

## סקר השכלה גבוהה (2018)

### 4. שיטות (מתודולוגיה)

#### 4.1 אוכלוסיית הסקר

אוכלוסיית הסקר כוללת את כל הסטודנטים שהחלו ללמוד לתואר ראשון בשנת תשע"א (2010/2011) במוסדות להשכלה גבוהה (אוניברסיטאות<sup>1</sup>, מכללות אקדמיות ומכללות אקדמיות להוראה) למעט האוניברסיטה הפתוחה. סטודנטים חדשים באוניברסיטאות ובמכללות האקדמיות הם סטודנטים שהמוסד הגדיר אותם כתלמידי שנה ראשונה וגם שלא הופיעו בשנים קודמות כסטודנטים בקבצים המנהליים.

#### 4.2 נושאי החקירה

נושאי החקירה של הסקר הם: סיום לימודים, שביעות הרצון מהלימודים במוסד, שיפור מיומנות וכישורים כתוצאה מהלימודים, התמחות בסטאג' (מתייחס למכללות להוראה בלבד), תעסוקה, הקשר בין הלימודים לעבודה ומיומנות נדרשת בעבודה, מאפייני עבודה ושביעות רצון מהעבודה.

#### 4.3 מסגרת הדגימה

בסקר מחושבים אומדנים נפרדים לסטודנטים הלומדים באוניברסיטאות ומכללות אקדמיות, ואומדנים נפרדים ללומדים במכללות הוראה, על כן נוצרו שתי מסגרות דגימה נפרדות וזרות המבוססות על קבצי סטודנטים לשנת הלימודים תשע"א שהתקבלו מהמוסדות (אוניברסיטאות ומכללות אקדמיות) וממשרד החינוך (מכללות אקדמיות להוראה). מתוך קבצים אלו אותרו סטודנטים לשנה א' ולאחר מכן נוכו סטודנטים שאינם שייכים לאוכלוסיית הסקר, אלו העונים על הקריטריונים הבאים:

1. נוכו סטודנטים שלמדו בשנים קודמות במוסדות להשכלה גבוהה (למעט הפתוחה).
2. נוכו כפילויות בתוך כל מוסד ובין המוסדות השונים בשנת הלימודים תשע"א.
3. נוכו סטודנטים זרים שאינם מופיעים במרשם האוכלוסין.
4. נוכו סטודנטים שלמדו בשלוחות של האוניברסיטאות.

סך הכל מסגרת הדגימה הראשונית הייתה 16,190 סטודנטים מהאוניברסיטאות, 22,686 מהמכללות האקדמיות ו-5,568 סטודנטים מהמכללות האקדמיות להוראה. ממנה הורדו מכללות בהם היו מספר קטן של סטודנטים וכן סטודנטים שגילם פחות מ-19 (המספרים הסופיים מופיעים להלן).

<sup>1</sup> בשנה זו מכללת אריאל טרם הוגדרה כאוניברסיטה.

## 4.4 שיטת הדגימה

המדגם לסקר הוצא כבר ב-2012 (המחזור הראשון לסקר), ושימש גם למחזור הפקידה השני שנערך ב-2018, בניכוי המקרים שהתגלו לקראת היציאה לפקידה ב-2018 כלא שייכים לאוכלוסייה. בפעולה זו הושמטו בסה"כ 49 סטודנטים.

### 4.4.1 אוניברסיטאות ומכללות אקדמיות

שיטת הדגימה הייתה מדגם מקרי-שיטתי חד-שלבי בשכבות. שכבות הדגימה הוגדרו להיות שילוב של מוסד הלימוד ומקצוע מקובץ. הקבצת מקצועות הלימוד הן תשע קבוצות של מקצועות על פי בקשת ות"ת עבור ועדת איכות (כלכלה, עבודה סוציאלית, פסיכולוגיה, מדעי המחשב, רפואה כללית, ביולוגיה, הנדסת חשמל ואלקטרוניקה והנדסת מחשבים, הנדסת תעשייה וניהול, אדריכלות), ועוד שתי הקבוצות של יתר המקצועות (תחומי לימוד "רכים" - מדעי הרוח והחברה, ותחומי לימוד "קשים" - מדעים והנדסה). סך הכל הוגדרו 146 שכבות דגימה. גודל המדגם נקבע להיות 9,300, וזאת כדי לעמוד בדרישות הדיוק לקבוצות הפרסום הראשיות של הסקר ותחת אילוצים תקציביים.

הקצאת המדגם לשכבות הדגימה (מס' הסטודנטים שיידגמו בכל שכבה) נעשתה בצורה שתבטיח דיוק דומה בכל השכבות של מקצועות הלימוד "הרכים" וה"קשים", כאשר הדיוק בשכבות של תשעת המקצועות המתוגברים נקבע להיות רב יותר וזאת בכדי להבטיח את תגבורם. שבר הדגימה בכל שכבת פרסום הוא היחס שבין גודל המדגם שהתקבל לבין מספר הסטודנטים במסגרת הדגימה.

לקראת הוצאת המדגם מוינו הרשומות בכל שכבת דגימה לפי מקצוע, קבוצת אוכלוסייה (יהודים, עולים, ערבים), מין וגיל.

### 4.4.2 מכללות הוראה

שיטת הדגימה הייתה מדגם מקרי-שיטתי חד-שלבי בשכבות. שכבות הדגימה הוגדרו להיות שילוב של המשתנה **פיקוח לאום** (ממלכתי עברי, ממלכתי דתי, ממלכתי ערבי) **ותחום לימוד** (אנגלית, מתמטיקה, מדעי המחשב, מדעים ב.צ.ג., חינוך, יתר מדעי הרוח, יתר מדעי הטבע) ובנוסף הוגדרה שכבה נוספת, בכל אחת מקבוצות פיקוח לאום עבור סטודנטים מצטיינים. סך הכל הוגדרו 22 שכבות דגימה (לא בכל קבוצות פיקוח לאום קיימים כל תחומי הלימוד). גודל המדגם נקבע להיות 1,000 נדגמים וזאת כדי לעמוד בדרישות הדיוק לקבוצות הפרסום הראשיות של הסקר ותחת אילוצים תקציביים.

הקצאת המדגם לשכבות הדגימה (מס' הסטודנטים שיידגמו בכל שכבה) נעשתה בצורה שתבטיח דיוק דומה בכל השכבות, כאשר הדיוק בשכבות של סטודנטים מצטיינים נקבע להיות רב יותר וזאת בכדי להבטיח תגבור שלהם. שבר הדגימה בכל שכבה הוא היחס שבין גודל המדגם שהתקבל לבין מספר הסטודנטים במסגרת הדגימה.

לקראת הוצאת המדגם מוינו הרשומות בכל שכבת דגימה לפי מסלול למ"ס (מסלול לגנות, לגיל הרך, לכיתות ג'-ו', לכיתות ז'-י', בין מסלולי, חינוך בלתי פורמאלי), מין וגיל.

#### 4.5 עבודת השדה

להלן תוצאות הפקידה :

סוג מוסד/מדגם	תיאור	מספרים מוחלטים	אחוזים מסך הנדגמים
אוניברסיטאות ומכללות אקדמיות	סה"כ נדגמים	9,300	100.0%
	מהם: אינם שייכים לאוכלוסיית הסקר	37	0.4%
	לא ידוע אם שייכים לאוכלוסייה (לא אותרו, לא נוצר קשר וכד')	462	5.0%
	שייכים לאוכלוסיית הסקר	8,801	94.6%
	מהם: משיבים	8,057	83.9%
	לא-משיבים	965	10.4%
מכללות אקדמיות להוראה	סה"כ נדגמים	1,000	100.0%
	מהם: אינם שייכים לאוכלוסיית הסקר	20	2%
	לא ידוע אם שייכים לאוכלוסייה (לא אותרו, לא נוצר קשר וכד')	90	9%
	שייכים לאוכלוסיית הסקר	890	89%
	מהם: משיבים	749	74.9%
	לא-משיבים	127	12.7%

דרך נוספת להעריך את טיב האומדנים על סמך תוצאות הפקידה היא לפזר את אלו שלא היה ידוע אם הם שייכים לאוכלוסיית הסקר בין המשיבים, הלא משיבים והלא שייכים לאוכלוסיית הסקר, זאת באופן יחסי לגדלם.

תוצאות הפקידה משוקללות בהיפוך הסתברות הדגימה :

סוג מוסד/מדגם	תיאור	מספרים מוחלטים משוקלים בהיפוך הסתברות הדגימה	אחוזים מתוך סך מסגרת הדגימה	אחוז מתוך השייכים לאוכלוסיית הסקר
אוניברסיטאות ומכללות אקדמיות	גודל מסגרת הדגימה	38,571	100.0%	
	לא שייכים לאוכלוסיית הסקר	197	0.5%	
	שייכים לאוכלוסיית הסקר	38,374	99.5%	100.0%
	מהם: משיבים	33,936	88.0%	88.4%
	לא-משיבים	4,438	11.5%	11.6%
מכללות אקדמיות להוראה	גודל מסגרת הדגימה	4,777	100.0%	
	לא שייכים לאוכלוסיית הסקר	142	3.0%	
	שייכים לאוכלוסיית הסקר	4,635	97.0%	100.0%
	מהם: משיבים	4,003	83.8%	86.4%
	לא-משיבים	632	13.2%	13.6%

בסקר המשיבים ענו בשיטות איסוף שונות. אחוזי השבה לפי שיטת האיסוף:

סוג מוסד/מדגם	דואר	אינטרנט	טלפון
אוניברסיטאות ומכללות אקדמיות	0.5%	42.7%	56.8%
מכללות אקדמיות להוראה	7.0%	23.1%	76.2%

#### 4.6 שיטת האמידה

בדומה למחזור הראשון בסקר, גם במחזור השני חושב משקל לכל אחד מהמשיבים. המשקל נקרא מקדם ניפוח, והוא מבטא את מספר הסטודנטים באוכלוסייה שאותם הוא "מייצג" במחזור החקירה השני. מקדם הניפוח מבוסס על שלושה שלבים: (א) היפוך הסתברות הדגימה, (ב) תיקון אי-השבה, (ג) כיוול (התאמה) לאומדנים חיצוניים. כללית, ככל

שהסתברות הדגימה גבוהה יותר ושיעורי ההשבה בשכבה גבוהים יותר, כך המשיב "מייצג" מספר קטן יותר של סטודנטים, ומקדם הניפוח נמוך יותר. באופן דומה, ככל שהסתברות הדגימה נמוכה יותר ושיעורי ההשבה בשכבה נמוכים יותר, כך המשיב קיבל מקדם ניפוח גבוה יותר.

במחזור הראשון בסקר נתגלתה תופעה שעלתה מהנתונים המנהליים והיא שמבין אלו שהפסיקו ללמוד ונדגמו לסקר נתגלו שיעורי אי השבה גבוהים יותר מאשר אלו שהמשיכו ללמוד. תופעה זו ניכרת גם במחזור השני והיא קיבלה התייחסות בשלב כיוול הנתונים כדי לצמצם הטיות באומדנים.

חישוב אומדני הסקר נעשה על ידי שקלול הנתונים המדווחים בשאלונים במקדמי הניפוח.

## 4.7 מהימנות האומדנים

האומדנים המוצגים בפרסום מבוססים על סקר מדגמי, ותיכנה בהם טעויות שונות.

### 4.7.1 טעויות דגימה והשימוש בהן

טעויות מדגמיות נובעות מכך שנחקר רק מדגם ולא כלל אוכלוסיית הסקר (להלן: האוכלוסייה). המדגם שעליו מבוסס הסקר הנוכחי הוא רק אחד מתוך מספר רב של מדגמים אפשריים שהיו יכולים להתקבל באותה שיטת דגימה ובאותו גודל. ברור כי אומדנים המבוססים על מדגמים שונים נבדלים זה מזה, וכמעט כולם שונים מהערך שהיה מתקבל אילו נאסף המידע מכל האוכלוסייה ולא ממדגם.

האומדן  $\hat{X}$  הוא הערך הנאמד לפי המדגם המסוים של סקר, במקום הערך המקביל לו שהיה מתקבל לו נאסף המידע מכל האוכלוסייה.

טעות דגימה של האומדן  $\sigma(\hat{X})$  היא מידה ממוצעת של השוני בין כל האומדנים השונים שהיו יכולים להתקבל מכל המדגמים האפשריים באותו גודל ובאותה שיטה, לבין הערך המתקבל ממידע הנאסף מכל האוכלוסייה, באותם התנאים שבסקר. טעות הדגימה ניתנת לאמידה מנתוני הסקר.

לעתים נוח להעריך את דיוק האומדן לפי **טעות הדגימה היחסית** המוגדרת כטעות הדגימה

של האומדן, חלקי הערך הנאמד –  $\frac{\sigma(\hat{X})}{\hat{X}}$ . נהוג להכפיל ערך זה ב-100 כדי לבטא אותו באחוזים. יש להיזהר משימוש באומדנים החשופים לטעויות דגימה יחסיות גדולות (מעל 25%).

### 4.7.2 רווח-סמך של אומדן

כיוון שברור שאומדן מסוים –  $\hat{X}$  – סוטה (כמעט בוודאות) מהערך  $X$ , רצוי להתייחס לא רק לערך הנאמד ממדגם אלא גם לתחום שבו הערך  $X$  עשוי להימצא בהסתברות נתונה (ברמת ביטחון מסוימת). ניתן לעשות זאת באמצעות טעות הדגימה.

רווח-סמך של אומדן הוא רווח שמכיל את  $X$  ברמת ביטחון מסוימת, נתונה מראש. רווח-הסמך נבנה על סמך האומדן  $\hat{X}$  והאומדן של טעות הדגימה שלו  $\sigma(\hat{X})$ , המתקבלים מהמדגם, ובהתאם לרמת הביטחון שנקבעה.

מקובל להציג רווח-סמך ברמת ביטחון של 95%, ורווח-סמך ברמת ביטחון זו הוא בין  $\hat{X} - 2\sigma(\hat{X})$  לבין  $\hat{X} + 2\sigma(\hat{X})$ . ניתן לומר בביטחון של 95% שהערך שהיה מתקבל מכל האוכלוסייה נמצא בתחום זה, בהנחה שאין טעויות אחרות.

ניתן לקבוע רמת ביטחון גבוהה או נמוכה יותר כפי שמראה הטבלה:

$1-\alpha$	67%	80%	90%	95%	99.5%
$K_{1-\alpha}$	1.0	1.3	1.7	2.0	2.8

$K_{1-\alpha}$  הוא מספר טעויות הדגימה שיש להוסיף לאומדן או להחסיר ממנו כדי לקבל רווח-סמך ברמת הביטחון הדרושה, אשר מקובל לסמן ב- $1-\alpha$ .

**דוגמה 1:** האומדן של אחוז המועסקים בעת ביצוע הסקר בקרב אלו שלמדו באוניברסיטאות הוא 82.6% (לוח 1) וטעות הדגימה של האומדן היא 1.0%. ניתן לומר בביטחון של 95% שאחוז המועסקים מבין אלו שלמדו באוניברסיטאות נמצא בתחום של  $82.6\% \pm 2 \cdot 1.0\%$ , כלומר בין 80.6% ל-84.6%.

רווחי-הסמך הם בדרך כלל סימטריים ביחס לאומדן, אך לגבי אומדנים המבוססים על מספר קטן של מקרים במדגם, רווחי-הסמך עלולים להיות לא סימטריים. במקרים אלה, הן האומדן עצמו והן האומדן של טעות הדגימה שלו חשופים לטעות גדולה. כמו כן יש להיזהר מרווחי-סמך של אומדני אחוזים אשר עלולים במקרים מיוחדים לכלול ערכים מחוץ לתחום של 0 עד 100.

#### 4.7.3 טעויות דגימה ורווחי-סמך של הפרשים בין קבוצות זרות (שאינן חופפות)

בהמשך מובאות הנחיות כיצד לחשב אומדנים של הפרשים בין קבוצות זרות, לבנות עבורם רווחי-סמך ולבחון את מובהקות ההפרשים על סמך האומדנים וטעויות הדגימה. ההנחיות הן עבור אומדני הפרשים בין שני אומדנים "מאותו סוג" המתייחסים לקבוצות אוכלוסייה זרות (כגון מוסדות לימוד ותחומי לימוד שונים). אין אפשרות לערוך את החישובים לקבוצות שאינן זרות על סמך האומדנים וטעויות הדגימה בלבד.

אם  $\hat{X}_1$  הוא האומדן של קבוצה 1 ו-  $\hat{X}_2$  הוא האומדן של קבוצה 2, אומדן ההפרש בין שתי הקבוצות הוא  $\hat{D} = \hat{X}_1 - \hat{X}_2$ .

כדי להסיק אם  $\hat{X}_1$  אכן שונה מ-  $\hat{X}_2$  באוכלוסייה, יש לחשב את טעות הדגימה של אומדן ההפרש  $\hat{D}$ :

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

בנוסחה זו  $\hat{\sigma}(\hat{X}_1)$  ו-  $\hat{\sigma}(\hat{X}_2)$  הן טעויות הדגימה של האומדנים  $\hat{X}_1$  ו-  $\hat{X}_2$ , בהתאמה.

אם טעות הדגימה היא  $\hat{\sigma}(\hat{D})$ , רווח-הסמך של אומדן ההפרש  $\hat{D}$  ברמת הביטחון  $1-\alpha$  הוא:

$$\hat{D} \pm K_{1-\alpha} \hat{\sigma}(\hat{D})$$

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  $\hat{X}_1$  אכן שונה מ-  $\hat{X}_2$  באוכלוסייה (אף על פי שבמדגם הם שונים זה מזה).

אם הרווח אינו מכיל את הערך אפס, נאמר כי קיים הפרש מובהק בין שתי הקבוצות, וברמת ביטחון של  $1-\alpha$  ההפרש יהיה בין  $\hat{D} - K_{1-\alpha} \hat{\sigma}(\hat{D})$  לבין  $\hat{D} + K_{1-\alpha} \hat{\sigma}(\hat{D})$ .

**דוגמה 2:** אומדן המועסקים בקרב הסטודנטים שתחום לימודיהם בתואר הראשון

היה מדעי החברה הוא 82.9% ובקרב אלו שלמדו הנדסה –85.1%. טעות

הדגימה של הנתון הראשון היא 1.3% ושל הנתון השני היא 1.3%. אומדן ההפרש הוא:

$$\hat{D} = \hat{X}_2 - \hat{X}_1 = 85.1\% - 82.9\% = 2.18\%$$

טעות הדגימה של ההפרש היא:

$$\hat{\sigma}(\hat{D}) = \sqrt{\hat{\sigma}^2(\hat{X}_1) + \hat{\sigma}^2(\hat{X}_2)} = \sqrt{1.3\%^2 + 1.3\%^2} = 1.83\%$$

רווח-סמך של אומדן ההפרש ברמת ביטחון של 95% הוא  $\hat{D} \pm 2\hat{\sigma}(\hat{D}) = 2.18\% \pm 2 * 1.83\%$ , כלומר בין -1.5% ל-5.8%. הרווח מכיל את הערך אפס, ולכן ניתן לומר כי ההפרש אינו מובהק.

**דוגמה 3:** אומדן המועסקים בקרב הסטודנטים שתחום לימודיהם בתואר הראשון היה רפואה הוא 70.5% ובקרב סטודנטים שלמדו מדעים פיזיקליים-90.8%. טעות הדגימה של הנתון הראשון היא 4.17% ושל הנתון השני היא 4.16% (לוח 12- טעויות דגימה). אומדן ההפרש הוא:

$$\hat{D} = \hat{X}_2 - \hat{X}_1 = 90.8\% - 70.5\% = 20.3\%$$

טעות הדגימה של ההפרש היא:

$$\hat{\sigma}(\hat{D}) = \sqrt{\hat{\sigma}^2(\hat{X}_1) + \hat{\sigma}^2(\hat{X}_2)} = \sqrt{4.16\%^2 + 4.17\%^2} = 5.89\%$$

רווח-סמך של אומדן ההפרש ברמת ביטחון של 95% הוא  $\hat{D} \pm 2\hat{\sigma}(\hat{D}) = 20.3\% \pm 2 * 5.89\%$ , כלומר בין 8.5% ל-32.1%. הרווח אינו מכיל את הערך אפס, ולכן ניתן לומר כי ההפרש מובהק.

הערה: יש להימנע מהסקת מסקנות כוללניות על הבדלים בין שתי קבוצות זרות בהסתמך על השוואות מרובות בין אומדנים. כך למשל, אם התקבלו אומדני הפרש רבים מובהקים במרכיבים שונים של שביעות רצון הסטודנטים בין האוניברסיטאות לבין המכללות האקדמאיות, אין להסיק באופן גורף כי שביעות הרצון באוניברסיטאות שונה משביעות הרצון במכללות אקדמיות. השוואות מרובות מרכיבים דורשות כלים אחרים להסקה, ואילו הכלים המתוארים בהנחיות כאן הם להשוואות פרטניות בלבד.

#### 4.7.4 טעויות לא-מדגמיות

טעויות לא-מדגמיות בסקר יכולות לנבוע מגורמים רבים בכל שלבי איסוף הנתונים ועיבודם, והן מצויות גם באיסוף מידע מכלל האוכלוסייה ולא רק ממדגם של יחידות.

טעויות לא-מדגמיות עיקריות בסקר הן:

(א) טעויות הנובעות מאי-השבה: טעויות הנובעות מכך שנדגמים אינם משיבים עקב היעדרות, סירוב או סיבות אחרות. דבר זה עלול לגרום להטיה מסוימת באומדנים,



משום שתכונותיהם של המשיבים עלולות להיות שונות משל אלה שלא השיבו לסקר. שיטת האמידה מיועדת לצמצם הטיות זו.

- (ב) טעויות תשובה: טעויות הנובעות מאי-הבנת השאלות, מחוסר רצון או מחוסר יכולת להשיב נכון או מניסוח או הצגת שאלה בדרך לא נכונה.
- (ג) טעויות בהקלדת הנתונים: הנתונים של הסטודנטים שהשיבו בדואר מוקלדים במשרד הראשי. כמו כן נתונים שנאספו בראיונות טלפוניים מוקלדים ישירות על ידי הסוקר. בהקלדת נתונים תיתכנה טעויות.
- (ד) טעויות בשלבי העיבוד השונים: טעויות בתהליך עיבוד החומר, כגון טעויות בסימול. חלק מהטעויות מתקנות באמצעות ביקורות של הנתונים המתקבלים.
- בעוד שטעויות הדגימה ניתנות לאמידה מתוך נתוני הסקר, קשה (ולעתים קרובות בלתי אפשרי) להעריך טעויות לא-מדגמיות, ואין בפרסום זה הערכות כמותיות של טעויות אלה. עם זאת חשוב להדגיש שבתכנון הסקר ובביצועו נעשו מאמצים להקטין ככל האפשר את מספר הטעויות מהסוגים השונים.