

נספח מס' 1

מתודולוגיית הסקר

א. שיטות הסקר

1. אוכלוסיית הסקר

הסקר הוא סקר דירות שבהן מתגוררים בני 60 ומעלה. לצורכי הסקר נקבע כי מי שנולד בשנת 1937 לכל המאוחר, הוא בן 60 ומעלה.

אוכלוסיית הסקר כללה את כל בני ה-60 ומעלה אשר בתקופת הסקר:

- ← גרו באופן קבוע ביישובים עירוניים, בכפרים, ביישובים קהילתיים וב"מקומות".
 - ← היו תושבים ותיקים, או עולים חדשים שעלו עד סוף מרץ 1997, או תיירים ותושבים ארעיים המתגוררים בארץ מעל שנה ברציפות, או תושבי קבע ששהו בחו"ל לכל היותר שנה ברציפות.
 - ← התגוררו בקביעות במשקי בית "רגילים", ובכלל זה אנשי צוות במוסדות ובמשקי בית במוסדות דיור מוגן ובמרכזי קליטה, או היו מאושפזים בבתי חולים (רגילים), לתקופה של 6 חודשים לכל היותר, במועד עריכת הסקר (עבורם נקבע בסקר כי מקום המגורים הוא כתובתם האזרחית).
- אוכלוסיית הסקר לא כללה את בני ה-59 ומטה וכן את בני ה-60 ומעלה השייכים לאחת מהקבוצות הבאות לפחות:

- ← תושבים קבועים הגרים במושבים, במושבים שיתופיים, בקיבוצים, גרים מחוץ ליישובים ובכלל זה שבטי הבדווים.
- ← תיירים ותושבים ארעיים המתגוררים בארץ פחות משנה.
- ← עולים חדשים שעלו לאחר מרץ 1997.
- ← שוהים בחו"ל מעל שנה ברציפות במועד עריכת הסקר.
- ← דיירים במוסדות שאינם דיור "מאורגן", כמו בתי אבות ומוסדות לאשפוז ממושך.
- ← מאושפזים לתקופות ארוכות מ-6 חודשים בעת עריכת הסקר.

2. המסגרת לדגימה

המסגרת העיקרית ששימשה להוצאת המדגם בסקר, התבססה על קובץ שהוכן על פי נתוני חלק א' (הפקידה המלאה) בשאלון מפקד האוכלוסין והדיור 1995.

הקובץ הבסיסי כלל את כל הנפשות בדירות שנפקדו במפקד 1995 והופק לאחר ביצוע הפעולות הבאות:

- ← **זיווג עם המרשם** – להוספת התכונות המרשמיות של הנפשות כמקור נוסף לקביעת התכונות הסופיות.
- ← **הוספת שאלוני נעדרים** – שאלוני מפקד שהתקבלו בדואר.
- ← **ניכוי רשומות כפולות** – לרשומות של נפשות שהופיעו במפקד במספר כתובות, נוכח הרשומות העודפות.

לאחר הכנת הקובץ הבסיסי של המפקד נעשו פעולות נוספות במטרה להתאימו, ככל האפשר, להגדרת אוכלוסיית הסקר ולצורכי תכנון המדגם:

◀ ניכוי לא שייכים – נפשות שנפטרו לאחר מועד המפקד ועד להכנת המסגרת הסופית, נפשות בדירות ביישובים ובמוסדות שאינם שייכים להגדרת הסקר, נפשות בדירות עם "צעירים" בלבד (ילידי 1938 ואילך).

◀ קביעת תכונות בסיסיות לתכנון המדגם – לנפשות ולדירות עם אי-התאמה בין המקורות (המפקד והמרשם), או עם תכונות בסיסיות חסרות.

בסופן של פעולות ההכנה, כלל קובץ המפקד לצורכי הסקר 674,662 בני 60 ומעלה ב-476,995 דירות.

בהגדרת הסקר נכללים, בין היתר, עולים חדשים שעלו לאחר מועד המפקד ועד לסוף מרץ 1997. עולים אלה אינם מכוסים במסגרת הדגימה העיקרית המבוססת על נתוני מפקד 1995.

על מנת להקטין את אי-הכיסוי של אוכלוסיית הסקר, הוכן קובץ משלים של עולים חדשים שעלו לאחר מועד עריכת מפקד 1995. מהקובץ נוכו נפשות בדירות עם ילידי 1938 ואילך ונפשות בדירות ביישובים שאינם כלולים בהגדרת הסקר.

בסך הכל, כלל קובץ העולים החדשים שעלו לאחר המפקד ועד לסוף מרץ 1997 18,783 בני 60 ומעלה ב-13,983 דירות.

מסגרת הדגימה של הסקר כללה את קובץ המפקד ואת קובץ העולים לאחר המפקד, ובה 693,445 בני 60 ומעלה ב-490,978 דירות.

3. אי התאמה בין אוכלוסיית הסקר למסגרת הדגימה

למרות הפעולות שנעשו על מנת להתאים את מסגרת הדגימה להגדרת אוכלוסיית הסקר, מסגרת הדגימה של הסקר הייתה חשופה לכיסוי עודף של מקרים מיותרים ולכיסוי חלקי של אוכלוסיית הסקר.

א. מקרים מיותרים במסגרת הדגימה

מסגרת הדגימה כללה מקרים שאינם שייכים לאוכלוסיית הסקר, מהסיבות הבאות:

1) שינויים שחלו באוכלוסייה מזמן עדכון מסגרת המפקד ועד מועד עריכת הסקר

- פטירות של בני 60 ומעלה שאירעו מאז המפקד ועד לתקופת הסקר – זאת למרות שרוב הנפטרים עד סוף פברואר 1997 נוכו מהמסגרת לפני שלב הדגימה. חלק מהנפטרים שנדגמו לסקר אותרו בקובצי הפטירות במהלך הסקר ולפני היציאה לשדה (שנחסכה עבורם). אך עבור אחרים התברר כי הם נפטרו רק בשלב עבודת השדה, או לאחריו, בקובצי פטירות מעודכנים למועד עריכת הסקר.

- מעבר של בני 60 ומעלה ממגורים במשקי בית "רגילים" למגורי קבע כדיירים במוסדות.

- מעבר של בני 60 ומעלה ממגורים ביישובים השייכים להגדרת הסקר למגורים ביישובים שאינם כלולים בהגדרת הסקר – אם כי ניתן להעריך שקבוצה זו קטנה יחסית.

- מעבר של בני 60 ומעלה לחו"ל לאחר המפקד ושהות של שנה לפחות ברציפות בחו"ל במועד עריכת הסקר – אם כי ניתן להעריך שקבוצה זו קטנה יחסית.

2) טעויות במסגרת הדגימה

למרות שנעשו מאמצים לאתר טעויות במסגרת הדגימה ולתקן, המסגרת הייתה חשופה לטעויות:

- **טעויות תשובה במפקד או טעויות ברישום שנת הלידה במרשם, או זיווג שגוי בין רשומות המפקד לבין מרשם התושבים** – הביאה לכך שבמסגרת נכללו נפשות עם רישום לשנת לידה 1937 או קודם לכן (שייכים לאוכלוסיית הסקר), אך למעשה הם היו ילידי 1938 ואילך, אשר אינם שייכים לאוכלוסיית הסקר.
- **נפשות שנפקדו בטעות יותר מפעם אחת במפקד או כפילות בין מסגרת המפקד למסגרת העולים לאחר המפקד.**
יש להדגיש, כי הכיסוי העודף, אינו גורר הטיה של אומדני הסקר, אך הוא מגדיל את עלות הסקר, אם לא ניתן לנכותו מראש מהמדגם. בנוסף, הכיסוי מגדיל את טעות הדגימה של האומדנים, בעיקר עבור אומדני שכיחויות.

ב. אי-כיסוי של המסגרת לדגימה

אי-כיסוי של המסגרת כתוצאה ממספר גורמים עיקריים:

1) שינויים שחלו באוכלוסייה מזמן עדכון מסגרת הדגימה ועד למועד עריכת הסקר

- מעבר של בני 60 ומעלה, שהיו דיירים במוסדות בזמן המפקד, למגורים במשקי בית "רגילים" בתקופת הסקר - ניתן להעריך שאי-כיסוי זה אינו גדול.
- מעבר של בני 60 ומעלה ממגורים ביישובים שאינם כלולים בהגדרת הסקר במועד עדכון המסגרת, למגורים ביישובים השייכים להגדרת הסקר בזמן עריכת הסקר.
- בני 60 ומעלה ששהו שנה לפחות ברציפות בחו"ל בעת המפקד וחזרו ארצה לפני עריכת הסקר - ניתן להעריך שאי-כיסוי זה אינו גדול.

2) טעויות וחוסרים של המסגרת לדגימה

- **נפשות שלא נכללו בסקר בגלל אי-כיסוי בסיסי במפקד או במסגרת המשלימה של עולים לאחר המפקד** – אי-כיסוי זה עשוי להיות חמור במיוחד עבור אוכלוסיית בני 60 ומעלה (בהשוואה לקבוצות גיל אחרות באוכלוסייה). לא נעשתה פעולה מיוחדת להערכת היקף אי-כיסוי זה ויש להניח כי הוא אינו זניח.
- **נפשות שנקבע בטעות כי הן מתחת לגיל 60 ("צעירים") בעוד שלמעשה הן היו בני 60 ומעלה** – זאת עקב טעות ברישום שנת הלידה במפקד ו/או במרשם, או שהיו ללא שנת לידה במפקד ולא הייתה אפשרות לערוך זיווג עם המרשם לצורך קבלת שנת הלידה המרשמית.
אי-כיסוי של המסגרת עלול לגרום להטיות של אומדני הסקר. עבור אומדני שכיחויות, אי-הכיסוי יגרום להטיה כלפי מטה של האומדנים. אם תכונותיה של האוכלוסייה שלא מכוסה במסגרת לדגימה שונות משאר האוכלוסייה, נקבל הטיית גם של אומדני הפרופורציה.
יש להעיר כי בשלב האמידה נעשו פעולות התאמה של התפלגות המדגם עם התפלגות אומדנים דמוגרפיים שוטפים של אוכלוסיית הסקר, לפי תכונות, במטרה לצמצם הטיית של אומדני הסקר, שחלקן נבע מאי-כיסוי של המסגרת לדגימה.

4. שיטת הדגימה

מהמסגרת שהוכנה הוצא מדגם של דירות שבהן מתגוררים בני 60 ומעלה, על-פי הרכב הנפשות בדירה, כפי שהיה במסגרת לדגימה. בכל דירה, כל בני ה-60 ומעלה היו מיועדים להיחקר.

כל דירה במסגרת אופיינה לפי תכונות בסיסיות (קבוצות ראשיות) ששימשו לתכנון המדגם:

קבוצה ראשית	תיאור הקבוצה	הערות
קבוצת אוכלוסייה לדירה	- ערבים - עולים חדשים - אחרים	-לפי הדת הרווחת ביישוב ודת הנפשות בדירה -עולי 1990 ואילך -רובם יהודים
"סוג" הדירה	- "קשיש" (בן 60 ומעלה) בודד - "קשיש" אחד + "צעיר/ים" - שני "קשישים" שוני מין ללא "צעיר/ים" - שני "קשישים" שוני מין עם "צעיר/ים" - אחרים	-ללא תלות במצב משפחתי -ללא תלות במצב משפחתי -כולל 3 "קשישים" ומעלה בדירה
הרכב הנפשות בדירה לפי קבוצות גיל ומין	- גברים בני 60-64 - גברים בני 65-69 - גברים בני 70-74 - גברים בני 75-79 - גברים בני 80 ומעלה - נשים בנות 60-64 - נשים בנות 65-69 - נשים בנות 70-74 - נשים בנות 75-79 - נשים בנות 80 ומעלה	

המדגם תוכנן במטרה לראיין כ-5,000 בני 60 ומעלה (מדגם נטו). הקצאת המדגם נטו בין הקבוצות הראשיות לא הייתה יחסית לגודל הקבוצות, אלא דיפרנציאלית, על מנת להבטיח ייצוג במדגם לתת-קבוצות בהתאם למטרות הסקר. כך למשל, הסקר תוכנן לתגבר את בני ה-80 ומעלה וכן את אוכלוסיות הערבים והעולים החדשים.

כדי לקבל את גודל המדגם ברוטו, הוערכו שיעורי אי-הפקידה הצפויים בסקר עבור כל קבוצת אוכלוסייה בנפרד ולכל קבוצת גיל ומין. הערכות אלו התבססו על שיעורי אי-הפקידה בסקר בני 60 ומעלה הקודם שנערך בשנת 1985, שיעורי תמותה מעודכנים והערכות נוספות. לפי הערכות אלו התקבל כי גודל המדגם ברוטו צפוי להיות כ-6,000 בני 60 ומעלה.

גודל המדגם ברוטו, המתוכנן, לא אפשר הקצאת המדגם לאוסף כל הקבוצות הראשיות במצולב (קבוצת אוכלוסייה לדירה, סוג הדירה והרכב הנפשות בדירה לפי קבוצת גיל ומין). כחלופה נבנה תהליך

איטראטיבי להקצאת המדגם, המבטיח התפלגויות נתונות של נפשות בקבוצות גיל ומין לפי קבוצת אוכלוסייה ולפי סוג דירה בנפרד. זאת תחת האילוץ כי בכל דירה מיועדים לחקירה כל בני 60 ומעלה.

ההסתברות של דירה להדגם היא היחס בין גודל המדגם ברוטו, שנקבע בתהליך האיטראטיבי לדירות עם התכונות הבסיסיות של הדירה, לבין מספר הדירות עם אותן תכונות בסיסיות במסגרת הדגימה. כל הנפשות השייכות לאוכלוסיית הסקר בדירה שנדגמה מיועדות לפקידה, ולכן ההסתברות של נפש להיכלל במדגם שווה להסתברות הדגימה של הדירה שאליה היא שייכת.

הדירות סווגו לשלוש שכבות ראשיות לפי קבוצת אוכלוסייה. בתוך כל שכבה סודרו הדירות לפי סוג הדירה והרכב הנפשות לפי קבוצות גיל ומין בדירה ולפי מאפיינים נוספים של הדירה. הוצא מדגם מקרי-שיטתי של דירות יחסית להסתברות הדגימה של הדירות.

בסך הכל נדגמו 4,275 דירות שבהן 5,967 בני 60 ומעלה.

5. עבודת השדה

ראה "איסוף הנתונים ותוצאות הפקידה" במבוא.

6. זקיפות והשלמות

במטרה להקטין הטיות באומדנים, העלולות להיגרם בגלל התפלגויות שונות בין המשיבים והלא-משיבים השייכים לאוכלוסיית הסקר ("נפל"), בקטגוריה של נפשות עם תכונות מיוחדות, נערכו זקיפות אינדיבידואליות לחלק מהלא-משיבים מהסיבות הבאות:

א. קשיי תקשורת ("סניליים")

בסקר התקבלו 31 לא-משיבים עקב קשיי תקשורת. עבורם נערך חיפוש של תורמים, מבין המשיבים לסקר בעלי אותן התכונות הבסיסיות לתכנון המדגם והם רואיינו באמצעות שאלון PROXY.

סך הכל נזקפו 24 לא-משיבים עם קשיי תקשורת.

ב. מאושפזים לתקופה קצרה בבתי חולים רגילים

בסקר התקבלו 12 לא-משיבים עקב אשפוז לתקופה קצרה בבתי חולים. עבורם נערך חיפוש של תורמים, מבין המשיבים לסקר, בעלי אותן התכונות הבסיסיות לתכנון המדגם, אשר עפ"י הקבצים המנהליים נפטרו תקופה קצרה לאחר מועד הריאיון בסקר.

סך הכל נזקפו 10 לא-משיבים עקב אשפוז בבתי חולים.

בנוסף, נערכו זקיפות אינדיבידואליות ללא-משיבים מסיבות אחרות (סרבנים, נעדרים וכו'), אשר גרים בדירות עם נפשות אחרות שהשיבו לסקר ("נפל חלקי"), וזאת כדי ליצור דירות עם השבה מלאה. בחירת התורמים נעשתה לפי דמיון בתכונות הבסיסיות לתכנון המדגם. בסך הכל נזקפו נתונים ל-48 לא-משיבים בדירות עם "נפל חלקי".

הרשומה שנוצרה בעקבות הזקיפות האינדיבידואליות הייתה שילוב בין המידע המוקדם אשר היה ידוע לגבי הלא-משיבים (בעיקר ממסגרת הדגימה) לבין נתוני התורמים (בעיקר משאלון הסקר).

7. קביעת תוצאות הפקידה

לאחר סיום עבודת השדה, זווג המדגם לפי מספר הזהות עם קובצי פטירות, ביקורת הגבולות ונתונים חלקיים של מפקד מוסדות 1997. הזיווג איפשר לקבוע באופן מדויק, עבור חלק מהלא-משיבים, (ובעיקר, עבור הלא-מאותרים, נעדרים מסיבה לא ידועה, שוהים בחו"ל תקופה לא ידועה וכו'), אם הם שייכים לאוכלוסיית הסקר (מקרי "נפל"), או אינם שייכים לאוכלוסיית הסקר (מקרי "אפס").

גם לאחר הזיווג עם הקבצים המנהליים, נותרו מקרים שעבורם לא ניתן לקבוע אם הם שייכים או לא שייכים לאוכלוסיית הסקר. זאת מאחר שמבין הלא-שייכים, הנפטרים בלבד מכוסים באופן כמעט מלא בקבצים המנהליים (על ידי קובצי הנפטרים) ואילו הקבוצות האחרות (השוהים במוסדות ובחו"ל שנה ומעלה) מכוסות באופן חלקי בלבד.

על סמך המידע שהתקבל בשדה ובקבצים המנהליים, נקבעה תוצאת הפקידה הסופית לנפש ולדירה. להלן מובאות תוצאות הפקידה הסופיות:

בסקר נדגמו 4,275 דירות.

ב-236 דירות (5.5% מכלל הדירות במדגם) כל הנדגמים לא היו שייכים לאוכלוסיית הסקר (מקרי "אפס"). בלמעלה ממחצית מהמקרים התברר כי כל הנדגמים בדירה נפטרו ובכשליש מהדירות עברו כל הנדגמים למוסדות שאינם כלולים בהגדרת אוכלוסיית הסקר.

עבור 124 דירות (2.9%) לא אותרה הכתובת, או שדייריה לא אותרו בכתובת, או שעברו לכתובת אחרת שלא נפקדה. יש להעיר כי ברשות הפוקד היו הכתובת במפקד וכן הכתובת אשר הייתה רשומה במרשם התושבים, סמוך למועד היציאה לשדה. יש להניח כי חלק מהדיירים לא היו שייכים לאוכלוסיית הסקר, למרות המאמצים שנעשו לאתר אותם בקבצים המנהליים השונים.

ב-362 דירות (8.5%) לא כל הדיירים השייכים לאוכלוסיית הסקר השיבו לראיון. הסיבה העיקרית הייתה סירוב (כ-60% מכלל הדירות של הלא-משיבים).

3,553 דירות (83.1%) נכללו בעיבודי הסקר.

בדירות שנדגמו לסקר היו על פי מסגרות הדגימה 5,967 בני 60 ומעלה. בשדה התברר כי 105 נפשות השייכות לאוכלוסיית הסקר נוספו לדירות שנדגמו מאז מועד העידכון של המסגרת (לכמחציתם התברר בדיעבד, כי נכללו במסגרת לדגימה בדירות אחרות).

סך הכל, המדגם כלל 6,072 נפשות.

389 נפשות (6.4% מכלל הנפשות במדגם) לא היו שייכות לאוכלוסיית הסקר (מקרי "אפס"). מהן קרוב לשני שלישי עקב פטירה, ומעל רבע עברו למוסד שאינו נכלל בהגדרת אוכלוסיית הסקר.

144 נפשות (2.4%) לא אותרו, למרות המאמצים שנעשו בשדה וסביר כי חלקם לא שייכים לאוכלוסיית הסקר, למרות שלא נמצאו בקבצים המנהליים.

484 נפשות (8.0%) לא השיבו לסקר (ולא נזקפו אינדיבידואלית), מהם כשני שלישי סירבו להשתתף בסקר.

5,055 נפשות (83.2%) השתתפו בעיבודי הסקר.

8. שיטת האמידה

שיטת האמידה מיועדת לצמצם הן את טעויות הדגימה והן את ההטיות העלולות לנבוע מכך, שדירות שדייריהן לא השיבו יכולות להיות שונות בתכונותיהן מאלו שהשיבו לסקר.

הדירות שעבורן לא ידוע אם הן שייכות לאוכלוסיית הסקר (לא מאותרים וכו'), גם לאחר ניסיון הזיווג לקבצים המנהליים, הוקצו בין השייכים ("נפל") והלא-שייכים להגדרת אוכלוסיית הסקר. ההקצאה נעשתה באופן פרופורציונלי למספר הלא שייכים מהשדה שלא הופיעו בקבצים המנהליים (חלק מהשוהים במוסדות ובחו"ל מעל שנה) והשייכים הידועים במדגם.

בנוסף, המסגרת לדגימה הייתה חשופה לאי-כיסוי או לטעויות (ר' סעיף 3 לעיל), אשר יתכן שהן דיפרנציאליות בין הקבוצות השונות.

כדי לקבל אומדנים המתייחסים לכלל אוכלוסיית הסקר (או תת-קבוצות מתוכה), נקבע מקדם ניפוח לכל דירה שנחקרה, כאשר לכל הנפשות השייכות לאותה דירה יש אותו מקדם ניפוח. מקדם הניפוח של הדירה מבטא את מספר הדירות ואת מספר הנפשות באוכלוסיית הסקר המיוצגת על ידי הדירה.

תהליך קביעת מקדמי הניפוח כלל את השלבים הבאים:

◀ בסקר זה הדירות השונות נדגמו בהסתברויות שונות. לכן, נקבע בשלב ראשון מקדם ניפוח ראשוני לדירה (ולנפשות) לפי היפוך הסתברות הדגימה.

◀ בשלב שני, בתוך קבוצות הומוגניות לניפוח שהתבססו על אוסף התכונות הבסיסיות להקצאת המדגם, ניתנו גורמי תיקון לדירות שהשיבו לסקר ואשר השתתפו באומדנים, במטרה לפצות עבור הלא-משיבים (**לפי היחס בין המדגם ברוטו למדגם נטו**). על ידי כך זוקפים גלובלית בתוך קבוצת הניפוח את הלא משיבים על פי המשיבים.

מקדם הניפוח של הדירה, בשלב זה, הוא המכפלה של מקדם הניפוח הראשוני עם גורם התיקון שחושב כדי לפצות עבור הלא משיבים בקבוצת הניפוח.

◀ בשלב השלישי, חושבה מערכת מקדמי ניפוח סופיים אשר מתאימה את התפלגות האומדנים המנופחים של הנפשות, על פי מקדמי הניפוח מהשלב הקודם, להתפלגות האומדנים הדמוגרפיים השוטפים של הלמ"ס לסוף שנת 1997. האומדנים הדמוגרפיים לסוף שנת 1997 הם על בסיס מפקד 1995, לאחר הורדת האוכלוסיות שאינן שייכות לאוכלוסיית הסקר (דיירי מוסדות, האוכלוסייה במושבים ובקיבוצים וכו'). נעיר, כי האומדן הדמוגרפי לסך הכל בני 60 ומעלה היה גבוה ב-6.1% מהאומדן המנופח של השלב הקודם.

מערכת מקדמי הניפוח הסופיים חושבה בשיטה איטראטיבית (RAKING) דו-מימדית. בשיטה זו הותאמה התפלגות המדגם המנופח של הנפשות על פי מקדמי הניפוח מהשלב הקודם, למספר התפלגויות חיצוניות של נפשות, לפי משתני חלוקה נבחרים, תוך שמירה על מקדם ניפוח אחיד בדירה:

- קבוצת אוכלוסייה, חתכים גיאוגרפיים, מין וקבוצות גיל.
- חתכים גיאוגרפיים מפורטים, מין וקבוצת גיל (מעבר לכל קבוצות האוכלוסייה).

הערות:

- ◀ קבוצות גיל ומין אשר עבורם נערכה ההתאמה היו לפי הפירוט של הקבוצות הראשיות (ראה סעיף 4).
- ◀ עבור קבוצת האוכלוסייה הערבית החתכים הגיאוגרפיים היו: העיר ירושלים; היישובים המונים 10,000 תושבים ומעלה במחוזות הצפון וחיפה; היישובים המונים פחות מ-10,000 תושבים במחוזות הצפון וחיפה; שאר המחוזות.
- ◀ עבור קבוצת האוכלוסייה עולי 1990 ומעלה, החתכים הגיאוגרפיים היו: שלוש הערים הגדולות והיישובים ביהודה ושומרון הצמודים לירושלים; מחוז הצפון; מחוז חיפה; נפות השרון ופתח תקווה ויתר היישובים ביהודה ושומרון; נפות רמלה ורחובות; מחוז תל אביב; מחוז הדרום ונפת עזה.
- ◀ עבור קבוצת האוכלוסייה של כל שאר האוכלוסייה, החתכים הגיאוגרפיים היו: מחוז ירושלים והיישובים ביהודה ושומרון הצמודים לירושלים; העיר תל אביב; העיר חיפה; מחוז הצפון; מחוז חיפה; נפות השרון ופתח תקווה ויתר היישובים ביהודה ושומרון; נפות רמלה ורחובות; נפות תל אביב ורמת גן; נפת חולון; מחוז הדרום ונפת עזה.
- ◀ לצורך ההתאמה לכל קבוצת האוכלוסייה יחדיו, החתכים הגיאוגרפיים היו הפירוט הרחב ביותר המתקבל מהחתכים הגיאוגרפיים של שלוש קבוצות האוכלוסייה.

ב. מהימנות האומדנים

האומדנים המוצגים בפרסום זה מבוססים על סקר מדגמי, ועלולים להיות חשופים לטעויות משני סוגים עיקריים:

- (1) **טעויות דגימה:** טעויות אלה נובעות מכך שנחקר רק מדגם אחד של משקי בית, על כל הנפשות שבהם ולא נחקרו כל משקי הבית וכל הנפשות באוכלוסייה.
- (2) **טעויות לא מדגמיות:** טעויות אלה יכולות לנבוע מגורמים אחרים, שעשויים להתקיים גם כאשר נערך מפקד מלא של כל האוכלוסייה.

(1) טעויות דגימה

המדגם שעליו מבוסס סקר זה הוא אחד מתוך מספר רב מאוד של מדגמים אפשריים באותו גודל, שהיו יכולים להיבחר באותה שיטת דגימה מאותה אוכלוסייה.

האומדן X' הוא הערך הנאמד לפי המדגם המסוים של סקר זה בעבור הערך המקביל לו X , שהיה מתקבל אילו נערך מפקד מלא.

טעות דגימה של אומדן $\sigma(X')$ היא מידה ממוצעת של השוני בין כל האומדנים השונים שהיו יכולים להתקבל מכל המדגמים האפשריים באותו גודל ובאותה שיטה, לבין הערך שהיה מתקבל לו נערך מפקד מלא, באותם תנאים של איסוף הנתונים.

רווח-סמך לאומדן הוא רווח שמכיל את הערך המפקדי X ברמת ביטחון מסוימת נתונה מראש. האומדן X' על פי המדגם והאומדן לטעות הדגימה שלו - $\sigma'(X')$, מאפשרים לבנות רווח-סמך עם רמת ביטחון מסוימת, נתונה מראש, כך שהרווח מכיל את הערך המפקדי X בביטחון שנקבע.

מקובל להציג רווחי סמך ברמת ביטחון של 95%, לכן רווח הסמך ברמת ביטחון זו מתקבל על ידי $X' \pm 2\sigma'(X')$. לכל לוח בפרסום זה, מובא ליד האומדן באותיות נטויות הערך של שתי טעויות הדגימה לאותו אומדן.

דוגמא: עפ"י לוח סיכום לפרק 1, "לוח א: בני 60 ומעלה, לפי מבחר משתנים, קבוצת גיל ומין", האומדן לאחוז בני 60 ומעלה המתגוררים לבד הוא 22.5 אלף. רווח הסמך 95% לאחוז זה הוא 22.5 ± 0.5 אלף, כלומר, ניתן לומר בביטחון של 95% כי אחוז בני 60 ומעלה המתגוררים לבד הוא בין 22.0 לבין 23.0 אלף.

ניתן לקבוע רמת ביטחון גבוהה יותר או נמוכה יותר, ולחשב את רווח-הסמך כך:

α	67%	80%	90%	95%	99.5%
$K(\alpha)$	1.0	1.3	1.7	2.0	2.8

כאשר $K(\alpha)$ (שהנו מספר טעויות הדגימה לכל כיוון) נקבע בהתאם לרמת הביטחון הדרושה α .

בהמשך לדוגמא הקודמת: אם מעוניינים בביטחון גבוה יותר, של 99.5% (כמעט בוודאות), ניתן לחלק

ב-2 את הערך של שתי טעויות דגימה (כדי לקבל את הערך של טעות דגימה אחת),

ולהכפיל את התוצאה ב- $K(\alpha) = 2.8$. בדוגמא זו מתקבלת טעות דגימה אחת של 0.25

אלף, ולכן רווח-סמך 99.5% יהיה:

$22.5 \pm 2.8 \times 0.25$ אלף, כלומר: 22.5 ± 0.7 אלף.

לכן, ניתן לומר בביטחון כמעט מלא, שאחוז בני 60 ומעלה המתגוררים לבד הוא בין 21.8

לבין 23.2 אלף.

הערה

כדי להזהיר את הקורא מין השימוש באומדנים החשופים לטעויות גבוהות, אומדנים עם טעויות דגימה יחסיות בין 25% לבין 40% הוצללו, ואומדנים עם טעויות דגימה יחסיות מעל 40% סומנו כ- "...".

במקרה של כשל המדגם (0 מקרים במדגם) האומדן קיבל סימון "--".

השוואות בין אומדנים של קבוצות זרות

באמצעות טעויות הדגימה לשני אומדנים שונים, המתייחסים לקבוצות אוכלוסייה זרות (למשל גברים ונשים), אפשר לערוך השוואה ולבחון האם ההבדל בין שתי הקבוצות אכן משמעותי מבחינה סטטיסטית.

אם $X'(1)$ הוא האומדן לקבוצה 1 ו- $X'(2)$ הוא האומדן לקבוצה 2, אזי האומדן להפרש בין שתי קבוצות אלה הוא: $D' = X'(1) - X'(2)$.

כדי לבדוק אם $X'(1)$ אכן שונה מ- $X'(2)$ באוכלוסייה עצמה, יש לבחון מהי טעות הדגימה של אומדן ההפרש D' : $\sigma'(D') = \sqrt{\sigma'(X'(1))^2 + \sigma'(X'(2))^2}$ כאשר טעות דגימה אחת של האומדנים מתקבלת ע"י חלוקה ב-2 של הערכים המוצגים בלוחות.

בהינתן $\sigma'(D')$, אפשר לבנות רווח-סמך להפרש ברמת ביטחון α : $D' \pm K(\alpha)\sigma'(D')$.

אם רווח-הסמך מכיל את הערך אפס, נאמר כי ההפרש D' אינו מובהק. כלומר, על פי המדגם המסוים בסקר, ברמת הביטחון שנקבעה, אי-אפשר לומר כי $X'(1)$ אכן שונה מ- $X'(2)$ - באוכלוסייה עצמה (למרות שבמדגם הם שונים זה מזה).

אם הרווח אינו מכיל את הערך אפס, נאמר כי קיים הפרש מובהק בין שתי הקבוצות, ובביטחון שנקבע ההפרש יהיה בין $D' - K(\alpha)\sigma'(D')$ לבין $D' + K(\alpha)\sigma'(D')$.

כדי להקל על הקורא, מצורף בסוף פרק המהימנות תרשים שלפיו אפשר למצוא רווח-סמך 95% להפרש בין קבוצות זרות. השימוש בתרשים נעשה כדלהלן:

עבור האומדנים וטעויות הדגימה המוצגים בפרסום, מוצאים בלוחות את טעויות הדגימה בעבור שני האומדנים $X'(1)$ ו- $X'(2)$. ערכים אלה מצוינים על גבי שתי העמודות הקיצוניות בתרשים. מעבירים ביניהם קו דמיוני כדי לקבל את $2\sigma'(D')$ בעמודה האמצעית בתרשים (לנוחות הקריאה ניתן לשנות את קנה המידה של הגבולות).

אם ההפרש D' קטן בערכו המוחלט מ- $2\sigma'(D')$, ההפרש אינו מובהק.

אם ההפרש D' גדול בערכו המוחלט מ- $2\sigma'(D')$, ההפרש מובהק והוא נמצא בתחום $D' \pm 2\sigma'(D')$.

לדוגמא: לפי אותו לוח סיכום לפרק 2, "לוח א': בני 60 ומעלה, לפי עיסוק בזמן הפנוי", נשווה בין אחוז הגברים והנשים בקבוצת הגיל 65-69 המאזינים לרדיו לפחות פעם ביום, כל זאת ברמת ביטחון של 95%.

אחוז המאזינים לרדיו כל יום בקבוצות גיל 65-69 הוא:

87.9% ± 3.2% לגברים.

82.7% ± 3.6% לנשים.

נשאלת השאלה, האם אכן ההפרש בין קבוצות אלה הנו מובהק. לכאורה לפי האומדנים בלבד, ניתן להניח כי יש הבדל בין המינים בקבוצת גיל זו. ההפרש בין אחוז הגברים המאזינים לרדיו כל יום לבין הנשים המאזינות לרדיו כל יום בקבוצת הגיל 65-69 נאמד ב- $D' = 5.2$.

עפ"י נוסחה ניתן לחשב את טעות הדגימה של ההפרש:

$$2\sigma'(D') = \sqrt{3.2^2 + 3.6^2} = 4.8$$
$$D' \pm 2\sigma(D') = 5.1 \pm 4.8$$

רווח זה אינו מכיל את הערך "אפס" ועל כן נאמר כי ההפרש D' מובהק, כלומר, על פי המדגם המסוים בסקר, ברמת ביטחון של 95%, ניתן לומר כי אחוז הגברים המאזינים לרדיו כל יום שונה מאחוז הנשים המאזינות לרדיו כל יום, בקבוצת הגיל 65-69. רווח הסמך להפרש ברמת ביטחון 95% הוא 0.3%-9.9%.

(2) טעויות לא-מדגמיות

כאמור, האומדנים המתקבלים מסקר מדגמי לוקים, בנוסף לטעויות הדגימה, גם בטעויות נוספות ממקורות שונים. את טעויות הדגימה ניתן לאמוד מתוך נתוני הסקר עצמו, בעוד שאת הטעויות הלא-מדגמיות קשה או לא ניתן כלל להעריך.

גם בסקר זה עשויים גורמים שונים להיות מקור לטעויות לא-מדגמיות. ביניהם:

(1) הטיות הנובעות מבעיות כיסוי במסגרת

אם מדגם המשיבים בסקר אינו "מייצג" את כל אוכלוסיית הסקר, אומדני הסקר חשופים להטיות. אי-כיסוי האוכלוסייה יגרום בוודאות להטייה כלפי מטה של אומדני השכיחויות, ויתקבלו אומדנים חסרים. בנוסף לכך, אם קבוצת אי-הכיסוי שונה בתכונותיה מקבוצת המקרים שכוסו בסקר, נקבל גם הטיה (בכיוון לא ידוע) של אומדנים אחרים, כמו אחוזים, ממוצעים וכדומה.

הסברים על בעיות כיסוי המסגרת הוצגו בסעיף 3 בפרק שיטות הסקר לעיל.

(2) הטיות הנובעות מאי השבה

בכל סקר (מדגמי או מפקדי), לא מצליחים לקבל שאלון מלא מכל הנדגמים. על חלק מאלה שאינם משיבים ידוע כי הם שייכים לאוכלוסיית הסקר ("נפל"), ועל חלק מהם אין כל מידע על שייכותם לאוכלוסיית הסקר ("נפס"). נתונים על היקף אי-ההשבה הוצגו בסעיף 7 בפרק שיטות הסקר.

כאמור, ייתכן שהנפל שונה בתכונותיו מהנחקרים, והדבר יכול לגרום להטיות באומדני הסקר. חלק מההטיות תוקנו באמצעות שיטת האמידה, אך לא כולן.

(3) טעויות הנובעות מנתונים שגויים בסקר

הנתונים שעל פיהם מתקבלים האומדנים של כל סקר אינם תמיד מדויקים ושלמים. נתונים שגויים יכולים להימצא בקובץ המשמש לקבלת האומדנים, למשל, בגלל טעויות תשובה בסקר, או טעויות בקליטת הנתונים במחשב. באמצעות ביקורות לוגיות מצמצמים את ממדי הבעיה.