

## מתודולוגיה של המפקד האוכלוסין

### כף

מפקד האוכלוסין 2008 היה "מפקד משולב", הראשון מסוגו שנערך בארץ. שיטה זו משלבת בין שיטת "המפקד המינהלי" ושיטת "המפקד המסורתי", ובה נעשה שימוש בקבצים מינהליים לצד הפקידה בשטח. הרעיון הכללי של המפקד המשולב הוא בניית "קובץ מרשם משופר" (הקמ"מ), שמספק תמונה של האוכלוסייה ה"קבועה" במועד הקובע<sup>1</sup> של מפקד האוכלוסין, בשילוב סקרים מדגמיים להערכת טיב המידע בקמ"מ ותיקונו באמצעות חישוב המשקלות. המפקד המשולב מתנהל במספר שלבים. במסמך זה מובא תיאור של תהליכי העבודה המרכזיים של מפקד האוכלוסין 2008:

1. בניית הקמ"מ – קובץ מרשמי משופר
  - 1.1 הגדרת המסגרת של האוכלוסייה המפקדית
  - 1.2 בניית משפחות מינהליות
  - 1.3 עיגון מרחבי
2. סקרי שדה
  - 2.1 סקר השטח
    - 2.1.1 תהליך הפקידה הכללי
    - 2.1.2 תהליך הפקידה של דיירי המוסדות
    - 2.1.3 נושאי השאלון
  - 2.2 סקר עודפי קמ"מ
3. שיטת האמידה
  - 3.1 שימוש במערכת הכפולה לאמידת הפרמטרים של כיסוי חסר וכיסוי עודף
  - 3.2 חישוב המשקלות בקובץ המפקד המשולב ICF
  - 3.3 אמידת המשקלות בקובץ החברתי-כלכלי (SEF) Social Economic File
4. פעולות נוספות בהפקת הקבצים המפקדיים
  - 4.1 בדיקות תהליכיות: בדיקת איכות אומדני האוכלוסייה
  - 4.2 עריכה וזקיפה
    - 4.2.1 עריכה וזקיפה בקובץ ICF
    - 4.2.2 עריכה וזקיפה של קובץ SEF
  - 4.3 הוספת משתנים חדשים
    - 4.3.1 משתנים שעברו סימול
    - 4.3.2 משתנים מחושבים
    - 4.3.3 משתנים מקבצים מינהליים
5. פרסום נתוני מפקד האוכלוסין 2008

<sup>1</sup> המועד הקובע במפקד האוכלוסין 2008 היה במוצאי שבת, ה-27.12.08 (ל" בכסלו תשס"ט).

## 1. כניית הקמ"מ – קובץ מרשמי משוכר

הקמ"מ שנבנה לצרכי המפקד המשולב בישראל, הוגדר כקובץ המכיל את מסגרת אוכלוסיית המדינה במועד הקובע של מפקד האוכלוסין. בבסיס הקמ"מ נמצא מרשם האוכלוסין. מרשם האוכלוסין מכיל מידע דמוגרפי-גאוגרפי על כל אזרח ובעל מעמד של "תושב קבוע" במדינה. על-מנת להפיק ממנו את הקמ"מ הכולל את האוכלוסייה הקבועה בוצעו מספר פעולות מתודולוגיות שעיקרן בהגדרת האוכלוסייה הקבועה במדינה ועיגון כתובות (הצמדת נ.צ. לכל כתובת) ברמה המדויקת ביותר האפשרית ובעיקר לרמת האזורים הסטטיסטיים שהם היחידה הגיאוגרפית הקטנה ביותר שעבורה מחושבים אומדני המפקד<sup>2</sup>.

### 1.1 הגדרת המסגרת האוכלוסייה המפקדית

האוכלוסייה הכלולה במפקד האוכלוסין היא "אוכלוסייה קבועה", כלומר אוכלוסיית האנשים הגרים בדרך-כלל במדינה. לעומת זאת, במרשם האוכלוסין מופיעים גם אנשים שאינם חלק מ"אוכלוסייה הקבועה", כאלה שעזבו את הארץ ואינם מתגוררים בה באופן קבוע, אנשים שמוכרים לרשויות כנעדרים, אנשים שנפטרו אך פטירתם לא עודכנה במרשם מסיבות שונות וכדומה.

הגדרת תושב ישראלי בחו"ל מתבססת על הגדרות האו"ם (המלצות האו"ם למפקדי 2010): "מהגר בין-לאומי הוא כל אדם המשנה את המדינה המשמשת למגוריו בשגרה. מהגר לטווח ארוך הוא אדם העובר למדינה שאיננה משמשת למגוריו בשגרה, לפרק זמן של שנה ויותר. מהגר לטווח קצר הוא אדם העובר למדינה שאיננה משמשת למגוריו בשגרה לפרק זמן של שלושה חודשים לכל הפחות ושנה לכל היותר".

ההגדרה האופרטיבית המתבססת על הגדרה זו חלה על כלל הפרטים הפעילים במרשם האוכלוסין (כאלה שלא נפטרו או חדלו להיות תושבים), ושיצאו את גבולות ישראל למשך שנה (12 חודשים) עם אפשרות לביקורי מולדת בני 90 יום לכל היותר (3 חודשים).

מאגר התושבים הישראליים בחו"ל מורכב ממידע הנגזר ממספר קבצים: מרשם האוכלוסין, קובץ בקורת הגבולות, קובץ "לא נפקדו במפקדים 1983 ו-1995", קבצים דמוגרפים המשמשים לאמידת האוכלוסייה בישראל (של הלמ"ס) וקובץ ביטוח לאומי שבו מוגדרים ה"לא תושבים" (שנעדרו מישראל חמש שנים רצופות, למעט ביקורים המצטברים מתחת ל-90 יום בכל שנה).

תכלית יצירת מאגר תושבים ישראליים בחו"ל היא איתור מנהלי, טרם יציאה לשטח, של פרטים עליהם ניתן לומר בוודאות גבוהה כי מרכז חייהם איננו בישראל.

<sup>2</sup> אזור סטטיסטי הוא יחידת שטח שנוצרת מחלוקה גאוגרפית-סטטיסטית של יישובים עירוניים המונים למעלה מ-10,000 תושבים. אזור סטטיסטי המיועד למגורים מונה בדרך-כלל בין 3,000 ל-5,000 נפשות.

לצורך בניית הקמ"מ בוצעה גריעה של "הלא-שייכים לאוכלוסייה המפקדית" ממרשם האוכלוסין. במסגרת פעולה זו, נגרעו ממרשם האוכלוסין גם רשומות שבהן היו פרטים בסיסיים לקויים (כגון שנת לידה חסרה, שם משפחה חסר או יישוב מגורים חסר), זאת בהנחה שהינם אינדיקטורים לכך שהרשומה אינה מייצגת תושב קבוע/אזרחי במדינה.

### להלן התפלגות של הלא-שייכים שנגרעו ממרשם האוכלוסין:

מספרים מוחלטים	לא-שייכים לאוכלוסיית המפקד, לפי סיבה
517,663	ישראלים המתגוררים בחו"ל
3,163	חסרים פרטים (שנת לידה, שם משפחה, יישוב מגורים)
61	נעדרים
12,117	אחר
533,004	סה"כ

### 1.2 בניית משפחות מינהליות

פעולה נוספת שבוצעה בקמ"מ שנועדה ליעל את סקרי השדה, היא בניית משפחות מינהליות. משפחה מינהלית מורכבת מאנשים בעלי קשרים משפחתיים מרשמיים הגרים באותה כתובת, על-פי מרשם האוכלוסין. משפחה מינהלית נבנתה על בסיס מרשם האוכלוסין ומהווה תשובה מינהלית להגדרה של משקי-בית משפחתיים. משק בית משפחתי הוא קבוצה של אנשים בעלי קרבה משפחתית, הגרים באותה דירה ויש להם תקציב הוצאות משותף למזון. איחוד הפרטים למשפחות מינהליות נעשה באמצעות מידע על קשרים משפחתיים המבוסס על הרשום של האב/האם ושל בן/בת הזוג בנתוני המרשם של הפרט, מהם נגזרים יחסי הקרבה ומידע על כתובות.

הגדרת המשפחות המינהליות תורמת לאורך כל שלבי העבודה במפקד המשולב: בשלב של סקר השדה הטלפוני, מתאפשר איתור של כל הפרטים במסגרת של משקי בית, דבר שמונע פניות חוזרות לאותו משק בית, כשנאספים הנתונים עבר כמה בני משק הבית (מפורט בהרחבה בסעיף 2.2).

בשלב של עיבוד המידע שנתקבל מהשטח, בתהליך של זקיפת הנתונים החסרים, משמשות המשפחות המינהליות כ"משקי בית תורמים" ומספקים "השלמות" של מידע חסר (מפורט בהרחבה בסעיפים 4.2.1, 4.2.2).

בשלב של בניית הקמ"מ, אחת הפעולות שנעשות היא קביעת כתובת יחידה עבור כל הפרטים במדינת ישראל. פעולה זו מבוצעת באמצעות העיגון המרחבי (מפורט בהרחבה בסעיף הבא, 1.3), שהופכת לפעולה פשוטה יחסית כשהפרטים מקובצים למשפחות.

## ציון מרחבי

עיון (Geocoding) הוא פעולת קישור הרשומות לישות גיאוגרפית (מרחבית), דרך מזהים מרחביים, כלומר, שיוך פרט כלשהו למיקומו הגיאוגרפי, ברמת דיוק גבוהה ככל הניתן, באמצעות מאפייני כתובת מלאה ככל האפשר (שם ישוב, שם רחוב ומספר בית). ובשאיפה למקם כל פרט באוכלוסייה לאזור הסטטיסטי שבו הוא רשום במרשם.

תהליך העיון המרחבי נערך בשני שלבים:

1. עיון אוטומטי – בשלב זה, מבוצע קישור רשומות בין הרשומות בקובץ לבין שכבות או/ו טבלאות שונות של ישויות גיאוגרפיות באמצעות תוכנה לעיון אוטומטי. התוכנה לעיון אוטומטי פועלת על-פי שלבים מוגדרים של התהליך, על-מנת לעגן באופן מדויק וברמת פירוט מרבית. רשומות שעבורן העיון האוטומטי לא צלח עוברות לשלב הבא.

2. עיון ידני – בשלב זה מבוצע עיון ידני תמוך מחשב על בסיס מידע ייחודי ממקורות אחרים, לרוב היסטורי, שאינו קיים במערכת האוטומטית.

בסוף התהליך של עיון מרחבי, עוגנו 98.7% מהרשומות בקובץ המינהלי המשופר לרמת אזור סטטיסטי, מתוכן 78.1% עוגנו לרמת הדיוק המרבית – המבנה.

פעולת העיון המרחבי של אוכלוסיית הקמ"מ משמשת לצורך אמידת כיסוי החסר וכיסוי העודף של אזורים סטטיסטיים<sup>3</sup>, ולכן העיון חייב להתבצע לרמה של אזור סטטיסטי.

פעולת העיון המרחבי (האוטומטי והידני) עוברת בהמשך סדרת בדיקות איכות שונות. במספר יישובים עירוניים המחולקים לאזורים סטטיסטיים אין מערך כתובות מסודר, ברובם יישובים ערביים ולכן אין אפשרות לעגן בהם את האוכלוסייה לאזורים סטטיסטיים. במקרים אלה בוצעה זקיפה של א"ס ע"י "משיכה" של תוצאות העיון המרחבי ממפקד 1995 באמצעות קישור לפי מספר זהות של הפרט שנפקד במפקד 1995 או אם לא נפקד ב-1995 לפי המיקום של בני משפחתו שנפקדו. לדוגמא עבור ילדים שנולדו אחרי מפקד 1995 נמשכו תוצאות העיון של הוריהם.

## 2. סקרי שדה

שני סקרי שטח שנועדו לאמידת חוסר כיסוי ועודף כיסוי ברמת האזור הסטטיסטי (פירוט מלא של שני הסקרים ראה בסעיפים 2.1, 2.2):

סקר שטח שבו המדגם נבחר מתוך של תאי שטח ושמשהן לאמידת הכיסוי החסר<sup>4</sup> של הקמ"מ והן לאיסוף הנתונים החברתיים-כלכליים של האוכלוסייה המתגוררת בתאי שטח אלו.

<sup>3</sup> ביישובים שאינם מחולקים לאזורים סטטיסטיים, נאמדו הכיסוי העודף והכיסוי החסר של הקמ"מ ברמה של כלל היישוב.

<sup>4</sup> כיסוי חסר של הקמ"מ מתייחס לאנשים הגרים באזור סטטיסטי מסוים, אך רשומים באזור סטטיסטי אחר בקמ"מ.

סקר טלפוני שבו המדגם נבחר מתוך פרטים של הקמ"מ ושמש לאמידת הכיסוי העודף<sup>5</sup> של קמ"מ.

## 2.1 סקר השטח

סקר השטח מכונה "סקר U" (Under-coverage). מדינת ישראל מחולקת לכ-3,000 אזורים סטטיסטיים. כל אזור סטטיסטי חולק לתאי שטח, כך שבכל תא שטח היו כ-50 משקי בית בממוצע. מכל אזור סטטיסטי נדגמו בצורה אקראית כ-17% מהתאים, ובתאים אלו בוצעה פקידה מלאה בשיטת "המפקד המסורתי". **בסקר השטח** נפקדה כל האוכלוסייה בדירות שנמצאו בתאי השטח הנדגמים באמצעות שאלון ממוחשב. שאלון סקר מפקד האוכלוסין 2008 הורכב מנושאי חקירה רבים: דמוגרפיה, מקום מגורים, מוגבלות, השכלה, עבודה, תנאי דיור והכנסות שאינן מעבודה.

הסקר נערך בתאריכים 7-18 בדצמבר 2008 (שלב קדם איסוף – ראה להלן) ומ-28 בדצמבר עד סוף פברואר 2009 (שלב איסוף והשלמת האיסוף – ראה להלן).

### 2.1.1 תהליך הפקידה הכללי

הסוקרים הגיעו לכל תא שטח במדגם מספר פעמים: בפעם הראשונה, **בשלב קדם איסוף**, סרקו הסוקרים את השטח ורשמו את כל המבנים והדירות שכלולים בשטח התאים, תוך מתן סיווג של שימוש המבנה (למשל: מבנה למגורים, מבנה בבנייה וכו'). כמו כן, הופצה איגרת פנייה לאוכלוסייה הכוללת הסבר על מטרות הסקר. בפעם השנייה, **בשלב איסוף**, חזרו הסוקרים לכל הדירות שרשמו בשלב הקודם וראיינו את האוכלוסייה במשקי הבית באמצעות שאלון ממוחשב ואספו נתונים דמוגרפיים, חברתיים וכלכליים עבור כל משקי הבית בתאים הדגומים. **בשלב השלמת האיסוף**, חזרו הסוקרים כדי לראיין את הדיירים בדירות שבהן עדיין לא מולא השאלון. בדירות שהיו סגורות ובדירות סרבנים הושאר שאלון להשבה עצמית על-ידי הדיירים<sup>6</sup>.

### 2.1.2 תהליך הפקידה של דיורי המוסדות

ההגדרה המפקדית של "מוסד" או בהגדרתו הרשמית "יחידת מגורים משותפים" (collective living quarters) היא "יחידה מינהלית המספקת שירותי לינת לילה, ובדרך-כלל גם שירותי אכילה, לחמישה דיירים לפחות". תהליך פקידה מיוחד ונפרד הוגדר עבור דיורי המוסדות, מכיוון שזו אוכלוסייה בעלת מאפיינים ייחודיים הן באורח החיים והן בצורת הדיור:

<sup>5</sup> כיסוי עודף מתייחס לאנשים הרשומים בקמ"מ באזור סטטיסטי מסוים, אך גרים באזור סטטיסטי אחר או בחו"ל.  
<sup>6</sup> שאלון השבה עצמית הוא שאלון נייר קצר שהופץ בשפות העברית, הערבית הרוסית. הוא כולל מספר שאלות דמוגרפיות ונועד למילוי עצמי על-ידי הדיירים. כל שאלון הגיע במעטפה שכללה את דף השאלון, נוסח פנייה לדיירים ומעטפה מבולת וממוענת ללמ"ס. הדיירים הונחו לשלוח את השאלון המלא בדואר.

1. מוסדות הם מסגרות של דיור מרוכז, המספקים מענה לצרכים מיוחדים של אוכלוסייה

בעלת מאפיינים הומוגניים: מעונות סטודנטים, פנימיות, הוסטלים לבעלי מוגבלויות, בתי אבות, מקבצי דיור לעולים, מסגרות שיקום, בתי כלא וכדומה. הנהלת המוסדות מנהלת ברוב המוסדות מערכת רישום של הדיירים מחד ומצד שני מנטרת את אפשרות הגישה הישירה לדיירים מסיבות בטיחותיות ומסיבות אחרות. כך שפקידת הדיירים מחייבת שיתוף הפעולה של הנהלת המוסד.

2. חלק גדול מדיירי המוסדות אינם במצב בריאותי או נפשי או במעמד חוקי שמאפשר בצוע ריאיון פנים אל פנים עם הנדגם.

לצורך פקידת דיירי המוסדות, בשלבי תכנון תהליכי הפקידה במפקד הורכבה מסגרת ארצית של "מוסדות לפקידה". המסגרת נבנתה באמצעות פנייה ל"ארגוני גג" (ארגונים ממשלתיים, ציבוריים, מלכ"רים וארגונים פרטיים המנהלים מוסדות). "ארגוני הגג" התבקשו למסור רשימות של מוסדות הנמצאים תחת חסותם. הרשימה עודכנה שוב סמוך לתהליך החקירה בשדה ע"י פנייה חוזרת לארגוני הגג ובשלב זה נתבקשה גם הסכמתם לבצוע תהליך הפקידה. בפנייה זו נמסר פירוט מלא של התהליך הצפוי, לוח הזמנים והנתונים המבוקשים. בשלב קדם-איסוף, ביקרו הסוקרים במוסדות, עדכנו נתונים ואספו, מכל מוסד, את רשימות הדיירים השוהים בהם.

#### דגימה ופקידה של דיירי המוסדות:

דיירי המוסדות חולקו לשתי קטגוריות, על-פי צורת הדיור שלהם:

**דיירים המתגוררים במסגרת משקי בית** בתוך המוסד – דיירים המתגוררים ביחידת דיור ומנהלים תקציב משק בית עצמאי (לדוגמה: דיור מוגן, מרכזי קליטה, מקבצי דיור לעולים). בקטגוריה זו נכללו גם אנשי צוות במוסדות.

**דיירים יחידים** – דיירים שאינם מנהלים תקציב הוצאות משק בית עצמאי והמוסד מספק את רוב צרכיהם (לדוגמה: אסירים, בעלי מוגבלויות, קשישים סיעודיים, פנימיות נוער וכד'). כל מוסד התבקש לספק שתי רשימות של דיירים: רשימת הדיירים המתגוררים במשקי בית ורשימת הדיירים היחידים.

מכל אחת מרשימות המוסד נדגמו 20% מהדיירים בני 15 ומעלה, למילוי שאלון חברתי-כלכלי. למדגם הדיירים המתגוררים במשקי הבית פנה סוקר וביצע ריאיון פנים אל פנים באמצעות המחשב. למדגם הדיירים היחידים הופץ שאלון נייר מותאם למילוי עצמי (שכולל גם שאלות חברתיות-כלכליות) והוא הוקלד למחשב על-ידי סוקר. בסיום תהליך הפקידה של כלל דיירי המוסדות, התקבל מידע דמוגרפי עבור 100% מאוכלוסיית המוסדות ומידע חברתי-כלכלי עבור 20% מהדיירים.

### 2.1.3 נושאי השאלון

לקראת כל מפקד (אחת לעשור בקירוב), מתגבשות באו"ם ההמלצות וההצעות לגבי נושאי החקירה שיופיעו בשאלון המפקד. המלצות האו"ם לגבי נושאי החקירה משרתות את הצרכים המקומיים והלאומיים של מדינות ומאפשרות להשוות בין נתוני המדינות השונות בעולם. לגבי כל נושא חקירה שממליץ האו"ם לכלול בשאלון, קיימים שני סוגי משתנים: משתנים בסיסיים (Basic Topics) ומשתנים נוספים (Additional Topics) המותאמים לצרכים האינדיווידואליים של כל מדינה ומדינה. כך, למשל, בקבוצת משתני דיור, "מספר החדרים בדירה" מופיע כמשתנה חקירה בסיסי, ואילו "שטח הדירה" הוא משתנה נוסף, שהכללתו בחקירה נתונה לשיקול דעת בהתאם לצרכיה של כל מדינה. לצורך גיבוש נושאי החקירה שיכללו בשאלון מפקד 2008 הוקמו וועדות היגוי במטרה לזהות ולאתר את הצרכים המקומיים. השאלות שנכללו בשאלון המפקד עברו התנסויות בשטח<sup>7</sup> שבוצעו במסגרת סקרי האוכלוסין לקראת המפקד ב-2008. כמו כן, נשמרה המסגרת הנושאת של שאלון מפקד 1995, זאת כדי לאפשר השוואת נתונים וזיהוי תמורות על פני הזמן שבין שני המפקדים. לגבי חלק מהנושאים (הכנסות מעבודה, קושי לראות) הוחלט שלא להכניסם לשאלון המפקד וזאת עקב זמינות המידע הנחוץ בקבצים המינהליים של משרדים ממשלתיים וגופים אחרים.

השאלון של מפקד האוכלוסין 2008 הורכב מנושאי חקירה רבים: דמוגרפיה, מקום מגורים, מוגבלות, השכלה, עבודה, תנאי דיור והכנסות שאינן מעבודה. פירוט השאלות נמצא ב"שאלון סקר מפקד האוכלוסין 2008".

### 2.2 סקר צדפי קמ"מ

מדגם מתוך האוכלוסייה בקמ"מ, המכונה "מדגם O" (Over-coverage). המדגם כלל כ-600,000 פרטים הרשומים בקמ"מ בתאים שנדגמו. הנדגמים נחקרו לגבי כתובתם במועד הקובע בלבד. בסקר זה לא נחקרו נושאים חברתיים כלכליים. מאחר והמדגמים היו תלויי כתובת מרשמית חלק מהנדגמים נפקדו בסקר השטח, ולכן לא היה צורך לחזור ולחקור אותם לגבי כתובתם במועד הקובע. אי לכך נדגמים אלה נגרעו ממדגם סקר עודפי מרשם. הנדגמים אשר לא נפקדו במדגם תאי השטח או שהמידע שנתקבל עבורם בשטח היה חלקי, נפקדו בסקר טלפוני, כאמור נדגמים אלה נחקרו אודות פרטי כתובת המגורים במועד הקובע, הכתובות שדווחו בסקר עוגנו מאוחר יותר לאזורים סטטיסטיים. הסקר הטלפוני נערך מ-25 במרס עד תחילת יולי 2009.

<sup>7</sup> ניסוי 2002 נערך ביישוב בית שמש; ניסוי 2004 נערך ביישובים: גבעתיים, טירה, נווה ימין, ירחיב, ואלישמע וניסוי 2006 נערך ב-25 יישובים: אלעד, באר יעקב, בן שמן, בני עי"ש, גלג'וליה, גדרה, כפר קאסם, לוד, מודיעין, מזכרת בתיה, רמלה, בית חשמונאי, משמר איילון, יד רמב"ם, עזריה, בארות יצחק, גבעת ברנר, גזר, משמר דוד, נחשונים, נצר סירני, עינת, צובה, חורה והיישוב המוסדי בן שמן. להרחבה ראה:

[http://www.cbs.gov.il/census/census/pnimi\\_page.html?id\\_topic=4](http://www.cbs.gov.il/census/census/pnimi_page.html?id_topic=4)

עם סיום עבודת השדה, התקבלו שלושה מסדי הנתונים: נתוני סקר תאי השטח, נתוני סקר המוסדות ונתוני סקר עודפי הקמ"מ. שלושת מסדי הנתונים יחד עם הקמ"מ למועד הקובע, שימשו בתהליך חישוב המשקלות לקבלת האומדנים המפקדיים כפי שיפורט בהמשך.

### 3. שיטת האמידה

בחלק זה מתואר תהליך בניית המשקלות עבור שני הקבצים המפקדיים:

#### א. קובץ המפקד המשולב – (ICF) Integrated Census File

קובץ זה כולל את כל הפרטים הרשומים בקמ"מ, (100% האוכלוסייה); לכל פרט בקמ"מ כלולים הנתונים הדמוגרפיות המופיעים במרשם האוכלוסין ונוסף לו משתנה "משקל". אומדן האוכלוסייה של כל יישוב ושל כל אזור סטטיסטי ביישוב מתקבל באמצעות סכימת המשקל של כל הפרטים שכתובתם בקמ"מ היא היישוב/האזור הסטטיסטי.

#### ב. קובץ חברתי-כלכלי – (SEF) Social Economic File

הקובץ החברתי-כלכלי מורכב ממשקי בית והפרטים שנפקדו בשטח (כ-14% מהאוכלוסייה); וכולל את התכונות הדמוגרפיות, החברתיות-כלכליות שנחקרו בשאלון המפקד ו"משקל". משקל זה בהגדרה שונה מהמשקל של קובץ המפקד המשולב. יחד עם זאת בוצע תהליך כיול כדי למזער את ההבדלים באומדני האוכלוסייה ברמות יישוב, לגבי שיטת הכיול ראה בהמשך מסמך זה.

### 3.1 שימוש במערכת הכפולה לאמידת הפרמטרים $\lambda$ כיסוי חסר וכיסוי

#### צדף

שיטת האמידה של סך האוכלוסייה מבוססת על מתודולוגיה של אומדן המערכת הכפולה, ה-Dual System (DS) Estimators. השיטה מתמקדת בבעיית כיסוי החסר ברשומות שנוצרו באופן בלתי תלוי, בקובץ המינהלי ובנתוני מדגם השטח. בהתבסס על אי תלות בין שתי הרשומות אומדים את ההסתברות של אדם השייך לאוכלוסייה (מתגורר בפועל בא"ס/יישוב) להיות רשום בקמ"מ בכתובת זו (פרמטר כיסוי החסר). היפוך אומדן ההסתברות (אחד חלקי אומדן ההסתברות) מאפשר "לנפח" את אוכלוסיית הקמ"מ בא"ס/יישוב בגין כיסוי החסר. במחקר מתודולוגי חדשני של הלמ"ס, הורחב המודל הקלאסי המטפל בכיסוי חסר למודל שנותן מענה לכיסוי העודף של הקמ"מ.

באופן כללי, ידוע כי הקמ"מ מכיל כיסוי עודף, הכולל את האנשים הרשומים בקמ"מ באזור מסוים, אך בפועל גרים במקום אחר או בחו"ל. בעזרת מדגם הקמ"מ, התקבל אומדן של מספר האנשים שרשומותיהם מהוות כיסוי עודף של הקמ"מ וחושב אומדן לפרמטר כיסוי העודף. אומדן זה מאפשר "לכווץ" את ספירות הקמ"מ בגין כיסוי עודף.



אומדן ה-DS המורחב של סך האנשים בקבוצה מסוימת מבוסס על סך הרשומים בקמ"מ בקבוצה לאחר שקלול באומדני טעויות הכיסוי.

קיום אומדן ה-DS המורחב מבוסס על ההנחות הבאות:

1. הספירות בשתי הרשימות, קמ"מ ומדגם שטח, בלתי תלויות.
2. הסתברויות הספירה הן הומוגניות, כלומר, לכל האנשים באוכלוסייה הסתברות זהה להיספר בקמ"מ. כמו כן, לכל האנשים הסתברות זהה להיספר במדגם.
3. התפלגות כיסוי העודף על פני תאי הפקידה פרופורציונלית להתפלגות באוכלוסייה.

כדי לקיים את אי-התלות בין שתי הספירות (ראו ההנחה הראשונה), נשמרה הפרדה מוחלטת בין הקמ"מ לעבודת השטח. שתי ההנחות האחרות מצריכות חלוקה של האוכלוסייה לקבוצות אמידה הומוגניות ביחס לטעויות הכיסוי על פי משתנים אשר יכולים להשפיע על כיסוי החסר והעודף.

תהליך האמידה מורכב משלבים אלה:

1. קביעת קבוצות אמידה הומוגניות מבחינת טעויות הכיסוי;
  2. חישוב אומדנים לפרמטרים של כיסוי חסר וכיסוי עודף;
  3. חישוב המשקל המפקדי לכל פרט בקובץ המפקד הסופי בהתאם לקבוצת האמידה שאליה הוא שייך.
- בסיום התהליך, מתקבלים אומדני סך האוכלוסייה בחתכים שונים וקבצים מפקדיים, שכוללים משקל מפקדי לכל רשומה בקמ"מ.

### 3.2 חישוב המשקלות בקובץ המפקד האמיתי ICF

קובץ ה-ICF של מפקד האוכלוסין 2008 מורכב מכלל הפרטים שרשומים בקמ"מ (100% האוכלוסייה, כולל דיירי המוסדות), בתוספת המשקל המפקדי של כל פרט. המשקל המפקדי של כל פרט הוא מקדם שמבטא את יחס מאזן העודף והחסר בכיסוי של הקמ"מ. הוא מייצג את הסבירות של כל פרט (ושל כל הפרטים שהוא מייצג, כלומר הפרטים השייכים לקבוצת האמידה שלו) להתגורר בכתובת שרשומה עבורו בקמ"מ. אומדנים מפקדיים מתקבלים מסכימת סך המשקלות של קבוצה נתונה.

**קבוצת האמידה** היא אוסף של פרטים בעלי תכונת הומוגניות בפרמטרים הבאים: כיסוי חסר מרשמי, כיסוי עודף מרשמי, כיסוי חסר שטח.

על סמך הניסיון שהצטבר בניסויים שקדמו לביצוע המפקד המשולב, הורכבו קבוצות האמידה מהשילוב של **יישוב, אזור סטטיסטי וקבוצות גיל**. ארבע קבוצות הגיל שנמצאו כמתאימות ביותר לקיום ההנחות הן: 0-19, 20-29, 30-39, +40.

עבור כל קבוצת אמידה חושב אומדן הגודל בשיטת המערכת הכפולה המורחבת, לפי הנוסחה:

$$\hat{N} = \frac{U_{1+}^G}{\hat{p}_{1+} + \hat{\lambda}}$$

כאשר:

$\hat{N}$  - אומדן לגודל קבוצת אמידה.

$U_{1+}^G$  - כל הפרטים הרשומים בקמ"מ בקבוצת האמידה.

$\hat{p}_{1+} = \frac{U_{11}^S}{U_{+1}^S}$  - אומדן להסתברות להיות רשום נכון בקמ"מ בקבוצת האמידה. הסימון  $S$  מסמל

את התאים שנדגמו.

כך ש:

$U_{11}^S$  - סך הפרטים הרשומים בקמ"מ בקבוצת האמידה בתאים שנדגמו ושנפקדו במדגם  $U$ .

$U_{+1}^S$  - סך הפרטים בקבוצת האמידה שנפקדו בתאים שנדגמו.

$\hat{\lambda} = \frac{X^S}{\left\{ \frac{U_{1+}^S}{\hat{p}_{1+}} \right\}}$  - אומדן לכיסוי העודף בקבוצת האמידה.

עם:

$X^S = \sum_{i \in S} x_i$  אומדן לסך הפרטים העודפים (רשומים באזור הסטטיסטי, אך לא גרים שם)

בתאים שנדגמו בקבוצת האמידה.

$$x_i = \begin{cases} 1 & \text{תוצאת הפקידה של נדגם } i \text{ במדגם } 0, \text{ כך ש:} \\ p & \left\{ \begin{array}{l} \text{אם נמצא כי פרט } i \text{ רשום בא"ס, אך לא גר בו} \\ \text{ההסתברות להיות רשום בא"ס, ללא מגורים בו} \end{array} \right. \\ 0 & \text{פרט } i \text{ רשום בא"ס וגם מתגורר בו} \end{cases}$$

ההסתברות להיות רשום באזור סטטיסטי, ללא המגורים בו, ה- $p$ , חושבה כפרופורציית הרשומים באזור סטטיסטי ללא מגורים מקרב פרטים בקבוצת האמידה שכן השיבו.

נסמן ב-  $U_{1+}^{GS}$  את סך כל הפרטים הרשומים בקמ"מ בתאים שנדגמו. אומדן לסך הפרטים

הרשומים בתאים נדגמים שאכן גרים שם  $U_{1+}^S$  מתקבל על-ידי:

$$U_{1+}^S = U_{1+}^{GS} - X^S$$

המשקל המפקדי עבור כל פרט חושב באופן הבא (i מסמן את קבוצת האמידה):

$$\hat{w}_i = \frac{1}{\hat{p}_{1+i} + \hat{\lambda}_i}$$

### כסיוט התהליך *de* חישוב המשקל

- קיבל כל פרט ב-ICF את המשקל המפקדי שלו; פרטים השייכים לאותה קבוצת אמידה קיבלו משקל זהה;
- כל הפרטים שנפקדו במוסדות, קיבלו משקל "1", מכיוון שנפקדו כל דיירי המוסדות במדינה;
- המשקל הממוצע בקובץ הדמוגרפי קרוב ל-1;
- אומדן האוכלוסייה של כל יישוב ושל כל אזור סטטיסטי ביישוב מתקבל באמצעות סכימת המשקל של היישוב/האזור הסטטיסטי.

### 3.3 אמידת המשקלות בקובץ החברתי-כלכלי (SEF) Social Economic File

הקובץ החברתי-כלכלי מורכב ממשקי הבית ומהפרטים שנפקדו בשטח (כ-14% מהאוכלוסייה), בתוספת תכונותיהם הדמוגרפיות ותכונותיהם החברתיות-כלכליות. לכל פרט ב-SEF מוצמד המשקל, אשר מבטא את אומדן הפרטים שהוא מייצג. הקובץ SEF מכיל מידע חברתי-כלכלי שנאסף בשאלון המפקד ברמת הפרט וברמת משק הבית: השכלה, עבודה, מוגבלות, הגירה פנימית, תנאי דיור והכנסות; הוא מהווה מקור מידע המאפשר לאפיין יחידות גאוגרפיות גדולות כקטנות לפי התכונות החברתיות-כלכליות.

בקובץ המדגם של מפקד 1995 משקלם של כל הפרטים היה זהה ושווה להיפוך שבר הדגימה, כלומר ל-5. מצב זה יצר טבלאות עם שכיחויות שונות, בהתאם לסוג הקובץ שממנו בוצעו החישובים: הקובץ המפקדי המלא של מפקד 1995 או קובץ המדגם. כך לדוגמה, גודלה של גבעתיים היה 48,851 על סמך החישוב שבוצע לפי הקובץ המפקדי המלא, לעומת 48,530 על סמך החישוב שבוצע לפי קובץ המדגם.

בקובץ החברתי-כלכלי של מפקד 2008, חושב המשקל לכל הפרטים ולמשקי הבית שנפקדו בשטח (כ-14% מהאוכלוסייה), כך שסכימת המשקלות ברמת היישובים מפיקה אומדני אוכלוסייה זהים בשני הקבצים.

המשקל בקובץ החברתי-כלכלי מייצג הן את הדגימה והן את מודל המערכת הכפולה. המשקל מקיים את התכונות הבאות:

1. המשקל של כל אחד מהפרטים במשק הבית הוא זהה;
2. המשקל מייצג הן את תכונות הפרט והן את תכונות משק הבית;
3. סכימת המשקל בקובץ ה-SEF תהיה זהה נומרית לסכימת המשקל בקובץ ה-ICF בקבוצות שייקבעו מראש (קבוצות כיול);
4. סכימת המשקל בקובץ ה-SEF תהיה קרובה ככל הניתן לסכימת המשקל בקובץ ה-ICF בחתכים שאינם "קבוצות כיול";

**קבוצות כיול** הן חתכי אוכלוסייה מוגדרים מראש. בקבוצות כיול מבקשים אמידת גודל בעלת הטיה ושונות מינימליות, ולכן דורשים שסכום משקולות הדגימה בהן יהיה זהה, נומרית לזה שמתקבל בקובץ ה-ICF, כלומר, זהה לאומדני המערכת הכפולה. בהגדרה לא ניתן ליצור משקל חברתי-כלכלי המפיק אומדנים זהים לאומדני ה-ICF עבור כל החתכים האפשריים.

היינו רוצים שקבוצות הכיול תהיינה קטנות ככל הניתן. ככל שקבוצות הכיול קטנות יותר קיימת קרבה גדולה יותר לקובץ ה-ICF, אך השונות של המשקל תהיה גדולה יותר (יתקבלו משקלות שליליים ומשקלות גבוהים). דבר זה פוגע באומדנים המתקבלים עבור חתכי אוכלוסייה שאינם מוכלים בקבוצות הכיול.

הוגדרו קבוצות כיול בשלוש רמות:

1. רמת משקי בית – לפי גודל של משקי בית ביישוב: 1, 2, 3, 4+
2. רמת אזור סטטיסטי.
3. רמת הפרטים ביישוב: לפי 7 קבוצות גיל X מין (14 קבוצות).

הכיול נעשה בשיטת "GREG". "GREG" היא שיטת כיול שלוקחת בחשבון אילוצים נתונים. השיטה מתחילה ממשקל התחלתי, G, השווה להיפוך שבר הדגימה (משקל שווה בכל אזור סטטיסטי). השיטה מאתרת משקל חדש, W, המקיים את האילוצים הדרושים וה"קרוב ככל האפשר" למשקל ההתחלתי. הקרבה ניתנת למדידה בפונקציות מרחק שונות, כאן נעשה שימוש במרחק חי-בריבוע:

$$Dist = \sum_i \frac{(W_i - G_i)^2}{G_i}$$

ניתן לבצע GREG רב-שלבי. לדוגמה, אם מעוניינים להתכיל גם לגדלי משקי הבית וגם לקבוצות גיל X מין, ניתן לבצע GREG ראשוני, המתכיל לגדלי משקי בית. התוצאה תניב משקל שימש לצורך המשקל הראשוני, ב-GREG חדש, שיבוצע, הפעם, מכויל לקבוצות גיל X מין. במפקד 2008 בוצע כיוול תלת-שלבי בתוך כל יישוב: כיוול ראשון – עבור גדלי אזורים סטטיסטיים; כיוול שני – עבור 4 גדלי משקי הבית; וכיוול שלישי – עבור 14 קבוצות גיל X מין. שיטה זו מבטיחה כי סכום המשקלות בקובץ ה-SEF תהיה זהה לסכום המשקלות בקובץ ה-ICF עבור כל יישוב ב-14 קבוצות גיל ומין.

### **כסיומ התהליך של חישוב המשקל**

- קיבל כל פרט ב-SEF את המשקל המפקדי שלו, פרטים השייכים למשק בית אחד קיבלו משקל זהה, משקל זהה ניתן לפרטים ולמשק בית שאליו הם שייכים;
- המשקל הממוצע בקובץ החברתי-כלכלי הוא 6.9;
- המשקל בקובץ ה-SEF מכויל למשקל בקובץ ה-ICF, כך שסכימת סך האוכלוסייה ברמת היישובים היא זהה בשני הקבצים. אומדן סך האוכלוסייה ברמה הארצית אינו זהה בשני הקבצים, זאת משום שבקובץ ה-SEF חסרים היישובים שלא נפקדו בסקר השדה השני.

### **4. כצולות נוספות בהפקת הקבצים המפקדיים**

לצורך ההפקה של הקובץ הדמוגרפי (ICF) ושל הקובץ החברתי-כלכלי (SEF) בוצעו מספר פעולות:

- בדיקות תהליכיות שנועדו לאתר ולתקן בעיות בדגימה או/ו בפקידה. בדיקות אלה נעשו לאחר קבלת אומדני האוכלוסייה;
- בדיקות נושאיות שנועדו לשפר את איכות הנתונים על-ידי עריכה וזקיפה של משתנים דמוגרפיים בקובץ ה-ICF ושל משתנים חברתיים-כלכליים בקובץ ה-SEF;
- הוספת משתנים חדשים (משתנים מחושבים, משתנים מסומלים, משתנים ממקורות מינהליים) לקובץ החברתי-כלכלי (SEF).

### **4.1 בדיקות תהליכיות: בדיקת איכות אומדני האוכלוסייה**

לאחר אמידת המשקל עבור כל פרט בקמ"מ, חושבו אומדני האוכלוסייה עבור כלל היישובים בארץ. אומדני האוכלוסייה שהתקבלו בכל יישוב עברו בדיקות, במטרה לאתר ולתקן בעיות בתהליכי המפקד – בדגימה או/ו בפקידה בשטח. הבדיקות בוצעו באמצעות השוואה בין האומדנים החדשים לבין האומדנים המפורסמים לפי סקרים שוטפים של הלמ"ס; הפרשים חריגים הוגדרו כ"חשודים" ונחקרו בצורה יסודית במטרה לגלות את מקורם. בחלק מהמקרים,

הפערים בין האומדנים התבררו כמוצדקים, אך במקרים אחרים נתגלתה הנחיצות של תיקון האומדנים החדשים; תיקונים אלה גררו, לרוב, גם ביטול פרסומם של כלל הנתונים או של הנתונים החברתיים-כלכליים של היחידה הגאוגרפית.

לדוגמה, האומדן החדש של היישוב עין תמר נפסל מכיוון שבשטח נפקדה אוכלוסייה מועטה ובלתי מייצגת. בהתאם לכך, בוטל גם פרסום הנתונים של כלל היישוב. דוגמה נוספת היא היישוב תל יצחק, שבו נדגמו תאי שטח שבהם שכנו בתי אבות בלבד. יישוב זה, על סמך האומדן החדש, הפך שלא בצדק, ל"יישוב מוסדי"<sup>8</sup>. כשבוצע תיקון האומדן, צורת היישוב השתנתה ל"מושב"<sup>9</sup>, אך הנתונים החברתיים-כלכליים שנאספו מדיירי המוסדות בשטח, נפסלו, מכיוון שאוכלוסיית המוסדות איננה מייצגת את אוכלוסיית היישוב, בשל מאפייניה המיוחדים. עם זאת, הותרו לפרסום הנתונים הדמוגרפיים של אוכלוסיית היישוב.

#### 4.2 צריכה ולקיפה

עריכה וזקיפה בוצעו הן למשתנים הדמוגרפיים שבקובץ ה-ICF שנגזרו מתוך מרשם האוכלוסין והן למשתנים חברתיים-כלכליים שבקובץ ה-SEF שנאספו במדגם השטח באמצעות ראיונות ממוחשבים. עריכה וזקיפה בוצעו במטרה לזהות ולאתר טעויות במידע ולטפל בטעויות אלה בצורה סיסטמטית באמצעות מחיקה, חישוב, השלמה או זקיפה של ערכים, זאת כדי לספק קבצים בעלי נתונים דמוגרפיים וחברתיים-כלכליים עם עקביות פנימית ברמת הרשומה הבודדת ועם השלמת מידע חסר כדי לספק קובץ ברמת איכות מרבית.

שני הסקרים של מפקד האוכלוסין 2008 בוצעו באמצעות שאלונים ממוחשבים, שבהם התקדמות הראיון הותוותה לפי תזרים שהוגדר מראש. סתירות לוגיות בין תשובותיו של המראיין, כמו גם טעויות שעשו הסוקרים בעת קלידת התשובות, הוצגו על-גבי מסך המחשב במהלך הראיון, כדי לאפשר לסוקרים לחזור לשאלות הרלוונטיות ולתקן טעויות במידת הצורך. בדיקה אוטומטית של עקביות והתאמה בין הנתונים, המובנית בראיון, צמצמה במידה ניכרת את היקף הערכים הבעייתיים שהגיעו להשלמות באמצעות עריכה וזקיפה. יתירה מכך, חלק מהמשתנים הוגדרו כ"משתנים קריטיים"; מערכת של בדיקות אוטומטיות שהופעלה על השאלות הרלוונטיות בשאלון הממוחשב מנעה את סיום הראיון בלי שהתקבל מענה על שאלות אלו. יחד עם זאת בדיקות מרובות במהלך הראיון קוטעות את רצף הראיון ולכן לא כל הבדיקות האפשריות נערכו במהלך הראיון ולא כל התשובות הוגדרו כקריטיות להמשך הראיון וכתוצאה מכך היה צורך בעריכה וזקיפה של תשובות שאינן עקביות והשלמת נתונים חסרים.

<sup>8</sup> יישוב מוסדי הוא מוסד בעל תכונות של יישוב, שאינו כלול בתחום המוניציפלי של יישוב אחר.  
<sup>9</sup> מושב הוא יישוב כפרי המאורגן כאגודה שיתופית ויש לו זכות לנחלות חקלאיות (לפי משמעות המונח במינהל מקרקעי ישראל). זהו יישוב המורכב מיחידות משפחתיות, אשר כל אחת מהן היא ישות כלכלית עצמאית. חלק מהייצור ומהניהול הכלכלי הוא בידי האגודה השיתופית, ורמת השיתוף נקבעת בידי התושבים.

המשתנים הדמוגרפיים הבאים מתוך המרשם הוגדרו כמשתנים "קריטיים" ועבורם בוצעו הזקיפות במקרים שהיו בהם ערכים חסרים (item non-response) או ערכים קיימים, אך שגויים (סתירות לוגיות, אי התאמה לערכים חוקיים או לנתונים אחרים):

- שנת לידה
- מין
- שנת עלייה/כניסה לארץ
- ארץ לידה
- ארץ לידת אב/אם
- מצב משפחתי

פעולת הזקיפה בוצעה לפי השלבים הבאים: תחילה הופעלו שיטות "זודאיות", עם מיצוין הופעלו זקיפות סטטיסטיות.

שלב 1 – זקיפות דטרמיניסטיות חזקות. זקיפות דטרמיניסטיות מתבצעות לפי תנאי לוגי שמופעל על נתון או על קבוצה של נתונים שחייבים להתקיים כדי שהנתון ייחשב כנכון. לדוגמה: זקיפת הנתון החסר "שנת עלייה" מתבצעת לפי הערך של שנת כניסה ראשונה לארץ.

שלב 2 – זקיפות לפי שיטת COLD DECK. בשיטת ה-Cold-Deck זקיפת הערך החסר מבוצעת באמצעות שימוש בערך "תורם" של אותו פרט מתוך מקור נתונים חיצוני. המקורות החיצוניים שבהם נעשה השימוש הם נתוני השטח של סקר האוכלוסין 2008 ונתוני המפקדים הקודמים (1972, 1983, 1995). לדוגמה: הערך החסר של המשתנה "ארץ לידה" ייזקף מתוך הערך התקין של אותו פרט מתוך נתוני השטח של סקר האוכלוסין 2008 או מתוך הערך התקין של נתוני המפקדים הקודמים.

שלב 3 – זקיפות דטרמיניסטיות חלשות. זקיפות דטרמיניסטיות חלשות מתבצעות באמצעות הפעלת תנאי לוגי שהוגדר. לדוגמה: זקיפה למשתנה "מצב משפחתי" מתבצעת לפי תנאי לוגי "אם גיל צעיר מ-15 אזי מצבו המשפחתי = רווק".

שלב 4 – זקיפות סטטיסטיות. זקיפות סטטיסטיות מתבצעות באמצעות חישוב ערך לזקיפת הערך החסר. לדוגמה: שימוש בערך ממוצע – המשתנה "שנת עלייה" של ילדים שנולדו בחו"ל מבוצעת על-פי גיל עלייה ממוצע מהארץ שבה נולדו.

שלב 5 – זקיפות סטטיסטיות בשיטת HOT DECK, "השכן הקרוב". זקיפה בשיטת ה-Hot Deck היא זקיפת הערך מ"תורם". בתור "תורם" נבחר משק בית עם דמיון מרבי למשק בית "נתרם". גילאים, ארצות מוצא, שנת עלייה של חברי משק הבית והיחסים ביניהם, יוצרים דפוסים משפחתיים שאפשר ליישםם בזקיפת נתונים חסרים עבור רשומת פרט של משק בית ספציפי.

מבין כל הרשומות במשק הבית התורם נבחרת הרשומה התורמת שדומה ביותר בנתונים הדמוגרפיים והדמיון נמדד על-ידי פונקציה מרחק<sup>10</sup>.

להלן לוח המתאר את התפלגות הזקיפות שנעשו למשתנים דמוגרפיים (כל הנתונים חושבו מתוך אוכלוסיית השייכים למפקד 2008):

משתנה	זקיפה לפי שיטה דטרמיניסטית (חזקה) (%)	זקיפה לפי נתוני שטח של סקר האוכלוסין 2008 (%)	זקיפה לפי נתונים ממפקדים קודמים (%)	זקיפה לפי שיטה דטרמיניסטית חלשה (%)	זקיפה לפי שיטה סטטיסטית (%)	זקיפה לפי שיטת Hot Deck (%)
שנת לידה	0.0	0.0	0.0	0.0	0.0	0.0
מין	0.0	0.0	0.0	0.0	0.0	0.0
מצב משפחתי	0.0	0.0	0.0	0.0	0.0	0.1
שנת עלייה <sup>11</sup>	8.8	0.3	0.8	0.1	0.8	0.3
ארץ לידה	0.0	0.0	0.0	0.0	0.0	0.0
ארץ לידת אב <sup>12</sup>	0.0	1.3	7.5	0.2	0.0	0.8
ארץ לידת אם <sup>12</sup>	0.0	1.3	7.7	0.0	0.0	0.7

#### 4.2.2 צריכה ולקיפה של קופת SEF

כדי לשמור על אחידות בנתונים, הועתקו לקובץ החברתי-כלכלי המשתנים הדמוגרפיים שנזקפו בקובץ הדמוגרפי: מין, שנת לידה, ארץ לידה, שנת עליה/כניסה לארץ, ארץ לידת אב/אם, מצב משפחתי.

המשתנים החברתיים-כלכליים הבאים שכאמור התקבלו בראיון ממוחשב במשק הבית הוגדרו כמשתנים "קריטיים" ועבורם בוצעו הזקיפות במקרים שהיו בהם ערכים חסרים (non-item response):

##### משתני השכלה:

- מצב לימודים (לומד כעת/למד בעבר/לא למד כלל)
- מצב לימודים (למד/לא למד) לפי סוג מוסד לימודים
- לימודים (למד/לא למד) בישיבה גבוהה
- מספר שנות לימוד
- מספר שנות לימוד בישיבה גבוהה

<sup>10</sup> פונקציה המרחק משקפת את מידת הקרבה המהותית בין ערכים שונים של משתנים. הרשומה הזקופה היא קומבינציה של רשומה נכשלת ורשומה תורמת, כך שהשינוי בנתונים הוא מינימלי ואין סתירות לוגיות.

<sup>11</sup> האחוזים מחושבים מתוך האוכלוסייה של ילידי חו"ל.

<sup>12</sup> האחוזים מחושבים מתוך האוכלוסייה של יהודים ואחרים, ילידי ישראל.



• התעודה הגבוהה

משתני עבודה:

- שייכות לכוח העבודה השנתי
- שייכות לכוח העבודה השבועי

תנאי דיור:

- מספר חדרים

**תהליך הזקיפה:**

תהליך הזקיפה של משתנים חברתיים-כלכליים מבוצע במתודולוגיה NIM, שמבוססת על מציאת "השכן הקרוב". הזקיפות מבוצעות בצורה אוטומטית באמצעות תוכנת ה-CANCEIS<sup>13</sup> שבה מוזנים הגדרות המשתנים, אילוצים לוגיים ופרמטרים אחרים. בתהליך הזקיפה, התוכנה מחפשת את "השכן הקרוב", שהוא משק בית עם דמיון מרבי למשק בית "נתרם"; מבין כל משקי הבית נבחר משק בית, הדומה ביותר הן במשתנים המסבירים את המשתנה הנזקף: דת, גיל, מין והן במשתנים החברתיים-כלכליים של חברי משק בית.

עבור המשתנה "מספר חדרים" הוגדר "השכן הקרוב" כרשומת משק בית הדומה במיקום גיאוגרפי ובמספר נפשות במשק בית.

המשתנה "מספר שנות לימוד בישיבה גבוהה" הוא משתנה מיוחד, שבניגוד למפקדים הקודמים, לא נאסף במסגרת השאלון. לצורך חישוב האומדן של "מספר שנות לימוד בישיבה גבוהה" פותח אלגוריתם שהופעל על רשומות שציינו כי למדו בישיבה גבוהה. האלגוריתם חישב "מספר שנות לימוד תקיני" – מספר שנות לימוד מינימלי הנדרש לקבלת תעודה במוסדות לימוד שונים. "סה"כ שנות לימוד" מנוכה מ"מספר שנות לימוד תקיני" (בכל מוסדות הלימוד, שאינם הישיבה הגבוהה) הוא האומדן ל"מספר שנות לימוד בישיבה גבוהה".

להלן לוח המתאר את התפלגות הזקיפות שנעשו למשתנים חברתיים-כלכליים (הנתונים חושבו מתוך 768,821 רשומות של בני 15 ומעלה, למעט המשתנה "מספר חדרים", שבו חושב הנתון מתוך 327,587 רשומות של משקי בית):

משתנה	% זקיפה
מצב לימודים (לומד כעת / למד בעבר / לא למד כלל)	1.06
למד / לא למד בבי"ס יסודי	1.14
למד / לא למד בבי"ס תיכון	1.16
למד / לא למד בבי"ס על-תיכוני	1.15
למד / לא למד במוסד אקדמי	1.15
למד / לא למד במוסד אחר	1.14
למד/ לא למד בישיבה גבוהה	1.14
מספר שנות לימוד	2.57
מספר שנות לימוד בישיבה גבוהה	3.30
התעודה הגבוהה	1.30
שייכות לכוח עבודה שנתי, שייכות לכוח עבודה שבועי	2.41
מספר חדרים	0.53

### 4.3 הוספת משתנים חדשים

לקובץ החברתי-כלכלי נוספו משתנים וכן סמלים למידע מילולי, בהתאם לצרכים שונים בעיבוד הנתונים:

- משתנים שעברו סימול (כתובות, ענף כלכלי ומשלח יד);
- משתנים מחושבים (מספר נפשות במשק בית);
- משתנים ממקורות מידע מנהליים (קבצי מס הכנסה, ביטוח לאומי, משרד הרווחה);

#### 4.3.1 משתנים שאגרו סימול

המידע שנאסף על כתובת מקום העבודה וכתובת לפני 5 שנים, על ענף כלכלי ועל משלח יד נכתב כמלל חופשי בהתאם לתשובותיהם של המרואיינים. במקרים של כתובת מקום העבודה וכתובת לפני 5 שנים בוצעו תהליך של סימול היישוב והרחוב ועיגון (ידיני ואוטומטי, בדומה למתואר בסעיף 1.3) לרמת היישוב ולרמת האזור הסטטיסטי. תהליך הסימול של הענף הכלכלי ומשלח היד בוצע על כ-435,000 רשומות והוא נערך בשני שלבים:

- בשלב הראשון, בוצע סימול אוטומטי באמצעות המערכת הממוחשבת שנבנתה במיוחד לצורך זה. בשלב זה סומלו 77% מהרשומות לענף כלכלי ו-72% מהרשומות למשלח יד.
- בשלב השני, בוצע סימול ידיני של רשומות שלא צלחו את הסימול האוטומטי. בשלב זה טופלו כ-180,000 רשומות.

הרשומות שנותרו בלתי מסומלות, קיבלו ערך XXX (ענף כלכלי/משלח יד לא ידוע).

### 4.3.2 משתנים מחושבים

משתנים מחושבים הם משתנים שלא נאספו בשאלון מפקד 2008, אך ניתן ליצור אותם באמצעות חישוב הנתונים של שניים או יותר מהמשתנים הקיימים בקובץ החברתי-כלכלי. משתנים מחושבים נוצרו כדי להקל את עיבוד הנתונים, כמו גם כדי לאפשר בסיס השוואתי עם נתוני מפקד 1995 כשמדובר על משתנים דומים. דוגמאות למשתנים מחושבים: משתנים המשווים את מקום המגורים לפני 5 שנים למקום המגורים במועד הקובע של המפקד (ברמות: מחוז, נפה, אזור טבעי ומטרופולין), משתנים שמסכמים את הכנסות הפרט ומשק הבית, משתנה טיפולוגיה של משק בית, מספר נפשות במשק בית וכדומה.

### 4.3.3 משתנים מקבצי מינהליים

כדי להקל את נטל ההשבה של האוכלוסייה, הוחלט שלא לאסוף באמצעות השאלון את המידע שנאסף בצורה שוטפת על-ידי גופים ממשלתיים לצרכי מינהל שונים. להלן שלושת המקורות העיקריים שמהם התקבל מידע מינהלי איכותי:

- מקבצי רשות המיסים התקבל מידע על הכנסות הפרטים מעבודה או עסק. מנתונים אלה חושבו משתנים חדשים (כמו: הכנסה ממוצעת חודשית) שנוספו אל הקובץ החברתי-כלכלי.
- מקבצי ביטוח לאומי התקבלו נתונים של הכנסות הפרטים מקצבאות של ביטוח לאומי (אבטלה, הבטחת הכנסה, מזונות, נכות, שארים, ילדים, זקנה, אחר). מנתונים אלה חושבו משתנים חדשים שנוספו לקובץ החברתי-כלכלי: סה"כ הכנסות מקצבאות – לפרט ולמשק בית (שנתי וממוצע חודשי) וחודשים של קבלת הקצבאות.
- קובץ עיוורים של משרד הרווחה: בשאלון נבדק הקושי בביצוע ארבע פעילויות יומיומיות (לשמוע, ללכת בתוך הבית ולעלות ולרדת במדרגות, לזכור או להתרכז, להתלבש או להתרחץ). מידע נוסף, הקושי בראייה, התקבל מתוך הקובץ ממשרד הרווחה ונוסף אל הקובץ החברתי-כלכלי.

### 5. פרסום נתוני מפקד האוכלוסין 2008

הקובץ הדמוגרפי והקובץ החברתי-כלכלי הם שני מקורות המשמשים להפקת פרסומי המפקד. עם זאת, הקובץ החברתי-כלכלי הוא המקור המפקדי היחיד המכיל את המידע החברתי-כלכלי של כ-14% מאוכלוסיית המדינה. קובץ זה מהווה בסיס בהפקת הקבצים: ה-PUF (Public Use File), קובץ שמיועד לשימוש הציבור הרחב וכן ה-MUC (Micro Data Under Contract), המשמש לצרכי החוקרים ולביצוע עיבודים מיוחדים בלמ"ס.