# מבוא לסטטיסטיקה למדעי המחשב

# תוכן העניינים

# תוכן העניינים

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 1 - סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה

תוכן העניינים

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

א. כמה תצפיות נאספו במחקר?
ב. סווגו את סולם המדידה של "מינון התרופה"?

**4)** לפניכם רשימה של משתנים.
ציינו 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-₪.
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. $n=7$.
ב. 4.
ג. 3.
ד. רווק, נשוי, גרוש, אלמן.

2) א. לקוחות סניף 543 של בנק "רווה".
ב. $n=80$.
ג. $X$ = מספר הפעמים בשבוע שלקוח נכנס לסניף. כמותי בדיד.

3) א. $n=5$.
ב. כמותי רציף.

4) א. כמותי רציף.
ב. כמותי בדיד.
ג. אין מספיק נתונים.
ד. סדר.
ה. אין מספיק נתונים.
ו. שמי.
ז. שמי.
ח. שמי.
ט. סדר.

5) א. רציף.
ב. בדיד.
ג. בדיד.
ד. רציף.
ה. רציף.

6) א. מספר הרמזורים בהם עוצרת גברת לוי ביום בדרך לעבודה.
ב. 0, 1, 2, 3.
ג. 4.
ד. כמותי בדיד.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 2 - סטטיסטיקה תיאורית- הצגה של נתונים

### תוכן העניינים

# 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: 4, 5, 3, 4, 3.

### טבלת שכיחויות בדידה:

| שם המשתנה- $X$ | שכיחות – $f(x)$ | שכיחות יחסית באחוזים |
|---|---|---|
| $X_1$ | $f_1$ | $\frac{f_1}{N}\cdot 100$ |
| $X_2$ | $f_2$ | $\frac{f_2}{N}\cdot 100$ |
| $X_3$ | $f_3$ | $\frac{f_3}{N}\cdot 100$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $X_k$ | $f_k$ | $\frac{f_x}{N}\cdot 100$ |
| **סה"כ** | $N=\sum_{i=1}^{k} f_i$ | 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 – 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 שווה לערך.

**טבלת שכיחויות במחלקות:**

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

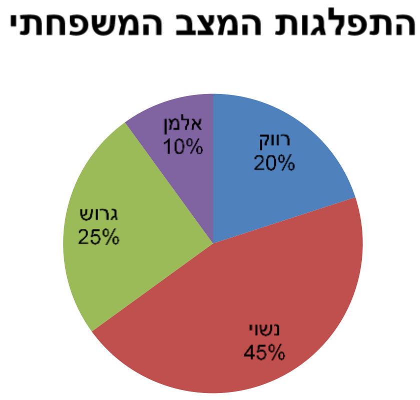
**דוגמה:**

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

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

## דיאגרמת עוגה:

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



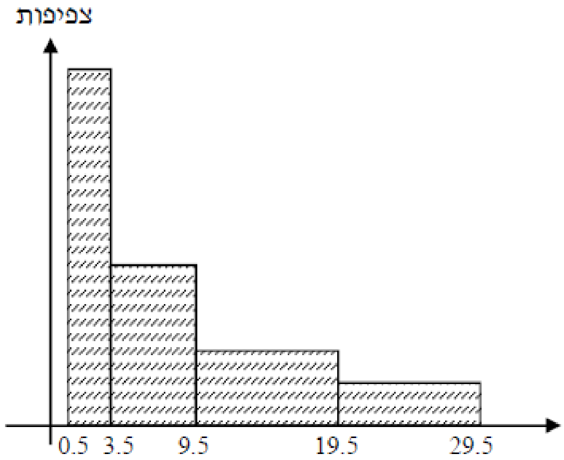

## דיאגרמת מקלות:

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 ליחדה. 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 |



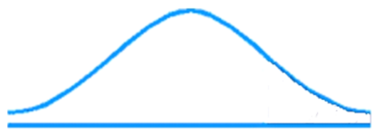

## פוליגון – מצולעון:

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

**צורות התפלגות נפוצות:**

**התפלגות סימטרית פעמונית**

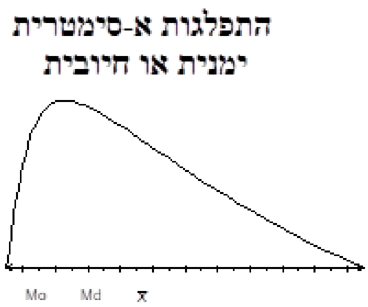
רוב התצפיות במרכז, וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי.
לדוגמה, ציוני IQ.



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 אסימטרית שמאלית (שלילית)**

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

## 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":

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

א. מהו המשתנה הנחקר?

ב. מהי פרופורציית התלמידים שמעדיפים תנ"ך?

3) להלן התפלגות ההשכלה במקום עבודה מסוים:

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

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, 9, 8, 6, 7, 6, 7, 8, 5, 4, 6, 10, 9, 8, 6, 7.

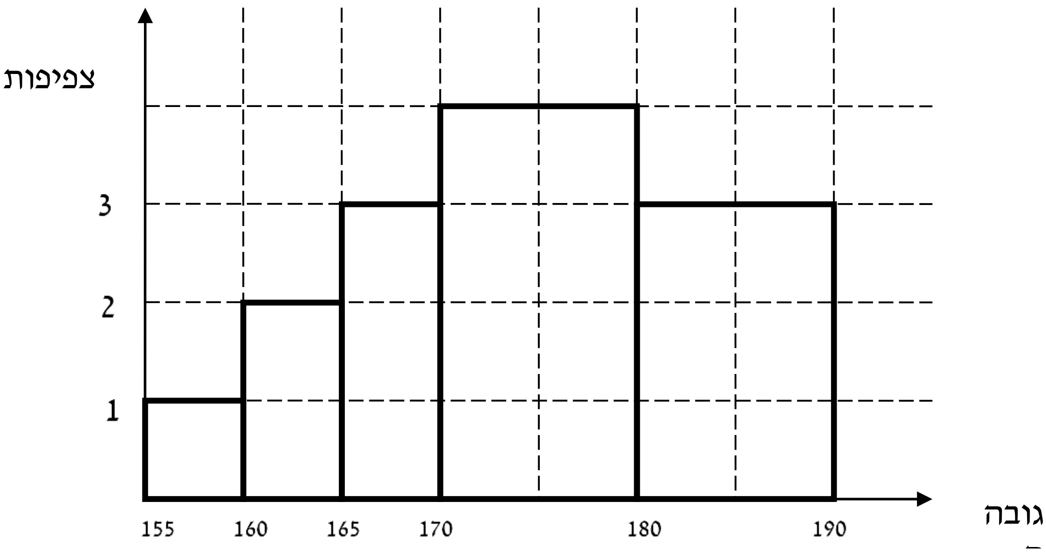
א. מהו המשתנה? האם הוא בדיד או רציף?

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:




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

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

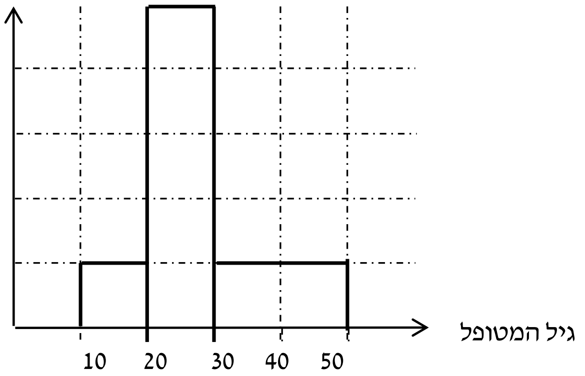
| משקל | מספר מקרים |
|---|---|
| 40-45 | 10 |
| 45-50 | 20 |
| 50-60 | 30 |
| 60-65 | 20 |
| 65-70 | 10 |

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

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

קנה מידה :

8 מטופלים=




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:

| $x$ | $f(x)$ | $\frac{f(x)}{n}$ | % |
|---|---|---|---|
| **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% |

ב. עיין גרף מלא בסרטון הוידאו.

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:

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

ה. 62.5%.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 3 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

# 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:

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

(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: $i=1,2,...,6$, 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. $\left(\sum_{i=1}^{n} X_i\right)^2 = \sum_{i=1}^{n} X_i^{\,2}$.

**4)** נתון: $\sum_{i=1}^{10} X_i^2 = 1640$ , $\sum_{i=1}^{10} X_i = 80$.

חשבו: $\sum_{i=1}^{10} (X_i - 4)^2$.

## 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. | |
| **4)** | 1160. | | | |

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 4 - מדדי מרכז

### תוכן העניינים

# מבוא לסטטיסטיקה למדעי המחשב

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

תוכן העניינים

# 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.

עבור סדרת נתונים: $S_x{}^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \dfrac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2$.

## דוגמאות:

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

עבור טבלת שכיחויות: $S_x{}^2 = \dfrac{\sum(x - \bar{x})^2 f}{n} = \dfrac{\sum x^2 \cdot f}{n} - \bar{x}^2$.

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

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

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

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

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, 9, 8, 6, 7, 6, 7, 8, 5, 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.

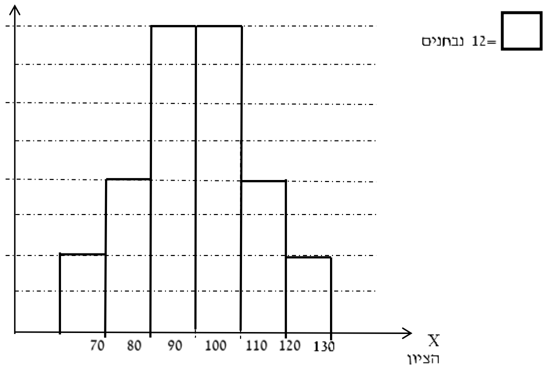
| מספר חדרים | פרופורציה |
|---|---|
| 1 | 0.1 |
| 2 | 0.2 |
| 3 | 0.4 |
| 4 | 0.15 |
| 5 | |

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.

6) להלן התפלגות המשקל של קבוצה מסוימת בק"ג:
מהי סטיית התקן של התפלגות המשקל?

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

7) להלן התפלגות הציונים במבחן אינטליגנציה:




א. מה הממוצע ומה החציון של ההתפלגות?
ב. חשבו את סטיית התקן של הציונים.
ג. מסתבר שיש להוסיף 20 תצפיות לכל אחת משתי המחלקות 90-100 ו- 100-110. כיצד הדבר ישתנה את כל אחד מהמדדים של הסעיפים 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.
6) 7.73.
7) א. 100. ב. 12.96. ג. ממוצע: לא ישתנה, סטיית תקן: תקטן.

# מבוא לסטטיסטיקה למדעי המחשב

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

תוכן העניינים

# 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 ו75% מהמקרים גבוהים או שווים לו.
סימון: $Q_1$.

רבעון עליון – ערך שמחלק את ההתפלגות לשניים.
75% מהמקרים נמוכים ממנו או שווים לו ו25% 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 החיסכון | $f(x)$ | שכיחות מצטברת | שכיחות יחסית מצטברת |
|---|---|---|---|
| 0 | 100 | | |
| 1 | 75 | | |
| 2 | 25 | | |
| 3 | 25 | | |
| 4 | 25 | | |

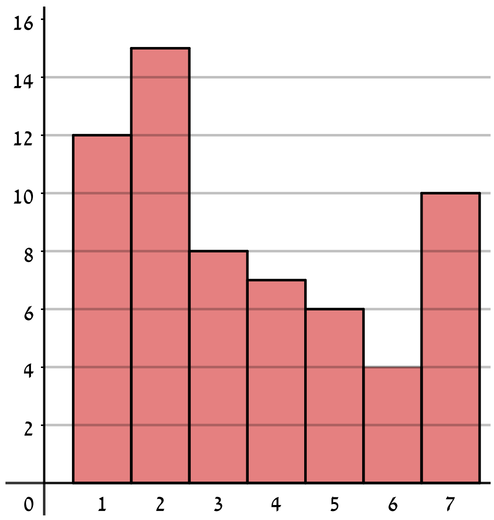
## שאלות:

**1)** להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

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

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

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



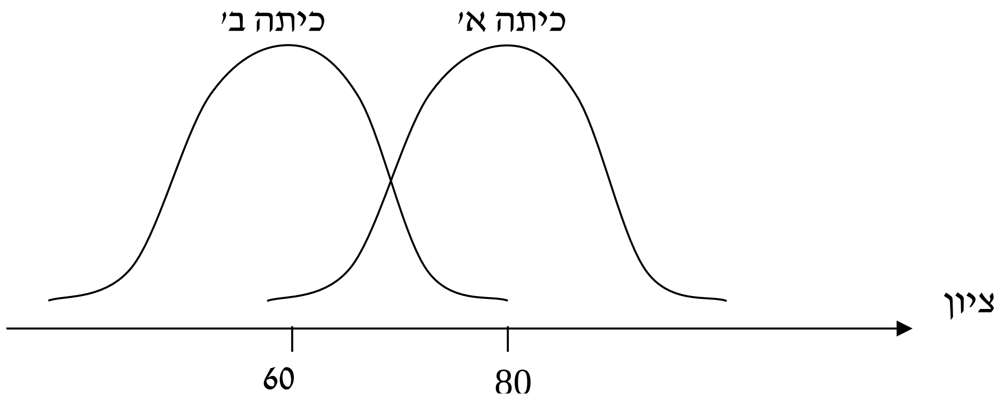

א. מה מייצגים הערכים בציר האופקי?

ב. מהו הטווח הבין-רבעוני של מספר ימי המחלה של המורים

ג. אם נוסיף 25 מורים אשר הצהירו שמספר ימי המחלה השנתיים שלהם הוא 4 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.

ד. 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 למדעי המחשב

## פרק 7 - סטטיסטיקה תיאורית - מדדי פיזור - ממוצע סטיות מוחלטות מהחציון

תוכן העניינים

# סטטיסטיקה תיאורית – מדדי פיזור – ממוצע סטיות מוחלטות מהחציון:

**רקע:**

מדד זה הוא מדד לפיזור בנוסף למדדים שנלמדו בפרקים הקודמים כמו סטיית התקן. המדד בודק את הפיזור הממוצע סביב החציון. הרעיון הוא למצוא בכמה התצפיות סוטות בערכן המוחלט מהחציון, בממוצע.
כדי לחשב את המדד יש לחשב קודם כל את החציון.

אם מדובר ברשימה של תצפיות, הנוסחה לחישוב המדד: $MAD=\frac{\sum_i |X_i-Md|}{n}$.

אם מדובר בטבלת שכיחויות, הנוסחה לחישוב המדד: $MAD=\frac{\sum |X_i-Md|\cdot f(X)}{n}$.

כאשר מדובר על טבלת שכיחויות במחלקות ניקח בתור $X$ את אמצע המחלקה.

**דוגמה (הפתרון בהקלטה):**

נתונה רשימת המספרים הבאה: 3, 6, 7, 8, 2.
מה ממוצע הסטיות המוחלטות מהחציון?

## שאלות:

1) נתונה רשימת המספרים הבאה: 3, 5, 6, 9, 12, 8.
מה ממוצע הסטיות המוחלטות מהחציון?

2) להלן התפלגות מספר מקלטי הטלוויזיה שנספרו עבור כל משפחה בישוב מסוים:

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

א. חשבו את החציון.
ב. חשב את ממוצע הסטיות המוחלטות מהחציון.
ג. הסבירו ללא חישוב כיצד כל מדד שחושב היה משתנה, אם 5 משפחות שהיה להם מקלט יחיד היו מוכרים אותו.

## תשובות סופיות:

1) 2.5.
2) א. 1.5. ב. 1.14.
ג. חציון לא ישתנה, ממוצע סטיות מוחלטות מהחציון יגדל.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 8 - סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן

תוכן העניינים

# 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: $Z = \frac{X - \overline{X}}{S}$.

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:

| | 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.
ערן התגייס כשגובהו 180 ס"מ ומשקלו 59 ק"ג.
א. במה ערן חריג יותר ביחס לשאר המתגייסים, גובהו או משקלו?
ב. כמה ערן אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

## תשובות סופיות:

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

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 9 - סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה

תוכן העניינים

# 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 :

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

א. מצאו את האחוזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

## שאלות:

**1)** להלן התפלגות של משתנה כלשהו:

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

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 למדעי המחשב

## פרק 10 - סטטיסטיקה תיאורית - טרנספורמציה לינארית

תוכן העניינים

# 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:

**מדדי המרכז:**

$$MR_y = a \cdot MR + b$$
$$MO_y = a \cdot MO + b$$
$$\bar{y} = a \cdot \bar{x} + b$$
$$Md_y = a \cdot \leq Md_x + b$$

**מדדי הפיזור:**

$$R_y = |a| R_x$$
$$S_y = |a| S_x$$
$$S_y^{\ 2} = a^2 S_x^{\ 2}$$

**מדדי המיקום היחסי:**

$$Y_p = a \cdot X_p + b$$
$$Z_Y = \frac{a}{|a|} Z_x$$

### 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 ₪. חשב 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 למדעי המחשב

## פרק 11 - סטטיסטיקה תיאורית- תרשים קופסא

תוכן העניינים

# סטטיסטיקה תיאורית – תרשים קופסא (Boxplot):

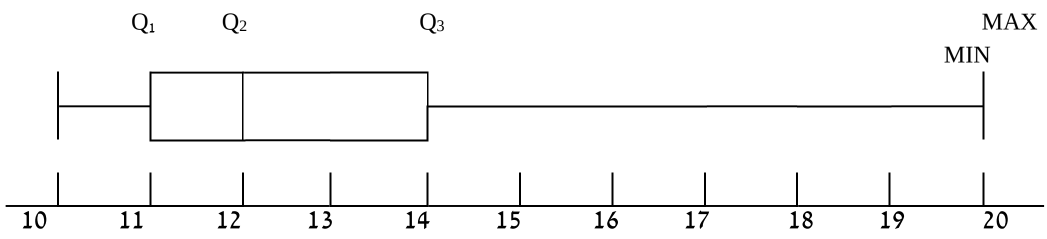
## רקע:

תרשים קופסא הינו תרשים שבעזרתו ניתן לבחון:

**1)** את המרכז של ההתפלגות על ידי החציון ($Q_2$).

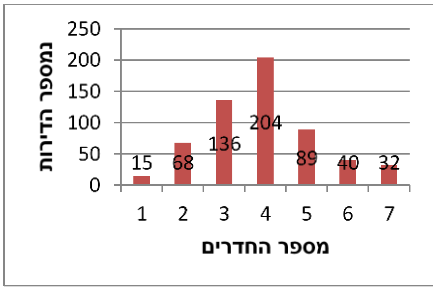
**2)** את הפיזור של הנתונים (הטווח והטווח הבין רבעוני).

**3)** את צורת ההתפלגות (סימטרית\אסימטרית ימנית או אסימטרית שמאלית).



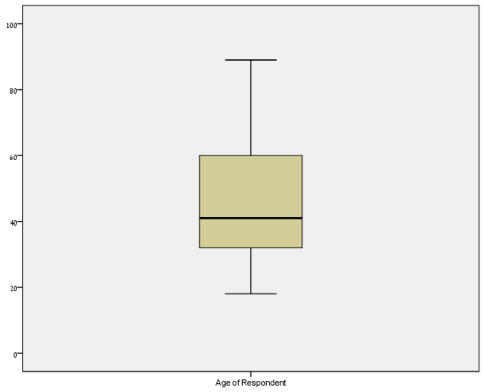
## שאלות:

**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 למדעי המחשב

## פרק 12 - סטטיסטיקה תיאורית - שאלות מסכמות

### תוכן העניינים

# 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 |

א. מהו המשתנה הנחקר? בדיד או רציף?

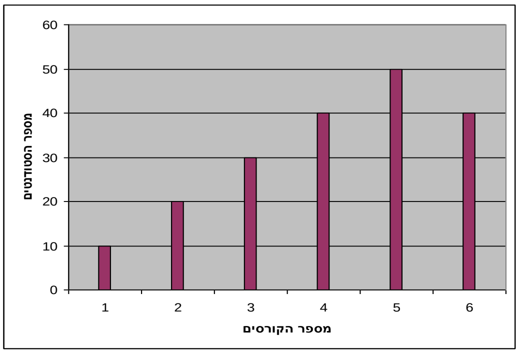
ב. מה המשקל החציוני, הממוצע והשכיח?

ג. מה הטווח וסטיית התקן של המשקל?

ד. לאותם תלמידים חישבו גם את הגובה בס"מ וקיבלו גובה ממוצע של 168 וסטיית תקן 6. במה תלמיד מספר 3, שגובהו 162, יותר חריג – במשקל או בגובה?

ה. 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:




א. מה המשתנה הנחקר? בדיד או רציף?

ב. מהי צורת ההתפלגות?

ג. תארו את הנתונים בטבלת שכיחויות.

ד. חשבו את השכיח, החציון והטווח.

**3)** להלן התפלגות הציונים בבחינה בלשון שנעשתה עבור תלמידי כיתות ד'.
במחקר השתתפו 150 תלמידים. ממוצע הציונים שהתקבל: $\overline{X} = 7\frac{1}{15}$.

| 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. $S=10.12$ , $R=28$ .

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.

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

**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 למדעי המחשב

## פרק 13 - התפלגות הדגימה ומשפט הגבול המרכזי

### תוכן העניינים

# 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 סטית התקן של ממוצע המדגם?

### דגימה מהתפלגות נורמאלית:

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

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

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

### 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.

א. מיהי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו תוחלת ממוצע המדגם?

ו. מהי טעות התקן?

**2)** להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב מסוים:

| מספר המשפחות | מספר מקלטים |
|---|---|
| 500 | 0 |
| 2500 | 1 |
| 3500 | 2 |
| 3000 | 3 |
| 500 | 4 |
| סך הכול<br>$N=10000$ | |

א. בנו את פונקציית ההסתברות של $X$.

ב. חשבו את התוחלת, השונות וסטיית התקן של $X$.

ג. אם נדגום 4 משפחות מהישוב עם החזרה מה תהיה התוחלת, מהי השונות ומהי סטיית התקן של ממוצע המדגם?

**3)** אם נטיל קובייה פעמיים ונתבונן בממוצע התוצאות שיתקבלו, מה תהיה התוחלת ומה תהיה סטיית התקן של ממוצע זה?

**4)** משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.

א. מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?

נתון כי ביום מסוים נולדו 4 תינוקות.

ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק״ג?

ג. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק״ג?

ד. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?

ה. הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?

**5)** הגובה של המתגייסים לצה"ל מתפלג נורמאלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.

א. מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?

ב. מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?

ג. מה ההסתברות שהגובה הממוצע שלהם יסטה מתחולת הגבהים בפחות מ-5 ס"מ?

ד. מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

**6)** 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 נורמאלית.

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?

**7)** נפח היין בבקבוק מתפלג נורמאלית 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. בקבוקיי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?

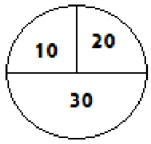
**8)** משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4.

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?

ג. הסבר את ההבדל בתשובות של שני הסעיפים.

**9)** בקזינו ישנה רולטה. על הרולטה רשומים המס' הבאים כמוראה בשרטוט:




אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.
א. בנו את פונקציית ההסתברות של סכום הזכיה במשחק בודד.
ב. מה התוחלת ומה השונות של סכום הזכייה?
ג. אם האדם ישחק את המשחק 5 פעמים מה התוחלת ומה השונות של ממוצע סכום הזכייה בחמשת המשחקים?
ד. אם האדם משחק את המשחק 50 פעם מה ההסתברות שבסה"כ יזכה ב-1050 ₪ ומעלה?

**10)** 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 ₪?

**11)** קובייה הוטלה 50 פעמים.
מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72?

**12)** 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.

**13)** נתון משתנה מקרי בדיד בעל פונקציית ההסתברות הבאה:

| $X$ | 8 | 6 | 4 | 2 |
|---|---|---|---|---|
| $P(X)$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

מתוך התפלגות זו נלקח מדגם מקרי בגודל 50.
מה הסיכוי שממוצע המדגם יהיה קטן מ-5?

**14)** נתון ש- $X \sim N\left(\mu,\sigma^2\right)$. דגמו 5 תצפיות מאותה התפלגות והתבוננו בממוצע המדגם $\bar{X}$. לכן: $P\left(\bar{X} > \mu\right)$ יהיה (בחרו בתשובה הנכונה):

א. 0.

ב. 0.5.

ג. 1.

ד. לא ניתן לדעת.

**15)** 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. $\bar{X} \sim N\left(\mu,\sigma^2\right)$.

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

**16)** נתון ש- $X \sim N\left(\mu,\sigma^2\right)$. 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.

**17)** משקל חפיסת שוקולד בקו ייצור מתפלג עם ממוצע 100 גרם. החפיסות נארזות בקרטון המכיל 36 חפיסות שוקולד אקראיות. ההסתברות שהמשקל הממוצע של חפיסות השוקולד בקרטון יהיה מעל 99 גרם הוא 0.9932.

א. מהי סטיית התקן של משקל חפיסת שוקולד בודדת?

ב. מה הסיכוי שמתוך 4 קרטונים בדיוק קרטון אחד יהיה עם משקל ממוצע לחפיסה הנמוך מ-100 גרם?

**18)** משתנה מקרי כלשהו מתפלג עם סטיית תקן של 20.
מה הסיכוי שאם נדגום 100 תצפיות בלתי תלויות מאותה התפלגות אזי ממוצע המדגם יסטה מתוחלתו בפחות מ-2?

**19)** מספר המכוניות הנכנסות לחניון "בציר" במשך היום מתפלג פואסונית עם קצב של מכונית אחת לדקה. שומר מסר נתונים על מספר המכוניות שנכנסות בכל שעה לגבי 40 שעות שאסף נתונים.
מה ההסתברות שממוצע מספר המכוניות שנכנסו לחניון לשעה בשעות אלה יהיה לפחות 63?

**20)** הוכיחו שאם משתנה מתפלג כלשהו עם תוחלת $\mu$ ושונות $\sigma^2$, ומבצעים מדגם בגודל $n$ של תצפיות בלתי תלויות מהמשתנה, אזי מתקיימות התכונות הבאות לגבי ממוצע המדגם: $E\left(\overline{x}\right)=\mu$ ו- $V\left(\overline{x}\right)=\dfrac{\sigma^2}{n}$.

## 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. להלן טבלה:

| $X$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P(X)$ | 0.05 | 0.25 | 0.35 | 0.3 | 0.05 |

ב. $\mu = 2.05$ , $\sigma^2 = 0.9475$ , $\sigma = 0.973$

ג. $\mu_{\bar{x}} = 2.05$ , $\sigma^2_{\bar{x}} = 0.2369$ , $\sigma\left(\overline{X}\right) = 0.486$

**3)** $\mu_{\bar{x}} = 3.5$ , $\sigma\left(\overline{X}\right) = 1.21$.

**4)** א. 0.8413. ב. 0.0013. ג. 0. ד. 0.1974.

**5)** א. 0. ב. 0. ג. 0.9544. ד. 178.205.

**6)** א. 0.0465. ב. 27.71. ג. 0.2628. ד. התשובה הייתה קטנה.

**7)** א. 0. ב. 0.1587. ג. 0.1587. ד. 0.5.

**8)** א. 0.5468. ב. 0.6826.

**9)** א. להלן טבלה:

| $X$ | 10 | 20 | 30 |
|---|---|---|---|
| $P(X)$ | 0.25 | 0.25 | 0.5 |

ב. התוחלת: 22.5, השונות: 68.75.

ג. התוחלת: 22.5, השונות: 13.75. ד. 0.8997.

**10)** 0.0475.

**11)** 0.1814.

**12)** א. 0.9772. ב. 0.0228. ג. 271.

**13)** 0.5.

**14)** ב'.

**15)** ד'.

**16)** ג'.

**17)** א. 2.429. ב. 0.25.

**18)** 0.6826.

**19)** 0.0071.

**20)** שאלת הוכחה.

# התפלגות סכום תצפיות בלתי תלויות ומשפט הגבול המרכזי:

## רקע:

כעת נדון בסטטיסטי המבטא את סכום התצפיות במדגם: $T=\sum_{i=1}^{n} X_i$.

כאשר כל התצפיות נדגמו באקראי מאותה אוכלוסייה, כלומר, היו: $X_1,...,X_n$ - משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה $\mu$ ושונותה $\sigma^2$ אזי:

התוחלת והשונות של סכום התצפיות: $E(T)=n\mu$ , $V(T)=n\sigma^2$.

### דגימה מתוך התפלגות נורמלית:

אם: $X\sim N(\mu,\sigma^2)$, אזי: $Z=\frac{T-n\mu}{\sqrt{n\sigma^2}}$ , $T\sim N(n\mu,n\sigma^2)$.

### משפט הגבול המרכזי:

אם $X$ מתפלג כלשהו וידוע כי: $V(X)=\sigma^2$, $E(X)=\mu$,

אזי עבור מדגם מספיק גדול (לפחות 30): $T\overset{\sim}{\succ} N(n\mu,n\sigma^2)$.

### דוגמה (פתרון בהקלטה):

בעיר מסוימת המשכורת הממוצעת של עובד הינה 8000 ₪. עם סטיית תקן של 2000 ₪. נדגמו 100 עובדים מהעיר שמפקידים את משכורותיהם לסניף בנק.

1. מה התוחלת וסטיית התקן של סך המשכורות שיופקדו לסניף הבנק על ידי העובדים הללו?
2. מה ההסתברות שלסניף יופקד פחות מ-780 אלף ₪ ע"י אותם עובדים?

## שאלות:

**1)** המשקל באוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת של 60 ק״ג וסטיית תקן של 10 ק״ג.

א. מה הסיכוי שאדם אקראי מהאוכלוסייה ישקול מתחת ל-65 ק״ג?

ב. מה הסיכוי שהמשקל הממוצע של 4 אנשים אקראיים יהיה מתחת ל-65 ק״ג?

ג. מה הסיכוי שהמשקל הכולל של 4 אנשים אקראיים יהיה מתחת ל-240 ק״ג?

**2)** נפח יין בבקבוק מתפלג נורמאלית עם תוחלת של 750 מ״ל וסטיית תקן של 20 מ״ל. אדם קנה מארז של 4 בקבוקי יין.

א. מהי התוחלת ומהי סטיית התקן של נפח היין במארז?

ב. את היין שבמארז האדם מזג לכלי שקיבולתו 3.1 ליטר. מה ההסתברות שהיין יגלוש מהכלי?

ג. אם לא היה נתון שנפח היין מתפלג נורמאלית. האם התשובה לסעיף א׳ הייתה משתנה? האם התשובה לסעיף ב׳ הייתה משתנה?

**3)** בספר כלשהו 500 עמודים. קצב הקריאה הממוצע הוא עמוד אחד ב-4 דקות עם סטיית תקן של 1 דקות.

א. מה ההסתברות לסיים את הפרק הראשון (40 עמודים) תוך שעתיים וחצי?

ב. מהו האחוזון ה-95 לזמן סיום קריאת הספר?

**4)** במגדל נבנו 40 יחידות דיור. כמו כן נבנו 135 מקומות חנייה לבניין. להלן פונקציית ההסתברות של מספר המכוניות ליחידת דיור:

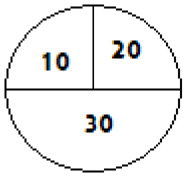
| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $P(X=x)$ | 0.1 | 0.2 | 0.3 | 0.25 | 0.15 |

נניח שמספר המכוניות ליחידת דיור בלתי תליות זו בזו ועם אותה פונקציית הסתברות לכל יחידת דיור (אין צורך בתיקון רציפות).

א. מהי ההסתברות שיהיה מקום בחניון המגדל לכל מכוניות הבניין?

ב. בהינתן ויש מקום במגדל לכל המכוניות, מה הסיכוי שבפועל מספר המכוניות נמוך מ-130?

5) בקזינו ישנה רולטה עליה מסומנים המספרים הבאים:




אדם מסובב את הרולטה וזוכה בסכום הרשום.

א. אם האדם משחק 50 פעמים, מה ההסתברות שבסך הכול יזכה בסכום של 1050 ₪ ומעלה?

ב. האדם מגיע בכל יום לקזינו ומשחק 50 פעם, עד אשר מגיע היום בו הוא יזכה ב-1050 ₪ ומעלה. מה התוחלת ומהי השונות של מספר הימים שיבלה בקזינו?

6) נתון ש- $X_i \sim \exp(\lambda=1)$, כאשר: $i=1,2...100$.

חשבו את הסיכוי: $P\left(\sum_i X_i \geq 115\right)$.

7) אורך חיי סוללה (בשעות) הוא בעל פונקציית הצפיפות הבאה: $0<x<1$, $f(x)=2x$. ברגע שסוללה מתרוקנת מחליפים אותה במיידית בסוללה אחרת. כמה סוללות יש להחזיק במלאי אם רוצים שבסיכוי של 90% לפחות המלאי יספיק עבור 35 שעות לפחות?

## תשובות סופיות:

1) א. 0.6915. ב. 0.8413. ג. 0.5.

2) א. תוחלת: 3000 מ"ל, סטיית תקן: 40 מ"ל. ב. 0.0062.
ג. סעיף א' - לא משתנה, סעיף ב' - לא פתיר, התבסס על התפלגות נורמלית.

3) א. 0.0571. ב. 2036.8.

4) א. 0.883. ב. 0.7949.

5) א. 0.8997. ב. תוחלת 1.111, שונות 0.1239.

6) 0.0668.

7) 56.

# התפלגות מספר ההצלחות במדגם – קירוב נורמלי להתפלגות הבינומית:

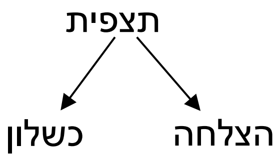
## רקע:

**תזכורת על התפלגות בינומית:**

בפרק זה נדון בהתפלגות מספר ההצלחות במדגם אקראי (תצפיות בלתי תלויות זו בזו).

את מספר ההצלחות במדגם נסמן ב- $Y$.

מחלקים כל תצפית במדגם להצלחה או כישלון.




כעת מה שמשתנה מתצפית לתצפית הוא משתנה דיכוטומי (משתנה שיש לו שני ערכים).

הסיכוי להצלחה יסומן עם הפרמטר $p$ וכישלון יסומן ע"י הפרמטר: $q=1-p$.

מבצעים מדגם אקראי בגודל $n$ : $Y \sim B(n,p)$.

פונקציית ההסתברות של ההתפלגות הבינומית היא: $p(y=k)=\binom{n}{k}p^k q^{n-k}$ ,

תוחלת: $E(y)=np$.

שונות: $V(y)=npq$.

**קירוב נורמלי עבור התפלגות בינומית:**

אם לפנינו התפלגות בינומית: $Y \sim B(n,p)$, ומתקיים ש:

1. $n \cdot p \geq 5$.
2. $n \cdot (1-p) \geq 5$.

$$y \sim\succ N(np, npq)$$

אז:

$$Z_y = \frac{y-np}{\sqrt{npq}}.$$

**תיקון רציפות:**

כאשר משתמשים בקירוב הנורמלי להתפלגות הבינומית יש לבצע תיקון רציפות.
הסיבה שעוברים כאן מהתפלגות בדידה להתפלגות נורמלית שהיא התפלגות רציפה.
על פי הכללים הבאים:

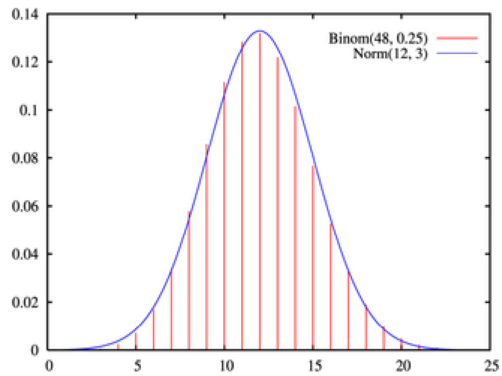
1. $p(Y=a)\cong p\left(a-\frac{1}{2}\le Y\le a+\frac{1}{2}\right)$.
2. $P(Y\le a)\cong P(Y\le a+0.5)$.
3. $P(Y\ge a)\cong P(Y\ge a-0.5)$.

**הערות:**

- התנאים למעבר מבינומי לנורמלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:
  1. $n\cdot p\ge 5$.
  2. $n\cdot(1-p)\ge 5$.
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
  1. $n\cdot p\ge 10$.
  2. $n\cdot(1-p)\ge 10$.
- וישנם מרצים שהתנאי שהם נותנים הוא: $(n\ge 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור מהתפלגות בינומית לנורמלית.
- הערה נוספת היא לגבי תיקון רציפות. ישנם מרצים שלא מחייבים לבצע תיקון רציפות שהמדגמים גדולים (בדרך כלל מעל 100 תצפיות) בפתרונות שאציג תמיד אבצע תיקון רציפות במעבר מבינומי לנורמלי כיוון שכך הפתרון יהיה יותר מדויק (בכל מקרה שהמדגמים גדולים העניין זניח).

**דוגמה (הפתרון בהקלטה):**
נתון שבקרב אוכלוסיית הנוער 25% זקוקים למשקפיים. נדגמו באקראי 48 בני נוער.

1. מה הסיכוי שבדיוק 14 מתוכם יהיו זקוקים למשקפיים?
2. מה הסיכוי שלכל היותר 13 מתוכם זקוקים למשקפיים?



## שאלות:

**1)** נתון ש-20% מאוכלוסייה מסוימת אקדמאית. נבחרו באקראי 10 אנשים באותה אוכלוסייה.

א. מה ההסתברות ששלושה מהם אקדמאים?

ב. מה ההסתברות שלכל היותר אחד מהם אקדמאי?

ג. מה התוחלת ומהי סטיית התקן של מספר האקדמאים במדגם?

**2)** במפעל 10% מהמוצרים פגומים. נלקחו 100 מוצרים באקראי מקו הייצור.

א. מה ההסתברות שנדגמו לפחות 6 מוצרים פגומים?

ב. מה ההסתברות שמספר המוצרים הפגומים יהיה לכל היותר 11 במדגם?

**3)** ציוני פסיכומטרי בקרב הנרשמים למוסד מסוים מתפלגים נורמאלית עם ממוצע 500 וסטיית תקן 100. למוסד מסוים הוחלט לקבל אך ורק סטודנטים שקיבלנו מעל 600 בפסיכומטרי. 100 סטודנטים אקראיים נרשמו למוסד. מה ההסתברות שלפחות 20 יתקבלו?

**4)** מטילים מטבע 50 פעמים.

א. מה ההסתברות לקבל לכל היותר 30 עצים?

ב. מה ההסתברות לקבל 28 עצים לפי התפלגות הבינומית ולפי הקירוב הנורמאלי?

**5)** במטוס מקום ל-400 נוסעים. נרשמו לטיסה 430 אנשים (overbooking). מנתונים סטטיסטיים ידוע שהסיכוי שאדם שנרשם לטיסה אכן יגיע הוא 0.9.

א. מה ההסתברות שלא יהיו מקומות ישיבה לכל האנשים שהגיעו לטיסה?

ב. מה צריך להיות גודל המטוס כדי שבסיכוי שלפחות 95% המטוס יספיק לכמות הנרשמים?

**6)** מפעל לייצור ארטיקים טוען שהסיכוי שארטיק שהוא מייצר יהיה פגום הוא 0.01. מוכר הזמין 1000 ארטיקים מהמפעל. מה ההסתברות שהמוכר יקבל לפחות 980 ארטיקים תקינים אם טענת המפעל מוצדקת?

**7)** מהמר מטיל קובייה הוגנת 100 פעמים. בכל הטלה, אם מתקבל תוצאה זוגית בקובייה המהמר זוכה בשקל. אחרת, המהמר משלם שקל. המהמר הטיל את הקובייה 100 פעמים מה הסיכוי שהרווח של המהמר יהיה לכל היותר 10?

## תשובות סופיות:

1) א. 0.201. ב. 0.3758. ג. התוחלת: 2, סטיית התקן: 1.2649.
2) א. 0.9332. ב. 0.6915.
3) 0.1611.
4) א. 0.9406. ב. בינומית - 0.0788, קירוב לנורמלית - 0.0778.
5) א. 0.015. ב. 398.
6) 0.9996.
7) 0.8643.

# התפלגות פרופורציית ההצלחות במדגם:

### רקע:

בפרק זה נדון בהתפלגות הדגימה של פרופורציית המדגם.

$Y$ - מספר ההצלחות במדגם (למשל, מספר המובטלים במדגם).

$\hat{p}=\frac{y}{n}$ - פרופורציית ההצלחות במדגם.

למשל, שיעור המובטלים במדגם - $n=200$ :

מספר המובטלים: $Y=20$.

פרופורציית המובטלים במדגם: $\hat{p}=\frac{20}{200}=0.1$.

נסמן ב- $p$ את שיעור ההצלחה באוכלוסייה וב- $q$ את שיעור הכישלונות באוכלוסייה.

נבצע מדגם מקרי (הנחה שהתצפיות בלתי תלויות זו בזו) ונתבונן בהתפלגות של פרופורציית המדגם.

**התוחלת, השונות וסטיית התקן של פרופורציית המדגם:**

$E(\hat{p})=p \quad , \quad V(\hat{P})=\frac{pq}{n}$.

משפט הגבול המרכזי עבור הפרופורציה המדגמית:

אם: $np\geq 5 \,\&\, nq\geq 5$, אזי: $\hat{p} \sim N\left(p,\frac{pq}{n}\right)$. $Z_{\hat{p}}=\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$.

**הערות:**

- התנאים לקרוב הנורמאלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:
  1. $n\cdot p\geq 5$.
  2. $n\cdot(1-p)\geq 5$.
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
  1. $n\cdot p\geq 10$.
  2. $n\cdot(1-p)\geq 10$.
- וישנם מרצים המשתמשים בתנאי: $(n\geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור לנורמלית.

- כיוון שפרופורציה אינה חייבת להיות מספר שלם בהכרח לא נהוג לבצע כאן תיקון רציפות.

**דוגמה (פתרון בהקלטה):**

לפי נתוני משרד החינוך בעיר ירושלים ל-60% מתלמידי התיכון זכאים לתעודת בגרות. נדגמו 200 תלמידי תיכון.

א. מה ההסתברות שהשכיחות היחסית $\left(\hat{p}\right)$ של הזכאים לבגרות במדגם תעלה על 60%?

ב. מה ההסתברות שפרופורציית הזכאים לבגרות במדגם תעלה על 70%?

## שאלות:

1) במדינה מסוימת 10% מכלל האוכלוסייה הינם מובטלים.
נדגמו באקראי 140 אנשים מהמדינה.
א. מה התוחלת ומהי השונות של פרופורציות המובטלים שנדגמו?
ב. מה ההסתברות שבמדגם לפחות 10% יהיו מובטלים?
ג. מה ההסתברות שלכל היותר 9% מהמדגם יהיו מובטלים?

2) נניח כי 30% מהאוכלוסייה תומכים בהצעת חוק מסוימת. אם נדגום מהאוכלוסייה 200 איש. חשבו את ההסתברויות הבאות:
א. לפחות 35% יתמכו בהצעת החוק במדגם.
ב. לכל היות 25% יתמכו בהצעת החוק במדגם.
ג. יותר מ-27% יתמכו בהצעת החוק במדגם.

3) לפי נתוני משרד התקשורת 40% מהאוכלוסייה מחזיקים בטלפון נייד מסוג "סמארטפון". נדגמו 400 אנשים מהאוכלוסייה.
א. מה ההסתברות שבמדגם לכל היותר ל-40% יש סמארטפון?
ב. מה ההסתברות שבמדגם לרוב יש סמאטרפון?
ג. מה ההסתברות שפרופורציית בעלי הסמרטפון במדגם תסטה מהפרופורציה באוכלוסייה בלא יותר מ-4%?
ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם הינו מגדילים את גודל המדגם?

4) נתון כי 80% מבתי האב מחוברים לאינטרנט. נדגמו 400 בתי אב אקראיים.
א. מה ההסתברות שלפחות 340 מהם מחוברים לאינטרנט?
ב. מה ההסתברות שפרופורציית המחוברים לאינטרנט במדגם תסטה מהפרופורציה האמיתית ביותר מ-4%?
ג. כמה בתי אב יש לדגום כדי שהסטייה בין הפרופורציה המדגמית לפרופורציה האמיתית לא תעלה על 3% בהסתברות של 90%?
ד. מהו העשירון התחתון של התפלגות פרופורציית המדגם?

5) נתון שציוני פסיכומטרי מתפלגים נורמלית עם תוחלת 500 וסטיית תקן 100.
ב"מועדון ה-700" נכללים נבחנים שמקבלים ציון מעל 700 בפסיכומטרי.
מה הסיכוי שבמועד בו נבחנו 2000 נבחנים אקראיים יהיו לפחות 3% המשתייכים למועדון?

6) נתון ש- $X \sim B(n,p)$ , ונגדיר את המשתנה הבא: $\hat{P}=\frac{X}{n}$ .

א. הוכיחו ש: $V(\hat{P})=\frac{P(1-P)}{n}$ , $E(\hat{P})=p$ .

ב. מה $p$ המביא את $V(\hat{P})$ להיות במקסימום?

## תשובות סופיות:

1) א. התוחלת: 0.1, השונות: 0.00064. ב. 0.5. ג. 0.3446.

2) א. 0.0618. ב. 0.0618. ג. 0.8238.

3) א. 0.5. ב. 0. ג. 0.8968. ד. גדלה.

4) א. 0.0062. ב. 0.0456. ג. 481. ד. 0.77436.

5) 0.0154.

6) א. שאלת הוכחה. ב. 0.5.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 14 - הסקה סטטיסטית - הקדמה

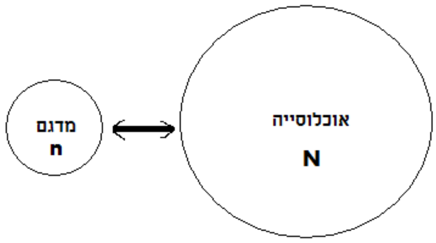
תוכן העניינים

# 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** | $\overline{X}$ | $\mu$ |
| **פרופורציה (שכיחות יחסית)** | $\hat{p}$ | $P$ |

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.
א. מי האוכלוסייה?
ב. מה המשתנה?
ג. מהם הפרמטרים?
ד. מהו גודל המדגם?

2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב "העוגן".
נגדיר את $X$ להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.
א. מיהי האוכלוסייה ומהו המשתנה הנחקר?
ב. 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.
א. מהי האוכלוסייה?
ב. מה המשתנה באוכלוסייה?
ג. מהם הפרמטרים?
ד. 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 בישוב 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 למדעי המחשב

## פרק 15 - אמידה נקודתית

### תוכן העניינים

# אומד חסר הטייה:

## רקע:

$\hat{\theta}$ יהיה אומד חסר הטיה ל-$\theta$, אם התוחלת של $\hat{\theta}$ תהיה שווה ל-$\theta$ : $E(\hat{\theta})=\theta$.

**דוגמה (פתרון בהקלטה):**

המשתנה $X$ הוא בעל פונקציית ההסתברות הבאה:

| $X$ | 1 | 2 | 3 |
|---|---|---|---|
| **הסתברות** | $2\theta$ | $1-60\theta$ | $4\theta$ |

מעוניינים לאמוד את $\theta$ על סמך שתי תצפיות מההתפלגות: $X_1$ ו-$X_2$.

א. הראו שהאומד: $T_1=\frac{2X_1+X_2}{2}$, הוא אומד מוטה ל-$\theta$.

הטיה של אומד היא: $E(\hat{\theta})-\theta$. כמובן שלאומד חסר הטיה אין הטיה.

ב. מהי ההטיה של האומד $T_1$?

ג. תקנו את $T_1$, כך שיהיה אומד חסר הטיה.

אם יש שני אומדים חסרי הטיה עדיף זה עם השונות היותר קטנה.

ד. מוצא האומד הבא: $T_3=1.5X_1-X_2-1$.

האם הוא עדיף על האומד שהצעת בסעיף ג'?

אם $\hat{\theta}$ אומד חסר הטיה ל-$\theta$, אז $g(\hat{\theta})$ יהיה אומד חסר הטיה עבור $g(\theta)$, רק אם $g$ תהיה לינארית.

ה. מצאו אומד חסר הטיה ל: $P(X=3)$.

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

ו. מצאו אומד חסר הטיה לשונות של $X$.

**תזכורות חשובות:**

אם: $Y=aX+b$, אזי: $E(Y)=aE(X)+b$ , $V(Y)=a^2\cdot V(X)$ , $\sigma_Y=|a|\sigma_x$.

אם: $X_1,X_2,.......,X_n$ משתנים מקריים, אזי:

$E(T)=E(X_1+X_2+......+X_n)=E(X_1)+E(X_2)+......+E(X_n)$.

אם: $X_1,X_2,.......,X_n$ משתנים מקריים בלתי תלויים בזוגות, אזי:

$V(T)=V(X_1+X_2+......+X_n)=V(X_1)+V(X_2)+......+V(X_n)$.

## שאלות:

**1)** הציון במבחן מסוים של תלמידי כתה ח׳ הנו משתנה מקרי בעל תוחלת $\mu$ וסטיית תקן 10. כדי לאמוד את התוחלת $\mu$, נלקח מדגם של 5 ציונים: $X_1,\dots,X_5$.
שלושה חוקרים הציעו אומדים לתוחלת על סמך מדגם זה:

חוקר א׳ הציע: $T_1=\dfrac{X_1+\dots+X_5}{5}$.

חוקר ב׳ הציע: $T_2=\dfrac{2X_1-X_3+X_4}{2}$.

חוקר ג׳ הציע: $T_3=\dfrac{2X_1+X_3}{2}$.

א. איזה מן האומדים הוא חסר הטיה?
ב. הציעו תיקון לאומד המוטה כך שיהיה חסר הטיה.
ג. במדגם התקבלו הציונים הבאים: 65, 78, 58, 82, 100.
חשבו את האומדנים המתקבלים עבור האומדים חסרי ההטיה.
ד. איזה מבין שני האומדים חסרי ההטיה עדיף? נמקו.

**2)** כדי לאמוד את המשקל הממוצע של הנשים בארה״ב, נבחר מדגם של $2n$ נשים.
נסמן את שונות הגובה ב- $\sigma^2$. הוצעו שני אומדים לממוצע המשקל על סמך מדגם זה: $T_1=\dfrac{1}{n}\sum_{i=1}^{n}X_i$ , $T_2=\dfrac{1}{2n}\sum_{i=1}^{2n}X_i$.

א. בדקו לגבי כל אומד אם הוא בלתי מוטה.
ב. איזה אומד עדיף? נמקו.

**3)** $X\sim B(n,p)$. כלומר, $X$ הינו משתנה מקרי המתפלג בינומית עם פרמטר $P$ (סיכוי להצלחה בניסיון בודד) במדגם בגודל $n$.

א. פתחו אומד חסר הטיה ל- $P$.
ב. מהו אומד חסר הטיה לסיכוי לכישלון בניסיון בודד?
ג. מהו אומד חסר הטיה ל- $E(X)$?
ד. מצאו אומד חסר הטיה ל- $E\left(X^2\right)$.

**4)** בתיק מניות שתי מניות. מספר המניות שיעלו ביום מסוים הוא משתנה מקרי התלוי בפרמטר לא ידוע: $\theta$, $0 \le \theta \le 2$.
פונקציית ההסתברות של $X$ - מספר המניות שיעלו ביום מסוים:
$$P(X=0)=1-\frac{\theta}{2}\ ,\ P(X=1)=\frac{\theta}{3}\ ,\ P(X=2)=\frac{\theta}{6}.$$
א. מְצאו אומד בלתי מוטה ל-$\theta$, שמתבסס על מספר המניות שיעלו ביום מסוים.
ב. מצאו אומד בלתי מוטה ל-$\theta$, שמתבסס על מספר המניות שעלו ביום, במשך שלושה ימים - $X_1, X_2, X_3$ (לכל אחד מהם אותה התפלגות כנ"ל והם בלתי תלויים).

**5)** בקרב המטפלות בת"א, מספר התינוקות שבטיפולן הוא משתנה מקרי בעל התפלגות התלויה בפרמטר $\theta$ באופן הבא:
הסיכוי שמטפלת תטפל בתינוק אחד בלבד הוא $3\theta$,
הסיכוי שמטפלת תטפל ב-2 תינוקות הוא $1-4\theta$,
הסיכוי שמטפלת תטפל ב-3 תינוקות הוא $\theta$.
במדגם מיקרי של 4 מטפלות מת"א, נמצא כי שתים מהם מטפלות בתינוק אחד בלבד, אחת מהן בשנים ואחת השלושה תינוקות.
א. מצאו אומד חסר הטיה לפרמטר $\theta$ על סמך תצפית בודדת.
ב. מצאו אומד חסר הטיה לפרמטר $\theta$ על סמך 4 תצפיות.
ג. מהו האומדן לפרמטר $\theta$ על סמך תוצאות המדגם.
ד. מצאו אומד חסר הטיה לסיכוי שלמטפלת בת"א תטפל בתינוק בודד אחד.
ה. מצאו אומדים חסרי הטיה לתוחלת ולשונות של מספר התינוקות בטיפול אצל מטפלת מת"א. חשבו אומדנים.

**6)** קבעו אילו מהטענות הבאות נכונות:
א. אם $T$ הוא אומד בלתי מוטה עבור פרמטר $\theta$, אז $5T$ אומד בלתי מוטה עבור הפרמטר $5\theta$.
ב. אם $T$ הוא אומד בלתי מוטה עבור פרמטר $\theta$, אז $T^2$ אומד בלתי מוטה עבור הפרמטר $\theta^2$.

**7)** במפעל שתי מכונות המייצרות מוצרים. במכונה הראשונה ההסתברות שמכשיר תקין היא $p$ , ובמכונה השנייה ההסתברות שמכשיר תקין היא $2p$.
דוגמים 20 מכשירים מהייצור של כל מכונה.
נסמן ב- $X$ את מספר המכשירים התקינים שיוצרו על ידי המכונה הראשונה,
וב- $Y$ את מספר המכשירים התקינים שיוצרו על ידי המכונה השנייה.
איזה מבין האומדים הבאים אינו אומד חסר הטיה ל- $p$ ?

א. $\frac{X}{20}$.

ב. $\frac{Y}{20}$.

ג. $\frac{X+Y}{60}$.

ד. $\frac{2X+Y}{80}$.

**8)** יהיו $T_1$ ו- $T_2$ אומדים חסרי הטיה ובלתי תלויים לפרמטר $\theta$.

א. מצאו אומד חסר הטיה ל- $\theta^2$, המתבסס על $T_1$ ו- $T_2$.

ב. מצאו אומד חסר הטיה ל- $\theta(1-\theta)$, המתבסס על $T_1$ ו- $T_2$.

**9)** נתון ש- $X$ הינו משתנה מקרי עם תוחלת $\mu$ ושונות $\sigma^2$.
נדגמו $n$ תצפיות בלתי תלויים מאותה אוכלוסיה.

א. הראו ש- $\sum_{i=1}^{n} p_i x_i$ אומד חסר הטייה ל- $\mu$ , כאשר: $\sum_{i=1}^{n} p_i = 1$.

ב. נתבונן במכפלת שתי התצפיות הראשונות: $X_1 \cdot X_2$.
הראו שהוא אומד חסרי הטיה ל- $\mu^2$.

**10)** $X_i \sim N(\mu,1)$ , כאשר: $i = 1,2...,n$.
נתון שהתצפיות הינן בלתי תלויות זו בזו.
מצאו אומד חסר הטיה ל- $\mu^2$.

**11)** נתונות $n$ תצפיות בלתי תלויות מתוך התפלגות בעלת הצפיפות הבאה:

$$f(x)=\begin{cases}\frac{1+\beta x}{2} & -1<x<1\\ 0 & else\end{cases}\ .$$

א. הראו כי האומד $3\bar{X}$ הנו אומד בלתי מוטה ל-$\beta$.

ב. מצאו את השונות של האומד מהסעיף הקודם.

**12)** $X_1,X_2,\ldots,X_n$ הינם משתנים מקריים רציפים בלתי תלויים בעלי פונקציית הצפיפות הבאה:

$$f(x)=\begin{cases}X\cdot A & 0\le x\le\theta\\ 0 & \text{אחרת}\end{cases}\ .$$

א. בטאו את ערכו של $A$ באמצעות $\theta$, כדי שפונקציית הצפיפות תהיה לגיטימית.

ב. מצאו אומד חסר הטיה ל-$\theta$, על סמך $n$ התצפיות.

## תשובות סופיות:

1) א. $T_1$ ו- $T_2$. ב. $\frac{2}{3}T_3$. ג. $T_1=76.6$ , $T_2=110$ ד. $T_1$.

2) א. ראו בווידאו. ב. $T_2$.

3) א. $\frac{x}{n}$. ב. $1-\frac{x}{n}$ ג. $X$. ד. $\theta$.

4) א. $\frac{3x}{2}$. ב. $\frac{3\overline{x}}{2}$.

5) א. $1-\frac{x}{2}$. ב. $1-\frac{1}{2}\overline{x}$. ג. 0.125. ד. $3\left(1-\frac{1}{2}\overline{x}\right)$.
ה. לשונות 0.917.

6) א. נכון. ב. לא נכון.

7) ב'.

8) א. $T_1 \cdot T_2$. ב. $T_1 - T_1 \cdot T_2$.

9) א. שאלת הוכחה. ב. שאלת הוכחה.

10) $\overline{X}^2-\frac{1}{n}$.

11) א. שאלת הוכחה. ב. $V\left(3\overline{X}\right)=\frac{3-\beta^2}{n}$.

12) א. $A=\frac{2}{\theta^2}$. ב. $\theta=\frac{3}{2}\overline{X}$.

# אומד נראות מקסימלית:

## רקע:

להלן נלמד את שיטת הנראות המקסימלית למציאת אומדים.

נניח ש- $X$ משתנה מקרי בדיד עם פונקציית הסתברות $P(x,\theta)$,

כאשר $\theta$ הפרמטר הבלתי ידוע.

יהיו: $X_1, X_2, ..., X_n$ תוצאות מדגם מקרי בגודל $n$ הנלקח מאוכלוסייה זו.

נבנה את פונקציית ההסתברות המשותפת (פונקציית הדגימה).

אם אנו יודעים את תוצאות המדגם, ולא את הפרמטר, קוראים לפונקציית הנראות שהיא פונקציה של הפרמטר.

נגדיר את פונקציית הנראות:

$$L(\theta)=P(x_1,\theta)\cdot P(x_2,\theta)\cdot ...\cdot P(x_n,\theta)=\prod_{i=1}^{n}P(x_i,\theta).$$

פונקציית הנראות היא ההסתברות לקבל את התצפית הראשונה (כפונקציה של $\theta$), כפול ההסתברות לקבל את התצפית השנייה, וכולי. כלומר, המשמעות של פונקציית הנראות היא ההסתברות לקבל את המדגם שהתקבל, כפונקציה של הפרמטר המבוקש $\theta$.

אם מדובר במשתנה רציף, נכפיל את פונקציות הצפיפות ולא את פונקציות ההסתברות:

$$L(\theta)=f(x_1,\theta)\cdot f(x_2,\theta)\cdot ...\cdot f(x_n,\theta)=\prod_{i=1}^{n}f(x_i,\theta).$$

**דוגמה (פתרון בהקלטה):**

הסיכוי של שחקן כדורסל לקלוע לסל הוא $p$ (לא ידוע). השחקן זורק כדורים לסל עד שהוא קולע בפעם הראשונה. נניח כי הזריקות בלתי תלויות זו בזו.
הכדור נכנס לסל לראשונה בניסיון השלישי.
השחקן חוזר על התהליך שוב, והפעם הכדור נכנס לסל בניסיון החמישי.
מצאו את פונקציית הנראות של $p$.

אומד נראות מקסימלית עבור $\theta$ הוא האומד $\hat{\theta}$, שממקסם את פונקציית הנראות $L(\theta)$.
כלומר, אנו מחפשים את האומד שיגרום לכך שהמדגם המקרי שקיבלנו יהיה כמה שיותר סביר.

**שלבים למציאת אומד נראות מקסימלית:**

- לוקחים את פונקציית ההסתברות המשותפת של המדגם (או צפיפות משותפת אם המשתנה רציף).
- מציבים את תוצאות המדגם ומקבלים את פונקציית הנראות (פונקציה של הפרמטר הנחקר).
- מוצאים מקסימום לפונקציית הנראות (לעיתים כדאי להוסיף $ln$ כדי להקל על המלאכה).

**המשך דוגמה:**

חשבו את אומדן הנראות המקסימלית עבור $p$.

**משפט:** אם $\hat{\theta}$ הוא אומד נראות מקסימלית עבור $\theta$, אזי $g\left(\hat{\theta}\right)$ הוא אומד נראות מקסימלית עבור $g\left(\hat{\theta}\right)$, בהנחה והפונקציה היא חד-חד ערכית (אינווריאנטיות).

**המשך דוגמה:**

מצאו אומדן נראות מקסימלית לסיכוי של שחקן הכדורסל לקלוע לסל פעמיים ברצף.

## שאלות:

**1)** הסיכוי של שחקן לנצח במשחק הוא $p$ (לא ידוע).
השחקן משחק במשחק עד אשר הוא מנצח בפעם הראשונה.
נתון שהשחקן ניצח לראשונה רק במשחק השני.
א. חשבו את פונקציית הנראות של $p$, וציירו גרף שלה.
ב. מצאו אומדן נראות מקסימלית עבור $p$.
ג. מצאו אומדן נראות מקסימלית ל- $p$, אם ביום אחד הוא נאלץ לשחק 4 פעמים וביום אחר הוא נאלץ לשחק 5 פעמים, עד אשר ניצח.

**2)** מספר הלקוחות שנכנסים לחנות מסוימת, מתפלג פואסונית עם תוחלת של $\lambda$ לקוחות ביום.
א. מצאו אומד נראות מקסימלית ל- $\lambda$, על סמך מספר הלקוחות שנכנסים ביום מסוים.
ב. מצאו אומד נראות מקסימלית ל- $\lambda$, על סמך מספר הלקוחות שנכנסים ב- $n$ ימים מסוימים.

**3)** הזמן שלוקח לאדם לחכות בתור מתפלג מעריכית עם פרמטר $\lambda$.
דגמו 4 אנשים מקריים שחיכו בתור ומדדו את זמני ההמתנה שלהם.
התוצאות שהתקבלו בדקות הן: 3, 7,5 ו-3.
א. פתחו אומד נראות מקסימלית לפרמטר זה על סמך $n$ תצפיות כלשהן.
ב. מהו האומדן לפרמטר?

**4)** משך זמן הכנת שיעורי הבית (בשעות) של בני נוער, ביום אחד, מתפלג אחיד: $U(0,q)$.
כדי לאמוד את $\theta$, נשאלו ביום מסוים מספר בני נוער כמה שעות הם הכינו שיעורי-בית באותו יום.
א. אלעד הכין ביום מסוים שעורי בית במשך שעה שלמה. חשבו את פונקציית הנראות של $\theta$ המתבססת על תצפית זו, וציירו את הגרף שלה.
ב. מצאו אומדן נראות מקסימלית ל- $\theta$ על סמך התצפית.
ג. משכי הכנת שיעורי בית (שעות) של 3 בני נוער היו 1,3,1.5.
מצאו אומדן נראות מקסימלית ל- $\theta$ על סמך המדגם הזה.
ד. מצאו באופן כללי אומד נראות מקסימלית ל- $\theta$, על סמך מדגם של $n$ בני נוער – $X_1,\dots,X_n$.

5) הגובה של אוכלוסייה מסוימת מתפלג נורמלית עם תוחלת ידועה של 170 ס"מ ושונות $\sigma^2$ לא ידועה.

א. מצאו אומד נראות מקסימלית עבור השונות על סמך מדגם $X_1,\dots,X_n$ מתצפיות מהאוכלוסייה.

ב. נדגמו 5 אנשים בלתי תלויים בעלי הגבהים: 170, 182, 165, 174, 174. מהו האומדן לשונות הגבהים באוכלוסייה?

6) פתחו אומד נראות מקסימלית לפרמטר $p$ בהתפלגות הבינומית, על סמך מדגם בגודל $n$, בו $X$ הוא מספר ההצלחות במדגם.

7) $X$ הוא משתנה מקרי בעל פונקציית הצפיפות: $f(x)=\begin{cases}2\theta xe^{-\theta x^2} & x>0\\ 0 & x\le 0\end{cases}$.

א. מצאו אומד נראות מקסימלית ל-$\theta$ על סמך $n$ תצפיות בלתי תלויות: $X_1,\dots,X_n$.

ב. מצאו אומד נראות מקסימלית ל-$\theta^2$.

8) בכד א' 10 כדורים שחורים ו-10 לבנים ובכד ב' 5 כדורים שחורים ו-15 לבנים. דוגמים באקראי כדור, בלי לדעת מאיזה כד.

א. מצא אומד נראות מקסימלית לכד שממנו הוצא הכדור על סמך הצבע של הכדור.

ב. מהו האומדן אם הצבע הוא שחור?

9) הזמן שלוקח ליוסי לפתור תשבץ מתפלג מעריכית עם תוחלת לא ידועה. נתנו ליוסי לפתור חמישה תשבצים ובממוצע לקח לו 32 דקות לפתור אותם.

א. מה אומדן הנראות המקסימלית לתוחלת זמן הפתרון של תשבץ על ידי יוסי (אין חובה לפתח).

ב. מה אומדן הנראות המקסימלית לסיכוי שייקח לו לפחות חצי שעה לפתור את התשבץ הבא?

**10)** מספר הלקוחות הממתינים בתור במוקד טלפוני הוא משתנה מקרי $X$, בעל התפלגות התלויה בפרמטר $\theta$, באופן הבא:

| $X$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X)$ | $4\theta^2$ | $4\theta-8\theta^2$ | $1-4\theta+4\theta^2$ |

בחמישה זמנים שונים שנבחרו באקראי נמצאו: 0, 0, 0, 1, 0 לקוחות ממתינים בתור.

א. מצאו אומדן בשיטת הנראות המקסימלית עבור הפרמטר $\theta$, על-סמך המדגם הנתון.

ב. מצאו אומדן בשיטת הנראות המקסימלית לסיכוי שלא יהיו לקוחות בתור.

**11)** אדם מחזיק בידו שני מטבעות: מטבע הוגן ומטבע שאינו הוגן – שהסיכוי לקבל בו תוצאה של עץ הוא 0.2. האדם מטיל את אחד המטבעות פעמיים ומודיע לך כמה פעמים הוא קיבל עץ. אתה צריך לנחש איזה מטבע הוא הטיל: את ההוגן או את זה שאינו הוגן.

א. מצא אומד בשיטת הנראות המקסימלית לסוג המטבע שהוטל.

ב. מהו האומדן אם האדם קיבל פעמיים עץ?

**12)** מעוניינים לאמוד את אחוז המובטלים באוכלוסייה.
דוגמים 50 אנשים אקראיים ומתקבל ש-4 מהם מובטלים.

א. מצא אומדן נראות מקסימלית לשיעור המובטלים באוכלוסייה.

ב. מצא אומדן לשיעור העובדים באוכלוסייה.

ג. מצא אומדן ליחס בין שיעור העובדים לשיעור המובטלים באוכלוסייה.

**13)** במשחק מחשב שלוש רמות משחק:
ברמה 1 הסיכוי של יוסי לסיים את המשחק הוא 0.9.
ברמה 2 הסיכוי של יוסי לסיים את המשחק הוא 0.7.
ברמה 3 הסיכוי של יוסי לסיים את המשחק הוא 0.4.
יוסי בחר ברמה מסוימת, אך אינו יודע באיזו רמה הוא בחר.
הוא משחק במשחק ברמה שבחר פעמיים.

א. הציעו א.נ.מ. לרמה של המשחק שיוסי שיחק, על סמך מספר הפעמים שסיים את המשחק.

ב. אם יוסי סיים את שני המשחקים, מה יהיה האומדן לרמה?

ג. מהו א.נ.מ. לסיכוי, שמתוך שני משחקים הוא יצליח בדיוק משחק אחד?

**14)** $X_1, X_2, ..., X_n$ מתפלגים אחיד בקטע: $[-\theta,\theta]$.
מצא אומד נראות מקסימלית לפרמטר $\theta$.

**15)** $X_1, X_2, ..., X_n$ מתפלגים בדיד לפי פונקציית ההסתברות הבאה:

$$K=1,2 \qquad P(X=k)=\frac{\binom{2}{k}\cdot P^k\cdot(1-P)^{2-k}}{1-(1-P)^2}.$$

הוכח שא.נ.מ ל- $P$, הינו: $2-\frac{2}{\overline{X}}$.

**16)** במכשיר חשמלי יש 2 סוללות שפועלות באופן ב״ת זו בזו, והוא מפסיק לפעול ברגע שאחת הסוללות מפסיקה לעבוד. הסיכוי של סוללה לתפקד לפחות חודש הוא $P$. כאשר המכשיר מפסיק לפעול מחליפים את שתי הסוללות שלו.
בתחילת הניסוי נלקחו 80 מכשירים כאלה עם סוללות חדשות ולאחר חודש נמצא ש-30 מהם עדיין פועלים.

א. מצא אומדן נראות מקסימלית עבור $P$.

ב. רשמו את האומד שבו השתמשתם בחלק א׳ באופן כללי, עבור מדגם של $n$ מכשירים שמתוכם נמצאו $Y$ מכשירים שעדיין פועלים לאחר חודש אחד.

ג. בהנחה שאורך החיים (בחודשים) של סוללה בודדת הוא מעריכי, עם פי צפיפות: $f(t)=\theta e^{-\theta t}$ עבור $t>0$.
מצא א.נ.מ. עבור $\theta$, המבוסס על $Y$.
מהו האומדן המתאים מן המדגם הנתון?

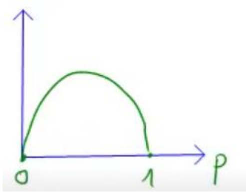
**17)** חיוג אוטומטי של מכשיר טלפון משדר אות אחת לשתי דקות. אם לאחר 20 דקות (10 אותות חיוג) המספר שאליו מטלפנים עדיין תפוס, החיוג האוטומטי נפסק.

א. רשמו את פונקציית ההסתברות של המשתנה $X$ – מספר הפעמים שהחייגן האוטומטי מחייג למספר הטלפון המבוקש, אם ההסתברות לקבלת צליל ״פנוי״ בשידור אחד של אות חיוג הוא $P$.

ב. מתוך 12 ניסיונות חיוג אוטומטי למשרד הרישוי בזמנים שונים במשך 5 ימים, התקבלו התוצאות הבאות: בשני ניסיונות הופסק החיוג האוטומטי ובשאר הניסיונות שבהם הצליח המטלפן להשיג את המספר המבוקש, מספר החיוגים האוטומטיים עד לקבל צליל ״פנוי״ היו: 1, 6, 2, 7, 3, 8, 2, 2, 1, 5.
מצאו אומדן נראות מקסימלית עבור $P$, על סמך התוצאות שהתקבלו.

## תשובות סופיות:

1) א. $L(p)=(1-p)\cdot p$. להלן גרף:




ב. 0.5.  ג. $\frac{2}{9}$.

2) א. $X$.  ב. $\overline{X}$.

3) א. $\frac{1}{\overline{X}}$.  ב. $\frac{2}{9}$.

4) א. 1.  ב. 1.  ג. 3.  ד. $\hat{\theta}=\max\{X_1,\ldots,X_n\}$.

5) א. $\frac{\sum_{i=1}^{n}(x_i-170)^2}{n}$.  ב. 40.2.

6) $\frac{x}{n}$.

7) א. $\frac{n}{\sum X_i^2}$.  ב. $\left(\frac{n}{\sum X_i^2}\right)^2$.

8) א. ראה סרטון.  ב. כד א'.

9) א. 32.  ב. 0.3916.

10) א. 0.45.  ב. 0.81.

11) א.ראה סרטון.  ב. הוגן.

12) א. 0.08.  ב.0.92.  ג. 11.5.

13) א. $\hat{\theta}=\begin{cases}3 & X=0,1\\ 1 & X=2\end{cases}$.  ב. 1.  ג. $\hat{p}=\begin{cases}2\cdot 0.4\cdot 0.6 & X=0,1\\ 2\cdot 0.9\cdot 0.1 & X=2\end{cases}$.

14) $\max|X_i|$.

15) שאלת הוכחה.

16) א. 0.6124  ב. $\hat{p}=\sqrt{\frac{y}{n}}$  ג.0.49.

17) א. $P(x)=\begin{cases}(1-p)^{x-1}p & 1\le x\le 9\\ (1-p)^9 & x=10\end{cases}$.  ב. 0.1818.

## נספח:

### התפלגויות רציפות

| ההתפלגות | פונקציית הצפיפות | פונקציית ההתפלגות המצטברת | תוחלת | שונות | הערות | אנ"מ |
|---|---|---|---|---|---|---|
| $X \sim U(a,b)$ | $f(x)=\frac{1}{b-a}$ <br> $a \le x \le b$ | $\frac{t-a}{b-a}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ | | $b=\max(X_i)$ <br> $a=\min(X_i)$ |
| $X \sim \exp(\lambda)$ | $f(x)=\lambda e^{-\lambda x}$ | $1-e^{-\lambda t}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | הזמן עד להתרחשות מאורע מסוים. $\lambda$ - הוא ממוצע האירועים ביחידת זמן. | $\hat{\lambda}=\frac{1}{\bar{X}}$ |
| $X \sim N(\mu,\sigma^2)$ | $f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $\Phi(t)$ | $\mu$ | $\sigma^2$ | $Z=\frac{X-\mu}{\sigma}$ | $\hat{\mu}=\bar{X}$ <br> $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ |

### התפלגויות בדידות

| ההתפלגות | פונקציית ההסתברות $P(X=k)$ | תוחלת | שונות | הערות | אנ"מ |
|---|---|---|---|---|---|
| בינומית <br> $B(n,p)$ <br> $0 \le p \le 1$ | $\binom{n}{k}p^k(1-p)^{n-k}$ <br> $k=0,1,\ldots,n$ | $np$ | $np(1-p)$ | (1) | $\hat{P}=\frac{Y}{n}$ |
| גיאומטרית <br> $G(p)$ <br> $0 < p \le 1$ | $(1-p)^{k-1}p$ <br> $k=1,2,\ldots,\infty$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ | (2) | $\hat{P}=\frac{1}{\bar{X}}$ |
| אחידה <br> $U(a,b)$ | $\frac{1}{b-a+1}$ <br> $K=a,\ldots,b$ | $\frac{a+b}{2}$ | $\frac{(b-a+1)^2-1}{12}$ | (3) | $b=\max(X_i)$ <br> $a=\min(X_i)$ |
| פואסונית <br> $P(\lambda)$ <br> $\lambda > 0$ | $\frac{e^{-\lambda}\lambda^k}{k!}$ <br> $k=0,1,\ldots,\infty$ | $\lambda$ | $\lambda$ | (4) | $\hat{\lambda}=\bar{X}$ |

(1) מספר ההצלחות ב- $n$ ניסויי ברנולי ב"ת. $p$ - ההסתברות להצלחה.

(2) מספר הניסויים עד להצלחה הראשונה בסדרת ניסויי ברנולי ב"ת, $p$ - ההסתברות להצלחה.

(3) בחירה אקראית של מספר בין $a$ ו- $b$.

(4) מספר אירועים ביחידת זמן, $\lambda$ - קצב האירועים.

# קריטריון MSE – תוחלת ריבוע הטעות:

## רקע:

הקריטריון הנפוץ ביותר כדי לבדוק את טיב האומד הוא קריטריון MSE:
Mean Squared Error – תוחלת ריבוע טעות האמידה.

$$MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = V(\hat{\theta}) + (E(\hat{\theta})-\theta)^2 .$$

כאשר: $V(\hat{\theta})$ - הינה שונות האומד.

$E(\hat{\theta})-\theta$ - הינה ההטיה של האומד.

אם $T_1$ ו- $T_2$ הינם אומדים לפרמטר $\theta$, האומד העדיף יהיה זה עם MSE קטן יותר.
כלומר, אם: $MSE(T_1) > MSE(T_2)$, אז $T_2$ עדיף על $T_1$.

### דוגמה (הפתרון בהקלטה):

נתון משתנה $X$ המתפלג אחיד רציף באופן הבא: $X \sim U(3,\theta)$.

מוצעים שני אומדים לפרמטר $\theta$ על סמך תצפית בודדת: $T_1 = 2X-3$ ו- $T_2 = \frac{3X-3}{2}$.

איזה אומד עדיף לאמידת הפרמטר $\theta$?

## שאלות:

1) מעוניינים לאמוד את התוחלת של התפלגות מסוימת.
מוצעים שני אומדים אפשריים ממוצע של שתי תצפיות וממוצע של שלוש תצפיות.
לפי קריטריון תוחלת ריבוע הטעות (MSE), איזה אומד עדיף? הסבירו.

2) בעיר מסוימת בשוויץ בכל $\theta$ דקות רכבת מגיעה לתחנה מסוימת.
דוד מגיע לתחנה בזמן אקראי ומודד את זמן ההמתנה לרכבת - $X$.
א. הצע אומד חסר הטיה ל-$\theta$, על סמך $X$.
ב. סטטיסטיקאי הציע לאמוד את $\theta$ על סמך האומד: $1.5X$.
האם האומד הנ"ל מוטה?
ג. איזה אומד מבין האומדים בסעיפים א' ו-ב' עדיף?

3) חוקר מעוניין לאמוד את הסיכוי לחלות במחלת השפעת בחורף
(להלן: הפרמטר $P$). הוא דוגם חמישה אנשים בריאים, ומתבונן
בסטטיסטי $X$ - מספר האנשים שחלו בשפעת בחורף.
הוא מתלבט בין שני אומדים: $T_1=\frac{X}{5}$ ו- $T_2=\frac{X+1}{7}$.
א. מי מבין האומדים הללו הוא חסר הטיה?
ב. מי מבין האומדים עדיף אם $P=0.5$?
ג. מי מבין האומדים עדיף אם $P=0.1$?

4) מספר השריפות המתרחשות בארץ בחודש אוקטובר מתפלג פואסונית עם
תוחלת $\lambda$. נלקח מדגם של 10 חודשי אוקטובר. להלן שני אומדים אפשריים:
$$\hat{\lambda}_1=\frac{\sum_{i=1}^{10}X_i}{10} \text{ ו- } \hat{\lambda}_2=\frac{2\cdot\sum_{i=1}^{5}X_i+2\cdot\sum_{i=6}^{10}X_i}{10}.$$
כאשר: $X_i$ = מספר השריפות בחודש אוקטובר ה-$i$.
איזה מהאומדים עדיף, לצורך אמידת הפרמטר $\lambda$?

5) הוכח ש: $E\left(\hat{\theta}-\theta\right)^2=V\left(\hat{\theta}\right)+\left(E\left(\hat{\theta}\right)-\theta\right)^2$.

## תשובות סופיות:

1) שלוש תצפיות.

2) א. $2x$.   ב. אומד מוטה.   ג. סעיף ב.

3) א. $T_1$.   ב. $T_2$.   ג. $T_1$.

4) $\hat{\lambda}_1$.

5) שאלת הוכחה.

# שיטת המומנטים:

### רקע:

מומנט מסדר ראשון של משתנה $X$ מוגדר להיות: $E(X)$.

מומנט מסדר שני של משתנה $X$ מוגדר להיות: $E\left(X^2\right)$.

באופן כלל, מומנט מסדר $r$ מוגדר להיות: $E\left(X^r\right)$.

מומנט מסדר ראשון של $n$ תצפיות בלתי תלויות מאותה התפלגות מוגדר להיות: $\frac{\sum X_i}{n}$ - זהו מומנט מסדר ראשון של המדגם.

מומנט מסדר שני של $n$ תצפיות בלתי תלויות מאותה התפלגות מוגדר להיות: $\frac{\sum X_i^2}{n}$ - זהו המומנט מסדר שני של המדגם.

באופן כללי, מומנט מסדר $r$ של $n$ תצפיות בלתי תלויות מאותה התפלגות מוגדר להיות: $\frac{\sum X_i^r}{n}$ - זהו מומנט ה-$r$ של המדגם.

השיטה: משווים את המומנט המתאים של ההתפלגות לפי המומנט המתאים של המדגם.

### דוגמה (פתרון בהקלטה):

נגיד שמספר הפעמים שאדם מתעטש ביום מתפלג פואסונית על ידי פרמטר $\lambda$ (קצב ההתעטשויות ביום). רוצים לאמוד את $\lambda$ בשיטת המומנטים.

## שאלות:

**1)** $X$ מתפלג אחיד רציף מהערך המינימלי $a$ לערך המכסימלי 20. מצא אומד לערך מינימלי $a$ לפי שיטת המומנטים על סמך $n$ תצפיות מההתפלגות.

**2)** דוגמים $n$ תצפיות בלתי תלויות מתוך התפלגות נורמאלית אשר תוחלתה היא $\mu$ והשונות שלה היא $\sigma^2$. מצא אומדים לפרמטרים אלה לפי שיטת המומנטים.

**3)** אדם מטיל מטבע רגיל $n$ פעמים. יש לאמוד את מספר הפעמים שהוא מטיל את המטבע וזאת על סמך $X$ – מספר העצים שהוא קיבל.

א. מצא אומד בשיטת המומנטים ל-$n$ על סמך $X$ בודד.

ב. מצא אומד בשיטת המומנטים ל-$n$ על סמך חזרה של $m$ פעמים על אותו תהליך בו מטילים את המטבע ההוגן n פעמים.

ג. מהו האומדן אם האדם חזר על התהליך שלוש פעמים:
פעם אחת קיבל 5 עצים, בפעם השנייה הוא קיבל 4 עצים ובפעם השלישית הוא קיבל 7 עצים.

**4)** נתון ש- $X_i \sim \exp(\lambda)$. מצא אומד בשיטת המומנטים לפרמטר $\lambda$ על סמך מדגם של $n$ תצפיות.

**5)** נתונה פונקציית הצפיפות הבאה: $f(x)=\begin{cases}\theta\cdot X^{\theta-1} & 0\le x\le 1\\ 0 & else\end{cases}$.

א. בטא את $E(X)$ כפונקציה של הפרמטר $\theta$.

ב. מצא אומד ל-$\theta$ על פי שיטת המומנטים.

**6)** הזמן בדקות להכנת לחם במאפייה מתפלג באופן הבא: $X_i \sim N(10,\sigma^2)$.
במדגם של הכנת ארבעה לחמים התקבלו התוצאות הבאות: 5, 10, 6, 4.

א. אמוד את $\sigma^2$ בשיטת המומנטים על סמך מדגם בגודל $n$.

ב. מצא את האומדן ל- $\sigma^2$. מה הבעייתיות בתשובה?

## תשובות סופיות:

1) $\hat{a} = 2\overline{X} - 20$ .

2) $\overline{X} = \hat{\mu}$ , $\hat{\sigma}^2 = \frac{\sum X_i^2}{n} - \overline{X}^2$ .

3) א. $\hat{n} = 2X$ . ב. $\hat{n} = 2\overline{X}_m$ . ג. $\hat{n} = 10\frac{2}{3}$ .

4) $\hat{\lambda} = \frac{1}{\overline{X}}$ .

5) א. $\frac{\theta}{\theta+1}$ . ב. $\hat{\theta} = \frac{\overline{X}}{1-\overline{X}}$ .

6) א. $\hat{\sigma}^2 = \frac{\sum X_i^2}{n} - 100$ . ב. -55.75.

# אומד עקיב:

## רקע:

יהי $\hat{\theta}_n$ אומד לפרמטר $\theta$, המתבסס על $n$ תצפיות.

אומד זה יקרא אומד עקיב, אם יתקיים ש: $\hat{\theta}_n \xrightarrow[n\to\infty]{} \theta$.

אם $\hat{\theta}_n$ אומד חסר הטיה לפרמטר $\theta$ ומתקיים ש: $\lim_{n\to\infty} V\left(\hat{\theta}_n\right)=0$,

אזי $\hat{\theta}_n$ אומד עקיב ל- $\theta$.

**דוגמה (פתרון בהקלטה):**

הסבר מדוע $\overline{X}$ אומד עקיב ל- $\mu$.

אם $\hat{\theta}_n$ אומד נראות מקסימלית לפרמטר $\theta$, מתקיים ש: $\hat{\theta}_n \xrightarrow[n\to\infty]{} \theta$.

כלומר, $\hat{\theta}_n$ אומד עקיב ל- $\theta$.

**דוגמה (פתרון בהקלטה):**

הסבר מדוע בהתפלגות גיאומטרית $\frac{1}{\overline{X}}$, אומד עקיב לפרמטר $P$.

## שאלות:

**1)** נתון כי: $X_i \sim U(0,\theta)$, כאשר: $i=1,2...,n$.
$2\bar{X}$ מוצע להיות האומד ל- $\theta$.
א. הראה שאומד זה הוא חסר הטיה.
ב. הסבר מדוע האומד הינו עקיב.

**2)** נתון ש- $X \sim B(n,p)$. כמו כן, נתון ש: $\hat{p}=\frac{X}{n}$ הינו אומד ל- $p$.
הוכח שאומד זה הינו אומד עקיב ל- $p$.

**3)** אורך חיי נורה מתפלג מעריכית עם קצב $\lambda$ לשנה.
נגדיר את: $W_1,W_2,...,W_n$ כסדרת זמנים (בשנים) של $n$ נורות בלתי תלויות.
א. מהו אומד הנראות המקסימלי עבור $\lambda$? האם האומד עקיב?
ב. מצא אומד עקיב לסיכוי שנורה כלשהי תישרף תוך פחות משנתיים.

**4)** נפח החלב בקרטון חלב מתפלג נורמלית עם תוחלת $\mu$ ושונות $\sigma^2$.
מצא אומד עקיב לפרמטר $\sigma^2$, המתבסס על $n$ תצפיות בלתי תלויות.

## תשובות סופיות:

**1)** א. שאלת הוכחה. ב. ראה סרטון.

**2)** שאלת הוכחה.

**3)** א. $\frac{1}{\bar{W}}$. ב. $1-e^{-\frac{2}{\bar{W}}}$.

**4)** $\frac{\sum_i (X_i-\bar{X})^2}{n}$.

# אומד חסר הטיה בעל שונות מינימלית:

אומד חסר הטיה יעיל ביותר – MVUE (Minimum-variance unbiased estimator).

**רקע:**

$T$ יהיה MVUE, אם מתקיים ש- $T$ אומד חסר הטיה ל-$\theta$, ובנוסף מתקיים ש:
$V(T) \le V(\hat{\theta})$, לכל $\hat{\theta}$ חסר הטיה אחר.

**דוגמה (פתרון בהקלטה):**

לרשת חנויות ישנם שני סניפים. מספר הלקוחות הנכנסים לכל סניף ביום מתפלג פואסונית עם קצב של $\lambda$ בסניף A וקצב של $2\lambda$ בסניף B.
נדגמו $n$ ימים מכל סניף, ונבדק בכל יום:
$X_i$ - מספר הלקוחות שנכנסו לסניף A ביום $i$.
$Y_j$ - מספר הלקוחות שנכנסו לסניף B ביום $j$.
על מנת לאמוד את $\lambda$, מוצע האומד: $\alpha\bar{X} + \beta\bar{Y}$.

א. מה התנאי, שצריך להתקיים על $\alpha$ ו- $\beta$, כדי שהאומד יהיה חסר הטיה?

ב. מה צריכים להיות $\alpha$ ו-$\beta$ כדי שהאומד יהיה גם בעל שונות מינימלית?

## שאלות:

**1)** $T_1$ ו- $T_2$ הינם אומדים חסרי הטיה ובלתי תלויים לפרמטר $\theta$.
כמו כן, נגדיר: $T = aT_1 + bT_2$.

א. מה צריך להיות התנאי על $a$ ו- $b$, כדי ש- $T$ יהיה אומד חסר הטיה?

ב. $\sigma_1^2$ ו- $\sigma_2^2$ הם השונויות של $T_1$ ו- $T_2$, בהתאמה.
מצאו $a$ ו- $b$, כך ש- $T$ יהיה אומד חסר הטיה ל- $\theta$, ובעל שונות מינימלית.

**2)** במפעל 3 מכונות המייצרות את אותו חלק.
תוחלת הקוטר של החלקים המיוצרים בכל מכונה זהה.
השונויות של כל מכונה שונות, ומקיימות: $\sigma_2^2 = 2\sigma_1^2$ , $\sigma_3^2 = 3\sigma_1^2$.
הוחלט לדגום $n$ חלקים מכל מכונה, ולחשב את ממוצע הקוטר המתקבל.
$\bar{X}_i$ - יהיה הממוצע המתקבל במכונה $i$.
יהי: $W = \sum_{i=1}^{3} a_i \bar{X}_i$ האומד לתוחלת קוטר החלקים המיוצרים על ידי מכונה כלשהי.

א. מה התנאי שצריך להתקיים על המשקלים $a_i$, כדי שהאומד המוצע יהיה בלתי-מוטה?

ב. נניח ש- $a_1 = a_2$.
מה במקרה זה המשקלים המביאים את האומד להיות MVUE?

## תשובות סופיות:

**1)** א. $a+b=1$. ב. $a = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ , $b = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$.

**2)** א. $\sum_{i=1}^{3} a_i = 1$. ב. $a_1 = a_2 = 0.4$, $a_3 = 0.2$.

# שאלות מסכמות:

## שאלות:

**1)** במפעל מייצרים מוצרים בשלוש מכונות שונות ובלתי תלויות.
במכונה הראשונה הסיכוי שמוצר יהיה תקין הוא $P$, במכונה השנייה
ההסתברות שמוצר יהיה תקין הוא $P^2$ ובמכונה השלישית הסיכוי הוא $2P$.
דוגמים 20 מוצרים מכל מכונה.
נסמן ב- $X$ את מספר המוצרים התקינים שיוצרו במכונה הראשונה,
ב- $Y$ את מספר המוצרים התקינים שיוצרו במכונה השנייה
וב- $Z$ את מספר המוצרים התקינים שיוצרו במכונה השלישית.

א. מהם הערכים האפשריים של הפרמטר $P$ ?
ב. מצאו אומד בלתי מוטה עבור הפרמטר $P$, על סמך $X$ ו- $Z$.
ג. אם התקבל ש- $Y=3$, $X=6$, מהו אומדן נראות מקסימלית ל- $P$ ?

**2)** מספר תאונות הדרכים בקטע כביש א' מתפלג פואסונית עם קצב של $\lambda$
תאונות בחודש, ומספר תאונות הדרכים בקטע כביש ב' מתפלג פואסונית
עם קצב של $2\lambda$ תאונות בחודש.
הוחלט לספור את כמות התאונות בחודש בכל אחד מקטעי הכביש.
נסמן ב- $X$ את מספר התאונות בחודש בקטע א' וב- $Y$ בקטע ב'.

א. מצאו אומד נראות מקסימלית לפרמטר $\lambda$, על סמך $X$ ו- $Y$.
ב. מצאו אומד נראות מקסימלית, לסיכוי שבקטע כביש א תהיה לפחות תאונה אחת בחודש.
ג. האם האומד שמצאת בסעיף א הוא חסר הטיה ל- $\lambda$ ?

**3)** זמן הייצור של מוצר מסוים בתהליך ייצור מתפלג נורמאלית, עם תוחלת ושונות שאינן ידועות.

א. הציעו אומדים חסרי הטיה לתוחלת והשונות של זמן הייצור של המוצר.
ב. הציעו אומדי נראות מקסימלית לתוחלת ולשונות של זמן הייצור של המוצר.
ג. הציעו אומד נראות מקסימלית לריבוע התוחלת של זמן הייצור.
ד. האם האומד מהסעיף הקודם הוא גם חסר הטיה?

4) בקזינו משחק, ובו 4 תאים ממוספרים מ-1 עד 4.
מפעיל המשחק שם כסף באחד מארבעת התאים והאדם המשתתף צריך לנחש באיזה תא הכסף מוחבא. מפעיל הקזינו מודיע שהסיכוי להחביא את הכסף בכל אחד משלושת התאים הראשונים שווה, אך לא בהכרח שווה לסיכוי להחביא אותו בתא הרביעי.
יש לאמוד את הסיכוי להחביא את הכסף בתא הראשון : $P$.
א. מצא את תחום ההגדרה של הפרמטר $P$.

יעל שיחקה את המשחק 3 פעמים וקיבלה שפעם אחת הכסף הוחבא בתא מספר 1 ובפעמים האחרות בתא מספר 2.
ב. מצאו אומדן ל- $P$ על סמך התוצאות הללו בשיטת הנראות המקסימלית.
ג. מצאו אומד חסר הטיה ל- $P$. מהו האומדן לפי התוצאות של יעל?
ד. מצאו אומדן חסר הטיה ונראות מקסימלית לסיכוי שהכסף יוחבא בתא מספר 4 על סמך התוצאות של יעל.

5) יהי : $X_1, X_{2,}...X_n$ מדגם מקרי מתוך ההתפלגות הבאה :

$$f(x)=\begin{cases}\dfrac{\theta}{\lambda}\left(\dfrac{x}{\lambda}\right)^{\theta-1} & 0<x<\lambda,\ \theta>0\\ 0 & else\end{cases}$$

א. מצא אח"ה ל- $\lambda$ (כאשר $\theta$ קבוע ידוע).
ב. מצא אנ"מ ל- $\theta$ (כאשר $\lambda$ קבוע ידוע).
ג. מצא אנ"מ ל- $\lambda$ (כאשר $\theta$ קבוע ידוע).

6) $X$ - משך זמן הפרסומות בערוץ 2 מתפלג אחיד רציף בתחום $(0,\theta)$.
$Y$ - משך זמן הפרסומות בערוץ 10 מתפלג אחיד רציף בתחום $(0,2\theta)$.
א. מצא אומד חסר הטיה ל- $\theta$, המשתמש במשך זמן אקראי של פרסומת בודדת בערוץ 2 ופרסומת בודדת בערוץ 10.
ב. מוצע האומד : $T_2 = X+0.5Y$. האם האומד הנ"ל הוא חסר הטיה?
ג. איזה אומד יותר עדיף זה של סעיף א או זה של סעיף ב'?
ד. מצא אומד נראות מקסימלית ל- $\theta$ על סמך $X$ ו- $Y$.

7) נדגמו 2 תצפיות, $X_1, X_2$, בלתי תלויות מהתפלגויות אחידות רציפות התלויות בפרמטר $\theta$.
ידוע כי : $X_1 \sim U(0,\theta)$ , $X_2 \sim U(0,a\theta)$ (כאשר $a$ קבוע ידוע וחיובי).
א. מצא אנ"מ ל- $\theta$, על סמך 2 התצפיות הנ"ל.
ב. חשב את תוחלת ושונות האנ"מ מסעיף א'. האם האנ"מ מוטה?
ג. מצא אח"ה ל- $\theta$ על סמך סכומן של 2 התצפיות הנ"ל. מהי שונותו?

## תשובות סופיות:

1) א. $0 \le P \le 0.5$. ב. $\hat{p} = \frac{x+z}{60}$. ג. 0.345.

2) א. $\frac{x+y}{3}$. ב. $1-e^{-\frac{x+y}{3}}$. ג. כן.

3) א. מאחר ולא התבקשתם לפתח, הרי שהאומד זהה לנוסחה הכללית (ראו נספח).
ב. כנ"ל. ג. כנ"ל (ראו הפרק על אומד נראות מקסימלי). ד. לא.

4) א. $0 \le P \le \frac{1}{3}$. ב. $\frac{1}{3}$. ג. 0.389. ד. 0.167-.

5) א. אח"ה יהיה: $\hat{\lambda} = \frac{\theta+1}{\theta}\bar{x}$. ב. $\hat{\theta} = \frac{n}{n\ln\lambda - \sum_{i=1}^{n}\ln x_i}$.
ג. $\hat{\lambda} = X_{\max}$.

6) א. $T_1 = (x+y)\frac{2}{3}$. ב. כן. ג. ב'. ד. $\hat{\theta} = \max\left\{x, \frac{1}{2}y\right\}$.

7) א. $\hat{\theta} = \max\left(X_1, \frac{X_2}{a}\right)$. ב. $E(\hat{\theta}) = \frac{2}{3}\theta$ , $V(\hat{\theta}) = \frac{1}{18}\theta^2$.
ג. $\tilde{\theta} = \left(\frac{2}{1+a}\right)(X_1 + X_2)$.

**נספח: אומדי נראות מקסימלית ואומדים חסרי הטיה בהתפלגויות השונות:**

**מודל בינומי**

נתון מדגם של משתנה בינומי: $X \sim B(n,p)$.

א.נ.מ עבור $p$ הוא: $\hat{p}=\frac{X}{n}$, והוא גם א.ח.ה.

**מודל אחיד (בדיד)**

נתון מדגם: $X_1,X_2,\ldots,X_n$ של משתנים אחידים: $X_i \sim U(1,N)$ בלתי-תלויים בזוגות.

א.נ.מ עבור $N$ הוא: $\hat{N}=\max\{X_1,\ldots,X_n\}$, ואינו א.ח.ה.

**מודל פואסוני**

נתון מדגם: $X_1,X_2,\ldots,X_n$ של משתנים פואסוניים: $X_i \sim P(\lambda)$ בלתי-תלויים בזוגות.

א.נ.מ עבור $\lambda$ הוא: $\hat{\lambda}=\bar{X}$ וגם א.ח.ה.

**מודל גיאומטרי**

נתון מדגם: $X_1,X_2,\ldots,X_n$ של משתנים גיאומטריים: $X_i \sim G(p)$ בלתי-תלויים בזוגות.

א.נ.מ עבור $p$ הוא $\hat{p}=\frac{1}{\bar{X}}$, אינו א.ח.ה. א.נ.מ עבור התוחלת $\frac{1}{p}$, הוא $\bar{X}$ והינו א.ח.ה.

**מודל נורמלי**

נתון מדגם: $X_1,X_2,\ldots,X_n$ של משתנים נורמליים: $X_i \sim N(\mu,\sigma^2)$ בלתי-תלויים בזוגות.

א.נ.מ עבור $\mu$, הוא: $\hat{\mu}=\bar{X}$.

כאשר $\mu$ ידוע, א.נ.מ עבור $\sigma^2$ הוא: $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ (אומד חסר-הטיה).

כאשר $\mu$ לא-ידוע, א.נ.מ עבור $\sigma^2$ הוא: $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ (אומד מוטה!!!).

אומד חסר-הטיה עבור $\sigma^2$:

כאשר $\mu$ ידוע: $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$.

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

**מודל מעריכי**

נתון מדגם: $X_1, X_2, \ldots, X_n$ של משתנים מעריכיים: $X_i \sim \exp(\theta)$ בלתי-תלויים בזוגות.

א.נ.מ עבור $\theta$ הוא: $\hat{\theta} = \frac{1}{\bar{X}}$ - מהווה אומד מוטה, וא.נ.מ עבור התוחלת $\frac{1}{\theta}$ הוא $\bar{X}$ א.ח.ה.

**מודל אחיד (רציף)**

נתון מדגם: $X_1, X_2, \ldots, X_n$ של משתנים אחידים: $X_i \sim U(0,\theta)$ בלתי-תלויים בזוגות.

א.נ.מ עבור $\theta$ הוא: $\hat{\theta} = \max\{X_1, \ldots, X_n\}$ אינו א.ח.ה.

**בכל התפלגות:**

א.ח.ה עבור $\mu$ הוא: $\hat{\mu} = \bar{X}$.

אומד חסר-הטיה עבור $\sigma^2$:

כאשר $\mu$ ידוע: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$.

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

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 16 - רווח סמך לתוחלת (ממוצע)

### תוכן העניינים

# רווח סמך כששונות האוכלוסייה ידועה:

## רקע:

ממוצע המדגם הוא אומד לממוצע האוכלוסייה, אך לא באמת ניתן להבין ממנו על גודלו של ממוצע האוכלוסייה. ההסתברות שממוצע המדגם יהיה בדיוק כמו הממוצע האמתי 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: $P(A \le \mu \le B) = 1-\alpha$.

$A$ - 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.

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.

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 ₪.

א. מי האוכלוסייה במחקר?
ב. מה המשתנה הנחקר?
ג. מה הפרמטר שאותו רוצים לאמוד?
ד. מה רווח הסמך לפרמטר?
ה. מהי רמת הסמך לפרמטר?
ו. מה אורך רווח הסמך?
ז. 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 נורמאלית עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.

א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
ג. 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 ₪.

א. בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.
ב. מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?
ג. מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?
ד. 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. $\overline{X}=6$.

ג. $\overline{X}=51$.

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.

**11)** חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא: $63<\mu<83$.
נתון שסטיית התקן בהתפלגות הייתה ידועה לו ושהמדגם התבסס על 40 תצפיות.

א. אם החוקר היה רוצה לבנות רווח סמך באורך 10.
כמה תצפיות עליו היה לדגום?

ב. רווח הסמך שנבנה על ידי החוקר היה ברמת סמך של 95%.
בנו את רווח הסמך שהיה מתקבל ברמת סמך של 98%.

**12)** נתון משתנה מקרי רציף מתפלג אחיד: $X_i \sim U(\mu-0.5, \mu+0.5)$.
נרצה לאמוד את $\mu$. מצאו רווח סמך ל- $\mu$ ברמת-בטחון של 0.95 אם במדגם של 45 תצפיות התקבל: $\bar{x}=74$.

(תזכורת על השונות בהתפלגות אחידה רציפה: $Var(X_i)=\frac{(b-a)^2}{12}$).

## תשובות סופיות:

**1)** א. העובדים במשק. ב. שכר ב-₪. 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. $9284 < \mu > 10{,}116$. ב. הסטיה המירבית בין $\bar{x}$ ל- $\mu$ היא 416 ₪ בביטחון של 95%.
ג. 800. ד. לא.

**5)** א. 102. ב. 3. ג. 0.9544.

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

**7)** א. 87. ב. 5. ג. 0.9544.

**8)** ב'.

**9)** ג'.

**10)** ד'.

**11)** א. 160. ב. $61.13 < \mu < 84.87$.

**12)** $74 \pm 0.084$.

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

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 האמתי 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:

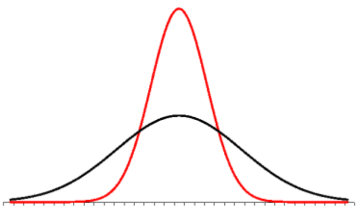
רווח סמך לתוחלת:

שונות האוכלוסייה ידועה

שונות האוכלוסייה אינה ידועה

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

מקרה יותר פרקטי.



**התנאי:** $X \sim N$ או שהמדגם גדול.

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

**האומד לשונות:** $S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} X^2_i - n\bar{X}^2}{n-1}$.

**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):**

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: 5.3, 4.6, 5.2, 4.7.

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: 84, 88, 84, 79, 89.
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: 30, 40, 32, 34, 27.

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 האמתית 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: $\sum_{i=1}^{60} X_i = 195$ , $\sum_{i=1}^{60} X^2{}_i = 643.19$.
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: $S=2$ , $\bar{x}=13.8$.
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 2 $\sigma=20$.
סטטיסטיקאי ב: חישב לפי המדגם וקיבל $S=20$.
למי משני הסטטיסטיקאים יהיה רווח סמך ארוך יותר?
א. סטטיסטיקאי א.
ב. סטטיסטיקאי ב.
ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.
ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

**8)** נתון ש: $X\sim N\left(\mu,\sigma^2\right)$ ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגמית 10.
אורך רווח הסמך שהתקבל הוא: 8.765. מהי רמת הביטחון של רווח הסמך?

## 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'.

**8)** 90%.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 17 - רווח סמך לפרופורציה

### תוכן העניינים

# רווח הסמך לפרופורציה:

**רקע:**

המטרה היא לאמוד את $P$ – פרופורציה באוכלוסייה.

**האומד הנקודתי:**

$\hat{p}=\frac{y}{n}$ ( $Y$ - מספר ההצלחות שבמדגם).

רווח הסמך ל- $p$ : $\hat{p}\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ .

**תנאי לבניית רווח הסמך:**

מדגם של לפחות 30 תצפיות (לעיתים נותנים תנאי של מספר הצלחות ומספר כשלונות לפחות 5 או לפחות 10).

**האומד לטעות התקן:** $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ .

**מתקיים ש:** $\hat{P}=\frac{A+B}{2}$ , $L=2\varepsilon$ .

**דוגמה (פתרון בהקלטה):**

במטרה לאמוד את אחוז המובטלים במשק נדגמו 200 אזרחים, מתוכם התקבל ש-24 היו מובטלים.

א. בנו רווח סמך לאחוז המובטלים באוכלוסייה ברמת סמך של 95%.

ב. מהו האומד לטעות התקן?

## שאלות:

**1)** נדגמו 200 דירות בעיר חיפה. 48 מתוכן נמצאו כבעלות ממ"ד.
א. בנו רווח סמך ברמת סמך של 95% לאחוז הדירות בחיפה עם ממ"ד.
ב. על סמך סעיף א' מה ניתן לומר על שגיאת האמידה המקסימאלית?
ג. בהנחה ובחיפה 80 אלף דירות, בנו רווח סמך ברמת סמך של 95% למספר הדירות בחיפה עם ממ"ד בפועל.

**2)** במדגם של 300 אנשי היי-טק התקבל ש-180 מהם אקדמאים.
א. בנו רווח סמך לפרופורציית אקדמאים ברמת סמך של 95% (בקרב אנשי הייטק).
ב. כיצד רווח הסמך של סעיף א' היה משתנה אם היינו מקטינים את רמת הסמך?
ג. כיצד רווח הסמך היה משתנה אם הינו מגדילים את גודל המדגם?

**3)** במדגם של 400 נהגים התקבל רווח סמך לפרופורציית הנהגים החדשים: $0.08 < p < 0.18$.
א. כמה נהגים במדגם היו נהגים חדשים?
ב. מהי רמת הסמך של רווח הסמך שנבנה?

**4)** במסגרת מערכת הבחירות בארה"ב נשאלו 840 אנשים עבור איזה מועמד יצביעו. 510 אנשים ענו כי יצביעו בעד ברק אובמה. בסקר פורסם שתתכן סטייה של $\pm 3\%$ מתוצאות האמת. באיזו רמת ביטחון הסקר השתמש?

**5)** במדגם של 300 נשים בגילאי 35-40 נמצא ש-140 היו נשואות, 80 היו גרושות, 60 רווקות והיתר אלמנות.
א. מצאו רווח סמך ברמה של 90% לאחוז הגרושות באוכלוסייה הנחקרת.
ב. מצאו רווח סמך ברמה של 99% לסיכוי שבאוכלוסייה הנחקרת תמצא אישה לא נשואה?

**6)** ביצעו מדגם באוכלוסייה. שיעור ההצלחות במדגם היה 10% ורווח הסמך ניבנה ברמת סמך של 95%. אורכו הינו 8.3156%. מהו גודל המדגם שנלקח?

## תשובות סופיות:

1) א. $18.1\% < p < 29.9\%$.
ב. בביטחון של 95% שגיאת האמידה היא לכל היותר 0.059.
ג. $14,480 < \mu < 23,920$.

2) א. $0.545 \le p \le 0.655$. ב. האורך שלו היה קטן. ג. לא ניתן לדעת.

3) א. 52. ב. 0.997.

4) 0.925.

5) א. $30.9\% > p > 22.5\%$. ב. $60.72\% > p > 45.91\%$.

6) 200.

# קביעת גודל מדגם:

## רקע:

בפרק זה נדון איך קובעים גודל מדגם שבאים לאמוד פרופורציה באוכלוסייה מסוימת:

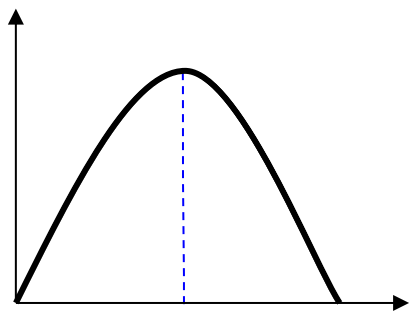
החוקר קובע מראש את רמת הסמך הרצויה: $1-\alpha$.

החוקר קובע מראש את הטעות הסטטיסטית המרבית שבה הוא מעוניין: $\varepsilon$
(או את אורך רווח הסמך).

$L = 2\varepsilon$ - אורך רווח הסמך.

$\varepsilon$ - טעות אמידה מרבית: המרחק המקסימאלי (הסטייה) בין הפרמטר ( $p$ ) לאומד ( $\hat{p}$ ).

$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$




ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

נקבל ש: $n \ge \left( \frac{2 \cdot Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\hat{p}(1-\hat{p})}}{L} \right)^2$.

הבעיה שאין אנו יודעים את $\hat{p}$.

נתבונן בביטוי: $\hat{p}(1-\hat{p})$.

כיוון שאין לנו ידע מוקדם על $\hat{p}$ נציב את המקרה השמרני ביותר שממקסם את הביטוי עבור: $\hat{p} = 0.5$.

$$n \ge \left( \frac{2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{0.5 \cdot 0.5}}{L} \right) \Rightarrow n \ge \left( \frac{z_{1-\frac{\alpha}{2}}}{L} \right).$$

אך אם תהיה לנו אינפורמציה מוקדמת על הפרופורציה נציב את הערך הקרוב ביותר ל-0.5 האפשרי.

### דוגמה (פתרון בהקלטה):

מעוניינים לאמוד את שיעור האבטלה במשק. האמידה צריכה להתבצע ברמת סמך של 90% ועם שגיאת אמידה שלא תעלה על 4%.

א. מהו גודל המדגם המינימאלי שיש לקחת?

ב. חזור לסעיף א' אם ידוע שהאבטלה לא אמורה לעלות על 20%.

## שאלות:

1) הממשלה אומדת מדי חודש את אחוז התמיכה בה.
מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמתי באוכלוסייה ביותר מ-3%, וזאת בביטחון של 95%?

2) משרד התקשורת מעוניין לדעת מה שיעור בתי האב עם אינטרנט.
א. כמה בתי אב יש לדגום אם מעוניינים שבביטחון של 90% אורך רווח הסמך לא יעלה על 8%?
ב. חזרו על סעיף א' אם ידעו שלפני חמש שנים ל-80% מבתי האב היה אינטרנט וכיום יש להניח שיש ליותר אינטרנט.

3) ערוץ טלוויזיה מעוניין לאמוד את הרייטינג של הערוץ בפריים טיים.
המטרה שבביטחון של 95% הסטייה המרבית בין האומד לרייטינג האמתי לא תעלה על 4%.
א. כמה מכשירי PEOPLE METER יש להתקין לצורך האמידה?
ב. לפי הערכה מוקדמת הרייטינג של הערוץ לא יכול לעלות על 20%. בהנחה ומכשיר כזה עולה 500 ₪ ליחידה מה החיסכון הכספי מאינפורמציה זאת?

4) ענו על הסעיפים הבאים:
א. כמה אזרחים יש לדגום כדי לאמוד את אחוז התמיכה בממשלה עם אורך רווח הסמך שלא עולה על 9% ברמת סמך של 90%?
ב. בהנחה ובוצע מדגם שאת גודלו חישבתם בסעיף א והתקבל שאחוז התמיכה בממשלה במדגם הנו 42%. בנו רווח סמך לאחוז התמיכה בממשלה ברמת סמך של 95%.
ג. על סמך סעיף ב', האם תקבלו את הטענה שמיעוט האוכלוסייה תומך הממשלה?

5) משרד הבריאות מתכנן לבצע מדגם שמטרתו לבדוק את הסיכוי לחלות בשפעת עם לקיחת חיסון נגד שפעת. הוא מעוניין שבסיכוי של 98% טעות האמידה לא תעלה על 3%.
א. כמה מחוסנים יש לדגום?
ב. משרד הבריאות ביצע את המדגם שאת גודלו חישבת בסעיף הקודם וקיבל ש-15% מבין אלה שקיבלו חיסון נגד שפעת בכל זאת חלו במשך החורף בשפעת. בנו ברמת סמך של 98% את הסיכוי לחלות בחורף בשפעת עם לקיחת חיסון נגד שפעת.
ג. בהמשך לסעיף הקודם. מהי טעות האמידה המרבית בביטחון של 98%? מדוע הוא קטן מ-3%?

## תשובות סופיות:

1) $1068$.

2) א. $423$. ב. $271$.

3) א. $601$. ב. $108,000$ש"ח.

4) א. $335$. ב. $0.367 < p < 0.473$.

ג. בביטחון של 0.95 ניתן להגיד שמיעוט באוכלוסייה תומך בממשלה.

5) א. $1509$. ב. $0.15 \pm 0.02$ ג. ראה סרטון.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 18 - רווח סמך להפרש פרופורציות

### תוכן העניינים

# רווח סמך להפרש פרופורציות:

**רקע:**

**המטרה:** לאמוד את $p_1 - p_2$ : הפרש פרופורציות בין שתי אוכלוסיות שונות.

**האומד הנקודתי:** $\hat{p}_1 - \hat{p}_2$ .

**התנאי לבניית רווח הסמך:** כל מדגם מעל 30 או לבדוק שמספר ההצלחות ומספר הכישלונות בכל מדגם לפחות 5 בכל מדגם (יש כאלה שבודקים לפחות 10).

**רווח סמך:** $(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ .

רק שאפס נופל בתחומי רווח הסמך להפרש הפרופורציה נאמר שלא ניתן לקבוע שקיים הבדל מובהק בין הפרופורציות באוכלוסיות.

**דוגמה (פתרון בהקלטה):**

במטרה להשוות בין שתי תרופות נדגמו 200 איש שלקחו תרופה $X$ , מתוכם 180 טענו שהתרופה עזרה להם. כמו כן, נלקחו 300 איש שלקחו את תרופה $Y$.
מתוכם 150 טענו שהתרופה עזרה להם. בנו רווח סמך להפרש אחוזי ההצלחה של התרופות ברמת סמך של 95%. מה ניתן לומר על סמך רווח הסמך על ההבדלים בין התרופות?

## שאלות:

1) מתוך 150 נשים שנדגמו באקראי 30% תמכו בהצעת חוק מסוימת.
מתוך 200 גברים שנדגמו באקראי 25% תמכו בהצעת החוק.
   א. בנו רווח סמך לפער בין אחוזי התמיכה של הנשים לעומת הגברים ברמת סמך של 96%.
   ב. בנו רווח סמך ברמת סמך של 95% לאחוז התמיכה בהצעת החוק .

2) במחקר רפואי השתתפו 200 אנשים הסובלים מכאבים כרוניים. הם חולקו באקראי ל-2 קבוצות שוות בגודלן. קבוצה 1 קיבלה את תרופה A וקבוצה שנייה קיבלה את תרופה B. בקרב לוקחי תרופה A 90 טענו שמצבם השתפר. בקרב לוקחי תרופה B 70 טענו שמצבם השתפר.
   א. בנו רווח סמך ברמת סמך של 95% להפרש בין שיעורי ההצלחה של שתי התרופות.
   ב. האם על סמך סעיף א' ניתן לקבוע שקיים הבדל בין התרופות מבחינת שיעורי ההצלחה?

3) נדגמו 200 משפחות מגוש דן. ל-70% מתוכן מכשיר DVD בבית.
נדגמו 300 משפחות מאזור הצפון ל-65% מתוכן מכשיר DVD בבית.
   א. בנו רווח סמך ברמת סמך של 98% לפרופורציות המשפחות בגוש דן עם DVD בבית.
   ב. בנו רווח סמך ברמת סמך של 95% להפרש בין פרופורציות המשפחות בגוש דן עם DVD לבין פרופורציות המשפחות בצפון עם DVD.

## תשובות סופיות:

1) א. $-4.9\% < P_F - P_M < 14.9\%$ . ב. $22.5\% < p < 31.8\%$.

2) א. $0.093 < P_A - P_B < 0.307$. ב. כן.

3) $0.625 < p < 0.7754$.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 19 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

## 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. ${\sigma^2}_1\ ,\ {\sigma^2}_2$ ידועות.
2. $X_1\ ,\ X_2\sim N$ או $n_1,n_2>30$.
3. שני מדגמים בלתי תלויים.

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

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. ${\sigma^2}_1={\sigma^2}_2$.
2. $X_1,X_2\sim N$.
3. מדגמים בלתי תלויים.

**השונות המשוקללת:** כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השונות הזו על ידי שקלול שתי השונויות של שני המדגמים על ידי הנוסחה הבאה: $S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

דרגות החופש: $d.f=n_1+n_2-2$.

רווח סמך: $(\bar{x}_1-\bar{x}_2)\pm t_{1-\frac{\alpha}{2}}^{n_1+n_2-2}\cdot\sqrt{\frac{{S^2}_p}{n_1}+\frac{{S^2}_p}{n_2}}$.

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:

| | תל אביב | באר שבע |
|---|---|---|
| **מספר האקדמאים** | 20 | 10 |
| **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: $N\left(\mu_x,\sigma^2\right)$, ומשתנה $Y$ שמתפלג: $N\left(\mu_y,\sigma^2\right)$.

| | | | | |
|---|---|---|---|---|
| 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:
1. השונות שווה.
2. שהציונים מתפלגים נורמלית.
3. המדגמים אינם תלויים זה בזה.

$-5.52<\mu_1-\mu_2<17.52$.

**2)** $-9.6<\mu_y-\mu_x<1.6$.

## כששונויות האוכלוסייה לא ידועות והמדגמים גדולים:

### 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. מדגמים גדולים.
2. מדגמים בלתי תלויים.

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

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

**דוגמה (פתרון בהקלטה):**

יבואן רכב מעוניין להשוות בין רכבים יפנים לעומת רכבים אמריקאים מבחינת תצרוכת הדלק שלהם. נדגמו 40 רכבים מכל סוג. לכל רכב נבדק כמות הק"מ שהמכונית נסעה עבור 1 ליטר דלק במהירות של 100 קמ"ש. התוצאות שהתקבלו במדגם הם: ברכבים היפנים ממוצע 12.5 עם סטיית תקן 1.5 וברכבים האמריקאים הממוצע 11.3 עם סטיית תקן של 1.4.

א. בנו רווח סמך ברמת סמך של 95% לפער הממוצעים של צריכת הדלק של מכוניות יפניות לעומת אמריקאיות.

ב. האם על סמך רווח הסמך ניתן לקבוע שקים הבדל בין שני סוגי הרכבים מבחינת צריכת הדלק?

## שאלות:

**1)** במדגם של 45 תלמידים הלומדים משפטים התקבל שבשבוע הם לומדים בממוצע 4 שעות עם סטיית תקן 2 שעות מעבר לשעות הלימוד בכיתה. במדגם של 55 תלמידים הלומדים הנדסה התקבל שבשבוע הם לומדים בממוצע 10 שעות עם סטיית תקן של 3 שעות.

א. אמדו את הפרש ממוצעי שעות הלמידה של סטודנטים למשפטים לעומת סטודנטים להנדסה ברמת בטחון של 90%.

ב. האם קיים הבדל בין סטודנטים למשפטים ולסטודנטים להנדסה מבחינת ההשקעה שלהם מעבר לשעות הלימוד בכיתה?

**2)** מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלו נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה. במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות וסטיית תקן של 0.7 שעות. במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות וסטיית תקן של 1.1 שעות.
מצאו רווח סמך לפער בין ממוצע שעות הצפייה בטלוויזיה בין שני האזורים בביטחון של 95%.

## תשובות סופיות:

**1)** א. $-6.83 < \mu_1 - \mu_2 < -5.17$. ב. כן.

**2)** כן, בביטחון של 90% סטודנטים להנדסה לומדים יותר מסטודנטים למשפטים.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 20 - רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים

תוכן העניינים

# 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. $x, y \sim N$.
2. המדגם מזווג.

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

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.

מִצאו 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':

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

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 למדעי המחשב

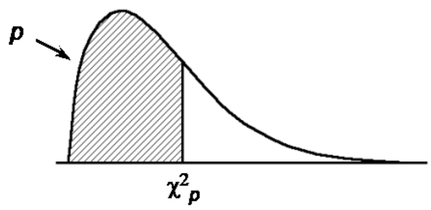
## פרק 21 - רווח סמך לשונות וסטיית תקן

### תוכן העניינים

# רווח סמך לשונות וסטיית תקן:

## רקע:

בפרק זה נדון על בניית רווח סמך לשונות האוכלוסייה.
התנאי לבניית רווח הסמך: המשתנה הנחקר מתפלג נורמלית,
למרות שנהוג לא לדרוש את התנאי הזה אם המדגם מספיק גדול.
רווח הסמך יתבסס על התפלגות הנקראת חי בריבוע. התפלגות זו היא התפלגות אסימטרית חיובית המתחילה מהערך אפס ותלויה בדרגות חופש.
דרגות החופש במקרה זה יהיו: $n-1$.




רווח הסמך לשונות: $\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$.

כאשר: $S^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}=\frac{\sum_{i=1}^{n}X^2_i-n\cdot\bar{X}^2}{n-1}$ אומד לשונות הלא-ידועה.
אם נרצה לבנות רווח סמך לסטיית תקן אז נוציא שורש לרווח סמך לשונות.

**דוגמה:**

זמן התגובה מתפלג נורמאלית. במטרה לאמוד את שונות זמן התגובה נדגמו 4 תצפיות.
להלן התוצאות בשניות: 5.3, 4.6, 5.2, 4.7.
בנו רווח סמך, ברמת סמך של 95%, לשונות זמן התגובה באוכלוסייה.

פתרון:

פרמטר: $\sigma^2$.

$X \sim N(\mu,\sigma^2)$ = זמן תגובה (בשניות).

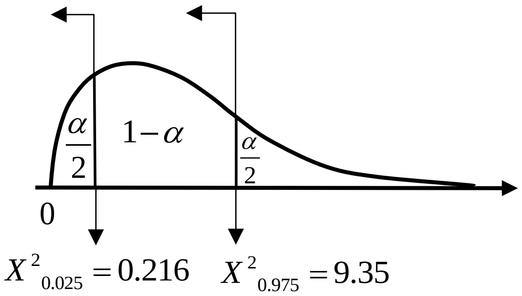
תוצאות מדגם: $n=4$.

$$\bar{X} = \frac{4.7+5.2+4.6+5.3}{4} = 4.95$$

$$d.f = n-1 = 4-1 = 3$$

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

$$S^2 = \frac{4.7^2+5.2^2 ... -4\cdot 4.95^2}{4-1} = 0.123$$




$$1-\alpha = 0.95$$
$$\alpha = 0.05$$
$$\frac{\alpha}{2} = 0.025$$

(טבלת התפלגות חי-בריבוע מופיעה בעמוד האחרון).

נציב: $\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}$

$$\frac{(4-1)\cdot 0.123}{9.35} < \sigma^2 < \frac{(4-1)\cdot 0.123}{0.216}$$

$$0.039 < \sigma^2 < 1.708$$

## שאלות:

**1)** חמישה מטופלים קבלו תרופה מסוימת. בדקו לכל מטופל את זמני התגובה שלו. להלן הזמנים שהתקבלו בדקות: 28, 26, 21, 17, 18.
בהנחה וזמני התגובה מתפלגים נורמאלית, בנו רווח סמך ברמת סמך של 95% לשונות זמן התגובה.

**2)** נדגמו 20 ימים אקראיים מחודשי יולי-אוגוסט ונמדדה בהם הטמפ' במעלות צלזיוס בת"א. במדגם התקבל טמפ' ממוצעת 30.8 וסטיית תקן מדגמית 1.1.
בהנחה והטמפ' מתפלגת נורמאלית:

א. בנו רווח סמך לתוחלת הטמפ' בחודשים אלה בת"א ברמת סמך של 95%.

ב. בנו רווח סמך לסטיית התקן של הטמפ' בחודשים אלה בת"א ברמת סמך של 95%.

**3)** ציוני IQ בארה"ב מתפלגים נורמאלית עם ממוצע 100 וסטיית תקן 5.

נבחנו 20 נבחנים ישראלים במבחן ה-IQ.

להלן התוצאות שהתקבלו: $\sum_{i=1}^{20} X_i^2 = 218,220$ , $\sum_{i=1}^{20} X_i = 2080$.

נניח שגם בישראל הציונים מתפלגים נורמאלית.

א. מצאו אומדנים לממוצע הציונים בישראל ולשונות הציונים בישראל באמצעות אומדנים חסרי הטיה.

ב. אמדו ברמת ביטחון של 95% את תוחלת הציונים של נבחנים בישראל.

ג. אמדו ברמת סמך של 90% את סטיית התקן של הציונים של נבחנים ישראלים.

ד. על סמך הסעיפים הקודמים, האם בישראל ממוצע הציונים וסטיית התקן של הציונים שונה מבארה"ב? הסבירו.

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

$\sum_{i=1}^{10} \left(X_i - \bar{X}\right)^2 = 900$ , $\sum_{i=1}^{10} X_i = 750$.

נתון ש: $X_i \sim N\left(\mu, \sigma^2\right)$.

א. בנו רווח סמך ל- $\mu$ ברמת סמך של 95%.

ב. בנו רווח סמך ל- $\sigma^2$ ברמת סמך של 95%.

## תשובות סופיות:

1) $8.4<\sigma^2<194.2$.

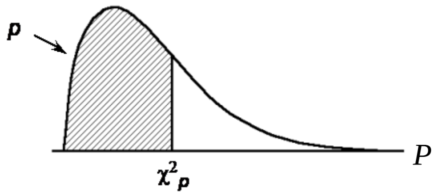
2) א. $30.285<\mu<31.315$. ב. $0.836<\sigma<1.606$.

3) א. ממוצע: 104, שונות: 100. ב. $99.32\le\mu\le108.68$. ג. $7.94<\sigma<13.7$.

ד. בביטחון של 95% ממוצע הציונים איננו שונה, ובביטחון של 90% סטיית התקן שונה.

4) א. $68.75<\mu<82.15$. ב. $47.4<\sigma^2<333.3$.

## נספח - טבלת התפלגות חי-בריבוע – ערכי החלוקה $\chi^2 p$ :




| df | .005 | .01 | .025 | .05 | .10 | .25 | .50 | .75 | .90 | .95 | .975 | .99 | .995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $0.0^4393$ | $0.0^3157$ | $0.0^3982$ | $0.0^2393$ | 0.0158 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 0.211 | 0.575 | 1.39 | 2.77 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 |
| 3 | 0.0717 | 0.115 | 0.216 | 0.352 | 0.584 | 1.21 | 2.37 | 4.11 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.06 | 1.92 | 3.36 | 5.39 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 |
| 5 | 0.412 | 0.554 | 0.831 | 1.15 | 1.61 | 2.67 | 4.35 | 6.63 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 3.45 | 5.35 | 7.84 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 4.25 | 6.35 | 9.04 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 5.07 | 7.34 | 10.2 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 5.90 | 8.34 | 11.4 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 6.74 | 9.34 | 12.5 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 7.58 | 10.3 | 13.7 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 8.44 | 11.3 | 14.8 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 9.30 | 12.3 | 16.0 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 10.2 | 13.3 | 17.1 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 11.0 | 14.3 | 18.2 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 11.9 | 15.3 | 19.4 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 12.8 | 16.3 | 20.5 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 13.7 | 17.3 | 21.6 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 14.6 | 18.3 | 22.7 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 15.5 | 19.3 | 23.8 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 16.3 | 20.3 | 24.9 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 17.2 | 21.3 | 26.0 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 18.1 | 22.3 | 27.1 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 19.0 | 23.3 | 28.2 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 19.9 | 24.3 | 29.3 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 20.8 | 25.3 | 30.4 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 21.7 | 26.3 | 31.5 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 22.7 | 27.3 | 32.6 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 23.6 | 28.3 | 33.7 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 24.5 | 29.3 | 34.8 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 |

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 22 - רווח סמך ליחס שונויות

תוכן העניינים

# רווח סמך ליחס שונויות:

## רקע:

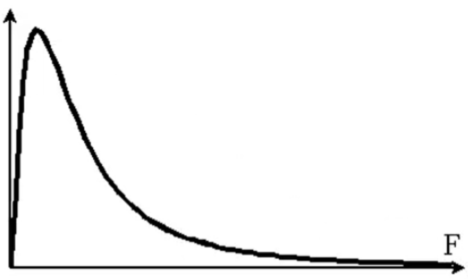
נרצה לאמוד את ההבדל בין שתי שונויות משתי אוכלוסיות שונות.

הפרמטר יהיה: $\frac{\sigma_1^2}{\sigma_2^2}$ : כלומר היחס בין השונויות.

התנאים:

- $X_1, X_2 \sim N$ או מדגמים גדולים.
- מדגמים בלתי תלויים.

רווח הסמך יבנה על סמך התפלגות הנקראת התפלגות F. התפלגות זו היא אסימטרית חיובית ומושפעת משתי דרגות החופש, זו של המונה וזו של המכנה.




$$df_1 = n_1 - 1$$
$$df_2 = n_2 - 1$$

רווח הסמך יהיה: $\frac{s_1^2}{s_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2} F_{1-\frac{\alpha}{2}}(n_2-1, n_1-1)$.

**דוגמה (פתרון בהקלטה):**

מחקר סוציולוגי מעוניין לחקור את הרגלי הבילויים בקבוצות גיל שונות:
במדגם שנעשה על סטודנטים בגילאי 21-26 התקבל אומד חוסר הטיה לשונות ההוצאה החודשית על בילויים 10,000. כמות הסטודנטים שנדגמו 16.
במדגם שנעשה על 11 מבוגרים בשנות השלושים התקבל אומד חסר הטיה לשונות ההוצאה החודשית על בילויים 490,000.
נניח שההוצאה החודשית לבילוי בכל קבוצת גיל מתפלג נורמאלית.
בנו רווח סמך ברמת סמך של 95% ליחס בין השונויות.

## שאלות:

**1)** בתחום הבינוי משתמשים בשני סוגי מתכות: מתכת A ומתכת B. מחקר מעוניין לבדוק האם קיים הבדל בין שני סוגי המתכות מבחינת שונות החוזק שלהן. דגמו מספר יחידות מתכת מכל סוג והתקבלו התוצאות הבאות:

| סוג המתכת | A | B |
|---|---|---|
| $N$ | 8 | 10 |
| $\sum X_i$ | 16 | 30 |
| $\sum X_i^2$ | 60 | 198 |

יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.

א. בנו רווח סמך ליחס השונויות של רמות החוזק בין שני סוגי המתכות ברמת סמך של 90%.

ב. בנו רווח סמך ליחס סטיות התקן של רמות החוזק בין שני סוגי המתכות ברמת סמך של 90%.

**2)** מעוניינים להשוות בין נשים וגברים מבחינת השונות בזמנים שלהם לבצע משימה מסוימת. במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמני ביצוע המשימה: $\sum(y_i-\bar{y})^2=204$. במדגם של 13 נשים התקבלו התוצאות הבאות: $\sum(x_i-\bar{x})^2=200$.

אמוד ברמת ביטחון של 95% פי כמה גדולה השונות של הגברים באוכלוסייה מהשונות של הנשים.

מה יש להניח לצורך פתרון?

## תשובות סופיות:

**1)** א. $0.1013<\frac{\sigma_A^2}{\sigma_B^2}<1.2267$.   ב. $0.318<\frac{\sigma_A}{\sigma_B}<1.108$.

**2)** $0.39\le\frac{\sigma_m^2}{\sigma_F^2}\le 5.26$.

**טבלת התפלגות** $F$. לפי זנב ימני של α

| | | *d.f.1.* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | *d.f.2* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 |
| .05 | 1 | 161 | 200 | 216 | 225 | 230 | 234 | 237 | 239 | 241 | 242 | 244 | 246 |
| .025 | | 648 | 800 | 864 | 900 | 922 | 937 | 948 | 957 | 963 | 969 | 977 | 985 |
| .01 | | 4052 | 5000 | 5403 | 5625 | 5764 | 5859 | 5928 | 5981 | 6022 | 6056 | 6106 | 6157 |
| .005 | | 16211 | 20000 | 21615 | 22500 | 23056 | 23437 | 23715 | 23925 | 24091 | 24224 | 24426 | 24630 |
| .05 | 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 | 19.41 | 19.43 |
| .025 | | 38.51 | 39.00 | 39.17 | 39.25 | 39.30 | 39.33 | 39.36 | 39.37 | 39.39 | 39.40 | 39.41 | 39.43 |
| .01 | | 98.50 | 99.00 | 99.16 | 99.25 | 99.30 | 99.33 | 99.36 | 99.38 | 99.39 | 99.40 | 99.42 | 99.43 |
| .005 | | 198.50 | 199.01 | 199.16 | 199.24 | 199.30 | 199.33 | 199.36 | 199.38 | 199.39 | 199.39 | 199.42 | 199.43 |
| .05 | 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 | 8.74 | 8.70 |
| .025 | | 17.44 | 16.04 | 15.44 | 15.10 | 14.88 | 14.73 | 14.62 | 14.54 | 14.47 | 14.42 | 14.34 | 14.25 |
| .01 | | 34.12 | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.34 | 27.23 | 27.05 | 26.87 |
| .005 | | 55.55 | 49.80 | 47.47 | 46.20 | 45.39 | 44.84 | 44.43 | 44.13 | 43.88 | 43.68 | 43.39 | 43.08 |
| .05 | 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 | 5.91 | 5.86 |
| .025 | | 12.22 | 10.65 | 9.98 | 9.60 | 9.36 | 9.20 | 9.07 | 8.98 | 8.90 | 8.84 | 8.75 | 8.66 |
| .01 | | 21.20 | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.55 | 14.37 | 14.20 |
| .005 | | 31.33 | 26.28 | 24.26 | 23.15 | 22.46 | 21.98 | 21.62 | 21.35 | 21.14 | 20.97 | 20.70 | 20.44 |
| .05 | 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 | 4.68 | 4.62 |
| .025 | | 10.01 | 8.43 | 7.76 | 7.39 | 7.15 | 6.98 | 6.85 | 6.76 | 6.68 | 6.62 | 6.52 | 6.43 |
| .01 | | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 | 10.16 | 10.05 | 9.89 | 9.72 |
| .005 | | 22.78 | 18.31 | 16.53 | 15.56 | 14.94 | 14.51 | 14.20 | 13.96 | 13.77 | 13.62 | 13.38 | 13.15 |
| .05 | 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 | 4.00 | 3.94 |
| .025 | | 8.81 | 7.26 | 6.60 | 6.23 | 5.99 | 5.82 | 5.70 | 5.60 | 5.52 | 5.46 | 5.37 | 5.27 |
| .01 | | 13.75 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 | 7.72 | 7.56 |
| .005 | | 18.63 | 14.54 | 12.92 | 12.03 | 11.46 | 11.07 | 10.79 | 10.57 | 10.39 | 10.25 | 10.03 | 9.81 |
| .05 | 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 | 3.57 | 3.51 |
| .025 | | 8.07 | 6.54 | 5.89 | 5.52 | 5.29 | 5.12 | 4.99 | 4.90 | 4.82 | 4.76 | 4.67 | 4.57 |
| .01 | | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 | 6.62 | 6.47 | 6.31 |
| .005 | | 16.24 | 12.40 | 10.88 | 10.05 | 9.52 | 9.16 | 8.89 | 8.68 | 8.51 | 8.38 | 8.18 | 7.97 |
| .05 | 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 | 3.28 | 3.22 |
| .025 | | 7.57 | 6.06 | 5.42 | 5.05 | 4.82 | 4.65 | 4.53 | 4.43 | 4.36 | 4.30 | 4.20 | 4.10 |
| .01 | | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 | 5.67 | 5.52 |
| .005 | | 14.69 | 11.04 | 9.60 | 8.81 | 8.30 | 7.95 | 7.69 | 7.50 | 7.34 | 7.21 | 7.01 | 6.81 |

| | | | | | | d.f.1. | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | d.f.2. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 |
| .05 | 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 | 3.07 | 3.01 |
| .025 | | 7.21 | 5.71 | 5.08 | 4.72 | 4.48 | 4.32 | 4.20 | 4.10 | 4.03 | 3.96 | 3.87 | 3.77 |
| .01 | | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 | 5.11 | 4.96 |
| .005 | | 13.61 | 10.11 | 8.72 | 7.96 | 7.47 | 7.13 | 6.88 | 6.69 | 6.54 | 6.42 | 6.23 | 6.03 |
| .05 | 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 | 2.91 | 2.85 |
| .025 | | 6.94 | 5.46 | 4.83 | 4.47 | 4.24 | 4.07 | 3.95 | 3.85 | 3.78 | 3.72 | 3.62 | 3.52 |
| .01 | | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 | 4.71 | 4.56 |
| .005 | | 12.83 | 9.43 | 8.08 | 7.34 | 6.87 | 6.54 | 6.30 | 6.12 | 5.97 | 5.85 | 5.66 | 5.47 |
| .05 | 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 | 2.69 | 2.62 |
| .025 | | 6.55 | 5.10 | 4.47 | 4.12 | 3.89 | 3.73 | 3.61 | 3.51 | 3.44 | 3.37 | 3.28 | 3.18 |
| .01 | | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 | 4.16 | 4.01 |
| .005 | | 11.75 | 8.51 | 7.23 | 6.52 | 6.07 | 5.76 | 5.52 | 5.35 | 5.20 | 5.09 | 4.91 | 4.72 |
| .05 | 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 | 2.48 | 2.40 |
| .025 | | 6.20 | 4.77 | 4.15 | 3.80 | 3.58 | 3.41 | 3.29 | 3.20 | 3.12 | 3.06 | 2.96 | 2.86 |
| .01 | | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 | 3.67 | 3.52 |
| .005 | | 10.80 | 7.70 | 6.48 | 5.80 | 5.37 | 5.07 | 4.85 | 4.67 | 4.54 | 4.42 | 4.25 | 4.07 |
| .05 | 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 | 2.28 | 2.20 |
| .025 | | 5.87 | 4.46 | 3.86 | 3.51 | 3.29 | 3.13 | 3.01 | 2.91 | 2.84 | 2.77 | 2.68 | 2.57 |
| .01 | | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 | 3.37 | 3.23 | 3.09 |
| .005 | | 9.94 | 6.99 | 5.82 | 5.17 | 4.76 | 4.47 | 4.26 | 4.09 | 3.96 | 3.85 | 3.68 | 3.50 |
| .05 | 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 | 2.09 | 2.01 |
| .025 | | 5.57 | 4.18 | 3.59 | 3.25 | 3.03 | 2.87 | 2.75 | 2.65 | 2.57 | 2.51 | 2.41 | 2.31 |
| .01 | | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 | 2.84 | 2.70 |
| .005 | | 9.18 | 6.35 | 5.24 | 4.62 | 4.23 | 3.95 | 3.74 | 3.58 | 3.45 | 3.34 | 3.18 | 3.01 |
| .05 | 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 | 1.92 | 1.84 |
| .025 | | 5.29 | 3.93 | 3.34 | 3.01 | 2.79 | 2.63 | 2.51 | 2.41 | 2.33 | 2.27 | 2.17 | 2.06 |
| .01 | | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 | 2.63 | 2.50 | 2.35 |
| .005 | | 8.49 | 5.79 | 4.73 | 4.14 | 3.76 | 3.49 | 3.29 | 3.13 | 3.01 | 2.90 | 2.74 | 2.57 |
| .05 | 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.18 | 2.09 | 2.02 | 1.96 | 1.91 | 1.83 | 1.75 |
| .025 | | 5.15 | 3.80 | 3.23 | 2.89 | 2.67 | 2.52 | 2.39 | 2.30 | 2.22 | 2.16 | 2.05 | 1.94 |
| .01 | | 6.85 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 | 2.47 | 2.34 | 2.19 |
| .005 | | 8.18 | 5.54 | 4.50 | 3.92 | 3.55 | 3.28 | 3.09 | 2.93 | 2.81 | 2.71 | 2.54 | 2.37 |
| .05 | $\infty$ | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 | 1.75 | 1.67 |
| .025 | | 5.02 | 3.69 | 3.12 | 2.79 | 2.57 | 2.41 | 2.29 | 2.19 | 2.11 | 2.05 | 1.94 | 1.83 |
| .01 | | 6.63 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 | 2.32 | 2.18 | 2.04 |
| .005 | | 7.88 | 5.30 | 4.28 | 3.72 | 3.35 | 3.09 | 2.90 | 2.74 | 2.62 | 2.52 | 2.36 | 2.19 |

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 23 - שאלות מסכמות על רווחי סמך

### תוכן העניינים

# שאלות מסכמות על רווחי סמך:

## שאלות:

**1)** מהירות הגלישה באינטרנט במקום מסוים מתפלגת נורמאלית. בדקו את מהירות הגלישה ב-30 זמנים אקראיים. מהירות הגלישה נמדדה ב-Mbps.
מהירות מתחת ל-10Mbps מוגדרת על ידי החברה כנמוכה.
התוצאות שהתקבלו במדגם: ממוצע היה 87 עם סטיית תקן 17 ו-12 פעמים המהירות הייתה נמוכה.
בנו רווחי סמך ברמת סמך של 95% לפרמטרים הבאים:
א. תוחלת מהירות הגלישה.
ב. סטיית תקן של מהירות הגלישה.
ג. הסיכוי שמהירות הגלישה תהיה נמוכה.

**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:

$$\sum_{i=1}^{20} x_i^2 = 210740\ ,\ \sum_{i=1}^{20} x_i = 2040 .$$

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?

5) להלן תוצאות מדגם שבדק עבור כל משפחה האם יש לה בבית מכשיר טאבלט:

| אזור מגורים | גוש דן | שאר הארץ |
|---|---|---|
| **גודל המדגם** | 200 | 240 |
| **מספר משפחות בעלי טאבלט** | 160 | 168 |

א. בנו רווח סמך להבדל בין אחוז המשפחות עם טאבלט בגוש דן ואחוז המשפחות בעלי טאבלט בשאר חלקי הארץ. ברמת סמך של 98%.

ב. בנו רווח סמך לפרופורציות משפחות בעלות טאבלט בכלל הארץ ברמת סמך של 95%.

6) הגובה של מתגייסים לצה"ל מתפלג נורמלית במדגם של 25 מתגייסים. התקבלו התוצאות הבאות: $\bar{x} = 176.2\text{cm}\ ,\ \sum(x_i - \bar{x})^2 = 2832\text{cm}^2 .$

א. אמדו את הגובה הממוצע של המתגייסים ברמת סמך של 98%.

ב. אמדו ברמת סמך של 90% את סטיית התקן של הגובה של מתגייסים של צה"ל.

7) בנק מתלבט האם לפתוח סניף באזור A או באזור B. לצורך פתרון נניח שסטיית התקן של המשכורת באזור A היא 1200 ובאזור B 1500. הבנק דגם 50 אנשים מאזור A, המשכורת הממוצעת שהתקבלה במדגם היא 6,800 ₪. כמו כן, נדגמו 40 אנשים מאזור B, המשכורת הממוצעת שהתקבלה במדגם היא 6,600 ₪.

א. בנו רווח סמך ברמת סמך של 95% להפרש הממוצעים של המשכורות בשני האזורים. האם על סמך רווח הסמך ניתן להמליץ לבנק היכן לפתוח את הסניף. אם כן, היכן?

ב. בנו רווח סמך לתוחלת המשכורת באזור A ברמת סמך של 95%.

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

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

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:

**1)** א. $80.65 \le \mu \le 93.35$. ב. $13.5 < \sigma < 22.9$. ג. $0.225 \le p \le 0.575$.

**2)** א. $1.21 \le \mu \le 1.65$. ב. $10.85\% \le p \le 19.15\%$.

**3)** א. 70. ב. 0.9988. ג. $83\% < p < 89\%$.

**4)** א. $97.4 \le \mu \le 106.6$. ב. לא. ג. יגדל.

**5)** א. $0.5\% \le p_1 - p_2 \le 19.5\%$. ב. $0.704 \le p \le 0.786$.

**6)** א. $170.8 \le \mu \le 181.6$. ב. $8.8 \le \sigma \le 14.3$.

**7)** א. $-372 \le \mu_A - \mu_B \le 772$, לא. ב. $6467 \le \mu_A \le 7133$.

**8)** $-21 \le \mu_D \le 821$.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 24 - בדיקת השערות כללית (סיכוי לטעויות ועוצמת מבחן)

תוכן העניינים

# בדיקת השערות כללית (סיכוי לטעויות ועוצמת מבחן):

**רקע:**

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטיקה.
בתהליך זה ישנן שתי השערות שנבדקות:

1. השערת האפס: המסומנות ב- $H_0$.
2. השערה אלטרנטיבית (השערת המחקר): המסומנת ב- $H_1$.

בדרך כלל השערת האפס מסמנת את אשר היה מקובל עד עכשיו, את השגרה הנורמה ואילו ההשערה האלטרנטיבית את החדשנות בעצם ההשערה האלטרנטיבית מדברת על הסיבה שהמחקר נעשה.

**דוגמה:**

ישנה תרופה קיימת למחלה A אשר גורמת ל-10% מהמשתמשים בה לתופעות לוואי. חברת תרופות טוענת שפיתחה תרופה שיעילה באותה מידה, אך מקטינה את הסיכוי לתופעות הלוואי. לכן יש לבצע מחקר שעל סמך תוצאותיו ננסה להכריע איזה השערה נקבל:

$H_0$: התרופה החדשה הנה קונבנציונאלית וגורמת ל-10% תופעות לוואי.

$H_1$: התרופה החדשה מקטינה את אחוז הסובלים מתופעות לוואי מתחת ל-10%.

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

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

**הגדרת הטעויות:**

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

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

**הגדרת הסתברויות:**

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):
(לדחות $H_0$) $= P_{H_0}$ ($H_0$ נכונה | לדחות את $H_0$) $= \alpha$.

הסיכוי לבצע טעות מסוג 2:
(לקבל $H_0$) $= P_{H_1}$ ($H_1$ נכונה | לקבל את $H_0$) $= \beta$.

רמת בטחון:
(לקבל $H_0$) $= P_{H_0}$ ($H_0$ נכונה | לקבל את $H_0$) $= (\alpha - 1)$.

עוצמה:
(לקבל $H_0$) $= P_{H_1}$ ($H_1$ נכונה | לדחות את $H_0$) $= \pi = (\beta - 1)$.

**דוגמה (פתרון בהקלטה):**

בכד יש 10 כדורים. יתכן ש-5 מהם לבנים והיתר שחורים (כד א' – השערת האפס) או ש-7 מהם לבנים והיתר שחורים (כד ב' – השערה אלטרנטיבית).
כדי להחליט איזה מהכדים ברשותנו, הוחלט להוציא כדור ולהשתמש בכלל ההחלטה הבא: אם הכדור שהוצא הוא לבן שזהו כד ב' $H_1$.

א. חשבו את רמת המובהקות ואת רמת הביטחון של המבחן המוצע.
ב. חשבו את הסיכוי לטעות מסוג שני והעוצמה של המבחן המוצע.

## שאלות:

**1)** אדם חשוד בביצוע פשע. מהן הטעויות האפשריות בהכרעת הדין?

**2)** ילד קנה שקית סוכריות אטומה שבה ציפה ל-10 סוכריות תות ו-5 לימון.
ישנה שקית אחרת אותה הוא לא רצה בה 6 סוכריות תות ו-9 לימון.
הוא החליט להוציא באקראי סוכרייה, אם היא תהיה לימון הוא יחזיר את השקית לחנות. מה הסיכויים לכל סוג של טעות בהכרעתו?

**3)** יהי $X$ מספר שלם הנבחר באקראי מבין המספרים השלמים. הסיכוי ש- $X$
יקבל ערך כלשהו נתון על ידי הנוסחה: $p\left(X=k\right)=\frac{1}{n}$ עבור: $k=1,2,......,n$.
נתונות ההשערות הבאות לגבי התפלגות של $X$: $H_0:n=4$ , $H_1:n=6$.
כמו כן נתון כלל ההכרעה הבא: נדחה את השערת האפס אם: $X>3$.
חשבו את הסיכוי לטעות מסוג ראשון וטעות מסוג שני ואת העוצמה?

**4)** איכות של מוצר מסווגת ל-4 רמות איכות: מצוין, טוב, בינוני וירוד.
להלן התפלגות טיב המוצר בשני מפעלים:

| מפעל / איכות | מצוין | טוב | בינוני | ירוד |
|---|---|---|---|---|
| "היוצר" | 0.6 | 0.2 | 0.2 | 0 |
| "שמשון" | 0.1 | 0.2 | 0.3 | 0.4 |

בוחרים ממשלוח מוצר באקראי, אך לא יודעים מאיזה מפעל המשלוח הגיע.
על סמך בדיקת האיכות מנסים להכריע האם מדובר במפעל "היוצר" (השערת האפס) או במפעל "שמשון" (השערה אלטרנטיבית).

א. להלן כלל החלטה: אם מדובר במוצר שטיבו "טוב" נכריע שהמוצר בא ממפעל "שמשון", מהן ההסתברויות לסוגי הטעויות השונים?
ב. להלן כלל החלטה: אם מדובר במוצר שטיבו "בנוני" או גרוע מכך נכריע שהמוצר בא ממפעל "שמשון", מה מהן ההסתברויות לסוגי הטעויות השונים?
ג. איזה כלל החלטה עדיף? נמקו!

**5)** במטרה לבדוק האם מטבע תקין הטילו אותו 8 פעמים. הוחלט שאם מספר העצים יהיה בין 1 ל-7 כולל יוחלט שהמטבע תקין, אחרת נחליט שהמטבע מזויף.

א. רשמו את השערות המחקר.
ב. מה ההסתברות לטעות מסוג ראשון?
ג. מהי עצמת המבחן אם במציאות אכן המטבע אינו תקין כי הסיכוי לעץ בו הוא 20%.

**6)** להלן השערות:

$H_0: X \sim t(5)$ - התפלגות T עם חמש דרגות חופש.

$H_1: X \sim Z$ - התפלגות נורמלית.

כלל החלטה: נדחה את השערת האפס אם $X$ גדול מ-2.015.

א. מהי רמת המובהקות של כלל ההחלטה?
ב. מהי העוצמה של כלל ההחלטה?

**7)** במפעל מסוים נפלטים לאוויר חומרים רעילים. במצב שיגרה העוצמה הממוצעת של החומר הרעיל אמורה להיות 6,000 יחידות עם סטיית תקן 900. במצב חירום העוצמה הממוצעת היא 7,000 עם סטיית תקן 900. במפעל מערכת התראה נתמכת על ידי 9 חיישנים. אם ממוצע העוצמה של החומר הרעיל לפי תשעת החיישנים עולה על 6,600 יחידות מופעלת מערכת ההתראה. נתון שעוצמת הזיהום מתפלגת נורמאלית.

א. מה הסיכוי להתראת שווא? (באיזה סוג טעות מדובר)?
ב. מה הסיכוי שבמצב חירום מערכת ההתראה לא תפעל? (באיזה סוג טעות מדובר)?
ג. מה ההסתברות שאם המצב הוא מצב חירום מערכת ההתראה תפעל? (איך קוראים להסתברות זו)?
ד. בסעיפים הבאים נשנה בכל סעיף נתון מסוים. כל סעיף עומד בפני עצמו, כיצד השינוי ישנה את הסיכוי לטעות מסוג ראשון ושני?

i. המפעל יקנה עוד 4 חיישנים.
ii. מצב חרום מוגדר כעת בתוחלת של 7,500 יחידות.
iii. מערכת ההתראה תופעל אם ממוצע של תשעת החיישנים יהיה מעל 6,700.

**8)** במטרה לבדוק האם במקום עבודה מסוים פרופורציית הבנים נמוכה מפרופורציית הבנות נדגמו באקראי 10 עובדים. הוחלט שאם מספר הבנים במדגם יהיה לכל היותר 2 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות.

א. מה רמת המובהקות של כלל ההכרעה הנ"ל?
ב. מהי העצמה בהנחה ובחברה 30% בנים?

9) זמן ההשפעה של משכך הכאבים "אופטלנוס" מתפלג נורמאלית עם תוחלת של 40 דקות וסטיית תקן של 12 דקות. חברת התרופות המייצרת את התרופה מנסה לשפר את התרופה כך שתוחלת הזמן עד להשפעה תתקצר. לצורך כך, דגמו 25 מטופלים שקיבלו את התרופה "אופטלנוס פורטה", ממוצע זמן התגובה של המטופלים היה 34.5 דקות. חברת התרופות החליטה מראש שאם ממוצע הזמן עד להשפעה יהיה נמוך מ-35 דקות, היא תמשיך בתהליך שיווק "אופטלנוס פורטה".

א. מהי רמת המובהקות של המבחן המוצע?

ב. על סמך תוצאות המדגם, מהי המסקנה ומהי הטעות האפשרית במסקנה?

ג. מהי עצמת המבחן המוצע אם במציאות התרופה "אופטלנוס פורטה" מפחיתה את התוחלת לכדי 32 דקות?

ד. כיצד תשתנה התשובה לסעיף ג' אם החברה הייתה מחליטה שהיא תמשיך בתהליך שיווק התרופה החדשה כאשר ממוצע המדגם יהיה נמוך מ-36 דקות?

10) ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 120.
מכון טוען שלימודים אצלו מעלים את ממוצע הציונים ביותר מ-30 נקודות. נלקחו 20 שלמדו במכון ו-20 שניגשו לבחינה בלמידה עצמית. הוחלט במשרד פרסום לקבל את טענת המכון רק אם במדגם ממוצע הציונים של אלה שלמדו במכון יהיה גבוהה בלפחות 50 נקודות מאלה שלא היו.

א. מהי רמת המובהקות של המחקר?

ב. מה הסיכוי לעשות טעות מסוג שני II בהנחה שהמכון מעלה את ממוצע הציונים ב-60 נקודות?

ג. כיצד התשובות לסעיף א ו ב' היו משתנות אם מסתבר שסטיית התקן בציוני הפסיכומטרי הינה 100. הסבירו ללא חישוב.

11) קו ייצור נחשב תקין אם יש בו לכל היותר 4% פגומים, ונחשב שאינו תקין אחרת. מנהל האיכות דוגם בכל יום מקו הייצור 500 מוצרים. אם במדגם יהיה לפחות 30 מוצרים פגומים יפסיקו באותו היום את קו הייצור.

א. מה ההסתברות להפסיק את קו הייצור כשהוא תקין. איך קוראים להסתברות זאת?

ב. מה ההסתברות להמשיך ביום מסוים את קו הייצור למרות שאינו תקין כי היו 8% פגומים בקו הייצור. איך קוראים להסתברות זאת?

12) מעוניינים לבדוק האם בפקולטה מסוימת ישנה העדפה לגברים. הוחלט לדגום 200 מתקבלים ועל סמך מספר הבנים לקבוע אם טענת המחקר מתקבלת. חוקר א' קבע רמת מובהקות של 5% וחוקר ב' החליט לקבל את טענת המחקר אם במדגם יהיו לפחות 120 בנים. למי מבין החוקרים רמת מובהקות גדולה יותר?

**13)** מספר המכוניות הנכנסות לחניון "עזרים" מתפלג פואסונית. בשנה שעברה המכוניות נכנסו לחניון בקצב של 2 מכוניות לדקה. בעקבות תלונות על עומס יתר בכניסה לחניון מעוניין מנהל החניון לבדוק האם קצב כניסת המכוניות לחניון גדל השנה. מנהל החניון החליט לספור את מספר המכוניות שיכנסו לחניון בדקה אקראית. אם מספר המכוניות שיספרו יהיה לפחות 4 יפתח מנהל החניון שער נוסף לחניון.

א. רשמו את השערות מנהל החניון ואת כלל ההחלטה שלו. האם כלל ההכרעה הגיוני?

ב. מהי רמת המובהקות של כלל ההכרעה?

ג. מהי העוצמה של כלל ההחלטה, אם כיום קצב כניסת המכוניות לחניון גדל ל-4 מכוניות בדקה?

**14)** עודד עובד במפעל שבו מתחילים לעבוד בשעה 8:00. עודד בדרך כלל מאחר לעבודה והמנהל החליט לרשום את שעת הגעתו. המנהל טוען שמשך האיחור של עודד (בדקות), $X$, הוא משתנה אחיד $U\left(0,60\right)$. עודד טוען שהוא לא מגיע באיחור כה גדול, אלא שהתפלגות $X$ היא בעלת התפלגות מעריכית עם תוחלת איחור של 20 דקות.

לבדיקת טענת המנהל ($H_0$) כנגד טענת עודד ($H_1$), המבוסס על משך האיחור של חגי ביום אחד. מוצאים שני כללי הכרעה:

כלל 1: דחה את השערת האפס אם משך האיחור יהיה לפחות 40 דקות.

כלל 2: דחה את השערת האפס אם משך האיחור יהיה לכל היותר 20 דקות.

חשבו את הסיכוי לטעות מסוג ראשון ושני לכל אחת מכללי ההכרעה. מי עדיף?

## תשובות סופיות:

1) ראה סרטון וידאו.

2) $\alpha=\frac{1}{3}$ , $\beta=\frac{2}{5}$ .

3) $\alpha=0.25$ , $\beta=0.5$ .

4) א. $\alpha=0.2$ , $\beta=0.8$ . ב. $\alpha=0.2$ , $\beta=0.3$ . ג. כלל ב'.

5) א. השערות: $H_0$ -מטבע תקין. $H_1$ -מטבע לא תקין. ב. 0.00781250. ג. 0.1678.

6) א. 0.05. ב. 0.022.

7) א. 0.0228. ב. 0.0918. ג. 0.9082.
ד. i. $\alpha,\beta$ יקטנו. ד. ii. $\alpha$ לא משתנה, $\beta$ קטנה.
ד. iii. $\alpha$ קטנה, $\beta$ גדלה.

8) א. 0.055. ב. 0.383.

9) א. 0.0188. ב. טעות מסוג I . ג. 0.8944.
ד. העוצמה תגדל.

10) א. 0.2981. ב. 0.3974. ג. קטן.

11) א. 0.0113. ב. 0.0495.

12) חוקר א'.

13) א. ראה סרטון וידאו. ב. 0.1428. ג. 0.566.

14) להלן טבלת טעויות, ממנה ניתן להסיק שכלל 2 עדיף.

| כלל | $\alpha$ | $\beta$ |
|---|---|---|
| **1** | $\frac{1}{3}$ | 0.865 |
| **2** | $\frac{1}{3}$ | 0.368 |

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 25 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

# הקדמה:

**רקע:**

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

**שלב א:** נזהה 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 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. אזור הדחיה 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:
$S = 280$ , $\overline{X} = 3120$ , $n = 20$ .

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

### הגדרת הטעויות:

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

טעות מסוג שני: להכריע לקבל את $H_0$ למרות שבמציאות $H_1$ 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. $\begin{matrix} H_0: \mu = 500 \\ H_1: \mu < 500 \end{matrix}$. 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. ד. $\begin{matrix} H_0: \mu = 2.3 \\ H_1: \mu < 2.3 \end{matrix}$.

ה. לא לדחות את $H_0$. ו. טעות מסוג שני.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 26 - הלמה של ניימן פירסון

### תוכן העניינים

# הלמה של ניימן פירסון:

## רקע:

שיטה זו עוזרת לנו לבנות מבחנים בעלי עוצמה מקסימאלית עבור $\alpha$ נתונה.
הלמה של ניימן פירסון אומרת שמבחן בעל עוצמה מקסימלית מתקבל כאשר אזור הדחיה שלו כולל את התוצאות שעבורן יחס הנראות הוא הגבוה ביותר.

נגדיר את יחס הנראות: $x_1, x_2, ..., x_n$ - תוצאות הניסוי.

ההשערות: $H_1 : x_i \sim p_1$ , $H_0 : x_i \sim p_0$.

$$\lambda(x_1, x_2, ..., x_n) = \frac{p_{H1}(x_1, x_2, ..., x_n)}{p_{H0}(x_1, x_2, ..., x_n)}$$

המשמעות של יחס הנראות היא פי כמה $H_1$ יותר סבירה מ- $H_0$.

**עבור משתנים שמתפלגים בדיד:**

**שלב א:** עבור כל תוצאות המדגם האפשריים מחשבים את הסיכויים בהנחת השערת האפס ובהנחת ההשערה האלטרנטיבית.

**שלב ב:** מחלקים את הסיכויים באופן הבא: $\lambda(x_1, x_2, ..., x_n) = \frac{p_{H1}(x_1, x_2, ..., x_n)}{p_{H0}(x_1, x_2, ..., x_n)}$.

ומקבלים את יחס הנראות.

**שלב ג:** מסדרים את תוצאות הניסוי על פי סדר יורד מהתוצאה שמניבה את ערך יחס הנראות הגבוה ביותר עד התוצאה שמניבה את ערך יחס הנראות הנמוך ביותר.

**שלב ד:** מכניסים את התוצאה שמניבה את יחס הנראות הגבוה ביותר לאזור הדחיה ובודקים תחת השערת האפס מהי רמת המובהקות המתקבלת. צוברים את התוצאות לפי העיקרון שהוצג עד שרמת המובהקות לא תעלה על ה- $\alpha$ הרצויה.

**דוגמה (הפתרון בהקלטה):**

איכות של מוצר מסווגת ל-4 רמות איכות: מצוין, טוב, בינוני וירוד.
להלן התפלגות טיב המוצר בשני מפעלים:

| מפעל / איכות | מצוין | טוב | בינוני | ירוד |
|---|---|---|---|---|
| "היוצר" | 0.6 | 0.15 | 0.25 | 0 |
| "שמשון" | 0.1 | 0.2 | 0.3 | 0.4 |

בוחרים ממשלוח מוצר באקראי, אך לא יודעים מאיזה מפעל המשלוח הגיע. על סמך בדיקת האיכות מנסים להכריע האם מדובר במפעל "היוצר" (השערת האפס) או במפעל "שמשון" (השערה אלטרנטיבית). צרו כלל הכרעה לפי הלמה של ניימן פירסון ברמת מובהקות שלא תעלה על 20%.

**עבור משתנים שמתפלגים רציף:**

**שלב א:** בונים את פונקציית הצפיפות המשותפת בהנחת השערת האפס ובהנחת ההשערה האלטרנטיבית.

**שלב ב:** מחלקים את שלב א באופן הבא: $\lambda\left(x_1,x_2,...,x_n\right)=\dfrac{f_{H1}\left(x_1,x_2,...,x_n\right)}{f_{H0}\left(x_1,x_2,...,x_n\right)}$.

ומקבלים את פונקציית יחס הנראות.

**שלב ג:** מזהים את האזור עבורו יחס הנראות הוא הגבוה ביותר.

**שלב ד:** לפי ההתפלגות של השערת האפס מוצאים את הערכים הקריטיים באזור שנקבע בסעיף הקודם כך שרמת המובהקות תהיה ה- $\alpha$ שנקבעה מראש.

**דוגמה (פתרון בהקלטה):**

נתון: $X\sim\exp\left(\lambda\right)$.

ההשערות הן: $H_1:\lambda=2$ , $H_0:\lambda=1$.

מצאו מבחן בעל עוצמה מקסימלית ברמת מובהקות של 5% על סמך תצפית בודדת.

## שאלות:

**1)** איכות של מוצר מסווגת ל-5 רמות איכות: מצוין, טוב, בינוני, ירוד ופסול.
להלן התפלגות טיב המוצר בשני מפעלים:

| מפעל / איכות | מצוין | טוב | בינוני | ירוד | פסול |
|---|---|---|---|---|---|
| "היוצר" | 0.4 | 0.2 | 0.1 | 0.1 | 0.2 |
| "שמשון" | 0.1 | 0.2 | 0.3 | 0.4 | 0 |

בוחרים משלוח מוצר באקראי, אך לא יודעים מאיזה מפעל המשלוח הגיע.
על סמך בדיקת האיכות מנסים להכריע האם מדובר במפעל "היוצר" (השערת האפס) או במפעל "שמשון" (השערה אלטרנטיבית).

א. חשבו את יחס הנראות עבור כל תוצאות המדגם האפשריים.

ב. צרו כלל הכרעה לפי הלמה של ניימן פירסון ברמת מובהקות שלא תעלה על 25%.

ג. מהי עוצמת המבחן שיצרת בסעיף הקודם?

**2)** מטבע הוטל 3 פעמים ומתבוננים במספר הפעמים שהתקבלה התוצאה ראש.
נסמן ב- $p$ את הסיכוי בהטלה בודדת לקבל את התוצאה ראש.
ההשערות הן: $H_0: p=0.5\ ,\ \ H_1: p=0.25$.
מצאו מבחן ברמת מובהקות שלא תעלה על 30% עם עוצמה מקסימלית.
מהי העוצמה?

**3)** בכד א' 7 כדורים לבנים ו-8 שחורים. בכד ב' 8 כדורים לבנים ו-7 שחורים.
אדם בוחר כד וממנו מוציא באקראי 4 כדורים ללא החזרה. הוא מתבוננן במספר הכדורים הלבנים שהוצאו ומודיע לך את המספר המתקבל.
יש לבנות כלל הכרעה על סמך המספר המתקבל שיכריע האם מדובר בהוצאה מכד א' (השערת האפס) או מכד ב' (השערה אלטרנטיבית).

א. בנו כלל הכרעה בעל עוצמה מקסימלית ברמת מובהקות שלא תעלה על 10%.

ב. מהי רמת המובהקות של כלל ההכרעה שבנית בסעיף הקודם?

ג. מה הסיכוי לטעות מסוג שני של כלל ההכרעה שבנית?

**4)** בצרור מפתחות 5 מפתחות שרק אחד פותח את הדלת. על סמך מספר הניסיונות לפתיחת הדלת יש להחליט האם הניסיונות נעשו ללא החזרה (השערת האפס) או עם החזרה (השערה אלטרנטיבית) של המפתחות לצרור.
מצאו מבחן לפי הלמה של ניימן פירסון ברמת מובהקות שלא עולה על 0.25.

**5)** מספר תאונות הדרכים בכביש 4 מתפלג פואסונית עם קצב של תאונה ביממה.
לאחרונה התעורר החשד שתוחלת מספר התאונות בכביש עלתה לקצב של שתי תאונות ביממה. דגמו 4 ימים אקראיים וקיבלו את מספר התאונות הבאות ליממה: 1, 0, 3, 3.

א. נסחו את הבעיה ובנו מבחן MP עם: $\alpha \le 0.1$.

ב. מהי מסקנתך ומהי הטעות האפשרית במסקנה?

ג. מה הסיכוי להכריע שכיום קצב תאונות הדרכים בכביש מספר 4 עלה לשתי תאונות ביממה שאכן כך הדבר באמת?
פתרו על סמך המבחן של סעיף א'.

**6)** התפלגות זמן ההמתנה לקופה בסופרמרקט מתפלג מעריכית.
בעל הסופרמרקט טוען שתוחלת זמן ההמתנה היא 5 דקות אך הלקוחות חושדים שהתוחלת גבוהה יותר ושווה ל-10 דקות.

א. רשמו את השערות המחקר וחשבו את פונקציית יחס הנראות על סמך זמן המתנה של לקוח אקראי לקופה.

ב. מה כיוון אזור הדחייה של השערת האפס?

ג. מצאו את אזור הדחייה עבור רמת מובהקות של 5%.

ד. בהמשך לסעיף הקודם, מה הסיכוי להכריע לטובת בעל הסופרמרקט בטעות?

**7)** יהי $X$ תצפית בודדת מפונקציית הצפיפות הבאה, כאשר $\theta$ פרמטר חיובי:

$$f_\theta(x)=\begin{cases}2\theta X+1-\theta & 0\le x\le 1\\ 0 & otherwise\end{cases} \qquad \frac{n!}{r!(n-r)!}$$

ההשערות הן: $H_1:\theta=1 \ , \ H_0:\theta=0$.

א. הוכיחו שמבחן MP (עוצמה מקסימלית) עם רמת מובהקות $\alpha$ יהיה: $C=\{X>1-\alpha\}$.

ב. הוכיחו שהעוצמה של המבחן שנמצאה היא: $2\alpha-\alpha^2$.

ג. מצאו את כלל ההכרעה והעוצמה עבור: $\alpha=0.05$.

**8)** מחשב חניון "אתרים" רושם את זמן כניסת כל מכונית לחניון. ישנו חשד שעקב תקלה המחשב מבצע את הרישום לכל מכונית שניה. נסמן ב- $X$ את הזמן בדקות בין רישום לרישום.

אם הרישום הוא תקין, ההתפלגות היא מעריכית: $f_0(x)=\lambda\cdot e^{-\lambda x}$.

אם הרישום הוא לפי החשד, פונקציית הצפיפות היא: $f_1(x)=\lambda^2\cdot x\cdot e^{-\lambda x}$.

כאשר מדובר באותו פרמטר $\lambda$.

א. רשמו את ההשערות.

ב. מצאו מבחן בעל עוצמה מקסימלית כדי לבדוק את ההשערות.

ג. פתרו עבור רמת מובהקות של 5% ו- $\lambda=1$.

ד. רשמו את איזו הדחייה עבור שתי תצפיות אקראיות של $X$.

**9)** יהי $X_1,...,X_n$ מדגם מקרי מההתפלגות בעלת פונקציית הצפיפות הבאה, כאשר $\theta$ פרמטר חיובי: $f_\theta(x)=\begin{cases}\theta e^{1-\theta x} & x\ge\frac{1}{\theta}\\ 0 & otherwise\end{cases}$.

א. מצאו מבחן בעל עוצמה מקסימאלית לבדיקת ההשערות: $H_0:\theta=1$ כנגד: $H_1:\theta=2$, על סמך תצפית בודדת (ברמת מובהקות 5%).

ב. מהי עוצמת המבחן שמצאת?

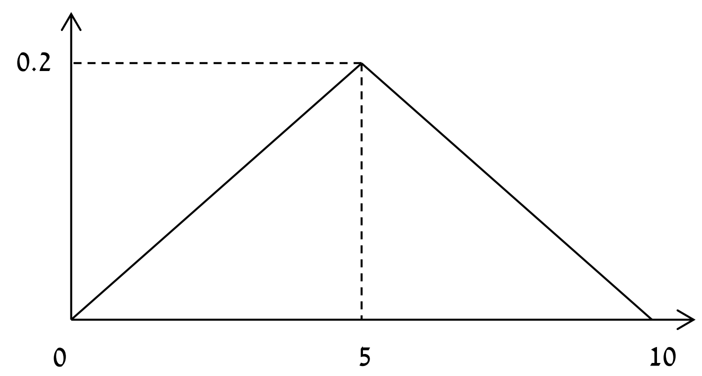
ג. כעת נשנה את הערך תחת האלטרנטיבה ל-3 במקום 2.
בחרו בתשובה הנכונה ונמקו:

i. אפשר לומר ללא חישוב נוסף שהעוצמה תגדל.

ii. אפשר לומר ללא חישוב נוסף שהעוצמה תקטן.

iii. יש לחשב כדי להחליט.

**10)** נתונה פונקציית הצפיפות הבאה שנסמנה ב- $f_1(x)$:



כמו כן נתון ש: $f_0(x)=\begin{cases}0.1 & 0\le x\le 10\\ 0 & else\end{cases}$ .

א. עבור השערות: $H_0: f=f_0$ כנגד: $H_1: f=f_1$, מצאו את צורת אזור הדחיה של מבחן בעל עצמה מקסימלית, על סמך תצפית בודדת.

ב. בהינתן: $\alpha=0.05$, מצאו את כלל הכרעה מתאים בעל עוצמה מקסימלית.

**11)** $X$ הוא משתנה רציף המוגדר בין 0 ל-20.

להלן השערות מחקר: $H_1: f(X)=\dfrac{e^{-\frac{X}{4}}}{4\left(1-e^{-5}\right)}$ , $H_0: X\sim U(0,20)$ .

יש לבנות מבחן בעל עוצמה מקסימלית ברמת מובהקות של 10% על סמך הממוצע 100 תצפיות אקראיות.

## תשובות סופיות:

1) א. להלן טבלה: ב. ראה סרטון. ג. 0.7.

| המפעל | ירוד | בינוני | טוב | מצוין | פסול |
|---|---|---|---|---|---|
| $H_0$ | 0.1 | 0.1 | 0.2 | 0.4 | 0.2 |
| $H_1$ | 0.4 | 0.3 | 0.2 | 0.1 | 0 |
| $\gamma(x)$ | 4 | 3 | 1 | 0.25 | 0 |

2) $C=\{X=0\}$, עוצמה $\frac{27}{64}$.

3) א. $C=\{X=4\}$. ב. 0.0256. ג. 0.9487.

4) $C=\{X=1 \ or \ \ X\geq 6\}$.

5) א. נדחה את השערת האפס אם מספר התאונות הכולל ב-4 הימים יהיה לפחות 8 תאונות. ב. נקבל את השערת האפס (טעות מסוג שני). ג. 0.5471.

6) א. $H_0$: התוחלת של זמן ההמתנה 5 דקות.
$H_1$: התוחלת של זמן המתנה 10 דקות.
פונקציית יחס הנראות: $0.5e^{0.1x}$.
ב. $c=\{x\geq k\}$. ג. $c=\{x\geq 14.98\}$. ד. 0.776.

7) א. שאלת הוכחה. ב. שאלת הוכחה. ג. $c=\{x>0.95\}$, 0.0975.

8) א. השערת האפס: שומר הלילה רושם כל מבקר אשר ניכנס לבניין.
ההשערה האלטרנטיבית: שומר הלילה מדלג ברישום על מבקר, רושם אחד כן ואחד לא.
ב. אזור הדחייה יהיה היכן שיחס הנראות הינו גבוה לכן הוא יהיה מאינסוף ועד ערך $k$ מסוים.
ג. נדחה את השערת האפס אם: $x>2.996$. ד. $x_1\cdot x_2>k$.

9) א. נדחה את $H_0$ עבור: $\frac{1}{2}\leq x\leq 1.0513$. ב. חישוב עוצמת המבחן: 0.668.
ג. i. אפשר לומר ללא חישוב נוסף שהעוצמה תגדל.

10) א. $c=\{|x-5|\leq a\}$. ב. $c=\{4.75\leq x\leq 5.25\}$.

11) כלל ההכרעה הוא נדחה את השערת האפס אם: $\bar{D}\leq 9.26$.

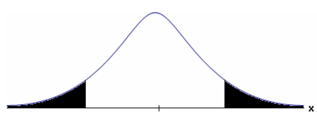
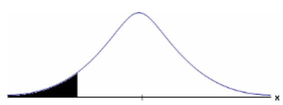
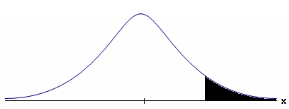
# מבוא לסטטיסטיקה למדעי המחשב

## פרק 27 - בדיקת השערות על תוחלת (ממוצע)

### תוכן העניינים

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

## רקע:

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

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

**חלופה אחרת לכלל הכרעה:**

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

**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 שזובלו 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.

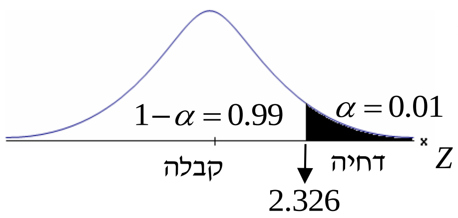
הפרמטר: $\mu$ = תוחלת היבול בשיטת הזיבול החדשה.

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: $\bar{x} = 12.5$ , $n = 4$

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

נציב: $Z_{\bar{x}} = \dfrac{1.25 - 10}{\frac{2.5}{\sqrt{4}}} = 2 < 2.326$

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

**2)** א. נדחה $H_0$, ברי"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 $H_0$, ברי"מ של 5% נקבע שהמכונה לא מאופסת.

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

**5)** נקבל $H_0$, ברי"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:

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

**הגדרת הסתברויות:**

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

$P_{H_0}$ (לדחות $H_0$) $= P($לדחות את $H_0$ | $H_0$ נכונה$) = \alpha$.

הסיכוי לבצע טעות מסוג 2:

$P_{H_1}$ (לקבל $H_0$) $= P($לקבל את $H_0$ | $H_1$ נכונה$) = \beta$.

רמת בטחון:

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

עוצמה:

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

### התהליך לחישוב סיכוי לטעות מסוג שני:

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

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

**התקנון:** $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$.

**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: $\begin{matrix} H_0: \mu = 200 \\ H_1: \mu < 200 \end{matrix}$.

תנאים:

1. $\mu = 200$.
2. $n = 36$.

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

$$\alpha = 0.05$$
$$Z_{1-\alpha} = Z_{0.95} = 1.645$$

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

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

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

ג. השערות: $\begin{matrix} H_0: \mu_0 = 200 \\ H_1: \mu < 200 \end{matrix}$.

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

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

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

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

ד. הקטנת $\alpha$ מגדילה את $\beta$.

## שאלות:

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

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:

$H_0:\mu=200$ , $H_1:\mu\neq200$ , $\sigma=30$ , $n=225$.

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 ומודדים את קוטרם, 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: $H_1:\mu=58$ , $H_0:\mu=50$.
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:

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

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

## תשובות סופיות:

**1)** א. מעל 6.645. ב. 0.3594.
ג. דחינו את $H_0$, תתכן טעות מסוג ראשון.

**2)** א. נדחה $H_0$ אם $\bar{X} < 2.24$. ב. נדחה $H_0$. ג. 1.

**3)** א. נדחה $H_0$ אם $\bar{X} > 203.29$ או $\bar{X} < 196.71$. ב. 0.8051. ג. תקטן.

**4)** א. נדחה $H_0$ אם $\bar{X} < 48.9$. ב. 0.0885. ג. תקטן. ד. תקטן.

**5)** א. 0.0033. ב. i. רמת המובהקות הייתה קטנה.
ב. ii. רמת המובהקות הייתה גדלה.

**6)** ד'.

**7)** ג'.

**8)** ג'.

**9)** א'.

**10)** ב'.

# קביעת גודל מדגם (ששונות האוכלוסייה ידועה):

**רקע:**

השערות המחקר הן: $H_1: \mu = \mu_1$ , $H_0: \mu = \mu_0$.

סטיית התקן של האוכלוסייה ידועה $\sigma$ ומעוניינים לבצע מחקר שרמת המובהקות לא תעלה על $\alpha$ והסיכוי לטעות מסוג שני לא יעלה על $\beta$.

הנוסחה הבאה נותנת את גודל המדגם הרצוי: $n \ge \left(\frac{\left(Z_{1-\alpha}+Z_{1-\beta}\right)\times\sigma}{\mu_0-\mu_1}\right)^2$.

**דוגמה:**

משרד החינוך מפעיל בגן חובה שיטת חינוך שפותחה בשנת 1995. לפי שיטת חינוך זו תוחלת הציון במבחן אוצר מילים לגיל הרך הוא 70. אנשי חינוך החליטו לבדוק שיטת חינוך שפותחה בהולנד הנותנת שם תוחלת ציון אוצר מילים של 80.
נניח שציוני מבחן זה מתפלגים נורמאלית עם $\sigma = 17$.
כדי לבדוק האם גם בישראל הפעלת שיטת החינוך ההולנדית תעבוד בגנים, רוצים לבנות מחקר ברמת מובהקות של 5%. כמו כן, מעוניינים שאם בהפעלת השיטה ההולנדית תוחלת הציונים תעלה לכדי 80, המחקר יגלה זאת בסיכוי של 90%.
כמה ילדי גן חובה דרושים למחקר?

**פתרון:**

<u>האוכלוסייה</u>: ילדי גן חובה.

<u>המשתנה</u>: $X$ = ציון במבחן אוצר מילים.

<u>הפרמטר</u>: $\mu$.

<u>השערות</u>: $\begin{matrix} H_0: \mu = 70 \\ H_1: \mu = 80 \end{matrix}$.

$X \sim N\left(\mu, \sigma^2 = 17^2\right)$.

אם בהפעלת השיטה ההולנדית התוחלת תעלה ל80, נגלה זאת בסיכוי 90%.

$$n \geq \left(\frac{\left(Z_{1-\alpha}+Z_{1-\beta}\right)\times\sigma}{\mu_0-\mu_1}\right)^2$$

$\alpha = 0.05$

$1-\beta = 0.9$

$\mu_0 = 70$

$\mu_1 = 80$

$\sigma = 17$

$Z_{1-\alpha} = Z_{0.95} = 1.645$
$Z_{1-\beta} = Z_{0.9} = 1.282$

**נציב**: $n \geq \left(\frac{(1.645+1.282)\times 17}{70-80}\right)^2 = 24.76$

**לכן**, $n_{\min} = 25$.

## שאלות:

1) במבחן אינטליגנציה הציונים מתפלגים נורמאלית עם סטיית תקן 8 וממוצע 100. פסיכולוג מעוניין לבדוק את הטענה שבאוכלוסיות במצב סוציו אקונומי נמוך תוחלת הציונים היא 95.
אם מעוניינים לגלות את הטענה בהסתברות של לפחות 99% כשרמת המובהקות היא 5% מהו גודל המדגם הדרוש?

2) משרד התקשורת טוענים שאדם מדבר בממוצע 180 דקות בחודש בטלפון הסלולרי. חברות הטלפון הסלולרי טוענות שאינפורמציה זו אינה נכונה ואדם מדבר בממוצע פחות: כ-160 דקות. לצורך פתרון נניח שסטיית התקן של זמן השיחה החודשי ידוע ושווה ל-60 דקות. כמה אנשים יש לדגום כך שאם טענת משרד התקשורת נכונה נדחה אותה בסיכוי של 5% (איך קוראים להסתברות זאת?) כמו כן אם טענת חברות הטלפון הסלולרית נכונה המחקר יגלה זאת בסיכוי של 90% (איך קוראים להסתברות זאת?).

3) השערות המחקר הן: $H_0:\mu=\mu_0$ , $H_1:\mu=\mu_1$.
כמו כן נתון שהמשתנה מתפלג נורמלית עם סטיית התקן ידועה $\sigma$ מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על $\alpha$ והסיכוי לטעות מסוג שני לא יעלה על $\beta$. הוכיחו שגודל המדגם הרצוי לכך יהיה: $n\ge\left(\frac{(Z_{1-\alpha}+Z_{1-\beta})\times\sigma}{\mu_0-\mu_1}\right)^2$.

## תשובות סופיות:

1) 41.
2) 78.
3) שאלת הוכחה.

# 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 $p_v \le \alpha$, דוחים את $H_0$.
מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

$p_v = P_{H_0}$ (לקבל את תוצאות המדגם וקיצוני).
אם ההשערה היא דו צדדית:
$p_v = 2P_{H_0}$ (לקבל את תוצאות המדגם וקיצוני).

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

| **השערת האפס:** **השערה אלטרנטיבה:** | $H_0: \mu = \mu_0$<br>$H_1: \mu \ne \mu_0$ | $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$ | $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$ |
|---|---|---|---|
| **תנאים:** | 1. $\sigma$ ידועה<br>2. $X \sim N$ או מדגם מספיק גדול | | |
| **p-value** | אם $\bar{x} > \mu_0 \Leftarrow 2 \cdot P_{H_0}\left(\bar{X} \ge \bar{x}\right)$<br>אם $\bar{x} < \mu_0 \Leftarrow 2 \cdot P_{H_0}\left(\bar{X} \le \bar{x}\right)$ | $P_{H_0}(\bar{X} \le \bar{x})$ | $P_{H_0}(\bar{X} \ge \bar{x})$ |

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

**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 נורמאלית 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. <u>אוכלוסייה</u>: המתגייסים לצבא כיום.

<u>משתנה</u>: $X$ = משקל בק"ג.

<u>פרמטר</u>: $\mu$.

<u>השערות</u>: $\begin{matrix} H_0 : \mu = 65 \\ H_1 : \mu > 65 \end{matrix}$.

<u>תנאים</u>:

1. $X \sim N$.
2. $\sigma = 12$.

<u>תוצאות מדגם</u>:

$$n = 16$$

$$\bar{X} = 71$$

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

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

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

## 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 נורמאלית 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- $p-value = 0.02$.
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: $p\ value = 0.0254$ 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. $\begin{array}{l} H_0: \mu = 100 \\ H_1: \mu < 100 \end{array}$. 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.

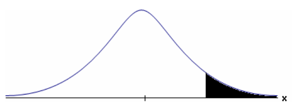
6) א'.

7) א'.

8) ג'.

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

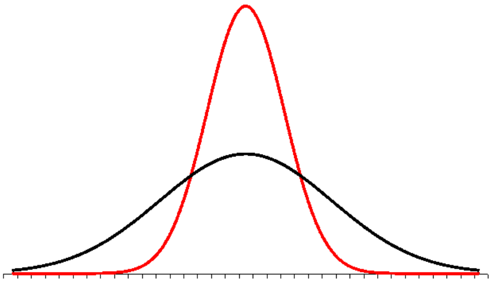
## רקע:

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

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

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

**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: 125, 80, 100, 95, 90 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:

$$\sum_{i=1}^{10} X_i = 750$$
$$\sum_{i=1}^{10} \left(X_i - \bar{X}\right)^2 = 900$$

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.

א. אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
ב. אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.
ד. 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: $\begin{matrix} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{matrix}$.
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'.

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

## רקע:

נזכיר שהמסקנה של המחקר תיקבע לפי העיקרון הבא: אם $p_v \le \alpha$ 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: השערה אלטרנטיבה:** | $H_0: \mu = \mu_0$<br>$H_1: \mu \ne \mu_0$ | $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$ | $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$ |
|---|---|---|---|
| **תנאים:** | 1. $\sigma$ אינה ידועה<br>או<br>2. מדגם מספיק גדול $X \sim N$ | | |
| **p-value** | אם $\bar{x} > \mu_0 \Leftarrow 2 \cdot P_{H_0}\left(\bar{X} \ge \bar{x}\right)$<br>אם $\bar{x} < \mu_0 \Leftarrow 2 \cdot P_{H_0}\left(\bar{X} \le \bar{x}\right)$ | $P_{H_0}\left(\bar{X} \le \bar{x}\right)$ | $P_{H_0}\left(\bar{X} \ge \bar{x}\right)$ |

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

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

$$d.f = n-1$$

**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: 27, 32, 30, 40, 34.
הניחו שזמן הנסיעה מתפלג נורמלית.

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: $\begin{matrix} H_0 : \mu = 40 \\ H_1 : \mu < 40 \end{matrix}$.

ב. תוצאות המדגם:

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

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

$$S = \sqrt{23.4}$$

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

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

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

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

$P_V < \alpha = 0.05$, 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: 997, 996, 1005, 1008, 1024.

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 גבוה מ-174 ס"מ באופן מובהק.

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

## תשובות סופיות:

**1)** א. $H_0: \mu=1000$ , $H_1: \mu\neq1000$ . ב. $20\% \le P_v \le 50\%$ .

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

**2)** 10%.

**3)** 1.01, נקבל את $H_0$ .

# 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 ← נקבל את $H_0$.

אם $\mu_0$ לא נופל ברווח ← 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 פנצילין 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. $112.87 \le \mu \le 118.13$.
ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.

3) א. $191.8 \le \mu \le 200.2$. ב. נכריע שיש אמת בפרסום.

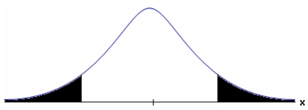
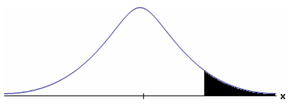
# מבוא לסטטיסטיקה למדעי המחשב

## פרק 28 - בדיקת השערות על פרופורציה

### תוכן העניינים

# התהליך:

## רקע:

| השערת האפס: השערה אלטרנטיבית: | $H_0: p = p_0$ <br> $H_1: p \neq p_0$ | $H_0: p = p_0$ <br> $H_1: p < p_0$ | $H_0: p = p_0$ <br> $H_1: p > p_0$ |
|---|---|---|---|
| **תנאים:** | $np_0 \geq 5 \,\&\, n(1-p_0) \geq 5$ | | |
| **כלל ההכרעה: אזור הדחייה של $H_0$:** | $Z_{\hat{p}} < -Z_{1-\frac{\alpha}{2}}$ או <br> $Z_{\hat{p}} > Z_{1-\frac{\alpha}{2}}$ <br>  <br> $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ <br> -דוחים את $H_0$ | $Z_{\hat{p}} < -Z_{1-\alpha}$ <br>  <br> $-Z_{1-\alpha}$ <br> -דוחים את $H_0$ | $Z_{\hat{p}} > Z_{1-\alpha}$ <br>  <br> $Z_{1-\alpha}$ <br> -דוחים את $H_0$ |

**סטטיסטי המבחן:** $Z_{\hat{p}} = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$.

**חלופה אחרת לכלל הכרעה:**

| כלל ההכרעה – אזור הדחייה של $H_0$: | | |
|---|---|---|
| $\hat{p} > p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{p_0(1-p_0)}{n}}$ <br> $\hat{p} < p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{p_0(1-p_0)}{n}}$ | $\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\dfrac{p_0(1-p_0)}{n}}$ | $\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\dfrac{p_0(1-p_0)}{n}}$ |

**דוגמה (פתרון בהקלטה):**

בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים.
בדקו ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

## שאלות:

1) במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%.
השנה מתוך מדגם של 120 מועמדים התקבלו 22.
ברמת מובהקות של 5% האם השנה הקשו על תנאי הקבלה?

2) במדגם של 300 אזרחים 57% מתנגדים להצעת חוק מסוימת.
לאור נתונים אלה האם רוב האזרחים מתנגדים להצעת החוק?
בדקו ברמת מובהקות של 10%.

3) הטילו מטבע 50 פעמים וקיבלו 28 פעמים עץ.
האם המטבע הוגן ברמת מובהקות של 5%?

4) קפיטריה במכללה מסוימת מעריכה כי אחוז הסטודנטים שקונים קפה בקפיטריה הינו 20%. נערך סקר אשר כלל 200 סטודנטים. התברר כי 33 מהם רוכשים קפה בקפיטריה. מטרת הסקר הייתה לבדוק את אמיתות הערכה של הקפיטריה.
א. רשמו את ההשערות.
ב. בדקו את ההשערות ברמת מובהקות של 10%.
ג. מה תהיה המסקנה אם נקטין את רמת המובהקות?

5) חבר כנסת רוצה להעביר חוק. לצורך כך הוא דוגם 400 אזרחים במטרה לבדוק האם רוב האזרחים תומכים בחוק. במדגם התקבל ש-276 אזרחים תומכים בחוק.
א. מה מסקנתכם ברמת מובהקות של 5%?
ב. האם ניתן לדעת מה תהיה המסקנה אם רמת המובהקות תהיה גדולה יותר? הסבירו.

6) שני חוקרים בדקו את ההשערות הבאות: $H_0: p = p_0$ , $H_1: p > p_0$.
חוקר א' השתמש ברמת מובהקות $\alpha_1$ וחוקר ב' ברמת מובהקות $\alpha_2$ החוקר הראשון דחה את $H_0$ ואילו החוקר השני קיבל את $H_0$.
שניהם התבססו על אותם תוצאות של מדגם. בחר בתשובה הנכונה:
א. $\alpha_1 = \alpha_2$.
ב. $\alpha_1 > \alpha_2$.
ג. $\alpha_1 < \alpha_2$.
ד. המצב המתואר לא אפשרי.

## תשובות סופיות:

1) נדחה $H_0$.

2) נדחה $H_0$.

3) נקבל $H_0$.

4) א. $\begin{matrix} H_0 : p = 0.2 \\ H_1 : p \neq 0.2 \end{matrix}$ .  ב. נקבל $H_0$.  ג. המסקנה לא תשתנה.

5) א. נדחה $H_0$.  ב. המסקנה לא תשתנה.

6) ג'.

# סיכוי לטעויות ועוצמה:

## רקע:

### הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):
$P_{H_0}$(לדחות $H_0$) $=$ $P($לדחות את $H_0$ | $H_0$ נכונה$) = \alpha$.

הסיכוי לבצע טעות מסוג 2:
$P_{H_1}$(לקבל $H_0$) $=$ $P($לקבל את $H_0$ | $H_1$ נכונה$) = \beta$.

רמת בטחון:
$P_{H_0}$(לקבל $H_0$) $=$ $P($לקבל את $H_0$ | $H_0$ נכונה$) = (1-\alpha)$.

עוצמה:
$P_{H_1}$(לדחות $H_0$) $=$ $P($לדחות את $H_0$ | $H_1$ נכונה$) = (1-\beta) = \pi$.

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

### התהליך לחישוב סיכוי לטעות מסוג שני:

| | | | |
|---|---|---|---|
| **השערת האפס: השערה אלטרנטיבית:** | $H_0: p = p_0$<br>$H_1: p \neq p_0$ | $H_0: p = p_0$<br>$H_1: p < p_0$ | $H_0: p = p_0$<br>$H_1: p > p_0$ |
| **תנאים:** | $np_0 \geq 5 \,\&\, n(1-p_0) \geq 5$ | | |
| **כלל ההכרעה: אזור הדחייה של $H_0$:** | $\hat{p} > p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$<br>או<br>$\hat{p} < p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ | $\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ | $\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ |

| חישוב $\beta$ : |
|---|
| $P_{H_1}\left(\hat{p} < p_0 + Z_{1-\alpha}\cdot\sqrt{\frac{p_0(1-p_0)}{n}}\right)$ |
| $P_{H_1}\left(p_0 - Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p} < p_0 + Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{p_0(1-p_0)}{n}}\right)$ |
| $P_{H_1}\left(\hat{p} > p_0 - Z_{1-\alpha}\cdot\sqrt{\frac{p_0(1-p_0)}{n}}\right)$ |

כאשר: $\hat{P} \sim N\left(p\ ,\frac{p(1-p)}{n}\right)$.

והתקנון: $Z_{\hat{p}} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$.

**דוגמה (פתרון בהקלטה):**

רופאי שיניים טוענים שיותר ממחצית האוכלוסייה הבוגרת בארץ אינם מבקרים אצל רופא שיניים באופן קבוע, כנדרש. כדי לבדוק טענה זו, נערך סקר בקרב 150 אנשים בוגרים.

א. רשמו את ההשערות וכלל הכרעה ברמת מובהקות של 10%.

ב. מהי עוצמת המבחן אם מסתבר ש 60% מהאוכלוסייה אינם מבקרים אצל רופא שיניים באופן קבוע.

## שאלות:

1) משרד הבריאות פרסם ש-10% מתושבי המדינה סובלים ממחלת האסטמה. מחקר דורש לבדוק האם בחיפה, בגלל זיהום האוויר, שיעור הסובלים מאסטמה גבוה יותר. לצורך המחקר נבדקו 260 מתושבי חיפה.

א. רשמו את השערות המחקר, וצרו מבחן ברמת מובהקות של 5% לבדיקתן.

ב. מהי עצמת המבחן של סעיף א' בהנחה ובחיפה 16% מהתושבים סובלים מאסטמה?

ג. כיצד תשנה התשובה לסעיף ב' אם מסתבר שבחיפה 18% סובלים מאסטמה?

ד. בהמשך לסעיף א' האם נכון לומר שבהסתברות של 5% ההשערה שבחיפה 10% מהתושבים סובלים מאסטמה אינה נכונה?

2) אחוז הסובלים מתופעות הלוואי מתרופה מסוימת הוא 15%. חברת תרופות טוענת שפיתחה תרופה שאמורה לצמצם את אחוז הסובלים מתופעות לוואי. לצורך בדיקת הטענה הוחלט לבצע מחקר שיכלול 120 חולים שיקבלו את התרופה הנבדקת.
נניח שהתרופה נבדקת אכן מורידה את פרופורציות הסובלים מתופעות הלוואי ל-10%, מהי עצמת המבחן עבור רמת מובהקות של 5%?

3) בעיר מסוימת היו 20% אקדמאים. בעקבות פתיחת מכללה בעיר לפני כמה שנים מעוניינים לבדוק האם אחוז האקדמאים גדל. מעוניינים שהמחקר יכלול 200 אנשים והוא יהיה ברמת מובהקות של 5%.

א. חשבו את הסיכוי לבצע טעות מסוג שני בהנחה והיום יש 28% אקדמאים.

ב. כיצד התשובה לסעיף הקודם תשתנה אם נגדיל את רמת המובהקות?

4) מעוניינים לבדוק האם בפקולטה מסוימת ישנה העדפה לגברים. הוחלט לדגום 200 מתקבלים ועל סמך מספר הבנים לקבוע אם טענת המחקר מתקבלת. חוקר א' קבע רמת מובהקות של 5% וחוקר ב' החליט לקבל את טענת המחקר אם במדגם יהיו לפחות 120 בנים. למי מבין החוקרים רמת מובהקות גדולה יותר?

5) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן (בחרו בתשובה הנכונה):

א. השערת האפס נכונה.

ב. השערת האפס נדחתה.

ג. השערת האפס לא נדחתה.

ד. אף אחת מהתושבות לא נכונה בהכרח.

6) קבעו אם הטענה הבאה נכונה: בבדיקת השערות לא ניתן לבצע בו זמנית טעות מסוג ראשון וטעות מסוג שני.

## תשובות סופיות:

1) א. $\begin{matrix} H_0: p = 0.1 \\ H_1: p > 0.1 \end{matrix}$ . ב. 0.9015. ג. תגדל. ד. טענה לא נכונה.

2) 0.4404.

3) א. 0.1446 ב. תקטן.

4) חוקר א'.

5) ג'.

6) נכונה.

# קביעת גודל מדגם:

## רקע:

השערות המחקר הן: $H_0: p = p_0 \quad , \quad H_1: p = p_1$

מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על $\alpha$ והסיכוי לטעות מסוג שני לא יעלה על $\beta$.

הנוסחה הבאה נותנת את גודל המדגם הרצוי: $n \geq \left( \dfrac{Z_{1-\alpha}\sqrt{p_0 q_0} + Z_{1-\beta}\sqrt{p_1 q_1}}{p_0 - p_1} \right)^2$.

**דוגמה (פתרון בהקלטה):**

רוצים לבדוק האם אחוז האנשים השוהים בשמש ללא הגנה ירד בעקבות הפרסומת על נזקי השמש.
בעבר 60% מהאוכלוסייה שהתה בשמש ללא הגנה. מה גודל המדגם המינימלי שיש לקחת כדי לבדוק שהאחוז הנ"ל ירד ל-48% אם מעוניינים שהסיכוי לטעות מסוג ראשון יהיה 5% והסיכוי לטעות מסוג שני יהיה 1%?

## שאלות:

**1)** משרד התמ"ת פרסם שאחוז האבטלה במשק היום עומד על 8%.
לעומתו, משרד הפנים טוען שחלה עלייה בשיעור האבטלה עד לכדי 11%.
כדי לבדוק מי מביניהם צודק, מה צריך להיות גודל המדגם שיענה על שני התנאים הבאים:

א. אם משרד התמ"ת צודק, נדחה את טענתו בסיכוי של 10%.

ב. אם משרד הפנים צודק, נדחה את טענתו בסיכוי של 4%.

**2)** מפעיל קזינו מפרסם שהסיכוי לזכות במכונת מזל הינו 0.42.
אדם טוען שהסיכויים לזכות במשחק נמוכים יותר.
כמה פעמים יש לשחק את המשחק כדי שאם טענת מפעיל הקזינו נכונה נקבל את טענת האדם בסיכוי של 1% ואם במציאות הסיכוי לזכות במכונה הוא 0.3 נקבל את מפעיל הקזינו בסיכוי של 8%?

## תשובות סופיות:

**1)** 891.

**2)** 224.

# מובהקות התוצאה – אלפא מינימלית:

## 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 $p_v \le \alpha$ 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.

| השערת האפס: השערה אלטרנטיבית: | $H_0: p = p_0$<br>$H_1: p \neq p_0$ | $H_0: p = p_0$<br>$H_1: p < p_0$ | $H_0: p = p_0$<br>$H_1: p > p_0$ |
|---|---|---|---|
| תנאים: | $np_0 \ge 5 \,\&\, n(1-p_0) \ge 5$ | | |
| p-value | אם $\hat{p} > p_0 \Leftarrow 2 \cdot P_{H_0}(\hat{P} \ge \hat{p})$<br>אם $\hat{p} < p_0 \Leftarrow 2 \cdot P_{H_0}(\hat{P} \le \hat{p})$ | $P_{H_0}(\hat{P} \le \hat{p})$ | $P_{H_0}(\hat{P} \ge \hat{p})$ |

כאשר בהנחת השערת האפס: $\hat{P} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$.

התקנון: $Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.

**דוגמה (פתרון בהקלטה):**

ישנה טענה שיש הבדל בין אחוז הבנים ואחוז הבנות הפונים ללמוד להנדסאי מחשבים. לשם כך נלקח מדגם מקרי של 200 תלמידים הלומדים מחשבים והתברר כי 112 מהם בנים.

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

## שאלות:

**1)** במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. רוצים לבדוק האם השנה הקשו על תנאי הקבלה.

א. מהי מובהקות התוצאה?

ב. מה תהיה המסקנה ברמת מובהקות של 1% וברמת מובהקות של 5%?

**2)** נהוג לחשוב ש-60% מהילדים בגיל שלוש קמים מהמיטה במהלך הלילה לפחות פעם אחת. ישנה טענה שללא שנת צהריים פחות מ-60% מהילדים בגיל זה יקומו לפחות פעם אחת במהלך הלילה. נדגמו 80 ילדים בגיל 3 אשר אינם ישנים בצהריים מתוכם התקבל ש-41 קמו במהלך הלילה.

א. מהי רמת המובהקות המינימלית עבורה תתקבל הטענה במחקר?

ב. מהי רמת המובהקות המקסימלית עבורה לא תתקבל טענת המחקר?

ג. עבור אילו רמות מובהקות נקבל את טענת המחקר?

ד. מה תהיה מסקנת המחקר ברמת מובהקות של 6%?

**3)** במטרה לבדוק האם מטבע הוא הוגן מטילים אותו 80 פעמים. התקבל ש-60 מההטלות הראו עץ. רשמו את השערות המחקר, חשבו את מובהקות התוצאה והסיקו מסקנה ברמת מובהקות של 5%.

**4)** בבדיקת השערות על פרופורציה התקבל שה- $p\text{-}value = 0.02$ .
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 5% :
(בחרו בתשובה הנכונה)

א. יקבל את השערת האפס

ב. ידחה את השערת האפס.

ג. לא ניתן לדעת כי אין מספיק נתונים.

**5)** קבעו אם הטענה הבאה נכונה :
"במבחן לבדיקת השערות חד-צדדי התקבל ערך $p-value$ של 3%,
לכן אם היינו מבצעים מבחן דו-צדדי (כאשר יתר הנתונים ללא שינוי),
היינו מקבלים ערך $p-value$ של 6%".

**6)** במפעל 10% מהעובדים נפגעים לפחות פעם אחת בשנה מתאונות עבודה. לאור זאת, המפעל החליט לצאת בתוכנית לצמצום שיעור הנפגעים. תכנית זו נוסתה על 100 עובדים. מתוכם 12 נפגעו בתאונות עבודה במשך השנה.
מהי רמת המובהקות הקטנה ביותר עבורה יוחלט שהתכנית יעילה?

## תשובות סופיות:

1) א. 0.0455.

ב. ברמת מובהקות של 1%: לא דוחים את $H_0$.

ברמת מובהקות של 5%: נדחה את $H_0$.

2) א. 0.0548. ב.0.0548. ג. מעל 0.0548.

ד. נכריע לטובת טענת המחקר.

3) $p_v = 0$, נדחה את $H_0$.

4) ב'.

5) הטענה נכונה.

6) 0.7486.

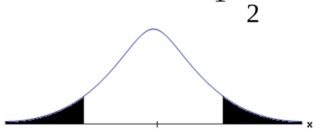
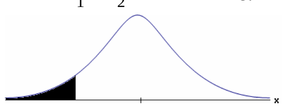
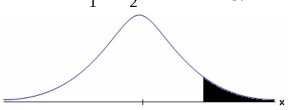
# מבוא לסטטיסטיקה למדעי המחשב

## פרק 29 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

### תוכן העניינים

# בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

## כשהשונויות של האוכלוסייה ידועות – רקע

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

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

**חלופה אחרת לכלל הכרעה:**

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

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

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

**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 מובהקות 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%.

**3)** במדגם אקראי של 20 ימים נבדקה התפוקה של מפעל ביום. התפוקה הממוצעת הייתה של 340 מוצרים ליום. במדגם אקראי של 20 ימים אחרים נבדקה התפוקה של המפעל בלילה והתפוקה הממוצעת הייתה 295. לצורך פתרון נניח שסטיית התקן של התפוקה ביום היא 40 מוצרים ובלילה 30 מוצרים.

א. מהי מובהקות התוצאה לבדיקה האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית.

ב. מה תהיה המסקנה ברמת מובהקות של 8%?

**4)** במחקר מקיף שנעשה באירופה נקבע שגברים גבוהים מנשים ב-8 ס"מ בממוצע. מחקר ישראלי מתעניין לבדוק האם בישראל הפער גדול יותר. לצורך המחקר נדגמו 40 גברים ו 40 נשים באקראי. כמו כן, נניח שסטיות התקן של הגברים והנשים ידועות ושוות ל-6 ס"מ אצל הנשים ו-12 ס"מ אצל הגברים.

א. מהן השערות המחקר ומהו כלל ההכרעה ברמת מובהקות של 10%?

ב. אם בישראל הפער בין גברים לנשים מבחינת הגובה הממוצע הוא 11 ס"מ, מה ההסתברות שהמחקר לא יגלה זאת? איך קוראים להסתברות הזאת?

## תשובות סופיות

**1)** נדחה $H_0$.

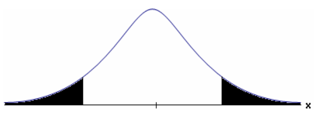
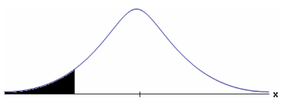
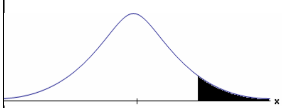
**2)** לא נדחה את $H_0$.

**3)** א. 0 ב. נדחה את $H_0$.

**4)** א. נדחה את $H_0$, אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותר מ-10.72 ס״מ.
ב. 0.6331

# בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

## כשהשונויות של האוכלוסייה לא ידועות והמדגמים גדולים – רקע

| | | | |
|---|---|---|---|
| **השערת האפס: השערה אלטרנטיבית:** | $H_0 \quad \mu_1-\mu_2=c$<br>$H_1 \quad \mu_1-\mu_2\neq c$ | $H_0 \quad \mu_1-\mu_2=c$<br>$H_1 \quad \mu_1-\mu_2<c$ | $H_0 \quad \mu_1-\mu_2=c$<br>$H_1 \quad \mu_1-\mu_2>c$ |
| **תנאים:** | 1. מדגמים בלתי תלויים<br>2. $\sigma_1,\sigma_2$ לא ידועות<br>3. מדגמים מספיק גדולים | | |
| **כלל ההכרעה:** אזור הדחייה של $H_0$: | $Z_{\bar{x}_1-\bar{x}_2}<-Z_{1-\frac{\alpha}{2}}$ או<br>$Z_{\bar{x}_1-\bar{x}_2}>Z_{1-\frac{\alpha}{2}}$<br><br>$-Z_{1-\frac{\alpha}{2}},\quad Z_{1-\frac{\alpha}{2}}$<br>דוחים את $H_0$ | $Z_{\bar{x}_1-\bar{x}_2}<-Z_{1-\alpha}$<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 $H_0$ |

**סטטיסטי המבחן:**

$$Z_{\bar{x}_1-\bar{x}_2}=\frac{\bar{x}_1-\bar{x}_2-c}{\sqrt{\frac{S_1^{\,2}}{n_1}+\frac{S_2^{\,2}}{n_2}}}$$

**חלופה אחרת לכלל הכרעה:**

<table>
<tr><th colspan="2">נדחה $H_0$ אם מתקיים:</th></tr>
<tr><td>$\bar{x}_1-\bar{x}_2>c+Z_{1-\alpha/2}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$<br>או<br>$\bar{x}_1-\bar{x}_2<c-Z_{1-\alpha/2}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$</td><td>$\bar{x}_1-\bar{x}_2>c+Z_{1-\alpha}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$</td></tr>
<tr><td colspan="2">$\bar{x}_1-\bar{x}_2<c-Z_{1-\alpha}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$</td></tr>
</table>

**דוגמה (פתרון בהקלטה) :**

נרצה לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. במדגם של 46 נבחנים חיילים התקבל ממוצע 543 וסטיית תקן 123. במדגם של 50 תלמידי תיכון התקבל ממוצע 508 וסטיית תקן 178. מה המסקנה ברמת מובהקות 5%?

## שאלות

**1)** חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. במדגם של 35 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות וסטיית תקן 4, ובמדגם של 45 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25 וסטיית תקן 6.
האם על סמך תוצאות המדגם יש הבדל בין סוגי הברגים מבחינת רמת הקשיות שלהם? בדקו ברמת מובהקות של 5%.

**2)** כדי לבדוק האם נהגים השותים תחת השפעת אלכוהול נוהגים מהר יותר מאלו שאינם שותים בוצע מדגם שבו בדקו את המהירות המקסימאלית של כל נהג בקמ"ש. להלן התוצאות:

| | גדול מדגם | $\bar{X}$ | $S$ |
|---|---|---|---|
| נהגים השותים אלכוהול | 70 | 80 | 20 |
| נהגים שאינם שותים אלכוהול | 100 | 60 | 15 |

א. מהי מובהקות התוצאה?
ב. מה המסקנה ברמת מובהקת של 5%?

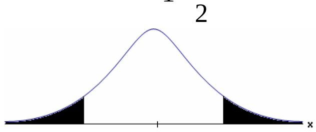
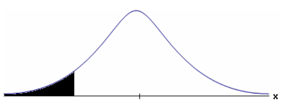
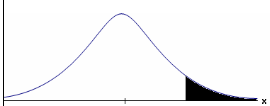
## תשובות סופיות

**1)** נדחה את $H_0$.

**2)** א. 0 ב. נדחה את $H_0$.

# בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

## כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

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

**סטטיסטי המבחן:**

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

**השונות המשוקללת:**

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

**חלופה אחרת לכלל הכרעה:**

<table>
<tr><th colspan="2">נדחה $H_0$ אם מתקיים:</th></tr>
<tr><td>$\bar{x}_1-\bar{x}_2>c+t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$<br>או<br>$\bar{x}_1-\bar{x}_2<c-t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$</td><td>$\bar{x}_1-\bar{x}_2<c-t_{1-\alpha}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$</td></tr>
<tr><td colspan="2">$\bar{x}_1-\bar{x}_2>c+t_{1-\alpha}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$</td></tr>
</table>

**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):

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

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 W 60דולקות בממוצע יותר מאשר נורות מסוג 100?

ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W 60 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 $H_0$.

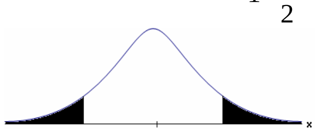
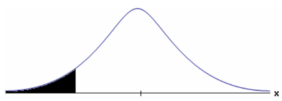
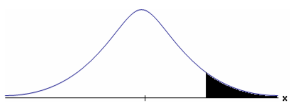
# מבוא לסטטיסטיקה למדעי המחשב

## פרק 30 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

### תוכן העניינים

# בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

## בדיקת השערות למדגמים מזווגים – רקע

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

**סטטיסטי המבחן:**

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

**דוגמה (פתרון בהקלטה):**

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה 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 בממוצע?

**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 בממוצע 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 נורמאלית 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 | 4.5 | 9.6 | 0.7 | 2.5 |
| משקל במכשיר 2 | 3.5 | 6.9 | 1.7 | 0.5 |

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.

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

## תשובות סופיות

**1)** לא נדחה $H_0$.
**2)** לא נדחה $H_0$.
**3)** א. $0.25 \le pv \le 0.5$ ב. 0.5 ג. לא נדחה $H_0$.
**4)** ד'.
**5)** ד'.
**6)** ג'.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 31 - הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות

תוכן העניינים

# הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

## רקע

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

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

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

אם $C$ נופל ברווח ← נקבל את $H_0$ .

אם $C$ לא נופל ברווח ← נדחה את $H_0$ .

**דוגמה (פתרון בהקלטה) :**

חוקר ביצע בדיקת השערות לתוחלת ההפרש במדגם מזווג.

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

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 נקי 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 :

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

## תשובות סופיות

**1)** א. $-3.8 \le \mu_D \le 19$ ב. נכריע שיש אמת בפרסום.

**2)** א. $-8.5 \le \mu_X - \mu_Y \le 26.5$ ב. נכריע שאין הבדל.

**3)** ג׳.

**4)** א׳.

# מבוא לסטטיסטיקה למדעי המחשב

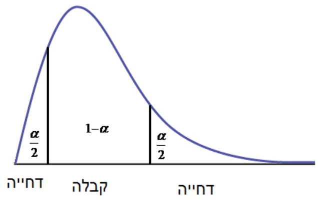
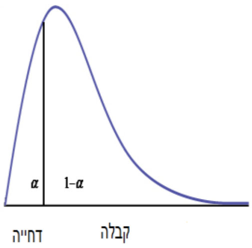
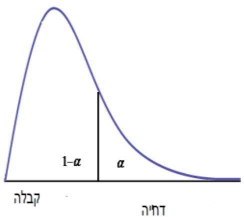
## פרק 32 - בדיקת השערות על שונויות

### תוכן העניינים

# בדיקת השערות על שונות וסטיית תקן:

## רקע:

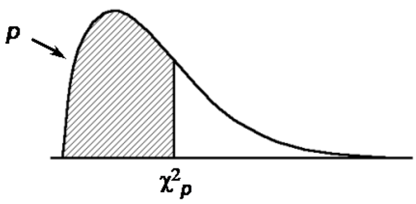
**בדיקת השערות על שונות האוכלוסייה כאשר התוחלת לא ידועה:**

| | | | |
|---|---|---|---|
| **השערת האפס: השערה אלטרנטיבית:** | $H_0:\sigma^2=\sigma_0^2$<br>$H_1:\sigma^2\neq\sigma_0^2$ | $H_0:\sigma^2=\sigma_0^2$<br>$H_1:\sigma^2<\sigma_0^2$ | $H_0:\sigma^2=\sigma_0^2$<br>$H_1:\sigma^2>\sigma_0^2$ |
| **תנאים:** | $X \sim N$ | | |
| **נדחה את השערת האפס אם:** | <br>$\chi^2<\chi^{2(n-1)}_{\frac{\alpha}{2}}$ או<br>$\chi^2>\chi^{2(n-1)}_{1-\frac{\alpha}{2}}$ | <br>$\chi^2<\chi^{2(n-1)}_{\alpha}$ | <br>$\chi^2>\chi^{2(n-1)}_{1-\alpha}$ |

**סטטיסטי המבחן:** $\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}$

**התפלגות חי בריבוע:**

אם $X_i\sim N(\mu,\sigma^2)$ , והפרמטר $\mu$ אינו ידוע, מתקיים ש: $\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^{2(n-1)}$.




התפלגות זו היא התפלגות אסימטרית חיובית המתחילה מהערך אפס וערכיה שואפים לאינסוף.
התפלגות זו תלויה בדרגות החופש.
אם $\mu$ אינו ידוע, אז: $d.f=n-1$.

**דוגמה:**

ציוני IQ לפי סטנדרטים אמריקאים מתפלגים נורמאלית עם $\sigma=15$.
מעוניינים לבדוק האם שונות הציונים של נבחנים ישראלים שונה מאמריקה.

במדגם של 20 ישראלים התקבל: $\sum_{i=1}^{20}(x_i-\bar{x})^2=3420$.

מה המסקנה ברמת מובהקות של 5%?

**פתרון:**

<u>האוכלוסייה</u>: נבחנים ישראלים במבחן I.Q

<u>המשתנה</u>: $X$ = ציון I.Q

<u>פרמטר</u>: $\sigma^2$

<u>השערות</u>:
$H_0:\sigma^2=15^2=225$
$H_1:\sigma^2\neq 225$

<u>הנחה</u>: $X\sim N$

<u>כלל הכרעה</u>: $d.f=n-1=20-1=19$

נדחה את $H_0$ אם $X^2>32.9$ או $X^2<8.91$

<u>תוצאות המדגם</u>: $n=20$

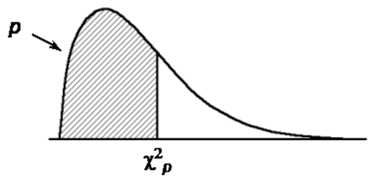
$$\sum(X_i-\bar{X})^2=3420$$

$$S^2=\frac{\sum(X_i-\bar{X})^2}{n-1}=\frac{3420}{20-1}=180$$

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}=\frac{19\cdot 180}{225}=15.2$$

<u>מסקנה</u>: לא נדחה את $H_0$, לא נסיק ששונות הציונים של נבחנים ישראלים במבחן I.Q שונה מזו של אמריקאים. ($\alpha=5\%$)

# טבלת התפלגות חי-בריבוע – ערכי החלוקה $\chi^2 p$




| df | .005 | .01 | .025 | .05 | .10 | .25 | .50 | .75 | .90 | .95 | .975 | .99 | .995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $0.0^4393$ | $0.0^3157$ | $0.0^3982$ | $0.0^2393$ | 0.0158 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 0.211 | 0.575 | 1.39 | 2.77 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 |
| 3 | 0.0717 | 0.115 | 0.216 | 0.352 | 0.584 | 1.21 | 2.37 | 4.11 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.06 | 1.92 | 3.36 | 5.39 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 |
| 5 | 0.412 | 0.554 | 0.831 | 1.15 | 1.61 | 2.67 | 4.35 | 6.63 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 3.45 | 5.35 | 7.84 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 4.25 | 6.35 | 9.04 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 5.07 | 7.34 | 10.2 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 5.90 | 8.34 | 11.4 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 6.74 | 9.34 | 12.5 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 7.58 | 10.3 | 13.7 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 8.44 | 11.3 | 14.8 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 9.30 | 12.3 | 16.0 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 10.2 | 13.3 | 17.1 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 11.0 | 14.3 | 18.2 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 11.9 | 15.3 | 19.4 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 12.8 | 16.3 | 20.5 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 13.7 | 17.3 | 21.6 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 14.6 | 18.3 | 22.7 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 15.5 | 19.3 | 23.8 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 16.3 | 20.3 | 24.9 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 17.2 | 21.3 | 26.0 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 18.1 | 22.3 | 27.1 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 19.0 | 23.3 | 28.2 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 19.9 | 24.3 | 29.3 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 20.8 | 25.3 | 30.4 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 21.7 | 26.3 | 31.5 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 22.7 | 27.3 | 32.6 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 23.6 | 28.3 | 33.7 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 24.5 | 29.3 | 34.8 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 |

## שאלות:

1) חברה אורזת סוכר במשקל עם סטיית תקן 20 גרם. משקל הסוכר באריזה מתפלג נורמאלית. החברה החליפה את מכונות האריזה במטרה לדייק יותר במשקל הנארז. (רוצים שסטיית התקן תהיה קטנה יותר).
לצורך בדיקה דגמו 5 אריזות סוכר ולהלן משקלן (בגרם):
997, 996, 1005, 1024, 1008.
מה המסקנה ברמת מובהקות של 5%?

2) זמן ההחלמה ממחלה מסוימת כאשר משתמשים בטיפול מסוים מתפלג נורמלית עם סטיית תקן של 80 שעות. תרופה חדשה נוסתה על 5 חולים. זמני ההחלמה שלהם בשעות היו: 38, 50, 72, 90, 110.
א. ברמת מובהקות של 5% בדקו האם סטיית התקן של זמן החלמה של התרופה החדשה נמוכה מהתרופה המקורית?
ב. האם ניתן לדעת מה תהיה התשובה לסעיף א',אם נגדיל את רמת המובהקות?
ג. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נקטין את רמת המובהקות?
ד. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נוסיף תצפית שערכה 70?

3) הגובה של אוכלוסייה מסוימת נחשב כמתפלג נורמלית עם ממוצע של 174 ס"מ וסטיית תקן 12. במדגם של 20 אנשים מהאוכלוסייה התקבל ממוצע 171 וסטיית תקן מדגמית 23.
א. בדקו ברמת מובהקות של 5% האם חל שינוי בשונות הגבהים באוכלוסייה.
ב. בדקו ברמת מובהקות של 5%, האם חל שינוי בתוחלת הגבהים באוכלוסייה, בבחירת המבחן המתאים הסתמך על המסקנה מסעיף א'.

4) השערות המחקר הן: $H_o:\sigma^2=100 \quad H_1:\sigma^2>100$
מתכננים לבצע מדגם בגודל 10 תצפיות. רמת המובהקות היא 5%.
א. מה תהיה עוצמת המבחן אם $\sigma_1^2=150$?
ב. איזו השערה אלטרנטיבית תיתן עוצמה של 90%?

5) השערות המחקר הן: $H_o:\sigma=2$ , $H_1:\sigma<2$.
במדגם של 21 תצפיות התקבל סטיית תקן 1.143.
תנו הערכה למובהקות התוצאה.

## תשובות סופיות:

1) לא נדחה $H_o$.

2) א. נדחה את $H_0$. ב. לא תשתנה. ג. לא ניתן לדעת.
ד. לא תשתנה.

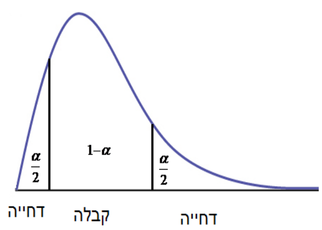
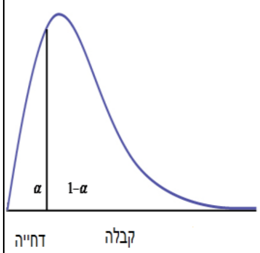
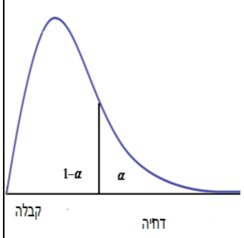
3) א. נדחה את $H_0$. ב. לא נדחה את $H_0$.

4) א. בין 25% ל-50%. ב. 405.3.

5) $0 < P_V < 0.005$.

# בדיקת השערות על שתי שונויות:

## רקע:

| | | | |
|---|---|---|---|
| **השערת האפס:** **השערה אלטרנטיבית:** | $H_0:\frac{\sigma_1^2}{\sigma_2^2}=1$ <br> $H_1:\frac{\sigma_1^2}{\sigma_2^2}\neq 1$ | $H_0:\frac{\sigma_1^2}{\sigma_2^2}=1$ <br> $H_1:\frac{\sigma_1^2}{\sigma_2^2}<1$ | $H_0:\frac{\sigma_1^2}{\sigma_2^2}=1$ <br> $H_1:\frac{\sigma_1^2}{\sigma_2^2}>1$ |
| **תנאים:** | 1.מדגמים בלתי תלויים <br> 2. $X_1, X_2 \sim N$ | | |
| **נדחה את השערת האפס אם:** |   <br> או $F \geq f_{1-\alpha/2}^{(n_1-1,n_2-1)}$ <br> $F \leq \frac{1}{f_{1-\alpha/2}^{(n_2-1,n_1-1)}}$ |   <br> $F \leq \frac{1}{f_{1-\alpha}^{(n_2-1,n_1-1)}}$ |   <br> $F \geq f_{1-\alpha}^{(n_1-1,n_2-1)}$ |

**סטטיסטי המבחן:** $F=\frac{S_1^{\ 2}}{S_2^{\ 2}}$

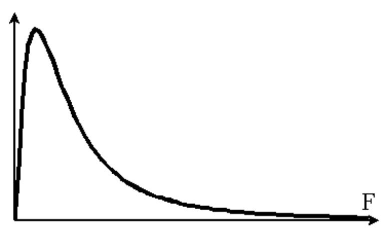
### התפלגות F:

אם $X_1 \sim N(\mu_1,\sigma^2)$ ו- $X_2 \sim N(\mu_2,\sigma^2)$ אזי: $\frac{S_1^2}{S_2^2} \sim F(n_1-1,n_2-1)$.

התפלגות F הינה התפלגות אסימטרית חיובית התלויה בדרגות חופש של המונה ושל המכנה.

כמו כן בהתפלגות F מתקיימת התכונה הבאה: $F_\alpha(n_1-1,n_2-1)=\frac{1}{F_{1-\alpha}^{(n_2-1,n_1-1)}}$.

$df_1=n_1-1$
$df_2=n_2-1$



**דוגמה:**

מעוניינים להשוות בין נשים וגברים מבחינת השונות בזמנים שלהם לבצע משימה מסוימת. במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמני ביצוע המשימה: $\sum(y_i-\bar{y})^2=204$ .

במדגם של 13 נשים התקבלו התוצאות הבאות: $\sum(x_i-\bar{x})^2=200$ .
בדקו ברמת מובהקות של 2% האם קיים הבדל בין השונויות? מה יש להניח?

**פתרון:**

<u>האוכלוסיות</u>: נשים מול גברים.

<u>משתנה</u>: $y$ = זמן ביצוע משימה של גבר , $x$ = זמן ביצוע משימה של אישה

<u>פרמטר</u>: $\dfrac{\sigma^2_x}{\sigma^2_y}$

<u>השערות</u>:

$$H_0:\frac{\sigma^2_x}{\sigma^2_y}=1$$
$$H_1:\frac{\sigma^2_x}{\sigma^2_y}\neq 1$$

<u>הנחות</u>: 1. מדגימים ב"ת 2. $x,y\sim N$

<u>כלל הכרעה</u>:

$$\alpha=2\%$$
$$n_1=13, d.f_1=n_1-1=12$$
$$n_2=10, d.f_2=n_2-1=9$$

<u>נדחה את</u> $H_0$ אם $F>5.11$ או $F<0.23$

<u>תוצאות המדגם</u>:

$$S^2_y=\frac{204}{10-1}=22\frac{2}{3}$$
$$S^2_x=\frac{200}{13-1}=16\frac{2}{3}$$

<u>סטטיסטי המבחן</u>: $F=\dfrac{S_1^2}{S_2^2}=\dfrac{16\frac{2}{3}}{22\frac{2}{3}}$

<u>מסקנה</u>: ברמת מובהקות של 2% נקבל את $H_0$ .

לא קיים הבדל מובהק בין גברים לנשים מבחינת השונויות שלהם.

$\alpha = 0.05$ טבלת ערכים קריטיים לפי התפלגות F

| ד"ח מנה/ד"ח מכנה | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 16 | 20 | 24 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 | 238.88 | 240.54 | 241.88 | 243.91 | 246.46 | 248.01 | 249.05 | 252.20 | 253.25 | 254.31 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 | 19.41 | 19.43 | 19.45 | 19.45 | 19.48 | 19.49 | 19.50 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 | 8.74 | 8.69 | 8.66 | 8.64 | 8.57 | 8.55 | 8.53 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 | 5.91 | 5.84 | 5.80 | 5.77 | 5.69 | 5.66 | 5.63 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 | 4.68 | 4.60 | 4.56 | 4.53 | 4.43 | 4.40 | 4.37 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 | 4.00 | 3.92 | 3.87 | 3.84 | 3.74 | 3.70 | 3.67 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 | 3.57 | 3.49 | 3.44 | 3.41 | 3.30 | 3.27 | 3.23 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 | 3.28 | 3.20 | 3.15 | 3.12 | 3.01 | 2.97 | 2.93 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 | 3.07 | 2.99 | 2.94 | 2.90 | 2.79 | 2.75 | 2.71 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 | 2.91 | 2.83 | 2.77 | 2.74 | 2.62 | 2.58 | 2.54 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 | 2.79 | 2.70 | 2.65 | 2.61 | 2.49 | 2.45 | 2.40 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 | 2.69 | 2.60 | 2.54 | 2.51 | 2.38 | 2.34 | 2.30 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 | 2.60 | 2.51 | 2.46 | 2.42 | 2.30 | 2.25 | 2.21 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 | 2.53 | 2.44 | 2.39 | 2.35 | 2.22 | 2.18 | 2.13 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 | 2.48 | 2.38 | 2.33 | 2.29 | 2.16 | 2.11 | 2.07 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 | 2.42 | 2.33 | 2.28 | 2.24 | 2.11 | 2.06 | 2.01 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 | 2.38 | 2.29 | 2.23 | 2.19 | 2.06 | 2.01 | 1.96 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 | 2.34 | 2.25 | 2.19 | 2.15 | 2.02 | 1.97 | 1.92 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 | 2.31 | 2.21 | 2.16 | 2.11 | 1.98 | 1.93 | 1.88 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 | 2.28 | 2.18 | 2.12 | 2.08 | 1.95 | 1.90 | 1.84 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 | 2.32 | 2.25 | 2.16 | 2.10 | 2.05 | 1.92 | 1.87 | 1.81 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 | 2.23 | 2.13 | 2.07 | 2.03 | 1.89 | 1.84 | 1.78 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 | 2.27 | 2.20 | 2.11 | 2.05 | 2.01 | 1.86 | 1.81 | 1.76 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 | 2.18 | 2.09 | 2.03 | 1.98 | 1.84 | 1.79 | 1.73 |
| 25 | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 | 2.24 | 2.16 | 2.07 | 2.01 | 1.96 | 1.82 | 1.77 | 1.71 |
| 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 | 2.22 | 2.15 | 2.05 | 1.99 | 1.95 | 1.80 | 1.75 | 1.69 |
| 27 | 4.21 | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.37 | 2.31 | 2.25 | 2.20 | 2.13 | 2.04 | 1.97 | 1.93 | 1.79 | 1.73 | 1.67 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 | 2.19 | 2.12 | 2.02 | 1.96 | 1.91 | 1.77 | 1.71 | 1.65 |
| 29 | 4.18 | 3.33 | 2.93 | 2.70 | 2.55 | 2.43 | 2.35 | 2.28 | 2.22 | 2.18 | 2.10 | 2.01 | 1.94 | 1.90 | 1.75 | 1.70 | 1.64 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 | 2.09 | 1.99 | 1.93 | 1.89 | 1.74 | 1.68 | 1.62 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 | 2.00 | 1.90 | 1.84 | 1.79 | 1.64 | 1.58 | 1.51 |
| 50 | 4.03 | 3.18 | 2.79 | 2.56 | 2.40 | 2.29 | 2.20 | 2.13 | 2.07 | 2.03 | 1.95 | 1.85 | 1.78 | 1.74 | 1.58 | 1.51 | 1.44 |
| 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 | 1.92 | 1.82 | 1.75 | 1.70 | 1.53 | 1.47 | 1.39 |
| 90 | 3.95 | 3.10 | 2.71 | 2.47 | 2.32 | 2.20 | 2.11 | 2.04 | 1.99 | 1.94 | 1.86 | 1.76 | 1.69 | 1.64 | 1.46 | 1.39 | 1.30 |
| 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.18 | 2.09 | 2.02 | 1.96 | 1.91 | 1.83 | 1.73 | 1.66 | 1.61 | 1.43 | 1.35 | 1.25 |
| ∞ | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 | 1.75 | 1.64 | 1.57 | 1.52 | 1.32 | 1.22 | 1.00 |

| טבלת ערכים קריטיים לפי התפלגות F | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | ראה איור מטה. | $\alpha = 0.01$ | | | | | | | | | | | | | |
| ד"ח מונה/ד"ח מכנה | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 16 | 20 | 24 | 60 | 120 | ∞ |
| 1 | 4052.18 | 4999.50 | 5403.35 | 5624.58 | 5763.65 | 5858.99 | 5928.36 | 5981.07 | 6022.47 | 6055.85 | 6106.32 | 6170.10 | 6208.73 | 6234.63 | 6313.03 | 6339.39 | 6365.86 |
| 2 | 98.50 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 | 99.39 | 99.40 | 99.42 | 99.44 | 99.45 | 99.46 | 99.48 | 99.49 | 99.50 |
| 3 | 34.12 | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.35 | 27.23 | 27.05 | 26.83 | 26.69 | 26.60 | 26.32 | 26.22 | 26.13 |
| 4 | 21.20 | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.55 | 14.37 | 14.15 | 14.02 | 13.93 | 13.65 | 13.56 | 13.46 |
| 5 | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 | 10.16 | 10.05 | 9.89 | 9.68 | 9.55 | 9.47 | 9.20 | 9.11 | 9.02 |
| 6 | 13.75 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 | 7.72 | 7.52 | 7.40 | 7.31 | 7.06 | 6.97 | 6.88 |
| 7 | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 | 6.62 | 6.47 | 6.28 | 6.16 | 6.07 | 5.82 | 5.74 | 5.65 |
| 8 | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 | 5.67 | 5.48 | 5.36 | 5.28 | 5.03 | 4.95 | 4.86 |
| 9 | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 | 5.11 | 4.92 | 4.81 | 4.73 | 4.48 | 4.40 | 4.31 |
| 10 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 | 4.71 | 4.52 | 4.41 | 4.33 | 4.08 | 4.00 | 3.91 |
| 11 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 | 4.40 | 4.21 | 4.10 | 4.02 | 3.78 | 3.69 | 3.60 |
| 12 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 | 4.16 | 3.97 | 3.86 | 3.78 | 3.54 | 3.45 | 3.36 |
| 13 | 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 | 4.10 | 3.96 | 3.78 | 3.66 | 3.59 | 3.34 | 3.25 | 3.17 |
| 14 | 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 | 3.80 | 3.62 | 3.51 | 3.43 | 3.18 | 3.09 | 3.00 |
| 15 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 | 3.67 | 3.49 | 3.37 | 3.29 | 3.05 | 2.96 | 2.87 |
| 16 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 | 3.55 | 3.37 | 3.26 | 3.18 | 2.93 | 2.84 | 2.75 |
| 17 | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 | 3.46 | 3.27 | 3.16 | 3.08 | 2.83 | 2.75 | 2.65 |
| 18 | 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 | 3.51 | 3.37 | 3.19 | 3.08 | 3.00 | 2.75 | 2.66 | 2.57 |
| 19 | 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 | 3.43 | 3.30 | 3.12 | 3.00 | 2.92 | 2.67 | 2.58 | 2.49 |
| 20 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 | 3.37 | 3.23 | 3.05 | 2.94 | 2.86 | 2.61 | 2.52 | 2.42 |
| 21 | 8.02 | 5.78 | 4.87 | 4.37 | 4.04 | 3.81 | 3.64 | 3.51 | 3.40 | 3.31 | 3.17 | 2.99 | 2.88 | 2.80 | 2.55 | 2.46 | 2.36 |
| 22 | 7.95 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 | 3.26 | 3.12 | 2.94 | 2.83 | 2.75 | 2.50 | 2.40 | 2.31 |
| 23 | 7.88 | 5.66 | 4.76 | 4.26 | 3.94 | 3.71 | 3.54 | 3.41 | 3.30 | 3.21 | 3.07 | 2.89 | 2.78 | 2.70 | 2.45 | 2.35 | 2.26 |
| 24 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 | 3.03 | 2.85 | 2.74 | 2.66 | 2.40 | 2.31 | 2.21 |
| 25 | 7.77 | 5.57 | 4.68 | 4.18 | 3.85 | 3.63 | 3.46 | 3.32 | 3.22 | 3.13 | 2.99 | 2.81 | 2.70 | 2.62 | 2.36 | 2.27 | 2.17 |
| 26 | 7.72 | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.42 | 3.29 | 3.18 | 3.09 | 2.96 | 2.78 | 2.66 | 2.58 | 2.33 | 2.23 | 2.13 |
| 27 | 7.68 | 5.49 | 4.60 | 4.11 | 3.78 | 3.56 | 3.39 | 3.26 | 3.15 | 3.06 | 2.93 | 2.75 | 2.63 | 2.55 | 2.29 | 2.20 | 2.10 |
| 28 | 7.64 | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.36 | 3.23 | 3.12 | 3.03 | 2.90 | 2.72 | 2.60 | 2.52 | 2.26 | 2.17 | 2.06 |
| 29 | 7.60 | 5.42 | 4.54 | 4.04 | 3.73 | 3.50 | 3.33 | 3.20 | 3.09 | 3.00 | 2.87 | 2.69 | 2.57 | 2.49 | 2.23 | 2.14 | 2.03 |
| 30 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 | 2.84 | 2.66 | 2.55 | 2.47 | 2.21 | 2.11 | 2.01 |
| 40 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 | 2.66 | 2.48 | 2.37 | 2.29 | 2.02 | 1.92 | 1.80 |
| 50 | 7.17 | 5.06 | 4.20 | 3.72 | 3.41 | 3.19 | 3.02 | 2.89 | 2.78 | 2.70 | 2.56 | 2.38 | 2.27 | 2.18 | 1.91 | 1.80 | 1.68 |
| 60 | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 | 2.63 | 2.50 | 2.31 | 2.20 | 2.12 | 1.84 | 1.73 | 1.60 |
| 90 | 6.93 | 4.85 | 4.01 | 3.53 | 3.23 | 3.01 | 2.84 | 2.72 | 2.61 | 2.52 | 2.39 | 2.21 | 2.09 | 2.00 | 1.72 | 1.60 | 1.46 |
| 120 | 6.85 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 | 2.47 | 2.34 | 2.15 | 2.03 | 1.95 | 1.66 | 1.53 | 1.38 |
| ∞ | 6.63 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 | 2.32 | 2.18 | 2.00 | 1.88 | 1.79 | 1.47 | 1.32 | 1.00 |

## שאלות:

**1)** להלן נתונים על שטחי דירות במ״ר עבור דירות חדשות שנבנו בשנת 2012 ובשנת 2013:

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

א. בדקו ברמת מובהקות של 10% את ההשערה ששונויות שטחי הדירות החדשות בשנת 2012 ובשנת 2013 שוות. מה הן ההנחות הדרושות לביצוע הבדיקה?

ב. האם וכיצד הייתה משתנה המסקנה מהסעיף הקודם אם מסתבר שחלה טעות ברישום ויש להפחית 10 מ״ר מכל הדירות שמופיעות במדגם?

**2)** בתחום הבינוי משתמשים בשני סוגי מתכות: מתכת A ומתכת B. מחקר מעוניין לבדוק האם קיים הבדל בין שני סוגי המתכות מבחינת החוזק שלהן. דגמו מספר יחידות מתכת מכל סוג והתקבלו התוצאות הבאות:

| סוג המתכת | A | B |
|---|---|---|
| $n$ | 8 | 10 |
| $\sum X_i$ | 16 | 30 |
| $\sum X_i^2$ | 60 | 198 |

יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.

א. האם קיים הבדל בין שונויות החוזק של מתכות?

ב. האם קיים הבדל בין תוחלות החוזק של מתכות?

בכל סעיף רמת מובהקות של 10%.

**3)** מחקר סוציולוגי מעוניין לחקור את הרגלי הבילויים בקבוצות גיל שונות. ידוע כי בקרב האוכלוסייה הבוגרת (מעל 18) ההוצאה החודשית על בילויים מתפלגת נורמאלית עם תוחלת של 500₪ וסטיית תקן של 300₪.

במדגם שנעשה על סטודנטים בגילאי 21-26 התקבל אומד חוסר הטיה לשונות ההוצאה החודשית על בילויים 10,000₪. כמות הסטודנטים שנדגמה 16.

במדגם שנעשה על 11 מבוגרים בשנות השלושים התקבל אומד חסר הטיה לשונות ההוצאה החודשית על בילויים 490,000₪.

א. בדקו ברמת מובהקות של 5% האם שונות ההוצאה על בילויים בקרב סטודנטים בקבוצת גילאי 21-26 נמוכה מהשונות אצל כלל המבוגרים.

ב. בדקו ברמת מובהקות של 1% האם הפיזור של ההוצאה החודשית לבילויים גדולה יותר בקבוצת גיל ה-30 מאשר בקבוצת גיל 21-26.

**4)** נתון $X_i \sim N(\mu_x,\sigma^2)$, וכמו כן $Y_i \sim N(\mu_y,\sigma^2)$.

מאוכלוסייה $X$ נדגמו 7 תצפיות ומאוכלוסיה $Y$ נדגמו 13 תצפיות.

א. כיצד $\frac{S_x^2}{S_Y^2}$ מתפלג?

ב. מה ההסתברות ש-$S_x^2$ גדולה ביותר מפי 3 מאשר $S_Y^2$?

## תשובות סופיות:

1) א. לא נדחה את $H_0$. ב. מסקנה לא תשתנה.

2) א. לא נדחה את $H_0$. ב. לא נדחה את $H_0$.

3) א. נדחה את $H_0$. ב. נדחה את $H_0$.

4) א. $F(6,12)$ ב. 5%

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 33 - שאלות מסכמות בבדיקת השערות

### תוכן העניינים

# שאלות מסכמות בבדיקת השערות על פרמטרים

## שאלות

**1)** שני חוקרים נתבקשו לבדוק את ההשערות הבאות: $H_1: \mu > 520 \quad H_0: \mu = 520$.
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 $\bar{X} > 522$.

א. למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?

ב. מהי ההסתברות לטעות מסוג שני של חוקר ב' עבור $\mu_1 = 525$.

ג. הסבר ללא חישוב נוסף, האם ההסתברות לטעות מסוג שני עבור $\mu_1 = 525$, של חוקר א' שווה/קטנה/גדולה לזו של חוקר ב'.

ד. 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.

א. מהי רמת המובהקות שכדאי לדני לדרוש, כדי שדנה תשתכנע בצדקת טענתו (שדני פופולרי יותר בכך שהוא מקבל יותר לייקים מדנה ביום).

ב. אם דני משער שתוחלת מספר ה"לייק"ים שהוא מקבל ביום היא 14 וקובע רמת מובהקות 2.5%, מהי עוצמת המבחן של דני?

**3)** ברצוננו להשוות בין רשתות A לבין B. לשם כך בחרנו 4 מוצרים, ובדקנו את מחיריהם בשתי הרשתות. להלן התוצאות:

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

הניחו כי המחירים מתפלגים נורמלית.
אם יש הנחות נוספות כדי לבצע את המבחן הפרמטרי רשמו אותן.

א. בדקו האם קיים הבדל בין הרשתות מבחינת תוחלת המחירים. רמת מובהקות של 5%.

ב. חזרו על הסעיף הקודם בהנחה ונבחרו בכל רשת מוצרים באקראי ולא בהכרח אותם מוצרים.

**4)** במדגם של 10 ישראלים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות:
$\sum X_i^2 = 105120 \quad \sum X_i = 1020, \quad n = 10$.

במדגם של 14 אמריקאים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות:
$\sum X_i^2 = 138644 \quad \sum X_i = 1386, \quad n = 14$.

נתון שציוני הבחינה מתפלגים נורמלית בכל מדינה.

א. בדקו ברמת מובהקות של 10% האם קיים שוויון שונויות בין אוכלוסיית אמריקה לאוכלוסיית ישראל?

ב. בדקו האם קיים הבדל בממוצע הציונים בבחינת ה-IQ בין ישראל לארה"ב. ברמת מובהקות של 5%?

**5)** 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:

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

א. נסחו את ההשערות ובדקו 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%.

**6)** במדינת טרפפו המשכורות במשק מתפלגות נורמלית עם ממוצע של 1 אלף דולר וסטיית תקן של 0.2 אלף דולר. בוצע מדגם מקרי בו השתתפו 5 נשים ו 5 גברים במדינת שומקום שבה המשכורות מתפלגות נורמאלית גם כן. להלן משכורותיהם באלפי דולר:

| | | | | | |
|---|---|---|---|---|---|
| גברים | 2 | 0.9 | 0.7 | 1.2 | 1.1 |
| נשים | 1.4 | 1.1 | 1.9 | 1.8 | 1.2 |

א. בדקו את הטענה שממוצע משכורותיהם של אזרחי שומקום גבוה מאשר ממוצע משכורותיהם של אזרחי טרפפו ברמת מובהקות של 5%. בהנחה שסטיית התקן זהה בשתי המדינות.

ב. חזרו על הסעיף הקודם ללא ההנחה הנ"ל.

ג. ישנה טענה שסטיית התקן במדינת שומקום גבוהה מזו של טרפפו. בדקו ברמת מובהקות של 5%.

7) במטרה להשוות בין אחוזי הצפייה של גברים ונשים בתוכנית טלוויזיה מסוימת בוצע סקר ובו התקבלו תוצאות הבאות:

| | צופים | לא צופים |
|---|---|---|
| נשים | 320 | 42 |
| גברים | 72 | 120 |

א. האם יש הבדל בין אחוזי הצפייה של גברים ונשים ברמת מובהקות של 1%?

ב. עבור רמת מובהקות של 5% בדוק טענה שמבין הצופים בתוכנית הטלוויזיה אחוז הנשים גדול פי 2 מאחוז הגברים.

8) בשנת 2000 ל-60% היה מדיח כלים בבית. מחקר רוצה לבדוק האם כיום פרופורציית המשפחות עם מדיח כלים עלה. הוחלט לבצע מדגם אקראי של 150 משפחות.

א. רשמו את השערות המחקר.

ב. מה היא מסקנת המחקר ברמת מובהקות של 5% אם במדגם ל-102 משפחות היה מדיח כלים.

ג. מהי הטעות האפשרית במסקנה מהסעיף הקודם.
האם ניתן לדעת את הסתברותה?

9) נערך מחקר על הקשר בין עישון ויתר לחץ דם. נבדק מדגם מקרי של 200 מעשנים ונמצא כי 30 סבלו מיתר לחץ דם.
ידוע שבאוכלוסייה 18% סובלים מיתר לחץ דם.

א. בדקו ברמת מובהקות 0.1 את ההשערה כי אחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.

ב. מהי רמת המובהקות המינימלית לקבלת הטענה שאחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.

ג. מהי עצמת המבחן, אם אחוז הסובלים מיתר לחץ דם בקרב אוכלוסיית המעשנים היא בפועל 25%.

10) להלן התפלגות מספר הנסיעות לחופשה השנתית במדגם של משפחות ישראליות. בדקו ברמת מובהקות של 5%:

| מספר הנסיעות | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| מספר המשפחות | 84 | 102 | 26 | 20 | 12 |

א. באיטליה משפחות נוסעות בממוצע פעמיים בשנה לחופשה. האם בישראל משפחות נוסעות פחות מאשר באיטליה?

ב. בהולנד 80% מהמשפחות נוסעות לפחות פעם אחת בשנה לחופשה, האם בישראל אחוז המשפחות שנוסעות לפחות פעם אחת בשנה לחופשה נמוך מאשר בהולנד?

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

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

דגמו 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: $H_1: \mu \neq 40$

**12)** ציוני בחינת הבגרות במתמטיקה מתפלגים נורמלית עם שונות 150. במדגם של 16 נבחנים מתל אביב התקבלה שונות מדגמית-190. במדגם של 25 ירושלמים התקבלה שונות מדגמית 118.

א. בדקו ברמת מובהקות של 2.5% האם שונות הציונים במתמטיקה בקרב נבחני תל אביב גבוהה מהשונות בכלל הארץ.

ב. בדקו ברמת מובהקות של 5% האם שונות ציונים במתמטיקה בקרב תלמידי תל אביב גבוהה מאשר בקרב תלמידי ירושלים.

## 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. לא נדחה $H_0$. | | |
| **5)** | א. לא נדחה $H_0$. | ב. בין 5% ל-10%. | ג. נדחה $H_0$. | |
| **6)** | א. נדחה $H_0$. | ב. נדחה $H_0$. | ג. נדחה $H_0$. | |
| **7)** | א. נדחה $H_0$. | ב. נדחה $H_0$. | | |
| **8)** | א. $H_0: p = 0.6$ <br> $H_1: p > 0.6$ | ב. נדחה $H_0$. | ג. טעות מסוג ראשון בסיכוי של 0.05. | |
| **9)** | א. לא נדחה $H_0$. | ב. 0.8643 | ג.0.8749. | |
| **10)** | א. נדחה $H_0$. | ב. נדחה $H_0$. | | |
| **11)** | א. 0.0062 | ב. 0.9938 | ג. 0.0124. | |
| **12)** | א. לא נדחה $H_0$. | ב. לא נדחה $H_0$. | | |

## 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 $\alpha = 0.05$?

א. השערת האפס הייתה נדחית.

ב. השערת האפס לא הייתה נדחית.

ג. ההשערה המחקרית הייתה נדחית.

ד. בהעדר נתונים נוספים, לא ניתן לדעת.

**2)** על מנת לבדוק האם ההסתברות ללידת בן הינה חצי, נבחר מדגם מקרי של 200 ילדים, ונמצא שישנם 120 בנים.
מהן ההשערות האלטרנטיבית להשערת האפס?

א. $H_1: p = 0.5$

ב. $H_1: p = 0.6$

ג. $H_1: p > 0.5$

ד. $H_1: p \neq 0.5$

**3)** 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:

א. מבחן Z למדגם יחיד.

ב. מבחן Tלמדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

**4)** 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 $X - Y$)

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

א. $H_1: \mu_d < 0,\ \ H_0: \mu_d = 0$

ב. $H_1: \mu_X - \mu_Y < 0,\ \ H_0: \mu_X - \mu_Y = 0$

ג. $H_1: \mu_X - \mu_Y < 0,\ \ H_0: \mu_X - \mu_Y = 0$

ד. $H_1: \mu_d > 0,\ \ H_0: \mu_d = 0$

**5)** 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.

**6)** 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.

**7)** 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.

**8)** 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:

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

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.

**9)** 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'?

א. לדחות את השערת האפס.
ב. לא לדחות את השערת האפס.
ג. ללא נתונים נוספים אי אפשר לדעת מה יחליט.

**10)** חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?

א. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
ב. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
ג. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.
ד. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

**11)** 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?

א. טעות מסוג ראשון.
ב. טעות מסוג שני.
ג. טעות מסוג שלישי.
ד. לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמתית אצל הצמחוניים.

**12)** בסקר שנערך התקבל ש 60% מתוך 220 נשאלים מבקרים אצל השיננית לפחות פעם אחת בשנה. עבור אילו רמות מובהקות ניתן יהיה לקבוע שרוב האוכלוסייה מבקרת אצל השיננית לפחות פעם בשנה?

א. רמת מובהקות הגדולה מ-5%.
ב. רמת מובהקות הקטנה מ-5%.
ג. רמת מובהקות הגדלה מ-0.0015.
ד. רמת מובהקות הקטנה מ-0.0015.

**13)** 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?

א. אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב' באותה רמת מובהקות.
ב. אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס באותה רמת מובהקות.
ג. אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת האפס באותה רמת מובהקות.
ד. אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת האפס בתנאי שרמת המובהקות כפולה בגודלה.

**14)** 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?

א. תוחלת ציוני העולים באוכלוסייה.
ב. ממוצע ציוני העולים במדגם.
ג. תוחלת ציוני האוכלוסייה מנתוני העבר.
ד. ממוצע ציוני שאר האוכלוסייה במדגם.

**15)** חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?

א. החוקר דחה את השערת $H_0$ כאשר היא הייתה נכונה.
ב. החוקר דחה את השערת $H_1$ כאשר היא הייתה נכונה.
ג. החוקר לא דחה את השערת $H_0$ כאשר היא הייתה לא נכונה.
ד. המדגם של החוקר שייך בפועל להתפלגות הדגימה של $H_1$.

**16)** 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?

א. מבחן T למדגמים בלתי תלויים עם 38 דרגות חופש.
ב. מבחן T למדגמים מזווגים, עם 39 דרגות חופש.
ג. מבחן T למדגמים בלתי תלויים עם 39 דרגות חופש.
ד. מבחן T למדגמים מזווגים עם 38 דרגות חופש.

**17)** בינואר השנה פורסם שהשכר הממוצע במשק הוא 8,900₪. במדגם שנעשה בחודש יוני על 60 עובדים נרשם עבור כל עובד במדגם האם השכר שלו נמוך או לא נמוך מהשכר הממוצע שפורסם בחודש ינואר. מהו המבחן המתאים כדי לבדוק שרוב העובדים בחודש יוני קיבלו שכר הנמוך מהשכר הממוצע שפורסם בחודש ינואר?

א. מבחן Z על פרופורציה.
ב. מבחן T על תוחלת אחת.
ג. מבחן T על שתי תוחלות במדגמים בלתי תלויים.
ד. מבחן T על שתי תוחלות במדגמים תלויים.

**18)** 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 :

א. שלושתם במבחן T למדגמים בלתי תלויים.
ב. עידו במבחן T למדגמים מזווגים, ורון ויואב במבחן T למדגמים בלתי תלויים.
ג. עידו במבחן T למדגמים מזווגים, רון במבחן T למדגמים בלתי תלויים ויואב במבחן T למדגם יחיד.
ד. עידו במבחן T למדגמים מזווגים, רון ויואב במבחן T למדגם יחיד.

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

א. הגדלת רמת המובהקות לא תשתנה 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.

**20)** 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:

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

**21)** 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?

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

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

**23)** 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.

**24)** 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.

**25)** מה המצב הרצוי לחוקר המבצע בדיקת השערה:

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

**26)** 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.

**27)** 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.

**28)** בבדיקת השערות התקבל שה- $p-value=0.02$.
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.

**29)** 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.

**30)** בבדיקת השערות מסוימת התקבל $p\ value = 0.0254$, 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 $H_0$.

**31)** 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.

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

**33)** 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

**34)** 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 | 4.5 | 9.6 | 0.7 | 2.5 |
| משקל במכשיר 2 | 3.5 | 6.9 | 1.7 | 0.5 |

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.

**35)** 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.

**36)** סטטיסטיקאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0.
המשמעות של תוצאה זו היא:
א. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
ב. תמיד לדחות את השערת האפס כאשר היא נכונה.
ג. לעולם לא לדחות את השערת האפס כאשר היא נכונה.
ד. תמיד לדחות את השערת האפס כאשר היא לא נכונה.

**37)** 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.

**38)** 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 מתפלגים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו: $-0.0293 < \mu_D < 0.2145$ רווח הסמך הוא ברמת סמך של 95%.
לכן מסקנת המחקר היא:

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

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

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

**40)** חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות: $H_0: \mu=10, \quad H_1: \mu\neq 10$.
לצורך בדיקה הוא לקח מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך תוצאות המדגם הוא חישב וקיבל: $t_{\bar{x}} = -2.63$. לכן המסקנה היא:

א. הוא ידחה $H_0$ ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
ב. הוא ידחה $H_0$ ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025.
ג. הוא ידחה $H_0$ ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01
ד. הוא לא ידחה $H_0$ ברמת מובהקות 0.1.

**41)** 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?

א. 0.005
ב. 0.03
ג. 0.0287
ד. 0.05

**42)** 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 **במציאות** השערתה של החוקרת נכונה, סביר כי אם היא תגדיל את גודל המדגם, אזי :

א. יקטן הסיכוי לטעות מסוג שני ($\beta$).

ב. תגדל רמת הביטחון ($1-\alpha$).

ג. אף תשובה לא נכונה.

ד. כל התשובות נכונות.

**43)** איזה מהמשפטים הבאים נכון תמיד?

א. $\text{POWER}+\alpha+\beta=1$

ב. $\text{POWER}=0.5-\beta$

ג. $\text{POWER}+\alpha=1$

ד. $\beta+\alpha=1$

ה. הכול לא נכון.

**44)** מה נכון לומר לגבי הנחת שיוויון השונויות במבחן T למדגמים בלתי תלויים?

א. היא אומרת שהשונויות המדגמיות שוות.

ב. בלעדיה אין שום דרך לבדוק השערה על הפרש בין תוחלות.

ג. היא חשובה הן עבור מדגמים מזווגים והן עבור מדגמים בלתי תלויים.

ד. אף תשובה אינה נכונה.

**45)** 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?

א. ההשערה תדחה.

ב. ההשערה לא תדחה.

ג. התשובה תלויה בעוצמת המבחן.

ד. לא ניתן לדעת בוודאות אם ההשערה תדחה או לא.

**46)** 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 :

א. אפס

ב. חיובי

ג. שלילי

ד. לא ניתן לדעת

**47)** עוצמה שווה ל-1 פרושה:
א. לעולם לא לדחות את השערת האפס כאשר היא נכונה.
ב. תמיד לדחות את השערת האפס כאשר היא נכונה.
ג. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

**48)** מה מהבאים **נכון** לגבי מבחן T מדגמים מזווגים?
א. כל התצפיות במחקר אינן תלויות זו בזו.
ב. כל התצפיות במחקר תלויות זו בזו.
ג. כל הצמדים של תצפיות במחקר אינם תלויים זה בזה.
ד. התצפיות בתוך כל צמד אינן תלויות זו בזו.

**49)** 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 $H_0: \mu = 10, \quad H_1: \mu \neq 10$, אז על סמך תוצאת אותו המדגם ברמת מובהקות 0.05:
א. ניתן להכריע בין ההשערות רק אם שונות האוכלוסייה נתונה.
ב. מקבלים את השערת האפס.
ג. דוחים את השערת האפס.
ד. לא ניתן להכריע בין ההשערות שכן חסרים נתונים.

**50)** לבדיקת ההשערה החד צדדית ימנית $H_0: \mu = 55, \quad H_1: \mu = 65$,
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: $H_0: \mu = 55, \quad H_1: \mu = 60$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות: $H_0: \mu = 55, \quad H_1: \mu = 65$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל $n$ ורמת מובהקות 10% לבדיקת ההשערות: $H_0: \mu = 55, \quad H_1: \mu = 65$ העוצמה תהיה קטנה מ-0.9.

א. שלושת הטענות אינן נכונות.
ב. טענות 2 ו-3 אינן נכונות וטענה 1 נכונה.
ג. טענות 1 ו-2 נכונות וטענה 3 אינה נכונה.
ד. טענות 1 ו-3 אינן נכונות וטענה 2 נכונה.

## תשובות סופיות:

| שאלה | תשובה | שאלה | תשובה |
|---|---|---|---|
| 1 | א | 26 | א |
| 2 | ד | 27 | ב |
| 3 | ד | 28 | א |
| 4 | א | 29 | א |
| 5 | ג | 30 | ג |
| 6 | ג | 31 | ד |
| 7 | א | 32 | א |
| 8 | ב | 33 | א |
| 9 | ג | 34 | ד |
| 10 | א | 35 | ג |
| 11 | א | 36 | א |
| 12 | ג | 37 | ג |
| 13 | א | 38 | א |
| 14 | א | 39 | ג |
| 15 | א | 40 | א |
| 16 | א | 41 | א |
| 17 | א | 42 | א |
| 18 | ג | 43 | ה |
| 19 | א | 44 | ד |
| 20 | ג | 45 | ד |
| 21 | ב | 46 | א |
| 22 | ג | 47 | ד |
| 23 | ד | 48 | ג |
| 24 | ג | 49 | ב |
| 25 | ג | 50 | ד |

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 34 - מבחני חי בריבוע

### תוכן העניינים

## מבחן טיב התאמה – רקע

מבחן זה בא לבדוק האם אוכלוסייה מסוימת מתפלגת לפי התפלגות נתונה.
המשתנה הנחקר מחולק למספר קטגוריות ויש לבדוק האם תוצאות המדגם תואמות להתפלגות הנתונה.

**מבנה המבחן:**

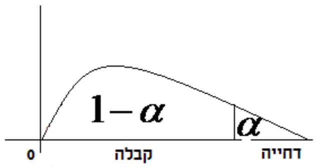
**השערות:**

המשתנה מתפלג לפי התפלגות מסוימת - $H_0$.

אחרת - $H_1$.

**כלל הכרעה:**

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש $d.f=K-1$, כאשר $K$ - מספר הקטגוריות.




הערך הקריטי הוא: $\chi^2_{1-\alpha,K-1}$, כלומר האחוזון ה- $1-\alpha$ בהתפלגות חי בריבוע שדרגות החופש הן $K-1$.
אם $\chi^2>\chi^2_{1-\alpha,K-1}$, דוחים את השערת האפס.

סטטיסטי המבחן: $\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}$ .

$O_i$ - השכיחות שנצפתה במדגם בקטגוריה $i$.

$p_i$ - הסתברות לקטגוריה i לפי השערת האפס.

$E_i=np_i$ - שכיחות צפויה במדגם לקטגוריה $i$ בהנחת השערת האפס.

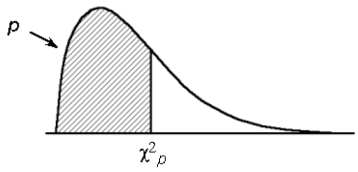
**הערה:**

תנאי כדי לבצע את המבחן הוא $E_i\geq 5$ לכל $i$. במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.

**דוגמה (פתרון בהקלטה) :**

במדינה מסוימת שלוש מפלגות. בפרלמנט הנוכחי התפלגות מספר המושבים היא 30% למפלגה A, ‏60% למפלגה B ו-10% למפלגה C.
לקראת הבחירות המתוכננות בשבוע הבא נעשה סקר שכלל 300 אזרחים.
בסקר התקבל ש40% יצביעו למפלגה A, ‏50% למפלגה B ו-10% למפלגה C.
האם תוצאות הסקר תואמות להתפלגות המושבים בפרלמנט הנוכחי?
בדקו ברמת מובהקות של 5%.

## טבלת התפלגות חי-בריבוע – ערכי החלוקה




| df | .005 | .01 | .025 | .05 | .10 | .25 | .50 | .75 | .90 | .95 | .975 | .99 | .995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $0.0^4393$ | $0.0^3157$ | $0.0^3982$ | $0.0^2393$ | 0.0158 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 0.211 | 0.575 | 1.39 | 2.77 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 |
| 3 | 0.0717 | 0.115 | 0.216 | 0.352 | 0.584 | 1.21 | 2.37 | 4.11 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.06 | 1.92 | 3.36 | 5.39 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 |
| 5 | 0.412 | 0.554 | 0.831 | 1.15 | 1.61 | 2.67 | 4.35 | 6.63 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 3.45 | 5.35 | 7.84 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 4.25 | 6.35 | 9.04 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 5.07 | 7.34 | 10.2 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 5.90 | 8.34 | 11.4 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 6.74 | 9.34 | 12.5 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 7.58 | 10.3 | 13.7 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 8.44 | 11.3 | 14.8 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 9.30 | 12.3 | 16.0 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 10.2 | 13.3 | 17.1 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 11.0 | 14.3 | 18.2 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 11.9 | 15.3 | 19.4 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 12.8 | 16.3 | 20.5 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 13.7 | 17.3 | 21.6 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 14.6 | 18.3 | 22.7 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 15.5 | 19.3 | 23.8 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 16.3 | 20.3 | 24.9 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 17.2 | 21.3 | 26.0 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 18.1 | 22.3 | 27.1 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 19.0 | 23.3 | 28.2 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 19.9 | 24.3 | 29.3 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 20.8 | 25.3 | 30.4 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 21.7 | 26.3 | 31.5 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 22.7 | 27.3 | 32.6 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 23.6 | 28.3 | 33.7 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 24.5 | 29.3 | 34.8 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 |

## שאלות

**1)** במטרה לבדוק האם קובייה הוגנת, מטילים אותה 120 פעמים.
התקבל 17 פעמים 1, 23 פעמים 2, 20 פעמים 3, 25 פעמים 4, 18 פעמים 5 ו-17 פעמים 6. מה מסקנתכם ברמת מובהקות של 5%?

**2)** מפעל מייצר סוכריות בצבעים כחול, אדום, ירוק וכתום. מעוניינים לבדוק שפרופורציית הסוכריות הכחולות גדולה פי 2 מכל צבע אחר. לצורך כך נדגמו באקראי 200 סוכריות והתקבל: 70 כחולות, 50 אדומות, 40 ירוקות והיתר כתומות. מה מסקנתכם ברמת מובהקות של 5%?

**3)** משרד החינוך טוען שבקרב השכירים במשק היחס בין השכירים בעלי השכלה נמוכה, תיכונית ואקדמאית הוא 1:2:1 בהתאמה. במדגם של 200 שכירים התקבלו 56 אנשים בעלי השכלה נמוכה, 105 בעלי השכלה תיכונית והיתר בעלי השכלה גבוהה.

א. על סמך תוצאות המדגם, האם התפלגות ההשכלה היא כמו שמשרד החינוך מפרסם? בדוק ברמת מובהקות של 5%.

ב. בנו רווח סמך ברמת סמך של 95% לפרופורציית השכירים במשק בעלי השכלה אקדמאית.

**4)** 200 איש נתבקשו לבחור ספרה באקראי והנה התוצאות שהתקבלו:
18 איש בחרו בספרה 0, 24 איש בחרו בספרה 1, 17 איש בחרו בספרה 2, 19 איש בחרו בספרה 3, 20 איש בחרו בספרה 4, 18 איש בחרו בספרה 5, 22 איש בחרו בספרה 6 והיתר בחרו בספרות 7-9.

א. על סמך התוצאות הללו האם בחירת הספרות אקראית?
בדקו ברמת מובהקות של 2.5%.

ב. תנו הערכה למובהקות התוצאה.

ג. אם נגדיל את גודל המדגם פי 2 ונשמור על אותם יחסים של כמות האנשים במדגם שבחרו בספרות, כיצד הדבר ישפיע על ערכו של הסטטיסטי $\chi^2$? מה תהיה המסקנה במקרה זה?

**5)** מעוניינים לבדוק האם קובייה היא הוגנת. הטילו את הקובייה פעמיים והתבוננו בסכום הוצאות. חזרו על התהליך 72 פעמים.
להלן התוצאות שהתקבלו במדגם:
מה המסקנה ברמת מובהקות של 5%?

| סכום התוצאות | מספר הטלות |
|---|---|
| 2-5 | 20 |
| 6-8 | 17 |
| 9-10 | 20 |
| 11-12 | 15 |

6) בפנס יש 4 סוללות. בבדיקה שנערכה ב-400 פנסים נמצאו סוללות פגומות לפי השכיחויות הבאות:

| מספר הסוללות הפגומות | 0 | 1 | 2 | 3 ומעלה |
|---|---|---|---|---|
| שכיחות | 276 | 104 | 12 | 8 |

מעוניינים לבדוק על סמך תוצאות מדגם אלה האם הסיכוי לסוללה פגומה הוא 20%. מהי רמת המובהקות המינימלית עבורה נכריע שהסיכוי לסוללה פגומה אינו 20%?

7) מטילים מטבע עד שלראשונה מתקבל "ראש". חוזרים על התהליך 120 פעמים. נסמן ב- $X$ את מספר ההטלות עד קבלת הראש. להלן התוצאות שהתקבלו:

| x | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| מספר החזרות על התהליך | 54 | 20 | 16 | 22 | 6 | 2 |

א. בהנחה והמטבע הוגן, מהי ההתפלגות של $X$?
ב. בדקו האם המטבע הוגן, על סמך תוצאות המדגם ברמת מובהקות של 5%.

8) להלן השערות מחקר: $H_0: X \sim N(40,2^2), \quad H_1: else$.
תוצאות המדגם הן:

| $X$ | מתחת 36 | 36-40 | 40-44 | מעל 44 |
|---|---|---|---|---|
| מספר הדגימות | 3A | 50A | 45A | 2A |

מהו ערכו המקסימלי של A עבורו נקבל את $H_0$ ברמת מובהקות של 5%?

## תשובות סופיות

1) לא נדחה $H_0$.

2) לא נדחה $H_0$.

3) א. לא נדחה $H_0$. ב. (0.14,0.25).

4) א. לא נדחה $H_0$. ב. בין 0.95 ל-0.975
ג. יגדל פי 2; המסקנה לא תשתנה.

5) נכריע שהקובייה אינה הוגנת.

6) 0.005.

7) א. $X \sim G(0.5)$ ב. נסיק שהמטבע לא הוגן.

8) 14.

## הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה – רקע

אם אנו רוצים לבצע מבחן טיב התאמה על משתנה שיש לו שתי קטגוריות בלבד (משתנה דיכוטומי), הדבר זהה לתהליך של בדיקת השערות דו צדדית על פרופורציה בודדת.

**דוגמה (פתרון בהקלטה) :**

הטילו מטבע 80 פעמים וקיבלו 48 פעמים את התוצאה "ראש".
בדקו האם המטבע הוא הוגן ברמת מובהקות של 5%.

א. באמצעות מבחן טיב התאמה.
ב. באמצעות מבחן Z לפרופורציה בודדת.

## שאלות

**1)** בסקר שנעשה על 320 נשאלים, 43.75% טענו שהחיה המועדפת עליהם היא כלב. עד היום היה נהוג לחשוב ש40% מהאנשים מעדיפים כלבים.
בדקו ברמת מובהקות של 5% האם הסקר יישנה את הסברה שהייתה נהוגה עד היום לגבי העדפת כלב.

א. באמצעות מבחן טיב התאמה.

ב. באמצעות מבחן על פרופורציה.

**2)** לסוכנות מכוניות שלושה סניפים ברחבי הארץ. המכוניות נמכרות בסניפים השונים. מתוך 100 מכוניות נמצא ש-65 נמכרו בסניף תל-אביב, 23 בסניף ירושלים והיתר בסניף חיפה.

א. בדקו ברמת מובהקות של 5% האם שיעור המכוניות שנמכרות בסניף ת"א גדול פי 2 מכל סניף אחר.

ב. בדקו באמצעות מבחן טיב התאמה האם 60% מהמכוניות נהוגות להימכר בסניף תל אביב. האם יש דרך אחרת לבדוק את ההשערה?

**3)** בתחרות ריצה בית ספרית שלושה מסלולי ריצה.
ב-50 תחרויות בדקו באיזה מסלול היה הניצחון.
התוצאות שהתקבלו מסוכמות בטבלה הבאה:

| המסלול | 1 | 2 | 3 |
|---|---|---|---|
| מספר הניצחונות | 20 | 15 | 15 |

א. בדקו ברמת מובהקות של 5% האם יש מסלול מועדף לניצחון.

ב. בדקו ברמת מובהקות של 5% האם הסיכוי לנצח במסלול מספר 1 גבוה מ- $\frac{1}{3}$.

## תשובות סופיות

**1)** לא נדחה $H_0$.

**2)** א. נדחה $H_0$. ב. לא נדחה $H_0$.

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

## מבחן חי בריבוע לאי תלות בין משתנים – רקע

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים.
שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות.

**מבנה המבחן:**

**השערות:**

אין תלות בין המשתנים $H_0$.

יש תלות בין המשתנים $H_1$.

**כלל הכרעה:**

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש $d.f=(r-1)(c-1)$.

כאשר: $r$ - מספר הקטגוריות של המשתנה שבשורות.
$c$ - מספר הקטגוריות של המשתנה שבעמודות.

הערך הקריטי הוא: $\chi^2_{1-\alpha,(r-1)(c-1)}$, כלומר האחוזון ה- $1-\alpha$ בהתפלגות חי בריבוע שדרגות החופש הן $(r-1)(c-1)$. אם $\chi^2>\chi^2_{1-\alpha,(r-1)(c-1)}$ אז דוחים את השערת האפס.

סטטיסטי המבחן: $\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}$.


כאשר:

$O_i$- השכיחות נצפית במדגם בתא $i$.
$E_i$- שכיחות צפויה במדגם בתא $i$ בהנחת השערת האפס.

$$E_i=\frac{f(x)\cdot f(y)}{n}$$

**הערה:**

תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל $i$. במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.

תנאי חלופי: אין $E$ קטן מ-1 וגם אין ביותר מ 20% מהתאמים $E$ קטן מ-5.

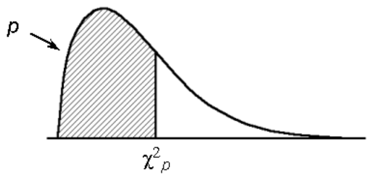
**דוגמה (הפתרון בהקלטה):**

האם יש תלות בין המגדר לבין דעה מסוימת?

יש לבדוק ברמת מובהקות של 5% על סמך תוצאות הסקר:

| המגדר / דעה | בעד | נגד | נמנע | סה"כ |
|---|---|---|---|---|
| גברים | 50 | 40 | 10 | |
| נשים | 20 | 60 | 20 | |
| סה"כ | | | | |

## טבלת התפלגות חי-בריבוע – ערכי החלוקה $\chi^2 p$




| df | .005 | .01 | .025 | .05 | .10 | .25 | .50 | .75 | .90 | .95 | .975 | .99 | .995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $0.0^4393$ | $0.0^3157$ | $0.0^3982$ | $0.0^2393$ | 0.0158 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 0.211 | 0.575 | 1.39 | 2.77 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 |
| 3 | 0.0717 | 0.115 | 0.216 | 0.352 | 0.584 | 1.21 | 2.37 | 4.11 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.06 | 1.92 | 3.36 | 5.39 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 |
| 5 | 0.412 | 0.554 | 0.831 | 1.15 | 1.61 | 2.67 | 4.35 | 6.63 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 3.45 | 5.35 | 7.84 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 4.25 | 6.35 | 9.04 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 5.07 | 7.34 | 10.2 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 5.90 | 8.34 | 11.4 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 6.74 | 9.34 | 12.5 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 7.58 | 10.3 | 13.7 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 8.44 | 11.3 | 14.8 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 9.30 | 12.3 | 16.0 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 10.2 | 13.3 | 17.1 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 11.0 | 14.3 | 18.2 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 11.9 | 15.3 | 19.4 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 12.8 | 16.3 | 20.5 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 13.7 | 17.3 | 21.6 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 14.6 | 18.3 | 22.7 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 15.5 | 19.3 | 23.8 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 16.3 | 20.3 | 24.9 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 17.2 | 21.3 | 26.0 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 18.1 | 22.3 | 27.1 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 19.0 | 23.3 | 28.2 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 19.9 | 24.3 | 29.3 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 20.8 | 25.3 | 30.4 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 21.7 | 26.3 | 31.5 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 22.7 | 27.3 | 32.6 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 23.6 | 28.3 | 33.7 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 24.5 | 29.3 | 34.8 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 |

## שאלות

**1)** נבדקה התלות בין גודל הארגון לבין שביעות הרצון של העובדים.
להלן התוצאות:

| שביעות רצון<br>גודל המפעל | נמוכה | בינונית | גבוהה | סה"כ |
|---|---|---|---|---|
| גדול | 182 | 203 | 215 | 600 |
| קטן | 154 | 110 | 136 | 400 |
| סה"כ | 336 | 313 | 351 | 1000 |

מה המסקנה ברמת מובהקות של 2.5%?

**2)** מפעל עובד בשלוש משמרות. להלן מספר המוצרים הפגומים והתקינים בכל אחת מן המשמרות לפי מדגם שנעשה:

| | יום | ערב | לילה |
|---|---|---|---|
| פגומים | 50 | 60 | 70 |
| תקינים | 600 | 700 | 800 |

האם יש הבדל בין שיעורי הפגומים במשמרות השונות?
הסיקו עבור רמת מובהקות $\alpha = 0.05$.

**3)** נדגמו 50 מוצרים ממפעל מסוים מתוך 30 מוצרים שיוצרו ביום 17 נבחרו לייצוא מתוך המוצרים שיוצרו בלילה 10 נבחרו לייצוא.
האם יש קשר בין היות מוצר לייצוא למועד שבו הוא יוצר?
בדקו ברמת בטחון של 95%.

**4)** במטרה לבדוק האם השתנו דפוסי ההצבעה למפלגות השונות בין שבוע שעבר לשבוע נלקחו שני סקרים אחד מהשבוע שעבר והאחר מהשבוע.
להלן דפוסי ההצבעה שהתקבלו בסקרים אלה.

א. מהי רמת המובהקות המינמלית עבורה ניתן להחליט שהשתנו דפוסי ההצבעה משבוע שעבר לשבוע באופן מובהק?

ב. כיצד הייתה התשובה לסעיף א משתנה אם כל השכיחויות בטבלה של תוצאות המדגם היו מוכפלות פי 2?

ג. בנו רווח סמך לשיעור המצביעים למפלגה א השבוע ברמת סמך של 95%.

| | מפלגה א | מפלגה ב | מפלגות אחרות | סה"כ |
|---|---|---|---|---|
| שבוע שעבר | | | | |
| השבוע | 143 | | 253 | 550 |
| סה"כ | 243 | 314 | | 1050 |

**5)** בחנות בגדים A בדקו את התפלגות הצבעים של הבגדים הנמכרים ביום מסוים. כמו כן בדקו את התפלגות הצבעים בחנות שכנה B:

| מספר פריטים / צבע | שחור | לבן | אדום | כחול |
|---|---|---|---|---|
| חנות A | 15 | 20 | 15 | 50 |
| חנות B | 60 | 20 | 10 | 20 |

א. בדקו ברמת מובהקות של 5% האם התפלגות הצבעים בחנות A היא ביחס של 1:1:1:3 לטובת הכחול.

ב. בדוק ברמת מובהקות של 2.5% האם קיים הבדל בין החניות מבחינת התפלגות הצבעים של הפריטים הנמכרים.

**6)** סטודנט קיבל בבדיקת השערות ערך $\chi^2$ (chi-square) השווה לאפס.
הסטודנט הסיק כי לא קיימת תלות בין שני המשתנים שבדקו, בכל רמת מובהקות. נכון / לא נכון? נמקו.

**7)** להלן טבלת O של שני משתנים שהתקבל במדגם כלשהו:

| | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $f(x)$ |
|---|---|---|---|---|---|
| $X_1$ | | | | | 200 |
| $X_2$ | | | | | 200 |
| $f(y)$ | 60 | 60 | 120 | 160 | |

מה צריכות להיות השכיחויות בתוך הטבלה כדי שמובהקות התוצאה (PV) תהיה 100%?

## תשובות סופיות

**1)** נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.

**2)** נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.

**3)** נסיק שאין קשר בין היות מוצא לייצוא למועד שבו הוא יוצר.

**4)** א. 10% ב. קטן ג. (0.223,0.297)

**5)** א. נסיק שהתפלגות הצבעים בחנות היא כמו שמצוין.
ב. נסיק שיש הבדל בין החנויות מבחינת התפלגות הצבעים.

**6)** נכון

**7)** להלן טבלה:

| | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $f(x)$ |
|---|---|---|---|---|---|
| $X_1$ | 30 | 30 | 60 | 80 | 200 |
| $X_2$ | 30 | 30 | 60 | -8 | 200 |
| $f(y)$ | 60 | 60 | 120 | 160 | 400 |

## הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות – רקע

מבחן לאי תלות שבו לכל משתנה יש שתי קטגוריות שקול לבדיקת השערות דו צדדית על הפרש פרופורציות כאשר השערת האפס היא שהפרופורציות שוות. כל זאת, כמובן, אם התנאים למבחנים מתקיימים.

**דוגמה (פתרון בהקלטה) :**

בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? האם אפשר לבדוק זאת גם על ידי מבחן לאי תלות וגם על ידי בדיקת השערות לשתי פרופורציות?

## שאלות

**1)** בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים?
בדוק ברמת מובהקות של 5%.

א. על ידי מבחן לאי תלות.

ב. על ידי בדיקת השערות לשתי פרופורציות.

**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 ניתן להגיד שהסיכוי להימצא במצב בריאותי תקין גבוה יותר כאשר עוסקים בפעילות גופנית סדירה לעומת המצב שלא עוסקים בפעילות גופנית סדירה? בדוק ברמת בטחון של 90%.

## תשובות סופיות

**1)** נדחה את השערת האפס לפירוט הדרך ראה וידאו.

**2)** א. להלן טבלה: ב. נדחה את השערת האפס.

| פעילות/מצב בריאותי | תקין | לא תקין | |
|---|---|---|---|
| סדירה | 50 | 10 | 60 |
| לא סדירה | 90 | 50 | 140 |
| | 140 | 60 | 200 |

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 35 - מבחנים אפרמטרים למדגמים מזווגים

### תוכן העניינים

# מבחנים אפרמטרים למדגמים מזווגים

## מבחן הסימן – רקע

מבחן הסימן הוא מבחן שמשתמשים בו כאשר לפנינו מדגם מזווג ולא ניתן להניח שהמשתנה הנחקר מתפלג נורמלית.

גם אם המשתנה הנחקר מתפלג נורמלית ניתן לבצע את מבחן הסימן אבל מבחן T למדגמים מזווגים יהיה מבחן עם עוצמה גבוהה יותר ולכן יש לבצע אותו.
מבחן הסימן נחשב למבחן אפרמטרי – מבחנים אפרמטרים הינם כל המבחנים שאינם דורשים שהמשתנה הנחקר יתפלג נורמלית.
מבחן הסימן נקרא כך כיוון שהוא דורש הגדרת סימן לכל תצפית:

$(+)$ – אם מצב $X$ גבוה ממצב $Y$.

$(-)$ – אם מצב $Y$ גבוה ממצב $X$.

$(0)$ – אין הבדל בין המצבים.

במבחן הסימן נתעלם מהפרשים שהם 0, ולכן נסמן את מספר ההפרשים האפקטיביים (השונים מאפס) ב- $n^*$.
תחת השערת האפס נאמר שהסיכוי לקבל הפרש חיובי $(p+)$ שווה לסיכוי לקבל הפרש שלילי $(p-)$.

**השערות המבחן:**

$$H_0: P_+ = 0.5 \qquad H_1: P_+ \neq,<,> 0.5$$

**או**

$$H_0: P_- = 0.5 \qquad H_1: P_- \neq,<,> 0.5$$

נסמן ב $n(+)$ או ב $S_+$ את מספר התצפיות שקיבלו את הערך (+), ובאופן דומה:
נסמן ב $n(-)$ או ב $S_-$ את מספר התצפיות שקיבלו את הערך (–).

ניתן לומר שבהנחת השערת האפס: $n(+)\setminus n(-) \sim B(n^*, 0.5)$.

נחשב את PV על סמך תוצאות המדגם בעזרת ההתפלגות הבינומית כך שאם $PV \le \alpha$ נדחה את השערת האפס.
במבחן הסימן אין התייחסות לגודל הפער בתצפיות אלא רק את כיוון ההבדל.

**דוגמה (פתרון בהקלטה):**

קופות החולים טוענות כי רכישת תרופות שאינן דורשות מרשם רופא, הינן זולות יותר אצלן מאשר מברשתות הפארם. דגמו 11 תרופות ובדקו את מחירן בבית המרקחת של קופות החולים וברשת הפארם. המחיר המוצג הינו עבור קפסולה בודדת:
בדקו ברמת מובהקות של 5% באמצעות מבחן הסימן.

| שם התרופה | קופת חולים | פארם |
|---|---|---|
| אדוויל | 1.2 | 1.5 |
| אקמול | 2.6 | 2.6 |
| אופטלגין | 0.9 | 1.4 |
| פוסטינור | 3.5 | 3.2 |
| סטרפסיל | 1.1 | 1.4 |
| נורפן | 1.7 | 1.8 |
| לורסטין | 0.8 | 1.1 |
| קולדקס | 1.5 | 2 |
| אלרגיז | 2 | 2.8 |
| נוסידקס | 2 | 2.5 |
| קורמיר | 3 | 3.3 |

## שאלות

**1)** רוצים לבדוק את הטענה שהציונים במבחן בסטטיסטיקה ב גבוהים מאשר בסטטיסטיקה א. נלקחו 10 סטודנטים שסיימו את סטטיסטיקה ב.
עבור כל סטודנט נבדק מה הציון בסטטיסטיקה א ומה הציון בסטטיסטיקה ב.
להלן התוצאות שהתקבלו :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| א | 62 | 74 | 68 | 94 | 82 | 67 | 65 | 84 | 78 | 80 |
| ב | 70 | 80 | 70 | 90 | 77 | 67 | 80 | 86 | 79 | 82 |

א. בדקו ברמת מובהקות של 5% באמצעות מבחן הסימן.

ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם יוחלט לתת פקטור של 2 נקודות לכל הסטודנטים בשני המועדים?

**2)** מעוניינים לבדוק האם ההוצאות על "ג'אנק פוד" בקרב הסטודנטים רבות יותר בזמן הלימודים לעומת ימי החופשה.
נדגמו 15 סטודנטים מקריים, אצל 13 ההוצאות בתקופת הלימודים היו גבוהות יותר מימיי החופשה ואצל 2 נמוכות יותר.
מה מסקנתך בר"מ של 0.05?

**3)** מעוניינים לבדוק האם סם מסוים משפיע על לחץ הדם. נלקחו 24 אנשים אשר נמדד להם לחץ הדם לאחר מכן ניתן להם הסם ושוב מדדו להם את לחץ הדם. לחמישה אנשים לחץ הדם לא השתנה ל 15 אנשים לחץ הדם עלה וליתר לחץ הדם ירד אחרי לקיחת הסם. מה מסקנתכם ברמת מובהקות של 5%?

4) במדגם שנעשה על 15 משפחות השוו את רמת הביטחון העצמי של הבכור במשפחה לעומת הצעיר שבמשפחה. תוצאות המדגם הראו שאצל 7 משפחות רמת הביטחון העצמי של הבכור הייתה גבוהה יותר, אצל 3 משפחות רמת הביטחון העצמי של הצעיר הייתה גבוהה יותר ואצל 5 משפחות לא נמצא הבדל בין האחים מבחינת רמת הביטחון העצמי. טענת החוקר הייתה שבמשפחות לבכור ביטחון עצמי גבוה מזה של הצעיר במשפחה.

א. מהי רמת המובהקות המינימלית עבורה יוחלט לקבל את טענת החוקר?

ב. רמת הביטחון הוערכה על ידי פסיכולוג זוטר. פסיכולוג בכיר ביצע הערכה מחודשת וקבע שלמשפחה אחת במדגם הייתה הערכה שגויה: הפסיכולוג הזוטר קבע שלצעיר במשפחה יש ביטחון עצמי יותר גבוה למרות שלאח הבכור יש ביטחון עצמי יותר גבוה במשפחה הזו. מה יקרה לרמת המובהקות המינימלית שחושבה בשאלה הקודמת?

5) איזה מהטענות הבאות נכונות?

א. $n(+)+n(-)=n^*$

ב. $n(+)+n(-)=n$

ג. $n(+)=n(-)$

ד. $n(+)-n(-)=n^*$

## תשובות סופיות

1) א. לא נדחה את $H_o$. ב. לא תשתנה המסקנה.

2) נדחה את $H_o$.

3) נדחה את $H_o$.

4) א. 0.172 ב. תקטן.

5) א'.

## מבחן הסימן (שימוש בקירוב הנורמלי) – רקע

מבחן הסימן הוא מבחן שמשתמשים בו כאשר לפנינו מדגם מזווג ולא ניתן להניח שהמשתנה הנחקר מתפלג נורמלית.

גם אם המשתנה הנחקר מתפלג נורמלית ניתן לבצע את מבחן הסימן אבל מבחן T למדגמים מזווגים יהיה מבחן עם עוצמה גבוהה יותר ולכן יש לבצע אותו.
מבחן הסימן נחשב למבחן אפרמטרי – מבחנים אפרמטרים הינם כל המבחנים שאינם דורשים שהמשתנה הנחקר יתפלג נורמלית.
מבחן הסימן נקרא כך כיוון שהוא דורש הגדרת סימן לכל תצפית:

$(+)$ – אם מצב $X$ גבוה ממצב $Y$.

$(-)$ – אם מצב $Y$ גבוה ממצב $X$.

$(0)$ – אין הבדל בין המצבים.

במבחן הסימן נתעלם מהפרשים שהם 0, ולכן נסמן את מספר ההפרשים האפקטיביים (השונים מאפס) ב- $n^*$.

תחת השערת האפס נאמר שהסיכוי לקבל הפרש חיובי $(p+)$ שווה לסיכוי לקבל הפרש שלילי $(p-)$.

**השערות המבחן:**

$$H_0: P_+ = 0.5 \qquad H_1: P_+ \neq,<,> 0.5$$

**או**

$$H_0: P_- = 0.5 \qquad H_1: P_- \neq,<,> 0.5$$

נסמן ב $n(+)$ או ב $S_+$ את מספר התצפיות שקיבלו את הערך (+), ובאופן דומה:
נסמן ב $n(-)$ או ב $S_-$ את מספר התצפיות שקיבלו את הערך (–).

ניתן לומר שבהנחת השערת האפס: $n(+) \backslash n(-) \sim B(n^*, 0.5)$.

נחשב את PV על סמך תוצאות המדגם בעזרת ההתפלגות הבינומית כך שאם $PV \le \alpha$ נדחה את השערת האפס.

במבחן הסימן אין התייחסות לגודל הפער בתצפיות אלא רק את כיוון ההבדל.
אם התנאים לקירוב נורמלי מתקיימים ניתן להמיר את ההתפלגות הבינומית להתפלגות נורמלית:

אם: $n^* \cdot 0.5 \ge 5$

אז: $S_\pm \sim N\left(n^* \cdot \frac{1}{2}, n^* \cdot \frac{1}{4}\right)$.

**דוגמה (פתרון בהקלטה) :**

קופות החולים טוענות כי רכישת תרופות שאינן דורשות מרשם רופא, הינן זולות יותר אצלן מאשר מברשתות הפארם. דגמו 11 תרופות ובדקו את מחירן בבית המרקחת של קופות החולים וברשת הפארם. המחיר המוצג הינו עבור קפסולה בודדת. בדקו ברמת מובהקות של 5% באמצעות מבחן הסימן.

| שם התרופה | קופת חולים | פארם |
|---|---|---|
| אדוויל | 1.2 | 1.5 |
| אקמול | 2.6 | 2.6 |
| אופטלגין | 0.9 | 1.4 |
| פוסטינור | 3.5 | 3.2 |
| סטרפסיל | 1.1 | 1.4 |
| נורפן | 1.7 | 1.8 |
| לורסטין | 0.8 | 1.1 |
| קולדקס | 1.5 | 2 |
| אלרגיז | 2 | 2.8 |
| נוסידקס | 2 | 2.5 |
| קורמיר | 3 | 3.3 |

## שאלות

**1)** רוצים לבדוק את הטענה שהציונים במבחן בסטטיסטיקה ב גבוהים מאשר בסטטיסטיקה א. נלקחו 10 סטודנטים שסיימו את סטטיסטיקה ב. עבור כל סטודנט נבדק מה הציון בסטטיסטיקה א ומה הציון בסטטיסטיקה ב. להלן התוצאות שהתקבלו:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| א | 62 | 74 | 68 | 94 | 82 | 67 | 65 | 84 | 78 | 80 |
| ב | 70 | 80 | 70 | 90 | 77 | 67 | 80 | 86 | 79 | 82 |

א. בדקו ברמת מובהקות של 5% באמצעות מבחן הסימן.

ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם יוחלט לתת פקטור של 2 נקודות לכל הסטודנטים בשני המועדים?

**2)** מעוניינים לבדוק האם ההוצאות על "ג'אנק פוד" בקרב הסטודנטים רבות יותר בזמן הלימודים לעומת ימי החופשה. נדגמו 15 סטודנטים מקריים, אצל 13 ההוצאות בתקופת הלימודים היו גבוהות יותר מימיי החופשה ואצל 2 נמוכות יותר. מה מסקנתך בר"מ של 0.05?

**3)** מעוניינים לבדוק האם סם מסוים משפיע על לחץ הדם. נלקחו 24 אנשים אשר נמדד להם לחץ הדם לאחר מכן ניתן להם הסם ושוב מדדו להם את לחץ הדם. לחמישה אנשים לחץ הדם עלה וליתר לחץ הדם ירד אחרי לקיחת הסם. מה מסקנתכם ברמת מובהקות של 5%?

**4)** במדגם שנעשה על 15 משפחות השוו את רמת הביטחון העצמי של הבכור במשפחה לעומת הצעיר שבמשפחה. תוצאות המדגם הראו שאצל 7 משפחות רמת הביטחון העצמי של הבכור הייתה גבוהה יותר, אצל 3 משפחות רמת הביטחון העצמי של הצעיר הייתה גבוהה יותר ואצל 5 משפחות לא נמצא הבדל בין האחים מבחינת רמת הביטחון העצמי. טענת החוקר הייתה שבמשפחות לבכור ביטחון עצמי גבוה מזה של הצעיר במשפחה.

א. מהי רמת המובהקות המינימלית עבורה יוחלט לקבל את טענת החוקר?

ב. רמת הביטחון הוערכה על ידי פסיכולוג זוטר. פסיכולוג בכיר ביצע הערכה מחודשת וקבע שלמשפחה אחת במדגם הייתה הערכה שגויה: הפסיכולוג הזוטר קבע שלצעיר במשפחה יש ביטחון עצמי יותר גבוה למרות שלאח הבכור יש ביטחון עצמי יותר גבוה במשפחה הזו. מה יקרה לרמת המובהקות המינימלית שחושבה בשאלה הקודמת?

**5)** איזה מהטענות הבאות נכונות?

א. $n(+)+n(-)=n^*$.

ב. $n(+)+n(-)=n$.

ג. $n(+)=n(-)$.

ד. $n(+)-n(-)=n^*$.

## תשובות סופיות

1) א. לא נדחה $H_0$. ב. לא תשתנה המסקנה.
2) נדחה $H_0$.
3) נדחה $H_0$.
4) א. 0.172 ב. תקטן.
5) א׳.

## מבחן ווילקוקסון למדגמים מזווגים לפי שיטת המנייה – רקע

**מתי נשתמש במבחן זה?**

מבחן זה לא דורש הנחה של התפלגות נורמלית, אולם דורש ערכים מספריים המאפשרים חישוב הפרש בין ערכי $X$ לערכי $Y$. מבחן זה הוא הגרסה הלא פרמטרית למבחן T למדגמים מזווג. נשתמש במבחן זה שיש משתנה כמותי שאינו מתפלג נורמלית או שיש משתנה מסולם סדר והמדגם הוא מזווג.

**דוגמה (פתרון בהקלטה):**

שני קונדיטורים מתחרים על מקום עבודה. נתנו לשניהם להכין 8 מאפים שונים כאשר כל אחד מהמאפים נאפה על ידי שניהם. בסופו של דבר בעל הקונדיטוריה נתן ציון לכל אחד מהאופים בעבור כל אחד מהמאפים. להלן הציונים שהתקבלו, ורוצים לבדוק האם אופה א טוב יותר מאופה ב.

א. מהו המבחן הסטטיסטי המתאים?

ב. מהן השערות המחקר?

| אופה א | אופה ב |
|---|---|
| 10 | 9 |
| 9 | 8 |
| 7 | 7 |
| 8 | 9 |
| 9 | 6 |
| 10 | 6 |
| 7 | 5 |
| 8 | 4 |

**חישוב סטטיסטי המבחן:**

1. נחשב את ההפרשים $D_i$ לכל תצפית.
2. נוציא מהמדגם את כל התצפיות עם ההפרשים ששווים ל-0.
3. נדרג את ההפרשים הנותרים מהקטן אל הגדול בלי להתייחס לסימן ההפרש, כלומר מדרגים את הערכים המוחלטים של ההפרשים. הפרשים זהים מקבלים דרגה זהה שהיא הדרגה הממוצעת של המקומות שהם תופסים.
4. מסכמים את הדרגות של ההפרשים החיוביים $(W+)$ ואת הדרגות של ההפרשים השליליים $(W-)$.

**דוגמה (פתרון בהקלטה):**

חשבו את $W$ על סמך תוצאות המדגם.

| אופה א | אופה ב |
|---|---|
| 10 | 9 |
| 9 | 8 |
| 7 | 7 |
| 8 | 9 |
| 9 | 6 |
| 10 | 6 |
| 7 | 5 |
| 8 | 4 |

**חישוב מובהקות התוצאה:**

כדי לחשב את מובהקות התוצאה נצטרך לדון בהתפלגות של $W$ בהנחת השערת האפס. אם השערת האפס נכונה, לכל $D_i$ יש את אותו סיכוי לקבל ערך חיובי או ערך שלילי כלומר 2 אפשרויות. ל $n*$ דרגות יש $2n*$ אפשרויות שונות, שוות הסתברות. לפי רעיון זה ניתן יהיה לחשב את מובהקות התוצאה.

**דוגמה (פתרון בהקלטה):**

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

## שאלות

**1)** נדגמו 8 לקוחות שקיבלו שירות ממוקד טלפוני. לקוחות אלה נתבקשו לתת הערכה על יעלות השירות ועל האדיבות שבשירות.
הציונים ניתנו בסקאלה מ-1 (הערכה הנמוכה) עד 10 (הערכה הגבוהה ביותר).
להלן התוצאות שהתקבלו:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $X$ | הערכה על יעילות השירות | 7 | 8 | 4 | 3 | 2 | 5 | 7 | 5 |
| $Y$ | הערכה על אדיבות השירות | 8 | 7 | 7 | 6 | 8 | 10 | 7 | 4 |

בדקו ברמת מובהקות של 5% האם קיים הבדל בין הערכה על יעילות השירות להערכה על אדיבות השירות?

**2)** רוצים לבדוק האם תרופה חדשה להקלת כאבי ראש יעילה יותר מתרופה מוכרת. לצורך כך נלקח מדגם בן 9 אנשים, שנתבקשו להשתמש בתרופה החדשה ובתרופה המוכרת, ולהשוות את יעילותה של התרופה החדשה ליעילות התרופה המוכרת. האנשים במחקר היו צריכים לתת הערכה של יעילות בסקלה של מ-1 עד 100.
התוצאות שקיבל היו:

| הנבדק | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| תרופה חדשה | 95 | 90 | 100 | 80 | 75 | 81 | 69 | 100 | 86 |
| תרופה מוכרת | 80 | 76 | 65 | 49 | 75 | 70 | 50 | 60 | 60 |

האם התרופה החדשה משפרת את היעילות ביותר מ 10 נקודות?
בדקו ברמת מובהקות של 1%.

## תשובות סופיות

**1)** לא נדחה $H_0$.

**2)** לא נדחה $H_0$.

## מבחן ווילקוקסון למדגמים מזווגים (על ידי שימוש בקירוב הנורמלי) – רקע

**מתי נשתמש במבחן זה?**

מבחן זה לא דורש הנחה של התפלגות נורמלית, אולם דורש ערכים מספריים המאפשרים חישוב הפרש בין ערכי $X$ לערכי $Y$. מבחן זה הוא הגרסה הלא פרמטרית למבחן T למדגמים מזווג. נשתמש במבחן זה שיש משתנה כמותי שאינו מתפלג נורמלית או שיש משתנה מסולם סדר במדגם מזווג.

**דוגמה (פתרון בהקלטה):**

שני קונדיטורים מתחרים על מקום עבודה. נתנו לשניהם להכין 8 מאפים שונים כאשר כל אחד מהמאפים נאפה על ידי שניהם. בסופו של דבר בעל הקונדיטוריה נתן ציון לכל אחד מהאופים בעבור כל אחד מהמאפים.
להלן הציונים שהתקבלו, ורוצים לבדוק את הטענה שאופה א טוב יותר מאופה ב.

א. מהו המבחן הסטטיסטי המתאים?
ב. מהן השערות המחקר?

| אופה א | אופה ב |
|---|---|
| 10 | 9 |
| 9 | 8 |
| 7 | 7 |
| 8 | 9 |
| 9 | 6 |
| 10 | 6 |
| 7 | 5 |
| 8 | 4 |

**חישוב סטטיסטי המבחן:**

1. נחשב את ההפרשים $D_i$ לכל תצפית.
2. נוציא מהמדגם את כל התצפיות עם ההפרשים ששווים ל-0.
3. נדרג את ההפרשים הנותרים מהקטן אל הגדול בלי להתייחס לסימן ההפרש, כלומר מדרגים את הערכים המוחלטים של ההפרשים. הפרשים זהים מקבלים דרגה זהה שהיא הדרגה הממוצעת של המקומות שהם תופסים.
4. מסכמים את הדרגות של ההפרשים החיוביים $(R+)$ ואת הדרגות של ההפרשים השליליים $(R-)$.

**דוגמה (פתרון בהקלטה) :**

חשבו את $W$ על סמך תוצאות המדגם.

| אופה א | אופה ב |
|---|---|
| 10 | 9 |
| 9 | 8 |
| 7 | 7 |
| 8 | 9 |
| 9 | 6 |
| 10 | 6 |
| 7 | 5 |
| 8 | 4 |

מתקיים תמיד ש : $R_+ + R_- = \frac{n^*(n^*+1)}{2}$ .

כמו כן, ניתן להגיד שהתוחלת והשונות של הסטטיסטיים הללו הם :

$$\mu_{R_\pm} = \frac{n^*(n^*+1)}{4} \qquad \sigma^2_{R_\pm} = \frac{n^*(n^*+1)(2n^*+1)}{24}$$

אם המדגם מספיק גדול, ניתן לבצע קירוב נורמלי לסטטיסטיים אלה באופן הבא :

$$R_\pm \sim N\left(\frac{n^*(n^*+1)}{4}, \frac{n^*(n^*+1)(2n^*+1)}{24}\right) \qquad Z_\pm = \frac{R_\pm - \mu_{R_\pm}}{\sqrt{\sigma^2_{R_\pm}}} \sim N(0,1)$$

**דוגמה (פתרון בהקלטה) :**

א. מהי מובהקות התוצאה של מבחן זה?

ב. מה המסקנה ברמת מובהקות של 5%?

## שאלות

**1)** נדגמו 8 לקוחות שקיבלו שירות ממוקד טלפוני. לקוחות אלה נתבקשו לתת הערכה על יעילות השירות ועל האדיבות שבשירות. הציונים ניתנו בסקאלה מ-1 (הערכה הנמוכה) עד 10 (הערכה הגבוהה ביותר). להלן התוצאות שהתקבלו:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| X | הערכה על יעילות השירות | 7 | 8 | 4 | 3 | 2 | 5 | 7 | 5 |
| Y | הערכה על אדיבות השירות | 8 | 7 | 7 | 6 | 8 | 10 | 7 | 4 |

בדקו ברמת מובהקות של 5% האם קיים הבדל בין הערכה על יעילות השירות להערכה על אדיבות השירות?

**2)** סטודנטים נתבקשו לתת חוות דעתם על רמת הקושי של הקורס (סקאלה של 1-5 כאשר 5=קשה ביותר) ועל רמת הקושי של הבחינות באותה סקאלה. הסטודנטים טוענים שהבחינה הייתה ברמה גבוהה יותר מהרמה של הקורס. להלן תוצאות המדגם:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| קושי קורס-1 | 4 | 3 | 2 | 4 | 3 | 2 | 1 | 5 | 4 |
| קושי בחינה-2 | 4 | 4 | 4 | 3 | 5 | 5 | 5 | 3 | 2 |

בדקו ברמת מובהקות של 5% את טענת הסטודנטים.

**3)** רוצים לבדוק את הטענה שהציונים במבחן בסטטיסטיקה ב גבוהים מאשר בסטטיסטיקה א. נלקחו 10 סטודנטים שסיימו את סטטיסטיקה ב. עבור כל סטודנט נבדק מה הציון בסטטיסטיקה א ומה הציון בסטטיסטיקה ב. להלן התוצאות שהתקבלו:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| א | 62 | 74 | 68 | 94 | 82 | 67 | 65 | 84 | 78 | 80 |
| ב | 70 | 80 | 80 | 90 | 77 | 67 | 80 | 86 | 79 | 82 |

א. בדקו ברמת מובהקות של 5% באמצעות מבחן ווילקוקסון.

ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם יוחלט לתת פקטור של 2 נקודות לכל הסטודנטים בשני המועדים?

ג. כיצד הייתה משתנה התשובה אם מסתבר שנפלה טעות ועבור הסטודנט הראשון ברשימה יש להחליף בנתונים את הציון של סטטיסטיקה ב עם סטטיסטיקה א?

## תשובות סופיות

**1)** לא נדחה $H_0$.

**2)** לא נדחה $H_0$.

**3)** א. לא נדחה $H_0$. ב. לא משתנה. ג. לא משתנה.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 36 - מקדם המתאם ( מדד קשר ) הלינארי ומובהקותו

תוכן העניינים

# 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 קשר לגביהם 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) :**

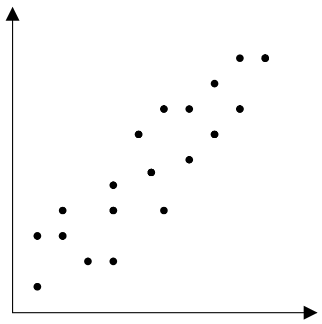
בבניין 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגרות בדירה. להלן התוצאות שהתקבלו :

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

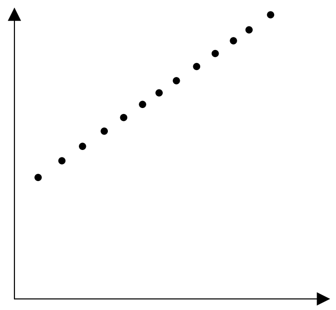
1) כמה תצפיות ישנן בדוגמה?
2) כמה משתנים ישנם בדוגמה, מי הם?
3) שרטטו לנתונים דיאגרמת פיזור.
4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

## דיאגרמות פיזור לקשר בין משתנים וניתוחם

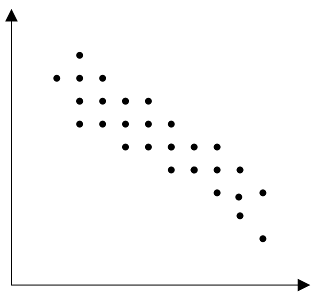
**קשר לנארי חיובי חלקי**



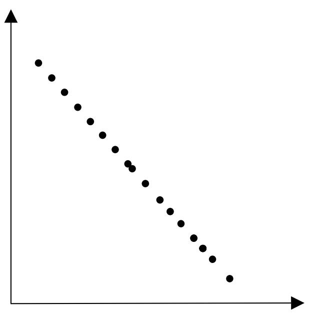
**קשר לנארי חיובי מלא**



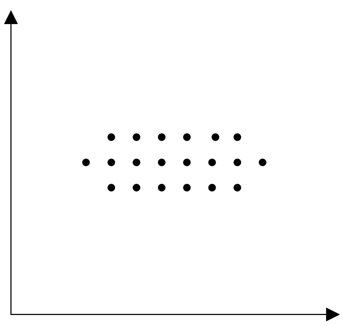
**קשר לינארי שלילי חלקי**



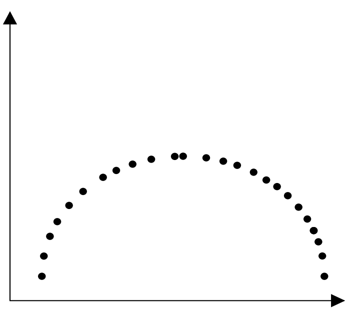
**קשר לינארי שלילי מלא**



**אין קשר**



**אין קשר לינארי**



**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 לנארי מלא בו השיפוע $a$ יהיה חיובי ואילו מתאם שלילי (מקדם מתאם-1) מלא אומר שקיים קשר לנארי 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 הלנארי, 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 $\rho$.

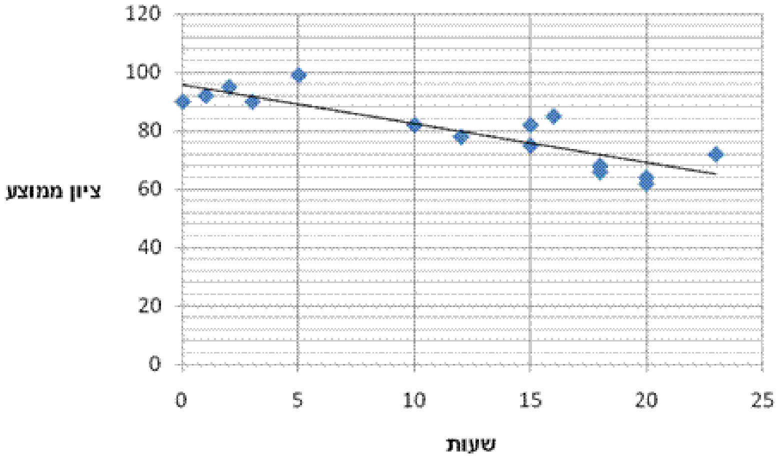
קיומו של מתאם בין שני משתנים אינו מצביע על סיבתיות בהכרח. למשל, אם נמצא מתאם חיובי בין כמות הסוכרזית שאדם אוכל לבין במשקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי, כלומר אם נחליף את $X$ ב- $Y$ התוצאה תהיה זהה.

**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 הליניארי בין ציוני מבחנים שונים שהתקבלו עבור תלמידים בכיתה מסוימת:

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

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?

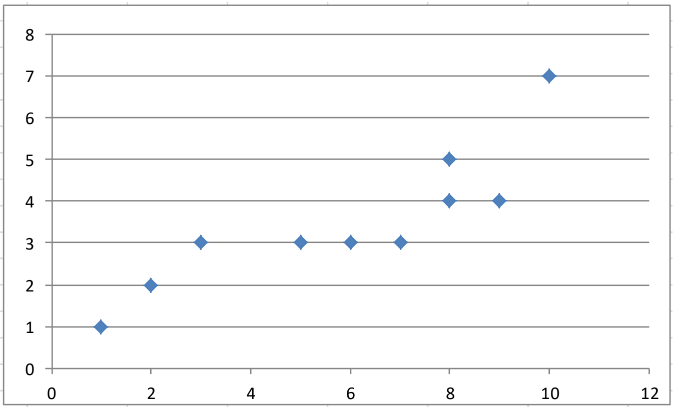
ד. מסתבר שבסופו של דבר נתנו פקטור של 5 נקודות לציון הסופי. כיצד הדבר היה משנה את מקדם המתאם של המדגם?

**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 פרנייט לכמות המשקעים במ"מ?

**5)** להלן דיאגרמת פיזור המראה קשר בין שני משנים :




א. השלימו : ניתן לראות שהקשר הוא ליניארי _________ (מלא\ 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 :

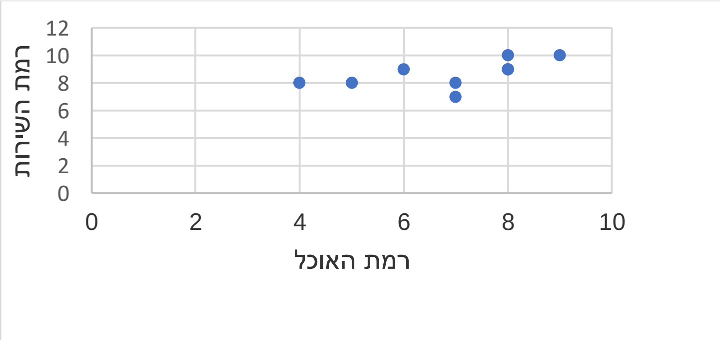
א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני בימים שנדגמו.

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.

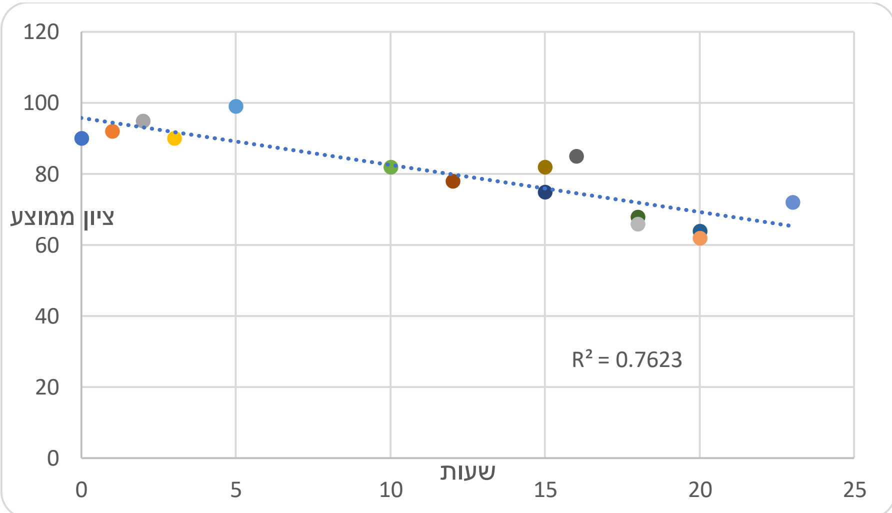
**7)** בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.




מה יהיה ערכו של מקדם המתאם ( $r$ )?

א. $r=-0.3$
ב. $r=0$
ג. $r=1.125$
ד. $r=0.593$

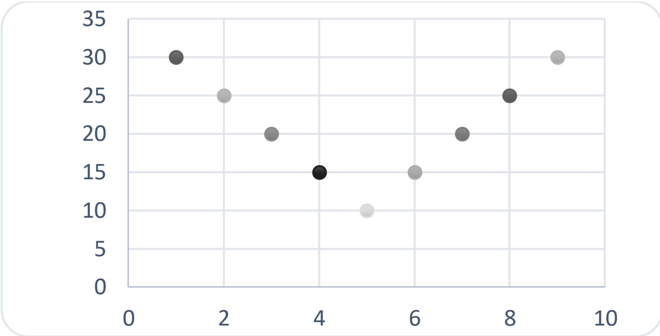
**8)** חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור.




מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?

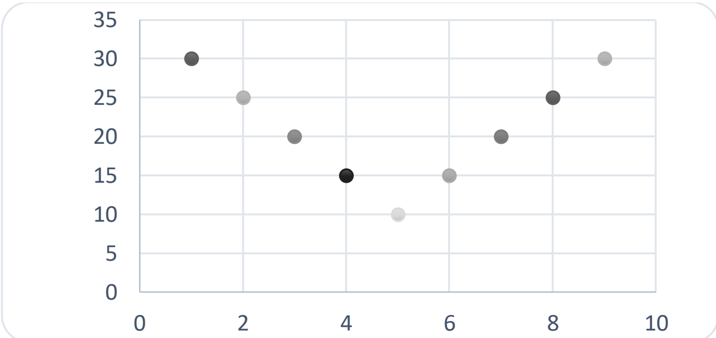
א. ככל שמבלים יותר הציון נוטה לרדת.
ב. אין קשר בין שעות הבילוי לציון.
ג. ככל שמבלים פחות הציון נוטה לרדת.
ד. ככל שהציון נוטה לרדת הסטודנט מבלה פחות.

**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 קווים 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?

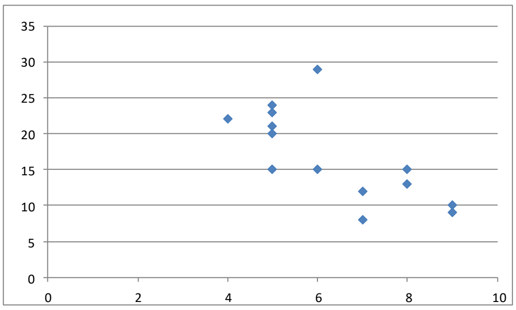
א. בתרשים מוצג הקשר בין שני משתנים.
ב. בתרשים מוצג הקשר בין 9 משתנים.
ג. בתרשים מוצג הקשר בין 10 משתנים.
ד. אין לדעת כמה משתנים מוצגים בתרשים.

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים:

$X$ - (משתנה בלתי תלוי בציר האופקי)

ו- $Y$ (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.

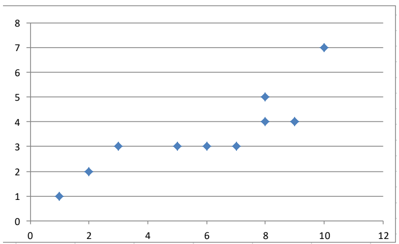
ב. עבור כל תצפית ערך של משתנה אחד שווה לערך של המשתנה השני.

ג. הקשר הליניארי 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.

| | ספורט | לשון | מתמטיקה |
|---|---|---|---|
| ספורט | 1 | -0.7 | 0.1 |
| לשון | -0.7 | 1 | 0.6 |
| מתמטיקה | -0.1 | 0.6 | 1 |

3) א. ב"ת- מס' שעות התרגול, תלוי- ציון. ב. ראה גרף בפתרון וידאו.
ג. קשר לינארי חיובי חלקי. ד. מקדם המתאם לא היה משתנה.

4) א. לעלות. ב. לא ישפיע על מקדם המתאם.

5) א. חלקי, חיובי. ב. קטן.

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:**

בבניין של 5 דירות בדקו את הנתונים הבאים:
$X$ -מס' חדרים בדירה. $Y$-מס' נפשות הגרות בדירה.
להלן התוצאות שהתקבלו:

| $Y$ | $X$ | מס' דירה |
|---|---|---|
| 2 | 3 | 1 |
| 2 | 2 | 2 |
| 3 | 4 | 3 |
| 3 | 3 | 4 |
| 4 | 5 | 5 |

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 לנארי מלא בו השיפוע $b$ יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לנארי מלא בו השיפוע $b$ שלילי (מקדם מתאם 1-).

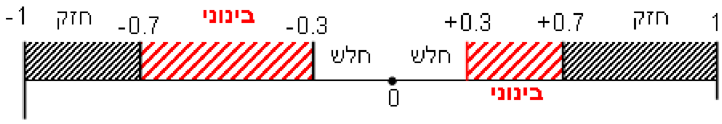
**מתאם חיובי חלקי:**

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה ליניארית שמקשרת את $X$ ל- $Y$ באופן מוחלט.

**מתאם שלילי חלקי:**

ככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה ליניארית 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.

שונות משותפת: $$COV(x,y)=\frac{\sum(x-\bar{x})(y-\bar{y})}{n}=\frac{\sum xy}{n}-\bar{x}\cdot\bar{y}$$

שונות של המשתנה $X$ : $$s_x^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}=\frac{\sum_{i=1}^{n}x_i^2}{n}-\bar{x}^2$$

שונות המשתנה $Y$ : $$S_Y^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}=\frac{\sum_{i=1}^{n}y_i^2}{n}-\bar{y}^2$$

מקדם המתאם הליניארי: $$r_{xy}=\frac{COV(x,y)}{S_x\cdot S_y}$$

GOOL בשביל התירגול

## 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 דיאגראמת 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?

**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:

| $X$ | $Y$ |
|---|---|
| 10 | 12 |
| 14 | 15 |
| 15 | 15 |
| 18 | 17 |
| 20 | 21 |

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:

$$\sum_{i=1}^{20} X_i = 240 \qquad \sum_{i=1}^{20} Y_i = 200$$

$$\sum_{i-1}^{20} (X_i - \bar{X})^2 = 76 \qquad \sum_{i-1}^{20} (Y_i - \bar{Y})^2 = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y - \bar{Y}) = 60.8$$

א. חשב את מדד הקשר הליניארי 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:

$$\sum_{i=1}^{20} X_i = 240 \qquad \sum_{i=1}^{20} Y_i = 200$$

$$\sum_{i-1}^{20} X_i^2 = 2960 \qquad \sum_{i=1}^{20} Y_i^2 = 2080$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

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.

א. 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.

ב. לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$ , $S_x = S_y = 1$.
לכן, מדד הקשר של פירסון יהיה 1.

ג. 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:

א. הדבר מעיד שהציונים בכתה היו שליליים.

ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.

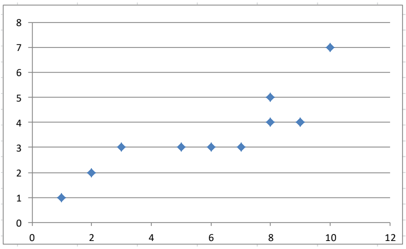
ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.

ד. אף אחת מהתשובות לא נכונה.

8) נלקחו 20 מוצרים ונבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו היום ערך הדולר היה-4.2₪).
מהו מקדם המתאם בין המחיר בדולר למחיר בש"ח?

א. 1
ב. 0
ג. 4.2
ד. לא ניתן לדעת.

9) להלן דיאגראמת 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 הליניארי – רקע

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

**השערת האפס:** תהיה שבאוכלוסייה לא קיים כלל קשר ליניארי בין שני המשתנים $H_0: \rho = 0$.

ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית.

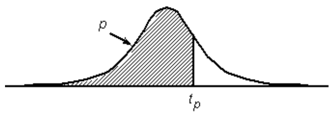
סטטיסטי המבחן: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t(n-2)$

סטטיסטי זה מתפלג $t$ עם $n-2$ דרגות חופש.

| | | | |
|---|---|---|---|
| השערת האפס: | $H_0: \rho = 0$ | $H_0: \rho = 0$ | $H_0: \rho = 0$ |
| השערת המחקר: | $H_1: \rho \neq 0$ | $H_1: \rho < 0$ | $H_1: \rho > 0$ |
| כלל ההכרעה:<br>אזור דחייה של השערת האפס | $t \geq t_{1-\alpha}$<br>או<br>$t \leq -t_{1-\alpha}$ | $t \leq -t_{1-\alpha}$ | $t \geq t_{1-\alpha}$ |

# 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 |
| ∞ | 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. בדוק 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

**1)** א. לא נדחה את $H_0$ . ב. לא תשתנה.

**2)** א. לא ב. לא נדחה את $H_0$ .

**3)** א. נדחה את $H_0$ . ב. לא ניתן לדעת.
ג. לפחות 0.005.

**4)** א. נדחה את $H_0$ . ב. $0.005 < P_V < 0.01$ .

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 37 - רגרסיה

תוכן העניינים

# מדדי קשר – רגרסיה ליניארית:

## רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבוי. לבנות קו ניבויים הנקרא גם קו רגרסיה המנבא משתנה אחד על סמך האחר. מדובר בקו שמנבא את $Y$ על סמך $X$.

השיטה למציאת הקו הנ"ל נקראת שיטת הריבועים הפחותים והקו המתקבל נקרא קו הרגרסיה או קו הניבויים או קו הריבועים הפחותים.

$a$ - נותן את ערך $Y$ כאשר $X$ הנו אפס על גבי קו הניבויים. הוא נקרא החותך של הקו.

$b$ - הוא שיפוע הקו נותן בכמה בעצם $Y$ משתנה כאשר $X$ גדל ביחידה אחת על גבי קו הניבויים.

להלן המשוואות למציאת הפרמטרים של קו הרגרסיה: $b = r\frac{S_r}{S_x}$ , $Y = bX + a$.

לצורך בניית קו ניבויים לניבוי $X$ על סמך $Y$ נצטרך לעדכן את הנוסחאות בהתאם.

## שאלות:

**1)** נסמן ב- $X$ את ההכנסה של משפחה באלפי ₪. נסמן ב- $Y$ את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} Y_i = 200 \ , \ \sum_{i=1}^{20} X_i = 240$$
$$\sum_{i=1}^{20}\left(X_i-\bar{X}\right)^2 = 76 \ , \ \sum_{i=1}^{20}\left(Y_i-\bar{Y}\right) = 76 \ .$$
$$\sum_{i=1}^{20}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right) = 60.8$$

א. חשבו את מדד הקשר הלינארי בין $X$ ל- $Y$. מיהו המשתנה התלוי?
ב. מצאו את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבירו את משמעות הפרמטרים של קו הרגרסיה.
ג. משפחת כהן הכניסה 15,000₪. מה ההוצאה הצפויה שלה?

**2)** נסמן ב- $X$ את ההשכלה של אדם בשנות לימוד.
נסמן ב- $Y$ את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:
$$S_X=2 \ , \ S_Y=5 \ , \ \bar{X}=14 \ , \ \bar{Y}=8 \ , \ \mathrm{COV}\left(X,Y\right)=7.5 \ .$$
א. חשבו את מדד הקשר של פירסון בין ההשכלה להכנסה.
ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?
ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000₪?

**3)** חוקר רצה לחקור את הקשר הקווי שבין הציון המבחן בסטטיסטיקה לבין מספר שעות ההכנה של הסטודנטים למבחן. במדגם של 100 סטודנטים שנבחנו בקורס נרשמו התוצאות הבאות: הציון הממוצע של הסטודנטים היה 65 עם סטיית תקן של 27. מספר שעות ההכנה הממוצע היה 30 עם סטיית תקן של 18. מקדם המתאם בין הציון לשעות ההכנה היה 0.8.
א. על פי משוואת הרגרסיה, שעת הכנה נוספת משפרת את ציון המבחן ב-?
ב. על פי משוואת הרגרסיה, תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון?
ג. מהו קו הרגרסיה לניבוי הציון לפי שעות ההכנה?

**4)** נתונים 2 משתנים $X$ ו- $Y$. כמו כן נתון: $\bar{X}=1.5$ , $S_x=S_y=4$ ,
וכן שקו הרגרסיה של $Y$ על בסיס $X$ הינו: $Y=-0.2X+0.5$ .
חשבו מהו מקדם המתאם בין $X$ ל- $Y$.

## תשובות סופיות:

1) א. 0.8. ב. $Y = 0.8X + 0.4$. ג. 12.4 אלפי ₪.
2) א. 0.75. ב. 4.25 אלפי ₪. ג. 14.6 שנים.
3) א. 1.2. ב. 29. ג. $Y = 1.2X + 29$.
4) -0.2.

# מבוא לסטטיסטיקה למדעי המחשב

## פרק 38 - מדדי קשר-רגרסיה -שונות מוסברת ושונות לא מוסברת

תוכן העניינים

# מדדי קשר – רגרסיה – שונות מוסברת ושונות לא מוסברת:

**רקע:**

המטרה ברגרסיה היא להסביר את השונות של המשתנה התלוי.
למשל, להסביר את השונות של המשכורת באמצעות הוותק או להסביר את השוני בציונים באמצעות כמות החיסורים.

$r^2$ - החלק מהשונות של המשתנה התלוי מוסבר. השונות המוסברת נקראת גם שונות ניבויים. השונות הלא מוסברת נקראת גם שונות טעויות.

## שאלות:

**1)** נמצא קשר חיובי בעוצמה של 0.7 בין שטח דירה למחירה. כמו כן, נתון שסטיית התקן של מחירי הדירות הינה 200.

א. איזה אחוז מהשונות של מחירי הדירות מוסבר על ידי שטח הדירה?

ב. איזה אחוז מהשונות של מחירי הדירות לא מוסבר על ידי שטח הדירה?

ג. מהי השונות המוסברות ומהי השונות הלא מוסברת של מחירי הדירות?

**2)** להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו!

א. אם שונות הטעויות שווה ל-0 (השונות הלא מוסברת) אז מקדם המתאם של פירסון יהיה 1.

ב. אם מקדם המתאם של פירסון בין שני משתנים הוא 1 אזי שונות הטעויות (השונות הלא מוסברת) תהיה 0.

ג. אם השונות המשותפת של $X$ ושל $Y$ היא 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

## שאלות רב-ברירה:

**3)** בקשר בין שני משתנים התקבל: $r^2 = 0.64$, לכן:

א. ללא יוצא מן הכלל ככל שערכי משתנה אחד עולה השני יעלה.

ב. 64% מהשונות של משתנה אחד מוסבר על ידי המשתנה השני.

ג. הקשר בין שני המשתנים הוא בעוצמה של 0.64.

ד. כל התשובות נכונות.

**4)** אם מגדילים את $r^2$, ניתן לומר כי:

א. אחוז השונות המוסברת יקטן.

ב. אחוז השונות המוסברת יגדל.

ג. אחוז השונות המוסברת יישאר ללא שינוי.

ד. סטיית התקן משתנה.

ה. לא ניתן לדעת.

**5)** בקורס מבוא לכלכלה ניתנו במשך השנה שני מבחנים: מבחן בסוף סמסטר א׳ $X$ ומבחן בסוף סמסטר ב׳ $Y$. כאשר בנו את קו הרגרסיה של הציון במבחן סוף סמסטר ב׳ לפי הציון במבחן סוף סמסטר א התקבלה שונות טעויות של 80, ושונות ניבויים של 20.
לפי נתונים אלו, מקדם המתאם בין הציון במבחן סוף סמסטר א׳ לבין הציון במבחן סוף סמסטר ב׳ הוא:

א. 0.44.

ב. 0.44 -.

ג. עוצמת ההקשר הלינארי היא 0.44, אך אין אפשרות לדעת את סימנה.

ד. אין אפשרות לחשב את מקדם המתאם.

ה. 0.35.

## תשובות סופיות:

**1)** א. 49%. ב. 51%.
ג. שונות מוסברת: 19,600, שונות לא מוסברת: 20,400.

**2)** א. לא נכון. ב. נכון. ג. נכון.

**3)** ב׳.

**4)** ב׳.

**5)** ג׳.