

ה. מהימנות האומדנים

הממצאים המובאים בפרסום זה הם **אומדנים** המבוססים על סקר מדגמי ולכן הם עלולים להיות חשופים לטעויות משני סוגים: טעויות מדגמיות וטעויות לא-מדגמיות.

1. טעויות מדגמיות

טעויות מדגמיות נובעות מכך שחוקרים רק מדגם ולא את כלל האוכלוסייה. המדגם שהסקר הנוכחי מבוסס עליו הוא רק אחד מתוך מספר רב של מדגמים אפשריים, שהיו יכולים להתקבל באותה שיטת דגימה ובאותו גודל. ברור כי אומדנים המבוססים על מדגמים שונים יהיו שונים זה מזה, וכמעט כולם יהיו שונים מהערך שהיה מתקבל אילו נערך מפקד מלא ("הערך המפקדי").

האומדן X' הוא הערך הנאמד לפי המדגם המסוים של סקר זה עבור הערך המקביל ל- 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.

כיוון שברור שאומדן מסוים סוטה (כמעט בוודאות) מהערך המפקדי, לא רצוי להתייחס בלעדית לערך הנאמד ממדגם אלא גם לרווח שבו הערך המפקדי עשוי להימצא בהסתברות נתונה (ברמת ביטחון מסוימת). אפשר לעשות זאת באמצעות **טעות הדגימה**.

טעות דגימה יחסית של האומדן נאמדת על-ידי היחס שבין טעות הדגימה לבין האומדן, כלומר $\frac{\sigma(X')}{X'}$.

רווח סמך לאומדן הוא רווח שמכיל את הערך המפקדי X ברמת ביטחון מסוימת, נתונה מראש. האומדן X' על-פי המדגם, והאומדן לטעות הדגימה שלו - $\sigma'(X')$, משמשים לבניית 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%; רווח הסמך ברמת ביטחון זו מתקבל מהרווח שבין $X' - 2\sigma'(X')$ לבין $X' + 2\sigma'(X')$. משמעותו היא שניתן לומר בביטחון של 95% שהערך המפקדי נמצא ברווח זה.

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

$k(\alpha)$ (שהוא מספר טעויות הדגימה לכל כיוון) ייקבע בהתאם לרמת הביטחון הדרושה α .

טעויות דגימה חושבו עבור כל האומדנים המובאים בפרסום זה. נוסף על כך, בתקליטור המצורף לפרסום זה מובאים לוחות שמקבילים ללוחות של האומדנים ושבהם שתי טעויות הדגימה (רווח סמך 95%).

כדי להזהיר את הקורא משימוש באומדנים החשופים לטעויות דגימה גבוהות, אומדנים עם טעויות דגימה יחסיות בין 25% ל-40% מוקפים בסוגריים כשהם מוצגים בלוחות הפרסום. ואילו אומדנים עם טעויות דגימה יחסיות מעל 40% אינם מוצגים, ובמקומם מופיע הסימן "...".

דוגמה:

בשנת 2002, אומדן המשפחות החד-הוריות עם ילדים עד גיל 17 הגרות במשקי בית שבהם יותר ממשפחה אחת הוא 19.7 אלף (ראה לוח 14). לפי הלוח המקביל בתקליטור המצ"ב, שתי טעויות הדגימה לאומדן הן 3.0 אלף ורווח הסמך ברמת ביטחון 95% הוא 19.7 ± 3.0 . כלומר, ניתן לומר בביטחון של 95% כי מספר המשפחות החד-הוריות עם ילדים עד גיל 17 הגרות במשקי בית שבהם יותר ממשפחה אחת הוא בין 16.7 לבין 22.7.

אם מעוניינים בביטחון גבוה מאוד של 99.5% יש לחלק את שתי טעויות הדגימה בשתיים ולכפול את טעויות הדגימה המתקבלות ב- $k(\alpha) = 2.8$. רווח הסמך הוא $19.7 \pm 2.8 * 1.5$. כלומר, ניתן לומר בביטחון כמעט מלא שמספר המשפחות החד-הוריות עם ילדים עד גיל 17 הגרות במשקי בית שבהם יותר ממשפחה אחת הוא בין 15.5 ל-23.9.

הערה: רווחי הסמך הם בדרך כלל סימטריים סביב האומדן, אך עבור אומדנים המבוססים על מספר קטן של מקרים במדגם (פחות מארבעים מקרים) רווחי הסמך אינם סימטריים. במקרים אלו, הן האומדן עצמו והן האומדן לטעות הדגימה שלו חשופים לטעות גבוהה.

השוואה בין אומדנים של קבוצות זרות

באמצעות טעויות הדגימה של אומדנים המתייחסים **לשתי קבוצות זרות** (למשל גברים ונשים), אפשר לבחון אם ההפרש ביניהן משמעותי מבחינה סטטיסטית.

אם X'_1 הוא האומדן לקבוצה 1 ו- X'_2 הוא האומדן לקבוצה 2, אז האומדן להפרש בין שתי קבוצות אלו הוא:

$$D' = X'_1 - X'_2$$

רווח סמך להפרש D' ברמת ביטחון α הוא: $D' \pm k(\alpha)\sigma(D')$.

אם מעוניינים ברמת ביטחון של 95%, רווח הסמך להפרש D' הוא: $D' \pm 2\sigma(D')$.

האומדן לשתי טעויות הדגימה של אומדן ההפרש מחושב כך:

$$2\sigma(D') = \sqrt{(2\sigma(X'_1))^2 + (2\sigma(X'_2))^2}$$

המצ"ב. $2\sigma(X'_1)$ ו- $2\sigma(X'_2)$ הן שתי טעויות הדגימה המתאימות לכל אחד מהאומדנים והן נמצאות בתקליטור

הערה: אפשר להשתמש בנוסחה שלעיל גם לקבלת שתי טעויות דגימה לאומדן ההפרש בין שתי שנים, אולם אם ההפרש הוא בין אומדנים **מקבילים של שתי שנים סמוכות**, יש לכפול את שתי טעויות הדגימה המתקבלות מהנוסחה בגורם תיקון השווה ל-0.7. במקרה זה, האומדן לשתי טעויות הדגימה של אומדן ההפרש יהיה:

$$2\sigma(D') = 0.7 \cdot \sqrt{(2\sigma(X'_1))^2 + (2\sigma(X'_2))^2}$$

להלן שני הקריטריונים להשוואה בין קבוצות. הקריטריונים **שקולים זה לזה**.

קריטריון א:

אם רווח הסמך מכיל את הערך אפס נאמר כי ההפרש אינו מובהק. כלומר, על-פי המדגם המסוים בסקר, **אי אפשר** לומר ברמת הביטחון של α , כי X_1 אכן שונה מ- X_2 באוכלוסייה עצמה (גם אם במדגם הם שונים זה מזה).

אם הרווח אינו מכיל את הערך אפס נאמר בביטחון של α כי קיים הפרש מובהק בין שתי הקבוצות. בכל מקרה, ניתן לומר ברמת ביטחון α כי ההפרש בין שתי הקבוצות הוא בין $D' - k(\alpha)\sigma(D')$ לבין $D' + k(\alpha)\sigma(D')$.

קריטריון ב:

אם ההפרש D' קטן (בערכו המוחלט) מ- $k(\alpha)\sigma(D')$, אזי ההפרש אינו מובהק ברמת ביטחון α .

אם ההפרש D' גדול (בערכו המוחלט) מ- $k(\alpha)\sigma(D')$, אזי ההפרש מובהק ברמת ביטחון α .

דוגמה:

בשנת 2002 גרו במשקי הבית כ-476.0 אלף גברים בגיל 25-34, 10.5% מהם במשקי בית לא-משפחתיים. מספר הנשים שגרו במשקי בית בקבוצת גיל זו עמד על 476.5 אלף, ו-7.6% מהן גרו במשקי בית לא-משפחתיים (ראה לוח 12).

אומדן ההפרש הוא: $D' = 10.5\% - 7.6\% = 2.9\%$.

על-פי הלוח המקביל של שתי טעויות הדגימה בתקליטור המצ"ב, שתי טעויות הדגימה של האומדנים הן 1.1% עבור גברים ו-0.9% עבור נשים.

לפי הנוסחה דלעיל, שתי טעויות הדגימה של אומדן ההפרש הן: $2\sigma(D') = \sqrt{1.1^2 + 0.9^2} = 1.4\%$, ולכן רווח הסמך 95% להפרש זה הוא: 2.9 ± 1.4 . כיוון שרווח הסמך אינו מכיל את הערך אפס, נאמר ברמת ביטחון של 95% שההפרש מובהק ושיש הבדל בין אחוז הגברים בגיל 25-34 הגרים במשקי בית לא-משפחתיים מבין כלל הגברים בקבוצת גיל זו לבין האחוז המקביל עבור נשים. ההפרש נמצא בתחום שבין 1.5% ($2.9 - 1.4 =$) לבין 4.3% ($2.9 + 1.4 =$).

כדי להקל על הקורא, מצורף בסוף פרק זה תרשים עזר לקבלת הערך של שתי טעויות דגימה לאומדן ההפרש בין קבוצות זרות. זאת כדי לבחון אם ההפרש בין שתי קבוצות זרות מובהק ברמת ביטחון של 95%, מבלי להזדקק לנוסחת החישוב המובאת לעיל. לשם כך יש לבצע את הפעולות הבאות:

1. יש למצוא בלוחות המקבילים בתקליטור המצ"ב את שתי טעויות הדגימה של האומדנים המשווים:
 $2\sigma(X'_1)$ ו- $2\sigma(X'_2)$.

2. יש להציג את ערכי שתי טעויות הדגימה שלהם, המתקבלים בסעיף 1 לעיל, כנקודות על שתי העמודות הקיצוניות בתרשים העזר המצורף לפרק זה.

3. יש למתוח קו בין שתי הנקודות שהתקבלו בסעיף 2. הקו חותך את העמודה האמצעית בתרשים. המספר המוצג על נקודת החיתוך של הקו עם העמודה האמצעית הוא ערכן של שתי טעויות דגימה של ההפרש בין האומדנים: $2\sigma(D')$.

4. 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, וניתן לומר שהוא ימצא בתחום: $D' \pm 2\sigma(D')$. לחלופין, אפשר לבנות רווח סמך להפרש בין האומדנים באמצעות שתי טעויות הדגימה להפרש האומדנים, ולהפעיל את קריטריון א המופיע לעיל.

אם המספרים בתרשים אינם מתאימים לערכי טעויות הדגימה של האומדנים המשווים, ניתן להכפיל את טעויות הדגימה בגורם כפל כלשהו (למשל 1/10, 10, 100) ולהציג את הערכים המתקבלים על העמודות הקיצוניות. את המספר המתקבל מהעמודה האמצעית בתרשים יש לחלק באותו גורם כפל, כדי לקבל את ערכן של שתי טעויות דגימה של ההפרש.

דוגמה:

בשנת 2002, אחוז משקי הבית שבהם שלושה דורות ויותר מכלל משקי הבית המשפחתיים הוא 5.2%. באוכלוסייה הערבית ו-4.5% באוכלוסייה היהודית (ראה לוח 6).

אומדן ההפרש הוא $D' = 5.2\% - 4.5\% = 0.7\%$.

על-פי הלוח המקביל של שתי טעויות הדגימה בתקליטור המצ"ב, שתי טעויות הדגימה של האומדנים הן 1.0% עבור האוכלוסייה הערבית ו-0.4% עבור האוכלוסייה היהודית.

כופלים את שתי טעויות הדגימה במאה כדי לקבל קנה מידה בתרשים. מעבירים קו ישר בין הנקודות 100 ו-40 בשתי העמודות הקיצוניות של התרשים. הערך המתקבל בנקודה שבה הקו חוצה את העמודה האמצעית הוא 107. מחלקים את 107 במאה והתוצאה המתקבלת עבור שתי טעויות הדגימה היא $2 \cdot \sigma(D') = 1.07\%$.

הערך שהתקבל גדול מאומדן ההפרש, ולכן ניתן לומר ברמת ביטחון של 95% שלא קיים הפרש מובהק בין היהודים לבין הערבים באחוז משקי הבית המשפחתיים שבהם שלושה דורות ויותר.

2. טעויות לא-מדגמיות

טעויות לא-מדגמיות בסקר יכולות לנבוע ממקורות רבים בכל שלבי איסוף הנתונים ועיבודם, והן מצויות גם במפקד מלא.

הטעויות הלא-מדגמיות העיקריות בסקר הן:

א. **טעויות הנובעות מאי-היענות:** טעויות הנובעות מכך שמשקי בית מסוימים אינם נחקרים בגלל היעדרות או סירוב. דבר זה עלול לגרום להטיה מסוימת באומדנים, משום שתכונותיהם של האנשים הכלולים במשקי בית אלו יכולות להיות שונות מאלה של אנשים שנחקרו.

ב. **טעויות תשובה:** טעויות הנובעות מאי-הבנת השאלות, מחוסר רצון או חוסר יכולת להשיב נכון או מהצגת שאלה בצורה לא נכונה.

ג. **טעויות בשלבי העיבוד:** טעויות בתהליך עיבוד החומר, כגון טעויות בסימול השאלונים ובהקלדתם. חלק מהטעויות מתוקנות באמצעות ביקורות הנערכות על החומר.

בעוד שאת טעויות הדגימה אפשר לאמוד מתוך נתוני הסקר עצמו, קשה (ואף בלתי אפשרי) להעריך טעויות לא-מדגמיות. בפרסום זה לא מובאות הערכות כמותיות על הטעויות הלא-מדגמיות. חשוב להדגיש שבתכנון הסקר ובביצועו נעשה מאמץ להקטין ככל האפשר את מספר הטעויות.