך בהם נפח תנועה "**נמוך**".

<u>בקטגוריית נפח תנועה "גבוה":</u>

- שיעור התאונות הצפוי להתרחש באזור **הצפון**, גדול משיעור התאונות הצפוי באזור **המרכז**- ממצא המתאים באופן מלא לתוצאות הניתוחים שתוארו בשלבים הקודמים.

<u>בקטגוריית נפח תנועה "בינוני":</u>

- שיעור התאונות הצפוי להתרחש באזור **הצפון**, **גדול** משיעור התאונות הצפוי באזור **המרכז**.
- שיעור התאונות הצפוי להתרחש באזור **הצפון**, **גדול** משיעור התאונות הצפוי באזור **הדרום**.
- שיעור התאונות הצפוי להתרחש באזור **הדרום**, **גדול** משיעור התאונות הצפוי באזור **המרכז** – שוב ממצא בעל מגמה דומה לזו שהתקבלה בניתוח השני שבוצע (בו כזכור, נבחנו הבדלים הסטטיסטיים בין האזורים ללא תלות בנפח התנועה). גם כאן, ההסבר המשוער לכך הינו, הפער באיכות התשתית, כאשר באזור הדרום איכות התשתית ירודה יותר מאשר באזור המרכז (לרבות כניסות רבות שאינן מוסדרות, המבוצעות ישירות לכבישים המהירים בדרום).

בקטגוריית נפח התנועה ה"נמוך":

- שיעור התאונות הצפוי להתרחש באזור הצפון, גדול משיעור התאונות הצפוי באזור המרכז – ממצא המתאים לתוצאות הניתוחים שתוארו בשלבים הקודמים.
- שיעור התאונות הצפוי להתרחש באזור הצפון, גדול משיעור התאונות הצפוי באזור הדרום- ממצא המתאים לתוצאות הניתוחים שתוארו בשלבים הקודמים.
- **שיעור התאונות הצפוי להתרחש באזור מרכז, גדול משיעור התאונות הצפוי באזור הדרום –** גם כאן, נמצאה התאמה לניתוחים שתוארו קודם לכן. שוב, אזור הדרום מאופיין בטווח נפחי התנועה הנמוכים יותר (ביחס למרכז), קרי מדובר בכבישים בהם התנועה דלילה מאוד, כאשר נראה כי הלוגיקה ולפיה, היכן שאין רכבים הרי שההסתברות לתאונה יורדת בצורה דרסטית, מתקיימת.

ממצאי ניתוח האינטראקציות, תואמים את תוצאות הניתוח הראשון שהתקיים במסגרת בניית מודל האימון, **וגם את תוצאות הניתוח השני**.

בניתוח האחרון שבוצע על משתנה ה"אזור", נבחן **אזור הצפון בצורה ייעודית, לאחר שבוצעה חלוקה לאזורי משנה** (כאמור בגין מספר התאונות הצפוי הגבוה שבו). תוצאות הניתוח הראו, כי שיעור התאונות הצפוי באזור ה"**גליל העליון המערבי**" גבוה פי 1.36 משיעור התאונות הצפוי באזור "חיפה", גבוה פי 2.2 משיעור התאונות באזור ה"עמקים", וגבוה פי 1.75 משיעור התאונות באזור ה"גליל התחתון המזרחי".

במקביל, שיעור התאונות באזור הגליל התחתון המערבי, עתיד להיות גבוה פי 1.87 משיעור התאונות ב"חיפה", ופי 1.59 משיעור התאונות באזור ה"עמקים".

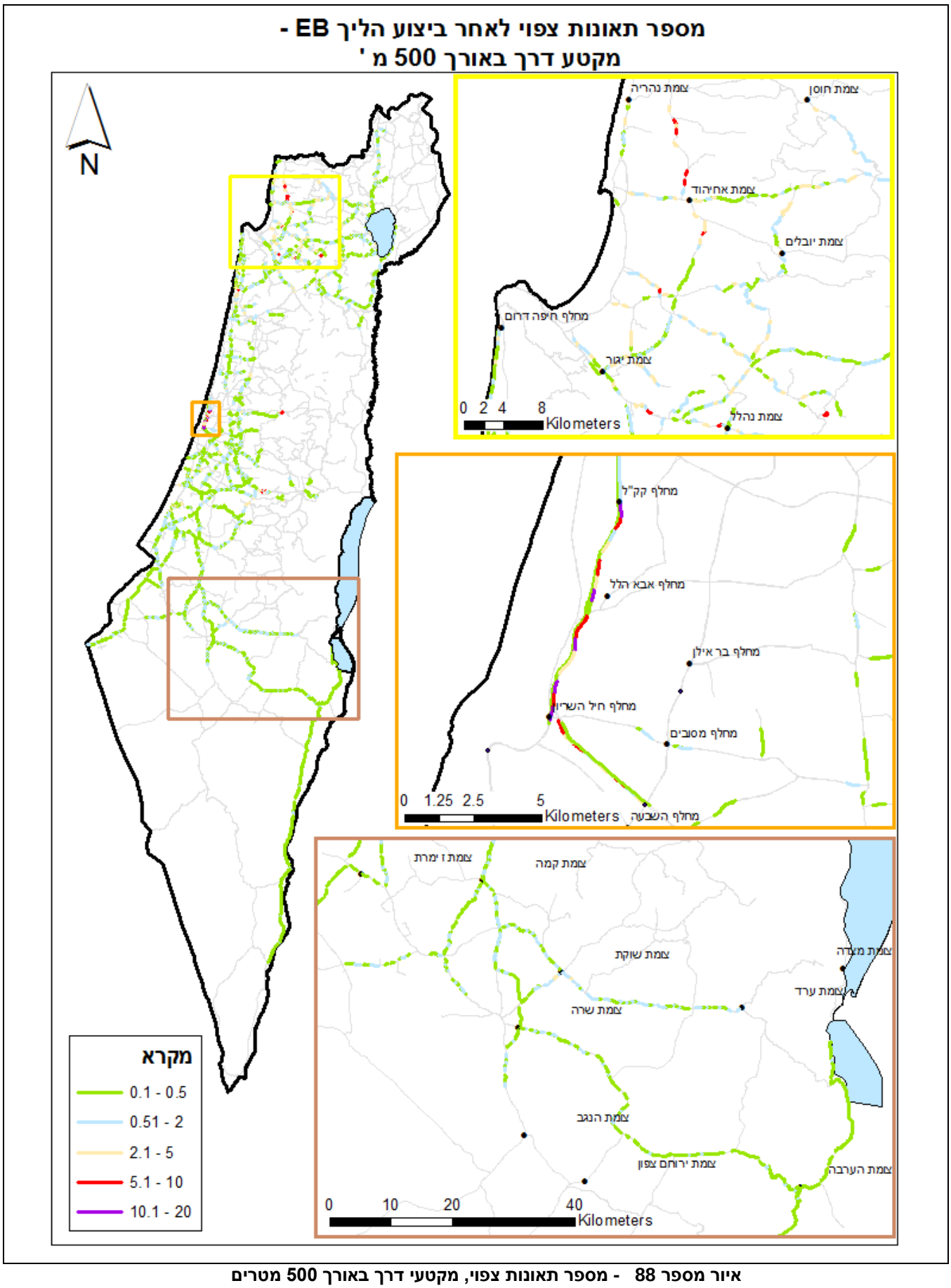
ממצא זה, ניתן לאפיין ולומר כי אזור "הגליל העליון המערבי", הינו האזור שנמצא כבעל הסיכון הרב ביותר לקיומו של אירועי תאונות דרכים, ושוב, ניתן בהמשך לממצא זה, לבצע פעולות מעשיות, החל מהכוונת אכיפה והשקעת כוח אדם, דרך שיפור תשתית, חינוך והסברה.

<u>מפות תוצר החיזוי</u>

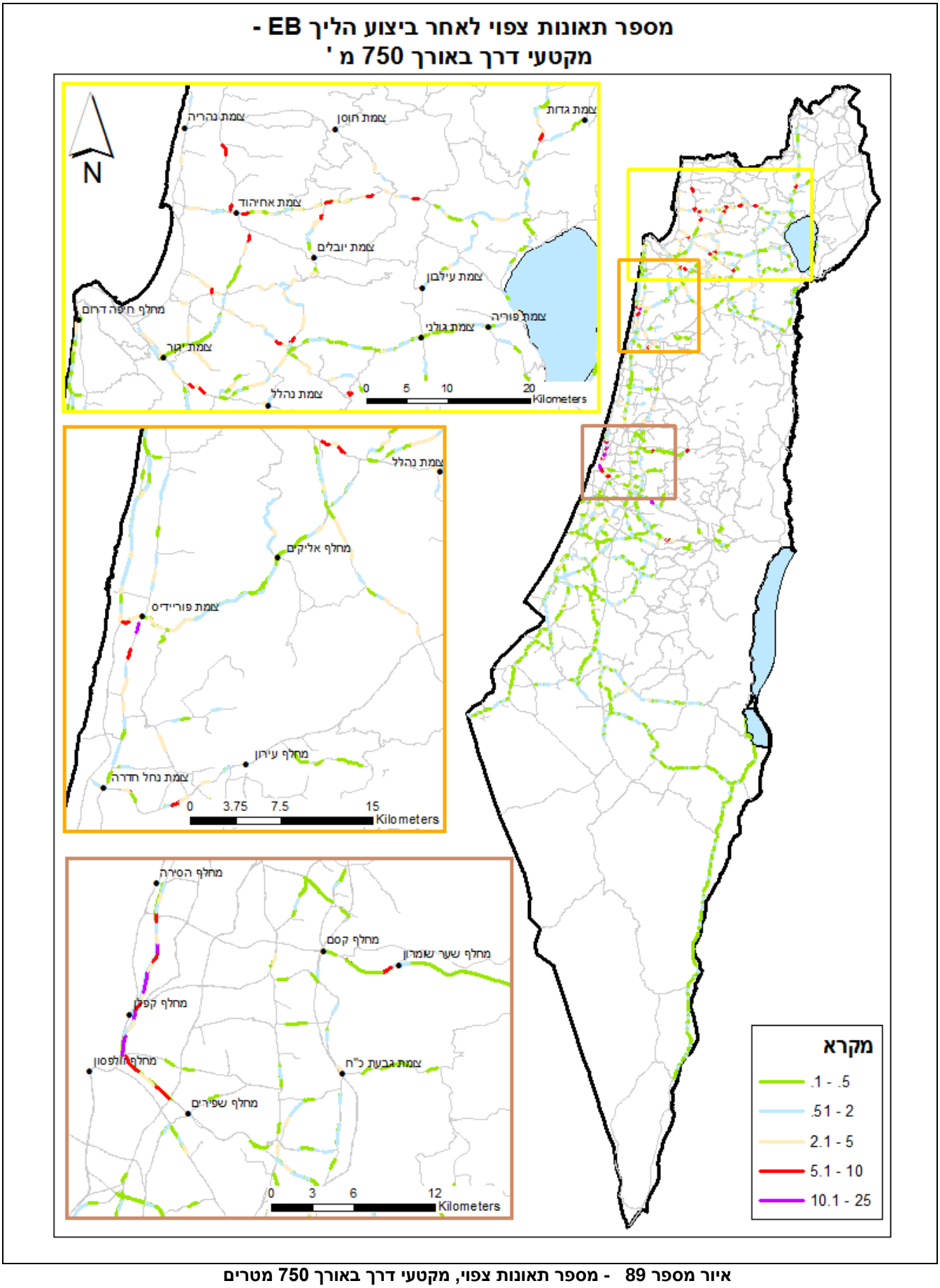
איורים מספר 88-91, מציגים את מפות התוצר של עבודת המחקר. במפות אלה **מוצג מספר התאונות הצפוי** במקטעי הדרך באורכים השונים. משרעת מספר התאונות הצפוי, מוצגת בסולם חזותי צבעוני, החל מהצבע ה"ירוק", המייצג מספר נמוך של תאונות למקטע דרך, וכלה באדום ובסגול המייצגים מקטעי דרך בעלי צפי למספר גבוה של אירועי תאונות דרכים.

יודגש, כי בעת הצבת התאונות במרחב, נבחר להציבן, באופן התואם את ערך הכיווניות שפורט בקובץ הלמ"ס, למרות חוסר הדיוק שבו. הרציונל היה כי לפחות בהיבט הפרמטריים

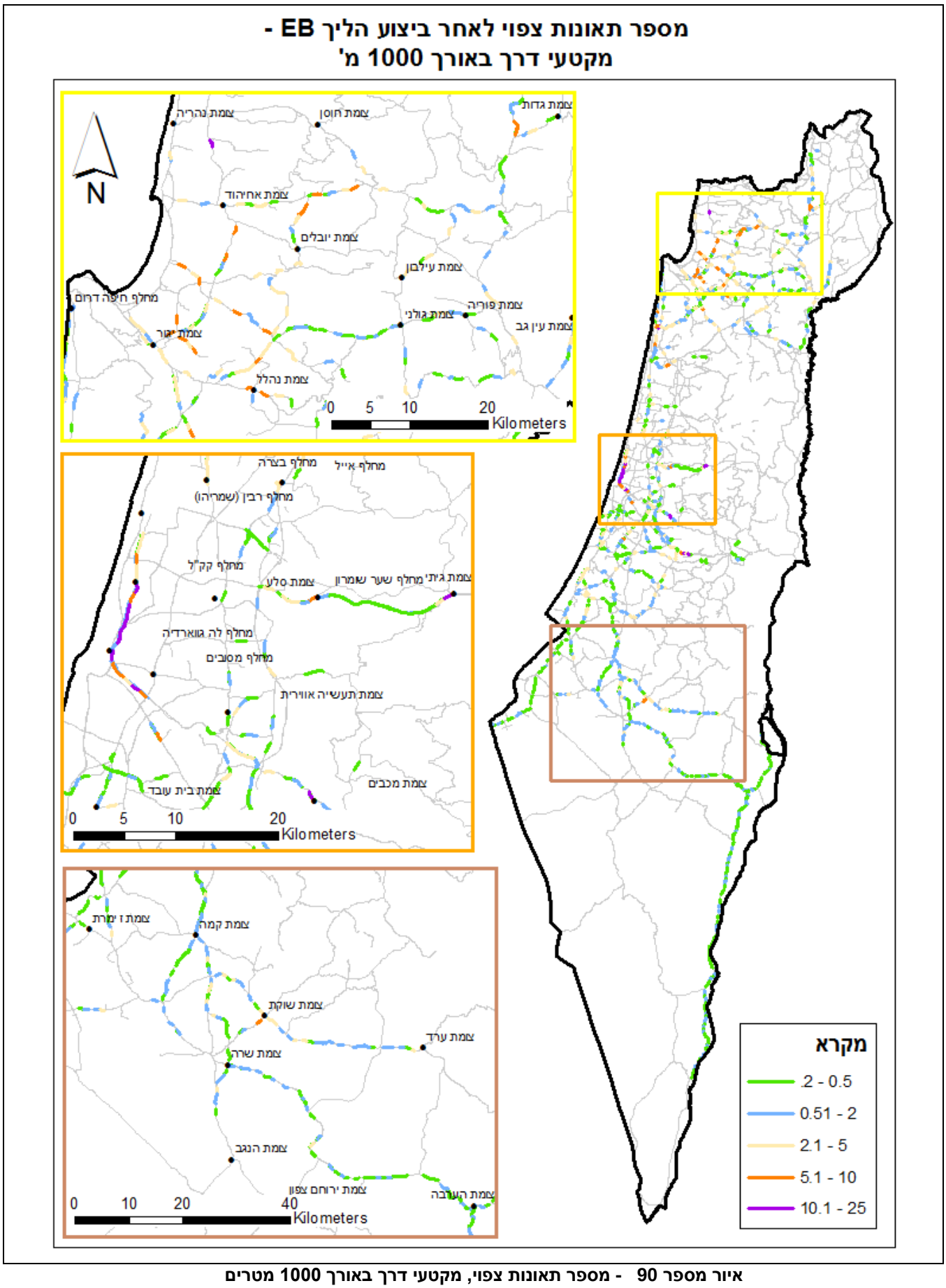
הפיזיים של הכביש, ניתן להניח כי מקטע דרך המסוכן בנתיב לכיוון אחד, יהיה מסוכן גם בנתיב הנגדי.



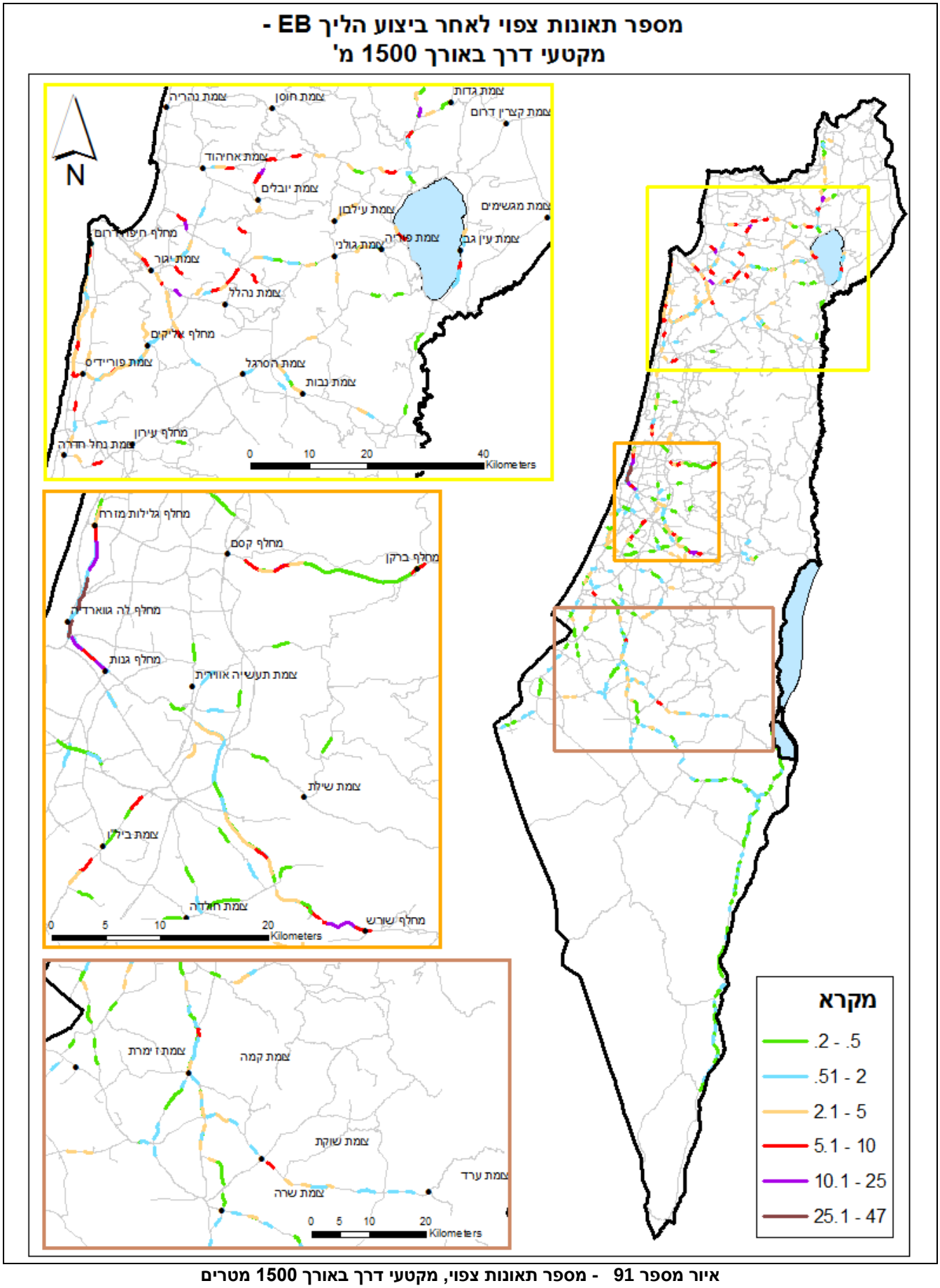
איור מספר 88 - מספר תאונות צפוי, מקטעי דרך באורך 500 מטרים



איור מספר 89 - מספר תאונות צפוי, מקטעי דרך באורך 750 מטרים



איור מספר 90 - מספר תאונות צפוי, מקטעי דרך באורך 1000 מטרים



איור מספר 91 - מספר תאונות צפוי, מקטעי דרך באורך 1500 מטרים

9. סיכום מסקנות והמלצות

בשנת 2006 אימצה ממשלת ישראל את דו"ח וועדת שיינין[30] והכנסת חוקקה את "חוק הרשות הלאומית לבטיחות בדרכים-הוראת שעה, התשס"ו-2006". החל משנת 2007, הרשות הלאומית לבטיחות בדרכים שמה לה כמטרה, להרחיב את בסיס המידע של נתוני תאונות הדרכים. בהתאם לחוזה שנחתם בשנת 2007 עם מכון גרטנר לחקר אפידמיולוגיה ומדיניות בריאות, מועברים ללמ"ס מדי חודש, נתונים נבחרים ממערכת רישום הטראומה (משך אשפוז, חומרת הפגיעה וכיוצא בזאת). לפי פרסומי הרשות, יש כוונה להוסיף לנתוני התאונות גם מידע רלוונטי על הנהגים המעורבים (תאונות ועבירות קודמות, נקודות שהצטברו) ומאפייני תשתית במקום התאונה. בשנת 2010 היתווספו למערכת המידע גם נתוני מד"א.

בפרק הזמן שחלף מראשיתה של עבודת מחקר זו ועד היום, ניתן (וגם אם לא נבחן העניין לעומקו במחקר סדור) להבחין בשינוי. כיום, בעת ביקור באתר הרשות הלאומית לבטיחות בדרכים ניתן למצוא בקלות את מפת תאונות הדרכים. קרי, לכל תאונה משוייך גם מיקום גיאוגרפי. בנוסף, קיימת מערכת המאפשרת לכל משתמש, לבצע פילוחים שונים של תאונות דרכים. בתחילת מחקר זה לא כך נראו הדברים. כזכור, חלק גדול מראשית המחקר, התמקד בהשמת התאונות במרחב. דרך השמת התאונות במרחב והצורך בקבלת ההחלטות, כפי שהוצגו בעבודה זו, מצביעה על מורכבת הנתונים ולצורך במידע משלים בבסיס הנתונים.

<u>המלצות</u>

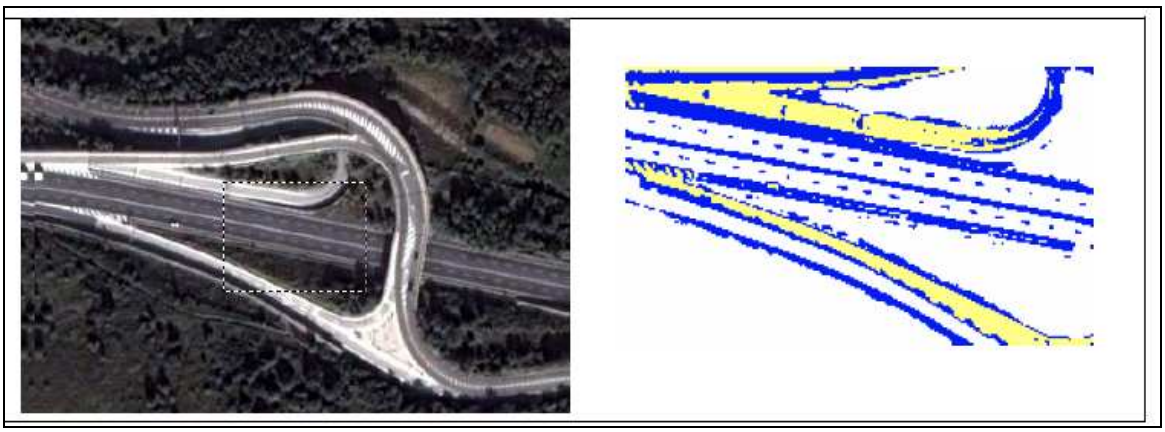
1. ההמלצה האופרטיבית הברורה העולה מדו"ח זה, הינה הצורך **בשיפור התשתיות, ההכוונה והאכיפה באזור הצפון.**
2. **הרחבת תקופת המחקר** – עבודת המחקר הנוכחית התבססה על נתוני שלוש שנים (2005-2007). ברור כי בפרק זמן זה בוצעו שיפורים רבים בהיבט התחבורתי. לפיכך, מומלץ לאמץ את מתודולוגיה המחקר המתוארת, ולבצעו על נתונים עדכניים. כמו כן, על מנת להגיע למודל מדוייק יותר, נהוג להרחיב את תקופת הבחינה ככל שניתן. אם כי, יודגש כי בשל הדינאמיות הרבה המאפיינת את המרחב הנבחן, עם זאת, .
3. **ביצוע מחשב מחודש בנוגע למבנה בסיס הנתונים של תאונות הדרכים**:

(א) אפיון דו"ח תאונה – מומלץ לבחון את הקטגוריות הקיימות, ולהגדירן באופן פשוט וברור, תוך פירוט רב. כך לדוגמא, מצבים דוגמת אלה שתוארו בניתוח נתוני ה"תאורה" , לא ישנו.

(ב) יש לפעול למיפוי מדוייק של רוחב השול. נכון להיום בנתוני מפ"י, נראה כי רוחב השול עודכן באופן אוטומאטי בהתאם לסדר הכביש. בנתוני חברת "מפה" (שבנתוני הכבישים הביעירוניים שלה עשינו שימוש) נתון זה כלל אינו מופיע.

[30] וועדה ציבורית להכנת תכנית לאומית רב-שנתית לבטיחות בדרכים

(ג) נדרש להשלים מידע מפורט בנוגע למפרדות, מספר נתיבים, קיום קו הפרדה לבן או קו הפרדה מפוספס. אלה ניתנים להשלמה גם באמצעים של חישה מרחוק, כמפורט במאמרם של Cefalo & Roberti [2007], וכמוצג דוגמא באיור מספר 92 להלן:



איור מספר 92 - חילוץ מאפייני הכביש באמצעים של חישה מרחוק.

(ד) עדכון "אמיתי" של פרמטר הכיווניות. בראש ובראשונה, בכדי שניתן יהיה להציב את התאונה במקומה האמיתי והנכון, בצד הכביש בו התרחשה התאונה בפועל. שנית, על מנת שבמחקר עתידי, יתאפשר לעשות שימוש בנתון מיקום זווית השמש, לאיתור צבירי תאונות דרכים, אשר יתכן והושפעו מסינוור. במחקר זה הוצג האופן בו ניתן הדבר לביצוע.

(ה) עבור אותם שדות או אותם משתנים מומלץ לוודא כי סוג המשתנה (DATA TYPE), מתואר באחידות בקרב כלל הגורמים העוסקים באיסוף ורישום נתוני תאונות דרכים. דבר שיקל על חיבור כלל הנתונים ויאפשר טיובם ובקרת איכות עליהם.

4. **איכון אוטמאטי של אירועי תאונות דרכים** – במשטרת ישראל ובמד"א, על אף קיומה של מערכות מידע גיאוגרפי ומכשירי GPS בכלי הרכב, עדין לא קיים איכון אוטומאטי של מיקום התאונה. מיקום מדוייק, משמעו שיפור משמעותי באיתור מקטע הדרך בו אירעה התאונה. פועל יוצא מכך, הינו העלאת דיוק הנתונים, והעלאת דיוקי המודלים הסטטיסטיים, בדומה למודלים שהוצגו במחקר זה. כפועל יוצא, איתור צבירי אירועי תאונות דרכים יהפוך מדוייק יותר, אף הוא. כמו כן, בהיבט הנתונים ה"גולמיים", ממצאי המחקר מצביעים על יישימות עקרונית לשימוש בנתוני מד"א למטרת טיוב נתוני המשטרה והלמ"ס באזורים עירוניים.

<u>השגי המחקר:</u>

ראשית, בהיבט הנתונים ה"גולמיים" – ממצאי המחקר מצביעים על יישימות עקרונית לשימוש בנתוני מד"א למטרת טיוב נתוני המשטרה והלמ"ס באזורים עירוניים.

שנית, במחקר שולבו פרמטרים מרחביים, בחלקם חדשניים בשימוש בחקר הבטיחות בדרכים. כך לדוגמא, הוצגה היכולת לבצע חישוב פרמטר זווית שמש למטרת בחינת תאונות שהתרחשו כתוצאה מסינוור. בנוסף, חושבו פרמטרים של שיפוע ועקמומיות מקטע דרך. כאשר במחקר זה פרמטרים אלה נדונו לעומקם, תוך בחינת מספר נוסחאות על תוצאותיהן, ובחירת הנוסחה שסיפקה את המענה המרחבי הטוב ביותר. כמו כן, בוצע שיוך של כל מקטע דרך לאזור גיאוגרפי: צפון, מרכז ודרום, באופן התואם את אזורי ה"פריפריה" (צפון ודרום) ואת אזור הליבה ("מטרופולין ת"א"). ניתוח משתנה זה (ה"אזור"), כפי שנבחר באופן החלוקה והשילוב במחקר, העלה ממצאים מעניינים לגבי שיעור התאונת הגבוה הצפוי באזור הצפון ובמיוחד באזור "הגליל העליון המערבי".
משתנים נוספים וייחודיים ששולבו במחקר זה הינם משתנים המייצגים אינטראקציות בין אזור לעקמומיות, בין אזור ושיפוע ובין נפח תנועה ואזור. אלה, איפשרו לבחון האם העקמומיות הינה בעלת השפעה שונה על מספר התאונות באזורים שונים, האם ניתן לאפיין מספר תאונות שונה עבור שיפועים באזורים שונים והאם השפעת נפח התנועה על שיעור תאונות הצפוי תלויה באזור בו נמצאים, בהתאמה.

ההיבט השלישי הינו ההיבט המתודולוגי. במהלך המחקר פותחה מתודולוגיה המאפשרת להצביע על מקטעי דרך קצרים, שהינם בעלי סיכון להתרחשות תאונת דרכים. להפנייה מדוייקת משמעות אופרטיבית של חיסכון במשאבים. יתר על כן, חלוקת בסיס הנתונים למקטעי דרך באורכים שונים, איפשרה לנטרל את פרמטר אורך הסגמנט, שבמחקרים רבים עלה כפרמטר משמעותי ברמת המודל. כמו כן, טוייב המודל על ידי שימוש בשיטת Empirical Bayes המעלה את דיוק ההערכה עקב שימוש במידע היסטורי ומתקנת את ההטייה עקב תופעת ה"רגרסייה אל הממוצע".
היבט מתודולוגי נוסף, הינו ההתייחסות לכיוון התאונה בעת הצבת התאונה במרחב. על אף שברור כי נתון הכיווניות הינו בעייתי (הרחבה בסעיף מספר 4.2.5), נעשה בו שימוש, הן בפיתוח שאילתות SQL המאפשרות לפתח מתודולוגיה של חישוב זווית שמש למטרת איתור תאונות המרחשות כתוצאה מסינוור, והן באיכון התאונות במרחב.
עוד, נכתב קוד SQL המאפשר לאתר צבירים במרחב הליניאמנטי המרחבי ולא במרחב הפוליגונאלי.

<u>מחקרי המשך</u>

כאמור, בחקר תאונות הדרכים קיימים עוד אפיקים רבים לשיפור, התקדמות והמשך מחקר. בהמשך למחקר זה, שבסיסו ניתוחים סטטיסטיים-מרחביים מתקדמים של כמות אירועי תאונות הדרכים במקטע דרך, קיימות עוד מספר אפשרויות לניתוח, לדוגמא:

(א) ניתוח אזורים - מוצע להמשיך ולבצע בחינה עבור תתי אזורים במרחב הדרום, ומרכז הארץ (בדומה, לניתוח שבוצע באזור הצפון).

(ב) באזורים בהם נמצא שיעור תאונות צפוי גבוה, מוצע לבחון משתנים מסבירים נוספים במגוון פילוחים : אנושיים-התנהגותים (דוגמת סוג אוכלוסיה, גיל הנהג, ותק הנהיגה), וגיאוגרפיים-מרחביים (דוגמת: שימושי קרקע, מקומות בילוי ועוד). עוד מומלץ לבחון שילוב פרמטרים הניתנים לחילוץ באמצעי חישה מרחוק למטרת חקר תאונות הדרכים, דוגמת מידע בנוגע לפסי הפרדה או מידע סביבתי נוסף (הצפות, אבק וכו'), בדומה למצויין במאמרם של **Fuller** et-al [2002].

(ג) שילוב מידע וניתוח אוטוקורלציה.

(ד) מחקרים בחיתוכים שונים של שעות היממה וימות השבוע.

10. רשימת מקורות

אנגלית

1. Aarts, L. & Van Schagen, I. [2006] "Driving speed and the risk of road crashes: a review", Accident Analysis and Prevention, Volume 38, Issue 2, March 2006, Pp 215-224.
2. Aguero V. J. & Jovanis P. P. [2008] "Analysis of Road Crash Frequency with Spatial Models", Transportation Research Record: Journal of the Transportation Research Board (TRB), Volume 2061, Pp 55-63.
3. Baruya, A. [1997], "Working Paper: A Review of Speed-Accident Relationship on European Roads", MASTER Working Paper R.1.1.1. (http://virtual.vtt.fi/master/deliver.htm)
4. Cadkin J. [2002], "Understanding Dynamic Segmentation", ArcUser, October-November, Pp: 40-43 (http://www.esri.com/news/arcuser/0702/files/dynseg.pdf)
5. Cadkin J. & Brennan P. [2002], "Dynamic Segmentation in ArcGIS", ArcUser, July-September, Pp: 52-54. (http://www.esri.com/news/arcuser/1002/files/dynseg_2.pdf).
6. Cefalo R. & Roberti R.[2007], "Extraction of Road Geometric Parameters from High Resolution Images Validated By Geodetic Techniques", 4th International SIIV Congress– PALERMO (ITALY), 12-14 September 2007
7. Chatterjee S., Hadi A. and Price B. [2000], Regression Analysis by Example, 3rd Edition John Wiley & Sons, New York.
8. Chiarella C. & Shenhuai G.[2002] "Type I *Spurious Regression* in Econometrics", School of Finance and Economics , UTS: Business, Working Paper No. 114. http://www.finance.uts.edu.au/research/wpapers/wp114.pdf
9. Cressie N.A.C [1993], Statistics for spatial data – revised edition, New York: John Wiley & Sons.
10. Cressie N. [1990] "The origins of kriging", Mathematical Geology, Volume 22, Number 3, Pp: 239-252.
11. Diggle P.J. [1983], Statistical Analysis of Spatial point Pattern, Academic Press, London, UK.
12. Dixon P.M [2002], Encyclopedia of Environmetrics, editors-in-chief: Abdel H. El-Shaarawi, Walter W. Piegorsch. Published John Wiley & Sons, Chichester c2002.

13. El-Basyouny K. & Sayed T. (2010) "Safety performance functions with measurement errors in traffic volume",Article in press, Safety Science, Elsevier Ltd.
14. Fuller D., Jeffe M.,. Williamson R. A. and James D. [2002] "Satellite Remote Sensing and Transportation Lifelines: Safety and Risk Analysis Along Rural Southwest Roads" , Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS 2002 Conference Proceedings.
15. Geurts K., Thomas I. and Geert Wets [2005] "Understanding spatial concentrations of road accidents using frequent item sets", Accident Analysis & Prevention , Volume 37, Issue 4, July 2005, Pages 787-799.
16. Guo B. & Kurt C.E. [2004], "Towards temporal dynamic segmentation", GeoInformatica.;8(3), Pp 265-283.
17. Hakim S., Shefer D., Hakkert S, and Hocherman I. [1991] "A critical review of macro models for road accidents",Accident Analysis & Prevention, Oct ;23(5):379-400.
18. Hauer E. [1992], "Empirical bayes approach to the estimation of “unsafety”: The multivariate regression method", Accident Analysis & Prevention, Volume 24(Issue 5), Pages 457-477.
19. Hauer E. [2001] "Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation", Accident Analysis & Prevention, Volume 33, Issue 6, November 2001, Pp 799-808.
20. Hauer E., Harwood D. W., Forrest M. and Michael S. G [2002] “Estimating Safety by the Empirical Bayes Method: A Tutorial”, TRB, 2002.
21. Hosmer W. D.& Lemeshow S.,[2000], Applied Logistic Regression , Second Edition, John Wiley & Sons, New York.
22. Jones A.P., Langford I.H. and Bentham G. (1996) "The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England", Social Science & Medicine, Vol. 42, No. 6, Pp 879-885.
23. Karlaftis M. G. & Tarko A. P. [1998], “Heterogeneity considerations in accident modeling“, Accident Analysis & Prevention, Vol. 30, Issue 4, July 1998, Pp 425-433.
24. Kelsall J. E & Diggle P.J [1995] Kernel estimation of relative risk, Bernoulli. Volume 1, Number 1-2 , Pp 3-16.
25. Kim K.; Takeyama D.; Nitz L. [1995] "Moped Safety in Honolulu, Hawaii", Journal of Safety Research, 1995, vol. 26, no. 3, pp. 177-185.

26. Kim K. & Levine N. [1996], "Using GIS to improve highway safety", Computers, Environment and Urban Systems, Volume 20, Number 4, July, Pp. 289-302 (14).
27. Levine N., Kim K.E & Nitz H. L. [1995] , "Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns", Accident Analysis & Prevention, Volume 27, Issue 5, October 1995, Pages 663-674.
28. Levine N., Kim K.E & Nitz H. L.[1995] , "Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators" ,Accident Analysis & Prevention, Volume 27, Issue 5, October 1995, Pages 675-685
29. Miaou S.P. [1994] "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions", Accident Analysis & Prevention, Volume 26, No. 4, Pp 471-482.
30. Miaou S.P. , Lum H. [1993] "Modeling vehicle accidents and highway geometric design relationships", Accident Analysis & Prevention, Volume 25, Issue 6, December, Pp: 689-709.
31. Miaou S.P., Song J.J., Mallick B. [2003] "Roadway traffic-crash mapping: a space-time modeling approach", Journal of Transportation and Statistics, No. 6, Pp: 33-57
32. Ng K.S, Hung WT and Wong WG (2002) "An algorithm for assessing the risk of traffic accident", Journal of Safety Research, Fall;33(3):387-410.
33. Nicholson A. [1999], "Analysis of spatial distributions of accidents", Safety Science, Volume 31, Number 1, March, Pp. 71-91(21)
34. Noronha V. & Church R.L [2002], Linear referencing and Alternative Expressions of location for Transportation, California Department of Transportation, Task Order 3021, Final Report.
http://www.ncgia.ucsb.edu/vital/research/pubs/200204-LRsynthesis.pdf
35. Okabe A. and Yamada I [2001], "The K Function method on a Network and its computational Implementation", Geographical Analysis 33 Vol.3, Pp 271-290.
36. Oppe [1991], “Development of Traffic and Traffic Safety: Global Trends and Incidental Fluctuations”, Accident Analysis and Prevention, 23 (5), Pp.413-422.
37. O'Sullivan D. & D. Unwin [2003] Geographic Information Analysis , published by Jhon Wiley & Sons Inc, US .
38. Page Y. [2001], "A statistical model to compare road mortality in OECD countries, Accident Analysis and Prevention, 33 (3), Pp.371-385, May.
39. Peled, A., Haj-Yehia, B., Hakkert, A.S.[1996] "ArcInfo-Based Geographical Information System for Road Safety Analysis and Improvement"

(http://www.esri.com/library/userconf/proc96/TO50/PAP005/P5.HTM)

40. Peled, A. & Hakkert, A.S., [1993], "A PC-Oriented GIS Application for Road Safety analysis and management", Traffic Engineering and Control, Vol. 34, July\August, Pp. 355-361.
41. Peltzman S. [1975] "The Effects of Automobile Safety Regulation", Journal of Political Economy, Vol. 83, No. 4, August, Pp: 677-726.
42. Powers M. & Carson J.[2004], Before-After Crash Analysis: A Primer for Using the Empirical Bayes Method – Tutorial, Reaserch Center – Montana Department of Transportation, May.
43. Qin X., Ivan J. N. , Ravishanker N. and Liu J. [2005] "Hierarchical Bayesian Estimation of Safety Performance Functions for Two-Lane Highways Using Markov Chain Monte Carlo Modeling", Journal of Transportation Engineering, May, Pp 345- 351.
44. Ripley, B. D. [1977] "Modelling spatial patterns (with discussion)", Journal of the Royal Statistical Society ,39, Pp172-212.
45. Road Research Laboratory [1965] Research on road traffic, Department of Scientific and Industrial Research, London.
46. Sabel C.E., Kingham S. Nicholson A. and Bartie P. [2005], "Road Traffic Accident Simulation Modelling - A Kernel Estimation Approach", SIRC (Spatial Information Research Centre) 24 November 2005, Otago University, New Zealand.
47. Sayed, T., & Rodriguez, F. [1999] "Accident prediction models for urban unsignalized intersections in British Columbia", Transportation Research Record, 1665, Pp 93–99.
48. Schneider R. J., Ryznar R. M. and Khattak J. A. [2004], 'An accident waiting to happen: a spatial approach to proactive pedestrian planning", Accident Analysis & Prevention, Volume 36, Issue 2, March 2004, Pp 193-211.
49. Smeed R.J. [1949] "Some Statistical Aspects of Road Safety Research", Journal of the Royal Statistical Society, Series A (General), Vol. 112, No. 1 , Pp. 1-34.
50. Solomon, D. [1964], Accidents on Main Rural Highways Related to Speed, Driver and Vehicle, Washington, DC: Bureau of Public Roads.
51. Song J.J., Ghosh M., Miaou S. and Mallick B. [2006], "Bayesian multivariate spatial models for roadway traffic crash mapping" , Journal of Multivariate Analysis, Volume 97, Issue 1, January 2006, Pages 246-273

52. Summala H. [1996], "Accident risk and driver behaviour", Safety Science , Vol. 22, No. 1-3, pp. 103-117.
53. Sutton J.C & Wyman M.M [2000], "dynamic location: am iconic model to synchronize temporal and spatial transportation data", Transportation Research, Part C 8, Pp: 37-52.
54. Tessa K. A. [2009] ," Kernel density estimation and K-means clustering to profile road accident hotspots", Accident Analysis & Prevention, Volume 41, Issue 3, May, Pp: 359-364
55. Thomas I. [1996], "Spatial Data Aggregation: Exploratory Analysis of Road Accidents", Accident Analysis and Prevention, Volume 28 (2),1996, Pp 251-264.
56. Ward J.H. [1963], "Hierarchical Grouping to Optimize an Objective Function", Journal of the American Statistical Association, Vol. 58, No. 301. (March), Pp. 236-244.
57. Wilde G. J. S. [1982], "The Theory of Risk Homeostasis: Implications for Safety and Health", Risk Analysis, Vol. 2, Issue 4, December, Pp 209-225
58. Winslott H. L. [2004] "Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows", Accident Analysis & Prevention, Volume 36, Issue 6, November, Pp: 985-992.
59. Yamada, I. & Thill J.C. [2004] "Comparison of Planar and Network K-functions in Traffic Accident Analysis." Journal of Transport Geography, 12, 2004, 149-158.

עברית

1. בינסטוק, מ., גולדין א., גפני-סרי ד. [1998], הקשר בין אכיפת תנועה ותאונות דרכים: ניתוח סטטיסטי, לשכת המדען הראשי, המשרד לבטחון הפנים, ירושלים.
2. בר- גרא הילל [2007], מודל סטטיסטי של תאונות דרכים בצמתים בין עירוניים – דו"ח מסכם לקרן רן נאור, אוניברסיטת בן גוריון.
3. גיטלמן ו., הקרט ש., דובא א. וכהן א. [2004], משתנים מסבירים למגמות בתאונות הדרכים (סקר ספרות ובחינת הנתונים הקיימים, פרוייקט מס' 5115-44, פרסום מספר 2004/0125.
4. לויתן ת. ורביב א. [1997], הסתברות, הוצאת עמיחי.
5. משרד התחבורה, הרשות הלאומית לבטיחות בדרכים [2005], תאונות דרכים בישראל 1995-2004, בעריכת: קרן טולדנו-גיא, דפוס ירושלים.
6. פלד אמציה [1996] "איתור בעיות בטיחות עירוניות באמצעות מערכת מידע גיאוגרפי (ממ"ג)", מתוך: חוברת מאמרים במסגרת יום עיון בטיחות בדרכים 16.5.1996, הטכניון, המכון לחקר התחבורה.

7. פרוטוקול מס' 51, מישיבת ועדת המדע והטכנולוגיה, מיום שלישי ביולי 2007, שעה 11:00, סדר היום: השימוש בטכנולוגיה לשיפור הבטיחות וצמצום תאונות הדרכים.
8. צדר א. וליבנה מ. [1977], הקשר בין צפיפות וקצב תאונות לבין עוצמת התנועה היומית (מחקר 681-015), המרכז לבטיחות בדרכים, משרד התחבורה והטכניון , פרסום מספר 77/7.
9. קלרמן אהרון [1986], פרקים בסטטיסטיקה של נתונים מרחביים: סטטיסטיקה תאורית, הסתברות וניתוח שונות, החוג לגיאוגרפיה, אוניברסיטת חיפה.
10. פרסום הלמ"ס : ספירות תנועה בדרכים לא-עירוניות 2001-2007, ירושלים, אדר ב' תשס"ח, מרס 2008.
11. פרסום הלמ"ס : נפח תנועה לפי דרך, קטע ושעת היממה (ממוצע לשעה לכל ימי השבוע), 2004-2009.

<u>אתרים מקוונים</u>

1. האתר הרשות הלאומית לבטיחות בדרכים: http://pasimlev.mot.gov.il/RoadSafety
2. עזרת תוכנת SAS : http://www2.sas.com/proceedings/forum2008/322-2008.pdf
3. משתנה ליטנטי http://www.bicohen.tau.ac.il/uploads/files/.pdf

נספח 1 – השמת תאונות הדרכים במרחב העירוני

הקושי העיקרי בהשמת תאונות במרחב הגיאוגרפי, הינה בסיס נתונים המגיע ללא שיוך קואורדינאטות למקום התרחשות התאונה, אם כי בנתוני הלמ"ס עבור כל תאונה מעודכן היישוב, מספר בית ורחוב.

על מנת להשים את התאונות במרחב, בוצעו שני שלבים. בשלב ראשון - בשכבת ה"מבנים", חושבו עבור כל מבנה, קואורדינאטות מרכז הפוליגון. נתונים אלה, חוברו בשלב מאוחר יותר, עם המידע הטבלאי של תאונות הדרכים. חלק זה בוצע תוך שימוש ב SQL. בשלב שני – בוצע "תיקון" לקואורדינאטות שחושבו בשלב הקודם, והנקודות הוסטו, כך שימוקמו על הכביש הקרוב ביותר. חלק זה בוצע ע"י הרצת קוד ב VBA בתוכנת ARCMAP. בסעיפים הבאים נציג את מהלך השמת התאונות העירוניות לא בצומת.

הנתונים בהם נעשה שימוש עבור ההשמה במרחב

(א) תאונות דרכים טבלאיים (כפי שהתקבלו מהלמ"ס)

נתוני תאונות הדרכים המקוריים, כפי שהתקבלו מהלמ"ס, תוך "חיתוך" התאונות שהתרחשו ביישוב מסויים.. עמודת "סמל יישוב" היא זו המבצעת את השיוך למתחם עירוני הרצוי.

(ב) נתונים גיאוגרפיים מרחביים

במחקר נעשה שימוש בשתי שכבות מידע גיאוגרפיות (קיבצי SHP) ובשכבת שמות רחובות טבלאית (address):

- שכבת מבנים של הממ"ג הלאומי (פוליגונאלית Layer 7) – בשכבה זו, עבור כל מבנה, חושבו קואורדינאטות X,Y של מרכז הפוליגון. נתונים אלו "יחוברו" בשלב מאוחר יותר עם המידע הטבלאי של תאונות הדרכים.
- שכבת כבישים של הממ"ג הלאומי (קווית Layer 1) – לשכבה זו יבוצע בשלב מאוחר יותר תיקון קואורדינטות, על מנת שהתאונות "תוצבנה" בהתאמה על הכבישים).
- שכבת שמות רחובות טבלאית (annotation.address), ושכבת streets שהיא שכבה טבלאית.

מהלך העבודה

(א) חילוץ נקודות מרכז הפוליגון משכבת המבנים ויצירת שכבה נקודתית

שלב זה בוצע בתוכנת ARCMAP. טרם תחילת המהלך, יש לוודא כי שכבת המבנים נמצאת בהיטל מטרי (Israel_TM_Grid).

- הוספת שתי עמודות Xcentral, Ycentral (מסוג: Long)
- חישוב עמודות (מרכז הפוליגון)

- הפיכת השכבה הפוליגונאלית לשכבה נקודתית, בהתאם לנקודות מרכז הפוליגון, תוך שימוש בכלי מדף Feature To Point. נשמור את השכבה הנקודתית בתוך ה mdb (שכבת Haifa_Buildings_ITM_Points)
- בסוף פרוצדורה זו, התקבלה טבלה בה עבור כל תאונה בבסיסי הנתונים מוצגות קואורדינאטות מרכז הפוליגון (Xcentral, Ycentral).

(ב) עריכת המידע הטבלאי של תאונות הדרכים (קובץ הלמ"ס) והרצת שאילתות SQL המחברות בין נתוני תאונות הדרכים לנתונים הגיאוגרפיים.

- כזכור, המטרה הינה שייוך X,Y לכל תאונה (כל שורה) בטבלת התאונות של הלמ"ס (Accidents2005).
- **הוספת עמודת Calc_Stree_ID** בטבלת התאונות. עמודה זו מייצגת סמל יישוב וסמל רחוב המחושבים על פי נוסחה מספר 1:

[1] Calc_Stree_ID = SEMEL_YISHUV * 10000 + REHOV1

עמודה זו, המייצגת כאמור, את סמל יישוב וסמל רחוב, בצירוף עמודה המייצגת את מספר הבית, ישמשו בשאילתה הבאה כשדות מקשרים בין נתוני התאונות הטבלאיים של הלמ"ס לנתונים הגיאוגרפיים (טבלת התוצר של Q1 הינה *Accidents2005_WithCalc_Street_Id*).

```
SELECT Accidents2005.*, ([SEMEL_YISHUV]*10000)+[REHOV1] AS
Calc_Stree_ID INTO Accidents2005_WithCalc_Street_Id
FROM Accidents2005;
```

- **עריכת שכבת האנוטציה (annotation) של המבנים.** כאמור, שכבת אנוטציה מכילה כתובות עבור כל פוליגון בבסיס המידע. במקרה בו בנוסף למספר הבית בשכבת האנוטציה (*bldg_num_annotation_address*) נכתבה אות (לדוגמא מספר בית 37 ב), נוצר מצב בו ה"רווח" מתפרש על ידי התוכנה כסימן □. כך37 ב מופיע כ: □7ב. על מנת שנוכל לקשר בין השכבות גיאוגרפיות לבין טבלאות הנתונים הוספנו בשכבת האנוטציה עמודה נוספת (Build_Num_Calc) מסוג Double. בשלב לאחר מכן, הורץ קוד PEARL העובר על עמודת ה TEXT ומחלץ את מספר הבית בלבד (ללא אות). לאחר מכן הורצה שאילתה (Q2) המעדכנת בטבלת (*bldg_num_annotation_address*) מספרים "נקיים" ללא אותיות.
- היות ובשכבת האנוטציה של המבנים "חתכנו" את ה"אותיות" והשארנו את הספרות, נוצר מצב בו עבור בית בו מספר כניסות (37א', 37ב' ו 37ג'), מיוצג בבסיס הנתונים על ידי מספר אחד (37) ושלושה קודים שונים ייחודיים **(UNIQUE7_ID).** דבר זה יצר כפילויות

בשלבים הבאים, לפיכך הורץ Q3 הבוחר את הערך אחד (המקסימאלי) עבור כל מבנה (יכולנו לבחור גם את המינימאלי או המרכזי– הבחירה היתה שרירותית).

- **יצירת טבלה מחיבור התאונות כפי שהתקבלו מהלמ"ס לשכבת המבנים הנקודתית (של מרכז הפוליגון). בכדי לבצע זאת הורץ Q4.**

(ג) ArcMap - יצירת שכבה גיאוגרפית נקודתית המייצגת את התאונות (SHP)

- ייבוא טבלת התוצר של הסעיף הקודם לתוך תוכנת ArcMap. ("import x,y")
- ייצוא השכבה לפורמט SHP בהיטל Israel TM grid.

(ד) תיקון הקואורדינאטות, כך שיוסטו למיקום על הכביש הקרוב (הרצת קוד VBA ב ArcMap) .

- נכתב קוד VBA המסיט את ה"נקודות", המייצגות תאונה, לכביש הקרוב ,והמוצג להלן:

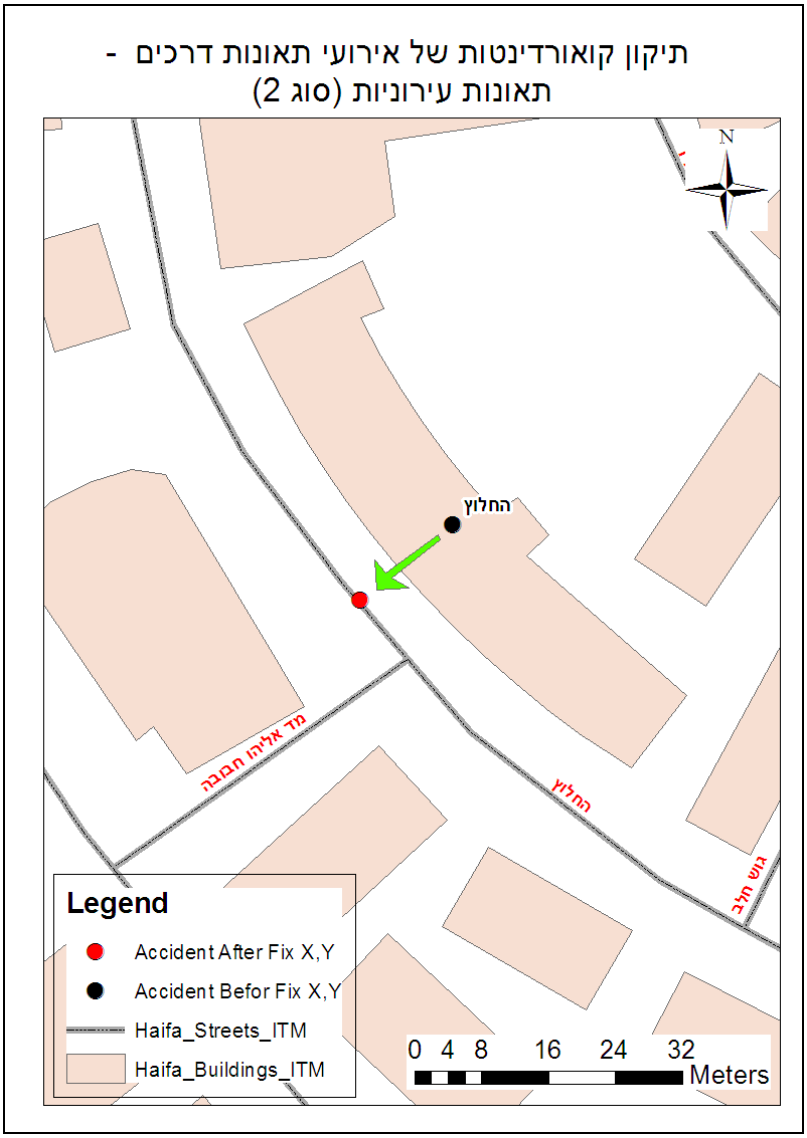
```
Option Explicit

Public Sub Iproximity_test()


Dim pMxDoc As IMxDocument
Set pMxDoc = ThisDocument

'set active data frame
Dim pDataFrame As IMap
Set pDataFrame = pMxDoc.FocusMap

'set the layer we want to work with on the map. Feature layer=map display
'Declare and set IFeatureLayer variable

Dim pFPointsLayer As IFeatureLayer
Set pFPointsLayer = pDataFrame.Layer(0(

Dim pFLineLayer As IFeatureLayer
Set pFLineLayer = pDataFrame.Layer(1(


'Declare and set IFeatureClass variable
'FeaturClass = the phisical location of shp

Dim pFClassPoint As IFeatureClass
Set pFClassPoint = pFPointsLayer.FeatureClass

Dim pFClassLine As IFeatureClass
Set pFClassLine = pFLineLayer.FeatureClass


'Create search cursor

Dim pFCursorPoint As IFeatureCursor
Set pFCursorPoint = pFClassPoint.Search(Nothing, False) 'now we have cursor that points to the top of the
colleciton features
'moving trought the features/rows
Dim pFeaturePoint As IFeature

Dim pFCursorLine As IFeatureCursor
Set pFCursorLine = pFClassLine.Search(Nothing, False) 'now we have cursor that points to the top of the colleciton
features
'moving trought the features/rows
Dim pFeatureLine As IFeature

----------------------------------------'

   Dim pPoint As IPoint
   Dim pPolyLine As IPolyline
   Dim pProxOp As IProximityOperator

  Dim min_dist As Double
```

```
  Dim near_pnt As IPoint

  Dim pFeatEdit As IFeatureEdit
  Dim str_id As Long
  Dim qf As IQueryFilter
  Set qf = New QueryFilter

  Set pFeaturePoint = pFCursorPoint.NextFeature

   Do While Not pFeaturePoint Is Nothing
   min_dist = 10000000
   Set pPoint = pFeaturePoint.ShapeCopy
   str_id = pFeaturePoint.Value(pFeaturePoint.Fields.FindField("Calc_Stree(("
   qf.WhereClause = "STREET_ID = " & str_id

   pPoint.Project pMxDoc.FocusMap.SpatialReference

   Set pFCursorLine = pFClassLine.Search(qf, False(

   Set pFCursorLine = pFClassLine.Search(Nothing, False) ' the 1st line
   Set pFeatureLine = pFCursorLine.NextFeature
   Do While Not pFeatureLine Is Nothing
      Set pPolyLine = pFeatureLine.ShapeCopy
      pPolyLine.Project pMxDoc.FocusMap.SpatialReference
      Set pProxOp = pPolyLine
      If (pProxOp.ReturnDistance(pPoint) < min_dist) Then
         Set near_pnt = pProxOp.ReturnNearestPoint(pPoint, esriNoExtension(
         min_dist = pProxOp.ReturnDistance(pPoint(
      End If

      pFeaturePoint.Value(pFeaturePoint.Fields.FindField("Xprox")) = near_pnt.X
      pFeaturePoint.Value(pFeaturePoint.Fields.FindField("Yprox")) = near_pnt.Y
      pFeaturePoint.Store
      Set pFeatureLine = pFCursorLine.NextFeature
   Loop
   Set pFeaturePoint = pFCursorPoint.NextFeature
Loop


End Sub
```

- נדגיש כי הקוד כתוב כך, שקיימת חשיבות לסדר פתיחת השכבות בפרוייקט. יש לפתוח את השכבה הנקודתית כך שהיא תהיה הראשונה (עליונה) ב TOC, ולאחר מכן את השכבה הקווית (שתהיה במקום השני).
- לאחר מכן נריץ: Tools→ macros→ macros

<u>התוצר</u>

איור מספר 93 מדגים את תוצר מיקום התאונה לאחר התיקון וההסטה לכביש הקרוב.



איור מספר 93 - דוגמא – השמת תאונות עירוניות. הסטת נקודת מרכז פוליגון לכביש הקרוב

<u>שימוש בנתוני מד"א למטרת טיוב נתוני תאונות הדרכים</u>

נכתבה סדרת שאילתות SQL המבצעות חיבור בין נתוני מד"א לבין נתוני תאונות דרכים של הלמ"ס. במהלך החיבור נדרשנו ל casting מרובה, עקב סוג הנתונים השונה מבסיס נתונים אחד למשנהו.

טבלה מספר J, מציגה את תוצרי החיבור הסופי בין נתוני הלמ"ס לבין נתוני מד"א. השדות המקשרים הינם: תאריך, סמל ישוב, שם רחוב ושעה. כאשר על מנת לחבר בצורה טובה יותר בין שני סטי הנתונים, הורחב טווח השעות בקובץ הלמ"ס לחצי שעה לפני השעה המקורית וחצי שעה אחרי השעה המקורית המדווחת בקובץ). זאת, כאמור מתוך הנחה כי פער שעות בין שני הקבצים הוא אפשרי ויתקיים.

טבלה מספר 94 מציגה, את תוצרי החיבור הסופי בין נתוני הלמ"ס לבין נתוני מד"א. בריבוע בכחול נתוני קובץ הלמ"ס (המבוססים על נתוני משטרת ישראל) באדום נתוני מד"א. הסימון הירוק עמודות *Between Range Result* - ערך 1 משמעו כי ערכי <u>השעה</u> נופלים בטווח תואם בין שני הקבצים. ערך 0 משמעו - ערכים שאינם תואמים. בעמודת : *Bayit NumComparision* - ערך 1 מצביע על ערכי <u>מספר בית</u> תואמים בין שני הקבצים. ערך 0 משמעו ערכים שאינם תואמים.

טבלה מספר1 – תוצר החיבור בין קובץ מד"א לקובץ הלמ"ס – אזיור חיפה (מידע של שלושה חודשים)

Number	Rehov	City	Date_	Time_	Build_Num_Calc	Date_Final	Val_min	Val_max	ValMinAdd3...	ValMaxA...	STR_NAME1	BetweenRange_Result	BayitNumCoparison
94	שד מוריה	חיפה	Feb 2 2005	10:57:44	94	Feb 2 2005	10:30	10:44	10:00:00	11:14:00	שד מוריה	1	1
127	דרך אלנבי	חיפה	Feb 21 2005	19:34:55	127	Feb 21 2005	19:30	19:44	19:00:00	20:14:00	דרך אלנבי	1	1
127	דרך אלנבי	חיפה	Feb 21 2005	19:36:16	127	Feb 21 2005	19:30	19:44	19:00:00	20:14:00	דרך אלנבי	1	1
67	ארלוורוב	חיפה	Feb 24 2005	13:05:29	60	Feb 24 2005	13:00	13:14	12:30:00	13:44:00	ארלוורוב	1	0
19	ארלוורוב	חיפה	Jan 2 2005	21:21:59	19	Jan 2 2005	21:15	21:29	20:45:00	21:59:00	ארלוורוב	1	1
77	הגליל	חיפה	Jan 3 2005	17:49:14	77	Jan 3 2005	11:15	11:29	10:45:00	11:59:00	הגליל	0	1
18	כורי	חיפה	Jan 3 2005	08:13:56	20	Jan 3 2005	08:00	08:14	07:30:00	08:44:00	כורי	1	0
33	דרך צרפת	חיפה	Jan 4 2005	17:16:34	38	Jan 4 2005	17:30	17:44	17:00:00	18:14:00	דרך צרפת	1	0
1	קליבנוב	חיפה	Jan 4 2005	20:11:00	1	Jan 4 2005	20:15	20:29	19:45:00	20:59:00	קליבנוב	1	1
114	דרך הים	חיפה	Jan 7 2005	03:08:56	114	Jan 7 2005	03:00	03:14	02:30:00	03:44:00	דרך הים	1	1
14	דרך הים	חיפה	Jan 7 2005	03:08:56	114	Jan 7 2005	03:00	03:14	02:30:00	03:44:00	דרך הים	1	0
14	דרך הים	חיפה	Jan 7 2005	03:13:27	114	Jan 7 2005	03:00	03:14	02:30:00	03:44:00	דרך הים	1	0
9	חטיבת גולני	חיפה	Jan 8 2005	21:14:32	16	Jan 8 2005	21:30	21:44	21:00:00	22:14:00	חטיבת גולני	1	0

כפי שניתן לראות ההבדלים בטבלת התוצר, מתקיימים בעיקר בין מספרי הבתים. משמע קיים הבדלים בין מספר הבית המדווח בקובץ מד"א, לבין מס' הבית המדווח על ידי הלמ"ס (למעשה המשטרה). בכדי לבחון את התוצאות (ברמה מדגמית), יצאנו לשטח בחיפה. טבלה מספר 95 מציגה את תוצאות הבחינה

טבלה 100 - תוצאות בחינת השטח - בחיפה

רחוב דיווח מד"א	רחוב – דיווח למ"ס	הערות
דרך צרפת 33-38	דרך צרפת 33-38	אחד מול השני
חטיבת גולני 9	חטיבת גולני 16	אחד מול השני
כורי 18	כורי 20	אחד מול השני
ארלוזרוב 67	ארלוזרוב 60	נמצאים במרחק של מספר בתים האחד מהשני
דרך הים 14 + 114	דרך הים 114	נראה כי בקובץ מד"א נשמט המספר 1 (נכתב 14 במקום 114). אותו הדיווח מופיע פעמים באותה השעה עם שני מספרי בתים שונים (המשמעות היא שנשלחו יותר משני אמבולנסים).

תוצאת הבחינה מצביעה על ישימות עקרונית לביצוע בקרת איכות על פי נתוני מד"א, בעת עדכון נתונים בתאונות במקטע העירוני.

נספח מספר 2 – חישוב גובה ואזימוט השמש – שאילתות SQL

שאילתה מספר 1 (*q1_Calcdate*):

```
SELECT Final_2005_NoError_wgs84.*, CDate([YOM_BE_HOD] & "/" & [HODESH_TEU] & "/" & [SHNAT_TEUN]) AS [Date] INTO
CalcDate_q1
FROM Final_2006_NoError_wgs84 AS Final_2005_NoError_wgs84;
```

שאילתה מספר 2 *(q2_Calc_Hour)*:

```
SELECT CalcDate_q1.*, DateNum.Date, DateNum.N, Left([Value],5) AS [Time], Left([time],2) AS HourOrig, Left([time],2) AS
HourFixSummerTime, Right([time],2) AS MinuteOrig, ((Right([time],2))+7)/60 AS MinuteMidlle4Furmula INTO q2_CalcHour_Res
FROM (CalcDate_q1 LEFT JOIN DateNum ON CalcDate_q1.Date = DateNum.Date) INNER JOIN [Time_table_ 93] ON
CalcDate_q1.SHAA = [Time_table_ 93].Code;
```

שאילתה מספר 3 (*q3_Update_Minus1Hour*):

```
UPDATE q2_CalcHour_Res SET q2_CalcHour_Res.HourFixSummerTime = [HourOrig]-"1"
WHERE (((q2_CalcHour_Res.CalcDate_q1_Date)
Between #3/31/2006# And #10/1/2006#) AND ((q2_CalcHour_Res.HourOrig)<>"00;(("
```

שאילתה מספר 4 (*q4_FinalHure4Furmula*):

```
SELECT q2_CalcHour_Res.*, [HourFixSummerTime]+[MinuteMidlle4Furmula] AS FinalHure4Fur
 INTO q3_FinalHure4Fur
FROM q2_CalcHour_Res;
```

שאילתה מספר 5:

```
SELECT q3_FinalHure4Fur.*, (360/365.25)*([N]+[FinalHure4Fur]/24) AS g, ([g]/360)*2*3.14159265358979 AS gRad, 0.396372-
22.91327*Cos([gRad])+4.02543*Sin([gRad])-0.387205*Cos(2*[gRad])+0.051967*Sin(2*[gRad])-
0.154527*Cos(3*[gRad])+0.084798*Sin(3*[gRad]) AS D, 0.004297+0.107029*Cos([gRad])-1.837877*Sin([gRad])-
0.837378*Cos(2*[gRad])-2.340475*Sin(2*[gRad]) AS TC, ([FinalHure4Fur]-12)*15+[Xdd]+[TC] AS SHA
INTO q4_res
FROM q3_FinalHure4Fur;
```

שאילתה מספר 6 א' (*q6A_updateSHA*):

```
UPDATE q4_res SET q4_res.SHA = [SHA]-360
WHERE (((q4_res.SHA)>180;((
```

שאילתה מספר 6 ב' (q6B_updateSHA):

```
UPDATE q4_res SET q4_res.SHA = [SHA]+360
WHERE (((q4_res.SHA)<-180;((
```

שאילתה מספר 7(q7):

```
SELECT q4_res.*, ([D]/360)*2*3.14159265358979 AS Drad, ([SHA]/360)*2*3.14159265358979 AS SHArad,
([Ydd]/360)*2*3.14159265358979 AS YddRad, Sin([YddRad])*Sin([Drad])+Cos([YddRad])*Cos([Drad])*Cos([SHArad]) AS
COS_SZA, 1.111 AS SZA
INTO q5_res
FROM q4_res;
```

שאילתה מספר 8 א' (*q8A_UpadetCos1*):

```
UPDATE q5_res SET q5_res.COS_SZA = 1
WHERE (((q5_res.COS_SZA)>1;((
```

שאילתה מספר 8ב' (*q8B_UpadetCos-1*)

```
UPDATE q5_res SET q5_res.COS_SZA = -1
WHERE (((q5_res.COS_SZA)<-1;((
```

שלב 9 (*q9_updae_SZAExcel*) - מבוצעת ב EXCEL, ומעודכן לאחר מכן ב MDB

העתקת עמודת ה COS_SZA ל EXCEL וחישוב זווית ה SZA על ידי חישוב ARC COS במעלות

COS_SZA	SZA
0.256837	75.11751953
0.136693	82.14346718
0.334743	70.44306158

→ SZA =DEGREES(ACOS(P2))

שאילתה מספר 10 (q10):

```
SELECT q5_res.*, 90-[SZA] AS SEA, ([SZA]/360)*2*3.14159265358979 AS SZArad, (Sin([Drad])-
Sin([YddRad])*Cos([SZArad]))/(Cos([YddRad])*Sin([SZArad])) AS COS_AZ, 1.111 AS AZ
 INTO 6_Res
FROM q5_res;
```

שאילתה מספר 11 א' (*q_11AupdateCos_AZ1*)

```
UPDATE 6_Res SET [6_Res].COS_AZ = -1
WHERE ((([6_Res].COS_AZ)<-1;((
```

שאילתה מספר 11 ב' (*q_11AupdateCos_AZ1*)

```
UPDATE 6_Res SET [6_Res].COS_AZ = 1
WHERE ((([6_Res].COS_AZ)>1;((
```

נספח מספר 3 – קוד SQL לחילוץ צבירים לאורך כביש

```
declare @TotalRows int,@counter int,@Gnamber int, @flag int
declare @concat nvarchar(200)

declare @TmpTbl table
(
	[pk_teuna_f] [float] NULL,
	[KVISH1] [float] NULL,
	[AutoNum] [smallint] NULL,
	[diff_km] [float] NULL,
	[Avg_calc] [float] NULL,
	[GroupNumber] [int] NULL
)


set @counter=-1
set @TotalRows=1667
set @Gnamber=0

declare @SumDist float, @CntTmpTblRows int
set @SumDist=0
set @CntTmpTblRows=0
declare @RoadNum int ,@RoadNumNextRow int

while @counter<=@TotalRows
begin

		set @counter=@counter+1
		set @concat = 'Outer counter: '+ cast(@counter as varchar)
		PRINT @concat
		set @concat = 'Outer @CntTmpTblRows: '+ cast(@CntTmpTblRows as varchar)
		PRINT @concat

		set @RoadNum=(select Kvish1 from q2_res where AutoNum=@counter)
		set @RoadNumNextRow=(select Kvish1 from q2_res where AutoNum=@counter+1)


		if   @RoadNum=@RoadNumNextRow
		begin
				declare @diff_km float
				set @diff_km=(select diff_km from q2_res where AutoNum=@counter)

				if @diff_km < 0.2
```

```
		begin
			INSERT INTO @TmpTbl (pk_teuna_f,KVISH1,AutoNum,diff_km)
			select pk_teuna_f,KVISH1,AutoNum,diff_km
			from dbo.q2_res
			where AutoNum=@counter

			set @SumDist=@SumDist+@diff_km
			set @CntTmpTblRows=@CntTmpTblRows+1
			set @Gnamber=@Gnamber+1

				set @concat = '@CntTmpTblRows @ diff_km < 0.2: '+ cast(@CntTmpTblRows as varchar)
				PRINT @concat
		end
		else
			if  @CntTmpTblRows >0
			begin
				declare @Avg_Diff float
				declare @Stdev_ float
				set @Avg_Diff=@SumDist/@CntTmpTblRows

				declare @rowind int --in order to know which row to update
				set @rowind=@counter-@CntTmpTblRows
					set @concat = 'Inner counter loop: '+ cast(@counter as varchar)
					PRINT @concat
					set @concat = 'CntTmpTblRows @ update table: '+ cast(@CntTmpTblRows as varchar)
					PRINT @concat
					set @concat = 'Rowindex: '+ cast(@rowind as varchar)
					PRINT @concat
				set @CntTmpTblRows=0
				set @SumDist=0
				while @rowind<@counter
				begin
					update @TmpTbl
					set Avg_calc=@Avg_Diff , GroupNumber=@Gnamber
					where AutoNum = @rowind
					set @rowind = @rowind+1
				end
			end
			else
				continue
end
else
	if  @CntTmpTblRows >0
	begin

			set @Avg_Diff=@SumDist/@CntTmpTblRows
			--declare @rowind int --in order to know which row to update
```

```
                                        set @rowind=@counter-@CntTmpTblRows
                                                set @concat = 'Inner counter loop: '+ cast(@counter as varchar)
                                                PRINT @concat
                                                set @concat = 'CntTmpTblRows @ update table: '+ cast(@CntTmpTblRows as varchar)
                                                PRINT @concat
                                                set @concat = 'Rowindex: '+ cast(@rowind as varchar)
                                                PRINT @concat
                                        set @CntTmpTblRows=0
                                        set @SumDist=0
                                        while @rowind<@counter
                                        begin
                                                update @TmpTbl
                                                set Avg_calc=@Avg_Diff , GroupNumber=@Gnamber
                                                where AutoNum = @rowind
                                                set @rowind = @rowind+1
                                end
                        end

end

select * from @TmpTbl



select * from (select KVISH1,GroupNumber, count(KVISH1)AS Cnt, Stdev(diff_km)as Stdev_a
from @TmpTbl
group by KVISH1,GroupNumber) as AA  INNER JOIN
@TmpTbl as Tmp_ ON aa.KVISH1=Tmp_.KVISH1 and aa.GroupNumber=Tmp_.Groupum
```