

# מבוא לסטטיסטיקה תיאורית והסקית



$$\{\sqrt{x}\}^2$$



## תוכן העניינים

|           |                                                            |     |
|-----------|------------------------------------------------------------|-----|
| 1.        | סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה              | 1   |
| 6.        | סטטיסטיקה תיאורית- הצגה של נתונים                          | 6   |
| 17.       | סטטיסטיקה תיאורית- סכימה                                   | 17  |
| 21.       | סטטיסטיקה תיאורית - מדדי מיקום מרכזי                       | 21  |
| 28.       | סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן | 28  |
| 31.       | סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני           | 31  |
| 35.       | סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן                | 35  |
| 37.       | סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה                     | 37  |
| 39.       | סטטיסטיקה תיאורית - טרנספורמציה לינארית                    | 39  |
| 42.       | סטטיסטיקה תיאורית-מקדם ההשתנות                             | 42  |
| 44.       | סטטיסטיקה תיאורית- תרשים קופסא                             | 44  |
| 46.       | סטטיסטיקה תיאורית שאלות אמריקאיות                          | 46  |
| 53.       | סטטיסטיקה תיאורית - שאלות מסכמות                           | 53  |
| 57.       | יסודות ההסתברות                                            | 57  |
| (ללא ספר) | 15. התפלגויות רציפות מיוחדות - התפלגות נורמלית             |     |
| 61.       | 16. הסקה סטטיסטית - הקדמה                                  | 61  |
| 64.       | 17. התפלגות הדגימה ומשפט הגבול המרכזי                      | 64  |
| 70.       | 18. מושגי יסוד באמידה                                      | 70  |
| 75.       | 19. רווח סמך לתוחלת (ממוצע)                                | 75  |
| 85.       | 20. מבוא לבדיקת השערות על פרמטרים                          | 85  |
| 91.       | 21. בדיקת השערות על תוחלת (ממוצע)                          | 91  |
| 115.      | 22. רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים    | 115 |
| 119.      | 23. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים        | 119 |

## תוכן העניינים

|     |                                                  |     |
|-----|--------------------------------------------------|-----|
| 24. | בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים. | 121 |
| 25. | בדיקת השערות לתוחלת ההפרש במדגמים מזווגים.       | 129 |
| 26. | הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות.    | 133 |
| 27. | שאלות מסכמות על רווחי סמך.                       | 136 |
| 28. | שאלות מסכמות בבדיקת השערות.                      | 138 |
| 29. | מבחני חי בריבוע (מדד הקשר של קרמר).              | 153 |
| 30. | מקדם המתאם ( מדד קשר ) הלינארי ומובהקותו.        | 155 |
| 31. | מדדי קשר - בחירת מדד מתאים.                      | 175 |

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 1 - סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה

תוכן העניינים

1. סולמות מדידה.....1

## סטטיסטיקה תיאורית – סיווג משתנים וסולמות מדידה:

### רקע:

סטטיסטיקה תיאורית הוא ענף בו לומדים כיצד לאסוף נתונים, להציג אותם ולנתח אותם.

בסטטיסטיקה תיאורית אנו פונים לקבוצה מסוימת. באותה קבוצה אנו אוספים נתונים על הישויות באותה קבוצה.

משתנה – תכונה שיכולה לקבל מספר ערכים : דעה פוליטית, מקום מגורים, גובה של אדם וכדומה.

כל ישות בקבוצה שאנו צופים בה ואוספים לגביה נתונים נקראת תצפית.

הנתונים שנאספים בדרך כלל מרוכזים במסד נתונים. במסד הנתונים כל שורה היא תצפית וכל עמודה מייצגת משתנה.

### דוגמה (פתרון בהקלטה):

למחלקת טראומה הגיעו 5 פצועים מתאונה שקרתה בכביש החוף. אספו נתונים לגבי אותם פצועים, הנתונים מרוכזים בטבלה הבאה :

| מין  | גיל  | מצב הפצוע | דופק |
|------|------|-----------|------|
| גבר  | 26.6 | אנוש      | 40   |
| גבר  | 24.5 | קשה       | 38   |
| אישה | 32.1 | קשה       | 50   |
| גבר  | 34.9 | בינוני    | 65   |
| אישה | 23.1 | קל        | 89   |

ענו על השאלות הבאות :

הגדירו את הקבוצה שבדוגמה.

כמה תצפיות בקבוצה?

כמה משתנים בקבוצה?

כמה ערכים יש למשתנה "מין"?

את המשתנים במחקר אנו מסווגים ל-"סולמות מדידה" הדבר חשוב לדרך שבה ננתח את הנתונים בהמשך.

### מיון משתנים לפי סולמות המדידה:

1. סולם שמי (nominal) – משתנה שלערכיו יש משמעות רק מבחינת הזהות ואין עניין של יותר או פחות לערכים שלהם לדוגמה: צבע מועדף.  
משתנה דיכוטומי (הינו מסולם שמי) אותם משתנים שיש להם רק שני ערכים אפשריים. למשל: אזרחות ישראלית: יש או אין.
2. סולם סדר (ordinal) – כאשר לערכים של המשתנה בנוסף לשם ישנה גם משמעות לסדר מי יותר או מי פחות אבל אין משמעות לגודל.  
משתנה מסולם סדר יכול לקבל ערכים מילוליים או מספריים.  
למשל, דרגה בצבא.
3. סולם כמותי (scale) – משתנה שחייב להיות מספרי, לערכים שלו בנוסף לשם ולסדר בניהם יש משמעות לערך המספרי. משתנה כמותי הוא משתנה שניתן בדרך כלל לספור או למדוד על ידי מכשיר מדידה. למשל, מספר המחשבים בדירה, שטח הדירה במ"ר.

את המשתנה הכמותי אנו מסווגים לשני סוגים:

משתנה בדיד:

משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים כמו: מספר המחשבים בדירה.

משתנה רציף:

משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים, הערכים מתקבלים ברצף וללא קפיצות של ערכים. למשל, שטח הדירה במ"ר.

### דוגמה (פתרון בהקלטה):

למחלקת טראומה הגיעו 5 פצועים מתאונה שקרתה בכביש החוף.  
אספו נתונים לגבי אותם פצועים, הנתונים מרוכזים בטבלה הבאה:

| מין  | גיל  | מצב הפצוע | דופק |
|------|------|-----------|------|
| גבר  | 26.6 | אנוש      | 40   |
| גבר  | 24.5 | קשה       | 38   |
| אישה | 32.1 | קשה       | 50   |
| גבר  | 34.9 | בינוני    | 65   |
| אישה | 23.1 | קל        | 89   |

סווגו כל משתנה במסד הנתונים: שמי, סדר, כמותי רציף, כמותי בדיד.

## שאלות:

(1) לפניכם טבלה המסכמת נתונים לגבי סקר שנעשה היום:

| שם משפחה  | מצב משפחתי | מידת דתיות | מספר רכבים |
|-----------|------------|------------|------------|
| כהן       | רווק       | חילוני     | 0          |
| חדד       | נשוי       | חילוני     | 1          |
| לביא      | גרוש       | מסורתי     | 1          |
| פיינגולד  | אלמן       | חילוני     | 2          |
| אבו שוקרא | נשוי       | דתי        | 1          |
| בן חיים   | נשוי       | מסורתי     | 0          |
| רוטשילד   | רווק       | חילוני     | 0          |

א. כמה תצפיות בדוגמה זו?

ב. כמה משתנים בדוגמה זו?

ג. כמה ערכים ישנם ל-"מידת דתיות"?

ד. מהם הערכים האפשריים למשתנה "מצב משפחתי"?

(2) סניף מספר 543 של בנק "רווה" בדק ל-80 לקוחות את מספר הפעמים שכל לקוח נכנס לסניף הבנק במשך שבוע. התוצאות שהתקבלו הן: 50 אנשים נכנסו 0 פעמים לסניף, 20 אנשים נכנסו פעם אחת לסניף, 5 אנשים נכנסו פעמיים לסניף, 5 אנשים נכנסו יותר מפעמיים.

א. הגדירו את הקבוצה בדוגמה זו.

ב. כמה תצפיות בדוגמה זו?

ג. הגדירו את המשתנה בדוגמה זו. מהו סולם המדידה שלו?

(3) במחקר רפואי התעניינו לדעת כיצד מינון תרופת "קופקס" משפיע על מספר שעות השינה של אדם. במחקר השתתף אדם אחד בשם דני שבכל יום ניתן לו מינון שונה של התרופה. הטבלה שלהלן מתארת בכל יום את מינון התרופה במ"ג שקיבל האדם וכמו כן את מספר שעות השינה שלו באותו הלילה:

| מספר שעות שינה | מינון התרופה | מספר היום |
|----------------|--------------|-----------|
| 6              | 12           | 1         |
| 7              | 14           | 2         |
| 7.5            | 16           | 3         |
| 6.5            | 18           | 4         |
| 8              | 20           | 5         |

א. כמה תצפיות נאספו במחקר?

ב. סווגו את סולם המדידה של "מינון התרופה" ?

- (4) לפניכם רשימה של משתנים.
- צינו באיזה סולם מדידה מדובר (שמי, סדר, כמותי בדיד, כמותי רציף):
- א. גובה אדם בס"מ.
  - ב. מספר ילדים למשפחה.
  - ג. מידת חרדה לפני מבחן.
  - ד. שביעות רצון משירות לקוחות בסקלה מ-1 עד 7 (1 כלל לא מרוצה עד 7 מרוצה מאד).
  - ה. השכלה.
  - ו. מספר אוטובוס.
  - ז. מקום מגורים.
  - ח. מין (1=גבר ו-2=אישה).
  - ט. מידת נעליים.
- (5) לפניכם רשימה של משתנים כמותיים.
- ציין ליד כל משתנה אם הוא רציף או בדיד:
- א. שכר עובד ב-ש.
  - ב. ציון בחינת בגרות.
  - ג. תוצאה בהטלת קובייה.
  - ד. מהירות ריצה במטר לשנייה בתחרות 100 מטר.
  - ה. שיעור התמיכה בממשלה בעיר.
- (6) גברת לוי החליטה לדגום 25 ימים של נסיעה לעבודה, כאשר בדרך לעבודתה יש 3 צמתים מרומזרים.
- ב-9 ימים הגיעה גברת לוי לעבודה מבלי לעצור באף צומת.
  - ב-9 ימים נוספים היא הצליחה לעבור בשני רמזורים ירוקים.
  - ב-5 ימים נוספים היא הצליח לעבור רק בירוק אחד.
- בשאר הימים, היא לא עברה באף רמזור ירוק.
- מעוניינים לחקור את מספר הרמזורים האדומים בהם עצרה גברת לוי.
- א. מהו המשתנה הנחקר בדוגמה זו?
  - ב. מהם הערכים האפשריים של משתנה זה?
  - ג. כמה ערכים אפשריים יש למשתנה?
  - ד. מהו סולם המדידה של המשתנה?

## תשובות סופיות:

- (1) א.  $n = 7$ .  
ג. 3.
- (2) א. לקוחות סניף 543 של בנק "רווה".  
ג.  $X =$  מספר הפעמים בשבוע שלקוח נכנס לסניף. כמותי בדיד.  
ב.  $n = 80$ .
- (3) א.  $n = 5$ .  
ב. כמותי רציף.
- (4) א. כמותי רציף.  
ג. אין מספיק נתונים.  
ה. אין מספיק נתונים.  
ז. שמי.  
ט. סדר.
- (5) א. רציף.  
ג. בדיד.  
ה. רציף.  
ב. בדיד.  
ד. רציף.
- (6) א. מספר הרמזורים בהם עוצרת גברת לוי ביום בדרך לעבודה.  
ב. 0, 1, 2, 3.  
ד. כמותי בדיד.  
ג. 4.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 2 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי ..... 6

## סטטיסטיקה תיאורית – הצגה של נתונים:

### רקע:

דרכים להצגת נתונים שנאספו :

### רשימה של תצפיות:

התצפית היא הערך שנצפה עבור ישות מסוימת בקבוצה. רושמים את התצפיות שהתקבלו כרשומה, יעיל שיש מספר מועט של תצפיות. ההצגה הזו רלבנטית לכל סוגי המשתנים. למשל, להלן מספר החדרים בבניין בן 5 דירות: 3, 4, 3, 5, 4.

### טבלת שכיחויות בדידה:

| שם המשתנה-<br>$X$ | שכיחות –<br>$f(x)$     | שכיחות יחסית<br>באחוזים   |
|-------------------|------------------------|---------------------------|
| $X_1$             | $f_1$                  | $\frac{f_1}{N} \cdot 100$ |
| $X_2$             | $f_2$                  | $\frac{f_2}{N} \cdot 100$ |
| $X_3$             | $f_3$                  | $\frac{f_3}{N} \cdot 100$ |
| $\vdots$          | $\vdots$               | $\vdots$                  |
| $X_k$             | $f_k$                  | $\frac{f_k}{N} \cdot 100$ |
| <b>סה"כ</b>       | $N = \sum_{i=1}^k f_i$ | 100%                      |

רושמים את התצפיות בטבלה שבה עמודה אחת מבטאת את ערכי המשתנה והשנייה את השכיחות. יעיל עבור משתנה איכותי וכמותי בדיד וכשיש מספר רב של תצפיות. לא יעיל למשתנה כמותי רציף.

**דוגמה:**

להלן התפלגות הציונים בכיתה מסוימת:

| הציון - $X$ | מספר התלמידים –<br>השכיחות - $f$ | $F_i$ | $\frac{f_i}{n}$ |
|-------------|----------------------------------|-------|-----------------|
| 5           | 2                                | 2     | $0.08=2/25$     |
| 6           | 4                                | 6     | $0.16=4/25$     |
| 7           | 8                                | 14    | $0.32=8/25$     |
| 8           | 5                                | 19    | $0.2=5/25$      |
| 9           | 4                                | 23    | $0.16=4/25$     |
| 10          | 2                                | 25    | $0.08=2/25$     |

שכיחות מצטברת – צבירה של השכיחויות.

השכיחויות  $F_i$  – השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי:

 $\frac{f_i}{n}$  – איזה חלק מהתצפיות בקבוצה שוות לערך.**טבלת שכיחויות במחלקות:**

משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד  
וטבלת שכיחויות תהיה ארוכה מידי.

**דוגמה:**

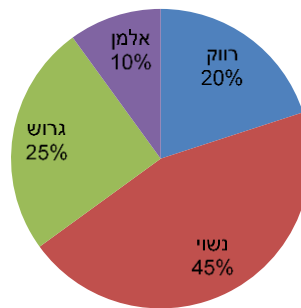
נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, בדקות.  
להלן ההתפלגות שהתקבלה:

| מספר הילדים | זמן בדקות |
|-------------|-----------|
| 20          | 0.5-3.5   |
| 18          | 3.5-9.5   |
| 14          | 9.5-19.5  |
| 8           | 19.5-29.5 |

### דיאגרמת עוגה:

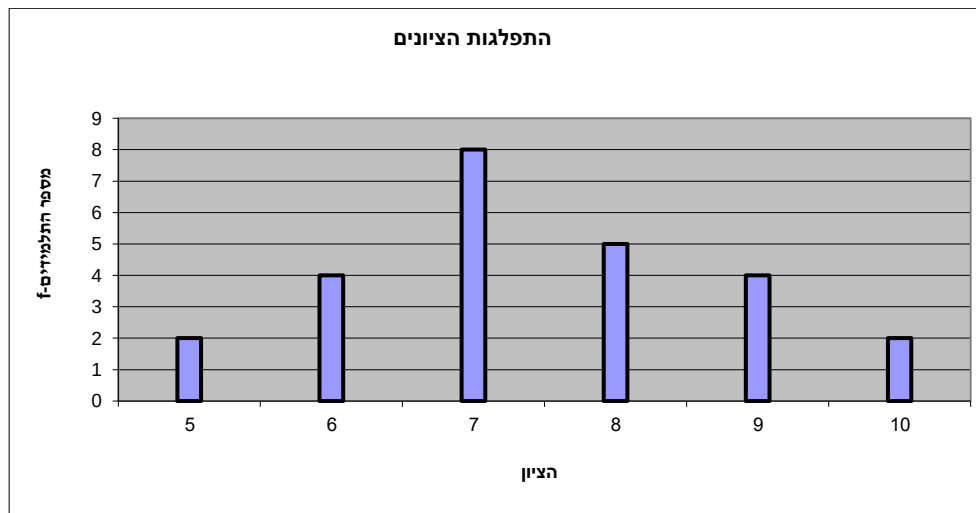
זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח", שהוא פרופורציונלי לשכיחות היחסית של ערך המשתנה בנתונים.

### התפלגות המצב המשפחתי



### דיאגרמת מקלות:

הציר האופקי הוא הציר של המשתנה והציר האנכי של השכיחות, כך שהגובה של המקל מעיד על השכיחות. רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף, וכן בסולמות מדידה עבור משתנה מסולם סדר.

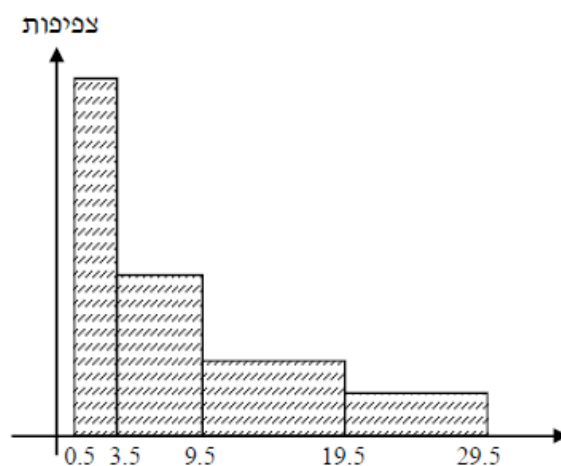


### היסטוגרמה:

היסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות, והיא רלוונטית למשתנה כמותי רציף.

בהיסטוגרמה הציר האופקי הוא הציר של המשתנה והציר האנכי הוא הציר של הצפיפות. הצפיפות מחושבת בכל מחלקה על ידי חלוקת השכיחות ברוחב של כל המחלקה, והיא נותנת את מספר התצפיות הממוצע בכל מחלקה ליחידה. אם המחלקות הן שוות ברוחב, ניתן לשרטט את ההיסטוגרמה לפי השכיחות ואין צורך בצפיפות.

| צפיפות | מצטברת | שכיחות | אמצע | רוחב | X           |
|--------|--------|--------|------|------|-------------|
| 6.6667 | 20     | 20     | 2    | 3    | 0.5 - 3.5   |
| 3      | 38     | 18     | 6.5  | 6    | 3.5 - 9.5   |
| 1.4    | 52     | 14     | 14.5 | 10   | 9.5 - 19.5  |
| 0.8    | 60     | 8      | 24.5 | 10   | 19.5 - 29.5 |



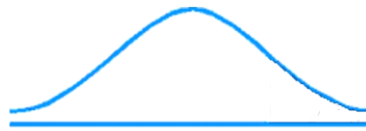
### פוליגון – מצולעון:

אם נחבר את אמצע קצה כל מלבן בקווים ישרים. נותן מראה חזותי לצורה של התפלגות המשתנה.

### צורות התפלגות נפוצות:

#### התפלגות סימטרית פעמונית

רוב התצפיות במרכז, וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. לדוגמה, ציוני IQ.

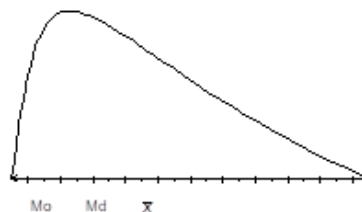


ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

#### התפלגות אסימטרית ימנית (חיובית)

רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. לדוגמה, שכר במשק.

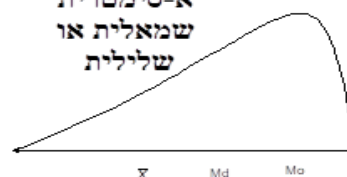
התפלגות א-סימטרית  
ימנית או חיובית



#### התפלגות אסימטרית שמאלית (שלילית)

רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. לדוגמה, אורך חיים.

התפלגות  
א-סימטרית  
שמאלית או  
שלילית



## שאלות:

- (1) בסקר צפייה בטלוויזיה התקבלו התוצאות הבאות : 25 צפו בערוץ הראשון, 25 צפו בערוץ 10, 75 צפו בערוץ השני, 50 צפו באחד מערוצי הכבלים ו-25 לא צפו בטלוויזיה בזמן הסקר.
- א. רשמו את טבלת השכיחות ואת השכיחות היחסית.
- ב. תארו את הנתונים באופן גרפי.

- (2) להלן נתונים על התפלגות המקצוע המועדף של תלמידי שכבה ו' בבית הספר "מעוף":

| מספר התלמידים | המקצוע   |
|---------------|----------|
| 44            | מתמטיקה  |
| 20            | תנ"ך     |
| 12            | אנגלית   |
| 26            | היסטוריה |

- א. מהו המשתנה הנחקר?
- ב. מהי פרופורציית התלמידים שמעדיפים תנ"ך?

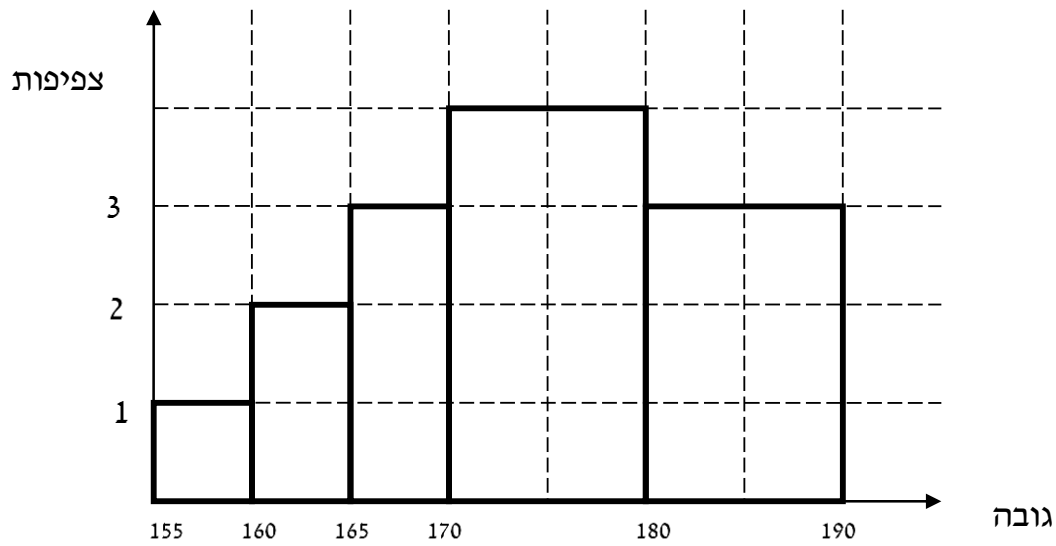
- (3) להלן התפלגות ההשכלה במקום עבודה מסוים:

| מספר העובדים | השכלה   |
|--------------|---------|
| 60           | נמוכה   |
| 120          | תיכונית |
| 20           | אקדמאית |

- א. מהו המשתנה הנחקר?
- מאיזה סולם הוא?
- ב. תארו את הנתונים באופן גרפי.

- (4) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
- 6, 5, 8, 7, 6, 7, 8, 6, 5, 4, 6, 10, 9, 8, 6, 7.
- א. מהו המשתנה? האם הוא בדיד או רציף?
- ב. תארו את הרשימה בטבלת שכיחויות.
- ג. הוסיפו שכיחויות יחסיות לטבלה.
- ד. תארו את הנתונים באופן גרפי.

5) להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת:



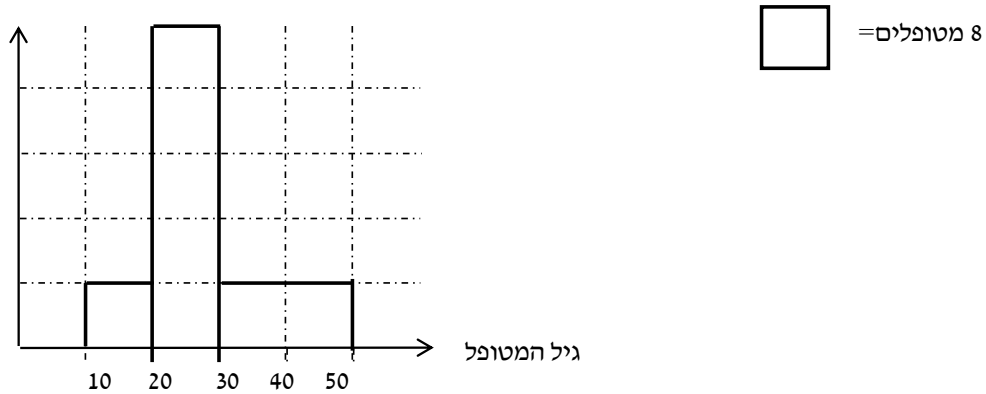
- מהו המשתנה הנחקר? האם הוא בדיד או רציף?
- תארו את הנתונים בטבלת שכיחויות במחלקות.
- הוסיפו שכיחות יחסית לטבלה.
- הוסיפו את הצפיפות של כל מחלקה לטבלה.
- מהי צורת ההתפלגות של הגבהים?

6) להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

| משקל  | מספר מקרים |
|-------|------------|
| 40-45 | 10         |
| 45-50 | 20         |
| 50-60 | 30         |
| 60-65 | 20         |
| 65-70 | 10         |

- תארו את ההתפלגות באופן גרפי.
- מה ניתן להגיד על צורת ההתפלגות?

7) להלן גיל המטופלים של ד"ר שוורץ בשנים :  
קנה מידה :



- א. מה המשתנה הנחקר? האם הוא בדיד או רציף?
- ב. מהי הקבוצה הנחקרת?
- ג. תרגמו את ההסיטוגרמה לטבלת שכיחות.
- ד. מהי הפרופורציה של המטופלים של ד"ר שוורץ בגילאים 20-30?

## תשובות סופיות:

1) א. להלן טבלה: ב. עיין גרף מלא בסרטון הוידאו.

| $x$     | $f(x)$ | $\frac{f(x)}{n}$ | %     |
|---------|--------|------------------|-------|
| ערוץ 1  | 25     | $\frac{25}{200}$ | 12.5% |
| ערוץ 10 | 25     | $\frac{25}{200}$ | 12.5% |
| ערוץ 2  | 75     | $\frac{75}{200}$ | 37.5% |
| כבלים   | 50     | $\frac{50}{200}$ | 25%   |
| לא צפו  | 25     | $\frac{25}{200}$ | 12.5% |
| סה"כ    | 200    | 1                | 100%  |

- 2) א. מקצוע מועדף. ב. 19.6%.
- 3) א. משתנה נחקר: השכלה, סוג: סדר. ב. עיין גרף מלא בסרטון הוידאו.

- (4) א. המשתנה : ציון, משתנה בדיד.  
ד. עיין גרף מלא בסרטון הוידאו.

ב+ג. להלן טבלה :

| $x$  | $f(x)$ | $\frac{f(x)}{n}$ | %    |
|------|--------|------------------|------|
| 4    | 1      | $\frac{1}{20}$   | 5%   |
| 5    | 2      | $\frac{2}{20}$   | 10%  |
| 6    | 6      | $\frac{6}{20}$   | 30%  |
| 7    | 4      | $\frac{4}{20}$   | 20%  |
| 8    | 4      | $\frac{4}{20}$   | 20%  |
| 9    | 2      | $\frac{2}{20}$   | 10%  |
| 10   | 1      | $\frac{1}{20}$   | 5%   |
| סה"כ | 20     | 20               | 100% |

- (5) א. גובה בס"מ, רציף.

ב+ג+ד. להלן טבלה : ה. אסימטרית.

| $x$     | $f(x)$ | $\frac{f(x)}{n}$ | %   | $d$ |
|---------|--------|------------------|-----|-----|
| 155-160 | 5      | $\frac{5}{100}$  | 5%  | 1   |
| 160-165 | 10     | $\frac{10}{100}$ | 10% | 2   |
| 165-170 | 15     | $\frac{15}{100}$ | 15% | 3   |
| 170-180 | 40     | $\frac{40}{100}$ | 40% | 4   |
| 180-190 | 30     | $\frac{30}{100}$ | 30% | 3   |

- 6) א. עיין גרף מלא בסרטון הוידאו. ב. סימטרית.
- 7) א. המשתנה: גיל בשנים, משתנה רציף. ב. המטופלים של ד"ר שוורץ.
- ד. להלן טבלה: ה. 62.5%.

| $f(x)$ | $x$   |
|--------|-------|
| 8      | 10-20 |
| 40     | 20-30 |
| 16     | 30-50 |

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 3 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי.....17

## סטטיסטיקה תיאורית – סכימה:

רקע:

בסטטיסטיקה ישנה צורת רישום מקובלת לסכום של תצפיות:  $\sum_{i=1}^n X_i$ .

נסביר את צורת הרישום על ידי הדוגמה הבאה:

| $i$ | $X_i$ |
|-----|-------|
| 1   | 5     |
| 2   | 0     |
| 3   | 1     |
| 4   | 3     |
| 5   | 2     |

(הסבר מלא מופיע בסרטונים באתר).

## שאלות:

- 1) בבניין 5 דירות. לכל דירה רשמו את מספר החדרים שיש בדירה ( $X$ ), ומספר הנפשות החיות בדירה ( $Y$ ) חשבו:

| מספר דירה | $X$ | $Y$ |
|-----------|-----|-----|
| 1         | 2   | 1   |
| 2         | 3   | 1   |
| 3         | 2   | 2   |
| 4         | 4   | 3   |
| 5         | 3   | 2   |

א.  $\sum_{i=1}^3 X_i$

ב.  $\sum_{i=1}^5 Y_i$

ג.  $\sum_{i=1}^4 X_i$

ד.  $\left( \sum_{i=1}^4 X_i \right)^2$

ה.  $\sum X_i$

ו.  $\sum X_i Y_i$

ז.  $\sum(X_i) \sum(Y_i)$

- (2) נתון לוח ערכי המשתנים  $X_i$  ו- $Y_i$ , כאשר:  $i = 1, 2, \dots, 6$ , ונתונים הקבועים:  
 $a = 2$ ,  $b = 5$ . חשבו את הנוסחאות הבאות:

|       |   |   |   |    |    |   |
|-------|---|---|---|----|----|---|
| $i$   | 1 | 2 | 3 | 4  | 5  | 6 |
| $X_i$ | 3 | 2 | 4 | -2 | 1  | 4 |
| $Y_i$ | 2 | 0 | 0 | 1  | -5 | 2 |

א.  $\sum_{i=1}^4 y_i$

ב.  $\sum_{i=1}^6 a$

ג.  $\sum_{i=1}^6 x_i y_i$

ד.  $\sum_{i=1}^6 (x_i + y_i)$

ה.  $\sum_{i=1}^6 x_i + a$

- (3) קבעו לכל זהות האם היא נכונה:

א.  $\sum_{i=1}^n bX_i = b \cdot \sum_{i=1}^n X_i$

ב.  $\sum_{i=1}^n a = a \cdot n$

ג.  $\left( \sum_{i=1}^n X_i \right)^2 = \sum_{i=1}^n X_i^2$

### תשובות סופיות:

- |              |           |              |         |
|--------------|-----------|--------------|---------|
| 1) א. 7.     | ב. 9.     | ג. 11.       | ד. 121. |
| ה. 14.       | ו. 27.    | ז. 126.      |         |
| 2) א. 3.     | ב. 12.    | ג. 7.        | ד. 12.  |
| ה. 14.       |           |              |         |
| 3) א. נכונה. | ב. נכונה. | ג. לא נכונה. |         |

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי

תוכן העניינים

1. כללי..... 21

## סטטיסטיקה תיאורית – מדדי מיקום מרכזי:

### רקע:

המטרה במדדי המיקום המרכזי היא למדוד את מרכז ההתפלגות של התצפיות.

### השכיח – Mode:

השכיח הוא הערך הנפוץ ביותר בהתפלגות.

### ברשימה

הערך החוזר על עצמו הכי הרבה פעמים: 7, 9, 4, 8, 4, 10, 6.

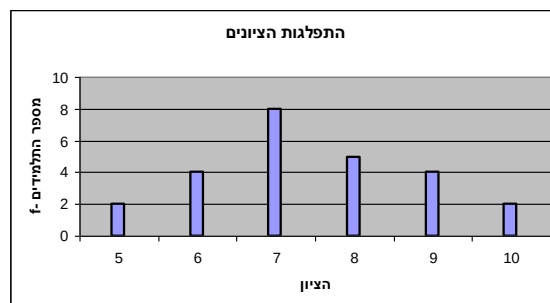
### בטבלת שכיחויות בדידה

הערך שהשכיחות שלו היא הגבוהה ביותר.

| $f(x)$ | # תוכניות החיסכון |
|--------|-------------------|
| 100    | 0                 |
| 75     | 1                 |
| 25     | 2                 |
| 25     | 3                 |
| 25     | 4                 |

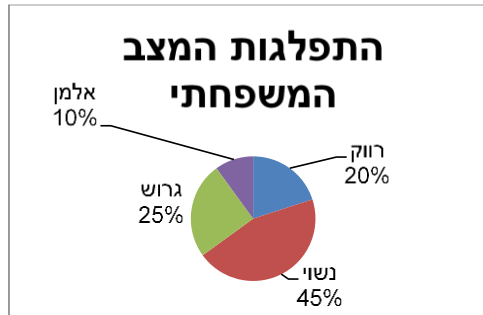
### בדיאגרמת מקלות

שיעור ה- $X$  של המקל הגבוה ביותר.



### בעוגה

הערך של הפלח הגדול ביותר.



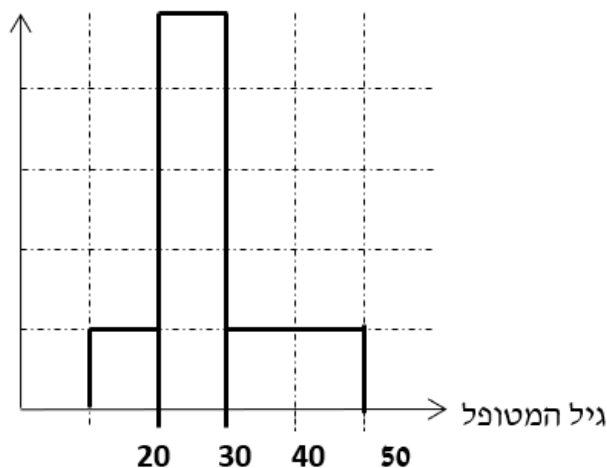
### בטבלת שכיחויות במחלקות

אמצע המחלקה עם הצפיפות הגבוהה ביותר.  
לדוגמה, התפלגות הציונים בכיתה :

| $f(x)$ | $X$    |
|--------|--------|
| 20     | 0-60   |
| 10     | 60-70  |
| 18     | 70-80  |
| 15     | 80-90  |
| 15     | 90-100 |

### בהיסטוגרמה

שיעור ה-  $X$  של אמצע המחלקה הגבוהה ביותר.  
לדוגמה, גיל המטופלים של ד"ר שוורץ בשנים :



= 8

### כללי

יתכן שלהתפלגות יותר משכיח אחד.  
השכיח הוא מדד הרלבנטי לכל סוגי המשתנים.

### אמצע תחום (טווח) – Midrange:

הממוצע בין התצפית הגבוהה ביותר לתצפית הנמוכה ביותר:

$$MR = \frac{X_{\min} + X_{\max}}{2}$$

### החציון – Median:

החציון הוא ערך שמחצית מהתצפיות קטנות או שוות לו ומחצית מהתצפיות גדולות או שוות לו.

### ברשימה

נסדר את התצפיות בסדר עולה.

אם יש מספר אי זוגי של איברים, מקומו של החציון יהיה התצפית שמיקומה:  $\frac{n+1}{2}$ .

אם יש מספר זוגי של איברים – החציון הוא ממוצע של האיבר ה- $\frac{n}{2}$ ,

והאיבר ה- $\frac{n}{2} + 1$ , כלומר שיש מספר אי-זוגי של תצפיות החציון יהיה:  $md = X_{\frac{n+1}{2}}$ ,

וכשיש מספר זוגי של תצפיות החציון יהיה:  $md = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$ .

### בטבלת שכיחויות בדידה

נעשה תהליך דומה אך נעזר בשכיחות המצטברת.

### דיאגרמת מקלות

נמיר לטבלת שכיחויות בדידה במטרה למצוא את החציון.

### בטבלת שכיחויות במחלקות

שלב א: נמצא את המחלקה החציונית שמיקומה יהיה  $\frac{n}{2}$ .

שלב ב: נציב בנוסחה הבאה:  $Md = L_0 + \frac{\frac{n}{2} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$

$F(x_{m-1})$  - שכיחות מצטברת של מחלקה אחת לפני המחלקה החציונית.  
 $f(x_m)$  - השכיחות של המחלקה החציונית.

$L_0$  - גבול התחתון של המחלקה.

$L_1$  - גבול העליון של המחלקה.

### היסטוגרמה

החציון הוא הערך על ציר ה-  $X$  שמחלק את ההיסטוגרמה לשני חלקים שווים בשטח.

### כללי

החציון אינו רלבנטי למשתנה מסולם שמי ולא רלבנטי למשתנה איכותי.

### הממוצע – Average:

הממוצע הוא מרכז הכובד של ההתפלגות.

### ברשימה

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

### בטבלת שכיחויות

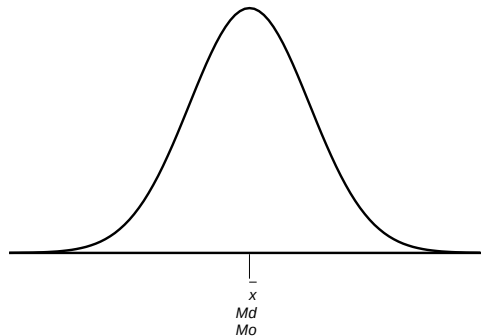
$$\bar{X} = \frac{\sum x \cdot f}{n}$$

**כללי**

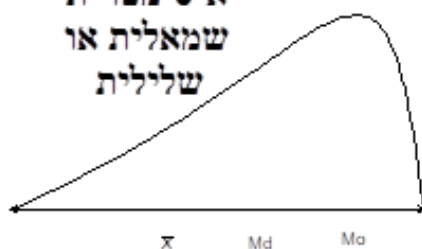
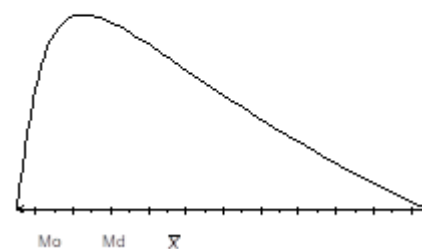
הממוצע רלבנטי רק למשתנה כמותי.

**מדדי המיקום המרכזי בהתפלגויות המיוחדות:**

בהתפלגות סימטרית פעמונית כל מדדי המרכז שווים זה לזה:

**התפלגות סימטרית**

בהתפלגות סימטרית השכיח לא חייב להיות במרכז:

**התפלגות U**
**התפלגות  
א-סימטרית  
שמאלית או  
שלילית**

**התפלגות א-סימטרית  
ימנית או חיובית**


## שאלות:

- (1) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
- 6, 5, 8, 7, 6, 8, 6, 7, 8, 5, 6, 4, 10, 9, 8, 6, 7.
- חשבו את החציון, השכיח, והממוצע של הציונים.
- (2) בדקו את מספר החדרים לדירה בבניין בן 5 דירות והתקבל ממוצע 3.8. לגבי 4 דירות נמצא מספר חדרים: 4, 3, 4, 5.
- א. כמה חדרים יש בדירה החמישית?
- ב. מהו השכיח ומהו החציון?
- (3) להלן התפלגות מספר מקלטי הטלויזיה שנספרו עבור כל משפחה ביישוב מסוים:

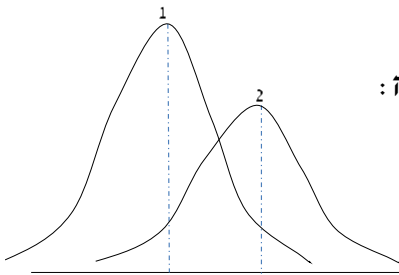
| מספר משפחות | מספר מקלטים |
|-------------|-------------|
| 22          | 0           |
| 28          | 1           |
| 18          | 2           |
| 22          | 3           |
| 10          | 4           |

- א. חשבו את הממוצע, החציון והשכיח של ההתפלגות.
- ב. הסבירו ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא כולן) שלא היה להם עד היום טלויזיה היו רוכשים מקלט אחד.

- (4) להלן התפלגות מספר המכוניות למשפחה ביישוב "הגורן":

| מספר מכוניות למשפחה | 1  | 2   | 3   | 4   | 5  |
|---------------------|----|-----|-----|-----|----|
| שכיחות              | 65 | 150 | 220 | 140 | 55 |

- א. כמה משפחות יש ביישוב?
- ב. מה אחוז המשפחות ביישוב עם לכל היותר 2 מכוניות?
- ג. חשבו את הממוצע, החציון והשכיח.
- הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!



- (5) מורה לימד 2 כיתות, הוא תיאר באותה מערכת צירים את התפלגות הציונים בכל כיתה. בחרו בתשובה הנכונה:
- בכיתה 1 השכיח גבוה יותר מכיתה 2.
  - בכיתה 2 השכיח גבוה יותר מכיתה 1.
  - בשתי הכיתות אותו שכיח.
  - לא ניתן לדעת באיזו כיתה השכיח גדול יותר.

### תשובות סופיות:

- חציון : 7, שכיח : 6, ממוצע : 6.9
- א. 3. ב. שכיח : 3.4, חציון : 4.
- א. ממוצע : 1.7, חציון : 1.5, שכיח : 1. ב. הממוצע יגדל ויתר המדדים לא ישתנו.
- א. 630. ב. 34.13%. ג. שכיח וחציון : 3, ממוצע : 2.952.
- ב'.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 5 - סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן

תוכן העניינים

1. כללי..... 28

## סטטיסטיקה תיאורית – מדדי פיזור – הטווח, השונות וסטיית התקן:

### רקע:

**המטרה:** למדוד את הפיזור של הנתונים, כלומר כמה הם רחוקים זה מזה ושונים זה מזה.

**הטווח / תחום (RANGE):**

ההפרש בין התצפית הגבוהה ביותר לנמוכה ביותר:  $R = X_{\max} - X_{\min}$ .

### שונות וסטיית תקן:

שונות היא ממוצע ריבועי של הסטיות מהממוצע וסטיית התקן היא שורש של השונות.

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 \quad \text{עבור סדרת נתונים}$$

### דוגמאות:

(1) נחשב את השונות של סדרת המספרים הבאה: 5, 4, 9.

$$S_x^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{\sum x^2 \cdot f}{n} - \bar{x}^2 \quad \text{עבור טבלת שכיחויות}$$

(2) להלן התפלגות הציונים בכיתה מסוימת בה ממוצע הציונים הוא 7.44.

| הציון<br>$X$ | השכיחות<br>$F$ | $x^2 \cdot F$ |
|--------------|----------------|---------------|
| 5            | 2              | 50            |
| 6            | 4              | 144           |
| 7            | 8              | 392           |
| 8            | 5              | 320           |
| 9            | 4              | 324           |
| 10           | 2              | 200           |
| סה"כ         |                | 1430          |

$$S_x^2 = \frac{\sum x^2 f(x)}{n} - \bar{x}^2 = \frac{1430}{25} - 7.44^2 = 1.8464$$

$$S = \sqrt{S_x^2} = \sqrt{1.8464} = 1.3588$$

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

## שאלות:

- (1) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:  
 6, 5, 8, 7, 6, 8, 7, 6, 5, 8, 4, 6, 10, 9, 8, 6, 7.  
 חשבו את השונות, סטיית התקן והטווח של הציונים.

- (2) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

| מספר מכוניות למשפחה | 1  | 2   | 3   | 4   | 5  |
|---------------------|----|-----|-----|-----|----|
| שכיחות              | 65 | 150 | 220 | 140 | 55 |

- א. חשבו סטיית התקן.  
 ב. חשבו את הטווח של הנתונים.  
 הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם.
- (3) בחברה העוסקת בטלמרקטינג בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4 שנים וסטיית התקן היא שנתיים.  
 א. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים עם וותק של 4 שנים להתפלגות?  
 ב. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים אשר אחד עם וותק של 0 שנים והשני עם וותק של 8 שנים להתפלגות?
- (4) נתונה רשימה של 5 תצפיות, אך רק עבור 4 מהן נרשמו הסטיות שלהן מהממוצע: 2, 3, 2, 1-. חשבו את השונות של חמש התצפיות.
- (5) בשכונה בדקו בכל דירה את מספר החדרים לדירה. בשכונה 200 דירות.

| מספר חדרים | פרופורציה |
|------------|-----------|
| 1          | 0.1       |
| 2          | 0.2       |
| 3          | 0.4       |
| 4          | 0.15      |
| 5          |           |

- א. מה הממוצע של מספר החדרים לשכונה בדירה?  
 ב. חשבו את סטיית התקן של מספר החדרים לדירה.  
 ג. חלק מבעלי הדירות בנות 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע(יקטין, יגדל, לא ישנה) על כל מדד שחישבתם בסעיפים הקודמים

### תשובות סופיות:

- (1) שונות : 2.19, סטיית תקן : 1.48, טווח : 6.
- (2) א. סטיית תקן : 1.106. ב. טווח : 4.
- (3) א. ממוצע לא ישתנה, סטיית התקן תקטן.  
 ב. ממוצע לא ישתנה, סטיית התקן תגדל.
- (4) 10.8.
- (5) א. 3.05. ב. 1.16. ג. ממוצע: יקטן, סטיית התקן: תגדל.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני

תוכן העניינים

1. טווח בינרבעוני..... 31

## סטטיסטיקה תיאורית – מדדי פיזור – טווח בין רבעוני:

### רקע:

הטווח הבין-רבעוני (יש הקוראים לו התחום הבין-רבעוני) נותן את הטווח בין הרבעונים בו נמצאים 50% מהתצפיות המרכזיות. הרעיון ליצור מדד פיזורי שלא רגיש לתצפיות החריגות ביותר. כדי לחשב את הטווח הבין-רבעוני יש למצוא את הרבעון התחתון והעליון של התפלגות התצפיות.

רבעון תחתון – ערך שמחלק את ההתפלגות לשניים.  
25% מהמקרים נמוכים ממנו או שווים לו ו 75% מהמקרים גבוהים או שווים לו.  
סימון:  $Q_1$ .

רבעון עליון – ערך שמחלק את ההתפלגות לשניים.  
75% מהמקרים נמוכים ממנו או שווים לו ו 25% מהמקרים גבוהים או שווים לו.  
סימון:  $Q_3$ .

הטווח הבין רבעוני הוא הפער בין שני הרבעונים:  $IQR = Q_3 - Q_1$ .

### שלבים במציאת טווח בין-רבעוני בטבלת שכיחויות:

שלב א: נמצא את הרבעון התחתון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 25%.

שלב ב: נמצא את הרבעון העליון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 75%.

שלב ג: נמצא את הטווח הבין-רבעוני: נחסר את הרבעונים:  $IQR = Q_3 - Q_1$ .

### דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו. מהו הטווח הבין-רבעוני של מספר תוכניות החיסכון בסניף?

| # תוכניות<br>החיסכון | $f(x)$ | שכיחות<br>מצטברת | שכיחות יחסית<br>מצטברת |
|----------------------|--------|------------------|------------------------|
| 0                    | 100    |                  |                        |
| 1                    | 75     |                  |                        |
| 2                    | 25     |                  |                        |
| 3                    | 25     |                  |                        |
| 4                    | 25     |                  |                        |

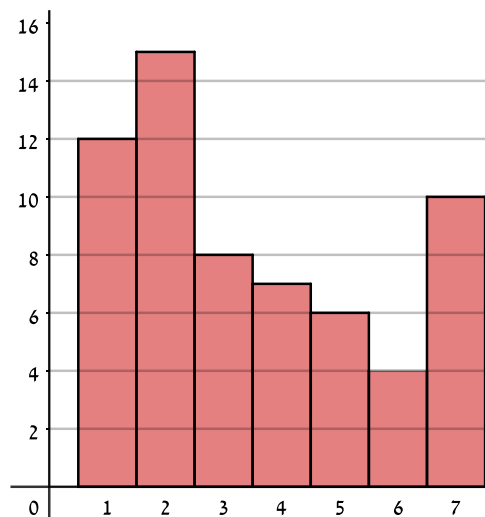
## שאלות:

1) להלן התפלגות מספר המכוניות למשפחה ב"הגורן":

| מספר מכוניות<br>למשפחה | 1  | 2   | 3   | 4   | 5  |
|------------------------|----|-----|-----|-----|----|
| שכיחות                 | 65 | 150 | 220 | 140 | 55 |

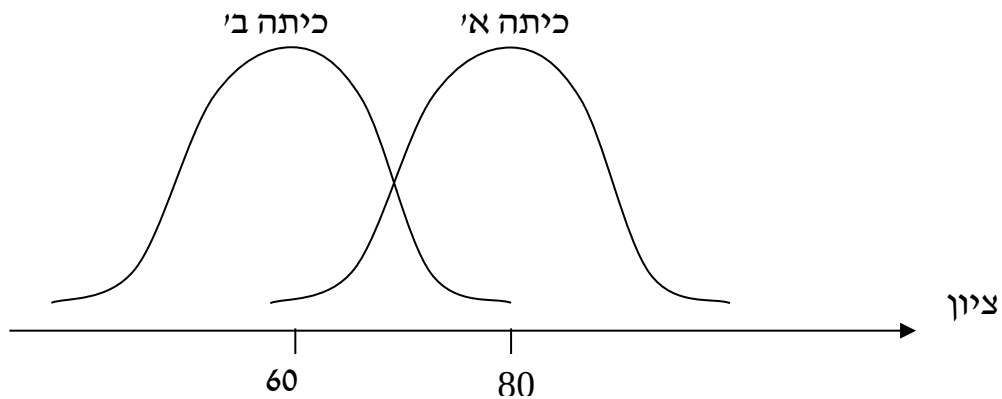
מהו הטווח הבין-רבעוני של מספר המכוניות למשפחה ב"הגורן"?

2) בסקר שנעשה בדקו את מספר ימי המחלה השנתיים של מורים בארץ.



- מה מייצגים הערכים בציר האופקי?
- מהו הטווח הבין-רבעוני של מספר ימי המחלה של המורים?
- אם נוסף 25 מורים אשר הצהירו שמספר ימי המחלה השנתיים שלהם הוא 4 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.
- אם מסתבר שחלק מהמורים בסקר הצהירו שהם חלו 7 ימים בשנה אבל בפועל הם חלו 8 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.

3) לפניך שתי עקומות המתארות את התפלגות הציונים בכל כיתה. באיזו כיתה הטווח הבין-רבעוני גדול יותר?



- א. כיתה א.
- ב. כיתה ב'.
- ג. לשתיהן אותו טווח בין-רבעוני.
- ד. לא ניתן לדעת, אין מספיק נתונים.

4) הוספת גודל קבוע לכל תצפיות סדרת נתונים :

- א. תגדיל את הטווח הבין-רבעוני.
- ב. תקטין את הטווח הבין-רבעוני.
- ג. לא תשנה הטווח הבין-רבעוני.
- ד. לא ניתן לדעת מה יקרה לטווח הבין-רבעוני.

5) חושב הטווח הבין-רבעוני עבור התפלגות מסוימת והתקבלה התוצאה אפס. לכן :

- א. לפחות 50% מהתצפיות זהות.
- ב. סטיית התקן היא אפס.
- ג. ההתפלגות היא סימטרית.
- ד. מצב זה כלל לא יתכן.

- (6) סניף מספר 543 של בנק "רווה" בדק ל-80 לקוחות את מספר הפעמים שכל לקוח נכנס לסניף הבנק במשך שבוע. התוצאות שהתקבלו הן:
- 50 אנשים נכנסו 0 פעמים לסניף.
  - 20 אנשים נכנסו פעם אחת לסניף.
  - 5 אנשים נכנסו פעמיים לסניף.
  - 5 אנשים נכנסו יותר מפעמיים.
- מהו הטווח הבין-רבעוני?
- א. 60.
  - ב. 2.
  - ג. 50.
  - ד. 1.

- (7) התפלגות הציונים במבחן ווקסלר היא סימטרית לכן:
- א. טווח הציונים הוא אפס.
  - ב. הטווח הבין רבעוני של הציונים אפס.
  - ג. סעיפים א ו-ב הם נכונים.
  - ד. אף סעיף אינו נכון.

### תשובות סופיות:

- (1) 2.
- (2) א. מספר ימי המחלה השנתיים. ב. 3. ג. יקטן. ד. לא ישתנה.
- (3) ג'.
- (4) ג'.
- (5) א'.
- (6) ד'.
- (7) ד'.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 7 - סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן

תוכן העניינים

1. כללי..... 35

## סטטיסטיקה תיאורית – מדדי מיקום יחסי – ציון תקן:

### רקע:

המטרה למדוד איך תצפית ממוקמת ביחס לשאר התצפיות בהתפלגות.

### ציון תקן:

הנוסחה לציון תקן של תצפית היא:  $Z = \frac{X - \bar{X}}{S}$ .

- ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע. כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע:
- ציון תקן חיובי אומר שהתצפית מעל הממוצע.
  - ציון תקן שלילי אומר שהתצפית מתחת לממוצע.
  - ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

### דוגמה (פתרון בהקלטה):

במקום עבודה מסוים, ממוצע המשכורות הוא 8 אלף ₪, עם סטית תקן של אלפיים ₪. באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים, עם סטית תקן של 1.5 שנים. ערן מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערן יותר, באופן יחסי, משכיל או משתכר?

## שאלות:

- (1) תלמידי כיתה ח' ניגשו למבחן בלשון ולמבחן במתמטיקה. להלן התוצאות שהתקבלו:

| המקצוע  | ממוצע | סטיית תקן |
|---------|-------|-----------|
| לשון    | 74    | 12        |
| מתמטיקה | 80    | 16        |

עודד קיבל: 68 בלשון ו-70 במתמטיקה.

- א. באיזה מקצוע עודד טוב יותר באופן יחסי לשכבה שלו?  
 ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שיהיה שקול לציונו בלשון?

- (2) במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום. להלן טבלה המסכמת את המידע שנאסף על שני המשתנים:

| ממוצע     | תפוקה | מספר פועלים |
|-----------|-------|-------------|
| ממוצע     | 48    | 15          |
| סטיית תקן | 10    | 2           |

- באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים.  
 מה יותר חריג באותו היום, יחסית לשאר הימים שנבדקו: נתוני התפוקה או כמות הפועלים?  
 א. התפוקה.  
 ב. כמות הפועלים.  
 ג. חריגים באותה מידה.  
 ד. חסרים נתונים כדי לדעת זאת.

- (3) הגובה הממוצע של המתגייסים לצבא הוא 175 סנטימטר עם סטיית תקן של 10 סנטימטר. המשקל הממוצע הוא 66 ק"ג עם סטיית תקן של 8 ק"ג. ערך התגייס כשגובהו 180 ס"מ ומשקלו 59 ק"ג.  
 א. במה ערך חריג יותר ביחס לשאר המתגייסים, גובהו או משקלו?  
 ב. כמה ערך אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

## תשובות סופיות:

- (1) א. לשון. ב. 72.  
 (2) ב'.  
 (3) א. משקל. ב. 70.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 8 - סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה

תוכן העניינים

1. כללי ..... 37

## סטטיסטיקה תיאורית – מדדי מיקום יחסי – אחוזונים בטבלה בדידה:

### רקע:

האחוזון (המאון) ה- $p$  הוא הערך בנתונים המחלק את הנתונים בצורה כזאת, שעד אליו (כולל) יש  $p\%$  מהנתונים. מסמנים את האחוזון ה- $p$  ב- $X_p$ .

### חישוב האחוזון מתוך נתונים בטבלת שכיחויות בדידה:

האחוזון הוא הערך שבו בפעם הראשונה השכיחות היחסית המצטברת (באחוזים) גדולה או שווה ל- $p\%$ .

### דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו:

| שכיחות<br>יחסית<br>מצטברת | שכיחות<br>מצטברת | $F(x)$ | # תוכניות<br>החיסכון |
|---------------------------|------------------|--------|----------------------|
|                           |                  | 100    | 0                    |
|                           |                  | 75     | 1                    |
|                           |                  | 25     | 2                    |
|                           |                  | 25     | 3                    |
|                           |                  | 25     | 4                    |

א. מצאו את האחוזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

## שאלות:

(1) להלן התפלגות של משתנה כלשהו:

| $F(x)$ | $X$ |
|--------|-----|
| 10     | 0   |
| 40     | 1   |
| 30     | 2   |
| 15     | 3   |
| 5      | 4   |

מצאו להתפלגות את:

- א. האחוזון ה-60.
- ב. המאון ה-40.
- ג. העשירון העליון.
- ד. הטווח בין הרבעונים.

(2) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

| מספר מכוניות למשפחה | 1  | 2   | 3   | 4   | 5  |
|---------------------|----|-----|-----|-----|----|
| שכיחות              | 65 | 150 | 220 | 140 | 55 |

חשבו את:

- א. העשירון התחתון.
- ב. האחוזון ה-30.
- ג. הערך ש-20% מהתצפית גדולות ממנו.
- ד. רבעון עליון.

## תשובות סופיות:

- (1) א. 2. ב. 1. ג. 3. ד. 1.
- (2) א. 1. ב. 2. ג. 4. ד. 4.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 9 - סטטיסטיקה תיאורית - טרנספורמציה לינארית

תוכן העניינים

1. כללי ..... 39

## סטטיסטיקה תיאורית – טרנספורמציה לינארית:

### רקע:

מצב שבו מבצעים שינוי מסוג הוספה (או החסרה) של קבוע, והכפלה (או חילוק) של קבוע, לכל התצפיות:  $y = a \cdot x + b$ . כך יושפעו המדדים השונים:

$$\begin{aligned} MR_y &= a \cdot MR + b \\ MO_y &= a \cdot MO + b \\ \bar{y} &= a \cdot \bar{x} + b \\ Md_y &= a \cdot Md_x + b \end{aligned} \quad \text{מדדי המרכז:}$$

$$\begin{aligned} R_y &= |a| R_x \\ S_y &= |a| S_x \\ S_y^2 &= a^2 S_x^2 \end{aligned} \quad \text{מדדי הפיזור:}$$

$$\begin{aligned} Y_p &= a \cdot X_p + b \\ Z_Y &= \frac{a}{|a|} Z_x \end{aligned} \quad \text{מדדי המיקום היחסי:}$$

### שלבי העבודה:

1. נזהה שמדובר בטרנספורמציה לינארית (שינוי קבוע לכל התצפיות).
2. נרשום את כלל הטרנספורמציה לפי נתוני השאלה.
3. נפשט את הכלל ונזהה את ערכי  $a$  ו- $b$ .
4. נציב בנוסחאות שלעיל בהתאם למדדים שנשאלים.

### דוגמה (פתרון בהקלטה):

השכר הממוצע של עובדים הינו 9000 ₪ וטווח 6000 ₪. חשבו את המדדים הללו לאחר שהעלו את כל המשכורות ב-10% ואחר כך קנסו אותם ב-100 ₪.

## שאלות:

- (1) עבור סדרת נתונים התקבל:  $\bar{x} = 80, S = 15, MO = 70$ . הוחלט להכפיל את כל התצפיות ב-4 ולהחסיר מהתוצאה 5. חשבו את המדדים הללו לאחר השינוי.
- (2) בחברה מסוימת השכר הממוצע הוא 40 ₪ לשעה עם סטיית תקן של 5 ₪ לשעה. הוחלט להעלות את כל המשכורות ב-10%, אך זה לא סיפק את העובדים ולכן הם קיבלו לאחר מכן תוספת של 2 ₪ לשעה. מה הממוצע ומהי השונות של השכר לשעה לאחר כל השינויים.
- (3) במבחן מסוים הציון החציוני היה 73, טווח הציונים היה 40 נקודות והעשירון העליון היה הציון 87. כיוון שהציונים בבחינה היו נמוכים, המורה החליט לתת פקטור של 4 נק' לכל התלמידים. חשבו את המדדים לאחר הפקטור.
- (4) דגמו מקו ייצור 50 קופסאות של גפרורים. בדקו בכל קופסא בה יש 40 גפרורים את כמות הגפרורים הפגומים. התקבל שבממוצע יש 3 גפרורים פגומים בקופסא, עם סטיית תקן של 1.5 גפרורים. מה יהיה הממוצע ומה תהיה סטיית התקן של מספר התקינים בקופסא?
- (5) חברת בזק הציעה את ההצעה הבאה: שלושים שקלים דמי מנוי חודשיים קבועים וכן 10 אגורות לכל דקה של שיחה יוצאת. אדם בדק במשך שנה את דקות השיחות היוצאות שלו, וקיבל שבממוצע חודשי יש לו 600 דקות שיחות יוצאות עם שונות של 2500 דקות רבועות, כמו כן בחודש ינואר ציון התקן היה 2. חשבו את המדדים הללו עבור חשבון הטלפון החודשי של אותו אדם בשקלים אם היה משתמש בחבילה המוצעת לו על ידי בזק.
- (6) הוכיחו שאם כל התצפיות בהתפלגות עברו טרנספורמציה לינארית:  $Y_i = a \cdot X_i + b$ , אזי הממוצע והשונות של כלל התצפיות לאחר הטרנספורמציה יהיו בהתאמה:  $\bar{y} = a \cdot \bar{x} + b, S_y^2 = a^2 S_x^2$ .

### תשובות סופיות:

- (1) ממוצע: 315, סטיית תקן: 60, שכיח: 275.
- (2) ממוצע: 46, שונות: 30.25.
- (3) טווח: 40, חציון: 77, עשירון עליון: 91.
- (4) ממוצע: 37, סטיית תקן: 1.5.
- (5) ממוצע: 90, שונות: 25, ציון תקן: 2.
- (6)  $a^2 \cdot S_x^2$

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 10 - סטטיסטיקה תיאורית-מקדם ההשתנות

תוכן העניינים

1. כללי..... 42

## סטטיסטיקה תיאורית – מקדם ההשתנות:

### רקע:

כאשר מחשבים סטיית תקן למספר קבוצות בעלי ממוצע שונה, השוואת מידת פיזור הנתונים אינה מתייחסת לערך מרכז הנתונים (לממוצע למשל).  
על מנת לתת מדד פיזור המתחשב בממוצע הנתונים נחשב את מקדם ההשתנות –

$$CV = \frac{S(X)}{\bar{X}} : \text{Coefficient of Variation}$$

ככל שמקדם ההשתנות נמוך יותר, כך המשתנה מרוכז יותר סביב הממוצע, וככל שמקדם ההשתנות גבוה יותר, מידת הפיזור סביב הממוצע גבוהה יותר.

## שאלות:

1) להלן נתונים לגבי ציונים במבחן באנגלית ב-3 כיתות מתוך שכבה י' בתיכון:

| כיתה | ממוצע | מס' תלמידים | סטיית תקן |
|------|-------|-------------|-----------|
| 1    | 76    | 40          | 12        |
| 2    | 68    | 20          | 15        |
| 3    | 82    | 30          | 10        |

א. חשבו את מקדם ההשתנות בכל כיתה.

ב. מהי הכיתה הכי הטרוגנית?

2) נתונות שתי קבוצות: הממוצע בקבוצה א' הוא 100 והשונות 100.

הממוצע בקבוצה ב' הוא 500 והשונות 400.

באיזו קבוצה מידת הפיזור יחסית קטן יותר?

3) במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית

(מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום.

להלן טבלה המסכמת את האינפורמציה שנאספה על שני המשתנים:

| מספר פועלים | תפוקה | ממוצע     |
|-------------|-------|-----------|
| 15          | 48    |           |
| 2           | 10    | סטיית תקן |

לפי קריטריון CV:

א. הפיזור באופן יחסי שווה בין התפוקה היומית לכמות הפועלים העובדים ביום.

ב. הפיזור יחסית יותר גדול עבור התפוקה היומית מאשר עבור מספר הפועלים ביום.

ג. הפיזור יחסית יותר גדול עבור מספר הפועלים ביום מאשר עבור התפוקה היומית.

ד. אין מספיק נתונים כדי לחשב את CV.

## תשובות סופיות:

1) א.  $\frac{\sigma}{\bar{X}}$ . ב. כיתה ב'.

2) קבוצה ב'.

3) ב'.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 11 - סטטיסטיקה תיאורית- תרשים קופסא

תוכן העניינים

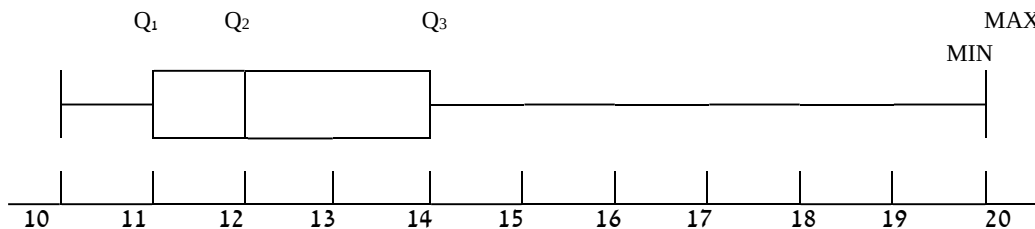
1. כללי..... 44

## סטטיסטיקה תיאורית – תרשים קופסא (Boxplot):

**רקע:**

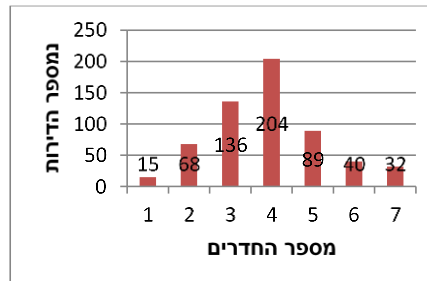
תרשים קופסא הינו תרשים שבעזרתו ניתן לבחון:

- (1) את המרכז של ההתפלגות על ידי החציון ( $Q_2$ ).
- (2) את הפיזור של הנתונים (הטווח והטווח הבין רבעוני).
- (3) את צורת ההתפלגות (סימטרית \אסימטרית ימנית או אסימטרית שמאלית).



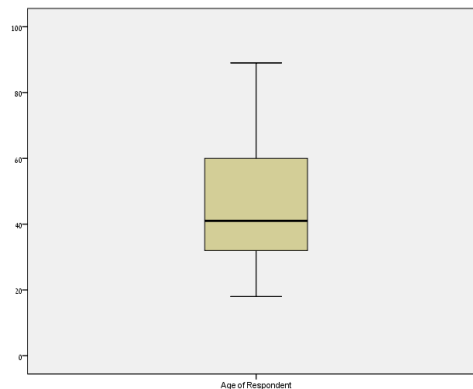
## שאלות:

1) להלן התפלגות מספר החדרים לדירות שנבנו בשנת 2009 בעיר אשדוד:



- א. מצאו את החציון, הרבעון התחתון והרבעון העליון של ההתפלגות.
- ב. שרטטו דיאגרמת קופסא להתפלגות.
- ג. מה ניתן לומר על צורת ההתפלגות?

2) להלן דיאגרמת קופסא המתארת את התפלגות הגיל (בשנים) באוכלוסייה מסוימת:



- א. מה הגיל החציוני?
- ב. מה בערך טווח הגילאים?
- ג. מה ניתן להגיד על צורת ההתפלגות?

## תשובות סופיות:

- 1) א. חציון: 4, רבעון תחתון: 3, רבעון עליון: 5.  
 ב. ראה גרף מלא בסרטון וידאו. ג. כמעט סימטרית.
- 2) א. חציון: 40. ב. טווח: 70. ג. התפלגות אסימטרית ימנית.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 12 - סטטיסטיקה תיאורית שאלות אמריקאיות

תוכן העניינים

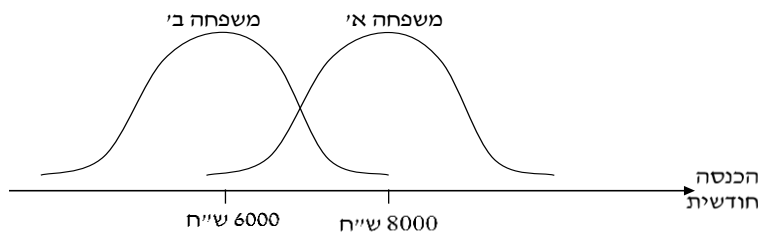
1. כללי..... 46

## סטטיסטיקה תיאורית – שאלות אמריקאיות:

### שאלות:

#### שאלות 1-3 מתייחסות לקטע הבא:

להלן שתי עקומות המתארות את התפלגות ההכנסות החודשיות של שתי משפחות שנבחרו באקראי:



- (1) לאיזו משפחה הכנסה שכיחה גבוהה יותר?
  - א. משפחה א'.
  - ב. משפחה ב'.
  - ג. לשתיהן אותה הכנסה שכיחה.
  - ד. לא ניתן לדעת – אין מספיק נתונים.
  
- (2) באיזו משפחה ההכנסה החציונית שווה להכנסה הממוצעת?
  - א. משפחה א'.
  - ב. משפחה ב'.
  - ג. בשתיהן ההכנסה החציונית שווה להכנסה הממוצעת.
  - ד. לא ניתן לדעת – אין מספיק נתונים.
  
- (3) באיזו משפחה סטית התקן של ההכנסה החודשית גבוהה יותר?
  - א. משפחה א'.
  - ב. משפחה ב'.
  - ג. לשתיהן אותה סטית תקן.
  - ד. לא ניתן לדעת – אין מספיק נתונים.

**הנתונים הבאים מתייחסים לשאלות 4-6:**

להלן נתונים חלקיים של טבלת שכיחויות:  
כמו כן, נתון כי הממוצע הוא 1.66.

| $F(x)$ | $x$  |
|--------|------|
| ?      | 0    |
| 10     | 1    |
| 6      | 2    |
| 15     | 3    |
| ?      | 4    |
| 50     | סה"כ |

(4) השכיח של הנתונים הוא:

- א. 0.
- ב. 15.
- ג. ישנם שני שכיחים: 0 ו-3.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של השכיח.

(5) חציון הנתונים הוא:

- א. 2.
- ב. 1.5.
- ג. 25.5.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(6) הטווח של הנתונים:

- א. 11.
- ב. 3.
- ג. 4.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(7) בהתפלגות אסימטרית ימנית של משתנה כמותי רציף, הערך המתאים למאון ה-30, ציון התקן שלו הוא בהכרח:

- א. שלילי.
- ב. חיובי.
- ג. אפס.
- ד. לא ניתן לדעת ללא הנתונים.

- 8) סדרת נתונים סטטיסטיים מונה 10 תצפיות. נתון כי סדרת הנתונים סימטרית סביב הממוצע. ממוצע הסדרה-40 ושונות הסדרה-100. בשלב מאוחר יותר נוספו שתי תצפיות נוספות לסדרה : 50 ו-30. השונות של 12 התצפיות :
- א. תקטן.
  - ב. תגדל.
  - ג. לא תשתנה.
  - ד. לא ניתן לחשב את השונות ללא ידיעת התצפיות.

#### הנתונים הבאים מתייחסים לשאלות 9-10:

בחברת "טיק" המשכורת הממוצעת היא 4,600 ₪ וסטיית התקן של משכורת זו הינה 200 ₪. לאחר מו"מ עם ועד עובדי ההנהלה סוכם כי המשכורת תוכפל פי 1.5.

- 9) מהי המשכורת הממוצעת החדשה (ב-₪)?
- א. 2,300.
  - ב. 6,900.
  - ג. 4,650.
  - ד. 4,600.
  - ה. חסרים נתונים כדי לדעת.
- 10) מהי סטיית התקן של המשכורת לאחר יישום המו"מ לגבי השכר (ב-₪)?
- א. 200.
  - ב. 300.
  - ג. 675.
  - ד. לא ניתן לדעת.
- 11) הוספת גודל קבוע לכל תצפיות סדרת נתונים :
- א. תגדיל את סטיית התקן.
  - ב. תקטין את סטיית התקן.
  - ג. לא תשנה את סטיית התקן.
  - ד. לא ניתן לדעת.

### הנתונים הבאים מתייחסים לשאלות 12-14:

להלן נתונים על ציוני תלמידים שנבחנו במועדים שונים בסטטיסטיקה:

| שם התלמיד | ציון | ממוצע הציונים במועד בו נבחן | סטיית התקן של הציונים במועד בו נבחן |
|-----------|------|-----------------------------|-------------------------------------|
| צבי       | 50   | 50                          | 12                                  |
| סטף       | 82   | 80                          | 5                                   |
| שרית      | 65   | 60                          | 15                                  |
| לובה      | 60   | 63                          | 1.5                                 |
| מיטב      | 70   | 70                          | 10                                  |

12) התלמיד הטוב ביותר ביחס לנבחנים באותו מועד בו נבחן הוא:

- א. מיטב.
- ב. צבי.
- ג. לובה.
- ד. שרית.
- ה. סטף.

13) פנינה נבחנה עם סטף וציון התקן שלה שווה לציון התקן של שרית לכן ציונה הוא:

- א. 80.55.
- ב. 65.
- ג. 80.
- ד. 81.66.

14) איזו כיתה היא ההומוגנית ביותר. הכיתה של:

- א. מיטב.
- ב. צבי.
- ג. לובה.
- ד. שרית.
- ה. סטף.

**הנתונים הבאים מתייחסים לשאלות 15-18:**

בבדיקת פתע של משרד הבריאות במפעל שוקולד, נמצא ש:

|   |    |    |    |    |    |    |    |             |
|---|----|----|----|----|----|----|----|-------------|
| 7 | 6  | 5  | 4  | 3  | 2  | 1  | 0  | שוקולד פגום |
| 8 | 10 | 11 | 13 | 12 | 48 | 63 | 35 | מס' קופסאות |

**15** מהו החציון של מספר הפגומים בקופסא:

- א. 1.
- ב. 2.
- ג. 4.
- ד. לא ניתן לדעת.

**16** מהו הרבעון התחתון של מספר הפגומים בקופסא?

- א. 1.
- ב. 2.
- ג. 3.
- ד. 4.
- ה. לא ניתן לדעת.

**17** מספר הפגומים בקופסא הוא משתנה:

- א. סדר.
- ב. שמי.
- ג. כמותי בדיד.
- ד. כמותי רציף.

**18** השכיח של מספר הפגומים בקופסא:

- א. 63.
- ב. 1.
- ג. 200.
- ד. לא ניתן לדעת.

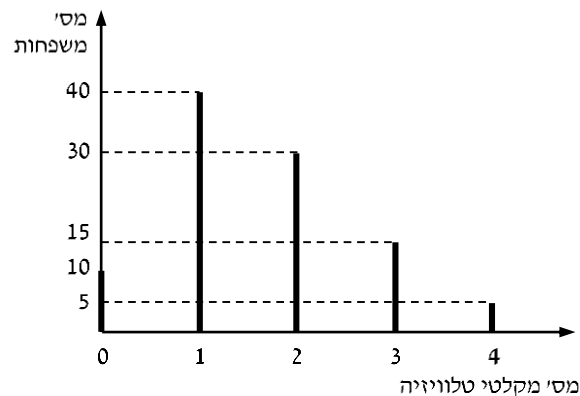
**19** ביחס לציר המספרים, רוב הערכים בהתפלגות א-סימטרית ימנית נמצאים:

- א. בערכים הגבוהים.
- ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
- ג. בערכים הנמוכים.
- ד. לא ניתן לדעת.
- ה. אף לא תשובה מהנ"ל נכונה.

- 20** בוצע מחקר על מספר העובדים בחברות מזון לעומת חברות תקשורת. החציון והממוצע בשתיהן שווה 8. איזה מהטענות הבאות היא הנכונה והמלאה ביותר:
- השכיחות ב-2 החברות זהה אך שונה מ-8.
  - השכיח ב-2 החברות זהה אך לא ניתן לדעת מהו.
  - השכיח בשתי חברות הינו בהכרח 8.
  - שכיח בחברה אחת שונה מ-8 ובשנייה הוא 8.
  - אף תשובה אינה נכונה.

### הנתונים הבאים מתייחסים לשאלות 21 עד 25:

נערך סקר על מספר מקלטי הטלוויזיה הנמצאים בבית. תוצאות הסקר נתונות בדיאגרמת מקלות הבאה:



### 21 המשתנה הנחקר כאן הוא:

- משתנה שמי.
- משתנה מסולם סדר.
- משתנה כמותי בדיד.
- משתנה כמותי רציף.

### 22 הטווח של ההתפלגות הוא:

- 35.
- 4.
- 3.
- 2.

**(23) ממוצע מספר מקלטי הטלויזיה למשפחה הוא :**

- א. 1.65
- ב. 1.5
- ג. 1
- ד. 2

**(24) השכיח של התפלגות זו היא :**

- א. 40
- ב. 1.5
- ג. 1
- ד. 2

**(25) מסתבר שיש בין 2 ל-5 משפחות נוספות שאין להם מקלטי טלויזיה ויש לצרף את המשפחות הללו להתפלגות. כיצד הנתון זה ישפיע על סטיית התקן?**

- א. יקטין אותו.
- ב. יגדיל אותו.
- ג. לא ישנה אותו.
- ד. אין לדעת.

### תשובות סופיות:

|        |        |        |        |        |
|--------|--------|--------|--------|--------|
| 1) א'  | 2) ג'  | 3) ג'  | 4) ג'  | 5) ב'  |
| 6) ג'  | 7) א'  | 8) ג'  | 9) ב'  | 10) ב' |
| 11) ג' | 12) ה' | 13) ד' | 14) ג' | 15) ב' |
| 16) א' | 17) ג' | 18) ב' | 19) ג' | 20) ה' |
| 21) ג' | 22) ב' | 23) א' | 24) ג' | 25) ב' |

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 13 - סטטיסטיקה תיאורית - שאלות מסכמות

תוכן העניינים

1. כללי..... 53

## סטטיסטיקה תיאורית – שאלות מסכמות:

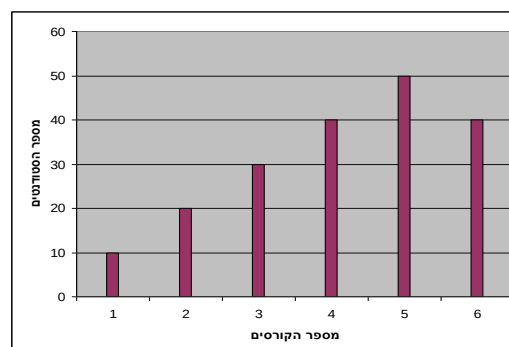
### שאלות:

(1) בדקו עבור 5 תלמידים את המשקל שלהם:

| מספר תלמיד | משקל בק"ג |
|------------|-----------|
| 1          | 58        |
| 2          | 62        |
| 3          | 48        |
| 4          | 34        |
| 5          | 58        |

- מהו המשתנה הנחקר? בדיד או רציף?
- מה המשקל החציוני, הממוצע והשכיח?
- מה הטווח וסטיית התקן של המשקל?
- לאותם תלמידים חישוב גם את הגובה בס"מ וקיבלו גובה ממוצע של 168 וסטיית תקן 6. במה תלמיד מספר 3, שגובהו 162, יותר חריג – במשקל או בגובה?
- הוסיפו עוד תלמיד השוקל 52 ק"ג בדיוק. הסבירו ללא חישוב כיצד הדבר ישפיע על הממוצע וסטיית התקן (יגדיל, יקטין או לא ישנה).

(2) בפקולטה להנדסה אספה המזכירות נתונים לגבי מס' הקורסים שכל סטודנט סיים בשנה הראשונה ללימודיו בשנת 2008. להלן התוצאות שהתקבלו:



- מה המשתנה הנחקר? בדיד או רציף?
- מהי צורת ההתפלגות?
- תארו את הנתונים בטבלת שכיחויות.
- חשבו את השכיח, החציון והטווח.

(3) להלן התפלגות הציונים בבחינה בלשון שנעשתה עבור תלמידי כיתות ד'.

במחקר השתתפו 150 תלמידים. ממוצע הציונים שהתקבל:  $\bar{X} = 7\frac{1}{15}$ .

| ציון | מספר התלמידים |
|------|---------------|
| 4    | 12            |
| 5    | 16            |
| 6    |               |
| 7    | 38            |
| 8    |               |
| 9    | 14            |
| 10   | 10            |

א. השלימו את השכיחויות החסרות בטבלה.

ב. חשבו את הציון החציוני, השכיח.

ג. חשב שונות וסטיית תקן להתפלגות הציונים.

(4) חברה סלולארית דגמה 200 אנשים. עבור כל אדם נבדקה מידת שביעות הרצון

של הלקוח מהחברה (1 - שביעות רצון נמוכה ו-5 - שביעות רצון גבוהה).  
להלן ההתפלגות שהתקבלה:

| שביעות רצון | מספר האנשים |
|-------------|-------------|
| 1           | 40          |
| 2           | 60          |
| 3           | 50          |
| 4           | 30          |
| 5           | 20          |

א. מה אחוז האנשים עם רמת שביעות רצון נמוכה?

ב. מה המשתנה הנחקר ומאיזה סוג הוא?

ג. מהי הדרך הגרפית המתאימה ביותר לתיאור הנתונים?

i. היסטוגרמה.

ii. דיאגרמת מקלות.

iii. דיאגרמת עוגה.

ד. חשבו את המדדים הבאים:

i. טווח.

ii. שכיח.

iii. חציון.

- (5) להלן מספר טענות, עבור כל טענה ציינו אם היא נכונה או לא נכונה ונמקו:
- א. בסדרה שבה כל התצפיות שוות זו לזו השונות הינה 0.
  - ב. ציון התקן של החציון תמיד יהיה 0.
  - ג. ציון התקן של האחוזון ה-70 בהתפלגות אסימטרית ימנית (חיובית) תמיד יהיה חיובי.
  - ד. אם נוסיף תצפיות לסדרה של תצפיות, הדבר בהכרח יגדיל את הממוצע של הסדרה.
  - ה. בסדרה החציון הינו 80. הוספו שתי תצפיות אחת 79 ואחת 100 לכן החציון יגדל.
  - ו. אם נוסיף את הערך 4 לכל התצפיות אז סטיית התקן לא תשתנה.
  - ז. אם נחלק את כל התצפיות בהתפלגות ב-2 אז השונות תקטן פי 2.
  - ח. אם נגדיל את ממוצע המשכורות של עובדים בחברה אז גם השונות תגדל.

## תשובות סופיות:

(1) א. המשתנה הנחקר: משקל תלמיד בק"ג, משתנה כמותי רציף.

ב.  $\bar{X} = 52$ ,  $Md = X_{\frac{n+1}{2}} = X_3 = 58$ , שכיח: 58.

ג.  $R = 28$ ,  $S = 10.12$ .

ד. הוא חריג יותר בגובה כי שם ציון התקן בערך מוחלט יותר גבוה.

ה. הממוצע לא ישתנה אך סטיית התקן תקטן.

(2) א. מספר הקורסים, בדיד. ב. התפלגות אסימטרית שמאלית.

ג. להלן טבלה: ד. שכיח: 5, טווח: 5.

| $f(x)$ | $x$  |
|--------|------|
| 10     | 1    |
| 20     | 2    |
| 30     | 3    |
| 40     | 4    |
| 50     | 5    |
| 40     | 6    |
| 190    | סה"כ |

(3) א. 20 תלמידים קיבלו ציון 6 ו-40 תלמידים קיבלו ציון 8.

ב. חציון: 7, שכיח: 8. ג. שונות: 2.533, סטיית תקן: 1.592.

(4) א. 20%. ב. שביעות רצון (סדר).

ג. 2. ד. טווח: 4, שכיח: 2, חציון: 2.

(5) א. נכון. ב. לא נכון. ג. לא נכון. ד. לא נכון. ה. לא נכון.

ו. נכון. ז. לא נכון. ח. לא נכון.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 14 - יסודות ההסתברות

תוכן העניינים

1. כללי ..... 57

## הגדרות יסודיות:

### רקע:

**ניסוי מקרי:** תהליך לו כמה תוצאות אפשריות. התוצאה המתקבלת נודעת רק לאחר ביצוע התהליך. למשל: תוצאה בהטלת קובייה, מזג האוויר בעוד שבועיים.

**מרחב מדגם:** כלל התוצאות האפשריות בניסוי המקרי. לדוגמה, בהטלת קובייה:  $\{1, 2, 3, 4, 5, 6\}$ , או: מזג האוויר בעוד שבועיים:  $\{\text{נאה, שרבי, מושלג, גשום, מעונן, חלקית, אביד}\}$ .

**מאורע:** תת קבוצה מתוך מרחב במדגם. מסומן באותיות:  $A, B, C$ . בהטלת קובייה למשל, המאורע 'לקבל לפחות 5' יסומן:  $A = \{5, 6\}$ . המאורע 'לקבל תוצאה זוגית' יסומן:  $B = \{2, 4, 6\}$ .

**גודל מרחב המדגם:** מספר התוצאות האפשריות במרחב המדגם. בהטלת קובייה למשל נקבל:  $|\Omega| = 6$ .

**גודל המאורע:** מספר התוצאות האפשריות במאורע עצמו. למשל, בהטלת הקובייה האירועים הקודמים יסומנו:  $|A| = 2, |B| = 3$ .

**מאורע משלים:** מאורע המכיל את כל התוצאות האפשריות במרחב המדגם פרט לתוצאות במאורע אותו הוא משלים. למשל, בהטלת הקובייה:  $\bar{A} = \{1, 2, 3, 4\}$ ,  $\bar{B} = \{1, 3, 5\}$ .

**מרחב מדגם אחיד (סימטרי):** מרחב מדגם בו לכל התוצאות במרחב המדגם יש את אותה עדיפות, אותה סבירות למשל, קובייה הוגנת, אך לא כמו מזג האוויר בשבוע הבא.

**הסתברות במרחב מדגם אחיד:** במרחב מדגם אחיד הסיכוי למאורע יהיה:  $P(A) = \frac{|A|}{|\Omega|}$ .

דוגמה: מה הסיכוי בהטלת קובייה לקבל לפחות 5?  $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6}$ .

דוגמה: מה הסיכוי בהטלת קובייה לקבל תוצאה זוגית?  $P(B) = \frac{|B|}{|\Omega|} = \frac{3}{6}$ .

**הסתברות במרחב לא אחיד:** תחושב לפי השכיחות היחסית:  $\frac{f}{n}$ .

**דוגמה:**

להלן התפלגות הציונים בכיתה מסוימת:

| הציון $x$ | מספר התלמידים – השכיחות $f$ |
|-----------|-----------------------------|
| 5         | 2                           |
| 6         | 4                           |
| 7         | 8                           |
| 8         | 5                           |
| 9         | 4                           |
| 10        | 2                           |

מה ההסתברות שתלמיד אקראי שניבחר בכיתה קיבל את הציון 8?  $\frac{f}{n} = \frac{5}{25} = 0.2$ .

מה ההסתברות שתלמיד אקראי שניבחר בכיתה יכשל?  $\frac{f}{n} = \frac{2}{25} = 0.08$ .

**הסתברות למאורע משלים:** הסתברות לקבוצת המשלים של המאורע ביחס למרחב המדגם:  $P(\bar{A}) = 1 - P(A)$ . למשל, בדוגמה הקודמת הסיכוי לעבור את הבחינה יכול

להיות מחושב לפי הסיכוי להיכשל:  $P(A) = 1 - \frac{2}{25} = \frac{23}{25}$ .

## שאלות:

(1) מהאותיות  $E, F$  ו- $G$  יש ליצור מילה בת 2 אותיות, לא בהכרח בת משמעות.

א. הרכיבו את כל המילים האפשריות.

ב. רשמו את המקרים למאורע:

i. במילה נמצאת האות  $E$ .

ii. במילה האותיות שונות.

ג. רשמו את המקרים למאורע  $\bar{A}$ .

(2) מטילים זוג קוביות.

א. רשמו את מרחב המדגם של הניסוי. האם מרחב המדגם אחיד?

ב. רשמו את כל האפשרויות לאירועים הבאים:

i. סכום התוצאות 7.

ii. מכפלת התוצאות 12.

ג. חשבו את הסיכויים לאירועים שהוגדרו בסעיף ב'.

(3) נבחר באקראי ספרה מבין הספרות 0-9.

א. מה ההסתברות שהספרה שנבחרה גדולה מ-5?

ב. מה ההסתברות שהספרה שנבחרה היא לכל היותר 3?

ג. מה ההסתברות שהספרה שנבחרה היא אי זוגית?

(4) להלן התפלגות מספר מקלטי הטלוויזיה עבור כל משפחה בישוב מסוים:

| מספר משפחות | 22 | 28 | 18 | 22 | 10 |
|-------------|----|----|----|----|----|
| מספר מקלטים | 0  | 1  | 2  | 3  | 4  |

נבחרה משפחה באקראי מהישוב.

א. מה ההסתברות שאין מקלטים למשפחה?

ב. מה ההסתברות שיש מקלטים למשפחה?

ג. מה ההסתברות שיש לפחות 3 מקלטים למשפחה?

(5) להלן התפלגות מספר המכוניות למשפחה ביישוב "עדן":

| מספר משפחות  | 20 | 40 | 100 | 30 | 10 |
|--------------|----|----|-----|----|----|
| מספר מכוניות | 0  | 1  | 2   | 3  | 4  |

נבחרה משפחה אקראית מן הישוב.

א. מה ההסתברות שאין לה מכוניות?

ב. מה ההסתברות שבבעלות המשפחה לפחות 3 מכוניות?

ג. מה הסיכוי שבבעלותה פחות מ-3 מכוניות?

- (6) נטיל מטבע רגיל 3 פעמים. בצד אחד של המטבע מוטבע עץ ובצד השני פלי.  
 א. רשמו את מרחב המדגם של הניסוי. האם מרחב המדגם הוא אחיד?  
 ב. רשמו את כל האפשרויות לאירועים הבאים:  
 i. התקבל פעם אחת עץ.  
 ii. התקבל לפחות פלי אחד.  
 ג. מהו המאורע המשלים ל-D?  
 ד. חשבו את הסיכויים לאירועים שהוגדרו בסעיפים ב-ג.

### תשובות סופיות:

- (1) א.  $\Omega = \{EE, EF, EG, FE, FF, FG, GE, GF, GG\}$   
 ב.  $A = \{EE, EF, EG, FE, GE\}$ ,  $B = \{EF, EG, FE, FG, GE, GF\}$   
 ג.  $\bar{A} = \{FF, FG, GF, GG\}$
- (2) א.  $\Omega = \left\{ \begin{pmatrix} (1,1) & (2,1) & (3,1) & (5,1) & (4,1) & (6,1) \\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{pmatrix} \right\}$   
 ב.  $A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$ ,  $C = \{(2,6), (3,4), (4,3), (6,2)\}$   
 ג. הסיכוי ל-A:  $\frac{1}{6}$ . הסיכוי ל-B:  $\frac{1}{9}$ .
- (3) א. 0.4 ב. 0.4 ג. 0.5  
 (4) א. 0.22 ב. 0.78 ג. 0.32  
 (5) א. 0.1 ב. 0.2 ג. 0.8  
 (6) א.  $\Omega = \{PPP, PPE, PEP, EPP, PEE, EPE, EEP, EEE\}$   
 ב.  $A = \{PPE, PEP, EPP\}$ ,  $D = \{PPP, PPE, PEP, EPP, PEE, EPE, EEP\}$   
 ג.  $\bar{D} = \{EEE\}$   
 ד.  $\frac{1}{8}$

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 15 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

תוכן העניינים

1. התפלגות נורמלית (טבלת  $z$  כוללת ערכים שליליים) ..... (ללא ספר)

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 16 - הסקה סטטיסטית - הקדמה

תוכן העניינים

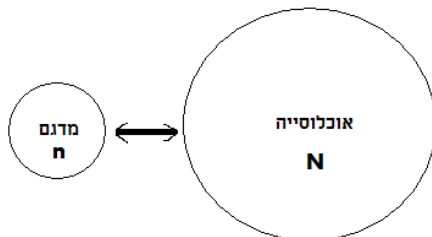
1. כללי..... 61

## הסקה סטטיסטית – הקדמה:

### רקע:

#### אוכלוסייה:

קבוצה שאליה מפנים שאלה מחקרית. למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסוכרת מתעניינת באוכלוסיית חולי הסוכרת בעולם.



#### מדגם:

חלק מתוך האוכלוסייה. למשל, אם נדגום באקראי 10 אנשים מתוך חולי הסוכרת אז זהו מדגם מתוך אוכלוסיית חולי הסוכרת.

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, אנו מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה. הדגימה בקורס תהיה דגימה מקרית - הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכוי להיכלל במדגם.

#### סטטיסטי:

גודל המחושב על המדגם.

#### פרמטר:

גודל המתאר את האוכלוסייה.

### הסימונים לפרמטר וסטטיסטי הם שונים:

| סטטיסטי (מדגם) | פרמטר (אוכלוסייה) | ממוצע                    |
|----------------|-------------------|--------------------------|
| $\bar{X}$      | $\mu$             |                          |
| $\hat{p}$      | $P$               | פרופורציה (שכיחות יחסית) |

פרמטר הוא גודל קבוע גם אם אנו לא יודעים אותו סטטיסטי הוא משתנה ממדגם למדגם ולכן יש לו התפלגות הנקראת התפלגות הדגימה.

### דוגמה (פתרון בהקלטה):

25% מאזרחי המדינה תומכים בהצעת החוק של חבר כנסת מסוים. הוחלט לדגום 200 אזרחים ומתוכם לבדוק מהו אחוז התומכים בהצעת החוק.

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

## שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מי האוכלוסייה?
  - מה המשתנה?
  - מהם הפרמטרים?
  - מהו גודל המדגם?
- (2) להלן התפלגות מספר מקלטי הטלויזיה למשפחה ב"העוגן". נגדיר את  $X$  להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלויזיה במדגם.
- מיהי האוכלוסייה ומהו המשתנה הנחקר?
  - מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

| מספר המשפחות       | מספר מקלטים |
|--------------------|-------------|
| 50                 | 0           |
| 250                | 1           |
| 350                | 2           |
| 300                | 3           |
| 50                 | 4           |
| סך הכול $N = 1000$ |             |

- (3) נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.
- מיהי האוכלוסייה?
  - מה המשתנה באוכלוסייה?
  - מהם הפרמטרים?
  - מהו הסטטיסטי?

## תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון. ג. ממוצע: 78, סטיית תקן: 15. ד. 2.
- (2) א. האוכלוסייה: 1000 משפחות ב"העוגן, המשתנה הנחקר: מס' מקלטים. ב.  $\bar{X}$  = ממוצע מדגם.
- (3) א. השכירים במדינה. ב. השכלה: אקדמאי, לא אקדמאי. ג. שיעור ההצלחות באוכלוסייה: 0.2. ג. מס' האקדמאים במדגם.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 17 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

1. התפלגות ממוצע המדגם ומשפט הגבול המרכזי ..... 64

## התפלגות ממוצע המדגם ומשפט הגבול המרכזי:

### רקע:

בפרק זה נדון בהתפלגות של ממוצע המדגם:  $\bar{x} = \frac{\sum x_i}{n}$ .

מכיוון שממדגם למדגם אנו יכולים לקבל ממוצע מדגם שונה, אזי ממוצע המדגם הוא משתנה מקרי ויש לו התפלגות.

גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים.

להלן רשימה של פרמטרים החשובים לפרק זה:

ממוצע האוכלוסייה נסמן ב- $\mu$  (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב- $\sigma^2$ .

סטיית תקן של אוכלוסייה:  $\sigma$ .

### תכונות התפלגות:

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה:  $E(\bar{x}) = \mu_{\bar{x}} = \mu$ .

שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב- $n$ .

תכונה זו נכונה רק במדגם מקרי:  $V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ .

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

אם נוציא שורש לשונות נקבל סטיית תקן של ממוצע המדגם שנקראת גם

טעות תקן:  $\sigma(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$ .

### דוגמה (פתרון בהקלטה):

השכר הממוצע במשק הינו 9000 ₪ עם סטיית תקן של 4000. דגמו באקראי 25 עובדים.

א. מיהי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

**דגימה מהתפלגות נורמאלית:**

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע  $\mu$  ושונות  $\sigma^2$ .

$$\text{ממוצע המדגם גם יתפלג נורמאלית: } Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}, \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

**דוגמה (פתרון בהקלטה):**

משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.  
מה ההסתברות שבמדגם של 4 תינוקות אקראיים בעת הולדתם המשקל הממוצע של התינוקות יהיה מתחת ל-3.5 ק"ג?

**משפט הגבול המרכזי:**

אם אוכלוסייה מתפלגת כלשהו עם ממוצע  $\mu$  ושונות  $\sigma^2$  אזי עבור מדגם מספיק

$$\text{גדול } (n \geq 30) \text{ ממוצע המדגם מתפלג בקירוב נורמאלי: } \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

**דוגמה (פתרון בהקלטה):**

משקל חפיסת שוקולד בקו ייצור מתפלג עם ממוצע 100 גרם וסטיית תקן של 4 גרם.  
דגמו מקו הייצור 36 חפיסות שוקולד אקראיות.  
מה ההסתברות שהמשקל הממוצע של חפיסות השוקולד שנדגמו יהיה מתחת ל-102 גרם?

## שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מיהי האוכלוסייה?
  - מה המשתנה?
  - מהם הפרמטרים?
  - מהו גודל המדגם?
  - מהו תוחלת ממוצע המדגם?
  - מהי טעות התקן?
- (2) משקל תינוק ביום היוולדו מתפלג נורמאלי עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
- מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם? נתון כי ביום מסוים נולדו 4 תינוקות.
  - מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?
  - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
  - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
  - הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?
- (3) הגובה של המתגייסים לצה"ל מתפלג נורמאלי עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.
- מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?
  - מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?
  - מה ההסתברות שהגובה הממוצע שלהם יסטה מתוחלת הגבהים בפחות מ-5 ס"מ?
  - מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

- (4) הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך הפתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלי.
- א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
- ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?
- ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
- ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?
- (5) נפח היין בבקבוק מתפלג נורמאלי עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.
- א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ"ק?
- ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ"ק?
- ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ"ק?
- ד. בקבוקי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?
- (6) משתנה מתפלג נורמאלי עם תוחלת 80 וסטיית תקן 4.
- א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
- ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
- הסבר את ההבדל בתשובות של שני הסעיפים.
- (7) לפי הערכות הלשכה המרכזית לסטטיסטיקה השכר הממוצע במשק הוא 8000 ₪ עם סטיית תקן של 3000 ₪.
- מה ההסתברות שבמדגם מקרי של 100 עובדים השכר הממוצע יהיה יותר מ-8500 ₪?

8) אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ.

א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל 78 ס"מ?

ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש 100 המוטות יספיקו למלאכה?

ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. היעזרו במשפט הגבול המרכזי.

9) נתון ש- $X \sim N(\mu, \sigma^2)$ . דגמו 5 תצפיות מאותה התפלגות והתבוננו בממוצע

המדגם  $\bar{X}$ . לכן:  $P(\bar{X} > \mu)$  יהיה (בחרו בתשובה הנכונה):

א. 0.

ב. 0.5.

ג. 1.

ד. לא ניתן לדעת.

10) נתון ש- $X$  מתפלג כלשהו עם תוחלת  $\mu$  ושונות  $\sigma^2$ .

החליטו לבצע מדגם בגודל 200 מתוך ההפלגות הנתונה לפי משפט הגבול המרכזי מתקיים (בחרו בתשובה הנכונה):

א.  $X \sim N\left(\mu, \frac{\sigma^2}{200}\right)$ .

ב.  $\mu \sim N\left(\mu, \frac{\sigma^2}{200}\right)$ .

ג.  $\bar{X} \sim N(\mu, \sigma^2)$ .

ד.  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{200}\right)$ .

11) נתון ש- $X \sim N(\mu, \sigma^2)$ . אם נדגום  $n$  תצפיות מתוך ההתפלגות ונגדיר:  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ ,

אזי (בחרו בתשובה הנכונה):

א.  $\mu$  ו- $\bar{X}$  יהיו משתנים מקריים.

ב.  $\mu$  יהיה משתנה מקרי ו- $\bar{X}$  קבוע.

ג.  $\bar{X}$  יהיה משתנה מקרי ו- $\mu$  קבוע.

ד.  $\mu$  ו- $\bar{X}$  יהיו קבועים.

## תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון.  
ג. ממוצע: 78, סטיית תקן: 15. ד. 2.  
ה. 78. ו. 10.6.
- (2) א. 0.8413 ב. 0.0013 ג. 0. ד. 0.1974.
- (3) א. 0 ב. 0 ג. 0.9544 ד. 178.205.
- (4) א. 0.0465 ב. 27.71 ג. 0.2628 ד. התשובה הייתה קטנה.
- (5) א. 0 ב. 0.1587 ג. 0.1587 ד. 0.5.
- (6) א. 0.5468 ב. 0.6826 ג. 0.271.
- (7) 0.0475
- (8) א. 0.9772 ב. 0.0228 ג. 0.271.
- (9) ב'.
- (10) ד'.
- (11) ג'.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 18 - מושגי יסוד באמידה

תוכן העניינים

1. כללי ..... 70

## מושגי יסוד באמידה:

### רקע:

כזכור מהמפגש הקודם, פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסוימת. כמו ממוצע הגבהים בקרב מתגייסים לצה"ל -  $\mu$ . כמו פרופורציית התומכים בממשלה בקרב אזרחי המדינה -  $p$ . בדרך כלל הפרמטרים הם גדלים שאינם ידועים באמת, ולכן מבצעים מדגמים במטרה לאמוד אותם. אין אפשרות לחשב אותם הניסיון הוא בלהעריך כמה הם שווים ככל שניתן.

- נסמן באופן כללי פרמטר באות  $\theta$  ואומד ב- $\hat{\theta}$ . הוא סטטיסטי המחושב על המדגם ובאמצעותו נאמוד את  $\theta$ .
- שגיאת אמידה:  $|\hat{\theta} - \theta|$  - ההפרש בין האומד לאמת (הפרמטר).

### דוגמה (פתרון בהקלטה):

בכנסת ה-19 קיבלה מפלגת העבודה 15 מנדטים. בערוץ 10 ברגע סגירת הקלפיות העריכו את מספר המנדטים של המפלגה להיות 17 מנדטים וזאת על סמך תוצאות מדגם של הערוץ.

- א. מה הפרמטר בדוגמה זו?
  - ב. מהי טעות האמידה של ערוץ 10?
- $\hat{\theta}$  יהיה אומד חסר הטיה ל- $\theta$  אם התוחלת של  $\hat{\theta}$  תהיה שווה ל- $\theta$ :  $E(\hat{\theta}) = \theta$ .
  - טעות התקן של אומד היא סטיית התקן שלו, כלומר:  $\sigma(\hat{\theta}) = S.E$ .

**פרמטרים מרכזיים והאומדים שלהם:****ממוצע האוכלוסייה  $\mu$ :**

האומד הנקודתי שלו יהיה: ממוצע המדגם:  $\bar{x} = \frac{\sum x}{n}$ .

$E(\bar{x}) = \mu$  לכן  $\bar{x}$  הינו אומד חסר הטיה ל- $\mu$ . כמו כן, טעות תקן:  $\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = SE$ .

**פרופורציה באוכלוסייה  $p$ :**

האומד הנקודתי שלו יהיה: פרופורציה במדגם:  $\hat{p} = \frac{y}{n}$ .

$E(\hat{p}) = p$ , לכן  $\hat{p}$  הינו אומד חסר הטיה ל- $p$ . כמו כן טעות התקן:  $\sigma(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}}$ .

**שונות האוכלוסייה  $\sigma^2$ :**

האומד הנקודתי שלו יהיה:  $S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ .

$E(S^2) = \sigma^2$  ולכן  $S^2$  הינו אומד חסר הטיה ל- $\sigma^2$ :  $S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$ .

**הערה:** אומד הוא הנוסחה הכללית לאמידת הפרמטר ואומדן הוא הערך הספציפי שהתקבל במדגם מסוים.

**דוגמה (פתרון בהקלטה):**

נדגמו 10 משפחות בתל אביב ונבדק עבור כל משפחה מספר הילדים שלה.  
להלן התוצאות שהתקבלו: 2, 1, 3, 2, 1, 4, 5, 2, 1, 3.  
אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים:

1. ממוצע מספר הילדים למשפחה בתל אביב.

2. שונות מספר הילדים למשפחה בתל אביב.

3. פרופורציית המשפחות בנות שני ילדים.

## שאלות:

- (1) מתוך 500 טירונים, נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
  - מהי טעות התקן של האומדן כשהמדגם בגודל 500?
  - מהו האומדן לפרמטר?
  - מהי טעות האמידה?
- (2) לפי נתוני היצרן, מקרר צורך בממוצע 2400 וואט לשעה עם סטיית תקן של 500 וואט לשעה.
- במדגם של 25 מקררים של היצרן התקבל ממוצע של 2342 וואט לשעה.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
  - מהי טעות התקן של האומדן?
  - מהו האומדן לפרמטר?
  - מהי טעות האמידה?
- (3) נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו: 175, 168, 184, 192, 171, 180, 177, 187, 168, 177.
- מצאו אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
  - מצאו אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
  - מצאו אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.
- (4) נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים.
- להלן התוצאות שהתקבלו:  $\sum_{i=1}^{20} X_i^2 = 1502.2$ ,  $\sum_{i=1}^{20} X_i = 162$ .
- אמדו את השכר הממוצע של השכירים במשק.
  - אמדו את סטיית התקן של שכר השכירים במשק.
- (5) במטרה לאמוד את ממוצע האוכלוסייה, דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישבו את הממוצע שלהם. מהי טעות התקן?
- סטיית התקן של האוכלוסייה.
  - סטיית התקן של ממוצע האוכלוסייה.
  - סטיית התקן של המדגם.
  - סטיית התקן של ממוצע המדגם.

6) משקל הממוצע של אוכלוסייה מסוימת הוא 75 ק"ג עם שונות של 25. אם יבחרו כל המדגמים האפשריים בגודל 10 מאוכלוסייה זו סטיית התקן של ממוצעי המדגמים תהייה:

- א. 3.
- ב. 2.5.
- ג. 1.581.
- ד. אין מספיק נתונים לדעת.

7) במדגם מקרי, מתי סכום ריבועי הסטיות מהממוצע,  $\sum_{i=1}^n (x_i - \bar{x})^2$ , מחולק ב- $n-1$ ?

- א. כאשר  $n$  קטן.
- ב. כאשר תצפיות המדגם אינן בלתי תלויות.
- ג. כאשר האוכלוסייה אינה מתפלגת נורמאלית.
- ד. כאשר מעוניינים באומד חסר הטיה לשונות האוכלוסייה ממנה הוצא המדגם.
- ה. כאשר מעוניינים לחשב את שונות התפלגות הדגימה של ממוצע המדגם.

8)  $X_1, X_2, \dots, X_{16}$  מדגם מקרי מתוך אוכלוסייה בעלת ממוצע  $\mu$  לא ידוע ושונות:  $\sigma^2 = 64$ . טעות התקן של האומד ל- $\mu$  היא:

- א. 16.
- ב. 8.
- ג. 4.
- ד. 2.

9) מהו אומד חסר הטיה?

- א. אומד שערכו שווה לממוצע התפלגות הדגימה שלו.
- ב. אומד שערכו שווה לערך הפרמטר באוכלוסייה.
- ג. אומד שממוצע התפלגות הדגימה שלו שווה לערך הפרמטר באוכלוסייה.
- ד. אומד שהסיכוי שערכו יהיה גבוה מערך הפרמטר באוכלוסייה שווה לסיכוי שיהיה נמוך ממנו.

|     |                                                               |          |         |         |
|-----|---------------------------------------------------------------|----------|---------|---------|
| (1) | א. 0.25                                                       | ב. 0.019 | ג. 0.24 | ד. 0.01 |
| (2) | א. אוכלוסייה : מקררים של יצרן, תוחלת : 2400, סטיית תקן : 500. |          |         |         |
|     | ב. 100                                                        |          | ג. 2342 | ד. 58   |
| (3) | א. 177.9                                                      | ב. 64.1  | ג. 0.4  |         |
| (4) | א. 8.1                                                        | ב. 3.16  |         |         |
| (5) | ד'                                                            |          |         |         |
| (6) | ג'                                                            |          |         |         |
| (7) | ד'                                                            |          |         |         |
| (8) | ד'                                                            |          |         |         |
| (9) | ג'                                                            |          |         |         |

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 19 - רווח סמך לתוחלת (ממוצע)

תוכן העניינים

1. רווח סמך כששונות האוכלוסיה ידועה ..... 75
2. קביעת גודל מדגם ..... 80
3. רווח סמך כששונות האוכלוסיה לא ידועה ..... 82

## רווח סמך כששונות האוכלוסייה ידועה:

### רקע:

ממוצע המדגם הוא אומדן לממוצע האוכלוסייה, אך לא באמת ניתן להבין ממנו על גודלו של ממוצע האוכלוסייה. ההסתברות שממוצע המדגם יהיה בדיוק כמו הממוצע האמיתי הוא אפסי.

מה שנהוג לעשות כדי לאמוד את ממוצע האוכלוסייה, זה לבנות רווח סמך.

נבנה מרווח בטחון שהסיכוי שהפרמטר  $\mu$  ייכלל בתוכו הוא:  $1-\alpha$ .

$1-\alpha$ : נקרא רמת בטחון או רמת סמך. כך ש:  $P(A \leq \mu \leq B) = 1-\alpha$ .

A - גבול התחתון של רווח הסמך.

B - הגבול העליון של רווח הסמך.

$L = B - A$  - אורך רווח הסמך.

### דוגמה (פתרון בהקלטה):

חוקר דגם 25 חיילים שנבחנו במבחן הפסיכומטרי. הוא בנה רווח סמך לממוצע הציונים במבחן הפסיכומטרי בקרב אוכלוסיית החיילים וקיבל בין 510 ל-590. רווח הסמך נבנה ברמת סמך של 95%.

1. מהי אוכלוסיית המחקר?
2. מה המשתנה באוכלוסייה?
3. מה הפרמטר שהחוקר רצה לאמוד?
4. מהו רווח הסמך?
5. מה אורך רווח הסמך?
6. מהי רמת הביטחון של רווח הסמך?

בפרק זה נרצה לבנות רווח סמך לתוחלת ( $\mu$ ) במקרה ש- $\sigma^2$  (שונות האוכלוסייה) ידועה.

פרמטר אותו נרצה לאמוד:  $\mu$ .

אומד נקודתי:  $\bar{x}$ .

תנאים לבניית רווח הסמך:  $X \sim N$  או  $n \geq 30$ .

$\sigma^2$  (שונות האוכלוסייה) ידועה.

$$\text{נוסחה לרווח הסמך: } \bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

### דוגמה (פתרון בהקלטה):

על פי נתוני היצרן אורך חיי סוללה מתפלג נורמאלית עם סטיית תקן של 1 שעה. מעוניינים לאמוד את תוחלת חיי סוללה. נדגמו באקראי 4 סוללות, אורך החיים הממוצע שהתקבל הוא 13.5 שעות. בנו רווח סמך ברמת סמך של 95% לתוחלת אורך חיי סוללה.

$$\text{שגיאת האמידה המקסימלית: } \varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$\varepsilon$  - נותן את שגיאת האמידה המקסימלית, דבר שנקרא גם טעות סטטיסטית, טעות דגימה.

### דוגמה (פתרון בהקלטה):

בהמשך לשאלה עם הסוללות. מה ניתן להגיד בביטחון של 95% על שגיאת האמידה?

קשרים מתמטיים ברווח הסמך:

- אורך רווח הסמך הוא פעמיים שגיאת האמידה המקסימלית:  $L = 2\varepsilon$ .
- ממוצע המדגם נופל תמיד באמצע רווח הסמך:  $\bar{X} = \frac{A+B}{2}$ .
- ככל שמספר התצפיות ( $n$ ) גבוה יותר, כך יש יותר אינפורמציה ולכן האומד יותר מדויק, ולכן נקבל רווח סמך יותר קצר.
- ככל שרמת הביטחון  $(1-\alpha)$  גבוהה יותר, כך:  $Z_{1-\frac{\alpha}{2}}$  יותר גבוה, ורווח הסמך יותר ארוך.

## שאלות:

- (1) חוקר התעניין לאמוד את השכר הממוצע במשק. על סמך מדגם הוא קבע שבביטחון של 95% כי השכר הממוצע במשק נע בין 9200 ל-9800 ₪.
- מי האוכלוסייה במחקר?
  - מה המשתנה הנחקר?
  - מה הפרמטר שאותו רוצים לאמוד?
  - מה רווח הסמך לפרמטר?
  - מהי רמת הסמך לפרמטר?
  - מה אורך רווח הסמך?
  - מה הסיכוי שטעות הדגימה תעלה על 300 ₪?
- (2) מעוניינים לאמוד את התפוקה היומית הממוצעת של מפעל מסוים ברמת סמך של 95%. במדגם אקראי של 100 ימים התקבלה תפוקה ממוצעת 4950 מוצרים ביום. לצורך פתרון הנח שסטיית התקן האמתית ידועה ושווה 150 מוצרים ביום. בנו את רווח הסמך.
- (3) מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלי עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
- בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
  - בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
  - הסבירו כיצד ומדוע השתנה רווח הסמך.
- (4) דגמו 200 עובדים מהמשק הישראלי. השכר הממוצע שלהם היה 9700 ₪. נניח שסטיית התקן של השכר במשק היא 3000 ₪.
- בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.
  - מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?
  - מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?
  - אם היינו מגדילים את גודל המדגם ובונים רווח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרווח הסמך מכיל את הפרמטר?

- (5) בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (105,99). שחזרו את:
- א. ממוצע המדגם.
  - ב. שגיאת האמידה המקסימאלית.
  - ג. רמת הסמך.
- (6) זמן החלמה מאנגינה מתפלג עם סטיית תקן של יומיים. חברת תרופות מעוניינת לחקור אנטיביוטיקה חדשה שהיא פיתחה. במחקר השתתפו 60 אנשים שחלו באנגינה וקיבלו את האנטיביוטיקה החדשה. בממוצע הם החלימו לאחר 4 ימים.
- א. בנו רווח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.
  - ב. מה היה קורה לאורך רווח הסמך אם היה תקציב להגדלת גודל המדגם פי 4? הסבירו.
  - ג. מה היה קורה לאורך רווח הסמך אם היינו בונים את רווח הסמך ברמת סמך גדולה יותר? הסבירו.
- (7) חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא:  $82 < \mu < 92$ . נתון שסטיית התקן בהתפלגות שווה ל-10 ושהמדגם מתבסס על 16 תצפיות. התפלגות המשתנה היא נורמאלית.
- א. מהו ממוצע המדגם?
  - ב. מהי רמת הסמך של רווח הסמך שנבנה?
  - ג. מה הסיכוי ששגיאת האמידה באמידת ממוצע האוכלוסייה תעלה על 5?
- (8) חוקר בנה רווח סמך לתוחלת כאשר השונות בהתפלגות ידועה ברמת סמך של 95%. אם החוקר כעת יבנה על סמך אותם נתונים רווח סמך ברמת סמך קטנה מ-95%, איזה מהמשפטים הבאים לא יהיה נכון.
- א. אורך רווח הסמך החדש יהיה קטן יותר.
  - ב. גודל המדגם יהיה כעת קטן יותר.
  - ג. המרחק בין ממוצע המדגם לקצות רווח הסמך יהיו קטנים יותר ברווח הסמך החדש.
  - ד. רמת הביטחון לבנות רווח הסמך החדש תהיה קטנה יותר.

(9) חוקר בנה רווח סמך ל- $\mu$  וקיבל:  $48 < \mu < 54$ . מה נכון בהכרח:

א.  $\mu = 51$ .

ב.  $\bar{X} = 6$ .

ג.  $\bar{X} = 51$ .

ד. אורך רווח הסמך הינו 3.

(10) איזה מהגורמים הבאים אינו משפיע על גודלו של רווח בר סמך, כאשר שונות האוכלוסייה ידועה (בחרו בתשובה הנכונה):

א. רמת הביטחון.

ב. סטיית התקן באוכלוסייה.

ג. מספר המשתתפים.

ד. סטיית התקן במדגם.

### תשובות סופיות:

(1) א. העובדים במשק. ב. שכר ב-ש. ג.  $\mu$ . ד.  $9200 < \mu < 9800$ .

ה. 0.95. ו. 600. ז. 0.05.

(2)  $4920.6 < \mu < 4979.4$

(3) א.  $223.42 < \mu < 236.58$ . ב.  $222.16 < \mu < 237.84$ .

ג. ראה סרטון.

(4) א.  $10,116 < \mu < 9,284$ . ב. הסטיה המירבית בין  $\bar{x}$  ל- $\mu$  היא 416 שם בבטחון של 95%.

ג. 800. ד. לא.

(5) א. 102. ב. 3. ג. 0.9544.

(6) א.  $4.42 < \mu < 83.5$ . ב. יקטן פי 2. ג. גדל.

(7) א. 87. ב. 5. ג. 0.9544.

(8) ב'.

(9) ג'.

(10) ד'.

## קביעת גודל מדגם:

### רקע:

אם מעוניינים לאמוד את ממוצע האוכלוסייה כאשר סטיית התקן של האוכלוסייה ידועה:  $\sigma$  ברמת סמך של  $1-\alpha$  ושגיאת אמידה שלא תעלה על  $\varepsilon$  מסוים, נציב

$$.n \geq \left( \frac{z_{1-\frac{\alpha}{2}} \cdot \sigma}{\varepsilon} \right)^2 \quad \text{בנוסחה הבאה:}$$

כדי להציב בנוסחה צריך שהמשתנה הנחקר יתפלג נורמלית או שהמדגם ייצא בגודל של לפחות 30 תצפיות.

### דוגמה (פתרון בהקלטה):

חברת תעופה מעוניינת לאמוד את תוחלת משקל המטען של נוסע. נניח שמשקל מטען של נוסע מתפלג נורמאלית עם סטיית תקן של 2 ק"ג. כמה נוסעים יש לדגום אם מעוניינים שבביטחון של 98% הסטייה המרבית בין ממוצע המדגם לממוצע האמיתי לא יעלה על 0.5 ק"ג? (תשובה: 87).

## שאלות:

- (1) משתנה מקרי מתפלג נורמאלית עם סטיית תקן ידועה 12. מה צריך להיות גודל המדגם כדי לבנות רווח סמך ברמת סמך של 98% שאורכו לא יעלה על 2?
- (2) מעוניינים לאמוד את הדופק הממוצע של מתגייסים לצבא. מעוניינים שבביטחון של 95% שגיאת האמידה המרבית תהיה 0.5. נניח שהדופק מתפלג נורמאלית על סטיית תקן של 3 פעימות לדקה.
- א. כמה מתגייסים יש לדגום?
- ב. אם ניקח מדגם הגדול פי 4 מהמדגם של סעיף א ונאמוד את הממוצע באותה רמת סמך כיצד הדבר ישפיע על שגיאת האמידה?
- (3) יהי  $X$  משתנה מקרי עם ממוצע  $\mu$  וסטיית תקן  $\sigma$ . חוקר רוצה לבנות רווח בר סמך ל- $\mu$  ברמת ביטחון של 0.95, כך שהאורך של הרווח יהיה  $0.5\sigma$ . מהו גודל המדגם הנדרש?

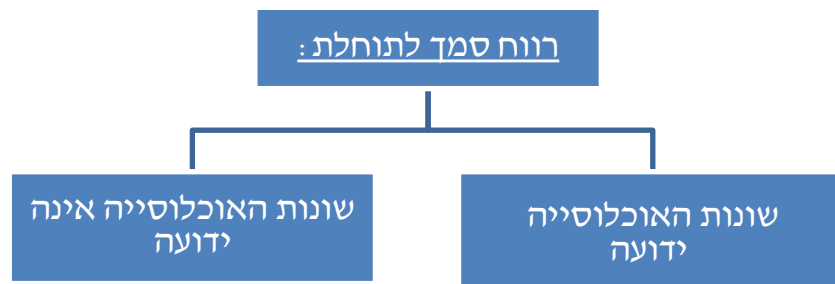
## תשובות סופיות:

- (1) 780.
- (2) א. 139. ב. הדבר יקטין את  $\varepsilon$  פי 2.
- (3)  $n = 62$ .

## רווח סמך כששונות האוכלוסייה לא ידועה:

רקע:

בבואנו לבנות רווח סמך לתוחלת אנו צריכים להתמקד בשני המצבים הבאים:

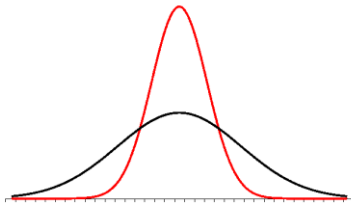


בפרק זה נעסוק במקרה ששונות האוכלוסייה  $(\sigma^2)$  אינה ידועה לנו.

מקרה יותר פרקטי.

**התנאי:**  $X \sim N$  או שהמדגם גדול.

**רווח סמך:**  $\bar{X} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \cdot \frac{S}{\sqrt{n}}$



$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1} \quad \text{האומד לשונות:}$$

**התפלגות T:**

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה

להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים.

התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן:  $df = n-1$ .

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה.

כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

**דוגמה (פתרון בהקלטה):**

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי כיתות ח' נורמאלית.

במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח'. להלן התוצאות

שהתקבלו בדקות: 4.7, 5.2, 4.6, 5.3.

בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

## שאלות:

- (1) מחקר מעוניין לדעת כיצד תרופה מסוימת משפיעה על קצב פעימות הלב. ל-5 אנשים שנטלו את התרופה מדדו את הדופק והתקבל מספר פעימות לדקה: 89, 79, 84, 88, 84. הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.
- א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הנ"ל.
- ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?
- ג. בהמשך לסעיף א', אם היינו בונים את רווח הסמך ברמת ביטחון של 99%, כיצד הדבר היה משפיע על רווח הסמך?
- (2) במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן  $S = 13$  ס"מ.
- בנו רווח סמך ברמת סמך של 90% לתוחלת גובה המתגייסים לצבא האמריקאי. מה יש להניח לצורך פתרון?
- (3) אדם מעוניין לאמוד את זמן הנסיעה הממוצע שלו לעבודה. לצורך כך הוא דוגם 5 ימים שזמן הנסיעה בהם בדקות הוא: 27, 34, 32, 40, 30.
- א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?
- ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים?
- (4) ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.
- א. בנו רווח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.
- ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמיתית של כלל הנבחנים.
- ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.
- (5) נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו:  $\sum_{i=1}^{60} X_i = 195$ ,  $\sum_{i=1}^{60} X_i^2 = 643.19$ .
- בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

- (6) נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו. להלן התוצאות שהתקבלו:  $\bar{x} = 13.8$ ,  $S = 2$ . בנו רווח סמך ברמת סמך של 96% לממוצע ההשכלה של אזרחים מעל גיל 50.
- (7) שני סטטיסטיקאים בנו רווח בר-סמך לאותו פרמטר  $\mu$ . לכל אחד מהסטטיסטיקאים מדגם אחר, אך באותו גודל 10. שניהם קבעו אותה רמת סמך. סטטיסטיקאי א': הניח  $\sigma = 20$ . סטטיסטיקאי ב': חישב לפי המדגם וקיבל  $S = 20$ . למי משני הסטטיסטיקאים יהיה רווח סמך ארוך יותר?
- סטטיסטיקאי א'.
  - סטטיסטיקאי ב'.
  - אותו אורך רווח סמך לשני הסטטיסטיקאים.
  - תלוי בתוצאות המדגם של כל סטטיסטיקאי.

### תשובות סופיות:

- (1) א.  $79.88 < \mu < 89.72$ . ב. כן. ג. הוא היה גדל.
- (2) ראה בסרטון.
- (3) א. צריך להניח שהמשתנה מתפלג נורמלית. ב. לא ניתן לדעת.
- (4) א.  $96.63 < \mu < 107.37$ . ב.  $96.90 < \mu < 107.10$ . ג. ראה בסרטון.
- (5)  $3.149 < \mu < 3.351$ .
- (6)  $13.42 < \mu < 14.18$ .
- (7) ב'.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 20 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

|    |                |
|----|----------------|
| 85 | 1. הקדמה       |
| 89 | 2. סוגי טעויות |

## הקדמה:

### רקע:

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטיקה. בבדיקת השערות על פרמטרים נעבוד לפי השלבים הבאים:

**שלב א:** נוהה את הפרמטר הנחקר.

**שלב ב:** נרשום את השערות המחקר.

השערת האפס המסומנות ב-  $H_0$ .

בדרך כלל השערת האפס מסמלת את אשר היה מקובל עד עכשיו, את השגרה, הנורמה.

השערה אלטרנטיבית (השערת המחקר) המסומנת ב-  $H_1$ .

ההשערה האלטרנטיבית מסמלת את החדשנות בעצם ההשערה האלטרנטיבית מדברת על הסיבה שהמחקר נעשה היא שאלת המחקר.

**שלב ג:** נבדוק האם התנאים לביצוע התהליך מתקיימים ונניח הנחות במידת הצורך.

**שלב ד:** נרשום את כלל ההכרעה. בתהליך של בדיקת השערות יוצרים כלל שנקרא כלל הכרעה. הכלל יוצר אזורי שנקראים:

1. **אזור דחייה:**

דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.

2. **אזור קבלה:**

קבלה של השערת האפס ודחייה של האלטרנטיבה. כלל ההכרעה מתבסס על איזשהו סטטיסטי. אזור הדחייה מוכתב על ידי סיכון שלוקח החוקר מראש שנקרא רמת מובהקות ומסומן ב-  $\alpha$ .

**שלב ה:** בתהליך יש ללכת לתוצאות המדגם ולחשב את הסטטיסטי המתאים ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה.

**שלב ו:** להסיק מסקנה בהתאם לתוצאות המדגם.

**דוגמה (פתרון בהקלטה):**

משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום לידתם בישראל 3300 גרם. משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$$n = 20, \bar{X} = 3120, S = 280$$

- א. מהי אוכלוסיית המחקר?
- ב. מה המשתנה הנחקר?
- ג. מה הפרמטר הנחקר?
- ד. מהן השערות המחקר?

## שאלות:

### בשאלות הבאות, ענו על הסעיפים הבאים:

- א. מהי אוכלוסיית המחקר?
- ב. מה המשתנה הנחקר?
- ג. מה הפרמטר הנחקר?
- ד. מהן השערות המחקר?

- (1) ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5.
- (2) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
- (3) במשך שנים אחוז המועמדים שהתקבל לפקולטה למשפטים היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. מחקר מעוניין לבדוק האם השנה מקשים על הקבלה לפקולטה למשפטים.
- (4) בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. רוצים לבדוק ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

### תשובות סופיות:

- (1) א. נבחנים בבגרות באנגלית.  
 ב. ציון.  
 ג. ממוצע הציונים בשיטת לימוד חדשה.  
 ד.  $H_0: \mu = 72$   
 $H_1: \mu > 72$
- (2) א. משקאות בבקבוק של חברה מסוימת.  
 ב. נפח משקה בסמ"ק.  
 ג. ממוצע נפח המשקה בבקבוק.  
 ד.  $H_0: \mu = 500$   
 $H_1: \mu < 500$
- (3) א. מועמדים לפקולטה למשפטים.  
 ב. משתנה דיכוטומי (התקבל, לא התקבל).  
 ג. אחוז הקבלה.  
 ד.  $H_0: p = 0.25$   
 $H_1: p < 0.25$
- (4) א. אזרחים בוגרים במשק.  
 ב. משתנה דיכוטומי (מובטל, עובד).  
 ג. אחוז האבטלה כיום.  
 ד.  $H_0: p = 0.08$   
 $H_1: p \neq 0.08$

## סוגי טעויות:

### רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה.  
הכלל יוצר אזורים שנקראים:

1. אזור דחייה – דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.
2. אזור קבלה – קבלה של השערת האפס ודחייה של האלטרנטיבה.

כלל ההכרעה מתבסס על איזשהו סטטיסטי.  
בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה – המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם. אם נשנה את כלל ההכרעה אז אנחנו יכולים לקבל מסקנה אחרת. אם נבצע מדגם חדש אז אנחנו עלולים לקבל תוצאה אחרת. לכן יתכנו טעויות במסקנות שלנו:

|  | הכרעה |             |             |
|--|-------|-------------|-------------|
|  |       | $H_0$       | $H_1$       |
|  | מצאות | $H_0$       | $H_1$       |
|  |       | אין טעות    | טעות מסוג 1 |
|  |       | טעות מסוג 2 | אין טעות    |

### הגדרת הטעויות:

טעות מסוג ראשון: להכריע לדחות את  $H_0$  למרות שבמצאות  $H_0$  נכונה.

טעות מסוג שני: להכריע לקבל את  $H_0$  למרות שבמצאות  $H_1$  נכונה.

### דוגמה (פתרון בהקלטה):

אדם חשוד בביצוע עבירה ונתבע בבית המשפט.  
אילו סוגי טעויות אפשריות בהכרעת הדין?

## שאלות:

- (1) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25. בסופו של דבר הוחלט להכריע לטובת חברת המשקאות.
- א. רשמו את השערות המחקר.  
 ב. מה מסקנת המחקר?  
 ג. איזו סוג טעות יתכן וביצעו במחקר?
- (2) במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
- א. האם ניתן לדעת אם בוצע טעות במחקר?  
 ב. מה סוג הטעות האפשרית?
- (3) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. ישנה טענה שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.
- א. מהי אוכלוסיית המחקר?  
 ב. מה המשתנה הנחקר?  
 ג. מה הפרמטר הנחקר?  
 ד. מה השערות המחקר?  
 ה. מה מסקנת המחקר?  
 ו. מהי סוג הטעות האפשרית במחקר?

## תשובות סופיות:

- (1) א.  $H_0: \mu = 500$   
            $H_1: \mu < 500$  . ב. לא דחינו את  $H_0$  . ג. טעות מסוג שני.
- (2) א. לא ניתן לדעת. ב. טעות מסוג ראשון.  
 (3) א. משפחות כיום. ב. מס' הילדים.
- ג. תוחלת מספר הילדים למשפחה כיום.  
 ה. לא לדחות את  $H_0$  . ו. טעות מסוג שני.
- ד.  $H_0: \mu = 2.3$   
        $H_1: \mu < 2.3$  .

# מבוא לסטטיסטיקה תיאורית והסקית

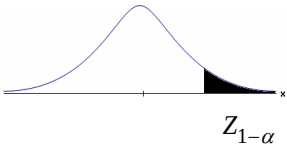
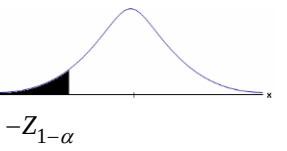
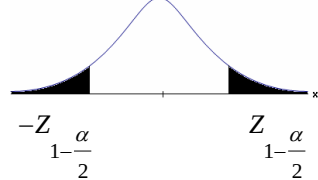
פרק 21 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

1. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה ידועה. .... 91
2. סיכוי לטעויות ועוצמה (ששונות האוכלוסיה ידועה). .... 95
3. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה ידועה). .... 101
4. בדיקת השערות על תוחלת ( ממוצע) כששונות האוכלוסיה לא ידועה. .... 106
5. הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע). .... 110
6. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה לא ידועה). .... 112

## בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה ידועה:

רקע:

| השערת האפס: השערה אלטרנטיבה:       | $H_0: \mu \leq \mu_0$<br>$H_1: \mu > \mu_0$                                                                                                            | $H_0: \mu \geq \mu_0$<br>$H_1: \mu < \mu_0$                                                                                                              | $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$                                                                                                                                                                                              |
|------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| תנאים:                             | 1. $\sigma$ ידועה<br>2. $X \sim N$ או מדגם מספיק גדול                                                                                                  |                                                                                                                                                          |                                                                                                                                                                                                                                          |
| כלל ההכרעה: אזור הדחייה של $H_0$ : | $Z_{\bar{x}} > Z_{1-\alpha}$<br><br>$Z_{1-\alpha}$<br>דוחים את $H_0$ | $Z_{\bar{x}} < -Z_{1-\alpha}$<br><br>$-Z_{1-\alpha}$<br>דוחים את $H_0$ | $Z_{\bar{x}} < -Z_{\frac{\alpha}{2}}$ או $Z_{\bar{x}} > Z_{\frac{\alpha}{2}}$<br><br>$-Z_{\frac{\alpha}{2}}$ $Z_{\frac{\alpha}{2}}$<br>דוחים את $H_0$ |

$$Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

סטטיסטי המבחן:

חלופה אחרת לכלל הכרעה:

|                                                                |                                                                |                                                                                                                                                        |                       |
|----------------------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| $\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$ | $\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$ | $\bar{X} > \mu_0 + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$<br>או<br>$\bar{X} < \mu_0 - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ | נדחה $H_0$ אם מתקיים: |
|----------------------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|

**דוגמה:**

יבול העגבניות מתפלג נורמלית עם תוחלת של 10 טון לדונם וסטיית תקן של 2.5 טון לדונם בעונה. משערים ששיטת זיבול חדשה תעלה את תוחלת היבול לעונה מבלי לשנות את סטיית התקן. נדגמו 4 חלקות שזוכלו בשיטה החדשה. היבול הממוצע שהתקבל היה 12.5 טון לדונם. בדקו את ההשערה ברמת מובהקות של 1%.

**פיתרון:**

אוכלוסייה: עגבניות.

המשתנה:  $X =$  יבול העגבניות בטון לעונה.

הפרמטר:  $\mu =$  תוחלת היבול בשיטת הזיבול החדשה.

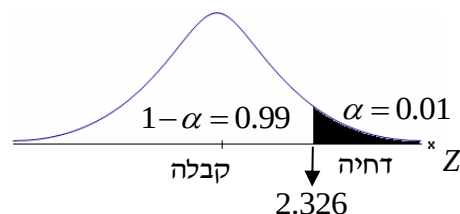
השערות:  
 $H_0: \mu = 10$   
 $H_1: \mu > 10$

תנאים:

1.  $X \sim N$ .

2.  $\sigma = 2.5$ .

כלל הכרעה:



נדחה את  $H_0$  אם  $Z_{\bar{X}} > 2.326$

תוצאות:  $n = 4$ ,  $\bar{x} = 12.5$

סטטיסטי המבחן:  $Z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

נציב:  $Z_{\bar{X}} = \frac{12.5 - 10}{\frac{2.5}{\sqrt{4}}} = 2 < 2.326$

מסקנה:

לא נדחה  $H_0$  (נקבל  $H_0$ ).

ברמת מובהקות של 1% לא נוכל לקבל את הטענה ששיטת הזיבול החדשה מעלה את תוחלת היבול של העגבניות.

## שאלות:

- (1) ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5. בהנחה שגם בשיטתו סטיית התקן תהיה 15 מה מסקנתכם ברמת מובהקות של 5%?
- (2) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
  - א. מה מסקנתכם ברמת מובהקות של 2.5%?
  - ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?
- (3) מהנדס האיכות מעוניין לבדוק אם מכונה מכיילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?
- (4) המשקל הממוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ובשימוש בדיאטה מסוימת שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם זה היה 84 ק"ג. יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.
- (5) לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדגם של 25 ברגים העובי הממוצע היה 4.07 מ"מ. קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט. הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.
- (6) במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5% מה תמיד נכון? בחרו בתשובה הנכונה.
  - א. הגדלת רמת המובהקות לא תשנה את מסקנת המחקר.
  - ב. הגדלת רמת המובהקות תשנה את מסקנת המחקר.
  - ג. הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
  - ד. הקטנת רמת המובהקות תשנה את מסקנת המחקר.

(7) חוקר ערך מבחן דו צדדי ברמת מובהקות של  $\alpha$  והחליט לדחות את השערת האפס.

אם החוקר היה עורך מבחן צדדי ברמת מובהקות של  $\frac{\alpha}{2}$  אזי בהכרח:

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס הייתה לא נדחית.
- ג. לא ניתן לדעת מה תהיה מסקנתו במקרה זה.

(8) שני סטטיסטיקאים בדקו השערות:  $H_0: \mu = \mu_0$  כנגד  $H_1: \mu > \mu_0$ ,

עבור שונות ידועה ובאותה רמת מובהקות.

שני החוקרים קבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.

- א. אם חוקר א' החליט לדחות את  $H_0$ , מה יחליט חוקר ב'? נמקו.
- ב. אם חוקר א' יחליט לא לדחות את  $H_0$ , מה יחליט חוקר ב'? נמקו.

### תשובות סופיות:

(1) נקבל  $H_0$ , בר"מ של 5% לא נקבל את הטענה של המורה ששיטת הלימוד שלו מעלה את ממוצע הציונים.

(2) א. נדחה  $H_0$ , בר"מ של 2.5% נקבל את תלונת אגודת הצרכנים בדבר הפחתת נפח המשקה בבקבוק.

ב. הגדלנו את רמת המובהקות לכן אנחנו נשארים בדחייה של  $H_0$  והמסקנה לא תשתנה.

(3) נדחה  $H_0$ , בר"מ של 5% נקבע שהמכונה לא מאופסת.

(4) נדחה  $H_0$ , בר"מ של 0.1 נקבל את הטענה שהדיאטה יעילה ומפחיתה את המשקל הממוצע.

(5) נקבל  $H_0$ , בר"מ של 0.05 נכריע שתוחלת עובי הבורג מתים למפרט.

(6) א'.

(7) ג'.

(8) א. לדחות. ב. לא ניתן לדעת.

## סיכוי לטעויות ועוצמה (ששונות האוכלוסייה ידועה):

רקע:

| מציאות | הכרעה |             |             |
|--------|-------|-------------|-------------|
|        |       | $H_0$       | $H_1$       |
|        |       | $H_0$       | $H_1$       |
|        | $H_0$ | אין טעות    | טעות מסוג 1 |
|        | $H_1$ | טעות מסוג 2 | אין טעות    |

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

$$\alpha = P(H_0 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_0}(H_0 \text{ נכונה})$$

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ נכונה} \mid \text{לקבל את } H_0) = P_{H_0}(H_0 \text{ נכונה})$$

רמת בטחון:

$$1 - \alpha = P(H_0 \text{ נכונה} \mid \text{לקבל את } H_0) = P_{H_0}(H_0 \text{ נכונה})$$

עוצמה:

$$1 - \beta = \pi = P(H_1 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_1}(H_0 \text{ נכונה})$$

## התהליך לחישוב סיכוי לטעות מסוג שני:

| השערת האפס:<br>השערה אלטרנטיבה:       | $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$                                                | $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$                                                | $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$                                                                                                                      |
|---------------------------------------|-----------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| תנאים:                                | 1. $\sigma$ ידועה<br>2. $X \sim N$ או מדגם מספיק גדול                                   |                                                                                         |                                                                                                                                                                  |
| כלל ההכרעה:<br>אזור הדחייה של $H_0$ : | $\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$                          | $\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$                          | $\bar{X} > \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$<br>או<br>$\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$       |
| חישוב $\beta$ :                       | $P_{\mu_1} \left( \bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$ | $P_{\mu_1} \left( \bar{X} > \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$ | $P_{\mu_1} \left( \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$ |

התפלגות ממוצע המדגם:  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

התקנון:

דוגמה:

בתחילת השנה חשבון הטלפון הסלולארי הממוצע לאדם היה 200 ₪ עם סטיית תקן של 80 ₪ לחודש. בעקבות כניסתן של חברות טלפון סלולארית חדשות מעוניינים לבדוק האם כיום ממוצע חשבון הטלפון הסלולארי פחת. לצורך בדיקה דגמו באקראי 36 אנשים וחשבון הטלפון הסלולארי שלהם היה 150 ₪ בממוצע לחודש.

א. רשמו את השערות המחקר ובנו כלל הכרעה במונחי חשבון ממוצע מדגמי ברמת מובהקות של 5%.

ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?

ג. נניח שבמציאות כיום החשבון הממוצע הוא 160 ₪. מה הסיכוי לבצע טעות מסוג שני?

ד. אם נקטין את רמת המובהקות מסעיף א', כיצד הדבר ישפיע על התשובה מסעיף ג'?

**פתרון:**

א. אוכלוסייה: משלמי חשבון טלפון סלולאר כיום.

המשתנה:  $X$  = חשבון הטלפון החודשי בשקלים.

הפרמטר:  $\mu$ .

השערות:  
 $H_0: \mu = 200$   
 $H_1: \mu < 200$

תנאים:

1.  $\mu = 200$ .

2.  $n = 36$ .

$$\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, K = \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\alpha = 0.05$$

$$Z_{1-\alpha} = Z_{0.95} = 1.645$$

$$\text{נציב: שקלים } K = 200 - 1.645 \cdot \frac{80}{\sqrt{36}} = 178.07$$

כלל ההכרעה: דחה את  $H_0$  אם שקלים  $\bar{X} < 178.07$ .

ב. ברמת מובהקות של 5% נכריע שאכן ממוצע חשבון הטלפון הסלולרי פחת מתחילת השנה.

ג. השערות:  
 $H_0: \mu_0 = 200$   
 $H_1: \mu < 200$

כלל ההכרעה: נדחה את  $H_0$  אם  $\bar{X} < 178.07$ .

$$H_1: \bar{X} \sim N\left(160, \frac{80^2}{36}\right)$$

$$Z = \frac{178.07 - 160}{\frac{80}{\sqrt{36}}} = 1.36$$

$$\beta = P_{H_1} \left( \overset{\text{לקבל את}}{H_0} \right) = P_{H_1} (\bar{X} > 178.07) = 1 - \phi(136) = 1 - 0.9131 = 0.0869$$

ד. הקטנת  $\alpha$  מגדילה את  $\beta$ .

## שאלות:

(1) נתון ש:  $X \sim N(\mu, \sigma^2 = 1)$ .

- להלן השערות של חוקר לגבי הפרמטר  $\mu$ :  $H_0: \mu = 5$ ,  $H_1: \mu = 7$ . מעוניינים ליצור כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.
- א. עבור אילו ערכים של  $X$  שידגם נדחית השערת  $H_0$ ?
- ב. מה הסיכוי לבצע טעות מסוג שני?
- ג. אם במדגם התקבל ש- $X = 6.9$  מה תהיה המסקנה ומה הטעות האפשרית?

- (2) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. מעוניינים לבדוק אם כיום ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה.
- א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי ברמת מובהקות של 5%.
- ב. בהמשך לסעיף א' מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?
- ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העצמה של הכלל מסעיף א'?

- (3) להלן נתונים על תהליך של בדיקת השערות על תוחלת:

$$H_0: \mu = 200, H_1: \mu \neq 200, \sigma = 30, n = 225$$

- א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.
- ב. בהמשך לסעיף א', מהי העצמה אם התוחלת שווה ל-195?
- ג. הסבירו, ללא חישוב, איך העצמה תשתנה אם רמת המובהקות תהיה 5%?

- (4) מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50

מ"מ וסטית תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומוודדים את קוטרים, בכדי לבדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוילת כנדרש או שקוטר הצינורות קטן מהדרוש.

- א. רשמו את ההשערות ואת כלל ההכרעה ברמת מובהקות של 5%.
- ב. אם ביום כלשהו מכונת הייצור התקלקלה והיא מייצרת את הצינורות בקוטר שתוחלתו 48 מ"מ בלבד (סטית התקן לא השתנתה), מה ההסתברות שהתקלה לא תתגלה בביקורת האיכות? כיצד נקראת הסתברות זו?
- ג. הסבירו ללא חישוב כיצד התשובה לסעיף ב' תשתנה אם רמת המובהקות תגדל.
- ד. הסבירו ללא חישוב כיצד התשובה לסעיף ב' תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

- (5) להלן השערות של מחקר:  $H_0: \mu = 50$ ,  $H_1: \mu = 58$ . מעוניינים לדגום 100 תצפיות. ידוע שסטיית התקן של ההתפלגות הינה 20.
- א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10%. מהי רמת המובהקות?
- ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו)?
- i. סטיית התקן הייתה יותר גדולה.
- ii. הסיכוי לטעות מסוג שני גדול יותר.

### השאלות שלהלן הן שאלות רב-ברירה, בחרו בתשובה הנכונה ביותר:

- (6) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:
- א. הסיכוי לטעות מסוג ראשון גדל.
- ב. העוצמה של המבחן גדלה.
- ג. הסיכוי לטעות מסוג שני גדל.
- ד. תשובות א' ו-ב' נכונות.
- (7) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן:
- א. השערת האפס נכונה.
- ב. השערת האפס נדחתה.
- ג. השערת האפס לא נדחתה.
- ד. אף אחת מהתשובות לא נכונה בהכרח.

- (8) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

| $1 - \beta$ | $\alpha$ |
|-------------|----------|
| גדולה       | א. גדולה |
| קטנה        | ב. גדולה |
| גדולה       | ג. קטנה  |
| קטנה        | ד. קטנה  |

- (9) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית  $H_0$  קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך:
- א. הן  $\alpha$ , והן  $1 - \beta$ , יקטנו.
- ב.  $\alpha$  יישאר ללא שינוי ואילו  $1 - \beta$  יגדל.
- ג.  $\alpha$  יגדל ואילו  $1 - \beta$  יקטן.
- ד. הן  $\alpha$  והן  $1 - \beta$  יגדלו.

- (10) ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח שלחץ הדם בקרב עיתונאים גבוה יותר מהממוצע באוכלוסייה. הוא לקח מדגם של 60 עיתונאים וקיבל ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?
- טעות מסוג ראשון.
  - טעות מסוג שני.
  - טעות מסוג שלישי.
  - אין טעות במסקנתו.

### תשובות סופיות:

- א. מעל 6.645. ב. 0.3594.  
ג. דחינו את  $H_0$ , תתכן טעות מסוג ראשון.
- א. נדחה  $H_0$  אם  $\bar{X} < 2.24$ . ב. נדחה  $H_0$  ג. 1.
- א. נדחה  $H_0$  אם  $\bar{X} > 203.29$  או  $\bar{X} < 196.71$ . ב. 0.8051. ג. תקטן.
- א. נדחה  $H_0$  אם  $\bar{X} < 48.9$ . ב. 0.0885. ג. תקטן. ד. תקטן.
- א. 0.0033. ב. i. רמת המובהקות הייתה קטנה.  
ב. ii. רמת המובהקות הייתה גדלה.
- ד'.
- ג'.
- ג'.
- א'.
- ב'.

## מובהקות תוצאה – אלפא מינימלית (ששונות האוכלוסייה ידועה):

### רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה:

באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- $p_v$ . את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות.

המסקנה של המחקר תקבע לפי העיקרון הבא: אם  $p_v \leq \alpha$ , דוחים את  $H_0$ . מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

(לקבל את תוצאות המדגם וקיצוני)  $p_v = P_{H_0}$ .

אם ההשערה היא דו צדדית:

(לקבל את תוצאות המדגם וקיצוני)  $p_v = 2P_{H_0}$ .

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

| שערת האפס:<br>השערה אלטרנטיבה:           |                                          | $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$                                                                                                    |  |
|------------------------------------------|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|--|
| $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$ | $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$ | $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$                                                                                                    |  |
| 1. $\sigma$ ידועה                        |                                          | תנאים:                                                                                                                                         |  |
| 2. $X \sim N$ או מדגם מספיק גדול         |                                          | p-value                                                                                                                                        |  |
| $P_{H_0}(\bar{X} \geq \bar{x})$          | $P_{H_0}(\bar{X} \leq \bar{x})$          | אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \Leftarrow \bar{x} > \mu_0$<br>אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \Leftarrow \bar{x} < \mu_0$ |  |

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}, \quad \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right) \quad \text{כאשר בהנחת השערת האפס:}$$

**דוגמה:**

המשקל הממוצע של מתגייסים לצבא לפני 20 שנה היה 65 ק"ג. מחקר מעוניין לבדוק האם כיום המשקל הממוצע של מתגייסים גבוה יותר. נניח שמשקל המתגייסים מתפלג נורמאלי עם סטיית תקן של 12 ק"ג. במדגם של 16 מתגייסים התקבל משקל ממוצע של 71 ק"ג.

א. מהי מובהקות התוצאה?

ב. מה המסקנה אם רמת המובהקות היא 5% ואם רמת המובהקות היא 1%?

**פתרון:**

א. אוכלוסייה: המתגייסים לצבא כיום.

משתנה:  $X$  = משקל בק"ג.

פרמטר:  $\mu$ .

השערות:  
 $H_0: \mu = 65$   
 $H_1: \mu > 65$

תנאים:

1.  $X \sim N$ .

2.  $\sigma = 12$ .

תוצאות מדגם:

$$n = 16$$

$$\bar{X} = 71$$

$$P_V = P_{H_0} \left( \begin{array}{c} \text{לתוצאות} \\ \text{המדגם} \\ \text{וקיצוני} \end{array} \right) = P_{H_0} (\bar{X} \geq 71) = 1 - \phi(2) = 1 - 0.9772 = 0.0228$$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{71 - 65}{\frac{12}{\sqrt{16}}} = 2$$

$$\alpha_{\min} = 0.0228$$

## שאלות:

- (1) להלן השערות של מחקר:  $H_0: \mu = 70$ ,  $H_1: \mu > 70$ . המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות:  $n = 100$ ,  $\bar{x} = 74$ . מהי מובהקות התוצאה?
- (2) השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪. מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?
- (3) אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ-100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטיית התקן של משקל החטיף היא 12 גרם. האדם מתכוון לשקול 100 חפיסות חטיפים ולאחר מכן להגיע להחלטה. לאחר הבדיקה הוא קיבל משקל הממוצע של 98.5 גרם.
  - א. רשמו את השערות המחקר.
  - ב. מהי רמת המובהקות המינימלית עבורה דוחים את השערת האפס?
  - ג. מהי רמת המובהקות המקסימלית עבורה נקבל את השערת האפס?
  - ד. מה המסקנה ברמת מובהקות של 5?
- (4) מכונה לחיתוך מוטות במפעל חותכת מוטות באורך שמתפלג נורמלית עם תוחלת אליה כוונה המכונה וסטיית תקן 2 ס"מ. ביום מסוים כוונה המכונה לחתוך מוטות באורך 80 ס"מ. אחראי האיכות מעוניין לבדוק האם המכונה מכוילת. לצורך כך נדגמו מקו הייצור 16 מוטות שנחתכו אורכן הממוצע היה 81.7 ס"מ.
  - א. מהי רמת המובהקות המינימלית עבורה נכריע שהמכונה לא מכוילת?
  - ב. אם נוסיף עוד תצפית שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?
  - ג. הכרע ברמת מובהקות של 5% האם המכונה מכוילת.
- (5) אם מקבלים בחישובים אלפא מינימלית (P value) קטנה מאד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון/לא נכון? נמק.

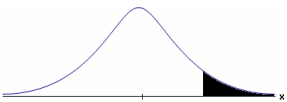
- (6) בבדיקת השערות התקבל שה-  $p\text{-value} = 0.02$ .  
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%?  
בחרו בתשובה הנכונה.
- א. יקבל את השערת האפס בכל מקרה.
  - ב. ידחה את השערת האפס מקרה.
  - ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
  - ד. לא ניתן לדעת כי אין מספיק נתונים.
- (7) מובהקות התוצאה (PV) היא גם (בחרו בתשובה הנכונה):
- א. רמת המובהקות המינימאלית לדחות השערת האפס.
  - ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
  - ג. רמת המובהקות שנקבעת מראש על ידי החוקר שטרם קיבל את תוצאות המחקר.
  - ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.
- (8) בבדיקת השערות מסוימת התקבל:  $p\text{ value} = 0.0254$  לכן (בחרו בתשובה הנכונה):
- א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את  $H_0$ .
  - ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את  $H_0$ .
  - ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את  $H_0$ .
  - ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את  $H_0$ .

### תשובות סופיות:

- (1) 0.0228.
- (2) עבור כל רמת מובהקות סבירה.
- (3) א.  $H_0: \mu = 100$   
 $H_1: \mu < 100$  . ב. 0.1056 . ג. 0.1056.
- ד. נכריע שיש עמידה בהתחייבות של החברה.
- (4) א. 0.0006 . ב. יקטן. ג. נכריע שאין כיול.
- (5) נכון.
- (6) א'.
- (7) א'.
- (8) ג'.

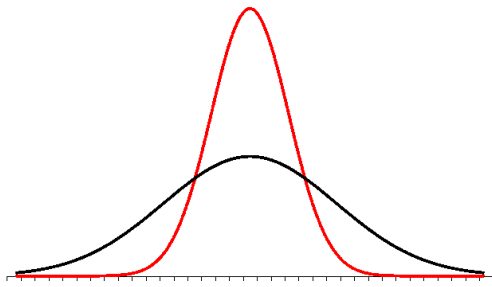
## בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה לא ידועה:

רקע:

| השערת האפס:<br>השערה<br>אלטרנטיבה:               | $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$                                                                                                                              | $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$                                                                                                                                | $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$                                                                                                                                                                                                                                 |
|--------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| תנאים:                                           | 1. $\sigma$ אינה ידועה<br>2. $X \sim N$ או מדגם מספיק גדול                                                                                                            |                                                                                                                                                                         |                                                                                                                                                                                                                                                                             |
| כלל ההכרעה:<br>אזור הדחייה<br>של $H_0$ :         | $t_{\bar{x}} > t_{1-\alpha}^{(n-1)}$<br><br>$t_{1-\alpha, n-1}$<br>- דוחים את $H_0$ | $t_{\bar{x}} < -t_{1-\alpha}^{(n-1)}$<br><br>$-t_{1-\alpha, n-1}$<br>- דוחים את $H_0$ | $t_{\bar{x}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{x}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$<br><br>$-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$<br>- דוחים את $H_0$ |
| חלופה לכלל<br>הכרעה:<br>נדחה $H_0$ אם<br>מתקיים: | $\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$                                                                                                       | $\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$                                                                                                         | $\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$<br>או<br>$\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$                                                                                                                |

$$. t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} : \text{סטטיסטי המבחן}$$

$$. S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$



### התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויה במושג שנקרא דרגות חופש.

דרגות החופש הן:  $df = n - 1$ .

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה. כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

### דוגמה (פתרון בהקלטה):

מפעל קיבל הזמנה לייצור משטחים בעובי של 0.1 ס"מ. כדי לבדוק האם המפעל עומד בדרישה נדגמו 10 משטחים ונמצא שהעובי הממוצע הוא 0.104 עם אומדן לסטיית תקן 0.002 ס"מ.

א. מהן השערות המחקר?

ב. מה ההנחה הדרושה לצורך פתרון?

ג. בדוק ברמת מובהקות של 5%.

## שאלות:

(1) משך זמן ההחלמה בלקיחת אנטיביוטיקה מסוימת הוא 120 שעות בממוצע עם סטיית תקן לא ידועה. מעוניינים לבדוק האם אנטיביוטיקה אחרת מקטינה את משך זמן ההחלמה. במדגם של 5 חולים שלקחו את האנטיביוטיקה האחרת התקבלו זמני ההחלמה הבאים: 90, 95, 100, 80, 125 שעות. מה מסקנתכם ברמת מובהקות של 5%. מהי ההנחה הדרושה לצורך הפתרון?

(2) משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום היוולדם בישראל 3300 גר'. משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$$n = 20$$

$$\bar{x} = 3120$$

$$S = 280$$

מה מסקנתכם ברמת מובהקות של 5% מה יש להניח לצורך פתרון?

(3) ציוני מבחן אינטליגנציה מתפלגים נורמלית. בארה"ב ממוצע הציונים הוא 100. במדגם שנעשה על 23 נבחנים ישראלים, התקבל ממוצע ציונים 104.5 וסטיית התקן המדגמית 16. האם בישראל ממוצע הציונים שונה מארה"ב? הסיקו ברמת מובהקות של 5%.

(4) באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{10} X_i = 750$$

$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$

נתון שההתפלגות היא נורמלית.

בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80.

- (5) ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם. ליאור השתמש בטבלה של התפלגות  $Z$ . רוני השתמשה בטבלה של התפלגות  $t$ . מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.
- אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
  - אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
  - שני החוקרים בהכרח יגיעו לאותה מסקנה.
  - לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

- (6) נתון ש:  $X \sim N(\mu, \sigma^2)$  כמו כן נתונות ההשערות הבאות:  $H_0: \mu = \mu_0$  .  $H_1: \mu < \mu_0$  . חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות.  $\sigma^2$  לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות. בחר בתשובה הנכונה:
- א. כעת בברור הוא ידחה את השערת האפס.
  - ב. כעת הוא דווקא יקבל את השערת האפס.
  - ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

### תשובות סופיות:

- (1) נדחה  $H_0$ .
- (2) נדחה  $H_0$ .
- (3) נקבל  $H_0$ .
- (4) נקבל  $H_0$ .
- (5) ב'.
- (6) ג'.

## הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע):

### רקע:

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות  $\alpha$  על  $\mu$  :

$$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$$

על ידי בניית רווח סמך ברמת סמך של  $1-\alpha$  ל- $\mu$  :

אם  $\mu_0$  נופל ברווח  $\leftarrow$  נקבל את  $H_0$ .

אם  $\mu_0$  לא נופל ברווח  $\leftarrow$  נדחה את  $H_0$ .

### דוגמה:

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו :

$$H_0: \mu = 80, H_1: \mu \neq 80, \alpha = 5\%$$

החוקר בנה רווח סמך ברמה של 90% וקיבל:  $79 < \mu < 84$ .  
האם אפשר לדעת מה מסקנתו, ואם כן מהי?

### פתרון (פתרון מלא בהקלטה):

רווח הסמך ברמת סמך של 90% מכיל "80".

ברמת סמך של 95% רווח הסמך יגדל ויכיל "80".

לכן, ברמת מובהקות של 5% נקבל  $H_0$ .

## שאלות:

- (1) חוקר רצה לבדוק את ההשערות הבאות:  $H_0: \mu = 90$ ,  $H_1: \mu \neq 90$ . החוקר בנה רווח סמך לתוחלת ברמת סמך של 95% וקיבל את רווח הסמך הבא: (87, 97).  
אם החוקר מעוניין לבצע בדיקת השערות ברמת מובהקות של 1% האם ניתן להגיע למסקנה ע"ס רווח הסמך? נמקו.
- (2) חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בדם. ידוע כי מספר מיליגרם הסוכר בסמ"ק דם הוא משתנה מקרי שמתפלג נורמלית עם סטיית תקן 10.4 מ"ג. נלקח מדגם של 60 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרם סוכר היה 115.5 מ"ג לסמ"ק.  
א. בנה רווח סמך ברמת סמך 95% לתוחלת רמת הסוכר בדם אצל הניזונים מדיאטה זו.  
ב. ידוע שתוחלת רמת הסוכר בדם באוכלוסיה היא 90 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א שהדיאטה משפיעה על רמת הסוכר בדם? הסבירו.
- (3) יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצלין היא 200 מ"ג לקפסולה. משרד הבריאות ביצע מדגם של 8 קפסולות אקראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצלין לקפסולה עם סטיית תקן מדגמית של 5 מ"ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.  
א. בנו רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.  
ב. בדקו ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

## תשובות סופיות:

- (1) נקבל השערת.  
(2) א.  $112.87 \leq \mu \leq 118.13$ .  
ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.  
(3) א.  $191.8 \leq \mu \leq 200.2$ . ב. נכריע שיש אמת בפרסום.

## מובהקות תוצאה – אלפא מינימלית (ששונות האוכלוסייה לא ידועה):

### רקע:

נזכיר שהמסקנה של המחקר תיקבע לפי העיקרון הבא: אם  $p_v \leq \alpha$  דוחים את  $H_0$ . מובהקות התוצאה היא הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

(לקבל את תוצאות המדגם וקיצוני)  $p_v = P_{H_0}$ .

אם ההשערה היא דו צדדית:

(לקבל את תוצאות המדגם וקיצוני)  $p_v = 2P_{H_0}$ .

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

| השערת האפס:<br>השערה אלטרנטיבה: | $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$             | $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$ | $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$                                                                                                    |
|---------------------------------|------------------------------------------------------|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| תנאים:                          | 1. אינה ידועה<br>או<br>2. מדגם מספיק גדול $X \sim N$ |                                          |                                                                                                                                                |
| p-value                         | $P_{H_0}(\bar{X} \geq \bar{x})$                      | $P_{H_0}(\bar{X} \leq \bar{x})$          | אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \Leftarrow \bar{x} > \mu_0$<br>אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \Leftarrow \bar{x} < \mu_0$ |

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$

$$d.f = n-1$$

**דוגמה:**

ממוצע זמן הנסיעה של אדם לעבודה הינו 40 דקות. הוא מעוניין לבדוק דרך חלופית שאמורה להיות יותר מהירה. לצורך כך הוא דוגם 5 ימים שבהם הוא נוסע בדרך החלופית. זמני הנסיעה שקיבל בדקות הם: 34, 40, 30, 32, 27. הניחו שזמן הנסיעה מתפלג נורמלי.

- א. רשמו את השערות המחקר.
- ב. מצאו חסמים למובהקות התוצאה.
- ג. מה המסקנה ברמת מובהקות של 5%?

**פתרון:**

אוכלוסייה: כלל הנסיעות לעבודה בדרך החלופית.

משתנה:  $X$  = זמן נסיעה בדקות.

תנאים:  $X \sim N$ .

פרמטר:  $\mu$ .

א. השערות:  
 $H_0: \mu = 40$   
 $H_1: \mu < 40$

ב. תוצאות המדגם:

$$n = 5, \bar{X} = \frac{\sum X_i}{n} = \frac{34 + 40 + \dots}{5} = 32.6$$

$$S^2 = \frac{\sum X_i^2 - n \cdot \bar{X}^2}{n-1} = \frac{34^2 + 40^2 - \dots - 5 \cdot 32.6^2}{5-1} = 23.4$$

$$S = \sqrt{23.4}$$

$$t_{\bar{X}} = \frac{\frac{\bar{X} - \mu_0}{S}}{\frac{1}{\sqrt{n}}} = \frac{32.6 - 40}{\frac{4.88}{\sqrt{5}}} = -3.39$$

$$P_V = P_{H_0} = ( \bar{X} \leq 32.6 ) = P(t \leq -3.39)$$

$$d.f = 5 - 1 = 4$$

$$1\% < P_V < 2.5\%$$

$P_V < \alpha = 0.05$ , לכן דוחים את  $H_0$ .

מסקנה: בר"מ של 5% נכריע שהדרך החלופית מהירה יותר.

## שאלות:

(1) קו ייצור אריזות סוכר נארזות כך שהמשקל הממוצע של אריזות הסוכר צריך להיות אחד קילוגרם. בכל יום דוגמים מקו הייצור 5 אריזות במטרה לבדוק האם קו הייצור תקין. בבדיקה דגמו 5 אריזות סוכר ולהלן משקלן בגרמים: 1024, 1008, 1005, 996, 997.

א. רשמו את השערות המחקר.

ב. מהי מובהקות התוצאה? הצג חסמים.

ג. מה המסקנה ברמת מובהקות של 5%?

(2) חוקר בדק את הטענה כי פועלים העובדים במשמרת לילה איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. במדגם מיקרי של 25 פועלים שעבדו במשמרת לילה נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטית תקן של 3 שעות. מהי ה- $\alpha$  המינימלית שלפיה ניתן להחליט שאכן העובדים במשמרת לילה איטיים יותר?

(3) הגובה של מתגייסים לצה"ל מתפלג נורמלית. במדגם של 25 מתגייסים מדדו את הגבהים שלהם בס"מ והתקבלו התוצאות הבאות:

$$\bar{x} = 176.2, \sum (x_i - \bar{x})^2 = 2832$$

מטרת המחקר היא לבדוק האם תוחלת הגבהים של המתגייסים גבוהה מ-174 ס"מ באופן מובהק.

מהי בקרוב מובהקות התוצאה ועל פיה מה תהיה המסקנה ברמת מובהקות של 6%?

## תשובות סופיות:

$$(1) \quad \begin{array}{ll} \text{א.} & H_0: \mu = 1000 \\ & H_1: \mu \neq 1000 \\ \text{ב.} & 20\% \leq P_v \leq 50\% \end{array}$$

ג. ברמת מובהקות של 5% לא נוכל לקבוע שקו הייצור אינו תקין.

(2) 10%.

(3) 1.01, נקבל את  $H_0$ .

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 22 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

1. כששוניות האוכלוסיה ידועות.....115
2. כששוניות האוכלוסיה לא ידועות ובהנחת שוויון שוניות.....117

## כששונויות האוכלוסייה ידועות:

### רקע:

המטרה היא לאמוד את פער התוחלות:  $\mu_1 - \mu_2$ , כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

האומד נקודתי:  $\bar{x}_1 - \bar{x}_2$ .

התנאים לבניית רווח הסמך:

1.  $\sigma_1^2, \sigma_2^2$  ידועות.

2.  $X_1, X_2 \sim N$  או  $n_1, n_2 > 30$ .

3. שני מדגמים בלתי תלויים.

רווח סמך:  $(\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ .

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של  $1-\alpha$ , לא קיים הבדל בין התוחלות.

### דוגמה (פתרון בהקלטה):

נדגמו 100 תושבים מאזור A והמשכורת הממוצעת הייתה שם 9200 ₪. כמו כן נדגמו 120 תושבים מאזור B וממוצע המשכורות שהתקבל שם 8700 ₪. לצורך פתרון נניח שסטיית התקן של המשכורות באוכלוסיית שני האזורים היא 1800 ₪. אמדו ברמת סמך של 90% את הפרש השכר הממוצע בין אזור A לאזור B.

## שאלות:

- (1) מעוניינים לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. ידוע שציוני הפסיכומטרי מתפלגים נורמאלית עם סטיית תקן 100. במדגם של 16 נבחנים חיילים התקבל ממוצע 543. במדגם של 20 תלמידי תיכון התקבל ממוצע 508. בנו רווח סמך לפער תוחלות הציונים בין חיילים לתלמידי תיכון ברמת סמך של 90%. מה ניתן להסיק מרווח סמך זה?
- (2) ציוני IQ מתוכננים כך שיתפלגו נורמאלית עם סטיית תקן של 15. במדגם של 20 נבחנים ישראלים התקבל ממוצע ציונים 104. במדגם של 23 נבחנים אמריקאיים התקבל ממוצע ציונים 99. א. בנו רווח סמך ברמת סמך של 95% לפער בין ישראל לארה"ב בממוצע הציונים במבחן ה-IQ. ב. האם קיים הבדל בין ישראלים לאמריקאים מבחינת ממוצע הציונים?
- (3) חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. ידוע שרמת הקשיות של ברגים מתפלגת נורמלית עם סטיית תקן של 4 יחידות. במדגם של 15 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות ובמדגם של 12 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25. עבור אילו רמות בטחון יקבע שאין הבדל בין שני סוגי הברגים מבחינת ממוצע רמת הקשיות שלהם?

## תשובות סופיות:

- (1)  $(-20, 90)$ .
- (2) א.  $-3.99 < \mu_1 - \mu_2 < 13.99$ .  
ב. לא נוכל לטעון בביטחון של 95% שקיים הבדל בין ישראל לארה"ב.
- (3) רמות בטחון הגבוהות מ-0.9476.

## כששונויות האוכלוסייה לא ידועות ובהנחת שוויון שונויות:

### רקע:

המטרה היא לאמוד את פער התוחלות:  $\mu_1 - \mu_2$ , כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

האומד נקודתי:  $\bar{X}_1 - \bar{X}_2$ .

התנאים לבניית רווח הסמך:

$$1. \sigma_1^2 = \sigma_2^2$$

$$2. X_1, X_2 \sim N$$

3. מדגמים בלתי תלויים.

השונויות המשוקללת: כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השונויות הזו על ידי שקלול שתי השונויות של שני המדגמים על ידי

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{הנוסחה הבאה:}$$

$$d.f = n_1 + n_2 - 2 \quad \text{דרגות החופש:}$$

$$\text{רווח סמך: } (\bar{X}_1 - \bar{X}_2) \pm t_{\frac{1-\alpha}{2}, n_1+n_2-2} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של  $1 - \alpha$ , לא קיים הבדל בין התוחלות.

### דוגמה (פתרון בהקלטה):

מחקר מעוניין לבדוק האם קיים הבדל בין תל אביב לבאר שבע מבחינת ההכנסה הממוצעת של אקדמאים. להלן תוצאות המדגם שנעשה:

| מספר האקדמאים                | תל אביב | באר שבע |
|------------------------------|---------|---------|
| ממוצע הכנסות של אקדמאים      | 11,000  | 9500    |
| סטיית התקן של הכנסות אקדמאים | 200     | 250     |

בנו רווח סמך ברמת ביטחון של 90% להפרש תוחלות ההכנסה בשני האזורים. הניחו שהשכר מתפלג נורמלית עם אותה שונות בכל אחד מהאזורים.

## שאלות:

- (1) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

| המדינה              | ישראל   | ארה"ב   |
|---------------------|---------|---------|
| גודל המדגם          | 15      | 15      |
| סכום הציונים        | 1560    | 1470    |
| סכום ריבועי הציונים | 165,390 | 147,560 |

מצאו רווח סמך ברמת סמך של 95% לסטייה בין ממוצע הציונים בישראל לממוצע הציונים בארה"ב. רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

- (2) להלן 4 תצפיות על משתנה  $X$  שמתפלג:  $N(\mu_x, \sigma^2)$ , ומשתנה  $Y$  שמתפלג:  $N(\mu_y, \sigma^2)$ .

|   |    |    |    |    |
|---|----|----|----|----|
| X | 22 | 20 | 21 | 25 |
| Y | 18 | 25 | 17 | 12 |

חשבו רווח סמך ל- $\mu_y - \mu_x$  ברמת הסמך 90%, בהנחה ששני המדגמים בלתי תלויים.

## תשובות סופיות:

- (1) הנחות:
- השונות שווה.
  - שהציונים מתפלגים נורמלית.
  - המדגמים אינם תלויים זה בזה.
- $$-5.52 < \mu_1 - \mu_2 < 17.52$$
- (2)
- $$-9.6 < \mu_y - \mu_x < 1.6$$

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 23 - רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים

תוכן העניינים

1. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים ..... 119

## רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים:

### רקע:

**מדגם מזווג:** מדגם אחד שבו יש  $n$  צמדים. כל תצפית במדגם תנפק זוג ערכים:  $X$  ו- $Y$ .

ניצור משתנה חדש:  $D = x - y$ .

הפרמטר שנרצה לאמוד:  $\mu_D$ .

התנאים לבניית רווח הסמך:

$$1. \quad x, y \sim N$$

2. המדגם מזווג.

$$\text{נוסחת רווח הסמך: } \bar{D} \pm t_{1-\frac{\alpha}{2}}^{n-1} \frac{S_D}{\sqrt{n}}$$

כאשר דרגות החופש:  $df = n - 1$ .

### דוגמה (פתרון בהקלטה):

מעוניינים לבדוק האם יש הבדל בין מהירות הריצות של שתי תוכנות מחשב. לקחו 5 קבצים אקראיים והריצו אותם בשתי התוכנות:

| הקובץ               | 1  | 2  | 3  | 4  | 5  |
|---------------------|----|----|----|----|----|
| הזמן בתוכנה הראשונה | 25 | 48 | 49 | 46 | 38 |
| הזמן בתוכנה השנייה  | 27 | 46 | 42 | 40 | 48 |

הניחו כי זמני הריצות מתפלגים נורמלית. מצאו רווח סמך של 95% להפרש תוחלת הזמן בין שתי התוכנות.

## שאלות:

- (1) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן הציונים בסמסטר א' ו-ב' :

|     |    |    |    |    |          |
|-----|----|----|----|----|----------|
| 82  | 75 | 90 | 68 | 74 | סמסטר א' |
| 100 | 76 | 87 | 84 | 80 | סמסטר ב' |

נניח שהציונים מתפלגים נורמאלית.

- א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א' לבין סמסטר ב'.
- ב. האם על סמך רווח הסמך קיים הבדל בין הסמסטרים מבחינת תוחלת הציונים?
- ג. מה צריך לשנות בנתונים כדי שהמדגמים יהיו בלתי תלויים?
- (2) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נדגמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות :

| חברה/ מדינה  | ארה"ב | קנדה | הולנד | פולין | מצרים | סין | יפן |
|--------------|-------|------|-------|-------|-------|-----|-----|
| בזק - X      | 1.5   | 2.1  | 2.2   | 3     | 3.5   | 3.2 | 4.2 |
| קווי זהב - Y | 1.4   | 2    | 1.9   | 3.1   | 3.3   | 3.2 | 4.2 |

בהנחה והמחירים מתפלגים נורמלית עבור כל חברה, בנו רווח סמך ברמת סמך של 90% לתוחלת הפרש המחירים של שתי החברות.

## תשובות סופיות:

- (1) א.  $-19 < \mu_0 < 38$  . ב. בביטחון של 95% לא קיים הבדל.
- ג. ראה הסבר בסרטון.
- (2)  $-0.013 < \mu < 0.185$  .

# מבוא לסטטיסטיקה תיאורית והסקית

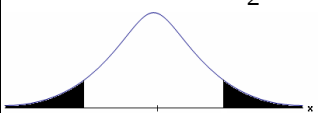
פרק 24 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסייה ידועות.....121
2. כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות.....125

## בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונות של האוכלוסייה ידועות – רקע

| $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 > c$                                                                                                      | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 < c$                                                                                                        | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 \neq c$                                                                                                                                                                                                                | השערת האפס:<br>השערה אלטרנטיבית:                                                                |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| מדגמים בלתי תלויים<br>$\sigma_1, \sigma_2$ ידועות<br>$X_1, X_2 \sim N$<br>או מדגמים מספיק גדולים                                                                    |                                                                                                                                                                       |                                                                                                                                                                                                                                                                                  | תנאים:                                                                                          |
| $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$<br><br>$Z_{1-\alpha}$<br>דוחים את $H_0$ | $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\alpha}$<br><br>$-Z_{1-\alpha}$<br>דוחים את $H_0$ | $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$<br><br>$-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$<br>דוחים את $H_0$ | כלל ההכרעה:<br>אזור הדחייה של $H_0$ :<br>$-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$ |

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{סטטיסטי המבחן:}$$

חלופה אחרת לכלל הכרעה:

| נדחה $H_0$ אם מתקיים :                                                                                  |                                                                                                                                                                                                                              |
|---------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ | $\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$<br>או<br>$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ |
| $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ |                                                                                                                                                                                                                              |

התפלגות הפרש הממוצעים:  $\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{התקנון:}$$

**דוגמה (פתרון בהקלטה) :**

בשנת 2004 הפער בין השכר הממוצע של הגברים לנשים היה 3000 ₪ לטובת הגברים. מעוניינים לבדוק האם כיום הצטמצם הפער בין הגברים לנשים מבחינת השכר הממוצע. נדגמו 100 עובדים גברים. שכרם הממוצע היה 9,072 ₪. נדגמו 80 עובדות, שכרן הממוצע היה 7809 ₪. לצורך פתרון נניח שסטיות התקן של השכר ידועות ושוות ל-2000 ₪ באוכלוסיית הנשים ו-3000 ₪ באוכלוסיית הגברים. מה המסקנה ברמת מבוהקות של 5%?

## שאלות

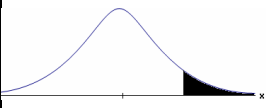
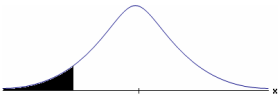
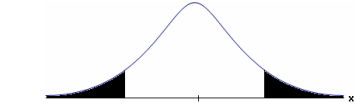
- (1) מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלה נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה. במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות. במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות. לצורך פתרון הניחו שבכל אזור, סטיית התקן היא שעה 1 ביום. בדקו את טענת המחקר ברמת מובהקות של 1%.
- (2) ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 100. מכון ללימוד פסיכומטרי טוען שהוא יכול לשפר את ממוצע הציונים ביותר מ-30 נקודות. במדגם של 20 נבחנים שניגשו למבחן ללא הכנה במכון התקבל ממוצע 508. במדגם של 25 נבחנים שעברו הכנה במכון התקבל ממוצע ציונים 561. מה מסקנתכם ברמת מובהקות של 5%.

**תשובות סופיות**

- (1) נדחה  $H_0$ .
- (2) לא נדחה את  $H_0$ .

## בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

| השערת האפס:<br>השערה אלטרנטיבית:<br>תנאים:                                                                     | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 > c$                                                                                                                                 | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 < c$                                                                                                                                  | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 \neq c$                                                                                                                                                                                                                                                               |
|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. מדגמים בלתי תלויים<br>2. $\sigma_1, \sigma_2$ לא ידועות אך שוות<br>3. המשתנים בכל אוכלוסייה מתפלגים נורמלית | $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$<br><br>$t_{1-\alpha, n_1+n_2-2}$<br>דוחים את $H_0$ | $t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$<br><br>$-t_{1-\alpha, n_1+n_2-2}$<br>דוחים את $H_0$ | $t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ או $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$<br><br>$-t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ $t_{1-\frac{\alpha}{2}, n_1+n_2-2}$<br>דוחים את $H_0$ |

$$t_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

סטטיסטי המבחן:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

השונות המשוקללת:

## חלופה אחרת לכלל הכרעה:

| נדחה $H_0$ אם מתקיים:                                                                                       |                                                                                                                                                                                                                                                      |
|-------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ | $\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$<br>או<br>$\bar{x}_1 - \bar{x}_2 < c - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ |
| $\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ |                                                                                                                                                                                                                                                      |

## דוגמה (פתרון בהקלטה):

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשתמשים בה כיום לבניית בניינים. לצורך בדיקת טענת המחקר נדגמו 10 יחידות של מתכות מהסוג הישן ו-12 יחידות של מתכות מהסוג החדש. להלן תוצאות המדגם:

טמפרטורת ההתכה הממוצעת במתכת הישנה 1170 מעלות עם אומד חסר הטיה לשונות  $S^2 = 200$ .

טמפרטורת ההתכה הממוצעת במתכת החדשה 1317 מעלות עם אומד חסר הטיה לשונות  $S^2 = 260$ .

נניח לצורך פתרון שטמפרטורת ההתכה מתפלגת נורמאלית עם אותה שונות במתכות השונות. בדקו ברמת מובהקות של 5%.

## שאלות

- (1) להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (במ"ר):

|     |    |    |     |    |     |     |      |
|-----|----|----|-----|----|-----|-----|------|
| 120 | 94 | 90 | 130 | 95 | 112 | 120 | 2012 |
|     | 69 | 74 | 105 | 91 | 82  | 100 | 2013 |

בדקו שבשנת 2013 הייתה ירידה משמעותית בשטחי הדירות לעומת שנת 2012 עבור רמת מובהקות של 5%.  
הניחו ששטחי הדירות בכל שנה מתפלגים נורמלית עם אותה שונות.

- (2) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

| המדינה              | ישראל   | ארה"ב   |
|---------------------|---------|---------|
| גודל המדגם          | 15      | 15      |
| סכום הציונים        | 1560    | 1470    |
| סכום ריבועי הציונים | 165,390 | 147,560 |

בדקו ברמת מובהקות של 5% האם קיים הבדל של נקודה בין ישראלים לאמריקאים מבחינת ממוצע הציונים במבחן ה-IQ לטובת ישראל.  
רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

- (3) להלן תוצאות מדגם הבדק אורך חיים של נורות מסוג W60 ומסוג W100. אורך החיים נמדד בשעות.

| הקבוצה    | 60W  | 100W |
|-----------|------|------|
| $\bar{X}$ | 1007 | 956  |
| $S$       | 80   | 72   |
| $n$       | 13   | 15   |

- א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפתרון.  
ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג 100?  
ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.

**תשובות סופיות**

- (1) נדחה את  $H_0$ .
- (2) הנחות:
1. סטיות התקן שוות.
  2. המשתנים מתפלגים נורמלית.
- נקבל את  $H_0$ .
- (3) א. נדחה את  $H_0$ . ב. רמת מובהקות של לפחות 5%.
- ג. לא נדחה את  $H_0$ .

# מבוא לסטטיסטיקה תיאורית והסקית

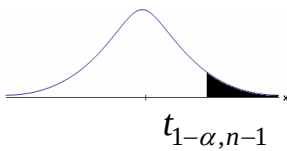
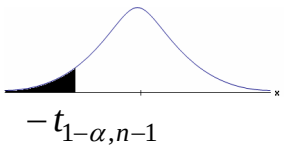
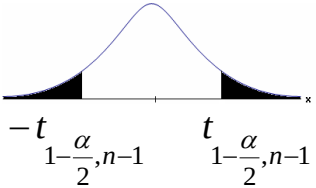
פרק 25 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

תוכן העניינים

1. בדיקת השערות למדגמים מזווגים.....129

## בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

### בדיקת השערות למדגמים מזווגים – רקע

| השערת האפס:<br>השערה אלטרנטיבית:           | $H_0: \mu_D = C$<br>$H_1: \mu_D > C$                                                                                                                                  | $H_0: \mu_D = C$<br>$H_1: \mu_D < C$                                                                                                                                    | $H_0: \mu_D = C$<br>$H_1: \mu_D \neq C$                                                                                                                                                                                                                                      |
|--------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| תנאים:                                     | 1. $\sigma_D$ אינה ידועה<br>2. $D \sim N$ או מדגם מספיק גדול                                                                                                          |                                                                                                                                                                         |                                                                                                                                                                                                                                                                              |
| כלל ההכרעה:<br>אזור הדחייה של $H_0$ :      | $t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$<br><br>$t_{1-\alpha, n-1}$<br>- דוחים את $H_0$ | $t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$<br><br>$-t_{1-\alpha, n-1}$<br>- דוחים את $H_0$ | $t_{\bar{D}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{D}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$<br><br>$-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$<br>- דוחים את $H_0$ |
| חלופה לכלל הכרעה:<br>נדחה $H_0$ אם מתקיים: | $\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$                                                                                                         | $\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$                                                                                                           | $\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$<br>ו<br>$\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$                                                                                                                      |

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}, \quad t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

סטטיסטי המבחן:

דוגמה (פתרון בהקלטה):

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחירה נמוכים מהמחירים מרשת השיווק "שופרסל". לצורך הבדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות. להלן המחירים:

| המוצר / רשת   | מגה בעיר | שופרסל |
|---------------|----------|--------|
| שמפו          | 17       | 18     |
| ג'ל כביסה     | 48       | 57     |
| עוגת גבינה    | 35       | 35     |
| לחם           | 12       | 10     |
| קפה נמס       | 49       | 47     |
| בקבוק יין     | 113      | 142    |
| גבינה בולגרית | 20       | 26     |

בהנחה והמחירים מתפלגים נורמאלית, בדקו ברמת מובהקות של 5% את טענת רשת "מגה בעיר".

## שאלות

- (1) במטרה לבדוק האם קיים הבדל בין חברת  $X$  לחברת  $Y$  מבחינת המחירים לשיחות בינ"ל. נדגמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

| חברה/מדינה | ארה"ב | קנדה | הולנד | פולין | מצרים | סין | יפן |
|------------|-------|------|-------|-------|-------|-----|-----|
| $X$        | 1.5   | 2.1  | 2.2   | 3     | 3.5   | 3.2 | 4.2 |
| $Y$        | 1.4   | 2    | 1.9   | 3.1   | 3.2   | 3.2 | 4.2 |

בהנחה והמחירים מתפלגים נורמלית בכל חברה, בדקו ברמת מובהקות של 5% האם קיים הבדל בין החברות מבחינת המחירים במוצא?

- (2) מכון המכין לפסיכומטרי טוען שהוא מעלה את ממוצע הציונים ביותר מ-30 נקודות. 8 נבחנים נבדקו לפני ואחרי שהם למדו במכון. להלן התוצאות שהתקבלו:

|      |     |     |     |     |     |     |     |     |
|------|-----|-----|-----|-----|-----|-----|-----|-----|
| לפני | 506 | 470 | 420 | 640 | 670 | 390 | 500 | 590 |
| אחרי | 570 | 540 | 430 | 610 | 680 | 510 | 520 | 580 |

מה מסקנתכם ברמת מובהקות 5%? הניחו שציוני פסיכומטרי מתפלגים נורמלית.

- (3) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן הציונים שלהם בסמסטר א' ו- ב':

|              |    |    |    |    |     |
|--------------|----|----|----|----|-----|
| סטטיסטיקה א' | 74 | 68 | 90 | 75 | 82  |
| סטטיסטיקה ב' | 80 | 84 | 87 | 76 | 100 |

פורסם שתלמידים שמסיימים את סמסטר ב' משפרים במוצא את הציונים ב-5 נקודות לעומת סמסטר א'. הניחו שהציונים מתפלגים נורמלית.

- א. מהי מובהקות התוצאה לבדיקת הטענה שהשיפור הוא יותר מ-5 נקודות?  
 ב. על סמך הסעיף הקודם, מהי רמת המובהקות המינימלית להכרעה שהשיפור הוא יותר מ-5 נקודות?  
 ג. לאור זאת, מה המסקנה ברמת מובהקות של 10%?

- (4) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמלית ללא ידיעת השונות האמתית. המבחן שיש לבצע כאן הוא:

- א. מבחן  $Z$  למדגם יחיד.  
 ב. מבחן  $T$  למדגם יחיד.  
 ג. מבחן  $T$  למדגמים בלתי תלויים.  
 ד. מבחן  $T$  למדגמים מזווגים.

- (5) בתחנת טיפת חלב מסוימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשיים נשקל בכל אחד מהמשקלים. להלן תוצאות השקילה (בק"ג):

|               |     |     |     |     |
|---------------|-----|-----|-----|-----|
| משקל במכשיר 1 | 2.5 | 0.7 | 9.6 | 4.5 |
| משקל במכשיר 2 | 0.5 | 1.7 | 6.9 | 3.5 |

נניח שהמשקלים מתפלגים נורמלית, המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
  - מבחן T למדגם יחיד.
  - מבחן T למדגמים בלתי תלויים.
  - מבחן T למדגמים מזווגים.
- (6) כדי להשוות בין שני אצנים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים. המבחן שיש לבצע כאן הוא:
- מבחן Z למדגם יחיד.
  - מבחן T למדגם יחיד.
  - מבחן T למדגמים בלתי תלויים.
  - מבחן T למדגמים מזווגים.

### תשובות סופיות

- לא נדחה  $H_0$ .
- לא נדחה  $H_0$ .
- א.  $0.25 \leq p \leq 0.5$     ב. 0.5    ג. לא נדחה  $H_0$ .
- ד'.
- ד'.
- ג'.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 26 - הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות

תוכן העניינים

1. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות ..... 133

## הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

### רקע

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות  $\alpha$  על  $\mu_1 - \mu_2$  :

$$H_0: \mu_1 - \mu_2 = C, \quad H_1: \mu_1 - \mu_2 \neq C$$

על ידי בניית רווח סמך ברמת סמך של  $1 - \alpha$  ל-  $\mu_1 - \mu_2$  :

אם  $C$  נופל ברווח  $\leftarrow$  נקבל את  $H_0$ .

אם  $C$  לא נופל ברווח  $\leftarrow$  נדחה את  $H_0$ .

### דוגמה (פתרון בהקלטה) :

חוקר ביצע בדיקת השערות לתוחלת ההפרש במדגם מזווג.

להלן השערותיו:  $\alpha = 5\%$ ,  $H_0: \mu_D = 80$ ,  $H_1: \mu_D \neq 80$ .

החוקר בנה רווח סמך ברמה של 90%,  $78 < \mu_D < 83$ .

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

## שאלות

(1) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'.

להלן ציוניהם בסמסטר א' ו- ב' :

| סמסטר א' | סמסטר ב' |
|----------|----------|
| 74       | 80       |
| 68       | 84       |
| 90       | 87       |
| 75       | 76       |
| 82       | 100      |

א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א' לבין סמסטר ב'.

ב. פורסם שתלמידים שמסיימים את סמסטר ב' משפרים בממוצע את הציונים ב-5 נק' לעומת סמסטר א'. האם יש אמת בפרסום?

(2) הוחלט להשוות הציונים אצל מרצה  $X$  ואצל מרצה  $Y$ . נבחרו באקראי 6 סטודנטים, 3 סטודנטים של מרצה  $X$  ו-3 סטודנטים של מרצה  $Y$ , עבורם התקבלו הציונים הבאים :

|          |    |    |    |
|----------|----|----|----|
| מרצה $X$ | 82 | 90 | 68 |
| מרצה $Y$ | 68 | 81 | 64 |

א. חשבו רווח סמך ברמת סמך 90% להפרש בין התוחלות של הציונים אצל שני המרצים.

ב. האם ברמת מובהקות של 10% נכריע שיש הבדל בין תוחלות הציונים אצל שני המרצים?

## שאלות רב-ברירה :

- (3) סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים.
- הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח  $-2 < \mu_1 - \mu_2 < 4.5$ .
- אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות :
- $H_0 : \mu_1 - \mu_2 = 0$  ;  $H_1 : \mu_1 - \mu_2 \neq 0$  ברמת מובהקות 0.05, מסקנתו תהיה :
- א. לדחות את השערת האפס.
- ב. לא לדחות את השערת האפס.
- ג. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.
- ד. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

- (4) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו :  $-0.0293 < \mu_D < 0.2145$ , רווח הסמך הוא ברמת סמך של 95%.  
לכן מסקנת המחקר היא :
- ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
  - ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
  - לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של  $D$ .

### תשובות סופיות

- (1) א.  $-3.8 \leq \mu_D \leq 19$  ב. נכריע שיש אמת בפרסום.
- (2) א.  $-8.5 \leq \mu_X - \mu_Y \leq 26.5$  ב. נכריע שאין הבדל.
- (3) ג'.
- (4) א'.

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 27 - שאלות מסכמות על רווחי סמך

תוכן העניינים

1. שאלות מסכמות על רווחי סמך.....136

## שאלות מסכמות על רווחי סמך:

### שאלות:

- (2) 200 אנשים נשאלו כמה פעמים ביום הם שותים כוס קפה. להלן התפלגות התשובות:

| מספר פעמים | 0  | 1  | 2  | 3  | 4  | 5  |
|------------|----|----|----|----|----|----|
| מספר אנשים | 86 | 34 | 28 | 22 | 20 | 10 |

- א. תנו רווח סמך לממוצע מספר כוסות הקפה שאנשים נוהגים לשתות ביום.  $\alpha = 0.05$ .
- ב. אדם השותה לפחות 4 כוסות קפה ביום נקרא "מכור לקפה". בנו רווח סמך לאחוז "המכורים לקפה".  $\alpha = 0.1$ .
- (3) חוקר בנה רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת בשנה. רווח הסמך שהתקבל הוא:  $81 < p < 91$ . רווח הסמך הנ"ל התבסס על מדגם של 500 איש.
- א. כמה אנשים במדגם טענו שכלל לא התקררו השנה?
- ב. באיזו רמת סמך נבנה רווח הסמך?
- ג. בנו רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת השנה ברמת סמך של 96% על סמך תוצאות המדגם.
- (4) ציוני IQ בארה"ב מתפלגים נורמאלית עם תוחלת 100. במדגם של 20 ישראלים שנבחנו במבחן ה-IQ התקבלו התוצאות הבאות:
- $$\sum_{i=1}^{20} x_i = 2040, \quad \sum_{i=1}^{20} x_i^2 = 210740$$
- א. אמדו ברמת ביטחון של 90% את ממוצע ציוני בחינת ה-IQ בישראל – מהי ההנחה הדרושה לפתרון?
- ב. על סמך רווח הסמך של סעיף א' האם תקבלו את הטענה שבישראל ממוצע הציונים שונה מארה"ב?
- ג. מה היה קורה לרווח הסמך אם היינו מגדילים את רמת הסמך שלו?

(8) להלן מדגם של שכר הדירה ב-9 של 5 דירות שלושה חדרים בשכונת בבלי בתל אביב:

|      |      |      |      |      |          |
|------|------|------|------|------|----------|
| 7500 | 6500 | 7000 | 7500 | 8000 | שנת 2012 |
| 7700 | 6800 | 7800 | 8200 | 8000 | שנת 2013 |

בנו רווח סמך ברמת סמך של 95% לתוחלת עליית שכר הדירה משנת 2012 לשנת 2013 בשכונת בבלי. ניתן להניח ששכר הדירה בשכונה מתפלג נורמלית.

### תשובות סופיות:

- (2) א.  $1.21 \leq \mu \leq 1.65$       ב.  $10.85\% \leq p \leq 19.15\%$
- (3) א. 70      ב. 0.9988      ג.  $83\% < p < 89\%$
- (4) א.  $97.4 \leq \mu \leq 106.6$       ב. לא.      ג. יגדל.
- (8)  $-21 \leq \mu_D \leq 821$

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 28 - שאלות מסכמות בבדיקת השערות

תוכן העניינים

1. שאלות פתוחות מסכמות..... 138
2. שאלות רב ברירה ( אמריקאיות)..... 140

## שאלות מסכמות בבדיקת השערות על פרמטרים

### שאלות

- (1) שני חוקרים נתבקשו לבדוק את ההשערות הבאות:  $H_0: \mu = 520$ ,  $H_1: \mu > 520$ . כל חוקר בדק מדגם של 225 נחקרים. ידוע ש- $\sigma = 20$ . חוקר א' קבע את כלל ההכרעה לפי  $\alpha = 0.05$ . חוקר ב' מחליט לדחות  $H_0$  אם  $\bar{X} > 522$ .

- למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?
- מהי ההסתברות לטעות מסוג שני של חוקר ב' עבור  $\mu_1 = 525$ ?
- הסבר ללא חישוב נוסף, האם ההסתברות לטעות מסוג שני עבור  $\mu_1 = 525$ , של חוקר א' שווה/קטנה/גדולה לזו של חוקר ב'?
- חוקר א' קיבל במדגם שלו  $\bar{X} = 523$ . מהי מסקנתו?

- (2) ידוע כי תוחלת מספר הלייקים היומי של דנה היא 12 עם סטיית תקן 5. דני טוען שהוא יותר פופולארי מדנה בכך שהוא מקבל יותר לייקים מדנה ביום. על-מנת לבדוק זאת ספר דני כמה לייקים הוא קיבל בכל יום במהלך 7 שבועות (כלומר, ב-49 ימים) וקיבל סך-הכול 637 לייקים. נניח כי סטיית התקן של מספר הלייקים שדני מקבל ביום זהה לסטיית התקן של דנה.
- מהי רמת המובהקות שכדאי לדני לדרוש, כדי שדנה תשתכנע בצדקת טענתו (שדני פופולרי יותר בכך שהוא מקבל יותר לייקים מדנה ביום)?
  - אם דני משער שתוחלת מספר ה"לייקים" שהוא מקבל ביום היא 14 וקובע רמת מובהקות 2.5%, מהי עוצמת המבחן של דני?

| מוצר / רשת | A | B |
|------------|---|---|
| 1          | 5 | 5 |
| 2          | 4 | 5 |
| 3          | 5 | 3 |
| 4          | 7 | 4 |

- (3) ברצוננו להשוות בין רשתות A לבין B. לשם כך בחרנו 4 מוצרים, ובדקנו את מחיריהם בשתי הרשתות. להלן התוצאות:
- הניחו כי המחירים מתפלגים נורמלית.
- אם יש הנחות נוספות כדי לבצע את המבחן הפרמטרי רשמו אותן.
- בדקו האם קיים הבדל בין הרשתות מבחינת תוחלת המחירים. רמת מובהקות של 5%.
  - חזרו על הסעיף הקודם בהנחה ונבחרו בכל רשת מוצרים באקראי ולא בהכרח אותם מוצרים.

(4) במטרה לבדוק האם סטודנטים הלומדים במכללות משקיעים יותר זמן ללימודים מאשר סטודנטים באוניברסיטאות נדגמו 12 סטודנטים ובדקו לכל סטודנט את הזמן שהוא משקיע ביום ללימודים. הזמנים נמדדו בדקות:

|     |     |     |     |     |     |                        |
|-----|-----|-----|-----|-----|-----|------------------------|
| 180 | 140 | 171 | 189 | 156 | 176 | סטודנטים באוניברסיטאות |
| 150 | 204 | 186 | 191 | 190 | 180 | סטודנטים במכללות       |

א. נסחו את ההשערות ובדוק אותן ברמת מובהקות של 5%. רשום את כלל ההכרעה ואת ההנחות הדרושות לביצוע המבחן הפרמטרי.

ב. חשבו את p-value.

ג. ישנה טענה שממוצע זמן ההשקעה בלימודים במכללות הוא 3.5 שעות ביום. בדוק את הטענה כאשר רמת המובהקות הינה 5%.

(5) נתון כי:  $X \sim N(\mu, \sigma^2 = 10^2)$ .

מעוניינים לבדוק את ההשערות:  $H_0: \mu = 40$ ,  $H_1: \mu > 40$ .

נדגמו 25 תצפיות מהאוכלוסייה והתקבל  $\bar{X} = 45$ .

א. חשבו את p-value (מובהקות התוצאה).

ב. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה:  $H_1: \mu < 40$ .

ג. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה:  $H_1: \mu \neq 40$ .

## תשובות סופיות

(1) א. חוקר א' ב. 0.0122 ג. גדלה. ד. נדחה  $H_0$ .

(2) א. לפחות 0.0808 ב. 0.7995.

(3) א. לא נדחה  $H_0$ . ב. לא נדחה  $H_0$ .

(4) א. לא נדחה  $H_0$ . ב. בין 5% ל-10%. ג. נדחה  $H_0$ .

(5) א. 0.0062 ב. 0.9938 ג. 0.0124.

### שאלות סיכום – שאלות רב ברירה על בדיקת השערות

(1) בבדיקת השערה חד-צדדית ימנית ברמת מובהקות  $\alpha = 0.01$ , נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות אותם נתונים ברמת מובהקות  $\alpha = 0.05$ ?

- השערת האפס הייתה נדחית.
- השערת האפס לא הייתה נדחית.
- ההשערה המחקרית הייתה נדחית.
- בהעדר נתונים נוספים, לא ניתן לדעת.

(2) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמיתית. המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
- מבחן T למדגם יחיד.
- מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים מזווגים.

(3) כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות:

מהן ההשערות הנבדקות? (ההפרש חושב  $X - Y$ )

|    |    |    |    |                   |
|----|----|----|----|-------------------|
| 68 | 82 | 93 | 69 | לפני הנישואין - X |
| 71 | 84 | 88 | 80 | לאחר הנישואין - Y |

- $H_1: \mu_d < 0, H_0: \mu_d = 0$
- $H_1: \mu_X - \mu_Y < 0, H_0: \mu_X - \mu_Y = 0$
- $H_1: \mu_X - \mu_Y < 0, H_0: \mu_X - \mu_Y = 0$
- $H_1: \mu_d > 0, H_0: \mu_d = 0$

(4) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן:

- השערת האפס נכונה.
- השערת האפס נדחתה.
- השערת האפס לא נדחתה.
- אף אחת מהתשובות לא נכונה בהכרח.

(5) ידוע כי ילד בגיל שנתיים ישן בממוצע 9 שעות בלילה. במדגם של 20 תינוקות בני שנתיים המתגוררים בצפון נמצא, כי ממוצע שעות השינה בלילה הינו 10 עם סטיית תקן של 1.1 במדגם של 10 תינוקות בדרום נמצא, כי ממוצע שעות השינה בלילה הינו 7.9 עם סטיית תקן של 1.1. על מנת להשוות בין ממוצע שעות השינה של ילדים מהצפון לבין זה של כלל הילדים יש לערוך \_\_\_\_\_, ועל מנת להשוות בין ממוצע שעות השינה של ילדים מהדרום לזה של ילדים המתגוררים בצפון יש לערוך \_\_\_\_\_.

יש להניח שההנחות הדרושות מתקיימות.

- א. מבחן Z למדגם יחיד ; מבחן T למדגם יחיד.
- ב. מבחן T למדגם יחיד ; מבחן T למדגמים תלויים.
- ג. מבחן T למדגם יחיד ; מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים בלתי תלויים ; מבחן T לממוצע יחיד.

(6) מובהקות התוצאה (PV) היא גם :

- א. רמת המובהקות המינימאלית לדחות השערת האפס.
- ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
- ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.

(7) כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות :

|    |    |    |    |               |
|----|----|----|----|---------------|
| 68 | 82 | 93 | 69 | לפני הנישואין |
| 71 | 84 | 88 | 80 | לאחר הנישואין |

באיזה התפלגות משתמשים לבדיקת ההשערות, ובכמה דרגות חופש :

- א. ההתפלגות Z ללא דרגות חופש.
- ב. ההתפלגות T ו-3 דרגות חופש.
- ג. ההתפלגות T ו-6 דרגות חופש.
- ד. ההתפלגות  $\chi^2$  ו-3 דרגות חופש.

- (8) שני סטטיסטיקאים בודקים השערות ברמת מובהקות  $\alpha = 0.05$  על סמך אותו מדגם. סטטיסטיקאי א' בודק את ההשערה:  $H_0: \mu = 20$  כנגד האלטרנטיבה  $H_1: \mu \neq 20$  ומחליט לא לדחות את השערת האפס. סטטיסטיקאי ב' בודק את ההשערה  $H_0: \mu \leq 20$  כנגד האלטרנטיבה  $H_1: \mu > 20$  מה יחליט סטטיסטיקאי ב'?
- לדחות את השערת האפס.
  - לא לדחות את השערת האפס.
  - ללא נתונים נוספים אי אפשר לדעת מה יחליט.
- (9) חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?
- הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
  - הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
  - הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.
  - הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
- (10) רמת הכולסטרול בדמם של אנשים מתפלג נורמאלית עם תוחלת של 180 מ"ג (ל 100 סמ"ק דם). וסטיית תקן של 10 מ"ג. מעוניינים לבדוק את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר. נניח שסטיית התקן אצל צמחונים זהה לסטיית התקן של כלל האנשים. במדגם של 20 צמחונים התקבל ממוצע רמת כולסטרול 174.5 מ"ג. אם הוחלט לקבל את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר איזה סוג טעות אפשרית במסקנה?
- טעות מסוג ראשון.
  - טעות מסוג שני.
  - טעות מסוג שלישי.
  - לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמתית אצל הצמחונים.

- 11** שני חוקרים העוסקים בתחום מחקרי משותף החליטו להסתמך על נתונים של מדגם שפורסם על ידי הלשכה המרכזית לסטטיסטיקה. חוקר א' ניסח השערה דו צדדית ואילו חוקר ב' ניסח השערה חד צדדית. מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?
- אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב' באותה רמת מובהקות.
  - אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס באותה רמת מובהקות.
  - אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת האפס באותה רמת מובהקות.
  - אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת האפס בתנאי שרמת המובהקות כפולה בגודלה.
- 12** ידוע מנתוני העבר כי תוחלת הציונים בבחינה בפסיכולוגיה היא 79. הועלתה השערה כי תוחלת הציונים בקרב העולים החדשים נמוכה יותר. לצורך בדיקת הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75. מה משמעות הפרמטר בניסוח ההשערות?
- תוחלת ציוני העולים באוכלוסייה.
  - ממוצע ציוני העולים במדגם.
  - תוחלת ציוני האוכלוסייה מנתוני העבר.
  - ממוצע ציוני שאר האוכלוסייה במדגם.
- 13** חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?
- החוקר דחה את השערת  $H_0$  כאשר היא הייתה נכונה.
  - החוקר דחה את השערת  $H_1$  כאשר היא הייתה נכונה.
  - החוקר לא דחה את השערת  $H_0$  כאשר היא הייתה לא נכונה.
  - המדגם של החוקר שייד בפועל להתפלגות הדגימה של  $H_1$ .
- 14** חוקר ביקש לבחון האם תאומים זהים אשר הופרדו בילדותם שונים מתאומים זהים אשר גדלו יחדיו מבחינת מידת הפער בין התאומים בלחץ הדם. הוא דגם 20 זוגות תאומים מכל אוכלוסייה ומדד את הפרש בין לחץ הדם בכל זוג תאומים. מהו המבחן הסטטיסטי המתאים?
- מבחן T למדגמים בלתי תלויים עם 38 דרגות חופש.
  - מבחן T למדגמים מזווגים, עם 39 דרגות חופש.
  - מבחן T למדגמים בלתי תלויים עם 39 דרגות חופש.
  - מבחן T למדגמים מזווגים עם 38 דרגות חופש.

**15** שלושה חוקרים רצו לבדוק את השפעתו של שידור פרסומות נגד תאונות דרכים על מהירות הנהיגה של נהגים בישראל (השונות של מהירות הנהיגה בישראל אינה ידועה). עידו השווה את מהירות הנהיגה של קבוצת נהגים אחת, חודש לפני שידור הפרסומות וחודש לאחר שידור הפרסומות. רון השווה את מהירות הנהיגה של קבוצת נהגים, שראו את הפרסומות, למהירות הנהיגה של קבוצת נהגים, שלא ראו את הפרסומות. יואב השווה את מהירות הנהיגה של קבוצת נהגים בחודש בו שודרו הפרסומות, למהירות הנהיגה הממוצעת בישראל על פי נתוני משרד התחבורה. המבחנים בהם צריכים החוקרים להשתמש הם:

- שלושתם במבחן T למדגמים בלתי תלויים.
- עידו במבחן T למדגמים מזווגים, ורון ויואב במבחן T למדגמים בלתי תלויים.
- עידו במבחן T למדגמים מזווגים, רון במבחן T למדגמים בלתי תלויים ויואב במבחן T למדגם יחיד.
- עידו במבחן T למדגמים מזווגים, רון ויואב במבחן T למדגם יחיד.

**16** במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5%. מה תמיד נכון?

- הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
- הגדלת רמת המובהקות תשנה את מסקנת המחקר.
- הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
- הקטנת רמת המובהקות תשנה את מסקנת המחקר.

**17** חוקר ערך מבחן דו צדדי ברמת מובהקות של  $\alpha$  והחליט לדחות את השערת האפס. אם החוקר היה עורך מבחן חד צדדי ברמת מובהקות של  $\frac{\alpha}{2}$  אזי בהכרח:

- השערת האפס הייתה נדחית.
- השערת האפס הייתה לא נדחית.
- לא ניתן לדעת מה תהיה מסקנתו במקרה זה.

**18** ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם. ליאור השתמש בטבלה של התפלגות Z. רוני השתמש בטבלה של התפלגות T. מה נוכל לומר בנוגע להחלטת המחקר שלהם?

- אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
- אם רוני ידחה את השערת האפס אז גם בהכרח ליאור.
- שני החוקרים בהכרח יגיעו לאותה מסקנה.
- לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

(19) נתון ש  $X \sim N(\mu, \sigma^2)$  כמו כן נתונות ההשערות הבאות:  $H_0: \mu = \mu_0$ ,  $H_1: \mu < \mu_0$ .

- חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות.  $\sigma^2$  לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות.
- א. כעת בברור הוא ידחה את השערת האפס.
  - ב. כעת הוא דווקא יקבל את השערת האפס.
  - ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

(20) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:

- א. הסיכוי לטעות מסוג ראשון גדל.
- ב. העוצמה של המבחן גדלה.
- ג. הסיכוי לטעות מסוג שני גדל.
- ד. תשובות א ו-ב נכונות.

(21) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן:

- א. השערת האפס נכונה.
- ב. השערת האפס נדחתה.
- ג. השערת האפס לא נדחתה.
- ד. אף אחת מהתשובות לא נכונה בהכרח.

(22) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

|             |          |
|-------------|----------|
| $1 - \beta$ | $\alpha$ |
| א. גדולה    | גדולה    |
| ב. גדולה    | קטנה     |
| ג. קטנה     | גדולה    |
| ד. קטנה     | קטנה     |

(23) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית  $H_0$  קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך:

- א. הן  $\alpha$ , והן  $(1 - \beta)$ , יקטנו.
- ב.  $\alpha$  יישאר ללא שינוי ואילו  $(1 - \beta)$  יגדל.
- ג.  $\alpha$  יגדל ואילו  $(1 - \beta)$  יקטן.
- ד. הן  $\alpha$  והן  $(1 - \beta)$  יגדלו.

**(24)** ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח שלחץ הדם בקרב עיתונאים גבוה יותר מהממוצע באוכלוסייה. הוא לקח מדגם של 60 עיתונאים וקיבל ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?

- א. טעות מסוג ראשון.
- ב. טעות מסוג שני.
- ג. טעות מסוג שלישי.
- ד. אין טעות במסקנתו.

**(25)** בבדיקת השערות התקבל ש-  $p\text{-value} = 0.02$ . מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה:

- א. יקבל את השערת האפס בכל מקרה.
- ב. ידחה את השערת האפס מקרה.
- ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
- ד. לא ניתן לדעת כי אין מספיק נתונים.

**(26)** מובהקות התוצאה (PV) היא גם:

- א. רמת המובהקות המינימאלית לדחות השערת האפס.
- ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
- ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.

**(27)** בבדיקת השערות מסוימת התקבל  $p\text{ value} = 0.0254$ , לכן:

- א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את  $H_0$ .
- ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את  $H_0$ .
- ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את  $H_0$ .
- ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את  $H_0$ .

**(28)** רמת המובהקות במחקר הייתה 2% לכן.

- א. בסיכוי של 2% נדחה את השערת האפס.
- ב. בסיכוי של 2% לא נדחה את השערת האפס.
- ג. בסיכוי של 2% השערת האפס לא נכונה.
- ד. אף תשובה לא נכונה.

- (29) נתון ש:  $X \sim N(\mu, \sigma^2)$ . כמו כן נתונות ההשערות הבאות:  $H_0: \mu = \mu_0$ ,  $H_1: \mu < \mu_0$ . חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות.  $\sigma^2$  לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5%. אם הוא היה מגדיל את רמת המובהקות ל-10% אזי:
- א. כעת בברור הוא ידחה את השערת האפס.
  - ב. כעת הוא דווקא יקבל את השערת האפס.
  - ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

- (30) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית. מספר דרגות החופש במבחן הוא:

- א. 9
- ב. 19
- ג. 18
- ד. 8

- (31) בתחנת טיפת חלב מסוימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשיים נשקל בכל אחד מהמשקלים. להלן תוצאות השקילה (בק"ג):

|               |     |     |     |     |
|---------------|-----|-----|-----|-----|
| משקל במכשיר 1 | 2.5 | 0.7 | 9.6 | 4.5 |
| משקל במכשיר 2 | 0.5 | 1.7 | 6.9 | 3.5 |

- נניח שהמשקלים מתפלגים נורמלית. המבחן שיש לבצע כאן הוא:
- א. מבחן Z למדגם יחיד.
  - ב. מבחן T למדגם יחיד.
  - ג. מבחן T למדגמים בלתי תלויים.
  - ד. מבחן T למדגמים מזווגים.

- (32) כדי להשוות בין שני אצים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים. המבחן שיש לבצע כאן הוא:

- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים מזווגים.

**(33)** סטטיסטיקאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0. המשמעות של תוצאה זו היא:

- לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
- תמיד לדחות את השערת האפס כאשר היא נכונה.
- לעולם לא לדחות את השערת האפס כאשר היא נכונה.
- תמיד לדחות את השערת האפס כאשר היא לא נכונה.

**(34)** סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים. הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח  $-2 < \mu_1 - \mu_2 < 4.5$ . אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות:  $H_0: \mu_1 - \mu_2 = 0$ ;  $H_1: \mu_1 - \mu_2 \neq 0$ ,

- ברמת מובהקות 0.05 מסקנתו תהיה:
- לדחות את השערת האפס.
  - לא לדחות את השערת האפס.
  - שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.
  - שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

**(35)** במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפללים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו:  $-0.0293 < \mu_D < 0.2145$ . רווח הסמך הוא ברמת סמך של 95%. לכן מסקנת המחקר היא:

- ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
- ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
- לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של  $D$ .

**(36)** אם רמת מובהקות של מבחן סטטיסטי הינה 0, הכוונה היא:

- תמיד נדחה  $H_0$  כאשר היא נכונה, אך לא תמיד נדחה אותה כאשר היא לא נכונה.
- לא נדחה את  $H_0$  אף פעם.
- לא נדחה את  $H_0$  כאשר היא נכונה אך יתכן ונדחה אותה כאשר היא לא נכונה.
- כל התשובות לא נכונות.

37) חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות:  $H_0: \mu = 10$ ,  $H_1: \mu \neq 10$ .

לצורך בדיקה הוא לקח מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך תוצאות המדגם הוא חישב וקיבל:  $t_x = -2.63$ . לכן המסקנה היא:

- הוא ידחה  $H_0$  ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
- הוא ידחה  $H_0$  ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025.
- הוא ידחה  $H_0$  ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01.
- הוא לא ידחה  $H_0$  ברמת מובהקות 0.1.

38) האיגוד האמריקני לרפואת ילדים מפרסם הנחיות חדשות הקובעות כי יש ליטול תוספת יוד במהלך תקופת ההיריון וההנקה. מחסור במינרל זה עלול לגרום לפגיעה מוחית אצל העובר והתינוק. החלטה זו נקבעה על סמך מחקר בו השתתפו 1050 נשים שנטלו יוד במהלך תקופת ההיריון וההנקה. מתוך הנשים שהשתתפו במחקר, רק ל-21 נמצאו ילדים בעלי פגיעה מוחית לעומת 3% באוכלוסייה הכללית. בנוסף, פורסם שהאיגוד האמריקאי מגיע למסקנותיו על סמך רמת מובהקות של 0.5%. מה הסיכוי לבצע טעות מסוג ראשון במחקר?

- 0.005
- 0.03
- 0.0287
- 0.05

39) חוקרת שיערה, כי משקלן של נשים כשנה לאחר החתונה גבוה ממשקלן בעת החתונה. החוקרת דגמה 15 נשים, ובדקה את משקלן בשתי נקודות הזמן (בעת החתונה, ושנה לאחריה), אך לא מצאה הבדל מובהק ברמת מובהקות 0.01. בהנחה, כי במצאות השערתה של החוקרת נכונה, סביר כי אם היא תגדיל את גודל המדגם, אזי:

- יקטן הסיכוי לטעות מסוג שני ( $\beta$ ).
- תגדל רמת הביטחון ( $1 - \alpha$ ).
- אף תשובה לא נכונה.
- כל התשובות נכונות.

40) איזה מהמשפטים הבאים נכון תמיד?

- $POWER + \alpha + \beta = 1$
- $POWER = 0.5 - \beta$
- $POWER + \alpha = 1$
- $\beta + \alpha = 1$
- הכול לא נכון.

- 41** מה נכון לומר לגבי הנחת שיוויון השונויות במבחן  $T$  למדגמים בלתי תלויים?
- היא אומרת שהשונויות המדגמיות שוות.
  - בלעדיה אין שום דרך לבדוק השערה על הפרש בין תוחלות.
  - היא חשובה הן עבור מדגמים מזווגים והן עבור מדגמים בלתי תלויים.
  - אף תשובה אינה נכונה.
- 42** חוקר החליט לא לדחות השערה ברמת מובהקות של  $\alpha$ . במידה וחוקר זה היה בודק השערה זו ברמת מובהקות של  $2\alpha$  על סמך אותם נתונים, האם ההשערה תדחה?
- ההשערה תדחה.
  - ההשערה לא תדחה.
  - התשובה תלויה בעוצמת המבחן.
  - לא ניתן לדעת בוודאות אם ההשערה תדחה או לא.
- 43** חוקרת שיערה, כי בגילאי הגן בנות יותר תקשורתיות מבנים. אם החוקרת תדגום אקראית 30 בנים ו-30 בנות, ובמדגם יתקבל אותו ממוצע של ציון תקשורת. סטטיסטי המבחן יהיה:
- אפס
  - חיובי
  - שלילי
  - לא ניתן לדעת
- 44** עוצמה שווה ל-1 פרושה:
- לעולם לא לדחות את השערת האפס כאשר היא נכונה.
  - תמיד לדחות את השערת האפס כאשר היא נכונה.
  - לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
- 45** מה מהבאים נכון לגבי מבחן  $T$  מדגמים מזווגים?
- כל התצפיות במחקר אינן תלויות זו בזו.
  - כל התצפיות במחקר תלויות זו בזו.
  - כל הצמדים של תצפיות במחקר אינם תלויים זה בזה.
  - התצפיות בתוך כל צמד אינן תלויות זו בזו.

- (46)** לבדיקת ההשערה החד צדדית על התוחלת של התפלגות נורמלית  $H_0: \mu \geq 10$ ,  $H_1: \mu < 10$ . נלקח מדגם והתקבלה רמת מובהקות מינימאלית לדחיית השערת האפס 0.058. לו רצינו לבדוק את ההשערה הדו צדדית  $H_0: \mu = 10$ ,  $H_1: \mu \neq 10$  אז על סמך תוצאת אותו המדגם ברמת מובהקות 0.05:
- ניתן להכריע בין ההשערות רק אם שונות האוכלוסייה נתונה.
  - מקבלים את השערת האפס.
  - דוחים את השערת האפס.
  - לא ניתן להכריע בין ההשערות שכן חסרים נתונים.

- (47)** לבדיקת ההשערה החד צדדית ימנית  $H_0: \mu = 55$ ,  $H_1: \mu = 65$ . נלקח מדגם מקרי בגודל  $n$  מאוכלוסייה בעלת התפלגות נורמלית ושונות  $\sigma^2$ . רמת המובהקות היא 5%. נמצא שהעוצמה היא 0.9. להלן 3 טענות:
- עבור מדגם בגודל  $n$  וברמת מובהקות 5% לבדיקת ההשערות:  $H_0: \mu = 55$ ,  $H_1: \mu = 60$  העוצמה תהיה גדולה מ-0.9.
  - עבור מדגם בגודל  $2n$  ורמת מובהקות 5% לבדיקת ההשערות:  $H_0: \mu = 55$ ,  $H_1: \mu = 65$  העוצמה תהיה גדולה מ-0.9.
  - עבור מדגם בגודל  $n$  ורמת מובהקות 10% לבדיקת ההשערות:  $H_0: \mu = 55$ ,  $H_1: \mu = 65$  העוצמה תהיה קטנה מ-0.9.
- שלושת הטענות אינן נכונות.
  - טענות 2 ו-3 אינן נכונות וטענה 1 נכונה.
  - טענות 1 ו-2 נכונות וטענה 3 אינה נכונה.
  - טענות 1 ו-3 אינן נכונות וטענה 2 נכונה.

## תשובות סופיות:

| שאלה | תשובה | שאלה | תשובה |
|------|-------|------|-------|
| 1    | א     | 25   | א     |
| 2    | ד     | 26   | א     |
| 3    | א     | 27   | ג     |
| 4    | ג     | 28   | ד     |
| 5    | ג     | 29   | א     |
| 6    | א     | 30   | א     |
| 7    | ב     | 31   | ד     |
| 8    | ג     | 32   | ג     |
| 9    | א     | 33   | א     |
| 10   | א     | 34   | ג     |
| 11   | א     | 35   | א     |
| 12   | א     | 36   | ג     |
| 13   | א     | 37   | א     |
| 14   | א     | 38   | א     |
| 15   | ג     | 39   | א     |
| 16   | א     | 40   | ה     |
| 17   | ג     | 41   | ד     |
| 18   | ב     | 42   | ד     |
| 19   | ג     | 43   | א     |
| 20   | ד     | 44   | ד     |
| 21   | ג     | 45   | ג     |
| 22   | ג     | 46   | ב     |
| 23   | א     | 47   | ד     |
| 24   | ב     |      |       |

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 29 - מבחני חי בריבוע (מדד הקשר של קרמר)

תוכן העניינים

1. מבחן לאי תלות..... (ללא ספר)
2. מקדם המתאם של קרמר..... 153

## מדדי קשר-מדד הקשר של קרמר – רקע

מתי משתמשים במדד הזה? - כאשר אחד המשתנים הוא מסולם שמי והשני מכל סולם אפשרי. מדד הקשר מקבל ערכים בין 0 ל-1. ככל שהמדד יותר קרוב לאחד קיים קשר בעוצמה יותר חזקה בין המשתנים.

### דוגמה (פתרון בהקלטה) :

במחקר רוצים לבדוק את הקשר בין מין לדעה בנושא מסוים, שאלו 100 גברים ו-100 נשים האם הם בעד/נגד/נמנעים באיזשהו נושא. להלן טבלת השכיחויות המשותפת שהתקבלה.

| $f(x)$    | נמנע | נגד | בעד | $X/Y$  |
|-----------|------|-----|-----|--------|
| 100       | 10   | 40  | 50  | גבר    |
| 100       | 10   | 60  | 30  | אישה   |
| $n = 200$ | 20   | 100 | 80  | $f(y)$ |

בהקשר של קרמר הטבלה נקראת טבלת O (observed)

$X$  - מין (גבר/אישה) – סולם שמי.

$Y$  - דעה (בעד/נמנע/נגד) – סולם שמי/סדר.

### שלבים בחישוב $r_c$ :

#### שלב א' :

נבנה את טבלת E (Expected)

נעתיק את המסגרת של טבלת O ואז כל  $E_i = \frac{f(x) \cdot f(y)}{n}$ .

| $f(x)$  | נמנע | נגד | בעד | $X/Y$  |
|---------|------|-----|-----|--------|
| 100     |      |     |     | גבר    |
| 100     |      |     |     | אישה   |
| $n=200$ | 20   | 100 | 80  | $f(y)$ |

#### שלב ב' :

נחשב  $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$ .

#### שלב ג' :

נחשב:  $r_c = \sqrt{\frac{1}{n(L-1)} \chi^2}$

כאשר  $L$  מבטא את המספר הקטן מבין מספר השורות או העמודות.

## שאלות

- (1) להלן תוצאות מחקר שבדק את הקשר בין מין להשכלה. לגבי כל נחקר נבדק המין שלו והשכלתו.

| מין / השכלה | נמוכה | תיכונית | גבוהה |
|-------------|-------|---------|-------|
| גבר         | 120   | 40      | 20    |
| אישה        | 20    | 20      | 80    |

להלן התוצאות:

האם קיים קשר בין מין להשכלה? נמקו!

- (2) נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.

א. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.

ב. האם קיים קשר בין פעילות גופנית למצב בריאותי?

חשבו לפי מדד הקשר של קרמר.

## תשובות סופיות

- (1) קיים קשר בעוצמה בינונית בין המין להשכלה מקדם המתאם של קרמר הוא 0.595.
- (2) א. להלן טבלה: ב. מדד קרמר 0.19 מעיד על קשר בעוצמה נמוכה.

| $y/x$  | תקין | לא תקין | $f(x)$ |
|--------|------|---------|--------|
| כן     | 50   | 10      | 60     |
| לא     | 90   | 50      | 140    |
| $f(y)$ | 140  | 60      | 200    |

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 30 - מקדם המתאם ( מדד קשר ) הלינארי ומובהקותו

תוכן העניינים

1. מקדם המתאם הלינארי ( פירסון) ..... 155
2. חישוב מקדם המתאם הלינארי (פירסון) ..... 166
3. בדיקת השערות על מקדם המתאם הלינארי..... 171

## מקדם המתאם (מדד קשר) הלינארי ומובהקותו

### מדד הקשר הלינארי (פירסון) – מבוא

מעוניינים לבדוק עד כמה קיים קשר מסוג קשר לינארי (קו ישר) בין שני משתנים. שני המשתנים שאנו בודקים לגביהם קשר צריכים להיות משתנים כמותיים. מבחינת סולמות מדידה כל משתנה נחקר צריך להיות מסולם רווחים או מנה. בדרך כלל המשתנה המוצג כ-  $Y$  הוא המשתנה התלוי והמשתנה המוצג ב-  $X$  הוא המשתנה הבלתי תלוי. תיאור גרפי לנתונים נעשה על ידי דיאגרמת פיזור. בדיאגרמת פיזור אנחנו מסמנים כל תצפית בנקודה לפי שיעור ה-  $X$  ושיעור ה-  $Y$  שלה. דיאגרמת הפיזור נותנת אינדיקציה גרפית על הקשר בין שני המשתנים.

### דוגמה (פתרון בהקלטה) :

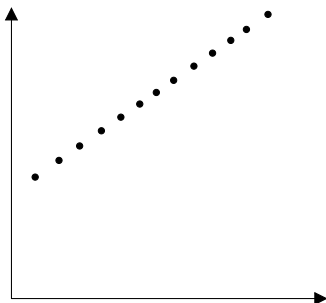
בבניין 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגרות בדירה. להלן התוצאות שהתקבלו :

|   |   |   |   |   |   |   |   |                   |
|---|---|---|---|---|---|---|---|-------------------|
| 4 | 4 | 3 | 3 | 2 | 3 | 2 | 2 | מספר חדרים בדירה  |
| 5 | 4 | 4 | 3 | 2 | 2 | 1 | 0 | מספר הנפשות בדירה |

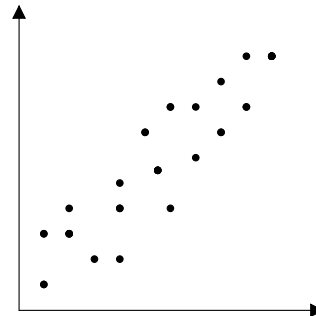
- (1) כמה תצפיות ישנן בדוגמה?
- (2) כמה משתנים ישנם בדוגמה, מי הם?
- (3) שרטטו לנתונים דיאגרמת פיזור.
- (4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

## דיאגרמות פיזור לקשר בין משתנים וניתוחם

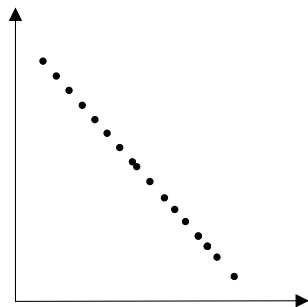
קשר לינארי חיובי מלא



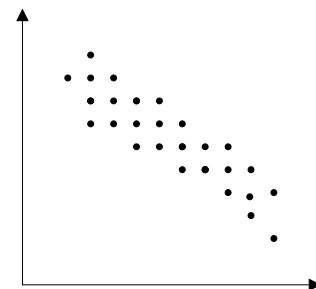
קשר לינארי חיובי חלקי



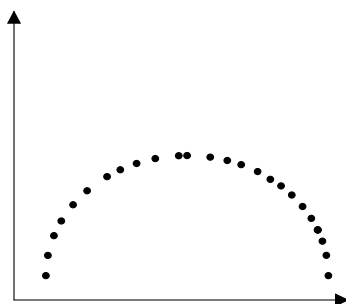
קשר לינארי שלילי מלא



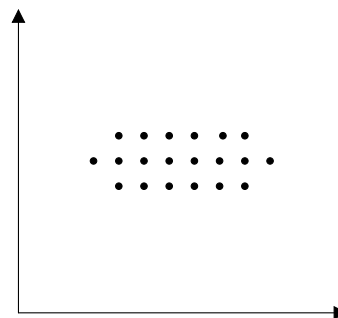
קשר לינארי שלילי חלקי



אין קשר לינארי



אין קשר



### משמעות מקדם המתאם:

כדי לבדוק עד כמה קיים קשר לינארי בין שני המשתנים ישנו מדד קשר שנקרא גם מקדם המתאם הלינארי הידוע גם בשם מקדם המתאם של פירסון. מקדם מתאם זה מקבל ערכים בין 1 ל-1.

-1

0

1

מקדם מתאם 1-או 1 אומר שקיים קשר לינארי מלא בין המשתנים שניתן לבטאו על ידי נוסחה של קו ישר:  $y = ax + b$ .

### מתאם חיובי מלא (מקדם מתאם 1):

קיים קשר לינארי מלא בו השיפוע  $a$  יהיה חיובי ואילו מתאם שלילי (מקדם מתאם-1) מלא אומר שקיים קשר לינארי מלא בו השיפוע  $a$  שלילי.

### מתאם חיובי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את  $X$  ל- $Y$  באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את  $X$  ל- $Y$  באופן מוחלט. ככל שמקדם המתאם קרוב לאפס עוצמת הקשר יותר חלשה וככל שהמדד רחוק יותר מהאפס העוצמה יותר חזקה. לסיכום, מקדם המתאם בודק את עוצמת הקשר הלינארי, ואת כיוון הקשר.

מקדם המתאם הלינארי אינו מושפע מיחידות המדידה. כל שינוי ביחידות המדידה של המשתנים, לא ישנה את מקדם המתאם.

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב:  $\rho$  - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:

$r$  - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר  $\rho$ .

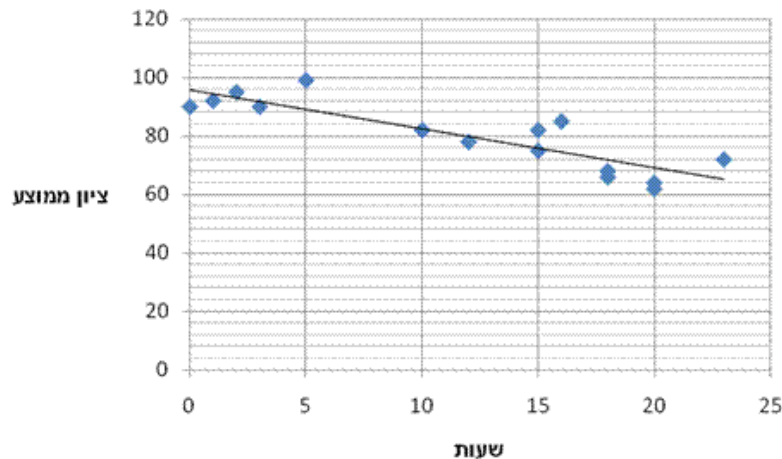
קיומו של מתאם בין שני משתנים אינו מצביע על סיבתיות בהכרח. למשל, אם נמצא מתאם חיובי בין כמות הסוכרזית שאדם אוכל לבין במשקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי, כלומר אם נחליף את  $X$  ב- $Y$  התוצאה תהיה זהה.

### דוגמה (פתרון בהקלטה):

- מה ניתן להגיד על מקדם המתאם של שני המשתנים על סמך דיאגרמת הפיזור ששרטטנו?
- אם היינו משנים את השרטוט כך שבציר האנכי היה המשתנה "מספר החדרים" ובציר האופקי היה "מספר הנפשות", האם הדבר היה משפיע על מדד הקשר של פירסון?

## שאלות

- (1) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור:



- א. מיהו המשתנה הבלתי תלוי?  
 ב. מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר? מה ניתן להגיד על עוצמת הקשר?
- (2) להלן טבלה המסכמת את מקדמי המתאם הלינארי בין ציוני מבחנים שונים שהתקבלו עבור תלמידים בכיתה מסוימת:

| מתמטיקה | לשון | ספורט |         |
|---------|------|-------|---------|
| ?       | -0.7 | ?     | ספורט   |
| 0.6     | ?    | ?     | לשון    |
| ?       | ?    | -0.1  | מתמטיקה |

- א. השלימו את מקדמי המתאם שמסומנים בסימן שאלה בטבלה.  
 ב. בין אילו שני ציוני מקצועות שונים קיים מתאם בעל העוצמה החזקה ביותר?

- (3) במחקר נתבקשו לבדוק את הקשר בין מספר שעות התרגול של קורס לבין הציון הסופי שלו. להלן תוצאות מדגם שהתקבל:

| שעות תרגול | ציון סופי |
|------------|-----------|
| 20         | 90        |
| 25         | 90        |
| 30         | 95        |
| 15         | 60        |
| 30         | 90        |
| 20         | 85        |
| 10         | 50        |

- א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי בדוגמה זו?  
 ב. שרטטו דיאגרמת פיזור לנתונים.  
 ג. מה ניתן לומר על הקשר בין המשתנים במדגם?  
 ד. מסתבר שבסופו של דבר נתנו פקטור של 5 נקודות לציון הסופי. כיצד הדבר היה משנה את מקדם המתאם של המדגם?

4) בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה 0.8.

א. השלימו את המשפט:

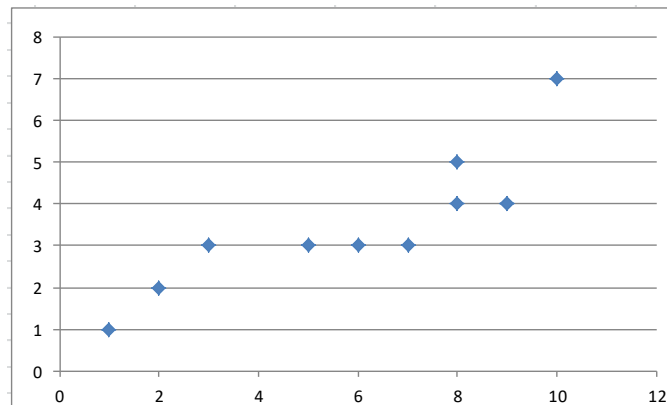
בחודש ינואר ככל שהטמפרטורה היומית נוטה לרדת, כך כמות המשקעים נוטה \_\_\_\_\_.

ב. הוחלט להעביר את הטמפרטורה למעלות פרנהייט על מנת שיוכלו

להשוות אותה לנתונים מארה"ב. נוסחת המעבר היא  $F^0 = 32 + \frac{9}{5}C^0$ .

כיצד הדבר ישפיע על מקדם המתאם בין הטמפרטורה במעלות פרנהייט לכמות המשקעים במ"מ?

5) להלן דיאגרמת פיזור המראה קשר בין שני משנים:



א. השלימו: ניתן לראות שהקשר הוא לינארי \_\_\_\_\_ (מלאו חלקי) כיוון הקשר הוא (חיובי/שלילי).

ב. השלימו: אם היינו מוסיפים תצפית שערך ה-  $X$  שלה הוא 4 וערך ה-  $Y$  שלה הוא 7, מקדם המתאם של פירסון היה \_\_\_\_\_ (גדלו קטן/לא משתנה).

**שאלות רב ברירה (יש לבחור את התשובה הנכונה):**

6) חוקר אקלים דגם כמה ימים בשנה ומדד את הטמפרטורה בטורונטו שבקנדה ואת הטמפרטורה בסידני שבאוסטרליה באותו היום. הוא חישב ומצא מקדם מתאם שלילי בין הטמפרטורה היומית בטורונטו לבין הטמפרטורה היומית בסידני. משמעות מקדם המתאם השלילי במדגם:

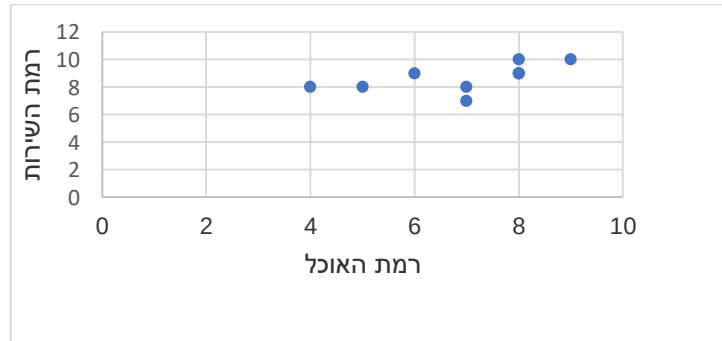
א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני בימים שנדגמו.

ב. במדגם, רוב הטמפרטורות בטורונטו היו שליליות.

ג. ההפרש בין הטמפרטורה בטורונטו לבין הטמפרטורה באוסטרליה, במדגם זה, הוא שלילי.

ד. במדגם יש נטייה שהטמפרטורה יורדת בטורונטו לטמפרטורה לעלות בסידני.

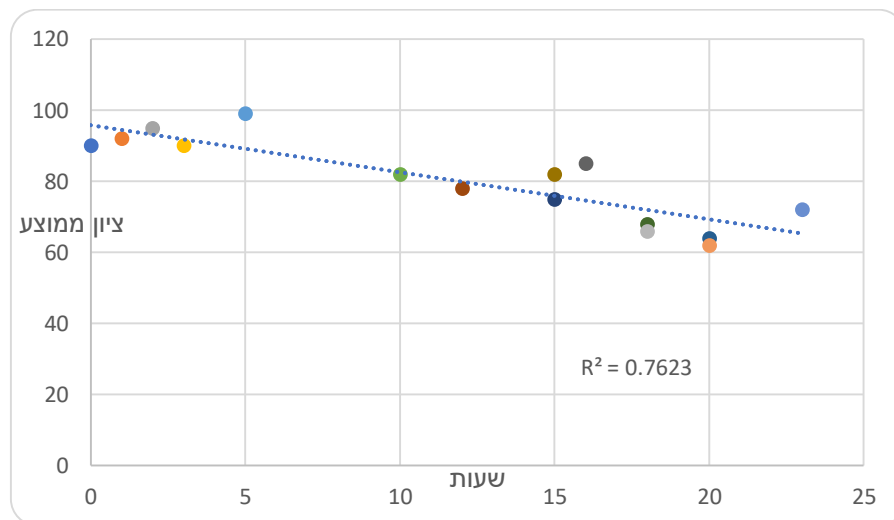
- 7) בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



מה יהיה ערכו של מקדם המתאם ( $r$ )?

- א.  $r = -0.3$
- ב.  $r = 0$
- ג.  $r = 1.125$
- ד.  $r = 0.593$

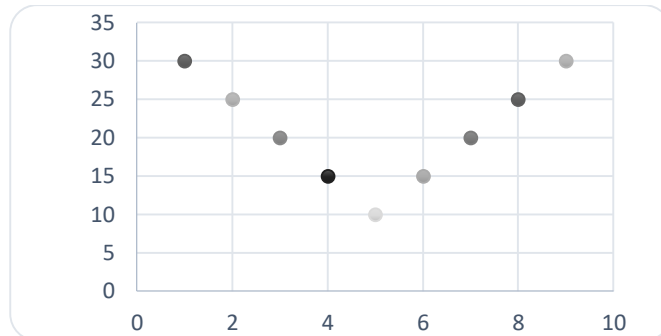
- 8) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור.



מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?

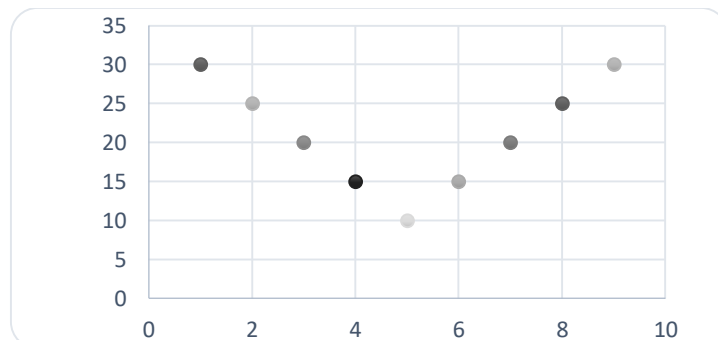
- א. ככל שמבלים יותר הציון נוטה לרדת.
- ב. אין קשר בין שעות הבילוי לציון.
- ג. ככל שמבלים פחות הציון נוטה לרדת.
- ד. ככל שהציון נוטה לרדת הסטודנט מבלה פחות.

9) התרשים הבא מתאר קשר בין שני משתנים, איזה מהמתאמים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?



- א.  $r = 1$  היות ושני המשתנים יוצרים קוים ישרים.
- ב.  $r = 2$  היות ויש שני קוים בעלי קשר מושלם.
- ג.  $r = 0$  היות והקו יורד ואחר כך עולה באותו האופן.
- ד.  $r = \pm 1$  היות ויש קו עולה וגם קו יורד.

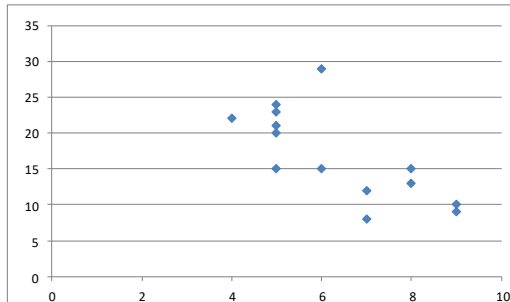
10) התרשים הבא מתאר דיאגרמת פיזור.



איזו טענה נכונה?

- א. בתרשים מוצג הקשר בין שני משתנים.
- ב. בתרשים מוצג הקשר בין 9 משתנים.
- ג. בתרשים מוצג הקשר בין 10 משתנים.
- ד. אין לדעת כמה משתנים מוצגים בתרשים.

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים :



$X$  - (משתנה בלתי תלוי בציר האופקי)

$Y$  - (משתנה תלוי).

במדגם התקבל  $r^2 = 0.52$ .

11) לאור הנתונים המופיעים בדיאגרמה, איזה מבין הערכים הבאים מתאים להיות התוצאה של  $r$  ?

א. -0.52

ב. 0.72

ג. -0.72

ד. 0.52

12) אם מקדם המתאם בין שני משתנים הוא 1, אזי :

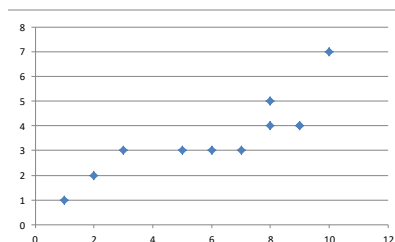
א. הערכים של המשתנים הם חיוביים.

ב. עבור כל תצפית ערך של משתנה אחד שווה לערך של המשתנה השני.

ג. הקשר הלינארי הוא בעוצמה חזקה.

ד. אף אחת מהתשובות לא בהכרח נכונה.

13) להלן דיאגרמת פיזור :



מה יהיה מקדם המתאם בין שני המשתנים ?

א. 1

ב. 0.85

ג. 0.15

ד. 0

14) בבדיקת קשר בין שני משתנים התקבל :  $r = -1$ .

א. קיימת נוסחה לינארית הקושרת בין כל התצפיות.

ב. לא קיים קשר בין שני המשתנים.

ג. ככל שמשתנה אחד נוטה לרדת גם לשני יש נטייה לרדת.

ד. קיים קשר בין שני המשתנים, אך לא ניתן לדעת מאיזה סוג.

15) לפי הפתגם "רחוק מהעין, רחוק מהלב", יש קשר \_\_\_\_\_ בין קרבה פיזית לקרבה נפשית.

- א. חיובי
- ב. שלילי
- ג. אפסי
- ד. לא ניתן לדעת.

16) מבחן אמי"ר הינו מבחן מיון באנגלית של המרכז הארצי לבחינות והערכה. הציון המינימלי בבחינה הינו 150 והמקסימלי הינו 250. בקורס הכנה למבחן השתתפו 19 תלמידים. להלן הציונים שלהם על פי פלט שהתקבל:

|        |         |
|--------|---------|
|        | 159     |
|        | 170     |
|        | 180     |
|        | 185     |
|        | 204     |
|        | 224     |
|        | 236     |
|        | 212     |
|        | 168     |
|        | 189     |
|        | 195     |
|        | 163     |
|        | 187     |
|        | 206     |
|        | 201     |
|        | 223     |
|        | 242     |
|        | 203     |
|        | 205     |
| 197.47 | AVERAGE |
| 536.25 | VARPA   |

יש להוסיף עמודה נוספת לצד עמודת הציונים שתראה לכל תלמיד כמה נקודות חסרות לו כדי להשלים לציון המקסימלי בבחינה.

מה יהיה מקדם המתאם בין שתי העמודות (כלומר, מקדם המתאם בין הציון לבין הנקודות החסרות)?

- א. -1
- ב. 1
- ג. -0.5
- ד. 0.5

17) מקדם המתאם בין שטחי דירה למחיר שלהם חושב ונמצא 1.2. מה נובע מכך?

- א. ככל שהדירה גדולה יותר בשטחה כך היא יקרה יותר.
- ב. ככל שהדירה קטנה יותר בשטחה כך היא זולה יותר.
- ג. לא קיים קשר בין שטח הדירה למחיר הדירה.
- ד. מצב כזה שמתואר הנתונים לא אפשרי.

18) אם ניקח 10 אנשים ונרשום לכל אדם את הגובה במטר וכמו כן את הגובה בס"מ. מה יהיה מקדם המתאם בין גובה האדם במטר לגובה האדם בס"מ?

- א. 1
- ב. 0
- ג. -1
- ד. לא ניתן לדעת.

- 19** נמצא מתאם חיובי בעוצמה גבוהה בין  $X$  – ציון בבגרות בלשון ל  $Y$  – ציון בבגרות במתמטיקה. אילו מהמשפטים הבאים נכון?
- א. ניתן לומר שאחת מהסיבות להבדלים שיש לסטודנטים במתמטיקה נובעים מההבדלים שיש להם בלשון.
- ב. קיימת נוסחה של קו ישר שקושרת בין ציון בבגרות במתמטיקה לציון בבגרות בלשון.
- ג. ללא יוצא מן הכלל, ניתן להגיד שכל תלמיד שמצליח יותר מתלמיד אחר בלשון גם יצליח יותר מאותו תלמיד במתמטיקה.
- ד. אף אחד מהטענות שהוצגו אינה בהכרח נכונה.

- 20** עבור סדרה של תצפיות מדדו את  $X$  ואת  $Y$ . נמצא שעבור כל התצפיות שהערך של  $Y$  ירד הערך של  $X$  בהכרח ירד ללא יוצא מן הכלל. מקדם המתאם של פירסון יהיה בהכרח :

- א. 1
- ב. -1
- ג. 0
- ד. אף אחת מהתשובות.

## תשובות סופיות

- (1) א. שעות בילוי.  
ב. הקשר חלקי, כיוון הקשר שלילי.  
(2) א. להלן טבלה :  
ב. ספורט ולשון.

| מתמטיקה | לשון | ספורט |         |
|---------|------|-------|---------|
| 0.1     | -0.7 | 1     | ספורט   |
| 0.6     | 1    | -0.7  | לשון    |
| 1       | 0.6  | -0.1  | מתמטיקה |

- (3) א. ב"ת- מס' שעות התרגול, תלוי- ציון.  
ג. קשר לינארי חיובי חלקי.  
(4) א. לעלות.  
(5) א. חלקי, חיובי.  
ב. קטן.  
ב. ראה גרף בפתרון וידאו.  
ד. מקדם המתאם לא היה משתנה.  
ב. לא ישפיע על מקדם המתאם.

(6) ד' (7) ד' (8) א' (9) ג' (10) א'

(11) ג' (12) ד' (13) ב' (14) א' (15) א'

(16) א' (17) ד' (18) א' (19) ד' (20) ד'

## מדדי קשר – מדד הקשר הלינארי (פירסון) – רקע

המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל,  $X$  הוא המשתנה המסביר (הבלתי תלוי) ו- $Y$  הוא המשתנה המוסבר (התלוי).

### דוגמה:

נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד  $X$  מסבירה את ההכנסה שלו  $Y$ . במקרה זה שנות ההשכלה זהו המשתנה המסביר (או הבלתי תלוי) ואנחנו מעוניינים לבדוק כיצד שינויים בשנות ההשכלה של אדם יכולים להסביר את השינויים שלו בהכנסה, ולכן רמת ההכנסה זהו המשתנה המוסבר התלוי במשתנה המסביר אותו.

**שלב ראשון:** נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים.

### דוגמה:

| מס' דירה | $X$ | $Y$ |
|----------|-----|-----|
| 1        | 3   | 2   |
| 2        | 2   | 2   |
| 3        | 4   | 3   |
| 4        | 3   | 3   |
| 5        | 5   | 4   |

בבניין של 5 דירות בדקו את הנתונים הבאים:  
 $X$  - מס' חדרים בדירה.  $Y$  - מס' נפשות הגרות בדירה.  
 להלן התוצאות שהתקבלו:

נשרטט מנתונים אלה דיאגרמת פיזור (הדיאגרמה המלאה בסרטון). נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן (הדיאגרמות המלאות בסרטון).

**שלב שני:** מחשבים את מקדם המתאם (מדד הקשר) שבודק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שניראה בשלב הראשון רק בעין.  
 המדד בודק את כיוון הקשר (חיובי או שלילי) ואת עוצמת הקשר (חלש עד חזק).  
 מקדם מתאם זה מקבל ערכים בין -1 ל-1.  
 מקדם מתאם -1 או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי הנוסחה:  $y = bx + a$ .

### מתאם חיובי מלא (מקדם מתאם 1):

קיים קשר לינארי מלא בו השיפוע  $b$  יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לינארי מלא בו השיפוע  $b$  שלילי (מקדם מתאם -1).

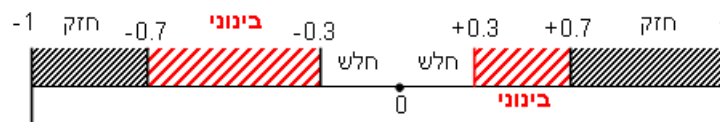
### מתאם חיובי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את  $X$  ל- $Y$  באופן מוחלט.

### מתאם שלילי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את  $X$  ל- $Y$  באופן מוחלט.

ככל שערך מקדם המתאם קרוב לאפס נאמר שעוצמת הקשר חלשה יותר וככל שמקדם המתאם רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר:



מקדם המתאם יסומן באות  $r$ .

כדי לחשב את מקדם המתאם, יש לחשב את סטיות התקן של כל משתנה ואת השונות המשותפת.

$$COV(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} = \frac{\sum xy}{n} - \bar{x} \cdot \bar{y} : \text{שונות משותפת}$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 : \text{שונות של המשתנה } X$$

$$S_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2 : \text{שונות המשתנה } Y$$

$$r_{xy} = \frac{COV(x, y)}{S_x \cdot S_y} : \text{מקדם המתאם הלינארי}$$

## שאלות

- 1) להלן נתונים לגבי שישה תלמידים שנגשו למבחן. בדקו לגבי כל תלמיד את הציון שלו בסוף הקורס וכמו כן את מספר החיסורים שלו מהקורס.

| מספר חיסורים | 2  | 1  | 0  | 2  | 3  | 4  |
|--------------|----|----|----|----|----|----|
| ציון         | 80 | 90 | 90 | 70 | 70 | 50 |

- א. שרטטו דיאגרמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר בין מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?
- ב. חשבו את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?
- ג. הסבירו, ללא חישוב, כיצד מקדם המתאם היה משתנה אם היה מתווסף תלמיד שהחסיר 4 פעמים וקיבל ציון 80?

| X  | Y  |
|----|----|
| 10 | 12 |
| 14 | 15 |
| 15 | 15 |
| 18 | 17 |
| 20 | 21 |

- 2) במחקר רפואי רצו לבדוק האם קיים קשר בין רמת ההורמון X בדם החולה לרמת ההורמון Y שלו. לצורך כך מדדו את רמת ההורמונים ההלו עבור חמישה חולים. להלן התוצאות שהתקבלו:
- א. מה הממוצע של כל רמת הורמון?
- ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

- 3) נסמן ב-X את ההכנסה של משפחה באלפי ₪. נסמן ב-Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} Y_i = 200 \qquad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 76 \qquad \sum_{i=1}^{20} (X_i - \bar{X})^2 = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y_i - \bar{Y}) = 60.8$$

- א. חשב את מדד הקשר הלינארי בין X ל-Y. מיהו המשתנה התלוי?
- ב. מה המשמעות של התוצאה שקיבלת בסעיף א'?

- (4) נסמן ב-  $X$  את ההכנסה של משפחה באלפי ₪. נסמן ב-  $Y$  את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} Y_i = 200 \quad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} Y_i^2 = 2080 \quad \sum_{i=1}^{20} X_i^2 = 2960$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

חשבו את מדד הקשר הלינארי בין  $X$  ל-  $Y$ .

- (5) במוסד אקדמי ציון ההתאמה מחושב כך: מכפילים את הציון הממוצע בבגרות ב-3 ומפחיתים 2 נקודות. ידוע שעבור 40 מועמדים סטיית התקן של ממוצע הציון בבגרות הייתה 2.  
מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם?

- (6) להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו.
- מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.
  - לסדרה של נתונים התקבל  $\bar{X} = \bar{Y} = 6$ ,  $S_x = S_y = 1$ . לכן, מדד הקשר של פירסון יהיה 1.
  - אם השונות המשותפת של  $X$  ושל  $Y$  הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

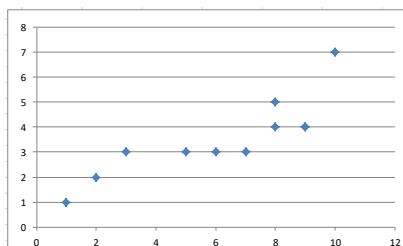
#### שאלות רב-ברירה:

- (7) נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן:
- הדבר מעיד שהציונים בכיתה היו שליליים.
  - ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
  - ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
  - אף אחת מהתשובות לא נכונה.

(8) נלקחו 20 מוצרים ונבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו היום ערך הדולר היה-4.2ש). מהו מקדם המתאם בין המחיר בדולר למחיר בש"ח?

- א. 1  
 ב. 0  
 ג. 4.2  
 ד. לא ניתן לדעת.

(9) להלן דיאגרמת פיזור: מה יהיה מקדם המתאם בין שני המשתנים?



- א. 1  
 ב. 0.85  
 ג. 0.15  
 ד. 0

## תשובות סופיות

- (1) א. משתנה תלוי: ציון, משתנה ב"ת: מס' חיסורים. ראה דיאגרמה בוידאו. ניתן להסיק שקיים קשר לינארי שלילי וחלקי בין מספר החיסורים לציון התלמיד.  
 ב. -0.9325.  
 ג. הקשר יישאר לינארי שלילי חלקי אך עוצמתו תחלש.
- (2) א.  $\bar{y} = 16$ ,  $\bar{x} = 15.4$  ב.  $r_{xy} = 0.96$ .
- (3) א. 0.8  
 (4) 0.8  
 (5) 1  
 (6) א. נכון. ב. לא נכון. ג. נכון.  
 (7) ג'.  
 (8) א'.  
 (9) ב'.

### בדיקת השערות על מקדם המתאם הלינארי – רקע

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב:  $\rho$  - פרמטר המאפיין את עוצמת הקשר הלינארי וכיוונו בין שני המשתנים הנחקרים באוכלוסייה. כאשר:  $r$  - מדד הקשר הלינארי במדגם שמהווה אומדן לפרמטר  $\rho$ .

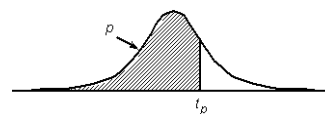
**השערת האפס:** תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים  $H_0: \rho = 0$ .  
ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t(n-2)$$

סטטיסטי המבחן:

סטטיסטי זה מתפלג  $t$  עם  $n-2$  דרגות חופש.

|                       |                        |                                                      |                                         |
|-----------------------|------------------------|------------------------------------------------------|-----------------------------------------|
| $H_0: \rho = 0$       | $H_0: \rho = 0$        | $H_0: \rho = 0$                                      | השערת האפס:                             |
| $H_1: \rho > 0$       | $H_1: \rho < 0$        | $H_1: \rho \neq 0$                                   | השערת המחקר:                            |
| $t \geq t_{1-\alpha}$ | $t \leq -t_{1-\alpha}$ | $t \geq t_{1-\alpha}$<br>ו<br>$t \leq -t_{1-\alpha}$ | כלל ההכרעה:<br>אזור דחייה של השערת האפס |

**טבלת ערכים קריטיים של  $t$  - נספח: טבלת התפלגות  $T$** 
 $P$ 

| דרגות חופש | 0.75  | 0.90  | 0.95  | 0.975  | 0.99   | 0.995  | 0.9995  |
|------------|-------|-------|-------|--------|--------|--------|---------|
| 1          | 1.000 | 3.078 | 6.314 | 12.709 | 31.821 | 63.657 | 636.619 |
| 2          | 0.816 | 1.886 | 2.920 | 4.303  | 6.965  | 9.925  | 31.598  |
| 3          | 0.765 | 1.638 | 2.353 | 3.182  | 4.541  | 5.841  | 12.941  |
| 4          | 0.741 | 1.533 | 2.132 | 2.776  | 3.747  | 4.604  | 8.610   |
| 5          | 0.727 | 1.476 | 2.015 | 2.571  | 3.365  | 4.032  | 6.859   |
| 6          | 0.718 | 1.440 | 1.943 | 2.447  | 3.143  | 3.707  | 5.959   |
| 7          | 0.711 | 1.415 | 1.895 | 2.365  | 2.998  | 3.499  | 5.405   |
| 8          | 0.706 | 1.397 | 1.860 | 2.306  | 2.896  | 3.355  | 5.041   |
| 9          | 0.703 | 1.383 | 1.833 | 2.262  | 2.821  | 3.250  | 4.781   |
| 10         | 0.700 | 1.372 | 1.812 | 2.228  | 2.764  | 3.169  | 4.587   |
| 11         | 0.697 | 1.363 | 1.796 | 2.201  | 2.718  | 3.106  | 4.437   |
| 12         | 0.695 | 1.356 | 1.782 | 2.179  | 2.681  | 3.055  | 4.318   |
| 13         | 0.694 | 1.350 | 1.771 | 2.160  | 2.650  | 3.012  | 4.221   |
| 14         | 0.692 | 1.345 | 1.761 | 2.145  | 2.624  | 2.977  | 4.140   |
| 15         | 0.691 | 1.341 | 1.753 | 2.131  | 2.602  | 2.947  | 4.073   |
| 16         | 0.690 | 1.337 | 1.746 | 2.120  | 2.583  | 2.921  | 4.015   |
| 17         | 0.689 | 1.333 | 1.740 | 2.110  | 2.567  | 2.898  | 3.965   |
| 18         | 0.688 | 1.330 | 1.734 | 2.101  | 2.552  | 2.878  | 3.922   |
| 19         | 0.688 | 1.328 | 1.729 | 2.093  | 2.539  | 2.861  | 3.883   |
| 20         | 0.687 | 1.325 | 1.725 | 2.086  | 2.528  | 2.845  | 3.850   |
| 21         | 0.686 | 1.323 | 1.721 | 2.080  | 2.518  | 2.831  | 3.819   |
| 22         | 0.686 | 1.321 | 1.717 | 2.074  | 2.508  | 2.819  | 3.792   |
| 23         | 0.685 | 1.319 | 1.714 | 2.069  | 2.500  | 2.807  | 3.767   |
| 24         | 0.685 | 1.318 | 1.711 | 2.064  | 2.492  | 2.797  | 3.745   |
| 25         | 0.684 | 1.316 | 1.708 | 2.060  | 2.485  | 2.787  | 3.725   |
| 26         | 0.684 | 1.315 | 1.706 | 2.056  | 2.479  | 2.779  | 3.707   |
| 27         | 0.684 | 1.314 | 1.703 | 2.052  | 2.473  | 2.771  | 3.690   |
| 28         | 0.683 | 1.313 | 1.701 | 2.048  | 2.467  | 2.763  | 3.674   |
| 29         | 0.683 | 1.311 | 1.699 | 2.045  | 2.462  | 2.756  | 3.659   |
| 30         | 0.683 | 1.310 | 1.697 | 2.042  | 2.457  | 2.750  | 3.646   |
| 40         | 0.681 | 1.303 | 1.684 | 2.021  | 2.423  | 2.704  | 3.551   |
| 60         | 0.679 | 1.296 | 1.671 | 2.000  | 2.390  | 2.660  | 3.460   |
| 120        | 0.677 | 1.289 | 1.658 | 1.980  | 2.358  | 2.617  | 3.373   |
| $\infty$   | 0.674 | 1.282 | 1.645 | 1.960  | 2.326  | 2.576  | 3.291   |

## שאלות

(1) להלן נתונים על הוותק בעבודה (בשנים) ועל השכלה (בשנים) במדגם של 10 עובדים:

| נחקר      | 1  | 2  | 3  | 4 | 5  | 6  | 7  | 8  | 9  | 10 |
|-----------|----|----|----|---|----|----|----|----|----|----|
| X - וותק  | 13 | 18 | 2  | 8 | 16 | 9  | 5  | 28 | 17 | 24 |
| Y - השכלה | 12 | 14 | 17 | 8 | 11 | 12 | 13 | 8  | 12 | 15 |

מקדם המתאם חושב והתקבל:  $-0.31$ .

- א. האם קיים מתאם בין וותק העובד להשכלתו? בדקו ברמת מובהקות של 5%?  
 ב. אם הוותק של העובד היה נמדד בחודשים האם התשובה לסעיף א' הייתה משתנה?

(2) מחקר התעניין לבדוק את הקשר בין גיל נשים בהריון לרמת ההמוגלובין שלהן בדם בזמן הריון. נדגמו 7 נשים והתקבלו התוצאות הבאות:

| נחקרת     | 1    | 2    | 3   | 4  | 5    | 6  | 7    |
|-----------|------|------|-----|----|------|----|------|
| המוגלובין | 14.7 | 13.5 | 9.7 | 12 | 10.8 | 13 | 10.3 |
| גיל       | 39   | 34   | 30  | 29 | 28   | 26 | 23   |

במדגם חושב מדד הקשר של פירסון להיות  $0.7$ .

- א. האם ניתן לומר שבמדגם אם אישה היא יותר מבוגרת אזי בהכרח יש לה יותר המוגלובין בדם?  
 ב. האם ניתן לומר, ברמת מובהקות של 5%, שקיים מתאם בין גיל האישה שבהריון לבין רמת ההמוגלובין שלה בדם?

(3) בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה  $-0.8$ .

- א. בדקו ברמת מובהקות של 2.5% האם קיים קשר לינארי שלילי בחודש ינואר בין הטמפרטורה במעלות צלזיוס לבין המשקעים במעלות צלזיוס.  
 ב. כיצד הייתה משתנה התשובה לסעיף א אם הינו מוסיפים עוד תצפיות למדגם?  
 ג. על סמך טבלת T המצורפת עבור אילו רמות מובהקות ניתן להחליט שקיים קשר לינארי שלילי מובהק?

(4) מתווך דירות חישב את מקדם המתאם בין שטח דירה במרכז תל אביב לבין המחיר של הדירה עבור 17 דירות. מקדם המתאם שקיבל היה  $0.6$ .

- א. בדקו ברמת מובהקות של 5% האם ניתן להגיד שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה במרכז תל אביב?  
 ב. מהי מובהקות התוצאה לבדיקת ההשערה שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה בתל אביב.

**תשובות סופיות**

- |                           |                           |
|---------------------------|---------------------------|
| ב. לא תשתנה.              | (1) א. לא נדחה את $H_0$ . |
| ב. לא נדחה את $H_0$ .     | (2) א. לא                 |
| ב. לא ניתן לדעת.          | (3) א. נדחה את $H_0$ .    |
|                           | ג. לפחות 0.005.           |
| ב. $0.005 < P_v < 0.01$ . | (4) א. נדחה את $H_0$ .    |

# מבוא לסטטיסטיקה תיאורית והסקית

פרק 31 - מדדי קשר - בחירת מדד מתאים

תוכן העניינים

1. בחירת מדד מתאים ..... 175

## מדדי קשר – בחירת מדד מתאים:

### רקע:

בפרק זה נתרגל את התהליך של בחירת מדד הקשר (מקדם המתאם) המתאים. נתרכז בשלושת מדדי הקשר הנפוצים ביותר:

- מדד הקשר של קרמר.
- מדד הקשר של ספירמן.
- מדד הקשר של פירסון (מדד הקשר הלינארי).

בחירת מדד הקשר נעשה לפי סולמות המדידה של שני המשתנים שאנחנו רוצים לבדוק את הקשר בינם. הנושא של סולמות מדידה נלמד כבר בפרק אחר, כמו כן כל מדד קשר נלמד בפרק נפרד. אנו מתרכזים ב 3 סולמות מדידה:

- סולם שמי/ זהות (nominal).
- סולם סדר (ordinal).
- סולם כמותי (scale): לכאן אנו מאחדים את סולם רווחים ומנה יחד.

שלושת מדדי הקשר שלעיל דנים בקשר בין שני משתנים. מדדי הקשר הם סימטריים, כלומר אין זה משנה איזה משתנה נגדיר בתור משתנה  $X$  ואיזה יוגדר בתור משתנה  $Y$ .

להלן טבלה שמסכמת את בחירת המדד המתאים:

| $X / Y$ | שמי  | סדר    | כמותי  |
|---------|------|--------|--------|
| שמי     | קרמר | קרמר   | קרמר   |
| סדר     | קרמר | ספירמן | ספירמן |
| כמותי   | קרמר | ספירמן | פירסון |

**דוגמה (פתרון בהקלטה):**

במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום. איזה מדד קשר מתאים כדי לבדוק האם קיים קשר בין התפוקה היומית לכמות עובדים באותו היום במפעל?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. חסרים נתונים כדי לדעת זאת.

## שאלות:

- (1) בקרב תלמידי כיתות א' בבית הספר גבריאלי אשר בתל אביב בדקו לכל תלמיד את גובהו בס"מ ואת משקלו בק"ג. מהו מדד הקשר המתאים כדי לבדוק האם קיים קשר בין גובה התלמיד למשקלו?
- פירסון.
  - ספירמן.
  - קרמר.
  - אין מספיק נתונים כדי לדעת.
- (2) בסקר שנעשה על אזרחים במדינה בדקו לכל אזרח את השכלתו ואת שכרו. מהו מדד הקשר המתאים כדי לבדוק האם קיים קשר בין השכלה לשכר?
- פירסון.
  - ספירמן.
  - קרמר.
  - אין מספיק נתונים כדי לדעת.
- (3) בגן הילדים של שולה אספו נתונים על 25 הילדים שבגן. על כל ילד בדקו את רמת הביטחון העצמי שלו ( $X$  = מדד שמקבל ערכים בין 1 - נמוך ועד 5 - גבוה), ואת אוצר המילים שלו ( $Y$  = לפי מבחן שנעשה לכל ילד בו ספרו את מספר המילים שידע מתוך רשימה של 20 מילים). איסוף הנתונים נעשה על ידי איש מקצוע שצפה בילדים ובחן אותן. מהו מקדם המתאם המתאים לבדיקת התלות בין  $X$  לבין  $Y$ ?
- פירסון.
  - ספירמן.
  - קרמר.
  - אין מספיק נתונים כדי לדעת.
- (4) הועלתה השערה בבית הספר למדעי ההתנהגות שיש קשר בין המרצה להצלחת הסטודנט. לצורך בדיקת הטענה בדקו לגבי כל סטודנט שלמד סטטיסטיקה אצל איזה מרצה הוא למד (היו 3 מרצים שונים) והאם הוא עבר את הבחינה. מהו מדד הקשר המתאים במקרה זה?
- פירסון.
  - ספירמן.
  - קרמר.
  - אין מספיק נתונים כדי לדעת.

(5) בבורסה בתל אביב רצו לבדוק את הקשר בין גובה הריבית במשק בסוף החודש (באחוזים), לבין תשואת מניית אקטר (באחוזים) בסוף החודש. מהו מדד הקשר המתאים?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

(6) בעל מסעדה ביצע סקר על לקוחותיו, בין השאלות שנשאלו בסקר :  
 • מה מידת שביעות הרצון של הלקוח מאדיבות השירות של המלצר בסקלה של 1 עד 5.  
 • מה גילו של הלקוח בשנים.  
 • מה גובה התשר (טיפ) ב-₪ אשר נתן הלקוח למלצר בלכתו מהמסעדה.

מהו המדד המתאים כדי לבדוק האם קיים מתאם חיובי בין מידת שביעות הרצון של הלקוח מאדיבות השירות לבין גובה התשר שהוא נתן למלצר?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

(7) בעל מסעדה ביצע סקר על לקוחותיו, בין השאלות שנשאלו בסקר :  
 • מה מידת שביעות הרצון של הלקוח מאדיבות השירות של המלצר בסקלה של 1 עד 5.  
 • מה גילו של הלקוח בשנים.  
 • מה גובה התשר (טיפ) ב-₪ אשר נתן הלקוח למלצר בלכתו מהמסעדה.

מהו המדד המתאים כדי לבדוק האם קיים מתאם בין גיל הלקוח לגובה התשר שהעניק לשירות?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

### הנתונים הבאים מתאימים ל-3 השאלות הבאות:

חוקרים ערכו מדגם של ילדים מכיתות ב' ו-ג' מ-4 בתי ספר שונים. הועבר לילדים שאלון בו תואר מצב מסוים והילדים התבקשו לציין את רמת החרדה שלהם באשר לאותו מצב. המשתנים שלגביהם נאספו נתונים:

- מגדר (1 - בן, 2 - בת).
- כיתה (0 - ג', 1 - ב').
- בית ספר (A, B, C, D).
- רמת חרדה (ציון שהילד היה צריך לתת בסקלה של 1 עד 10).
- גיל התלמיד בחודשים.

8) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין גיל התלמיד לבין רמת החרדה שלו מהמצב?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

9) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין המגדר לבין רמת החרדה שלו מהמצב?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

10) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין המגדר לבין בית הספר?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

11) בטבלה שלהלן נתונות שכיחויות ההצלחה והכישלון של 150 חולים :

| תוצאה/ התרופה | A  | B  | C  |
|---------------|----|----|----|
| נרפא          | 35 | 13 | 45 |
| לא נרפא       | 15 | 37 | 5  |

החולים קיבלו 3 תרופות שונות ובדקו עבור כל חולה אם התרופה הצליחה בריפוי. מהו מדד הקשר המתאים?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

12) שני מוסיקאים מפורסמים נתנו ציון בסולם של 1-10 לקולם של 8 מתמודדים בתוכנית ריאליטי ידועה. ציון 10 ניתן לקול שמצא חן ביותר בעיני המוסיקאי. מפיק התוכנית רצה לבדוק האם יש קורלציה בין המוסיקאים מבחינת הטעם. בטבלה הבאה נתונים הציונים של כל אחד מהמוסיקאים את שמונת המתמודדים :

|            | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|------------|---|---|---|---|---|---|---|---|
| מוסיקאי א' | 6 | 5 | 7 | 4 | 3 | 1 | 1 | 4 |
| מוסיקאי ב' | 5 | 7 | 5 | 2 | 3 | 3 | 2 | 7 |

מהו מדד הקשר המתאים?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

13) להלן טבלה המסכמת את השכר באלפי ₪ של עובדים בחברה ואת רמת המוטיבציה שלהם מ-1 עד 5 :

| שכר      | 12 | 18 | 20 | 15 | 30 |
|----------|----|----|----|----|----|
| מוטיבציה | 4  | 4  | 5  | 3  | 5  |

מהו מקדם המתאם המתאים לבדיקת רמת ההתאמה בין המוטיבציה לשכר של העובד?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

14) להלן טבלה על נתונים שנאספו על מספר תצפיות:

|    |    |    |    |    |   |
|----|----|----|----|----|---|
| 5  | 4  | 3  | 2  | 1  | X |
| 20 | 17 | 17 | 14 | 12 | Y |

אם מעוניינים לבדוק עד כמה קיים קשר לנארי בין שני המשתנים.  
מהו המדד המתאים?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

### תשובות סופיות:

|        |        |        |        |        |
|--------|--------|--------|--------|--------|
| א' (5  | ג' (4  | ב' (3  | ד' (2  | א' (1  |
| ג' (10 | ג' (9  | ב' (8  | א' (7  | ב' (6  |
|        | א' (14 | ב' (13 | ב' (12 | ג' (11 |