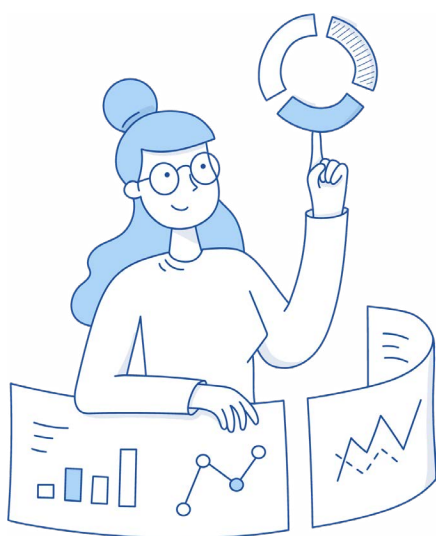




סטטיסטיקה תיאורית על קצה המזלג

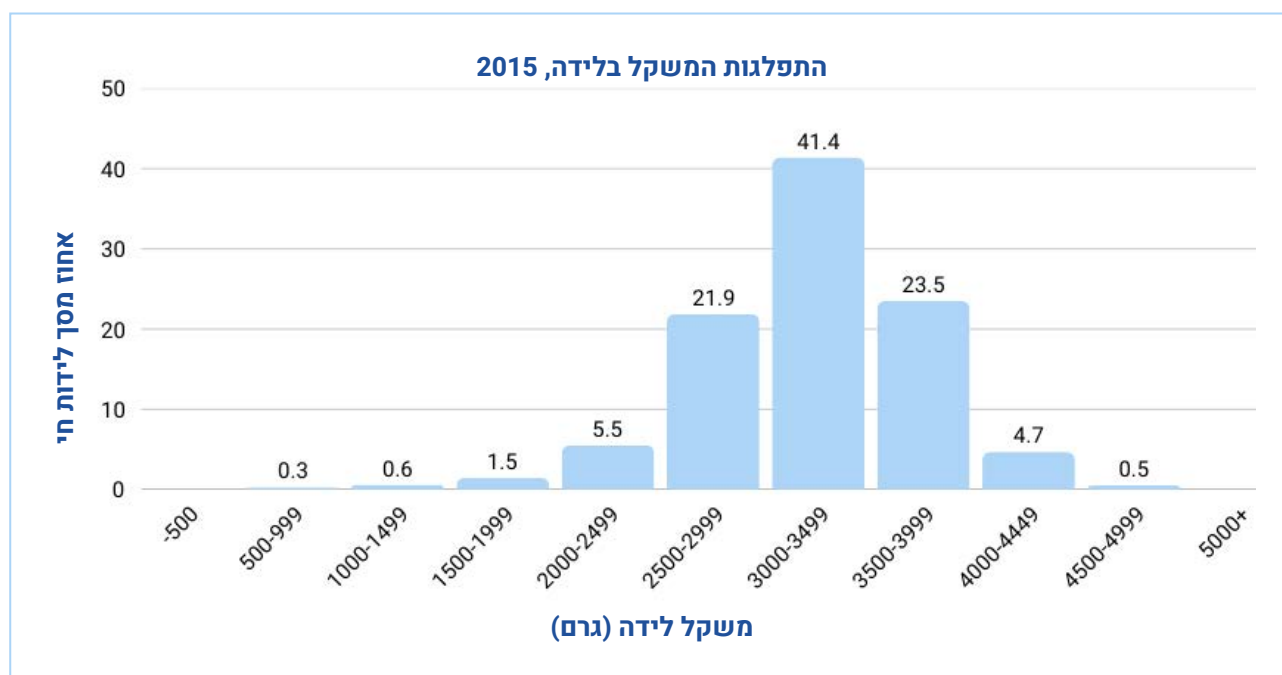


מטרתה של הסטטיסטיקה התיאורית היא לעזור לנו להציג את הערכים של התפלגות **משתנה** באופן תמציתי ואמין.

משתנה = מושג או מאפיין אשר יכולים לקבל ערכים שונים.
דוגמאות למשתנים **כמותיים**: ציון, מספר ילדים בכיתה,
שכר ברוטו, שנות הכשרה.
דוגמאות למשתנים **איכותיים**: מקצוע נלמד, סוג לקות למידה.

לכל משתנה יש **התפלגות** מסוימת והיא זו שמתארת את התנהגותו של המשתנה, כלומר את הצורה שבה מסודרות התצפיות או המופעים שלו.

הנה **דוגמה** להתפלגות של המשתנה "משקל בלידה" מתוך דוח לידות חי בישראל בשנת 2015, משרד הבריאות.
ניתן לראות שרוב היילודים החיים נולדים במשקל שבין 2,500 גרם ל-3,999 גרם.



הסטטיסטיקה התיאורית עוזרת להציג התפלגויות של משתנה והיא מתבססת על **מדדי מיקום מרכזי ומדדי פיזור**.

מדדי מיקום מרכזי

מדדי מיקום מרכזי נותנים מידע באשר לערך שסביבו מתרכזים מרבית הנתונים. המדד המוכר ביותר לכולנו הוא **הממוצע**. למשל, ממוצע הציונים בכיתה ה-2 הוא 85 וממוצע הציונים בכיתה ה-3 הוא 90. ואולם יש מדדי מיקום מרכזי נוספים. למשל, **שכיח** – מדד המבטא את הערך התדיר ביותר בהתפלגות. למשל: ממוצע הציונים בכיתה ה-2 הוא 85, ואולם הציון השכיח ביותר (כלומר הציון היחיד שהמספר הרב ביותר של תלמידים בכיתה קיבלו) הוא 88.

* לעיתים אין כל אפשרות לדווח על ממוצע של משתנה, אלא רק על הערך השכיח ביותר בהתפלגות שלו. כך הדבר למשל במדידה של משתנה איכותני כמו לקות למידה (משתנה הכולל למשל את הערכים: הפרעת קשב, דיסלקציה, דיסקלקוליה). בהקשרו של משתנה לקות למידה נוכל לדווח על סוג הלקות השכיח ביותר בכיתה או בבית הספר כולו.

מדדי פיזור

מדדי פיזור מדווחים עד כמה הערכים שונים זה מזה. מדד הפיזור האינטואיטיבי ביותר הוא **הטווח** והוא מייצג את ההפרש בין שני הערכים הקיצוניים ביותר בהתפלגות. ישנם שני מדדי פיזור נוספים – **שונות**, **סטיית תקן** – שאפשר לחשב בקלות יחסית בתוכנה דוגמת אקסל; ככל שערכם גדול יותר, כך הפיזור בין הערכים הוא רב יותר.

* כאשר ערך השונות שווה אפס, משמעות הדבר היא שאין פיזור כלל וכי כל הנתונים נמצאים בדיוק באותו הערך.

קשר בין משתנים נותן מידע אם התנהגותו של משתנה אחד קשורה בהתנהגותו של משתנה אחר וכיצד. כאשר קיים קשר בין משתנים, ידע על ערך של משתנה אחד מאפשר ידע (בסבירות מסוימת) על ערכו של משתנה אחר. חשוב לציין כי גם אם נמצא קשר בין משתנה א' למשתנה ב', לא נוכל להסיק גם על קשר סיבתי ביניהם. כך למשל, נוכל לומר כי כאשר מופיע ערך מסוים במשתנה א', אז קיימת סבירות גבוהה שיופיע ערך מסוים במשתנה ב', אך לא נוכל להסיק כי משתנה א' משפיע על משתנה ב', ולהיפך.

דוגמה: נניח ששאלנו מדגם מייצג של מתבגרים עם הפרעות אכילה באשר לסיבה המרכזית שהביאה להתפתחות ההפרעה בקרבם, וגילינו את הנתונים האלה:

בנות	בנים	
15%	5%	יחס ביקורתי מצד ההורה
30%	15%	היבטים הקשורים בדימוי גוף
30%	10%	השפעות המדיה
20%	20%	לחצים אקדמיים
5%	50%	דאגות הקשורות בשירות צבאי

במקרה זה, ידיעת המגדר של המתבגר/ת הסובל/ת מהפרעת אכילה יכולה לתת לנו כיוון ראשוני באשר לטיפול הנדרש, גם לו היה נמנע מאיתנו לשוחח עימו/עימה ולהבין את מקור הבעיה.

קשר חיובי

קשר חיובי בין משתנים מעיד שעלייה בערכיו של משתנה אחד קשורה בעלייה בערכיו של המשתנה השני. לדוגמה: מחקר מצא כי יש קשר חיובי בין מספר שעות ההכשרה הנוספת שמורה משלימה במהלך השנים לבין מגוון אסטרטגיות ההוראה שברשותה.

קשר שלילי

קשר שלילי בין משתנים מעיד שעלייה בערכיו של משתנה אחד קשורה בירידה בערכיו של המשתנה השני. לדוגמה: מחקר מצא כי יש קשר שלילי בין מספר התלמידים בכיתה לבין מספר הדיונים המתקיימים בכיתה.

ניתן לדבר על כיוון הקשר (חיובי או שלילי) רק כאשר שני המשתנים הנמדדים הם כמותיים. בדוגמה על הקשר בין מגדר לבין מקור הפרעות האכילה, ניתן לראות כי קיים קשר בין המשתנים, ניתן גם לתאר אותו (למשל: במחצית מהמקרים מקור ההפרעה בקרב בנים קשור בדאגות בשירות הצבאי), אך אין אפשרות לדבר על כיוון הקשר.

תוכנות כמו אקסל יכולות לחשב את עוצמת הקשר (מתאם) וכיוון הקשר המתקיים בין שני משתנים. ערכי המתאם ינועו בין (-1) לבין (+1). ככל שערכו של המתאם קרוב ל-0 אז משמעות הדבר שהקשר בין המשתנים אינו קיים או חלש מאוד. ככל שערכו של המתאם קרוב לערך המוחלט של 1 אז המשמעות היא כי קיים קשר חזק (חיובי או שלילי) בין המשתנים.

יתרונותיה של סטטיסטיקה תיאורית

1

מאפשרת למסור כמות רבה של מידע ביעילות ובפשטות.

לדוגמה: בכנס מקצועות המחר שהתקיים השבוע השתתפו 21 בעלי מקצוע (6 נשים ו-15 גברים). ארבעה משתתפים במקצועות הרפואיים, שניים בתחום ההסעדה והמלונאות ושניים במערכת האכיפה. ותק ממוצע במקצוע 6 שנים ושביעות רצון ממוצעת (על פני סולם הנע בין 1 ל-5) עמדה על 3.86.

2

מאפשרת השוואה בין נתונים לאורך זמן או בין קבוצות.

כך למשל, לפני שינוי מדיניות האיחורים בבית הספר, 20% מהתלמידים היו מאחרים בכל יום, ואולם עם השינוי במדיניות, כמות האיחורים פחתה וכעת רק 3% מהתלמידים מאחרים בכל יום.

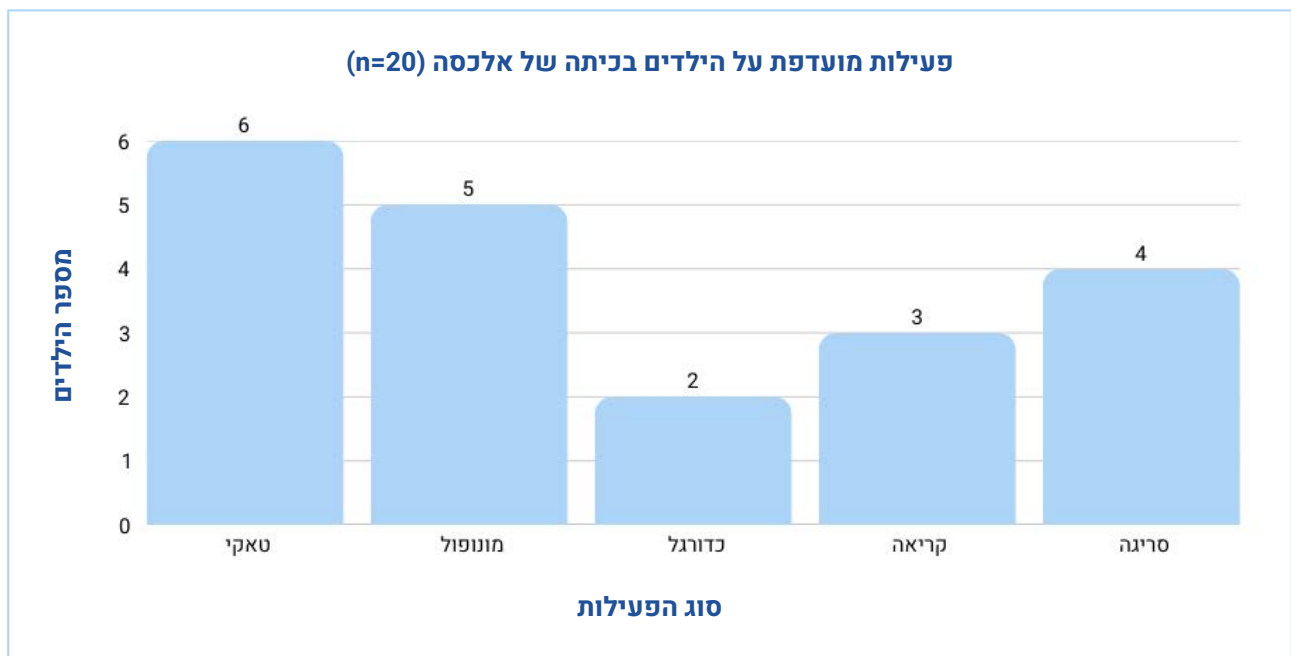
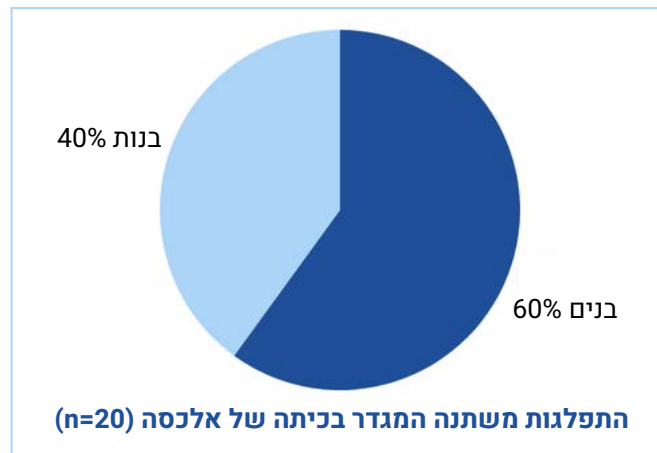
ואולם, חשוב לדעת שסטטיסטיקה תיאורית עשויה גם "להסתיר" מידע חשוב.

למשל, מה דעתנו על הנתונים האלה:

- ממוצע הציונים במתמטיקה בכיתה ה-2 הוא 75.
- ממוצע הציונים במתמטיקה בכיתה ה-2 הוא 75 ובכיתה ה-1 הוא 98.
- בכיתה ה-2, בחלוקה לפי מגדר, אנו רואים כי ממוצע הציונים של הבנות הוא 92 וממוצע הציונים של הבנים הוא 60.

טבלאות וגרפים מסייעים לנו בארגון ובהצגה של התפלגויות המשתנים באופן מרוכז ומפורט בו זמנית.

הנה כמה דוגמאות:



רמת השכלת האם	שכיחות
ללא השכלה תיכונית	2
פחות מתואר ראשון	3
תואר ראשון	9
תואר שני	5
דוקטורט ומעלה	1
סה"כ	20