ביוסטטיסטיקה

תוכן העניינים

תוכן העניינים

ביוסטטיסטיקה

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 ביניהם 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 "רווהה".
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.

ביוסטטיסטיקה

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 שווה 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.

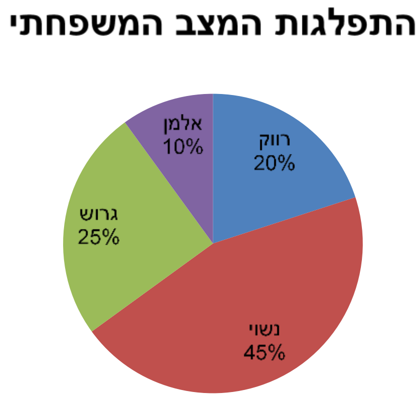
דוגמה:

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.



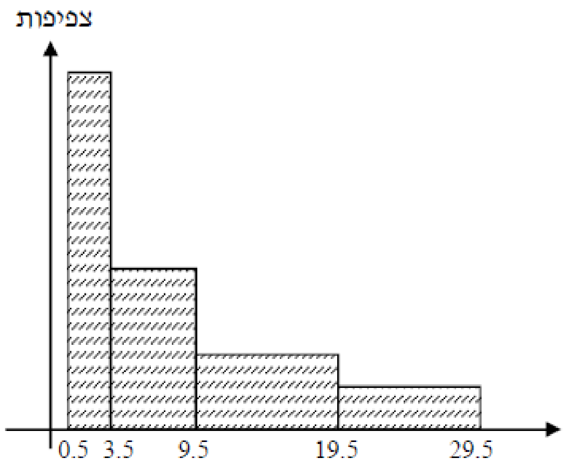
דיאגרמת מקלות:

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.

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



פוליגון – מצולעון:

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

צורות התפלגות נפוצות:

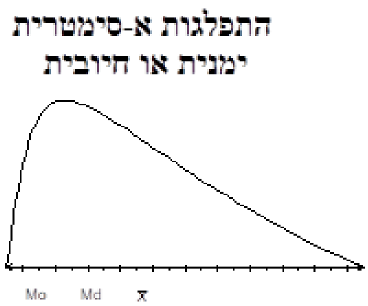
התפלגות סימטרית פעמונית

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.

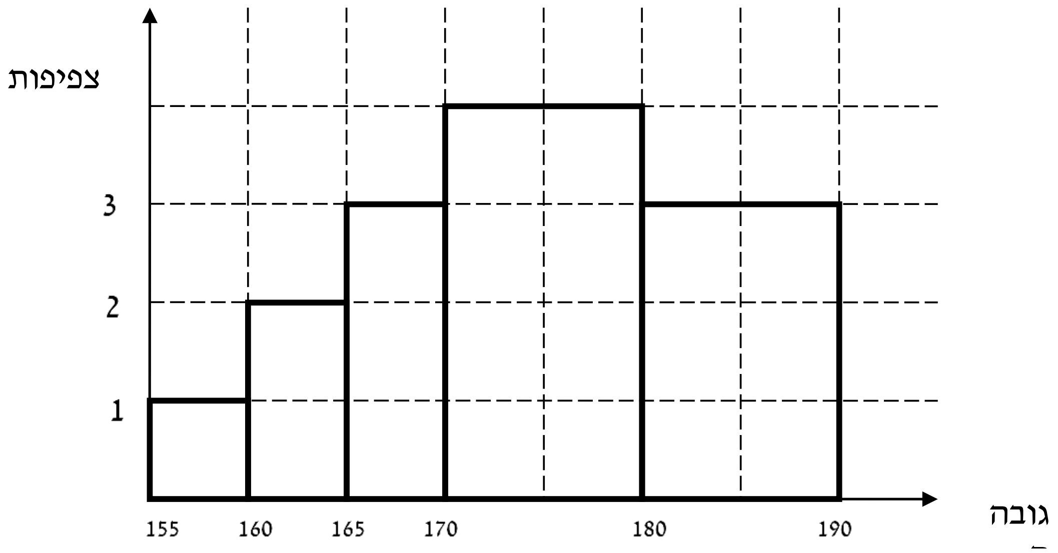
א. מהו המשתנה? האם הוא בדיד או רציף?

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?

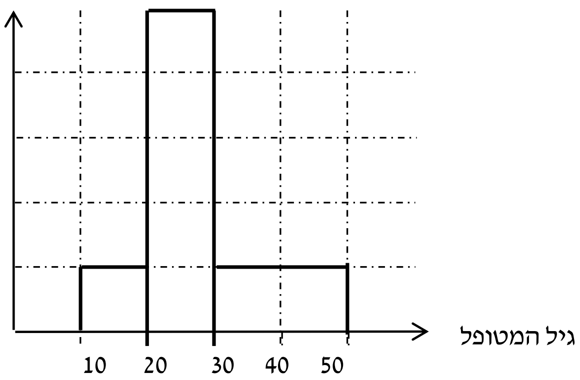
6) להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

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: ה. 62.5%.

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

ביוסטטיסטיקה

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:

i	X_i
1	5
2	0
3	1
4	3
5	2

(הסבר מלא מופיע בסרטונים באתר).

שאלות:

1) בבניין 5 דירות. לכל דירה רשמו את מספר החדרים שיש בדירה $\left(X\right)$,
ומספר הנפשות החיות בדירה $\left(Y\right)$ 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. $\sum\left(X_i\right)\sum\left(Y_i\right)$.

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

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

א. $\sum_{i=1}^{4} y_i$.

ב. $\sum_{i=1}^{6} a$.

ג. $\sum_{i=1}^{6} x_i y_i$.

ד. $\sum_{i=1}^{6} (x_i + y_i)$.

ה. $\sum_{i=1}^{6} x_i + a$.

3) קבעו לכל זהות האם היא נכונה:

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

ב. $\sum_{i=1}^{n} a = a \cdot n$.

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

תשובות סופיות:

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

ביוסטטיסטיקה

פרק 4 - סטטיסטיקה תיאורית- מדדי מרכז

תוכן העניינים

ביוסטטיסטיקה

פרק 5 - סטטיסטיקה תיאורית- מדדי פיזור

תוכן העניינים

ביוסטטיסטיקה

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		

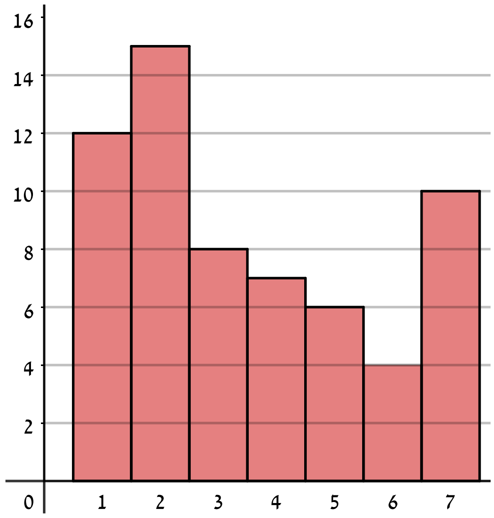
שאלות:

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

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

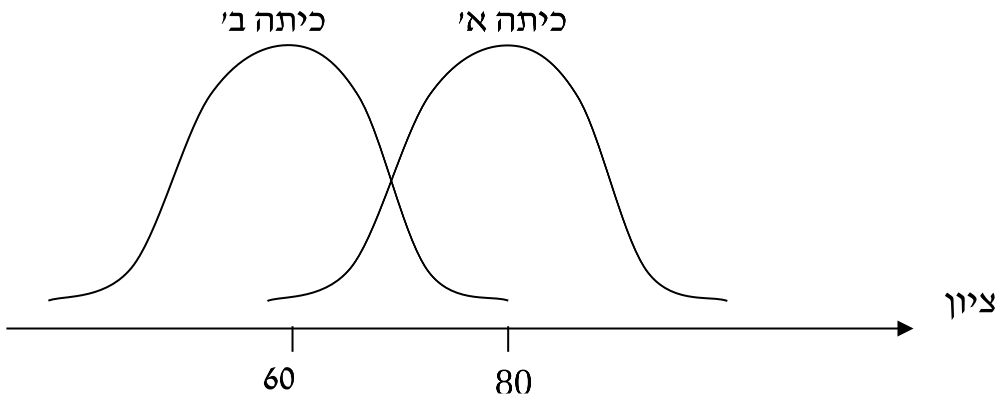
מהו הטווח הבין-רבעוני של מספר המכוניות למשפחה בישוב "הגורן"?

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 "רווהה" 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'.

ביוסטטיסטיקה

פרק 7 - 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.

ביוסטטיסטיקה

פרק 8 - סטטיסטיקה תיאורית- תרשים קופסא

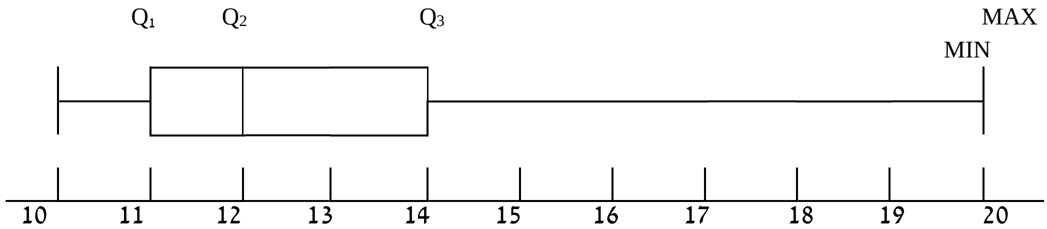
תוכן העניינים

סטטיסטיקה תיאורית – תרשים קופסא (Boxplot):

רקע:

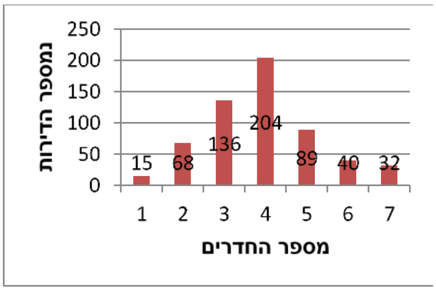
תרשים קופסא הינו תרשים שבעזרתו ניתן לבחון:

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).



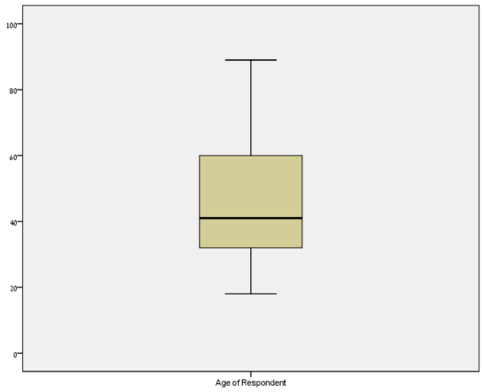
שאלות:

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.

ביוסטטיסטיקה

פרק 9 - פלטים בסטטיסטיקה תיאורית

תוכן העניינים

סטטיסטיקה תיאורית – ניתוח פלטים:

שאלות:

1) להלן פלט על התפלגות הגילאים באוכלוסייה מסוימת:

		Statistic
Age of Respondent	Mean	45.63
	Median	41.00
	Variance	317.140
	Std. Deviation	a
	Minimum	18
	Maximum	b
	Range	71
	Interquartile Range	28

א. מצאו את הערכים בטבלה המסומנים ב- a ו- b.
ב. נתון שההתפלגות היא אסימטרית, האם היא נוטה ימינה או שמאלה?

2) להלן התפלגות ההשכלה של העובדים בחברת "מתאר":

		Statistic
years of education	Mean	?
	Median	12.0000
	Variance	?
	Std. Deviation	2.54786
	Minimum	?
	Maximum	?
	Range	?
	Interquartile Range	?

מלאו את הערכים המסומנים בסימני שאלה.

תשובות סופיות:

1) א. $b=89$, $a=17.81$.   ב. אסימטרית ימנית.

2) ממוצע: 11.909, שונות: 6.492, טווח: 10, טב״ר: 3.

ביוסטטיסטיקה

פרק 10 - סטטיסטיקה תיאורית שאלות אמריקאיות

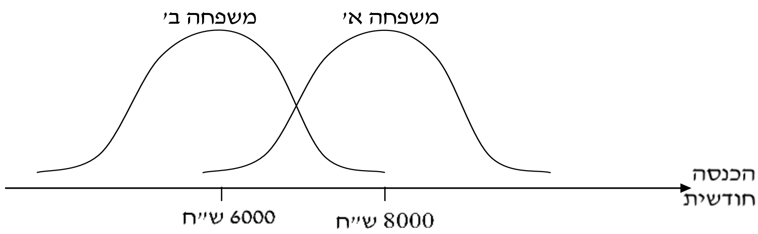
תוכן העניינים

סטטיסטיקה תיאורית – שאלות אמריקאיות:

שאלות:

שאלות 1-3 מתייחסות לקטע הבא:

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



1) לאיזו משפחה הכנסה שכיחה גבוהה יותר?
א. משפחה א'.
ב. משפחה ב'.
ג. לשתיהן אותה הכנסה שכיחה.
ד. לא ניתן לדעת – אין מספיק נתונים.

2) באיזו משפחה ההכנסה החציונית שווה להכנסה הממוצעת?
א. משפחה א'.
ב. משפחה ב'.
ג. בשתיהן ההכנסה החציונית שווה להכנסה הממוצעת.
ד. לא ניתן לדעת – אין מספיק נתונים.

3) באיזו משפחה סטית התקן של ההכנסה החודשית גבוהה יותר?
א. משפחה א'.
ב. משפחה ב'.
ג. לשתיהן אותה סטית תקן.
ד. לא ניתן לדעת – אין מספיק נתונים.

הנתונים הבאים מתייחסים לשאלות 4-6:

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

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

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

5) חציון הנתונים הוא:
א. 2.
ב. 1.5.
ג. 25.5.
ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

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

7) בהתפלגות אסימטרית ימנית של משתנה כמותי רציף, הערך המתאים למאון ה-30, ציון התקן שלו הוא בהכרח:
א. שלילי.
ב. חיובי.
ג. אפס.
ד. לא ניתן לדעת ללא הנתונים.

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

9) הוספת גודל קבוע לכל תצפיות סדרת נתונים:
א. תגדיל את סטיית התקן.
ב. תקטין את סטיית התקן.
ג. לא תשנה את סטיית התקן.
ד. לא ניתן לדעת.

הנתונים הבאים מתייחסים לשאלות 10-11:

להלן נתונים על ציוני תלמידים שנבחנו במועדים שונים בסטטיסטיקה:

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

10) התלמיד הטוב ביותר ביחס לנבחנים באותו מועד בו נבחן הוא:
א. מיטב.
ב. צבי.
ג. לובה.
ד. שרית.
ה. סטף.

11) פנינה נבחנה עם סטף וציון התקן שלה שווה לציון התקן של שרית לכן ציונה הוא:
א. 80.55.
ב. 65.
ג. 80.
ד. 81.66.

הנתונים הבאים מתייחסים לשאלות 12-15:

בבדיקת פתע של משרד הבריאות במפעל שוקולד, נמצא ש:

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

12) מהו החציון של מספר הפגומים בקופסא:
א. 1.
ב. 2.
ג. 4.
ד. לא ניתן לדעת.

13) מהו הרבעון התחתון של מספר הפגומים בקופסא?
א. 1.
ב. 2.
ג. 3.
ד. 4.
ה. לא ניתן לדעת.

14) מספר הפגומים בקופסא הוא משתנה:
א. סדר.
ב. שמי.
ג. כמותי בדיד.
ד. כמותי רציף.

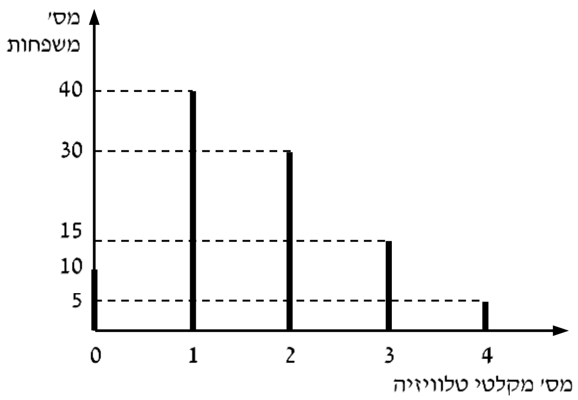
15) השכיח של מספר הפגומים בקופסא:
א. 63.
ב. 1.
ג. 200.
ד. לא ניתן לדעת.

16) ביחס לציר המספרים, רוב הערכים בהתפלגות א-סימטרית ימנית נמצאים:
א. בערכים הגבוהים.
ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
ג. בערכים הנמוכים.
ד. לא ניתן לדעת.
ה. אף לא תשובה מהנ"ל נכונה.

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

הנתונים הבאים מתייחסים לשאלות 18 עד 22:

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



18) המשתנה הנחקר כאן הוא:
א. משתנה שמי.
ב. משתנה מסולם סדר.
ג. משתנה כמותי בדיד.
ד. משתנה כמותי רציף.

19) הטווח של ההתפלגות הוא:
א. 35.
ב. 4.
ג. 3.
ד. 2.

20) ממוצע מספר מקלטי הטלוויזיה למשפחה הוא:
א. 1.65.
ב. 1.5.
ג. 1.
ד. 2.

21) השכיח של התפלגות זו היא:
א. 40.
ב. 1.5.
ג. 1.
ד. 2.

22) מסתבר שיש בין 2 ל-5 משפחות נוספות שאין להם מקלטי טלוויזיה ויש לצרף את המשפחות הללו להתפלגות. כיצד הנתון זה ישפיע על סטיית התקן?
א. יקטין אותו.
ב. יגדיל אותו.
ג. לא ישנה אותו.
ד. אין לדעת.

תשובות סופיות:

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

ביוסטטיסטיקה

פרק 11 - יסודות ההסתברות

תוכן העניינים

הגדרות יסודיות:

רקע:

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

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

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

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

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

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

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

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

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

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

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

דוגמה:

להלן התפלגות הציונים בכיתה מסוימת:

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

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

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

הסתברות למאורע משלים: הסתברות לקבוצת המשלים של המאורע ביחס למרחב המדגם: $P(\bar{A})=1-P(A)$. למשל, בדוגמה הקודמת הסיכוי לעבור את הבחינה יכול להיות מחושב לפי הסיכוי להיכשל: $P(A)=1-\frac{2}{25}=\frac{23}{25}$.

שאלות:

1) מהאותיות E, F ו-G יש ליצור מילה בת 2 אותיות, לא בהכרח בת משמעות.
א. הרכיבו את כל המילים האפשריות.
ב. רשמו את המקרים למאורע:
i. במילה נמצאת האות E.
ii. במילה האותיות שונות.
ג. רשמו את המקרים למאורע $\bar{A}$.

2) מטילים זוג קוביות.
א. רשמו את מרחב המדגם של הניסוי. האם מרחב המדגם אחיד?
ב. רשמו את כל האפשרויות לאירועים הבאים:
i. סכום התוצאות 7.
ii. מכפלת התוצאות 12.
ג. חשבו את הסיכויים לאירועים שהוגדרו בסעיף ב'.

3) נבחר באקראי ספרה מבין הספרות 0-9.
א. מה ההסברות שהספרה שנבחרה גדולה מ-5?
ב. מה ההסתברות שהספרה שנבחרה היא לכל היותר 3?
ג. מה ההסתברות שהספרה שנבחרה היא אי זוגית?

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

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

נבחרה משפחה באקראי מהישוב.
א. מה ההסתברות שאין מקלטים למשפחה?
ב. מה ההסתברות שיש מקלטים למשפחה?
ג. מה ההסתברות שיש לפחות 3 מקלטים למשפחה?

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

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

נבחרה משפחה אקראית מן הישוב.
א. מה ההסתברות שאין לה מכוניות?
ב. מה ההסתברות שבבעלות המשפחה לפחות 3 מכוניות?
ג. מה הסיכוי שבבעלותה פחות מ-3 מכוניות?

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

תשובות סופיות:

1) א. $\Omega=\{EE,EF,EG,FE,FF,FG,GE,GF,GG\}$.
ב. $A=\{EE,EF,EG,FE,GE\},\ B\{EF,EG,FE,FG,GE,GF\}$.
ג. $\bar{A}=\{FF,FG,GF,GG\}$.

2) א. $\Omega=\begin{Bmatrix}(1,1) & (2,1) & (3,1) & (5,1) & (4,1) & (6,1)\\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2)\\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3)\\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4)\\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5)\\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6)\end{Bmatrix}$.
ב. $A=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}\ ,\ C=\{(2,6),(3,4),(4,3),(6,2)\}$.
ג. הסיכוי ל-A: $\frac{1}{6}$. הסיכוי ל-B: $\frac{1}{9}$.

3) א. 0.4. ב. 0.4. ג. 0.5.
4) א. 0.22. ב. 0.78. ג. 0.32.
5) א. 0.1. ב. 0.2. ג. 0.8.
6) א. $\Omega=\{PPP,PPE,PEP,EPP,PEE,EPE,EEP,EEE\}$.
ב. $A=\{PPE,PEP,EPP\}\ ,\ D=\{PPP,PPE,PEP,EPP,PEE,EPE,EEP\}$.
ג. $\bar{D}=\{EEE\}$.
ד. $\frac{1}{8}$.

ביוסטטיסטיקה

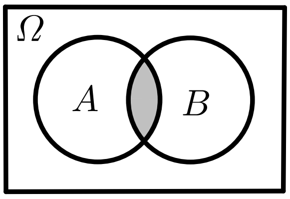
פרק 12 - פעולות בין מאורעות (חיתוך ואיחוד) - מאורעות זרים ומכילים

תוכן העניינים

פעולות בין מאורעות (חיתוך ואיחוד) – מאורעות זרים ומכילים:

רקע:

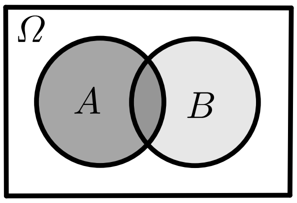
פעולת חיתוך:



נותנת את המשותף בין המאורעות הנחתכים.
חיתוך בין המאורע A למאורע B יסומן כך: $A\cap B$.
מדובר בתוצאות שנמצאות ב-A וגם ב-B.

דוגמה:
בהטלת קובייה, למשל, האפשרויות לקבל לפחות 5 הן: $A=\{5,6\}$.
האפשרויות לקבל תוצאה זוגית הן: $B=\{2,4,6\}$.
החיתוך שביניהם הוא: $A\cap B=\{6\}$.

פעולת איחוד:

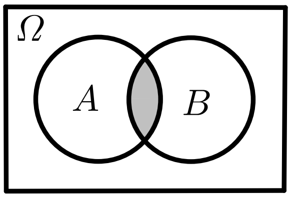


נותנת את כל האפשריות שנמצאות לפחות באחת מהמאורעות, ומסומנת: $A\cup B$.
הפעולה נותנת את אשר נמצא ב-A או B.
כלומר, לפחות אחד מהמאורעות קורה.

דוגמה:
בהטלת קובייה האפשרויות לקבל לפחות 5 הן: $A=\{5,6\}$.
האפשרויות לקבל תוצאה זוגית: $B=\{2,4,6\}$.
האפשרויות לקבל לפחות 5 וגם תוצאה זוגית: $A\cup B=\{2,4,5,6\}$.

דוגמה (הפתרון נמצא בהקלטה):
סטודנט ניגש בסמסטר לשני מבחנים. מבחן בסטטיסטיקה ומבחן בכלכלה. ההסתברות שלו לעבור את המבחן בסטטיסטיקה הוא 0.9, ההסתברות שלו לעבור את המבחן בכלכלה הוא 0.8 וההסתברות לעבור את המבחן בסטטיסטיקה ובכלכלה היא 0.75.
מה ההסתברות שלו לעבור את המבחן בסטטיסטיקה בלבד?
מה ההסתברות שלו להיכשל בשני המבחנים?
מה ההסתברות לעבור לפחות מבחן אחד?

נוסחת החיבור לשני מאורעות:



ההסתברות של איחוד מאורעות תחושב ע"י הקשר הבא:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

חוקי דה מורגן לשני מאורעות:

$$\overline{A\bigcup B}=\overline{A}\bigcap\overline{B}$$

$$\overline{A\bigcap B}=\overline{A}\bigcup\overline{B}$$

$$P(A\bigcap B)=1-P\left(\overline{A}\bigcup\overline{B}\right)$$

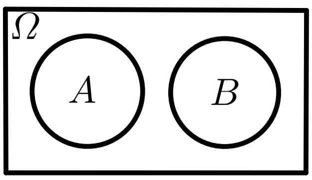
$$P(A\bigcup B)=1-P\left(\overline{A}\bigcap\overline{B}\right)$$

שיטת ריבוע הקסם:

השיטה רלבנטית רק אם יש שני מאורעות במקביל בדומה לתרגיל הקודם:

	$\overline{A}$	A	
B	$P\left(\overline{A}\cap B\right)$	$P(A\cap B)$	$P(B)$
$\overline{B}$	$P\left(\overline{A}\cap\overline{B}\right)$	$P\left(A\cap\overline{B}\right)$	$P\left(\overline{B}\right)$
	$P\left(\overline{A}\right)$	$P(A)$	1

מאורעות זרים:



מאורעות זרים הם כאשר אין להם אף איבר משותף:
$A\cap B=\{\ \}$. כלומר, הם לא יכולים להתרחש בו זמנית.

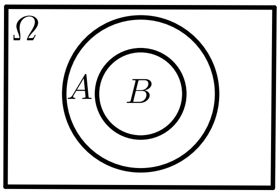
ההסתברות של חיתוך המאורעות היא אפס: $P(A\cap B)=0$.

ההסתברות של איחוד המאורעות תחושב: $P(A\cup B)=P(A)+P(B)$.

דוגמה:

בהטלת קובייה, האפשרויות לקבל לפחות 5 הן: $A=\{5,6\}$ והאפשרות לקבל 3 היא: $B=\{3\}$, ולכן החיתוך ביניהם הוא אפס, כלומר: $A\cap B=\{\ \}$.

מאורעות מוכלים:



נתונים שני מאורעות A ו-B, השונים מאפס.
נאמר שהמאורע B מוכל במאורע A אם כל איברי המאורע B כלולים במאורע A ונרשום: $B \subset A$.

מאורע A מכיל את מאורע B כל התוצאות שנמצאות ב-B מוכלות בתוך מאורע A.

קשר זה מסומן באופן הבא: $B \subset A$.

$A \cap B = B \quad P(A \cap B) = P(B)$

$A \cup B = A \quad P(A \cup B) = P(A)$

למשל: $A = \{2,4,6\}$, $B = \{2,4\}$.

שאלות:

1) מהאותיות E, F ו-G יוצרים מילה בת 2 אותיות – לא בהכרח בת משמעות.
נגדיר את המאורעות הבאים:
A - במילה נמצאת האות E.
B - במילה אותיות שונות.
א. רשמו את כל האפשרויות לחיתוך A עם B.
ב. רשמו את כל האפשרויות לאיחוד של A עם B.

2) תלמיד ניגש בסמסטר לשני מבחנים מבחן בכלכלה ומבחן בסטטיסטיקה.
נגדיר את המאורעות הבאים:
A - לעבור את המבחן בסטטיסטיקה.
B - לעבור את המבחן בכלכלה.
היעזרו בפעולות חיתוך, איחוד ומשלים בלבד כדי להגדיר את המאורעות הבאים וסמנו בדיאגראמת וון את השטח המתאים:
א. התלמיד עבר רק את המבחן בכלכלה.
ב. התלמיד עבר רק את המבחן בסטטיסטיקה.
ג. התלמיד עבר את שני המבחנים.
ד. התלמיד עבר לפחות מבחן אחד.
ה. התלמיד נכשל בשני המבחנים.
ו. התלמיד נכשל בכלכלה.

3) נתבקשתם לבחור ספרה באקראי. נגדיר את A להיות הספרה שנבחרה היא זוגית. נגדיר את B להיות הספרה שנבחרה קטנה מ-5.
א. רשמו את כל התוצאות למאורעות הבאים:
A , B , $\overline{B}$, $A\cap B$, $A\cup B$.
ב. חשבו את ההסתברויות לכל המאורעות מהסעיף הקודם.

4) נסמן ב-Ω את מרחב המדגם וב-ϕ קבוצה ריקה.
נתון כי A הינו מאורע בתוך מרחב המדגם.
להלן מוגדרים מאורעות שפתרונם הוא Ω או ϕ או A.
קבעו עבור כל מאורע מה הפתרון שלו:
$\overline{\overline{A}}$, $A\cap\phi$, $A\cup\phi$, $A\cap\Omega$, $A\cup\Omega$, $A\cap\overline{A}$, $\overline{\phi}$, $A\cup\overline{A}$.

5) הוגדרו המאורעות הבאים:

A - אדם שגובהו מעל 1.7 מטר

B - אדם שגובהו מתחת ל-1.8 מטר.

קבעו את גובהם של האנשים הבאים:

א. $A \cap B$.

ב. $A \cup B$.

ג. $\overline{A} \cap B$.

ד. $\overline{A} \cup \overline{B}$.

ה. $\overline{\overline{A}}$.

6) נגדיר את המאורעות הבאים:

A - אדם דובר עברית.

B - אדם דובר ערבית.

C - אדם דובר אנגלית.

השתמשו בפעולות איחוד, חיתוך והשלמה לתיאור המאורעות הבאים:

א. אדם דובר את כל שלוש השפות.

ב. אדם דובר רק עברית.

ג. אדם דובר לפחות שפה אחת מתוך השפות הללו.

ד. אדם אינו דובר אנגלית.

ה. קבוצת התלמידים שדוברים שתי שפות בדיוק (מהשפות הנ"ל).

7) שתי מפלגות רצות לכנסת הבאה. מפלגת "גדר" תעבור את אחוז החסימה בהסתברות של 0.08 ומפלגת "עתיד" תעבור את אחוז החסימה בהסתברות של 0.20. בהסתברות של 76% שתי המפלגות לא תעבורנה את אחוז החסימה.

א. מה ההסתברות שלפחות אחת מהמפלגות תעבור את אחוז החסימה?

ב. מה ההסתברות ששתי המפלגות תעבורנה את אחוז החסימה?

ג. מה ההסתברות שרק מפלגות "עתיד" תעבור את אחוז החסימה?

8) במקום עבודה מסוים 40% מהעובדים הם גברים. כמו כן, 20% מהעובדים הם אקדמאים. 10% מהעובדים הינן נשים אקדמאיות.

א. איזה אחוז מהעובדים הם גברים אקדמאיים?

ב. איזה אחוז מהעובדים הם גברים או אקדמאיים?

ג. איזה אחוז מהעובדים הם נשים לא אקדמאיות?

9) הסיכוי של מניה A לעלות הנו 0.5 ביום מסוים והסיכוי של מניה B לעלות ביום מסוים הנו 0.4. בסיכוי של 0.7 לפחות אחת מהמניות תעלה ביום מסוים.
חשבו את ההסתברויות הבאות לגבי שתי המניות הללו ביום מסוים:

א. ששתי המניות תעלנה.
ב. שאף אחת מהמניות לא תעלנה.
ג. שמניה A בלבד תעלה.

10) מטילים זוג קוביות, אדומה ושחורה. נגדיר את המאורעות הבאים:
A - בקובייה האדומה התקבלה התוצאה 4 ובשחורה 2.
B - סכום התוצאות משתי הקוביות הוא 6.
C - מכפלת התוצאות בשתי הקוביות היא 10.

א. האם A ו-B מאורעות זרים?
ב. האם המאורע B מכיל את המאורע A?
ג. האם A ו-C מאורעות זרים?
ד. האם A ו-C מאורעות משלימים?

11) עבור המאורעות A ו-B ידועות ההסתברויות הבאות: $P(A)=0.6$,
$P(B)=0.3$, $P(\bar{A}\cap\bar{B})=0.1$.

א. האם A ו-B מאורעות זרים?
ב. חשבו את $P(\bar{A}\cap B)$.

12) מטבע הוטל פעמיים. נגדיר את המאורעות הבאים:
A - קיבלנו עץ בהטלה הראשונה.
B - קיבלנו לפחות עץ אחד בשתי ההטלות.
איזו טענה נכונה?

א. A ו-B מאורעות זרים.
ב. A ו-B מאורעות משלימים.
ג. B מכיל את A.
ד. A מכיל את B.

13) בהגרלה חולקו 100 כרטיסים. על 3 מהם רשום חופשה ועל 2 מהם רשום מחשב. שאר הכרטיסים ריקים. אדם קיבל כרטיס אקראי.

א. מה הסיכוי לזכות בחופשה או במחשב? האם המאורעות הללו זרים?
ב. מה ההסתברות לא לזכות בפרס?

14) נתון כי: $P(A)=0.3\ ,\ P(B)=0.25\ ,\ P(A\cup B)=0.49$

א. חשבו את הסיכוי ל- $P(A\cap B)$.

ב. האם A ו- B מאורעות זרים?

ג. מה ההסתברות שרק A יקרה או שרק B יקרה?

15) A ו- B מאורעות זרים. נתון ש: $2\cdot P(B\cap\overline{A})=P(A\cap\overline{B})=P(\overline{A}\cap\overline{B})$.

מה הסיכוי למאורע A ומה ההסתברות למאורע B?

16) קבעו אילו מהטענות הבאות נכונות:

א. $A\cap B=B\cap A$.

ב. $\overline{A\cup B}=A\cap B$.

ג. $A\cap B\cap C=A\cap B\cap(C\cup B)$.

ד. $\overline{A\cap B\cap C}=\overline{A}\cup\overline{B}\cup\overline{C}$.

תשובות סופיות:

1) א. $A\cap B=\{EG,EF,FE,GE\}$.
ב. $A\cup B=\{EG,EF,EE,FE,GE,EG,GF\}$.

2) א. $B\cap\bar{A}$. ב. $A\cap\bar{B}$. ג. $A\cap B$. ד. $A\cup B$. ה. $\bar{A}\cap\bar{B}$. ו. $\bar{B}$.

3) א. $A=0, 2, 4, 6, 8$, $B=0, 1, 2, 3, 4$, $\bar{B}=5, 6, 7, 8, 9$,
$A\cap B=0, 2, 4$, $A\cup B=0, 2, 4, 6, 8, 1, 3$.
ב. $P(A)=0.5$, $P(B)=0.5$, $P(\bar{B})=0.5$, $P(A\cap B)=0.3$, $P(A\cup B)=0.7$.

4) $\bar{\bar{A}}=A$, $A\cap\phi=\phi$, $A\cup\phi=A$, $A\cap\Omega=A$, $A\cup\Omega=\Omega$,
$A\cap\bar{A}=\phi$, $\bar{\phi}=\Omega$, $A\cup\bar{A}=\Omega$.

5) א. $A\cap B$: גובה בין 1.7 ל-1.8. ב. $A\cup B$: כל גובה אפשרי.
ג. $\bar{A}=\bar{A}\cap B$: גובה לכל היותר 1.7. ד. $\bar{A}\cup\bar{B}$: לכל היותר 1.7 או לפחות 1.8.
ה. $A=\bar{\bar{A}}$: גובה מעל 1.7.

6) א. $A\cap B\cap C$. ב. $A\cap\bar{B}\cap\bar{C}$. ג. $A\cup B\cup C$.
ד. $\bar{C}$. ה. $(A\cap B\cap\bar{C})\cup(B\cap C\cap\bar{A})\cup(A\cap C\cap\bar{B})$.

7) א. $P(A\cup B)=0.24$. ב. $P(A\cap B)=0.04$. ג. $P(B\cap\bar{A})=0.16$.

8) א. $P(A\cap B)=10\%$. ב. $P(A\cup B)=50\%$. ג. $P(\bar{A}\cap\bar{B})=50\%$.

9) א. $P(A\cap B)=0.2$. ב. $P(\bar{A}\cap\bar{B})=0.3$. ג. $P(A\cup\bar{B})=0.3$.

10) א. לא. ב. כן. ג. כן. ד. לא.

11) א. כן. ב. $P(\bar{A}\cap B)=0.3$.

12) הטענה הנכונה היא ג'.

13) א. 0.05. ב. 0.95.

14) א. $P(A\cap B)=0.06$. ב. לא. ג. $P((A\cap\bar{B})\cup(B\cap\bar{A}))=0.43$.

15) $P(A)=\frac{2}{5}$, $P(B)=\frac{1}{5}$.

16) א. נכון. ב. לא נכון. ג. לא נכון. ד. נכון.

ביוסטטיסטיקה

פרק 13 - הסתברות מותנית-במרחב מדגם אחיד

תוכן העניינים

הסתברות מותנית – במרחב מדגם אחיד:

רקע:

לעיתים אנו נדרשים לחשב הסתברות למאורע כלשהו כאשר ברשותנו אינפורמציה לגבי מאורע אחר. הסתברות מותנית הינה סיכוי להתרחשות מאורע כלשהו כאשר ידוע שמאורע אחר התרחש/ לא התרחש.

ההסתברות של A בהינתן ש־B כבר קרה: $P(A|B)$.

כשמרחב המדגם אחיד: $P(A|B)=\frac{|A\cap B|}{|B|}$.

דוגמה (פתרון בהקלטה):

נטיל קובייה.

נגדיר:

A - התוצאה זוגית.

B - התוצאה גדולה מ-3.

נרצה לחשב את: $P(A|B)$.

שאלות:

1) נבחרה ספרה זוגית באקראי. מה הסיכוי שהספרה גדולה מ-6?

2) יוסי הטיל קובייה.
מה הסיכוי שקיבל את התוצאה 4, אם ידוע שהתוצאה שהתקבלה זוגית?

3) הוטלו צמד קוביות. נגדיר:
A - סכום התוצאות בשתי ההטלות הינו 7.
B - מכפלת התוצאות 12.
חשבו את $P(A|B)$.

4) מטבע הוטל פעמיים.
ידוע שהתקבל לכל היותר ראש אחד, מה הסיכוי שהתקבלו שני ראשים?

5) זוג קוביות הוטלו והתקבל שהתוצאות זהות.
מה הסיכוי שלפחות אחת התוצאות 5?

6) זוג קוביות הוטלו והתקבל לפחות פעם אחת 4.
מה הסיכוי שאחת התוצאות 5?

7) נבחרה משפחה בת שני ילדים, שמהם אחד הוא בן.
מה ההסתברות שבמשפחה שני בנים בקרב הילדים?

תשובות סופיות:

1) 0.2.
2) $\frac{1}{3}$.
3) 0.5.
4) 0.
5) $\frac{1}{6}$.
6) $\frac{2}{11}$.
7) $\frac{1}{3}$.

ביוסטטיסטיקה

פרק 14 - הסתברות מותנית - מרחב לא אחיד

תוכן העניינים

הסתברות מותנית – מרחב לא אחיד:

רקע:

הסיכוי שמאורע A יתרחש, בהינתן שמאורע B כבר קרה: $P(A|B)=\frac{P(A\cap B)}{P(B)}$.

במונה: הסיכוי לחיתוך של שני המאורעות, זה הנשאל וזה הנתון שהתרחש.
במכנה: הסיכוי למאורע נתון שהתרחש.

דוגמה (פתרון בהקלטה):

נבחרו משפחות שיש להם שתי מכוניות. ל-30% מהמשפחות הללו המכונית הישנה יותר היא מתוצרת אירופה ואצל 60% מהמשפחות הללו המכונית החדשה יותר מתוצרת אירופה. כמו כן, בקרב 15% מהמשפחות הללו שתי המכוניות הן מתוצרת אירופאית.
אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית?

שאלות:

1) תלמיד ניגש בסמסטר לשני מבחנים: מבחן בכלכלה ומבחן בסטטיסטיקה:
נגדיר את המאורעות הבאים:
A - לעבור את המבחן בסטטיסטיקה.
B - לעבור את המבחן בכלכלה.
כמו כן נתון שהסיכוי לעבור את המבחן בכלכלה הנו 0.8, הסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9 והסיכוי לעבור את שני המבחנים הנו 0.75.
חשבו את הסיכויים למאורעות הבאים:

א. התלמיד עבר בסטטיסטיקה, מה ההסתברות שהוא עבר בכלכלה?
ב. התלמיד עבר בכלכלה, מה ההסתברות שהוא עבר בסטטיסטיקה?
ג. התלמיד עבר בכלכלה, מה ההסתברות שהוא נכשל בסטטיסטיקה?
ד. התלמיד נכשל בסטטיסטיקה, מה ההסתברות שהוא נכשל בכלכלה?
ה. התלמיד עבר לפחות מבחן אחד, מה ההסתברות שהוא יעבור את שניהם?

2) במדינה שתי חברות טלפון סלולארי: "סופט" ו"בלי". 30% מהתושבים הבוגרים רשומים אצל חברת "בלי", 60% מהתושבים הבוגרים רשומים אצל חברת "סופט" ול-15% מהתושבים הבוגרים אין טלפון סלולארי כלל.

א. איזה אחוז מהתושבים הבוגרים רשומים אצל שתי החברות?
ב. נבחר אדם שרשום אצל חברת "סופט", מה ההסתברות שהוא רשום גם אצל חברת "בלי"?
ג. אם אדם לא רשום אצל חברת "בלי", מה ההסתברות שהוא כן רשום בחברת "סופט"?
ד. אם אדם רשום אצל חברה אחת בלבד, מה ההסתברות שהוא רשום בחברת "סופט"?

3) במכללה שני חניונים: חניון קטן וחניון גדול. בשעה 08:00 יש סיכוי של 60% שבחניון הגדול יש מקום, סיכוי של 30% שבחניון הקטן יש מקום וסיכוי של 20% שבשני החניונים יש מקום.

א. מה ההסתברות שיש מקום בשעה 08:00 רק בחניון הגדול של המכללה?
ב. ידוע שבחניון הקטן יש מקום בשעה 08:00, מה הסיכוי שבחניון הגדול יש מקום?
ג. אם בשעה 08:00 בחניון הגדול אין מקום, מה ההסתברות שבחניון הקטן יהיה מקום?
ד. נתון שלפחות באחד מהחניונים יש מקום בשעה 08:00, מה ההסתברות שבחניון הגדול יש מקום?

4) נלקחו 200 שכירים ו-100 עצמאים. מתוך השכירים 20 הם אקדמאיים, ומתוך העצמאיים 30 הם אקדמאיים.
א. בנו טבלת שכיחות משותפת לנתונים.
ב. נבחר אדם אקראי מה ההסתברות שהוא שכיר?
ג. מה ההסתברות שהוא שכיר ולא אקדמאי?
ד. מה ההסתברות שהוא שכיר או אקדמאי?
ה. אם האדם שנבחר הוא עצמאי מהי ההסתברות שהוא אקדמאי?
ו. אם האדם שנבחר הוא לא אקדמאי, מה ההסתברות שהוא שכיר?

תשובות סופיות:

1) א. 0.833. ב. 0.9375. ג. 0.0625. ד. 0.5. ה. 0.789.

2) א. 5%. ב. 0.0833. ג. 0.786. ד. 0.6875.

3) א. 0.4. ב. $\frac{2}{3}$. ג. 0.25. ד. $\frac{6}{7}$.

4) א. להלן טבלה: ב. $\frac{2}{3}$. ג. 0.6. ד. $\frac{23}{30}$.

	אקדמאי	לא אקדמאי	סה"כ
שכיר	20	180	200
עצמאי	30	70	100
סה"כ	50	250	300

ה. 0.3. ו. 0.72.

ביוסטטיסטיקה

פרק 15 - הערכת כלים אבחנתיים

תוכן העניינים

הערכת כלים אבחנתיים:

רקע:

אנו מנסים לאבחן תכונה מסוימת באמצעות כלי מסוים (למשל, לאבחן האם לאדם יש קורונה באמצעות ערכת אבחון ביתית). נגדיר מדדים סטטיסטיים שונים שנותנים אינדיקציה לאיכות כלי האבחנה.
נסמן ב-A: האדם קיבל תשובה חיובית, כלומר אובחן כבעל התכונה.

נסמן ב-B: האדם הוא בעל התכונה.

רגישות – Sensitivity:

ההסתברות שאדם בעל התכונה יקבל תשובה חיובית, כלומר: $Sensitivity = P\left(A|B\right)$.

סגוליות – Specificity:

ההסתברות שאדם ללא התכונה יקבל תשובה שלילית, כלומר: $Specificity = P\left(\bar{A}|\bar{B}\right)$.

דוגמה (פתרון בהקלטה):

5,000 נשים השתתפו במחקר שבו הן השתמשו בערכה לבדיקת היריון ביתית של חברה מסוימת ביום ה-14 של המחזור החודשי. בדיעבד מתוך 5,000 הנשים 80 היו בהיריון.
70 נשים קיבלו תשובה חיובית מערכת הבדיקה הביתית. 65 מהן אכן היו בהיריון.
מה הסגוליות ומה הרגישות של בדיקת ההיריון הביתית?

הערך המנבא החיובי – Positive Predictive Value:

ההסתברות שאדם שקיבל תשובה חיובית הוא אכן בעל התכונה.
הערך המנבא החיובי נקרא גם יכולת אבחון (יכולת דיאגנוסטית): $PPV = P\left(B|A\right)$.

הערך המנבא השלילי – Negative Predictive Value:

ההסתברות שאדם שקיבל תשובה שלילית אינו בעל התכונה: $NPV = P\left(\bar{B}|\bar{A}\right)$.

המשך הדוגמה (פתרון בהקלטה):

מהו הערך המנבא החיובי ומהו הערך המנבא השלילי של ערכת הבדיקה הביתית?

ההסתברות לתוצאה חיובית מדומה – False Positive:

ההסתברות שאדם שאינו בעל התכונה יקבל תשובה חיובית: $FP = P\left(A|\bar{B}\right)$.

ההסתברות לתוצאה שלילית מדומה – False Negative:

ההסתברות שאדם בעל התכונה יקבל תשובה שלילית: $FN = P\left(\bar{A}|B\right)$.

המשך הדוגמה (פתרון בהקלטה):

מה ההסתברות לתוצאה חיובית מדומה ומה ההסתברות לתוצאה שלילית מדומה בבדיקה באמצעות ערכת הבדיקה הביתית?

שאלות:

1) 2,000 גברים נבדקו בשיטה חדשה לגילוי סרטן המעי הגס. מתוך 150 גברים שביופסיה הוכיחה בוודאות שהם חולים 100 נמצאו חולים באמצעות הבדיקה החדשה. 80 גברים שהוכח שהם בריאים באמצעות ביופסיה קיבלו תשובה חיובית באמצעות השיטה החדשה.
מצאו את הסגוליות, הרגישות, הערך המנבא החיובי, הערך המנבא השלילי, ההסתברות לתוצאה חיובית מדומה.

2) הסיכוי של ערכת בדיקה ביתית של חברת קאנז לגלות מחלה מסוימת הוא 98%. לאדם בריא יש סיכוי של 5% לקבל בבדיקה תשובה חיובית. 5% ממשתמשי הערכה חולים במחלה.
א. מה הרגישות של הערכה?
ב. מה הסגוליות של הערכה?
ג. מה יכולת האבחון של הערכה?

3) הסיכוי של תהליך אבחון של הפרעת קשב לזהות סטודנטים בעלי הפרעת קשב הוא 0.95. הסיכוי שלו לזהות בטעות גם סטודנטים ללא הפרעת קשב כבעלי הפרעת קשב הוא 0.01. ידוע ש-7% מהסטודנטים סובלים מהפרעת קשב.
א. מה ההסתברות לתשובה חיובית מדומה?
ב. מה ההסתברות לתשובה שלילית מדומה?

4) נערכה בדיקה בשיטה חדשה לגילוי מלנומה בקרב 800 נבדקים – 400 נבדקים עם ביופסיה מוכחת של מלנומה ויתר הנבדקים ידועים ללא מלנומה. 30 מהנבדקים נמצאו בבדיקה החדשה חולים במלנומה. מתוך ה-30 רק 10 חולים באמת.
א. מה הרגישות של הבדיקה החדשה?
ב. מה הסגוליות של הבדיקה החדשה?
ג. מה היכולת הדיאגנוסטית של הבדיקה החדשה?

5) קבוצה של נבדקים שידוע ש-25% מהם נשאי HIV נבדקה בבדיקה חדשה המאפשרת קבלת תוצאה באופן מיידי. בבדיקה החדשה נמצאו 20% מהנבדקים כנשאי HIV. הסגוליות של הבדיקה החדשה היא 76%.
א. מה הרגישות של הבדיקה החדשה?
ב. מה ה-PPV של הבדיקה החדשה?
ג. מה ה-NPV של הבדיקה החדשה?

תשובות סופיות:

1) סגוליות: 0.9568, רגישות: $\frac{2}{3}$, הערך המנבא החיובי: $\frac{5}{9}$,
הערך המנבא השלילי: 0.9725, ההסתברות לתוצאה חיובית מדומה: 0.0043.

2) א. 0.98. ב. 0.95. ג. 0.5078.

3) א. 0.01. ב. 0.05.

4) א. 0.25. ב. 0.95. ג. $\frac{1}{3}$.

5) א. 0.08. ב. 0.1. ג. 0.7125.

ביוסטטיסטיקה

פרק 16 - דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה

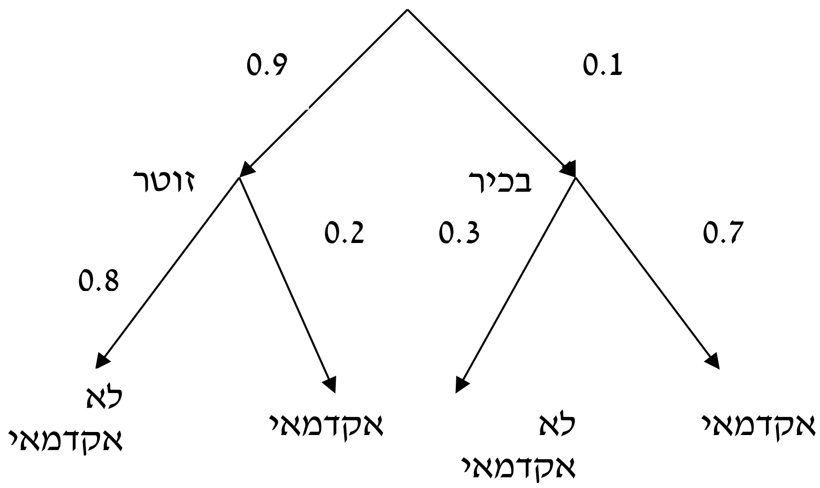
תוכן העניינים

דיאגרמת עצים – נוסחת בייס ונוסחת ההסתברות השלמה:

נשתמש בשיטה זו כאשר יש תרגיל שבו התרחשות המאורעות היא בשלבים, כך שכל תוצאה של כל שלב תלויה בשלב הקודם, פרט לשלב הראשון:

דוגמה:

בחברה מסוימת 10% מוגדרים בכירים והיתר מוגדרים זוטרים. מבין הבכירים 70% הם אקדמאים ומבין הזוטרים 20% הם אקדמאים. נשרטט עץ שיתאר את הנתונים, השלב הראשון של העץ אינו מותנה בכלום ואילו השלב השני מותנה בשלב הראשון.



כדי לקבל את הסיכוי לענף מסוים נכפיל את כל ההסתברויות על אותו ענף.
נבחר אדם באקראי מאותה חברה.

1) מה הסיכוי שהוא בכיר אקדמאי? $0.1\cdot 0.7=0.07$.
2) מה הסיכוי שהוא זוטר לא אקדמאי? $0.9\cdot 0.8=0.72$.

כדי לקבל את הסיכוי לכמה ענפים נחבר את הסיכויים של כל ענף (רק אחרי שבתוך הענף הכפלנו את ההסתברויות).

3) מה הסיכוי שהוא אקדמאי? $0.1\cdot 0.7+0.9\cdot 0.2=0.25$.
4) נבחר אקדמאי מה ההסתברות שהוא עובד זוטר?
מדובר כאן על שאלה בהסתברות מותנה ולכן נשתמש בעיקרון של הסתברות מותנה: $P\left(zutar\,|\,academay\right)=\frac{0.9\cdot 0.2}{0.25}=\frac{0.18}{0.25}=0.72$.

נוסחת ההסתברות השלמה:

בהינתן B, מאורע כלשהו, וחלוקה של מרחב המדגם Ω ל- $A_1,\dots,A_n$ כך ש- $\Omega=\bigcup_i A_i$,

אזי: $P(B)=\sum_{i=1}^{n}P(A_i)\cdot P\left(\frac{B}{A_i}\right)$.

נוסחת בייס:

$$P\left(\frac{A_j}{B}\right)=\frac{P(A_j)P\left(\frac{B}{A_j}\right)}{\sum^{n}P(A_i)\cdot P\left(\frac{B}{A_i}\right)}.$$

שאלות:

1) בשקית סוכריות 4 סוכריות תות ו-3 לימון. מוציאים באקראי סוכרייה. אם היא בטעם תות אוכלים אותה ומוציאים סוכרייה נוספת, ואם היא בטעם לימון מחזירים אותה לשקית ומוציאים סוכרייה נוספת.

א. מה ההסתברות שהסוכריה הראשונה שהוצאה בטעם תות והשנייה בטעם לימון?

ב. מה ההסתברות שהסוכריה השנייה בטעם לימון?

2) באוכלוסייה מסוימת 30% הם ילדים, 50% בוגרים והיתר קשישים. לפי נתוני משרד הבריאות הסיכוי שילד יחלה בשפעת במשך החורף הוא 80%, הסיכוי שמבוגר יחלה בשפעת במשך החורף הוא 40% והסיכוי שקשיש יחלה בשפעת במשך החורף הוא 70%.

א. איזה אחוז מהאוכלוסייה הינו קשישים שלא יחלו בשפעת במשך החורף?

ב. מה אחוז האנשים שיחלו בשפעת במשך החורף?

ג. נבחר אדם שחלה במשך החורף בשפעת, מה ההסתברות שהוא קשיש?

ד. נבחר ילד, מה ההסתברות שהוא לא יחלה בשפעת במשך החורף?

3) בכד א' 5 כדורים כחולים ו-5 כדורים אדומים. בכד ב' 6 כדורים כחולים ו-4 כדורים אדומים. בוחרים באקראי כד, מוציאים ממנו כדור ומבלי להחזירו מוציאים כדור נוסף.

א. מה ההסתברות ששני הכדורים שיוצאו יהיו בצבעים שונים?

ב. אם הכדורים שהוצאו הם בצבעים שונים, מה ההסתברות שהכדור השני שהוצא יהיה בצבע אדום?

4) חברת סלולר מסווגת את לקוחותיה לפי 3 קבוצות גיל: נוער, בוגרים ופנסיונרים. נתון כי: 10% מהלקוחות בני נוער, 70% מהלקוחות בוגרים והיתר פנסיונרים. מתוך בני הנוער 90% מחזיקים בסמארט-פון, מתוך האוכלוסייה הבוגרת ל-70% יש סמארט-פון ומתוך אוכלוסיית הפנסיונרים 30% מחזיקים בסמארט-פון.

א. איזה אחוז מלקוחות החברה הם בני נוער עם סמארט-פון?

ב. נבחר לקוח אקראי ונתון שיש לו סמארט-פון. מה ההסתברות שהוא פנסיונר?

ג. אם ללקוח אין סמארט-פון, מה ההסתברות שהוא לא בן נוער?

5) כדי להתקבל למקום עבודה יש לעבור שלושה מבחנים. המבחנים הם בשלבים, כלומר לאחר כישלון במבחן מסוים אין אפשרות לגשת למבחן הבא אחריו. 70% מהמועמדים עוברים את המבחן הראשון. מתוכם, 50% עוברים את המבחן השני. מבין אלה שעוברים את המבחן השני 40% עוברים את המבחן השלישי.

א. מה ההסתברות להתקבל לעבודה?

ב. מועמד לא התקבל לעבודה. מה ההסתברות שהוא נכשל במבחן הראשון?

ג. מועמד לא התקבל לעבודה. מה ההסתברות שהוא עבר את המבחן השני?

6) משרד הבריאות פרסם את הנתונים הבאים:
מתוך אוכלוסיית הילדים והנוער 80% חולים בשפעת בזמן החורף.
מתוך אוכלוסיית המבוגרים (עד גיל 65) 60% חולים בשפעת בזמן החורף.
30% מהתושבים הם ילדים ונוער. 50% הם מבוגרים. היתר קשישים.
כמו כן נתון ש68% מהאוכלוסייה תחלה בשפעת בחורף.

א. מה אחוז החולים בשפעת בקרב האוכלוסייה הקשישה?

ב. נבחר אדם שלא חלה בשפעת, מה ההסתברות שהוא לא קשיש?

7) רדאר שנמצא על החוף צריך לקלוט אנייה הנמצאת ב-1 מ-4 האזורים: A, B, C, D. אם האנייה נמצאת באזור A הרדאר מזהה אותה בסיכוי 0.8, סיכוי זה פוחת ב-0.1 ככל שהאנייה מתקדמת באזור. כמו כן נתון שבהסתברות חצי האנייה נמצאת באזור D, בהסתברות 0.3 באזור C, באזור B היא נמצאת בסיכוי 0.2, אחרת היא נמצאת באזור A.

א. מה הסיכוי שהאנייה תתגלה ע"י הרדאר?

ב. אם האנייה התגלתה ע"י הרדאר, מה ההסתברות שהיא נמצאת באזור C?

ג. אם האנייה התגלתה ע"י הרדאר, מה הסיכוי שהיא לא נמצאת באזור B?

8) סימפטום X מופיע בהסתברות של 0.4 במחלה A, בהסתברות של 0.6 במחלה B ובהסתברות של 0.5 במחלה C. סימפטום X מופיע אך ורק במחלות הללו, אדם לא יכול לחלות ביותר ממחלה אחת מבין המחלות הללו. לקליניקה מגיעים אנשים כדלקמן: 8% חולים במחלה A, 10% במחלה B, 2% במחלה C והיתר בריאים. כמו כן נתון שבמחלה A, סימפטום X מתגלה בסיכוי של 80%, ובמחלות B, C הסימפטום מתגלה בסיכוי של 90% בכל מחלה.

א. מה ההסתברות שאדם הגיע לקליניקה וגילו אצלו את סימפטום X?

ב. אם התגלה אצל אדם סימפטום X, מה ההסתברות שהוא חולה במחלה A?

ג. אם לאדם יש את סימפטום X, מה ההסתברות שהוא חולה במחלה A?

ד. אם לא גילו אצל אדם את סימפטום X, מה ההסתברות שהוא בריא?

תשובות סופיות:

1) א. $\frac{2}{7}$. ב. $\frac{23}{49}$.

2) א. 6%. ב. 58%. ג. 0.241. ד. 0.2.

3) א. 0.544. ב. 0.5.

4) א. 9%. ב. 0.09375. ג. 0.9722.

5) א. 0.14. ב. 0.3488. ג. 0.2442.

6) א. 70%. ב. 0.8125.

7) א. 0.57. ב. 0.3158. ג. 0.7543.

8) א. 0.0886. ב. 0.2889. ג. 0.3137. ד. 0.8778.

ביוסטטיסטיקה

פרק 17 - תלות ואי תלות בין מאורעות

תוכן העניינים

תלות ואי תלות בין מאורעות:

רקע:

אם מתקיים ש: $P(B|A)=P(B)$, נגיד שמאורע B בלתי תלוי ב- A.

הדבר גורר גם ההפך: $P(A|B)=P(A)$, כלומר, גם A אינו תלוי ב- B.

כשהמאורעות בלתי תלויים מתקיים ש: $P(A\cap B)=P(A)\cdot P(B)$.

הוכחה לכך: $P(A/B)=P(A)=\dfrac{P(A\cap B)}{P(B)}\Rightarrow P(A\cap B)=P(A)\cdot P(B)$.

נשתמש בנוסחאות של מאורעות בלתי תלויים רק אם נאמר במפורש שהמאורעות בלתי תלויים בתרגיל או שמההקשר אפשר להבין ללא צל של ספק שהמאורעות בלתי תלויים.

למשל,

חוקר מבצע שני ניסויים בלתי תלויים הסיכוי להצליח בניסוי הראשון הנו 0.7 והסיכוי להצליח בניסוי השני הוא 0.4.

א. מה הסיכוי להצליח בשני הניסויים יחדו?

כיוון שהמאורעות הללו בלתי תלויים:

$$P(A\cap B)=P(A)\cdot P(B)=0.7\cdot 0.4=0.28$$

ב. מה הסיכוי להיכשל בשני הניסויים?

באופן דומה:

$$P(\bar{A}\cap \bar{B})=P(\bar{A})\cdot P(\bar{B})=(1-0.7)(1-0.4)=0.18$$

הרחבה: אי תלות בין n מאורעות:

n מאורעות $A_1,\dots,A_n$ הם בלתי תלויים אם ורק אם: $P\left(\bigcap_{i=1}^{n} A_i\right)=\prod_{i=1}^{n} P(A_i)$.

שאלות:

1) נתון: $P(A)=0.2$, $P(B)=0.5$, $P(A\cup B)=0.6$.
האם המאורעות הללו בלתי תלויים?

2) תלמיד ניגש לשני מבחנים שהצלחתם לא תלויה זו בזו.
הסיכוי שלו להצליח במבחן הראשון הוא 0.7 והשני 0.4.
א. מה הסיכוי להצליח בשני המבחנים יחדו?
ב. מה הסיכוי שניכשל בשני המבחנים?

3) במדינה מסוימת יש 8% אבטלה, נבחרו באקראי שני אנשים מהמדינה.
א. מה ההסתברות ששניהם מובטלים?
ב. מה ההסתברות שלפחות אחד מהם מובטל?

4) מוצר צריך לעבור בהצלחה ארבע בדיקות בלתי תלויות לפני שיווקו, אחרת הוא נפסל ולא יוצא לשוק. הסיכוי לעבור בהצלחה כל אחת מהבדיקות הוא 0.8. בכל מקרה מבוצעות כל 4 הבדיקות.
א. מה הסיכוי שהמוצר יפסל?
ב. מה ההסתברות שהמוצר יעבור בהצלחה לפחות בדיקה אחת?

5) במדינה מסוימת יש 8% אבטלה, נבחרו באקראי חמישה אנשים מהמדינה.
א. מה ההסתברות שכולם מובטלים במדגם?
ב. מה ההסתברות שלפחות אחד מהם מובטל?

6) עבור שני מאורעות A ו- B המוגדרים על אותו מרחב מדגם נתון ש:
$P(A|B)=0.6$, $P(A\cap\overline{B})=0.3$, $P(A\cup B)=0.9$.
האם A ו- B מאורעות בלתי תלויים?

7) הוכיח שאם: $P(A/B)=P(B/A)$, אז: $P(A)=P(B)$.

8) קבעו אילו מהטענות הבאות נכונות. נמק!

א. אם: $P(A\cup B)=P(A)\cdot P(B)$, אזי המאורעות בלתי תלויים.

ב. מאורע A כלול במאורע B: $0<P(B)<1$, $P(A)>0$, לכן: $P(A/B)<P(A)$.

ג. A ו- B מאורעות זרים שסיכוייהם חיובים לכן הם מאורעות תלויים.

ד. A ו- B מאורעות תלויים שסיכוייהם חיובים לכן A ו- B מאורעות זרים.

ה. $\overline{P}\left(\overline{A}\cap\overline{B}\right)=1-P(A)-P(B)$ לכן A ו- B מאורעות זרים.

תשובות סופיות:

1) כן.
2) א. 0.28. ב. 0.18.
3) א. 0.0064. ב. 0.1536.
4) א. 0.5904. ב. 0.9984.
5) א. 0.08^5. ב. 0.3409.
6) לא, הם תלויים.
7) שאלת הוכחה.
8) א. לא נכון. ב. לא נכון. ג. נכון. ד. לא נכון. ה. נכון.

ביוסטטיסטיקה

פרק 18 - התפלגויות בדידות מיוחדות -התפלגות בינומית

תוכן העניינים

התפלגויות בדידות מיוחדות – התפלגות בינומית:

רקע:

נגדיר את המושג ניסוי ברנולי:
ניסוי ברנולי הנו ניסוי שיש לו שתי תוצאות אפשריות: "הצלחה" ו"כישלון".
למשל מוצר פגום או תקין, אדם עובד או מובטל, עץ או פלי בהטלת מטבע וכדומה.
בהתפלגות בינומית חוזרים על אותו ניסוי ברנולי n פעמים באופן בלתי תלוי זה בזה.
מגדירים את X להיות מספר ההצלחות שהתקבלו בסך הכול. נסמן ב- P את הסיכוי להצלחה בניסוי בודד, וב- Q את הסיכוי לכישלון בניסוי בודד.
ואז נגיד ש: $X \sim B(n,p)$.

פונקציית ההסתברות של X:

לכל $k=0,1,2,....,n$ $\quad P(X=K)=\binom{n}{k}p^k(1-p)^{n-k}$

כאשר: $0!=1$; $n!=n\cdot(n-1)\cdot(n-2)\cdot....\cdot1$; $\binom{n}{k}=\dfrac{n!}{k!(n-k)!}$

לגודל: $\binom{n}{k}$ ניתן לחשב באמצעות המחשבון.

תוחלת: $E(X)=np$.

שונות: $V(X)=npq$.

שימו לב, כדי לזהות שמדובר בהתפלגות בינומית צריכים להתקיים כל התנאים הבאים:

1) חוזרים על אותו ניסוי ברנולי באופן בלתי תלוי זה בזה.
2) חוזרים על הניסוי n פעמים.
3) X – מוגדר כמספר ההצלחות המתקבלות בסך הכול.

דוגמה (פתרון בהקלטה):

במדינה מסוימת ל-80% מהתושבים יש רישיון נהיגה.
נבחרו 10 תושבים אקראיים מהמדינה.

א. מה ההסתברות שבדיוק ל-9 מהם יש רישיון נהיגה?
ב. מה ההסתברות שלפחות ל-9 מהם יש רישיון נהיגה?
ג. מהי התוחלת ומהי סטיית התקן של מספר התושבים שנדגמו ושיש להם רישיון נהיגה?

שאלות:

1) במדינה 10% מהאוכלוסייה מובטלת. נבחרו 5 אנשים באקראי מאותה אוכלוסייה. נגדיר את X להיות מספר המובטלים שהתקבלו במדגם.
   א. מהי ההתפלגות של X ?
   ב. מה ההסתברות שיהיה בדיוק מובטל אחד?
   ג. מה ההסתברות שכולם יעבדו במדגם?
   ד. מה ההסתברות ששלושה יעבדו במדגם?
   ה. מה ההסתברות שלפחות אחד יהיה מובטל?
   ו. מה תוחלת ומהי השונות של מספר המובטלים במדגם?

2) על פי נתוני משרד התקשורת ל-70% מהאוכלוסייה יש סמארטפון. נבחרו 10 אנשים באקראי. נגדיר את X כמספר האנשים שנדגמו עם סמארטפון.
   א. מהי ההתפלגות של X ? הסבירו.
   ב. מה ההסתברות שבמדגם ל-8 אנשים יש סמארט-פון?
   ג. מה ההסתברות שבמדגם לפחות ל-9 יהיו סמארט-פון?
   ד. מה התוחלת ומה סטיית התקן של מספר האנשים שנדגמו ולהם סמארט-פון?

3) במדינה מסוימת התפלגות ההשכלה בקרב האוכלוסייה מעל גיל 30 היא כזו:

השכלה	**נמוכה**	**תיכונית**	**תואר I**	**תואר II ומעלה**
פרופורציה	0.1	0.6	0.2	0.1

   נבחרו 20 אנשים אקראיים מעל גיל 30.
   א. מה ההסתברות ש-5 מהם אקדמאים?
   ב. מה התוחלת של מס' בעלי ההשכלה הנמוכה?

4) במכללה מסוימת 20% מהסטודנטים גרים בת"א. מבין הסטודנטים שגרים בת"א 30% מגיעים ברכבם, ומבין הסטודנטים שלא גרים בת"א 50% מגיעים ברכבם למכללה.
   א. השומר בשער המכללה בודק לכל סטודנט את תיקו בהיכנסו למכללה. מה ההסתברות שבקרב 5 סטודנטים שנבדקו ע"י השומר רק 1 מתוכם הגיע למכללה ברכבו?
   ב. בהמשך לסעיף הקודם מה ההסתברות שרוב הסטודנטים בקרב ה-5 הגיעו למכללה ברכבם?

תשובות סופיות:

1) א. $X \sim B(n=5, p=0.1)$. ב. 0.32805. ג. 0.59049.
ד. 0.0729. ה. 0.40954. ו. תוחלת: 0.5, שונות: 0.45.

2) ב. 0.2335. ג. 0.1493 ד. תוחלת: 7, סטיית תקן: 1.449.

3) א. 0.1789. ב. 2.

4) א. 0.1956. ב. 0.4253.

ביוסטטיסטיקה

פרק 19 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

תוכן העניינים

ביוסטטיסטיקה

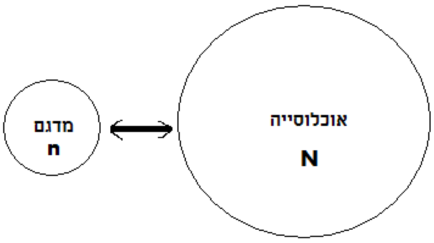
פרק 20 - הסקה סטטיסטית - הקדמה

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	$\overline{X}$	μ
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.

ביוסטטיסטיקה

פרק 21 - 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- μ (נקרא גם תוחלת).
שונות אוכלוסייה נסמן ב- σ^2.
סטיית תקן של אוכלוסייה: σ.

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 μ ושונות σ^2.

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

דוגמה (פתרון בהקלטה):

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

משפט הגבול המרכזי:

אם אוכלוסייה מתפלגת כלשהו עם ממוצע μ ושונות σ^2 אזי עבור מדגם מספיק גדול $(n \geq 30)$ ממוצע המדגם מתפלג בקירוב נורמאלי: $\bar{x} \overset{\sim}{\succ} N\left(\mu, \frac{\sigma^2}{n}\right)$.

דוגמה (פתרון בהקלטה):

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

שאלות:

1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
   א. מיהי האוכלוסייה?
   ב. מה המשתנה?
   ג. מהם הפרמטרים?
   ד. מהו גודל המדגם?
   ה. מהו תוחלת ממוצע המדגם?
   ו. מהי טעות התקן?

2) 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?

3) 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?

4) הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים.
לצורך הפתרון הניחו שזמן הנסיעה לעבודה מתפלג 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?

5) 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?

6) 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?
הסבר את ההבדל בתשובות של שני הסעיפים.

7) לפי הערכות הלשכה המרכזית לסטטיסטיקה השכר הממוצע במשק הוא 8000 ₪ עם סטיית תקן של 3000 ₪.
מה ההסתברות שבמדגם מקרי של 100 עובדים השכר הממוצע יהיה יותר מ-8500 ₪?

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

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

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

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

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

10) נתון ש- X מתפלג כלשהו עם תוחלת μ ושונות σ^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. $\bar{X} \sim N\left(\mu,\frac{\sigma^2}{200}\right)$.

11) 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. μ ו- $\bar{X}$ יהיו משתנים מקריים.

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

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

ד. μ ו- $\bar{X}$ יהיו קבועים.

תשובות סופיות:

1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון.
ג. ממוצע: 78, סטיית תקן: 15. ד. 2.
ה. 78. ו. 10.6.

2) א. 0.8413. ב. 0.0013. ג.0. ד.0.1974.

3) א. 0. ב. 0. ג. 0.9544. ד. 178.205.

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

5) א. 0. ב. 0.1587. ג. 0.1587. ד. 0.5.

6) א. 0.5468. ב. 0.6826.

7) 0.0475.

8) א. 0.9772. ב. 0.0228. ג. 271.

9) ב'.

10) ד'.

11) ג'.

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 μ ושונותה σ^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: $E(X)=\mu$, $V(X)=\sigma^2$,

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

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?

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.

התפלגות מספר ההצלחות במדגם – קירוב נורמלי להתפלגות הבינומית:

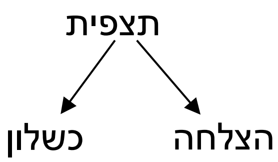
רקע:

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

תיקון רציפות:

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

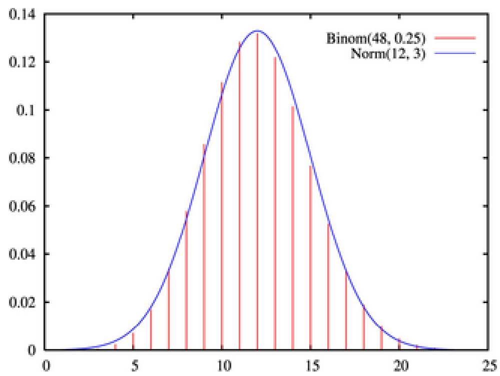
1. $p(Y=a)\cong p\left(a-\frac{1}{2}\le Y\le a+\frac{1}{2}\right)$.
2. $P(Y\le a)\cong P(Y\le a+0.5)$.
3. $P(Y\ge a)\cong P(Y\ge a-0.5)$.

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:

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

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

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.

ביוסטטיסטיקה

פרק 22 - מושגי יסוד באמידה

תוכן העניינים

מושגי יסוד באמידה:

רקע:

כזכור מהמפגש הקודם, פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסוימת. כמו ממוצע הגבהים בקרב מתגייסים לצה"ל - μ.

כמו פרופורציית התומכים בממשלה בקרב אזרחי המדינה - p.

בדרך כלל הפרמטרים הם גדלים שאינם ידועים באמת, ולכן מבצעים מדגמים במטרה לאמוד אותם. אין אפשרות לחשב אותם הניסיון הוא בלהעריך כמה הם שווים ככל שניתן.

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

דוגמה (פתרון בהקלטה):

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

א. מה הפרמטר בדוגמה זו?

ב. מהי טעות האמידה של ערוץ 10?

- $\hat{\theta}$ יהיה אומד חסר הטיה ל-θ אם התוחלת של $\hat{\theta}$ תהיה שווה ל-θ: $E\left(\hat{\theta}\right)=\theta$.
- טעות התקן של אומד היא סטיית התקן שלו, כלומר: $\sigma\left(\hat{\theta}\right)=S.E$

פרמטרים מרכזיים והאומדים שלהם:

ממוצע האוכלוסייה μ:

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

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

פרופורציה באוכלוסייה p:

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

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

שונות האוכלוסייה σ^2:

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

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

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

דוגמה (פתרון בהקלטה):

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

1. ממוצע מספר הילדים למשפחה בתל אביב.
2. שונות מספר הילדים למשפחה בתל אביב.
3. פרופורציית המשפחות בנות שני ילדים.

שאלות:

1) מתוך 500 טירונים, נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.

א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?

ב. מהי טעות התקן של האומד כשהמדגם בגודל 500?

ג. מהו האומדן לפרמטר?

ד. מהי טעות האמידה?

2) לפי נתוני היצרן, מקרר צורך בממוצע 2400 וואט לשעה עם סטיית תקן של 500 וואט לשעה .
במדגם של 25 מקררים של היצרן התקבל ממוצע של 2342 וואט לשעה.

א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?

ב. מהי טעות התקן של האומד?

ג. מהו האומדן לפרמטר?

ד. מהי טעות האמידה?

3) נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו: 168, 184, 192, 171, 180, 177, 187, 168, 177 ו-175.

א. מצאו אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.

ב. מצאו אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.

ג. מצאו אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.

4) נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים.
להלן התוצאות שהתקבלו: $\sum_{i=1}^{20} X_i = 162$, $\sum_{i=1}^{20} X^2{}_i = 1502.2$.

א. אמדו את השכר הממוצע של השכירים במשק.

ב. אמדו את סטיית התקן של שכר השכירים במשק.

5) במטרה לאמוד את ממוצע האוכלוסייה, דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישבו את הממוצע שלהם. מהי טעות התקן?

א. סטיית התקן של האוכלוסייה.

ב. סטיית התקן של ממוצע האוכלוסייה.

ג. סטיית התקן של המדגם.

ד. סטיית התקן של ממוצע המדגם.

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

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

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

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

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

א. 16.
ב. 8.
ג. 4.
ד. 2.

9) מהו אומד חסר הטיה?

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

תשובות סופיות:

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

ביוסטטיסטיקה

פרק 23 - 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 μ 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 (μ) במקרה ש- σ^2 (שונות האוכלוסייה) ידועה.

פרמטר אותו נרצה לאמוד: μ.

אומד נקודתי: $\bar{x}$.

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

σ^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: $\varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

ε - 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- μ 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.

תשובות סופיות:

1) א. העובדים במשק. ב. שכר ב-₪. ג. μ. 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- μ היא 416 ₪ בבטחון 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'.

קביעת גודל מדגם:

רקע:

אם מעוניינים לאמוד את ממוצע האוכלוסייה כאשר סטיית התקן של האוכלוסייה ידועה: σ ברמת סמך של $1-\alpha$ ושגיאת אמידה שלא תעלה על ε 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 μ וסטיית תקן σ. חוקר רוצה לבנות רווח בר סמך ל- μ ברמת ביטחון של 0.95, כך שהאורך של הרווח יהיה 0.5σ.
מהו גודל המדגם הנדרש?

תשובות סופיות:

1) 780.

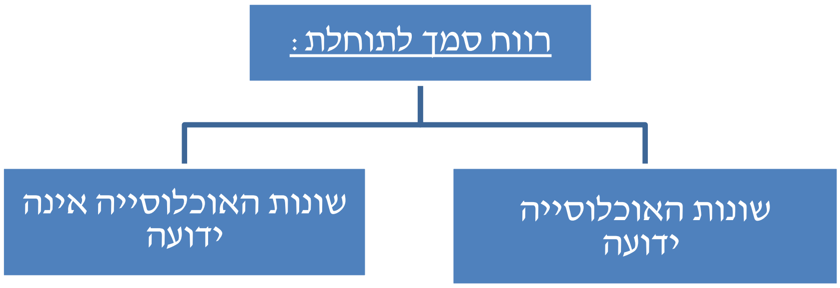
2) א. 139. ב. הדבר יקטין את ε פי 2.

3) $n = 62$.

רווח סמך כששונות האוכלוסייה לא ידועה:

רקע:

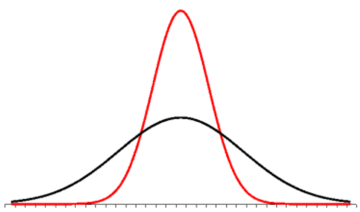
בבואנו לבנות רווח סמך לתוחלת אנו צריכים להתמקד בשני המצבים הבאים:



בפרק זה נעסוק במקרה ששונות האוכלוסייה $\left(\sigma^2\right)$ 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: $S^2 = \dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} = \dfrac{\sum_{i=1}^{n} X^2_i - n\bar{X}^2}{n-1}$.

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 μ.
לכל אחד מהסטטיסטיקאים מדגם אחר, אך באותו גודל 10.
שניהם קבעו אותה רמת סמך.
סטטיסטיקאי א: הניח $\sigma=20$ 2.
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%.

ביוסטטיסטיקה

פרק 24 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

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

דרגות החופש: $d.f=n_1+n_2-2$.

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

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 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 9200 ₪.
כמו כן נדגמו 120 תושבים מאזור B וממוצע המשכורות שהתקבל 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.

ביוסטטיסטיקה

פרק 25 - 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: μ_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) $-0.013 < \mu < 0.185$.

ביוסטטיסטיקה

פרק 26 - 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:

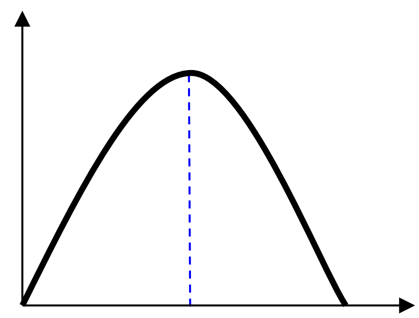
בפרק זה נדון איך קובעים גודל מדגם שבאים לאמוד פרופורציה באוכלוסייה מסוימת:
החוקר קובע מראש את רמת הסמך הרצויה: $1-\alpha$.
החוקר קובע מראש את הטעות הסטטיסטית המרבית שבה הוא מעוניין: ε
(או את אורך רווח הסמך).

$L=2\varepsilon$ - אורך רווח הסמך.
ε - 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: $n \geq \left(\frac{2 \cdot Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\hat{p}(1-\hat{p})}}{L} \right)^2$.



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: $\hat{p}=0.5$.

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

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₪.
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 ± 0.02 ג. ראה סרטון.

ביוסטטיסטיקה

פרק 27 - 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) $0.625 < p < 0.7754$.

ביוסטטיסטיקה

פרק 28 - שאלות מסכמות על רווחי סמך

תוכן העניינים

שאלות מסכמות על רווחי סמך:

שאלות:

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

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

א. תנו רווח סמך לממוצע מספר כוסות הקפה שאנשים נוהגים לשתות ביום. $\alpha = 0.05$.

ב. אדם השותה לפחות 4 כוסות קפה ביום נקרא "מכור לקפה". בנו רווח סמך לאחוז "המכורים לקפה". $\alpha = 0.1$.

3) חוקר בנה רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת בשנה.
רווח הסמך שהתקבל הוא: $81 < p < 91$. רווח הסמך הנ"ל התבסס על מדגם של 500 איש.

א. כמה אנשים במדגם טענו שכלל לא התקררו השנה?

ב. באיזו רמת סמך נבנה רווח הסמך?

ג. בנו רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת השנה ברמת סמך של 96% על סמך תוצאות המדגם.

4) ציוני IQ בארה"ב מתפלגים נורמאלית עם תוחלת 100. במדגם של 20 ישראלים שנבחנו במבחן ה-IQ התקבלו התוצאות הבאות:

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

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%.

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:

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. $0.704 \le p \le 0.786$.

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

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

ביוסטטיסטיקה

פרק 29 - 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- α.

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. $\begin{matrix} H_0: \mu = 72 \\ H_1: \mu > 72 \end{matrix}$.

2) א. משקאות בבקבוק של חברה מסוימת.
ב. נפח משקה בסמ"ק.
ג. ממוצע נפח המשקה בבקבוק.
ד. $\begin{matrix} H_0: \mu = 500 \\ H_1: \mu < 500 \end{matrix}$.

3) א. מועמדים לפקולטה למשפטים.
ב. משתנה דיכוטומי (התקבל, לא התקבל).
ג. אחוז הקבלה.
ד. $\begin{matrix} H_0: p = 0.25 \\ H_1: p < 0.25 \end{matrix}$.

4) א. אזרחים בוגרים במשק.
ב. משתנה דיכוטומי (מובטל, עובד).
ג. אחוז האבטלה כיום.
ד. $\begin{matrix} H_0: p = 0.08 \\ H_1: p \neq 0.08 \end{matrix}$.

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.

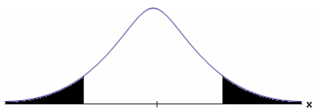
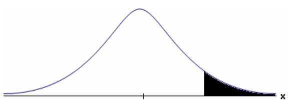
ביוסטטיסטיקה

פרק 30 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה ידועה:

רקע:

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

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:

<u>אוכלוסייה</u>: עגבניות.

<u>המשתנה</u>: X = יבול העגבניות בטון לעונה.

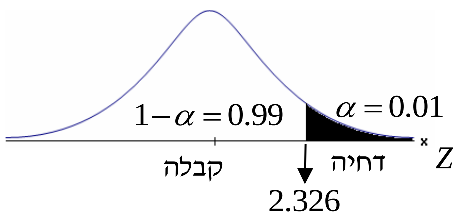
<u>הפרמטר</u>: μ = תוחלת היבול בשיטת הזיבול החדשה.

<u>השערות</u>: $\begin{array}{l} H_0: \mu = 10 \\ H_1: \mu > 10 \end{array}$.

<u>תנאים</u>:

1. $X \sim N$.
2. $\sigma = 2.5$.

<u>כלל הכרעה</u>:



נדחה את H_0 אם $Z_{\bar{x}} > 2.326$

<u>תוצאות</u>: $\bar{x} = 12.5$, $n = 4$

<u>סטטיסטי המבחן</u>: $Z_{\bar{x}} = \dfrac{\bar{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$.

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

<u>מסקנה</u>:

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 α והחליט לדחות את השערת האפס.
אם החוקר היה עורך מבחן צדדי ברמת מובהקות של $\frac{\alpha}{2}$ אזי בהכרח:
א. השערת האפס הייתה נדחית.
ב. השערת האפס הייתה לא נדחית.
ג. לא ניתן לדעת מה תהיה מסקנתו במקרה זה.

8) שני סטטיסטיקאים בדקו השערות: $H_0: \mu = \mu_0$ כנגד $H_1: \mu > \mu_0$, עבור שונות ידועה ובאותה רמת מובהקות.
שני החוקרים קבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.
א. אם חוקר א' החליט לדחות את H_0, מה יחליט חוקר ב'? נמקו.
ב. אם חוקר א' יחליט לא לדחות את H_0, מה יחליט חוקר ב'? נמקו.

תשובות סופיות:

1) נקבל H_0, בר"מ של 5% לא נקבל את הטענה של המורה ששיטת הלימוד שלו מעלה את ממוצע הציונים.
2) א. נדחה H_0, בר"מ של 2.5% נקבל את תלונת אגודת הצרכנים בדבר הפחתת נפח המשקה בבקבוק.
ב. הגדלנו את רמת המובהקות לכן אנחנו נשארים בדחייה של H_0 והמסקנה לא תשתנה.
3) נדחה H_0, בר"מ של 5% נקבע שהמכונה לא מאופסת.
4) נדחה H_0, בר"מ של 0.1 נקבל את הטענה שהדיאטה יעילה ומפחיתה את המשקל הממוצע.
5) נקבל H_0, בר"מ של 0.05 נכריע שתוחלת עובי הבורג מתים למפרט.
6) א'.
7) ג'.
8) א. לדחות. ב. לא ניתן לדעת.

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.

(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:	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
תנאים:	1. σ ידועה 2. $X \sim N$ או מדגם מספיק גדול		
p-value	אם $\bar{x} > \mu_0 \Leftarrow 2 \cdot P_{H_0}\left(\bar{X} \ge \bar{x}\right)$ 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>: μ.

<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: $\bar{x}=74$, $n=100$.
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. $H_0: \mu = 100$, $H_1: \mu < 100$. ב. 0.1056. ג. 0.1056.

ד. נכריע שיש עמידה בהתחייבות של החברה.

4) א. 0.0006. ב. יקטן. ג. נכריע שאין כיול.

5) נכון.

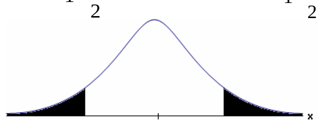
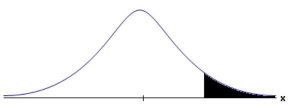
6) א'.

7) א'.

8) ג'.

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

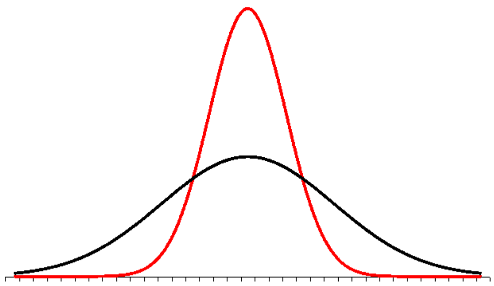
רקע:

השערת האפס: השערה אלטרנטיבה:	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
תנאים:	1. σ אינה ידועה 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 $t_{\bar{x}} < -t^{(n-1)}_{1-\frac{\alpha}{2}}$  $-t_{1-\frac{\alpha}{2},n-1}$ $t_{1-\frac{\alpha}{2},n-1}$ ■ - דוחים את H_0	$t_{\bar{x}} < -t^{(n-1)}_{1-\alpha}$  $-t_{1-\alpha,n-1}$ ■ - דוחים את H_0	$t_{\bar{x}} > t^{(n-1)}_{1-\alpha}$  $t_{1-\alpha,n-1}$ ■ - דוחים את H_0
חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:	$\bar{X} > \mu_0 + t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ או $\bar{X} < \mu_0 - t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - t^{n-1}_{1-\alpha} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t^{n-1}_{1-\alpha} \cdot \frac{S}{\sqrt{n}}$

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

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

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: $X \sim N\left(\mu,\sigma^2\right)$ 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.
σ^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 תוצאת 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:	$H_0: \mu = \mu_0$ $H_1: \mu \ne \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
תנאים:	1. σ אינה ידועה או 2. מדגם מספיק גדול $X \sim N$		
p-value	אם $\bar{x} > \mu_0 \Leftarrow 2 \cdot P_{H_0}\left(\bar{X} \ge \bar{x}\right)$ אם $\bar{x} < \mu_0 \Leftarrow 2 \cdot P_{H_0}\left(\bar{X} \le \bar{x}\right)$	$P_{H_0}\left(\bar{X} \le \bar{x}\right)$	$P_{H_0}\left(\bar{X} \ge \bar{x}\right)$

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

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

$$d.f = n-1$$

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:

<u>אוכלוסייה</u>: כלל הנסיעות לעבודה בדרך החלופית.

<u>משתנה</u>: X = זמן נסיעה בדקות.

<u>תנאים</u>: $X \sim N$.

<u>פרמטר</u>: μ.

א. <u>השערות</u>: $\begin{matrix} H_0 : \mu = 40 \\ H_1 : \mu < 40 \end{matrix}$.

ב. <u>תוצאות המדגם</u>:

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

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

$$S = \sqrt{23.4}$$

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

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

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

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

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

<u>מסקנה</u>: 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- α 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. $H_0: \mu=1000$, $H_1: \mu\neq 1000$. ב. $20\% \le P_v \le 50\%$.
ג. ברמת מובהקות של 5% לא נוכל לקבוע שקו הייצור אינו תקין.

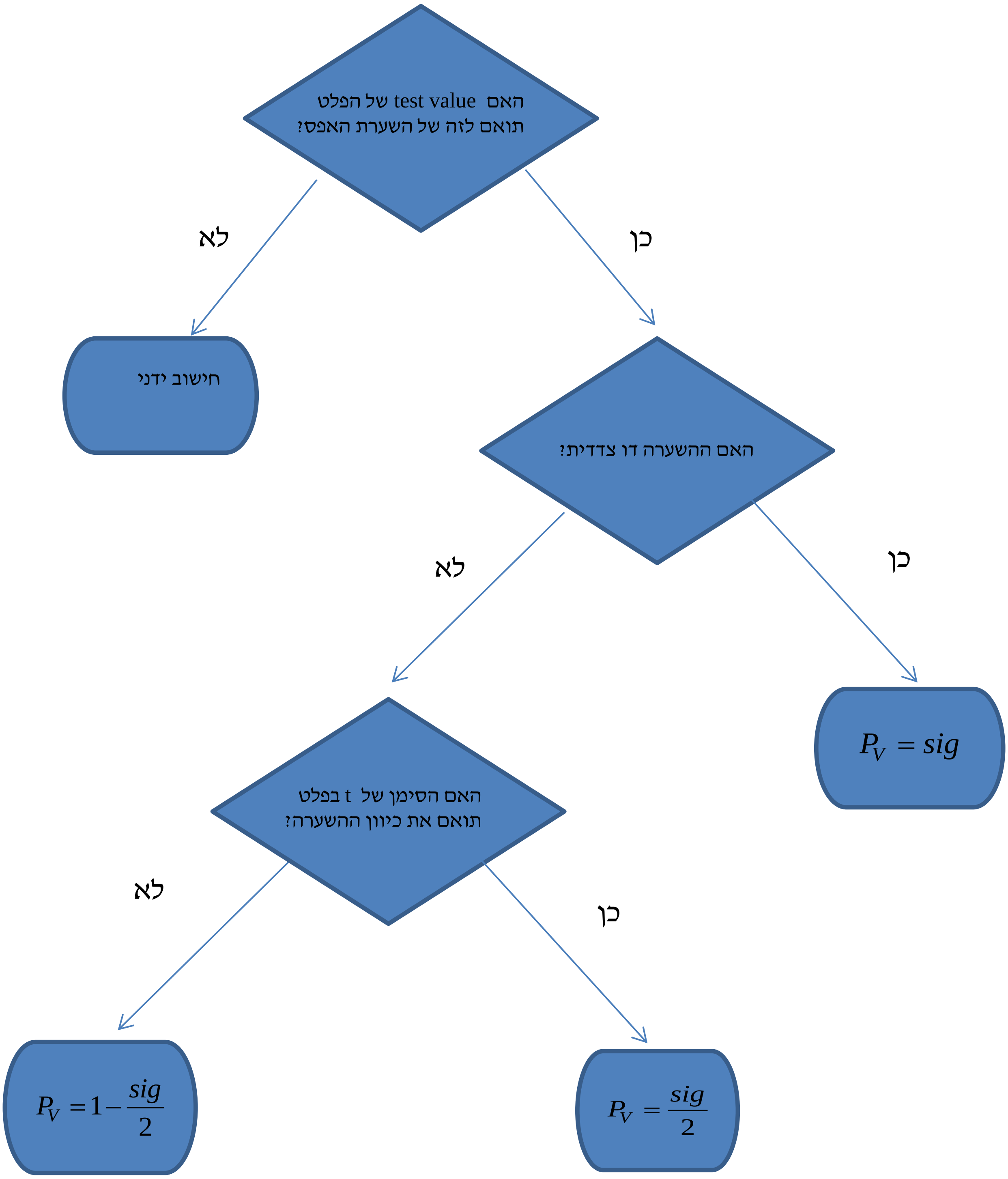
2) 10%.

3) 1.01, נקבל את H_0.

ניתוח פלטים:

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS :



דוגמה (פתרון בהקלטה):

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
X	25	87.6400	64.90434	12.98087

One-Sample Test

	Test Value = 60					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
X	2.129	24	.044	27.64000	.8488	54.4312

ממוצע הציונים במבחן המיצב בחשבון הוא 60. הוחלט לדגום כיתה אקראית של 25 תלמידים וללמד אותם בשיטת לימוד חדשה.

א. מהו רווח הסמך לממוצע הציונים בחשבון אם יוחלט ליישם את שיטת הלימוד החדשה?

ב. מהו P_V לבדיקת יעילותה של שיטת הלימוד החדשה?

ג. מה יוכרע ברמת מובהקות של 5% לגבי יעילותה של שיטת הלימוד החדשה?

שאלות:

1) באוניברסיטה גדולה גיל הסטודנטים לתואר ראשון מתפלג נורמאלית.
בעבר פורסם שהגיל הממוצע של הסטודנטים הינו 23. להלן פלט תוכנת SPSS על מדגם של 16 סטודנטים אקראיים מתואר ראשון:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
age	16	23.4375	2.50250	.62562

One-Sample Test

	Test Value = 23					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
age	.699	15	.495	.43750	-.8960	1.7710

א. מהו המבחן הסטטיסטי שנעשה כאן?
ב. מה ערכו של הפרמטר לפי השערת האפס?
ג. רשום רווח סמך ברמת סמך של 95% לתוחלת גיל הסטודנטים באוניברסיטה לתואר ראשון.
ד. בדוק ברמת מובהקות של 5% האם הגיל הממוצע כיום שונה מבעבר?

2) קבוצת ילדים בגיל 6 קיבלה משימה לביצוע. עבור כל ילד בדקו כמה זמן לקח לו לסיים את המשימה בדקות. להלן תוצאות הניתוח הסטטיסטי:

One-Sample Test

	Test Value = 4.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
time	-1.853	24	.076	-.09200	-.1944	.0104

א. כמה ילדים השתתפו במחקר?
ב. מצא רווח סמך ברמת סמך של 95% לתוחלת זמן ביצוע המשימה עבור ילדים בני 6.
ג. מה יש להניח כדי שרווח הסמך מסעיף א' יהיה תקף?
ד. בדוק ברמת מובהקות של 5% שזמן ביצוע המשימה הממוצע נמוך מ-4.5 דקות.

3) להלן פלט מחשב עבור ניתוח סטטיסטי שנעשה בתוכנת SPSS.
הניתוח הוא עבור מדגם אקראי של קבוצת נבחנים בבגרות באנגלית.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
grade	???	???	19.62787	2.95901

One-Sample Test

	Test Value = 75					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
grade	???	43	.017	-7.34091	-13.3083	????

א. השלימו את הגדלים החסרים המסומנים בסמני שאלה בפלט.
ב. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים שונה מ-75?
ג. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים קטנה מ-75?
ד. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים גדולה מ-75?

4) יצרן סיגריות מפרסם כי תוחלת הניקוטין בסיגריות שהוא מיצר קטנה מ-27 מ"ג. בבדיקה מקרית של 5 סיגריות מתוצרתו נמצאו כמויות הניקוטין הבאות: 21, 21, 20, 24, 22 מ"ג.
הניחו כי כמות הניקוטין בסיגריות מפולג נורמאלי.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
nicotine	5	21.6000	1.51658	.67823

One-Sample Test

	Test Value = 27					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
nicotine	-7.962	4	.001	-5.40000	-7.2831	-3.5169

א. האם ברמת מובהקות של 5% ניתן להסיק שיש אמת בפרסום?
ב. אם היינו מוסיפים עוד תצפית שערכה 20. כיצד הדבר היה משפיע על הערך Sig ועל המסקנה?
ג. בדקו האם ניתן להגיד שתוחלת רמת הניקוטין שונה מ-26 ברמת מובהקות של 5%.

תשובות סופיות:

1) א. הסקה של תוחלת אחת. ב. 23. ג. (22.104, 24.771).
ד. נקבל H_0.

2) א. 25. ב. (4.3056, 4.5104). ג. המשתנה מתפלג נורמלית.
ד. נדחה H_0.

3) א. $n=44$, $\bar{X}=67.66$, $t=-2.48$, $upper=-1.3735$.
ב. 0.017. ג. 0.0085. ד. 0.9915.

4) א. נכריע שיש אמת בפרסום. ב. המסקנה לא תשתנה.
ג. נכריע שהתוחלת שונה מ-26.

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

רקע:

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α על μ:

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

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

אם μ_0 נופל ברווח ← נקבל את H_0.

אם μ_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. $112.87\leq\mu\leq118.13$.

ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.

3) א. $191.8\leq\mu\leq200.2$. ב. נכריע שיש אמת בפרסום.

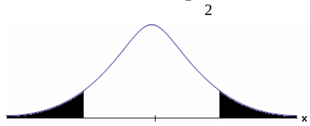
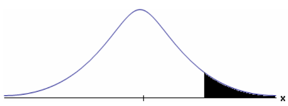
ביוסטטיסטיקה

פרק 31 - בדיקת השערות על פרופורציה

תוכן העניינים

התהליך:

רקע:

השערת האפס: השערה אלטרנטיבית:	$H_0: p = p_0$ $H_1: p \neq p_0$	$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p > p_0$
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 $Z_{\hat{p}} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ ■-דוחים את H_0	$Z_{\hat{p}} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ ■-דוחים את H_0	$Z_{\hat{p}} > Z_{1-\alpha}$  $Z_{1-\alpha}$ ■-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:		
$\hat{p} > p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{p_0(1-p_0)}{n}}$ $\hat{p} < p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\dfrac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\dfrac{p_0(1-p_0)}{n}}$

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 α_1 וחוקר ב' ברמת מובהקות α_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. $\begin{array}{l} H_0 : p = 0.2 \\ H_1 : p \neq 0.2 \end{array}$. ב. נקבל H_0. ג. המסקנה לא תשתנה.

5) א. נדחה H_0. ב. המסקנה לא תשתנה.

6) ג'.

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:	$H_0: p = p_0$ $H_1: p \ne p_0$	$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p > p_0$
תנאים:	$np_0 \ge 5 \,\&\, n(1-p_0) \ge 5$		
p-value	אם $\hat{p} > p_0 \Leftarrow 2 \cdot P_{H_0}\left(\hat{P} \ge \hat{p}\right)$ אם $\hat{p} < p_0 \Leftarrow 2 \cdot P_{H_0}\left(\hat{P} \le \hat{p}\right)$	$P_{H_0}\left(\hat{P} \le \hat{p}\right)$	$P_{H_0}\left(\hat{P} \ge \hat{p}\right)$

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.

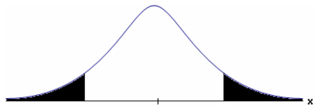
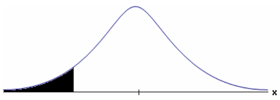
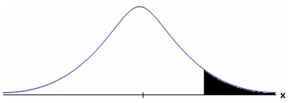
ביוסטטיסטיקה

פרק 32 - בדיקת השערות על הפרש פרופורציות

תוכן העניינים

בדיקת השערות על הפרש פרופורציות

רקע

השערת האפס השערה אלטרנטיבית:	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 \neq 0$	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 < 0$	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 > 0$
תנאים:	1.מדגמים בלתי תלויים	2. מדגמים גדולים	
כלל ההכרעה: אזור הדחייה של:	$Z_{\hat{p}_1-\hat{p}_2} < -Z_{1-\frac{\alpha}{2}}$ או $Z_{\hat{p}_1-\hat{p}_2} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ ■ - דוחים את H_0	$Z_{\hat{p}_1-\hat{p}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ ■ - דוחים את H_0	$Z_{\hat{p}_1-\hat{p}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ ■ - דוחים את H_0

סטטיסטי המבחן: $Z_{\hat{p}_1-\hat{p}_2} \underset{H_0}{=} \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}\hat{q}}{n_1} + \dfrac{\hat{p}\hat{q}}{n_2}}}$

כאשר הפרופורציה המשוקללת: $\hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

חלופה אחרת לכלל הכרעה:

כלל ההכרעה: אזור הדחייה של H_0	
$\hat{p}_1-\hat{p}_2>0+Z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}\ \hat{q}}{n_1}+\frac{\hat{p}\ \hat{q}}{n_2}}$ או $\hat{p}_1-\hat{p}_2<0-Z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n_1}+\frac{\hat{p}\ \hat{q}}{n_2}}$	$\hat{p}_1-\hat{p}_2<0-Z_{1-\alpha}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n_1}+\frac{\hat{p}\ \hat{q}}{n_2}}$
$\hat{p}_1-\hat{p}_2>0+Z_{1-\alpha}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n_1}+\frac{\hat{p}\ \hat{q}}{n_2}}$	

התפלגות של $\hat{p}_1-\hat{p}_2$: $\hat{p}_1-\hat{p}_2\sim N(p_1-p_2,\frac{p_1\cdot q_1}{n_1}+\frac{p_2\cdot q_2}{n_2})$

תקנון:

$$Z_{\hat{p}_1-\hat{p}_2}=\frac{\hat{p}_1-\hat{p}_2-(p_1-p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

$$Z_{\hat{p}_1-\hat{p}_2}\underset{H_0}{=}\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1}+\frac{\hat{p}\hat{q}}{n_2}}}$$

דוגמה (פתרון בהקלטה) :

נדגמו 80 סטודנטים שנבחנו במיקרו-כלכלה. מתוכם 60 עברו את הבחינה.
נדגמו 100 סטודנטים שנבחנו בסטטיסטיקה א'. מתוכם 82 עברו את הבחינה.
האם שיעור העוברים את הבחינה בסטטיסטיקה גבוה מאשר מהבחינה במיקרו כלכלה? בדקו ברמת מבוהקות של 10%.

שאלות

1) במדגם של 200 גברים, 8% היו מובטלים. במדגם של 180 נשים, 10% מהן היו מובטלות.
האם קיים הבדל מובהק בין פרופורציית המובטלים לפרופורציית המובטלות?
בדקו ברמת מובהקות של 5%.

2) אחוז בעלי רישיון נהיגה בקרב האוכלוסייה הבוגרת הינו 60%. במדגם של 300 בוגרים מתל אביב 204 היו בעלי רישיון נהיגה.
במדגם של 220 בוגרים מירושלים 100 היו בעלי רישיון נהיגה.
א. ברמת מובהקות של 5% האם תקבלו את הטענה שאחוז בעלי הרישיון בתל אביב גבוה מהאחוז הארצי?
ב. ברמת מובהקות של 10% האם תקבלו את הטענה שאחוז בעלי הרישיון נהיגה בתל אביב גבוה מאחוז בעלי רישיון הנהיגה בירושלים?

3) נדגמו 500 בוגרים מתוכם 200 גברים והיתר נשים. במדגם התקבל: מתוך הגברים ל-48% תעודת בגרות. מתוך הנשים ל-58% תעודת בגרות. מטרת המחקר היא לבדוק האם שיעור הזכאיות לבגרות גבוה משיעור הזכאים.
א. מהי מובהקות התוצאה?
ב. מה תהיה המסקנה ברמת מובהקות של 8%?

4) במדגם שנערך על 100 פרות מחוות בדרום הארץ התקבל כי 20 פרות נושאות וירוס מסוים. במדגם שנערך על 200 פרות מחוות בצפון הארץ התקבל כי 10 מתוכן נושאות וירוס גם כן.
א. בנו מבחן ברמת מובהקות של 5% לבדיקת הטענה כי הווירוס תקף את פרות הדרום באופן משמעותי יותר מאשר את הפרות בצפון הארץ.
ב. מהי המסקנה לבדיקת הטענה של סעיף א ומהי הטעות האפשרית במסקנה?
ג. מהי עוצמת המבחן אם שיעור הפרות בדרום עם הווירוס גבוה ב-10% משיעור הפרות בצפון עם הווירוס?
ד. כיצד העוצמה תשתנה אם נגדיל את רמת המובהקות?

תשובות סופיות

1) לא נדחה את H_0.

2) א. נדחה H_0. ב. נדחה H_0.

3) א. 0.0139 ב. נדחה H_0.

4) א. ראה סרטון. ב. נדחה H_0. ג. 0.8238
ד. תגדל.

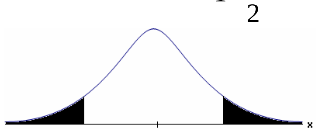
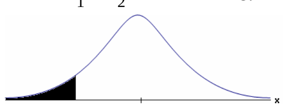
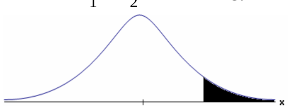
ביוסטטיסטיקה

פרק 33 - 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:	$H_0 \quad \mu_1-\mu_2=c$ $H_1 \quad \mu_1-\mu_2\neq c$	$H_0 \quad \mu_1-\mu_2=c$ $H_1 \quad \mu_1-\mu_2<c$	$H_0 \quad \mu_1-\mu_2=c$ $H_1 \quad \mu_1-\mu_2>c$
תנאים:	מדגמים בלתי תלויים σ_1,σ_2 ידועות $X_1,X_2\sim N$ או מדגמים מספיק גדולים		
כלל ההכרעה: אזור הדחייה של H_0: $-Z_{1-\frac{\alpha}{2}}$ $\quad$ $Z_{1-\frac{\alpha}{2}}$	$Z_{\bar{x}_1-\bar{x}_2}<-Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}_1-\bar{x}_2}>Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $\quad$ $Z_{1-\frac{\alpha}{2}}$ ■-דוחים את H_0	$Z_{\bar{x}_1-\bar{x}_2}<-Z_{1-\alpha}$  $-Z_{1-\alpha}$ ■-דוחים את H_0	$Z_{\bar{x}_1-\bar{x}_2}>Z_{1-\alpha}$  $Z_{1-\alpha}$ ■-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:

<table>
<tr><th colspan="2">נדחה H_0 אם מתקיים :</th></tr>
<tr><td>$\bar{x}_1-\bar{x}_2>c+Z_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
או
$\bar{x}_1-\bar{x}_2<c-Z_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$</td><td>$\bar{x}_1-\bar{x}_2>c+Z_{1-\alpha}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$</td></tr>
<tr><td colspan="2">$\bar{x}_1-\bar{x}_2<c-Z_{1-\alpha}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$</td></tr>
</table>

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

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%.

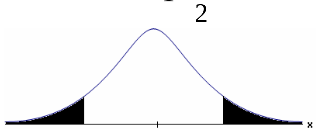
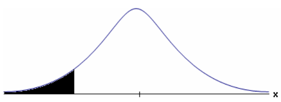
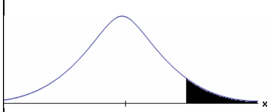
תשובות סופיות

1) נדחה H_0.

2) לא נדחה את H_0.

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

השערת האפס: השערה אלטרנטיבית:	$H_0 \quad \mu_1-\mu_2=c$ $H_1 \quad \mu_1-\mu_2\neq c$	$H_0 \quad \mu_1-\mu_2=c$ $H_1 \quad \mu_1-\mu_2<c$	$H_0 \quad \mu_1-\mu_2=c$ $H_1 \quad \mu_1-\mu_2>c$
תנאים:	1. מדגמים בלתי תלויים 2. σ_1, σ_2 לא ידועות אך שוות 3. המשתנים בכל אוכלוסייה מתפלגים נורמלית		
אזור הדחייה של H_0:	$t_{\bar{x}_1-\bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ או $t_{\bar{x}_1-\bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  $-t_{1-\frac{\alpha}{2},n_1+n_2-2}$ $\quad t_{1-\frac{\alpha}{2},n_1+n_2-2}$ ■-דוחים את H_0	$t_{\bar{x}_1-\bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$  $-t_{1-\alpha,n_1+n_2-2}$ ■-דוחים את H_0	$t_{\bar{x}_1-\bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  $t_{1-\alpha,n_1+n_2-2}$ ■-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:

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

דוגמה (פתרון בהקלטה) :

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

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

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

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

שאלות

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

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.

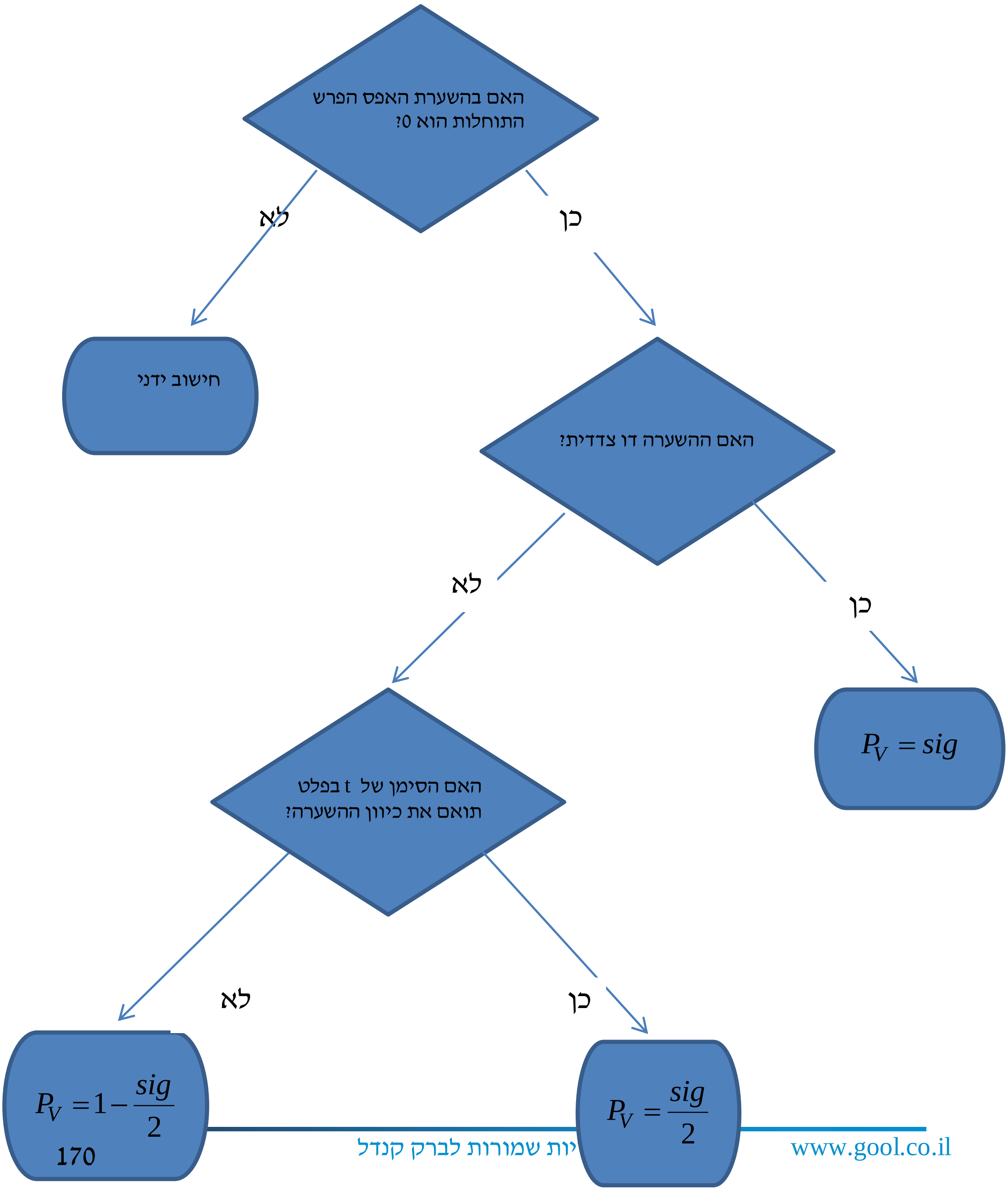
3) א. נדחה את H_0. ב. רמת מובהקות של לפחות 5%.

ג. לא נדחה את H_0.

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

ניתוח פלטים – רקע

מובהקות התוצאה על סמך הפלט :



דוגמה (פתרון בהקלטה):

בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים משני אזורים שונים במדינה על מס' האחים והאחיות שלהם. להלן הפלט שהתקבל:

Group Statistics

	Region of the United States	N	Mean	Std. Deviation	Std. Error Mean
Number of Brothers and Sisters	North East	676	3.76	2.939	.113
	South East	410	4.05	2.993	.148

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Number of Brothers and Sisters	Equal variances assumed	.173	.677	-1.583	1084	.114	-.293	.185	-.657	.070
	Equal variances not assumed			-1.576	850.945	.115	-.293	.186	-.658	.072

א. מהו המבחן הסטטיסטי שנעשה כאן?

ב. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?

ג. בדוק האם קיים הבדל בין "South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.

ד. מהי מובהקות התוצאה לבדיקת הטענה שההפרש הממוצע בין "South East" לבין "North East" חיובי?

שאלות

1) להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים. רמת האופטימיות נמדדה בסולם ציונים של 1 עד 5.

Group Statistics

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
optimizm	MALE	633	2.6053	.49781	.01979
	FEMALE	568	2.5503	.48483	.02034

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	1190.977	.053	.05500	.02838	-.00068	.11067

א. האם ניתן להניח ששונות האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?

ב. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?

ג. מצא את הגבול העליון של רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.

ד. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

2) פסיכולוגים טוענים שאנשים שניגשים למבחן אינטליגנציה יותר מפעם אחת נוטים לקבל ציונים גבוהים יותר. להלן הפלט שהתקבל:

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
grade	A	9	96.8889	9.40006	3.13335
	B	11	108.4545	11.46616	3.45718

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
grade	Equal variances assumed	.206	.656	-2.428	18	.026	-11.56566	4.76333	-21.57304	-1.55828
	Equal variances not assumed			-2.479	17.997	.023	-11.56566	4.66583	-21.36832	-1.76299

T-Test

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.

ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?

ג. האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של $\alpha = 0.05$.

ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha = 0.01$.

3) כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל בממוצע גיל הנישואין (הראשונים), בין נשים הגרות בערים מרכזיות לבין נשים הגרות בערים מרוחקות מהמרכז.
לשם כך נדגמו 50 כלות מכל אחת משתי ערים עיר א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שלהלן:

T-Test

Group Statistics

	מקום המגורים	N	Mean	Std. Deviation	Std. Error Mean
גיל הנישואין'	עיר א	50	24.8072	1.38978	.19654
	עיר ב	50	23.0131	1.62070	.22920

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
גיל הנישואין'	Equal variances assumed	.330	.567	5.942	98	.000	1.79415	.30193	1.19497	2.39332
	Equal variances not assumed			5.942	95.772	.000	1.79415	.30193	1.19480	2.39350

א. מהו המבחן הסטטיסטי שנעשה כאן?
ב. מצא רווח סמך ברמת סמך של 95% להפרש בין עיר א לעיר ב מבחינת גיל הנשים הממוצע בנישואין הראשונים.
ג. האם ניתן לומר ברמת מובהקות של 1% שנשים בערים מרכזיות מתחתנות בגיל מאוחר יותר מאשר נשים הגרות בערים מרוחקות?

4) להלן פלט של תוכנת SPSS :

T-Test

	N	Mean	Std. Deviation	Std. Error Mean
X	26	36.3077	13.23259	2.59513
Y	24	46.4583	20.96369	4.27920

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-ailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	4.446	.040	-2.164	???	.044	-10.15064	???	-20.03781	-.26347
Equal variances not assumed			-2.038	38.267	.048	???	5.00462	-20.27964	-.02164

א. השלימו את סימני השאלה בטבלה.
ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של X לזה של Y?
ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X גדולה מהתוחלת של Y?
ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X קטנה מהתוחלת של Y?

תשובות סופיות

1) א. נקבל את H_0 ונכריע שיש שוויון שוניות.

ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.

ג. 0.11076 ד. $2.5665 \le \mu \le 2.6441$.

2) א. מבחן T להפרש ממוצעים במדגמים בלתי תלויים.

ב. מבחן T למדגמים מזווגים.

ג. נקבל את H_0, נקבע לקיום שוויון שוניות.

ד. נקבל את H_0, לא נקבל את טענת הפסיכולוגים.

3) א. מבחן T להשוואת תוחלת במדגמים בלתי תלויים.

ב. $1.19497 \le \mu_1 - \mu_2 \le 2.39332$ ג. כן.

4) א. 10.15, 4.69,-48 ב. 0.04

ג. 0.978 ד. 0.022.

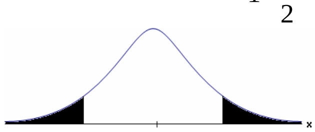
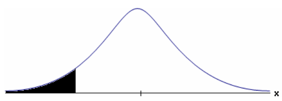
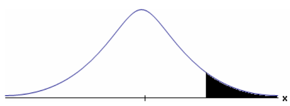
ביוסטטיסטיקה

פרק 34 - 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:	$H_0:\ \mu_D = C$ $H_1:\ \mu_D \neq C$	$H_0:\ \mu_D = C$ $H_1:\ \mu_D < C$	$H_0:\ \mu_D = C$ $H_1:\ \mu_D > C$
תנאים:	1. σ_D אינה ידועה 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 $t_{\bar{D}} < -t^{(n-1)}_{1-\frac{\alpha}{2}}$  $-t_{1-\frac{\alpha}{2},n-1}$ $t_{1-\frac{\alpha}{2},n-1}$ ■ - דוחים את H_0	$t_{\bar{D}} < -t^{(n-1)}_{1-\alpha}$  $-t_{1-\alpha,n-1}$ ■ - דוחים את H_0	$t_{\bar{D}} > t^{(n-1)}_{1-\alpha}$  $t_{1-\alpha,n-1}$ ■ - דוחים את H_0
חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:	$\bar{D} > C + t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{S_D}{\sqrt{n}}$ א ו $\bar{D} < C - t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t^{n-1}_{1-\alpha} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t^{n-1}_{1-\alpha} \cdot \frac{S_D}{\sqrt{n}}$

סטטיסטי המבחן: $t_{\bar{D}} = \dfrac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$, $S_D^{\ 2} = \dfrac{\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}{n-1} = \dfrac{\sum_{i=1}^{n} D^2_{\ i} - n\bar{D}^2}{n-1}$

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	2.5	0.7	9.6	4.5
משקל במכשיר 2	0.5	1.7	6.9	3.5

נניח שהמשקלים מתפלגים נורמלית, המבחן שיש לבצע כאן הוא:

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

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:

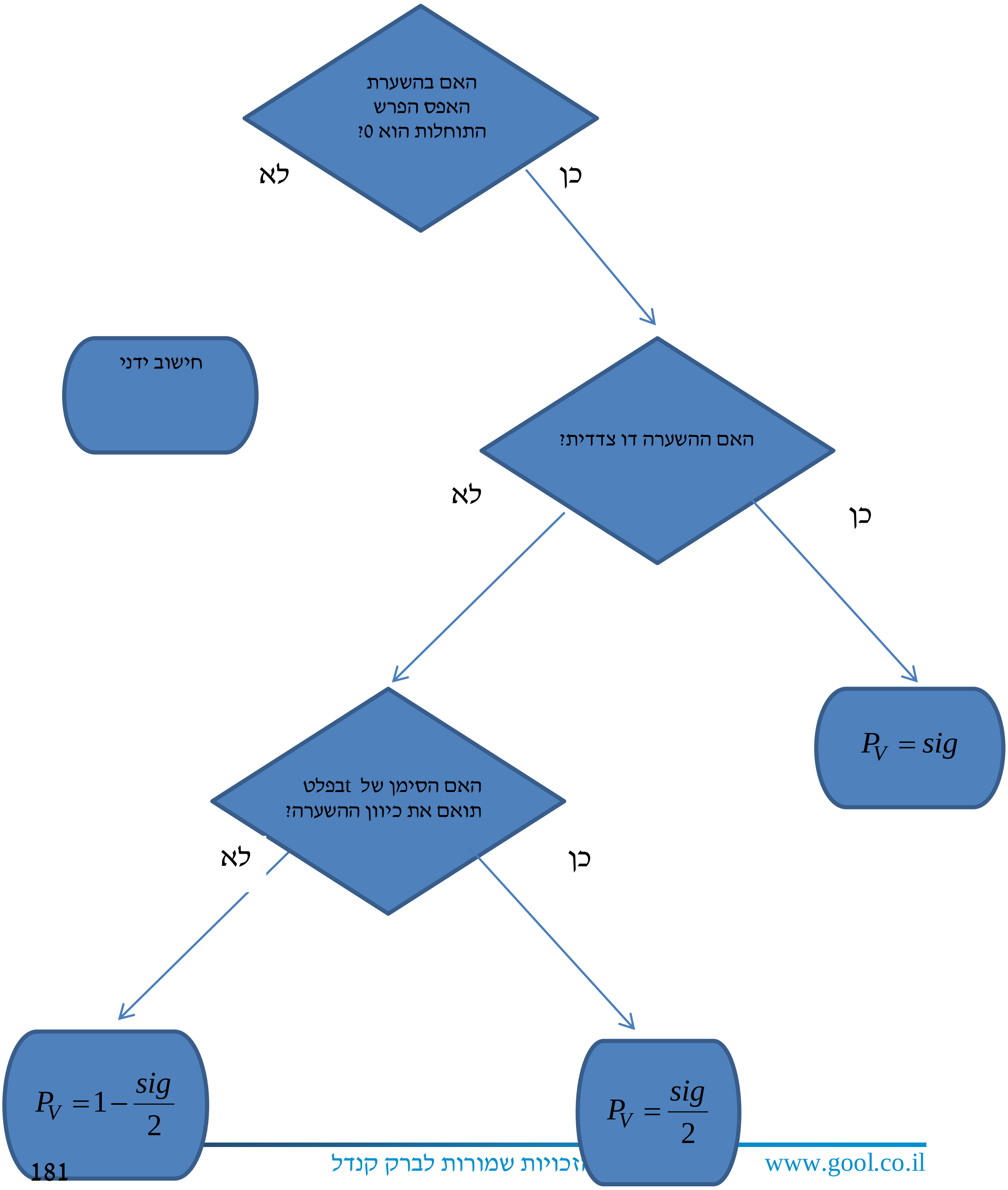
א. מבחן Z למדגם יחיד.
ב. מבחן T למדגם יחיד.
ג. מבחן T למדגמים בלתי תלויים.
ד. מבחן T למדגמים מזווגים.

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'.

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

מדגמים מזווגים – ניתוח פלטים – רקע



דוגמה (פתרון בהקלטה) :

כדי לבדוק את ההשפעה של קורס לגמילה מעישון נלקח מדגם מקרי של 5 נבדקים. עבור כל אחד מהם נמדדה צריכת הסיגריות היומית לפני הקורס וחודשיים אחריו. הניחו שצריכת הסיגריות מתפלגת נורמלית. להלן התוצאות :

נבדק	1	2	3	4	5
לפני	40	22	25	28	30
אחרי	30	24	13	10	12

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	BEFORE	29.0000	5	6.85565	3.06594
	AFTER	17.8000	5	8.72926	3.90384

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					90% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	BEFORE - AFTER	11.20000	8.19756	3.66606	3.38452	19.01548	3.055	4	.038

בדקו ברמת מובהקות של 5% האם הקורס יעיל.

שאלות

1) בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים על השכלת הוריהם, להלן הפלט שהתקבל:

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Highest Year School Completed, Father - Highest Year School Completed, Mother	-.007	3.115	.100	-.203	.189	-.072	973	.943

א. תנו אומדן להפרש הממוצעים.
ב. תנו אומדן לטעות התקן של הפרש הממוצעים.
ג. האם קיים הבדל מובהק בין השכלת האבות להשכלת האימהות ברמת מובהקות של 5%?

2) בתחרות קפיצה למים שופטים באופן קבוע שופט איטלקי ושופט דרום קוריאני. להלן פלט המנתח את הציונים ששופטים אלה נתנו בתחרויות השונות:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Italy	???	300	.86742	.05008
	South Korea	8.9183	???	.81992	.04734

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Italy - South Korea	-.42233	.36153	.02087	-.46341	-.38126	-20.234	???	???

א. השלימו את החלקים החסרים בפלט (מסומנים בסימני שאלה).
ב. בדקו את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.
ג. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים?
ד. בנו את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.

3) בדקו את ציוניהם של 44 נבדקים אקראיים במבחן הפסיכומטרי. פעם אחת לפני הכנה (Before) ופעם אחת אחרי הכנה (After).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Before - After	-7.45455	19.28303	2.90703	-13.31712	-1.59197	-2.564	43	.014

א. רשמו מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.

ב. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.

ג. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.

ד. מצאו רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

4) להלן פלט של תכנת SPSS:

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	x	54.0000	6	5.86515	2.39444
	y	46.5000	6	10.72847	4.37988

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	x - y	7.50000	??	4.72405	-4.64356	19.64356	??	5	.173

א. מלא את החלקים החסרים בטבלה.
ב. מהי רמת המובהקות המינימלית לקבלת הטענה שיש הבדל בין X ל- Y בממוצע?
ג. האם התשובה לסעיף הקודם הייתה משתנה, ואם כן גדלה או קטנה, אם היינו מוסיפים עוד תצפית שההפרש בין X ל- Y הוא 0.
ד. מהי מובהקות התוצאה לבדיקת הטענה ש X גדול מ- Y בממוצע?
ה. מהי מובהקות התוצאה לבדיקת הטענה ש X קטן מ- Y בממוצע?
ו. בנו רווח סמך לתוחלת של X ברמת סמך של 90%.

תשובות סופיות

1) א. -0.007 ב. 0.1 ג. אין הבדל מובהק.

2) א. $d.f = 299$, $n = 300$, $\bar{X} = 8.496$, $\text{Sig} = 0$.
ב. נדחה את H_0. ג. $(0.46341, 0.38126)$ ד. $(0.4708, 0.3738)$.

3) א. ראה וידאו. ב. נדחה את H_0. ג. לא נדחה את H_0.
ד. (1.592,13.317).

4) א. 1.5876, 11.5715 ב. 0.173 ג. יגדל.
ד. 0.0865 ה. 0.9135 ו. $49.18 < \mu < 58.82$.

ביוסטטיסטיקה

פרק 35 - 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 α 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 נק' לעומת סמסטר א'. האם יש אמת בפרסום?

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

מרצה X	82	90	68
מרצה Y	68	81	64

א. חשבן 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'.

ביוסטטיסטיקה

פרק 36 - שאלות מסכמות בבדיקת השערות

תוכן העניינים

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. $H_1: \mu_d < 0, \quad H_0: \mu_d = 0$
ב. $H_1: \mu_X - \mu_Y < 0, \quad H_0: \mu_X - \mu_Y = 0$
ג. $H_1: \mu_X - \mu_Y < 0, \quad H_0: \mu_X - \mu_Y = 0$
ד. $H_1: \mu_d > 0, \quad H_0: \mu_d = 0$

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 מהתושבות 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 χ^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 α 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. σ^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:

	α	$1-\beta$
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 α, והן $(1-\beta)$, יקטנו.
ב. α יישאר ללא שינוי ואילו $(1-\beta)$ יגדל.
ג. α יגדל ואילו $(1-\beta)$ יקטן.
ד. הן α 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 ש- $p-value=0.02$.
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.
σ^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 **<u>במציאות</u>** השערתה של החוקרת נכונה, <u>סביר</u> כי אם היא תגדיל את גודל המדגם, אזי:

א. יקטן הסיכוי לטעות מסוג שני (β).

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 α. במידה וחוקר זה היה בודק השערה זו ברמת מובהקות של 2α 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 σ^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

ביוסטטיסטיקה

פרק 37 - מבחני חי בריבוע

תוכן העניינים

מבחני חי בריבוע

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:

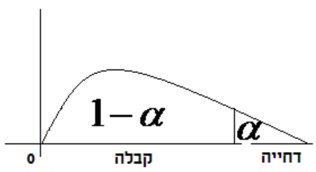
השערות:

H_0 - המשתנה מתפלג לפי התפלגות מסוימת.

H_1 - אחרת.

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.

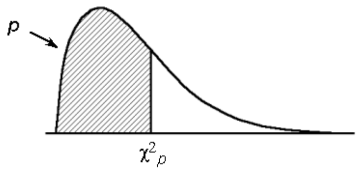
הערה:

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	$0.0^{4}393$	$0.0^{3}157$	$0.0^{3}982$	$0.0^{2}393$	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 χ^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) להלן השערות מחקר: $H_1: else$, $H_0: X \sim N(40,2^2)$.

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) 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.

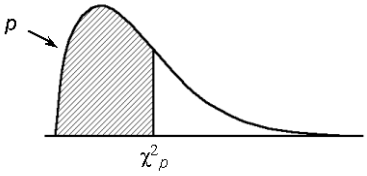
דוגמה (הפתרון בהקלטה):

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 $\chi^2 p$



	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 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 χ^2 (chi-square) השווה לאפס. הסטודנט הסיק כי לא קיימת תלות בין שני המשתנים שבדק, 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

פלטים על מבחן לאי תלות – 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.

דרגות חופש: $d.f=(r-1)(c-1)$.

כאשר: r - מספר הקטגוריות של המשתנה שבשורות.

c - ספר הקטגוריות של המשתנה שבעמודות.

סטטיסטי המבחן: $\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}$.

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% מהתאים E קטן מ-5.

דוגמה (פתרון בהקלטה) :

במחקר רצו לבדוק את הקשר בין צבע שיער לבין צבע עיניים של אנשים.
הפלטים שהתקבלו מצורפים.

א. מהו המבחן הסטטיסטי המתאים?
ב. כמה קטגוריות יש לכל משתנה?
ג. רשמו את השערות המחקר.
ד. מה מספר דרגות החופש?
ה. כמה אנשים במדגם נמצאו עם שיער חום?
ו. כמה אנשים היית מצפה במדגם שיהיה להם שיער חום ועיניים ירוקות בהנחה ואין קשר בין צבע שיער לצבע עיניים?
ז. מתוך הבלונדינים מה אחוז בעלי עיניים כחולות במדגם?
ח. מתוך בעלי עיניים ירוקות מה אחוז הבלונדינים במדגם?
ט. מה ערכו של סטטיסטי המבחן ומהי מובהקות התוצאה?
י. מה מסקנת המחקר? $\alpha = 5\%$

להלן הפלטים שהתקבלו :

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
hair_color * eye_color	78	100.0%	0	0.0%	78	100.0%

hair_color * eye_color Crosstabulation

			eye_color			Total
			brown	green	Blue	
hair_color	black	Count	13	7	7	27
		Expected Count	10.7	8.3	8.0	27.0
		% within hair_color	48.1%	25.9%	25.9%	100.0%
		% within eye_color	41.9%	29.2%	30.4%	34.6%
	brown	Count	12	12	6	30
		Expected Count	11.9	9.2	8.8	30.0
		% within hair_color	40.0%	40.0%	20.0%	100.0%
		% within eye_color	38.7%	50.0%	26.1%	38.5%
	blond	Count	6	5	10	21
		Expected Count	8.3	6.5	6.2	21.0
		% within hair_color	28.6%	23.8%	47.6%	100.0%
		% within eye_color	19.4%	20.8%	43.5%	26.9%
Total		Count	31	24	23	78
		Expected Count	31.0	24.0	23.0	78.0
		% within hair_color	39.7%	30.8%	29.5%	100.0%
		% within eye_color	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.880[a]	4	.208
Likelihood Ratio	5.641	4	.228
Linear-by-Linear Association	2.682	1	.101
N of Valid Cases	78		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.19.

שאלות

1) בסקר שנעשה על ידי משרד ראש הממשלה נדגמו 60 אזרחים.
כל אזרח נשאל על מגדרו והאם הוא בעד הקמת מדינה פלסטינית.

א. מה ההשערות הנבדקות ומהו סטטיסטי המבחן?

ב. אם סטטיסטי המבחן היה גדל כיצד הדבר היה משפיע על SIG שבפלט.

ג. האם קיים קשר בין מגדר ודעה ברמת מובהקות של 5%?

ד. מהו האומדן לאחוז התומכים במדינה פלסטינית מתוך הגברים?

ה. איזה אחוז מהנשאלים שהיו בעד מדינה פלשתינית הם גברים?

להלן הפלטים :

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.973[a]	2	.373
Likelihood Ratio	1.987	2	.370
Linear-by-Linear Association	1.882	1	.170
N of Valid Cases	60		

a. 0 cells (.0%) have expected count less than 5.

b. The minimum expected count is 7.25.

gender * opinion Crosstabulation

			opinion			Total
			yes	now	no opinion	
gender	Male	Count	10	10	9	29
		Expected Count	12.6	9.2	7.3	29.0
		% within gender	34.5%	34.5%	31.0%	100.0%
		% within opinion	38.5%	52.6%	60.0%	48.3%
		% of Total	16.7%	16.7%	15.0%	48.3%
	female	Count	16	9	6	31
		Expected Count	13.4	9.8	7.8	31.0
		% within gender	51.6%	29.0%	19.4%	100.0%
		% within opinion	61.5%	47.4%	40.0%	51.7%
		% of Total	26.7%	15.0%	10.0%	51.7%
Total		Count	26	19	15	60
		Expected Count	26.0	19.0	15.0	60.0
		% within gender	43.3%	31.7%	25.0%	100.0%
		% within opinion	100.0%	100.0%	100.0%	100.0%
		% of Total	43.3%	31.7%	25.0%	100.0%

2) להלן פלט על סמך סקר שנעשה בקרב סטודנטים, בסקר נשאלו הסטודנטים על המוזיקה אותה הם מעדיפים וצורת הבילוי המועדפת עליהם.

Crosstab

Count

		בילוי			Total
		קריאה	ספורט	מועדון	
מוזיקה	רוק	0	0	11	11
	פופ	1	6	8	15
	קלאסי	5	6	9	20
Total		6	12	28	46

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.929[a]	?	.018
N of Valid Cases	46		

a. 5 cells (55.6%) have expected count less than 5.

b. The minimum expected count is 1.43.

א. בין אלו משתנים נבדק הקשר? כמה קטגוריות לכל משתנה?

ב. האם התנאים של המודל מתקיימים?

ג. מה מספר דרגות החופש במבחן הנ"ל?

ד. מה ההשערות של המבחן?

3) מחקר התעניין לבדוק את הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות. הפלטים מצורפים.

א. השלימו את שלושת המספרים החסרים בטבלה (היכן שיש סימני שאלה).

ב. מה ערכו של חי בריבוע הסטטיסטי.

ג. תנו הערכה למובהקות התוצאה לבדיקת הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות.

רמת_הכנסה * צרכן עגבניות Crosstabulation

			צרכן עגבניות		Total
			אורגני	לא אורגני	
רמת_הכנסה	הרבה מתחת לממוצע	Count	17	42	59
		% within רמת_הכנסה	28.8%	?	100.0%
		% within צרכן עגבניות	13.6%	33.6%	23.6%
	מתחת לממוצע	Count	27	22	49
		% within רמת_הכנסה	55.1%	44.9%	100.0%
		% within צרכן עגבניות	?	17.6%	19.6%
	ממוצע	Count	31	29	60
		% within רמת_הכנסה	51.7%	48.3%	100.0%
		% within צרכן עגבניות	24.8%	23.2%	24.0%
	מעל הממוצע	Count	44	26	70
		% within רמת_הכנסה	62.9%	37.1%	100.0%
		% within צרכן עגבניות	35.2%	20.8%	28.0%
	הרבה מעל הממוצע	Count	?	6	12
		% within רמת_הכנסה	50.0%	50.0%	100.0%
		% within צרכן עגבניות	4.8%	4.8%	4.8%
Total		Count	125	125	250

4) חוקר בדק את הקשר בין צבע השיער לבין צבע העיניים בעזרת מבחן חי בריבוע בקרב 52 נבדקים. תוצאות המבחן מוצגות בטבלה. בנוסף ידוע כי סטטיסטי המבחן שהתקבל מעיבוד הנתונים הוא 8.08.

א. מה תהיה מסקנת המחקר ברמת מובהקות של 1%?
ב. מה ערכו של E עבור עיניים כחולות וצבע שיער כהה.
ג. מה יהיה בקירוב ערכו של מקדם המתאם של קרמר?
ד. מהי פרופורציית בעלי צבע השיער הבהיר מקרב בעלי העיניים הירוקות?

להלן הפלט:

צבע עיניים * צבע שיער Crosstabulation

			צבע שיער		Total
			כהה	בהיר	
צבע עיניים	כחול	Count			
		% within צבע עיניים	50.0%	50.0%	100.0%
		% within צבע שיער	21.6%	53.3%	30.8%
		% of Total	15.4%	15.4%	30.8%
	חום	Count			
		% within צבע עיניים	83.3%	16.7%	100.0%
		% within צבע שיער	27.0%	13.3%	23.1%
		% of Total	19.2%	3.8%	23.1%
	ירוק	Count			
		% within צבע עיניים	79.2%	20.8%	100.0%
		% within צבע שיער	51.4%	33.3%	46.2%
		% of Total	36.5%	9.6%	46.2%
Total		Count			
		% within צבע עיניים	71.2%	28.8%	100.0%
		% within צבע שיער	100.0%	100.0%	100.0%
		% of Total	71.2%	28.8%	100.0%

5) במחקר מסוים רצו לבדוק האם יש קשר בין המגדר להוצאה על לבוש במשך שנה. דגמו באופן מקרי גברים ונשים ובדקו את רמת ההוצאה שלהם על לבוש בשנה האחרונה. חוקר א' בדק האם קיים הבדל בתוחלות ההוצאה בין גברים לנשים. חוקר ב' קיבץ את ההוצאה לקטגוריות ובאופן הזה בדק האם קיים הבדל בהתפלגות ההוצאה בין גברים לנשים.
הקטגוריות חולקו לשלוש קבוצות הוצאה.

א. איזה פלט מתאים לאיזה אחד מהחוקרים? נמקו.
ב. מה מסקנתו של חוקר א'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ותנו מסקנה במונחי המשתנים).
ג. איזו טעות יכולה להיות במסקנתו של חוקר א'? נסחו את הטעות במונחי השאלה.
ד. מהי מסקנתו של חוקר ב'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ורשמו מסקנה במונחי המשתנים).
ה. איזו טעות יכולה להיות במסקנתו של חוקר ב'? נסחו זאת במונחי השאלה.
ו. כיצד ניתן ליישב את מסקנות שני החוקרים?

להלן פלט ראשון:

T-Test

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
dimension1	female	40	2.9000	1.15025	.18187
	male	40	2.6000	2.52982	.40000

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
expese	Equal variances assumed	16.805	.000	.683	78	.497	.30000	.43941	-.57479	1.17479
	Equal variances not assumed			.683	54.464	.498	.30000	.43941	-.58078	1.18078

להלן פלט שני:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * category	80	100.0%	0	.0%	80	100.0%

gender * category Crosstabulation

			category			Total
			a	b	c	
gender	Female	Count	2	30	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	5.0%	75.0%	20.0%	100.0%
		% within category	9.1%	71.4%	50.0%	50.0%
	Male	Count	20	12	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	50.0%	30.0%	20.0%	100.0%
		% within category	90.9%	28.6%	50.0%	50.0%
Total		Count	22	42	16	80
		Expected Count	22.0	42.0	16.0	80.0
		% within gender	27.5%	52.5%	20.0%	100.0%
		% within category	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.442[a]	2	.000
Likelihood Ratio	25.064	2	.000
N of Valid Cases	80		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.00.

תשובות סופיות

1) א. 1.973 ב. קטן . ג. לא נדחה H_0 .
ד. 34.5% ה. 38.5%.

2) א. בילוי מועדף ומוזיקה מועדפת עם 3 קטגוריות לכל משתנה.
ב. לא ג. 4.
ד. H_0 אין תלות בין בילוי מועדף למוזיקה מועדפת.
H_1 יש תלות בין בילוי מועדף למוזיקה מועדפת.

3) א. 71.2%, 21.6%, 6 ב. 15.8 ג. קטן מ-0.005.

4) א. לא נדחה H_0 . ב. 11.4 ג. 0.394
ד. 20.8%.

5) א. פלא א'- חוקר א', פלא ב'- חוקר ב'. ב. נקבל את H_0 .
ג. טעות מסוג שני- הכרענו שאין הבדל בין גברים לנשים למרות שיש במציאות הבדל.
ד. נקבל את H_1 .
ה. טעות מסוג ראשון- הכרענו שיש קשר בין מין להוצאה למרות שבמציאות אין קשר.
ו. כל חוקר פעל בשיטה סטטיסטית שונה ובמצב כזה יתכן מסקנות סותרות.

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

מדדי קשר-מדד הקשר של קרמר – רקע

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

דוגמה (פתרון בהקלטה) :

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

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

בהקשר של קרמר הטבלה נקראת טבלת O (observed)

X - מין (גבר/אישה) – סולם שמי.

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

שלבים בחישוב r_c :

שלב א׳:

נבנה את טבלת E (Expected)

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

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

שלב ב׳:

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

שלב ג׳:

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

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

שאלות

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

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

האם קיים קשר בין מין להשכלה? נמקו!

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

א. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.

ב. האם קיים קשר בין פעילות גופנית למצב בריאותי?
חשבו לפי מדד הקשר של קרמר.

תשובות סופיות

1) קיים קשר בעוצמה בינונית בין המין להשכלה מקדם המתאם של קרמר הוא 0.595.

2) א. להלן טבלה: ב. מדד קרמר 0.19 מעיד על קשר בעוצמה נמוכה.

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

ביוסטטיסטיקה

פרק 38 - 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) :

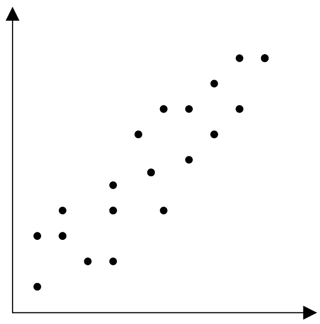
בבניין 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגרות בדירה. להלן התוצאות שהתקבלו :

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

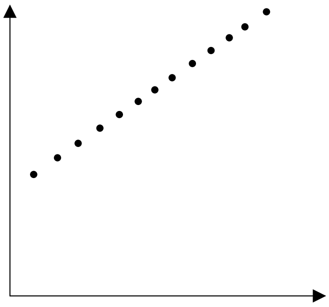
1) כמה תצפיות ישנן בדוגמה?
2) כמה משתנים ישנם בדוגמה, מי הם?
3) שרטטו לנתונים דיאגרמת פיזור.
4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

דיאגרמות פיזור לקשר בין משתנים וניתוחם

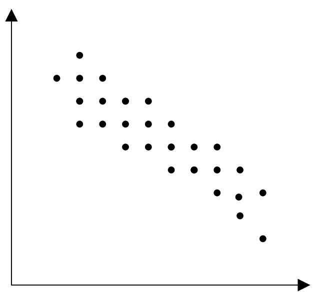
קשר לנארי חיובי חלקי



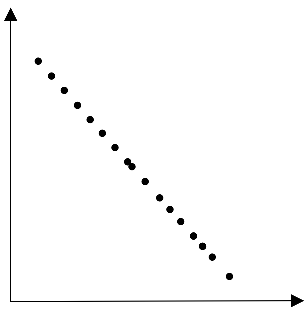
קשר לנארי חיובי מלא



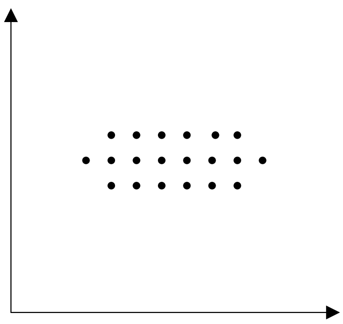
קשר לינארי שלילי חלקי



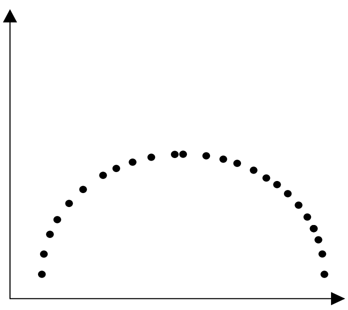
קשר לינארי שלילי מלא



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 : ρ - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר :

r - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר ρ.

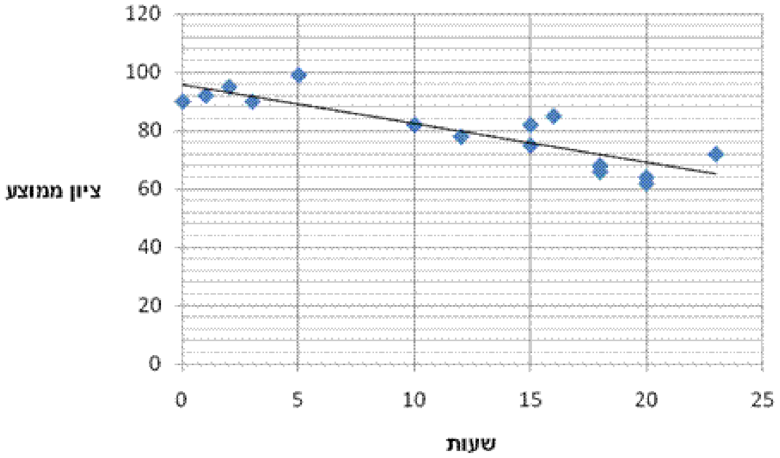
קיומו של מתאם בין שני משתנים אינו מצביע על סיבתיות בהכרח. למשל, אם נמצא מתאם חיובי בין כמות הסוכרזית שאדם אוכל לבין במשקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי, כלומר אם נחליף את X ב- Y התוצאה תהיה זהה.

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 נתבקש 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?

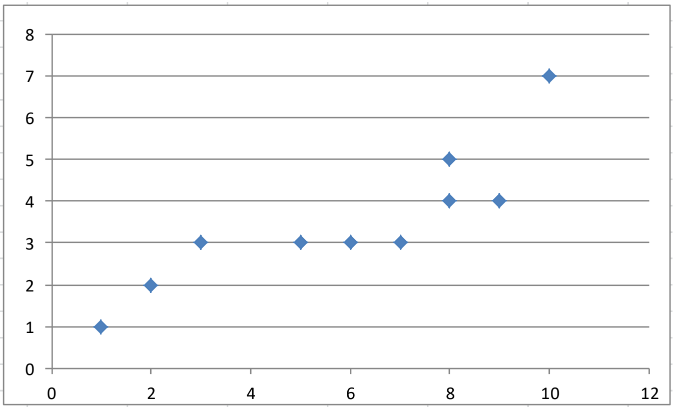
ד. מסתבר שבסופו של דבר נתנו פקטור של 5 נקודות לציון הסופי. כיצד הדבר היה משנה את מקדם המתאם של המדגם?

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 לינארי _________ (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 :

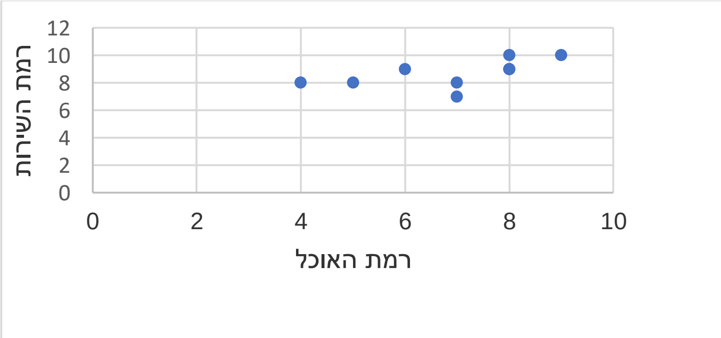
א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני בימים שנדגמו.

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.

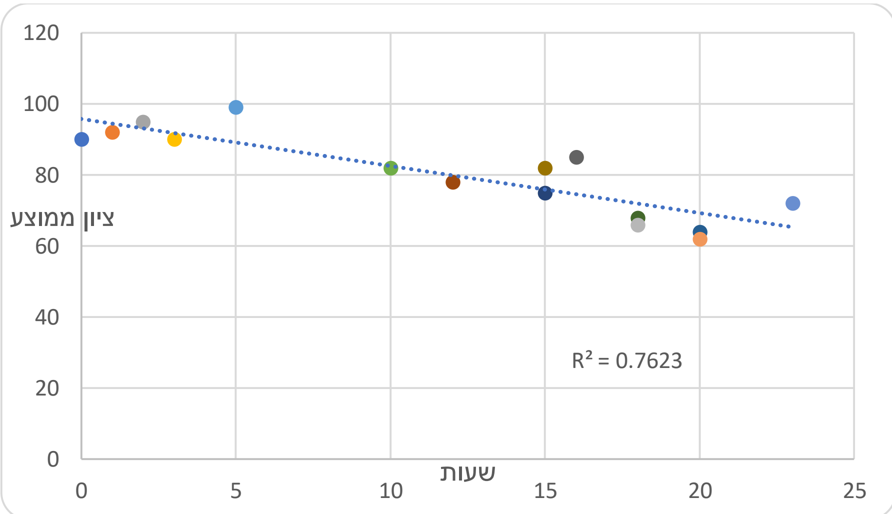
7) בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



מה יהיה ערכו של מקדם המתאם (r)?

א. $r=-0.3$
ב. $r=0$
ג. $r=1.125$
ד. $r=0.593$

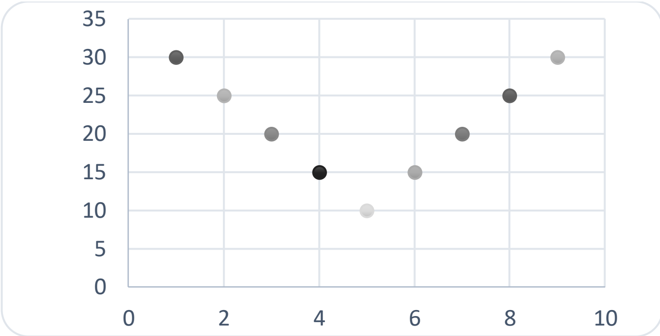
8) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור.



מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?

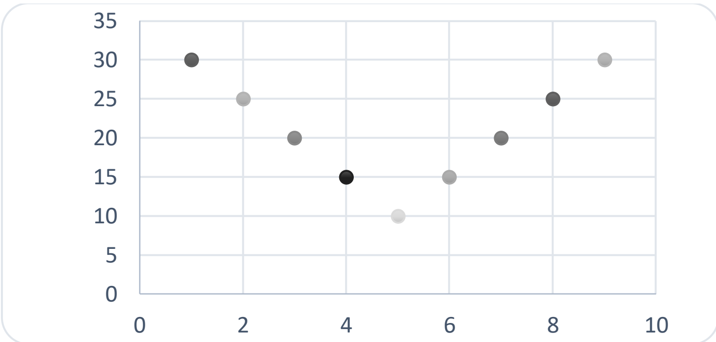
א. ככל שמבלים יותר הציון נוטה לרדת.
ב. אין קשר בין שעות הבילוי לציון.
ג. ככל שמבלים פחות הציון נוטה לרדת.
ד. ככל שהציון נוטה לרדת הסטודנט מבלה פחות.

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 קוים ישרים.
ב. $r = 2$ היות ויש שני קוים בעלי קשר מושלם.
ג. $r = 0$ היות והקו יורד ואחר כך עולה באותו האופן.
ד. $r = \pm1$ היות ויש קו עולה וגם קו יורד.

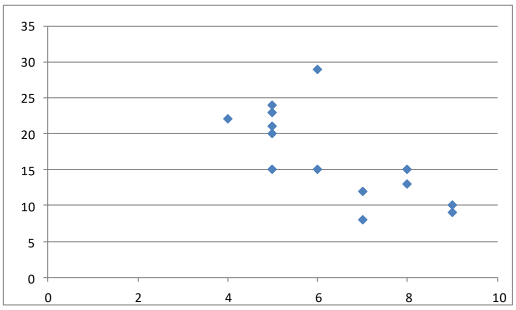
10) התרשים הבא מתאר דיאגרמת פיזור.



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.

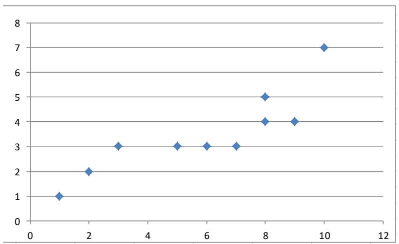
ב. עבור כל תצפית ערך של משתנה אחד שווה לערך של המשתנה השני.

ג. הקשר הלינארי הוא בעוצמה חזקה.

ד. אף אחת מהתשובות לא בהכרח נכונה.

13) להלן דיאגרמת פיזור:

מה יהיה מקדם המתאם בין שני המשתנים?



א. 1

ב. 0.85

ג. 0.15

ד. 0

14) בבדיקת קשר בין שני משתנים התקבל: $r = -1$.

א. קיימת נוסחה לינארית הקושרת בין כל התצפיות.

ב. לא קיים קשר בין שני המשתנים.

ג. ככל שמשתנה אחד נוטה לרדת גם לשני יש נטייה לרדת.

ד. קיים קשר בין שני המשתנים, אך לא ניתן לדעת מאיזה סוג.

15) לפי הפתגם "רחוק מהעין, רחוק מהלב", יש קשר ____ בין קרבה פיזית לקרבה נפשית.
א. חיובי
ב. שלילי
ג. אפסי
ד. לא ניתן לדעת.

16) מבחן אמי"ר הינו מבחן מיון באנגלית של המרכז הארצי לבחינות והערכה. הציון המינימלי בבחינה הינו 150 והמקסימלי הינו 250. בקורס הכנה למבחן השתתפו 19 תלמידים. להלן הציונים שלהם על פי פלט שהתקבל:

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

		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	

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:

מס' דירה	X	Y
1	3	2
2	2	2
3	4	3
4	3	3
5	5	4

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-).

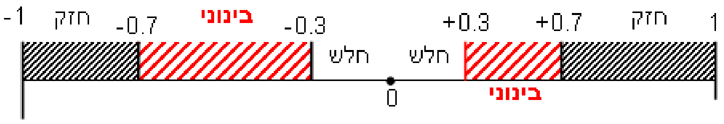
מתאם חיובי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

מתאם שלילי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

ככל שערך מקדם המתאם קרוב לאפס נאמר שעוצמת הקשר חלשה יותר וככל שמקדם המתאם רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר :



מקדם המתאם יסומן באות r.

כדי לחשב את מקדם המתאם, יש לחשב את סטיות התקן של כל משתנה ואת השונות המשותפת.

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 הלינארי: $$r_{xy}=\frac{COV(x,y)}{S_x\cdot S_y}$$

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:

Y	X
12	10
15	14
15	15
17	18
21	20

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. מתוך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5₪. אם מתוך 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.

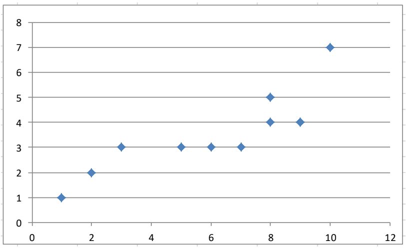
ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.

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 דיאגרמת פיזור:
מה יהיה מקדם המתאם בין שני המשתנים?
א. 1
ב. 0.85
ג. 0.15
ד. 0



תשובות סופיות

1) א. משתנה תלוי: ציון, משתנה ב״ת: מס׳ חיסורים. ראה דיאגרמה בוידאו. ניתן להסיק שקיים קשר לינארי שלילי וחלקי בין מספר החיסורים לציון התלמיד.
ב. ‎-0.9325‎. ג. הקשר יישאר לינארי שלילי חלקי אך עוצמתו תחלש.

2) א. $\bar{y}=16$, $\bar{x}=15.4$ ב. $r_{xy}=0.96$.

3) א. 0.8.

4) 0.8.

5) 1.

6) א. נכון. ב. לא נכון. ג. נכון.

7) ג׳.

8) א׳.

9) ב׳.

בדיקת השערות על מקדם המתאם הלינארי – רקע

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב : ρ - פרמטר המאפיין את עוצמת הקשר הלינארי וכיוונו בין שני המשתנים הנחקרים באוכלוסייה. כאשר :
r - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר ρ .

השערת האפס: תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים $H_0:\rho=0$.

ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית.

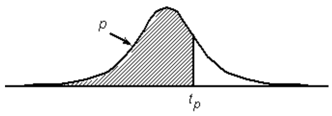
סטטיסטי המבחן : $t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\sim t(n-2)$

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 : אזור דחייה של השערת האפס	$t\geq t_{1-\alpha}$ א ו $t\leq -t_{1-\alpha}$	$t\leq -t_{1-\alpha}$	$t\geq t_{1-\alpha}$

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 ליניארי 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 ליניארי 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. $0.005 < P_V < 0.01$.

בדיקת השערות על מקדם המתאם הלינארי (באמצעות טבלה של ערכים קריטיים) – רקע

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב : ρ - פרמטר המאפיין את עוצמת הקשר הלינארי וכיוונו בין שני המשתנים הנחקרים באוכלוסייה. כאשר:
r - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר ρ.

השערת האפס: תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים: $H_0: \rho = 0$.

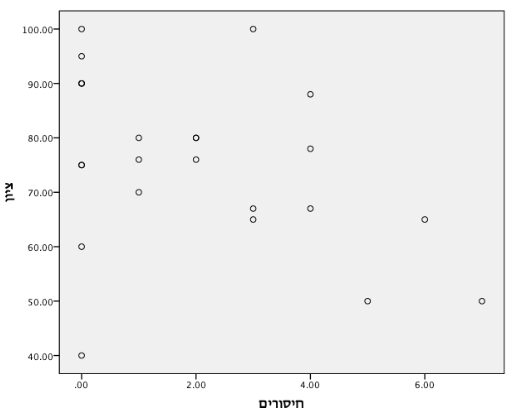
ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו-נורמלית.

את מקדם המדגם הקריטי, שנסמן ב- r_c, נוציא מתוך טבלה של ערכים קריטיים שמצורפת בהמשך.

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: אזור דחייה של השערת האפס	$r \geq r_c$ או $r \leq -r_c$	$r \leq -r_c$	$r \geq r_c$

דוגמה (פתרון בהקלטה) :

הדיקן ביקש לדגום סטודנטים כדי לבדוק את הקשר בין ציון הסטודנט בקורס למספר הפעמים שהוא החסיר שיעור בקורס.
דיאגרמת הפיזור שהתקבלה במדגם שבוצע:



מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי במחקר?

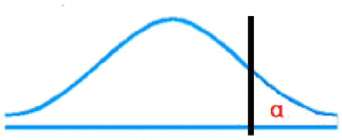
מה ניתן לראות לגבי הקשר הלינארי בין המשתנים שהתקבל במדגם?

חושב האומד למקדם המתאם הלינארי על סמך 24 הסטודנטים שנדגמו והתקבל: -0.389.

מה משמעות של מקדם המתאם שהתקבל במדגם?

האם ניתן להגיד ברמת מובהקות של 5% שקיים מתאם לינארי שלילי בין מספר החיסורים של הסטודנטים מהקורס לבין הציון של הסטודנטים בקורס?

טבלת ערכים קריטיים של מקדם המתאם הליניארי



n \ α	0.05	0.025	0.005	0.0005
4	0.900	0.950	0.990	0.999
5	0.805	0.878	0.959	0.991
6	0.729	0.811	0.917	0.974
7	0.669	0.754	0.875	0.951
8	0.621	0.707	0.834	0.925
9	0.582	0.666	0.798	0.898
10	0.549	0.632	0.765	0.872
11	0.521	0.602	0.735	0.847
12	0.497	0.576	0.708	0.823
13	0.476	0.553	0.684	0.801
14	0.458	0.532	0.661	0.780
15	0.441	0.514	0.641	0.760
16	0.426	0.497	0.623	0.742
17	0.412	0.482	0.606	0.725
18	0.400	0.468	0.590	0.708
19	0.389	0.456	0.575	0.693
20	0.378	0.444	0.561	0.679
21	0.369	0.433	0.549	0.665
22	0.360	0.423	0.537	0.652
23	0.352	0.413	0.526	0.640
24	0.344	0.404	0.515	0.629
25	0.337	0.396	0.505	0.618
26	0.330	0.388	0.496	0.607
27	0.323	0.381	0.487	0.597
28	0.317	0.374	0.479	0.588
29	0.311	0.367	0.471	0.579
30	0.306	0.361	0.463	0.570
35	0.283	0.334	0.430	0.532

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 היא בהכרח יש לה יותר המוגלובין בדם?

ב. האם ניתן לומר, ברמת מובהקות של 5%, שהמתאם בין גיל האישה שבהריון לבין רמת ההמוגלובין שלה בדם הוא חיובי?

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. בדוק ברמת מובהקות של 2.5% האם קיים קשר לינארי שלילי בחודש ינואר בין הטמפרטורה במעלות צלזיוס לבין המשקעים במ"מ?

ב. כיצד הייתה משתנה התשובה לסעיף א אם הינו מוסיפים עוד תצפיות למדגם?

תשובות סופיות

1) א. לא נדחה את H_0. ב.לא תשתנה.

2) א. לא. ב. נדחה את H_0.

3) א. נדחה את H_0. ב. לא ניתן לדעת.

מדד הקשר הליניארי – ניתוח פלטים – רקע

מדד הקשר הליניארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הליניארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:
r - מדד הקשר הליניארי במדגם שמהווה אומד לפרמטר ρ.

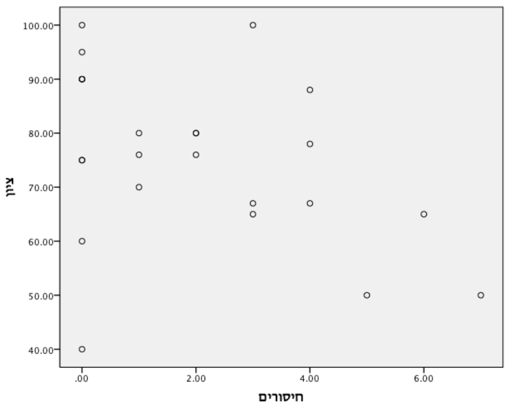
השערת האפס: תהיה שבאוכלוסייה לא קיים כלל קשר ליניארי בין שני המשתנים: $H_0:\rho=0$.

ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית.

דוגמה (פתרון בהקלטה):

הדיקן ביקש לדגום סטודנטים כדי לבדוק את הקשר בין ציון הסטודנט בקורס למספר הפעמים שהוא החסיר שיעור בקורס. דיאגרמת הפיזור שהתקבלה במדגם שבוצע:

א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי במחקר?
ב. מה ניתן לראות לגבי הקשר הליניארי בין המשתנים שהתקבל במדגם?



Correlations

		חיסורים	ציון
חיסורים	Pearson Correlation	1	-.389
	Sig. (2-tailed)		.060
	N	24	24
ציון	Pearson Correlation	-.389	1
	Sig. (2-tailed)	.060	
	N	24	24

ג. מהו מקדם המתאם שהתקבל במדגם? מה המשמעות שלו?
ד. האם ניתן להגיד ברמת מובהקות של 5% שקיים מתאם ליניארי שלילי בין מספר החיסורים של הסטודנטים מהקורס לבין הציון של הסטודנטים בקורס?

שאלות

1) מחקר רפואי התעניין לבדוק האם קיים קשר לינארי בין גיל האישה בהריון לרמת ההמוגלובין שלה. להלן תוצאות מדגם שהתקבלו, עבור נשים בהריון :

Correlations

		age	hemoglobin
age	Pearson Correlation	1	.565
	Sig. (2-tailed)		.005
	N	23	23
hemoglobin	Pearson Correlation	.565	1
	Sig. (2-tailed)	.005	
	N	23	23

א. מהי אוכלוסיית המחקר?
ב. מהן השערות המחקר?
ג. מהו המשתנה הבלתי תלוי ומהו המשתנה התלוי במחקר?
ד. מהי מסקנת המחקר ברמת מובהקות של 5%?

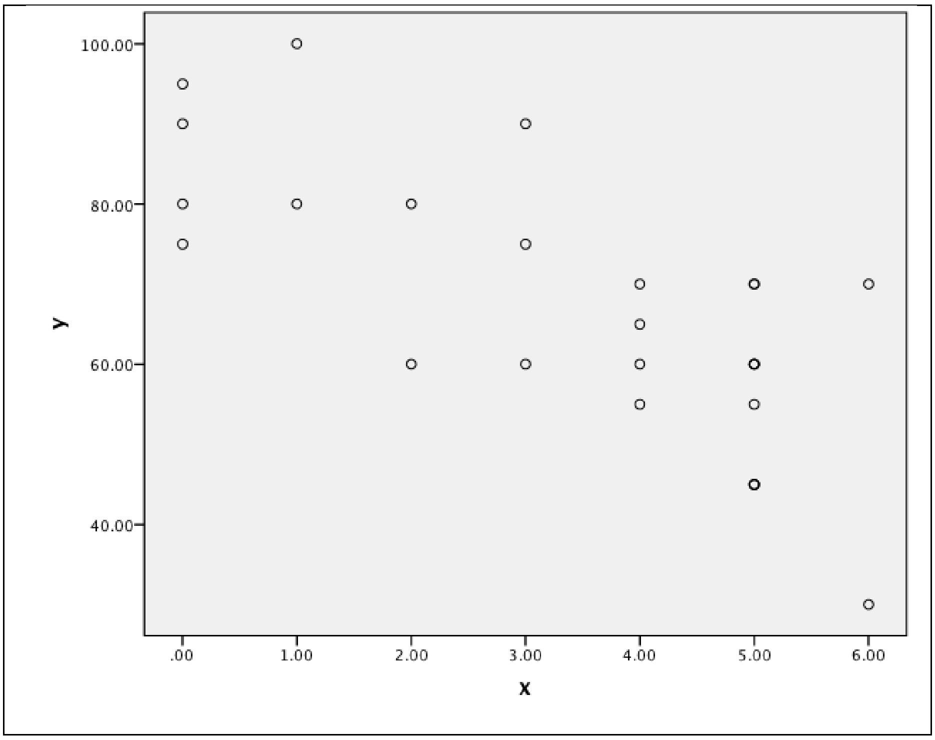
2) במדגם שנעשה נבדקו מספר משתנים על התצפיות שנדגמו. להלן פלט שהופק על המדגם?

Correlations

		x	y	z	w
x	Pearson Correlation	???	-.682	.134	.176
	Sig. (2-tailed)		.005	.634	.530
	N	15	15	15	15
y	Pearson Correlation	-.682	1	???	-.555
	Sig. (2-tailed)	.005		.544	.032
	N	15	15	15	15
z	Pearson Correlation	.134	.170	1	-.247
	Sig. (2-tailed)	???	.544		.374
	N	15	15	15	15
w	Pearson Correlation	.176	-.555	-.247	1
	Sig. (2-tailed)	.530	.032	.374	
	N	15	15	15	15

א. בין אילו שני משתנים שונים הקשר הליניארי במדגם נמצא עם העוצמה הכי חזקה?
ב. ברמת מובהקות של 5%, אילו שני משתנים בעל קשר ליניארי מובהק?
ג. השלימו את המספרים המסומנים בפלט בסימני שאלה.

3) נדגמו מספר תלמידים בכיתה יב׳ ובדקו לכל תלמיד : X - מספר שעות שבועיות שהתלמיד צופה בטלוויזיה ביום Y - ציון הבגרות שלו במתמטיקה. להלן התוצאות שהתקבלו במחקר :



Correlations

		x	y
x	Pearson Correlation	1	-.741**
	Sig. (2-tailed)		.000
	N	26	26
y	Pearson Correlation	-.741**	1
	Sig. (2-tailed)	.000	
	N	26	26

**. Correlation is significant at the 0.01 level (2-tailed).

א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי?
ב. מהו כיוון הקשר שהתקבל במדגם ומהו עוצמתו?
ג. האם ניתן להגיד שבאופן מובהק ככל שתלמיד צופה יותר בטלוויזיה הוא מצליח פחות בבגרות במתמטיקה?
ד. בהמשך לסעיף הקודם, האם ניתן להגיד שהסיבה להצלחה או אי הצלחה בבגרות במתמטיקה היא זמן הצפייה בטלוויזיה?

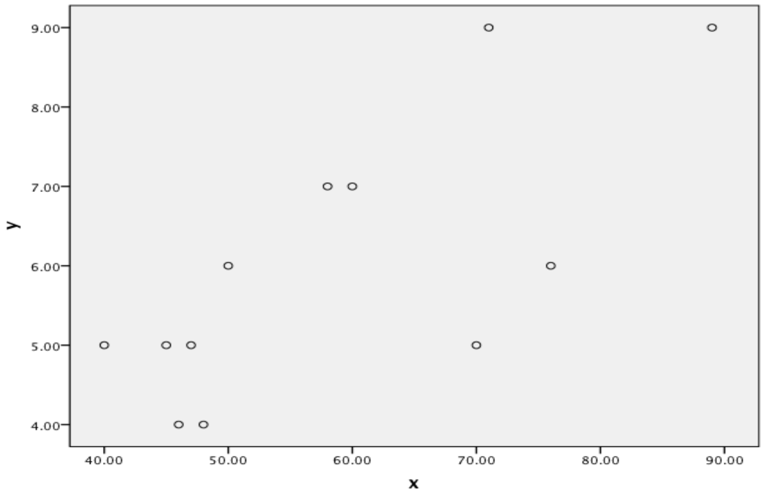
4) נדגמו ילדים בגיל 8 ונבדק עבור כל ילד גובהו בס"מ ומשקלו בק"ג. להלן הפלט שהתקבל עבור תוצאות המדגם :

Correlations

		גובה	משקל
גובה	Pearson Correlation	1	.552
	Sig. (2-tailed)		.062
	N	12	12
משקל	Pearson Correlation	.552	1
	Sig. (2-tailed)	.062	
	N	12	12

א. בדקו ברמת מובהקות של 5% האם קיים קשר לנארי חיובי בין המשקל והגובה.
ב. באילו רמות מובהקות ניתן לקבוע שקיים קשר לנארי חיובי בין במשקל והגובה?
ג. כיצד התשובה לסעיף הקודם הייתה משתנה אם היו מתווספות עוד 3 תצפיות למדגם?

5) בתהליך כימי מסוים חוקר בדק את הקשר בין הטמפרטורה בתהליך (X) לבין אחוז החומר (Y) בתהליך. דיאגרמת הפיזור שהתקבלה היא:



Correlations

		x	y
x	Pearson Correlation	1	.732**
	Sig. (2-tailed)		.007
	N	12	12
y	Pearson Correlation	.732**	1
	Sig. (2-tailed)	.007	
	N	12	12

**. Correlation is significant at the 0.01 level (2-tailed).

א. מה ניתן להגיד על סמך הפלט על הקשר שנימצא במדגם בין הטמפרטורה בתהליך לאחוז החומר?

ב. האם הקשר בין הטמפרטורה בתהליך לבין אחוז החומר הוא קווי חיובי מובהק? בדקו ברמת מובהקות של 5%.

ג. מה היה קורה למקדם המתאם במדגם ומובהקות התוצאה אם הייתה מתווספת תצפיות שבה הטמפרטורה היא 40 ואחוז החומר 9?

תשובות סופיות

1) א. נשים בהריון.

ב. $H_0 : p = 0$
$H_1 : p \neq 0$

ג. משתנה תלוי – רמת ההמוגלובין, משתנה בלתי תלוי- גיל.

ד. קיים קשר לינארי בין גיל האישה בהריון לרמת ההמוגלובין שלה בדם.

2) א. בין X ל- Y. ב. X ו- Y. כמו כן, W ו- Y. ג. ראה וידאו.

3) א. משתנה תלוי – ציון בבגרות במתמטיקה, משתנה בלתי תלוי- שעות צפייה.

ב. כיוון שלילי ועוצמה של 0.741. ג. כן. ד. לא.

4) א. נדחה את H_0. ב. לפחות 0.032. ג. לא ניתן לדעת.

5) א. קיים קשר לינארי חיובי וחלקי שעוצמתו : 0.732. ב. נדחה את H_0.

ג. מקדם המתאם קטן ומובהקות התוצאה גדלה.

ביוסטטיסטיקה

פרק 39 - רגרסיה לינארית פשוטה

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:

$$\sum_{i=1}^{20} Y_i = 200 \ , \ \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20}\left(X_i-\overline{X}\right)^2 = 76 \ , \ \sum_{i=1}^{20}\left(Y_i-\overline{Y}\right) = 76 \ .$$

$$\sum_{i=1}^{20}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right) = 60.8$$

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:
$S_X = 2 \ , \ S_Y = 5 \ , \ \overline{X} = 14 \ , \ \overline{Y} = 8 \ , \ \text{COV}(X,Y) = 7.5$.

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: $S_x = S_y = 4 \ , \ \overline{X} = 1.5$,
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.

ביוסטטיסטיקה

פרק 40 - 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 הקשר 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'.

ביוסטטיסטיקה

פרק 41 - ניתוח שונות חד כיוונית

תוכן העניינים

ניתוח שונות חד כיוונית

רקע תיאורטי

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות $\left(\mu_{1\ldots},\mu_k\right)$ של k אוכלוסיות שונות.

לכן, בניתוח שונות, השערות המחקר הן :

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ (התוחלות של כל האוכלוסיות שוות)

$H_1 :$ אחרת (לפחות שתיים מהתוחלות שונות)

ההנחות הדרושות לביצוע התהליך :

2) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.
3) כל האוכלוסיות הן עם אותה שונות σ^2.
4) המדגמים בלתי תלויים זה בזה.

ישנו משתנה המבדיל בין הקבוצות השונות, הוא המשתנה הבלתי תלוי הנקרא גורם (factor). משתנה זה הוא קטגוריאלי עם k רמות (levels).
כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה :
נסמן ב- n_i את גודל המדגם בקבוצה i.

$n = \sum_{1}^{k} n_i$ - מספר התצפיות סך הכול (בכל המדגמים).

$\overline{X}_1$ - ממוצע המדגם הראשון,......., $\overline{X}_k$-ממוצע המדגם ה- k-י.

$\overline{X}$ - ממוצע כללי (של כל המדגמים).

סכום ריבועים בין הקבוצות : $SS_B = \sum_{i=1}^{k} n_i \left[\overline{X}_i - \overline{X}\right]^2$

סכום ריבועים בתוך הקבוצות : $SS_W = \sum_{i=1}^{k} n_i [n_i - 1] \bullet \hat{S}_i^2$

סכום ריבועים כללי : $SS_T = \sum_{i=1}^{k} \sum_{j=1}^{n_j} [X_{ij} - \overline{X}]^2$

$SST = SSB + SSW$

יש למלא את טבלת ניתוח השונות הבאה :

מקור השונות	סכום הריבועים SS	דרגות חופש df	ממוצע הריבועים MS	F
B - בין הקבוצות	SSB	$k-1$	$\frac{SSB}{k-1}$	$\frac{MSB}{MSW}$
W - בתוך הקבוצות	SSW	$n-k$	$\frac{SSW}{n-k}$	
T - סה"כ	SST	$n-1$		

$$F=\frac{\frac{SSB}{k-1}}{\frac{SSW}{n-k}}\sim F(k-1,n-k)$$

אזור דחיית H_0 : $1-\alpha$: $F>F_{(k-1,n-k)}$

שאלות

1) מחקר מעוניין להשוות בין שלוש תרופות לשיכוך כאבים במטרה לבדוק האם קיים הבדל בין התרופות מבחינת הזמן בדקות שלוקח עד שהתרופה משפיעה. לצורך הבדיקה נלקחו 15 אנשים שסובלים מכאבי ראש. אנשים אלה חולקו באקראי לשלוש: קבוצה 1קיבלה "אקמולי" קבוצה 2 קיבלה "אופטלגין" קבוצה 3 קיבלה "נורופן".
כל אדם במחקר מסר את מספר הדקות עד שהתרופה השפיעה עליו.

א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר?
מהו ה"גורם" וכמה רמות יש לו?

ב. מהו המבחן הסטטיסטי המתאים כאן? רשמו את ההשערות.

ג. מה הן ההנחות הדרושות כדי לבצע את המבחן הסטטיסטי שהצעת בסעיף הקודם?

2) בעיר מסוימת שלושה בתי ספר תיכון. ראש העיר התעניין לבדוק האם קיים הבדל בהצלחה של בתי הספר במקצוע מתמטיקה. לצורך כך הוא דגם מספר תלמידים שנבחנו במבחן הבגרות במתמטיקה ברמה של 3 יחידות בעירו ובדק עבור כל תלמיד מה ציון הבגרות שלו במתמטיקה.
להלן הציונים שהתקבלו:

"המתמטיד"	"רבין"	"הסי"
78	98	85
65	62	83
70	55	74
90	80	85
56		75

א. מהו המבחן הסטטיסטי המתאים?
רשמו את ההשערות ואת ההנחות של המבחן.

ב. מהו גודל המדגם? מהו המשתנה הבלתי תלוי (factor) כמה רמות יש לו?

ג. חשבו את הממוצע ואת סטיית התקן של הציונים בכל אחד מהמדגמים.

ד. מלאו את טבלת ANOVA.

ה. רשמו את כלל ההכרעה למבחן שהוצע בסעיף א ברמת מובהקות של 5%.

ו. האם קיים הבדל בין בתי הספר בעיר מבחינת רמת הצלחת התלמידים במקצוע המתמטיקה? ענה על סמך הסעיפים הקודמים.

3) מעוניינים לבדוק האם יש הבדל בהשפעה של שיטות טפול שונות על לחץ הדם הסיסטולי (SBP) באוכלוסייה של קשישים. נבדקו 4 שיטות שונות.
בטבלה המצורפת מרוכזים ממצאי המחקר.

השיטה	A	B	C	D
גודל המדגם	12	14	8	12
הממוצע	178	172	180	182
סטיית התקן	4	8	5	3

א. רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.

ב. מה מסקנת המחקר ברמת מובהקות של 5%?

ג. האם יש צורך לבצע השוואות מרובות?

4) שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות.
האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות?
בדקו ברמת מובהקות של 5%.

האופה	ניר	מוזס	שלום
סכום הזמנים	206	212	182
סכום ריבועי הזמנים	10644	11250	8982

5) להלן טבלת ניתוח שונות חד כיוונית. במחקר בחנו 4 סוגי סוללות. רצו לבדוק האם לסוג הסוללה השפעה על תוחלת אורך החיים שלה. הפעילו את כל הסוללות על אותו מכשיר ובדקו את אורך החיים של כל סוללה בשעות.
מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.317	3	3.439	1.361	.279
Within Groups	60.648	24	2.527		
Total	70.964	27			

6) להלן טבלת ANOVA בטבלה הושמטו חלקים.
השלימו את החלקים בטבלה שהושמטו ומסומנים באותיות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450	ב	ג	ה	.000
Within Groups	א	17	ד		
Total	522.950	19			

7) חברת תרופות לקחה 15 אנשים ברמת בריאות דומה. החברה חילקה את האנשים ל שלוש קבוצות שוות בגודלן. לכל קבוצה ניתנה אותה תרופה במינון שונה (dosage). המינונים שניתנו הם: 10 מ"ג, 20 מ"ג ו-30 מ"ג. לאחר שעה מזמן לקיחת התרופה ניבדק קצב פעימות הלב של כל אדם (pulse). הנתונים הוזנו לתוכנה סטטיסטית והתקבלו התוצאות הבאות:

ANOVA

pulse

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	414.400	2	207.200	19.733	.000
Within Groups	126.000	12	10.500		
Total	540.400	14			

pulse

Tukey HSD[a]

dosage	N	Subset for alpha = 0.05	
		1	2
30.00	5	71.0000	
20.00	5		80.2000
10.00	5		83.4000
Sig.		1.000	.299

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

Post Hoc Tests

Multiple Comparisons

pulse
Tukey HSD

(I) dosage	(J) dosage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
10.00	20.00	3.20000	2.04939	.299	-2.2675	8.6675
	30.00	12.40000*	2.04939	.000	6.9325	17.8675
20.00	10.00	-3.20000	2.04939	.299	-8.6675	2.2675
	30.00	9.20000*	2.04939	.002	3.7325	14.6675
30.00	10.00	-12.40000*	2.04939	.000	-17.8675	-6.9325
	20.00	-9.20000*	2.04939	.002	-14.6675	-3.7325

*. The mean difference is significant at the 0.05 level.

א. בדקו ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.

ב. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.

ג. האם יש צורך במחקר בהשוואות מרובות. נמקו!

ד. לטבלת הANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "טוקיי". ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

8) בעיר מסוימת רצו לבדוק האם קיים הבדל ברמה של התלמידים בין בתי הספר השונים בעיר. ביצעו מדגם מכל בית ספר ונתנו מבחן זהה לכל הנדגמים. לאחר מכן ריכזו את הנתונים בתוכנה סטטיסטית והפעילו ניתוח שונות. מצורפים הפלטים שהתקבלו. ענו על הסעיפים הבאים:

א. כמה בתי ספר יש בעיר?

ב. כמה תלמידים השתתפו בסך הכול במחקר?

ג. האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%

ד. בביטחון של 95% אילו בתי ספר שונים זה מזה ברמת התלמידים? נמקו והסבירו.

Oneway

ANOVA

grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7799.600	4	1949.900	13.586	.000
Within Groups	2870.400	20	143.520		
Total	10670.000	24			

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

(I) school	(J) school	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.00	2.00	5.40000	7.57681	.971	-20.2543	31.0543
	3.00	36.80000*	7.57681	.003	11.1457	62.4543
	4.00	36.40000*	7.57681	.003	10.7457	62.0543
	5.00	-2.60000	7.57681	.998	-28.2543	23.0543
2.00	1.00	-5.40000	7.57681	.971	-31.0543	20.2543
	3.00	31.40000*	7.57681	.011	5.7457	57.0543
	4.00	31.00000*	7.57681	.013	5.3457	56.6543
	5.00	-8.00000	7.57681	.888	-33.6543	17.6543
3.00	1.00	-36.80000*	7.57681	.003	-62.4543	-11.1457
	2.00	-31.40000*	7.57681	.011	-57.0543	-5.7457
	4.00	-.40000	7.57681	1.000	-26.0543	25.2543
	5.00	-39.40000*	7.57681	.001	-65.0543	-13.7457
4.00	1.00	-36.40000*	7.57681	.003	-62.0543	-10.7457
	2.00	-31.00000*	7.57681	.013	-56.6543	-5.3457
	3.00	.40000	7.57681	1.000	-25.2543	26.0543
	5.00	-39.00000*	7.57681	.001	-64.6543	-13.3457
5.00	1.00	2.60000	7.57681	.998	-23.0543	28.2543
	2.00	8.00000	7.57681	.888	-17.6543	33.6543
	3.00	39.40000*	7.57681	.001	13.7457	65.0543
	4.00	39.00000*	7.57681	.001	13.3457	64.6543

*. The mean difference is significant at the 0.05 level.

Homogeneous Subsets

grade

Scheffe[a]

school	N	Subset for alpha = 0.05 1	Subset for alpha = 0.05 2
3.00	5	45.0000	
4.00	5	45.4000	
2.00	5		76.4000
1.00	5		81.8000
5.00	5		84.4000
Sig.		1.000	.888

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 5.000.

תשובות סופיות

1) א. משתנה בלתי תלוי: סוג התרופה.

משתנה תלוי: הזמן עד להשפעת התרופה בדקות.

ב. ניתוח שונות חד כיווני

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: otherwise$.

ג. 1. מדגמים בלתי תלויים.

2. שווין שונויות.

3. משתנים מתפלגים נורמלית.

2) א. המבחן לניתוח שונות חד כיוונית.

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: otherwise$

הנחות:

1. מדגמים בלתי תלויים.
2. משתנים מתפלגים נורמלית.
3. שוויון שונוית.

ב. גודל המדגם: 14. משתנה ב"ת: בית הספר, בעל 3 רמות.

ג. $\bar{X} = 71.8, \hat{S} = 12.93$, $\bar{X} = 73.75, \hat{S} = 19.29$, $\bar{X} = 80.4, \hat{S} = 5.46$.

ד. להלן טבלה:

מקור השונות	SS	df	MS	F
B	200.6	2	100.3	
W	1904.75	11	173.2	
סה"כ	2105.35	13		0.58

ה. $F > 3.98$.

ו. נקבל את H_0 .

3) א. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1: otherwise$

ב. נדחה את H_0 .

ג. כן.

הנחות:

1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

4) נקבל את H_0 : נכריע שאין הבדל מובהק בין האופים מבחינת תוחלת זמן הכנה.

5) $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_1: otherwise$

הנחות :
1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

נקבל את H_0 : לסוג סוללה אין השפעה של תוחלת החיים ברמת ביטחון של 10%.

6) להלן טבלה :

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450	ב 2	ג 178.725	ה 18.36	.000
Within Groups	א 165.5	17	ד 9.735		
Total	522.950	19			

7) א. $H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: otherwise$

הנחות :
1. מדגמים בלתי תלויים.
2. משתנים מתפלגים נורמלית.
3. שוויון שונויות.

נדחה את H_0 : ברמת ביטחון של 5% קיים הבדל במינונים השונים מבחינת תוחלת הדופק.

ב. ראה וידאו. ג. כן. ד. $\mu_{20} = \mu_{10} > \mu_{30}$.

8) א. 5 ב. 25

ג. נדחה את H_0 : יש לפחות שני בתי ספר בעיר עם תוחלת רמת ציונים שונה.

ד. $(\mu_3 = \mu_4) < (\mu_1 = \mu_2 = \mu_3)$.

ביוסטטיסטיקה

פרק 42 - ניתוח שונות דו כיווני

תוכן העניינים

ניתוח שונות דו-כיווני - הקדמה

רקע

ראשית, נחזור על עיקרי ניתוח השונות החד-כיווני (חד-גורמי).

בניתוח שונות חד-כיווני יש משתנה תלוי יחיד, שהוא כמותי, ומשתנה בלתי תלוי יחיד, שהוא משתנה קטגוריאלי (משתנה שהערכים שלו שייכים למספר סופי של קטגוריות). המשתנה הקטגוריאלי נקרא לעתים גם גורם (פקטור), והקטגוריות שלו נקראות רמות. המטרה בניתוח שונות חד-כיווני היא לבדוק האם לגורם יש השפעה מובהקת על המשתנה התלוי. השערת האפס של המחקר בניתוח שונות חד-כיווני היא שבכל הקטגוריות יש אותה התוחלת, והשערת המחקר טוענת שיש לפחות שתי קטגוריות שבהן התוחלות שונות.

<u>דוגמה :(פתרון בהקלטה)</u>

נבדקו שלושה סוגי דיאטות על אנשים בעלי משקל עודף. נבחרו 30 מטופלים בעלי משקל עודף, והם חולקו באקראי לשלוש קבוצות שוות בגודלן, כך שכל קבוצה קיבלה דיאטה נחקרת אחרת. כעבור שלושה חודשים בדקו את מספר הקילוגרמים שהפחית כל מטופל ממשקלו בתקופה זו. מטרת המחקר הייתה לבדוק האם קיים הבדל בין הדיאטות מבחינת ההפחתה במשקל.

- מהו המשתנה התלוי במחקר?
- מהו המשתנה הבלתי תלוי במחקר? כמה רמות יש לו?
- מה הן השערות המחקר?
- מהו המבחן הסטטיסטי המתאים?

בניתוח שונות דו-כיווני אנו מוסיפים עוד משתנה בלתי תלוי למחקר,
כלומר עוד גורם שאנו רוצים לבדוק איך הוא משפיע על המשתנה התלוי.
לכן בניתוח שונות דו-כיווני יש משתנה תלוי כמותי יחיד ושני משתנים בלתי תלויים שכל אחד מהם קטגוריאלי. כזכור, למשתנים הבלתי תלויים אנו קוראים גם גורמים (פקטורים), ומספר הקטגוריות של כל גורם נקרא גם מספר הרמות שלו.
ניתוח שונות רב-כיווני או רב-גורמי הוא ניתוח שונות שבו יש יותר מגורם אחד, כלומר יותר ממשתנה בלתי תלוי קטגוריאלי אחד. בניתוח שונות דו-כיווני יש שני גורמים, בניתוח שונות תלת-גורמי יש שלושה גורמים וכו'. ככל שנוסיף גורמים, הניתוח הסטטיסטי יהיה מורכב יותר ויידרשו יותר תצפיות למחקר אבל כיוון שהוא יקטין את שונות הטעויות (שונות מקרית) וייתן יותר הסבר לשונות הכללית, כך שהמבחן יהיה עוצמתי יותר.

המשך הדוגמה:

מבין 30 המטופלים שבמחקר 15 היו גברים ו-15 היו נשים. המטופלים חולקו כך שבכל דיאטה השתתפו 5 גברים ו-5 נשים.

מה הם המשתנים הבלתי תלויים? כמה רמות יש לכל משתנה?

בניתוח שונות דו-כיווני אנו בעצם רוצים לבדוק סימולטנית שלוש שאלות מחקר על אוכלוסיית כבדי המשקל:

- האם יש הבדלים משמעותיים בין שיעורי הפחתת המשקל של מטופלים כבדי משקל כתוצאה משימוש בדיאטות שונות?
- האם יש הבדלים משמעותיים בין שיעורי הפחתת המשקל של מטופלים כבדי משקל כתוצאה ממגדר שונה?
- האם יש השפעה משולבת (אינטראקציה) של שני הגורמים הנבדקים על הפחתת המשקל של מטופלים כבדי משקל, כלומר האם צירוף של דיאטה מסוימת ומגדר מסוים מביא להפחתת משקל גדולה יותר או קטנה יותר מצירופים אחרים?

נסמן גורם אחד ב-a ואת מספר הרמות שלו ב-A. באותו האופן הגורם האחר יסומן ב-b, ואת מספר הרמות שלו נסמן ב-B. מספר הקבוצות הכולל שאנו יוצרים הוא $A \cdot B$.

המשך הדוגמה:

- בחרו גורם אחד להיות a וגורם אחר להיות b. מהו A ומהו B?
- כמה קבוצות שונות נוצרו במחקר?

נסמן ב- m את מספר התצפיות בכל תא (בהנחה שהוא יהיה מספר קבוע). תא הוא שילוב של רמה מסוימת של גורם a עם רמה מסוימת של גורם b.

המשך הדוגמה:

- כמה תאים (קבוצות) יש במחקר?
- מה ערכו של m?
- מהו הקשר המתמטי בין m לבין n, גודל המדגם?

נסמן ב- a_1 את הרמה הראשונה של a, ב- a_2 את הרמה השנייה שלו וכך הלאה.

נסמן ב- b_1 את הרמה הראשונה של b, ב- b_2 את הרמה השנייה שלו וכך הלאה.

נסמן ב- $\mu_{i.}$ את התוחלת ברמה a_i. נסמן ב- $\mu_{.j}$ את התוחלת ברמה b_j. נסמן ב- μ_{ij} את התוחלת של תא ij.

<u>המשך הדוגמה:</u>

- מה המשמעות של $\mu_{1.}$ ושל $\mu_{.2}$?
- מה המשמעות של μ_{12} ושל μ_{21}?

<u>השערות המחקר בניתוח שונות דו-כיווני</u>

את השערות המחקר בניתוח שונות דו-כיווני אפשר לרשום בצורות רבות:

H_0: לגורם a אין השפעה על המשתנה התלוי

H_1: אחרת

H_0: לגורם b אין השפעה על המשתנה התלוי

H_1: אחרת

H_0: אין אינטראקציה בין שני הגורמים

H_1: אחרת

דרך אחרת היא שימוש בתוחלות:

$H_0: \mu_{1.} = \mu_{2.} = \cdots = \mu_{A.}$

$H_1:$ אחרת

$H_0: \mu_{.1} = \mu_{.2} = \cdots = \mu_{.B}$

$H_1:$ אחרת

$H_0:$ אין אינטראקציה בין שני הגורמים

$H_1:$ אחרת

<u>המשך הדוגמה:</u>

אם אנחנו מעוניינים לבצע ניתוח שונות דו-כיווני, מה הן ההשערות הנחקרות?

שאלות

1) בחברת טקסטיל בחנו 4 סוגי בדים שונים מבחינת חוזקם.
דגמו 5 חתיכות בד מכל סוג ובדקו את חוזק הקריעה של כל סוג בד.

א. מהו המשתנה התלוי במחקר?
ב. כמה משתנים בלתי תלויים יש במחקר? מה הם?
ג. מהו המבחן הסטטיסטי המתאים במקרה זה?

2) במחקר בתחום הפסיכולוגיה נדגמו אנשים הסובלים מחרדות מסוגים שונים. כל מטופל סווג כסובל מאחד מסוגי החרדות הבאים: חרדה חברתית, חרדה כללית או אגורפוביה. במחקר השתתפו 6 מטופלים מכל סוג חרדה שצוין. המטופלים במחקר חולקו כך שכל מטופל היה צריך לעבור במשך שנה את אחד מהטיפולים הבאים: טיפול קוגניטיבי התנהגותי (CBT), טיפול קבוצתי או טיפול דיאלקטי התנהגותי (DBT). בכל סוג טיפול השתתפו 2 מטופלים מכל סוג חרדה. בסוף השנה נבדקו כל המטופלים וקיבלו ציון כמותי על השיפור במצבם הנפשי (משתנה כמותי). מטרת המחקר הייתה לבדוק האם סוג החרדה, סוג הטיפול והשילוב ביניהם משפיעים על המצב הנפשי של המטופלים.

א. מהו גודל המדגם?
ב. מהו המשתנה התלוי במחקר הזה ומה הם המשתנים הבלתי תלויים?
ג. כמה קטגוריות יש לכל משתנה בלתי תלוי?
ד. כמה קבוצות שונות יש במערך המחקרי?
ה. מהו המבחן הסטטיסטי המתאים במערך מחקרי זה?

3) מחקר שיווקי בדק את השפעת גובה המדף בסופרמרקט והשפעת החומר שממנו עשוי הבקבוק (זכוכית או פלסטיק) על היקף המכירות של משקאות קלים. נבדקו שני סופרמרקטים. בכל סופרמרקט נבחן כל צירוף אפשרי של גובה המדף וחומר הבקבוק, ועבור כל צירוף כזה נבדק מספר בקבוקי המשקה הקל שנמכרו באותו סופרמרקט ביום מסוים. הנה התוצאות שהתקבלו :

סוג בקבוק / גובה המדף	זכוכית	פלסטיק
נמוך	23 32	59 63
בינוני	47 55	88 90
גבוה	40 48	51 56

א. מהו המבחן הסטטיסטי המתאים? נמקו.
ב. מהו מספר הרמות של כל גורם מחקרי?
ג. מה יהיו השערות המחקר אם יתבצע ניתוח שונות דו-כיווני?
ד. מהו ערכו של m ומהו ערכו של n?

4) יצרן של נוזל כביסה מעוניין לבחון שני נוזלי ניקוי מבחינת יעילותם בהסרת כתמים בשלוש רמות טמפרטורה. בכל אחד משׁשת הצירופים של סוג נוזל וטמפרטורה נבחנה יכולת הסרת הכתמים מבדים דומים, וניתן ציון בין 0 ל-15 (הציון הטוב ביותר).

מספר סידורי	סוג הנוזל	טמפרטורה במעלות צלזיוס	ציון הסרת כתמים
1	C	30	4
2	C	30	5
3	C	30	4
4	C	30	6
5	C	40	6
6	C	40	6
7	C	40	7
8	C	40	6
9	C	60	9
10	C	60	8
11	C	60	7
12	C	60	10
13	w	30	9
14	w	30	9
15	w	30	9
16	w	30	10
17	w	40	12
18	w	40	13
19	w	40	11
20	w	40	11
21	w	60	14
22	w	60	14
23	w	60	15
24	w	60	13

א. כמה משתנים יש במחקר?
ב. לגבי כל משתנה קבעו האם הוא משתנה תלוי או בלתי תלוי.
ג. כמה רמות יש לכל גורם?
ד. אם נבצע ניתוח שונות דו-כיווני, מה יהיו השערות המחקר?
ה. רכזו את נתוני המחקר בטבלה שבה בשורות גורם אחד, בעמודות גורם שני ובתאים התוצאות שהתקבלו למשתנה התלוי.

5) קבעו לגבי כל אחד מהבאים האם הוא משתנה קטגוריאלי:
א. מספר הניתוחים שעבר אדם בחייו.
ב. אחוז האבטלה בישראל בחודש זה.
ג. סוג הדם של חולה.
ד. שונות הציונים בבחינת הבגרות באנגלית במועד האחרון.
ה. משקל חבילה בדואר בגרמים.
ו. היבשת שאירחה את משחקי המונדיאל.

בשאלות הבאות יש לבחור את התשובה הנכונה ביותר:

6) משרד החינוך רוצה לבדוק עד כמה שיטת הוראה (יש 3 שיטות הוראה מקובלות) ומגדר משפיעים על ציוני הבגרות בהיסטוריה. מהו המבחן הסטטיסטי המתאים למחקר זה?
א. מבחן T להשוואת תוחלות.
ב. ניתוח שונות חד-כיווני.
ג. ניתוח שונות דו-כיווני.
ד. מבחן T לתוחלת אחת.

7) מחלקת שירות הלקוחות של חברת החשמל דגמה עובדים כדי לבחון האם ככל שמספר שנות הוותק של נותן השירות גדול יותר גדל גם מספר הלקוחות שבו הוא מטפל במהלך משמרת. מהו המבחן הסטטיסטי שיכול לבדוק זאת?
א. מבחן T להשוואת תוחלות.
ב. ניתוח שונות חד-כיווני.
ג. ניתוח שונות דו-כיווני.
ד. אף אחת מהאפשרויות שלעיל.

8) האיחוד האירופי המשותף דגם 10 עובדים מתחום ההוראה בכל אחת מהמדינות הבאות: הולנד, צרפת, בלגיה, גרמניה ואוסטריה. לכל עובד בדקו את גובה המשכורת החודשית שלו ביורו. אם נרצה להשוות בין המדינות הללו מבחינת תוחלת השכר של עובדי ההוראה במדינה, מה יהיה המבחן הסטטיסטי המתאים?

א. מבחן T להשוואת תוחלות
ב. ניתוח שונות חד-כיווני
ג. ניתוח שונות דו-כיווני
ד. אף אחת מהאפשרויות שלעיל

9) בקו ייצור 2 סוגים של מכונות ו-3 רמות ותק של מפעיל המכונה (עד שנתיים במפעל, בין שנתיים ל- חמש שנים במפעל, יותר מחמש שנים במפעל). מנהל הייצור רוצה לבדוק אם קיימת השפעה של סוג המכונה והוותק של המפעיל על מספר המוצרים הפגומים שיוצאים מהמכונה. מה יהיה המבחן הסטטיסטי המתאים במקרה זה?

א. מבחן T להשוואת תוחלות.
ב. ניתוח שונות חד-כיווני.
ג. ניתוח שונות דו-כיווני.
ד. ניתוח שונות תלת-כיווני.

10) במחקר נאספו הנתונים הבאים על קבוצת נחקרים:

1. כמה כוסות קפה הנחקר שותה ביום: לא שותה / 1-2 כוסות/ יותר מ-2 כוסות.
2. מין הנחקר: גבר/אישה.
3. דופק (מספר פעימות בדקה) שעתיים אחרי הקימה.

מטרת המחקר הייתה לבדוק האם מספר כוסות הקפה שאדם שותה ביום משפיע על הדופק אצל גברים אחרת מאשר אצל נשים.
מה יהיה המבחן הסטטיסטי המתאים במקרה זה?

א. מבחן T להשוואת תוחלות.
ב. ניתוח שונות חד-כיווני.
ג. ניתוח שונות דו-כיווני.
ד. ניתוח שונות תלת-כיווני.

11) במחקר יש משתנה כמותי אחד ושני גורמים שלכל אחד מהם שתי רמות.
אילו מהמשפטים הבאים אינו נכון?
א. אפשר מבחינה טכנית לבדוק כיצד כל גורם בנפרד משפיע על המשתנה התלוי באמצעות ניתוח שונות חד-כיווני שייערך לכל גורם בנפרד.
ב. אפשר מבחינה טכנית להשוות בין התוחלות של כל רמה בגורם הראשון על ידי מבחן T להשוואת תוחלות.
ג. אפשר מבחינה טכנית לבצע ניתוח שונות דו-כיווני במערך מחקרי זה.
ד. כיוון שבמחקר יש בסך הכול שלושה משתנים, אפשר מבחינה טכנית לבצע ניתוח שונות תלת-כיווני.

תשובות סופיות

1) א. חוזק הקריעה. ב. אחד. סוג הבד.
ג. ניתוח שונות חד גורמי.

2) א. 18
ב. המשתנה התלוי: ציון במצב הנפש. המשתנים הב"ת: סוג חרדה, סוג הטיפול.
ג. 3,3 ד. 9
ה. ניתוח שונות דו גורמי.

3) א. ניתוח שונות דו גורמי. ב. 3,2
ג. H_0: אין אינטראקציה, H_1: יש אינטראקציה. ד. $m=2,\ n=12$

4) א. 3
ב. משתנים ב"ת: סוג הנוזל, טמפרטורה. משתנה תלוי: ציון הסרת כתמים.
ג. 3,2
ד. H_0: אין אינטראקציה בין הגורמים, H_1: אחרת.
ה. עיין בסרטון הוידאו.

5) א. כן. ב. לא.
ג. כן. ד. לא.
ה. תלוי. ו. כן.

6) ג.

7) ד

8) ב

9) ג

10) ג

11) ד

אפקטים פשוטים, עיקריים ואינטראקציה

רקע

בניתוח שונות דו-כיווני אנו דנים במשתנה כמותי תלוי יחיד ובשני משתנים בלתי תלויים (גורמים) המחולקים כל אחד למספר רמות.
מטרת המחקר היא לבדוק שלוש השערות שונות:

H_0: לגורם a אין השפעה על המשתנה התלוי

H_1: אחרת

H_0: לגורם b אין השפעה על המשתנה התלוי

H_1: אחרת

H_0: אין אינטראקציה בין שני הגורמים

H_1: אחרת

נרצה להבין מה בדיוק כל השערה בודקת לגבי האוכלוסייה הנחקרת.

<u>אפקט עיקרי</u>: אם יש שתי קטגוריות (רמות) לפחות של גורם מסוים שהתוחלות שלהן שונות, נאמר שלגורם זה יש השפעה על המשתנה התלוי. השפעה זאת נקראת "אפקט עיקרי". למשל, אם יימצאו לפחות שתי תרופות נוגדות דיכאון שונות שמביאות לתוחלות שונות במצב הנפשי, נגיד שלסוג התרופה יש השפעה על המצב הנפשי, כלומר יש אפקט עיקרי. כמות האפקטים העיקריים שאפשר למצוא היא ככמות הגורמים במחקר.

אפקט אינטראקציה: מצב שבו גורם אחד משפיע על המשתנה התלוי באופן שונה בקטגוריות שונות של הגורם השני. למשל, תרופה נוגדת דיכאון אחת מביאה את הגברים למצב רוח טוב יותר מאשר את הנשים לעומת תרופה אחרת שמביאה דווקא את הנשים למצב רוח טוב יותר מאשר את הגברים. אפקט האינטראקציה הוא יחיד, כלומר נאמר אם יש או אין אינטראקציה. כמו כן הוא אפקט סימטרי: אם קיימת אינטראקציה בין מגדר לסוג התרופה, יש גם אינטראקציה בין סוג התרופה למגדר.

אפקט פשוט: אפקט פשוט מתייחס להשפעת גורם אחד על המשתנה התלוי בתוך קטגוריה מסוימת של הגורם השני. למשל, נרצה לבדוק רק בקטגוריה של הגברים האם קיים הבדל בין התרופות נוגדות הדיכאון. אם נמצא הבדלים כאלה נאמר שיש

אפקט פשוט של סוג התרופה בקרב אוכלוסיית הגברים. כמות האפקטים הפשוטים שאפשר למצוא היא סכום מספר הקטגוריות (רמות) של כל גורם. למשל, אם יש שלושה סוגי תרופות ושתי אפשרויות למגדר, בסך הכול נוכל לבדוק 5 אפקטים פשוטים.

<u>דוגמה</u>

נבדקו שלושה סוגי דיאטות על אנשים בעלי משקל עודף. כעבור שלושה חודשים בדקו כמה קילוגרמים הפחית כל מטופל ממשקלו באותה התקופה. נניח שאנו יודעים את תוחלת הפחתת המשקל של כל דיאטה בחלוקה למגדרים.

נתאר כמה מצבים אפשריים לגבי האוכלוסייה הנחקרת וננתח כל מצב מבחינת ההשפעה של כל גורם על תוחלת המשתנה התלוי ומבחינת אפקט האינטראקציה.

שימו לב שהמצבים שנתאר להלן מתייחסים לתוחלות האמיתיות. בניתוח שונות אין לנו נתוני אמת, אלא רק נתוני מדגם, ונרצה לבדוק האם האפקטים שהתקבלו במדגם הם מובהקים, כנדרש בכל תהליך של הסקה סטטיסטית.

אם התוצאות שלנו יהיו ממוצעי מדגם ולא תוחלות, נוכל לבדוק אם קיימים אפקטים במדגם, אך אין זה אומר שקיימים אפקטים באוכלוסייה, כלומר לא נוכל לדעת אם האפקטים במדגם הם מובהקים. כדי לבדוק אם האפקטים הם מובהקים נצטרך לעשות את מבחן ניתוח השונות.

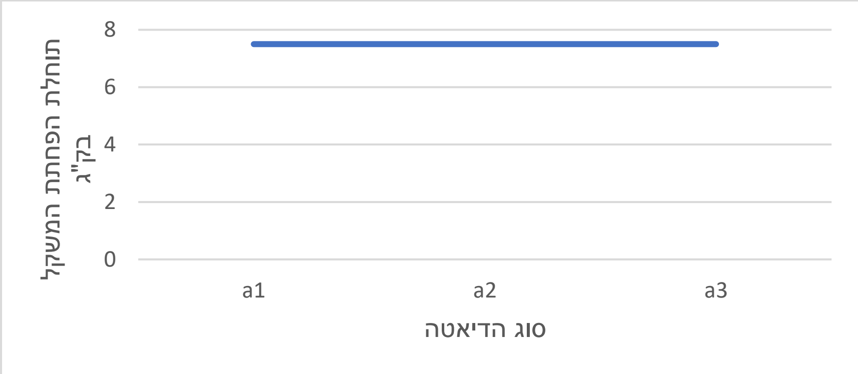
<u>מצב א:</u>

הטבלה הבאה מתארת את תוחלת הפחתת המשקל בק״ג לכל קבוצה:

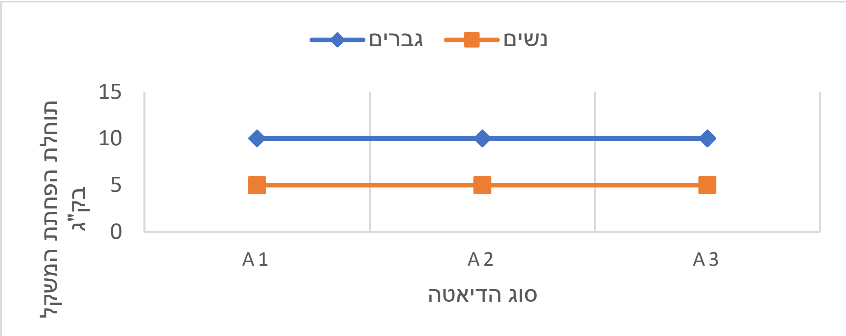
	גברים	נשים
a_1	10	5
a_2	10	5
a_3	10	5

<u>תיאור גרפי לבדיקת אפקט למגדר</u>

<u>תיאור גרפי לבדיקת אפקט לסוג הדיאטה</u>



<u>גרף אפקטים פשוטים</u>



<u>ניתוח המצב</u>: למגדר יש אפקט, לסוג הדיאטה אין אפקט, אין אפקט אינטראקציה.

הערה: אם הקווים הנוצרים בגרף האפקטים הפשוטים מקבילים או מתלכדים, אנו אומרים שאין אפקט אינטראקציה.

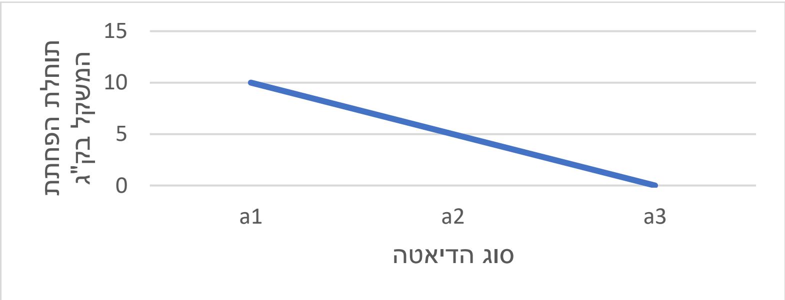
<u>מצב ב</u>

הטבלה הבאה מתארת את תוחלת הפחתת המשקל בק״ג לכל קבוצה :

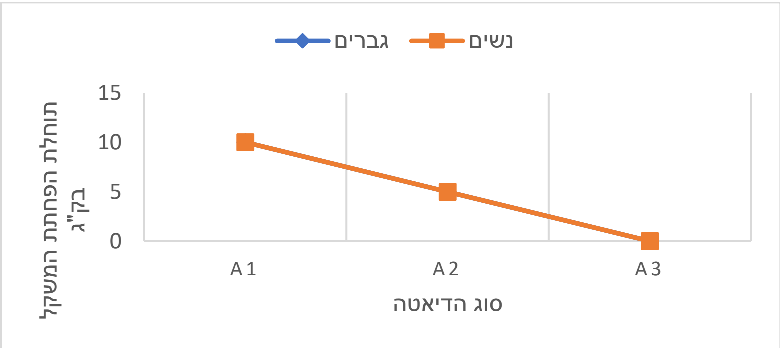
	גברים	נשים
a_1	10	10
a_2	5	5
a_3	0	0

<u>תיאור גרפי לבדיקת אפקט למגדר</u>

<u>תיאור גרפי לבדיקת אפקט לסוג הדיאטה</u>



<u>גרף אפקטים פשוטים</u>



<u>ניתוח המצב</u> : למגדר אין אפקט, לסוג הדיאטה יש אפקט, אין אפקט אינטראקציה.

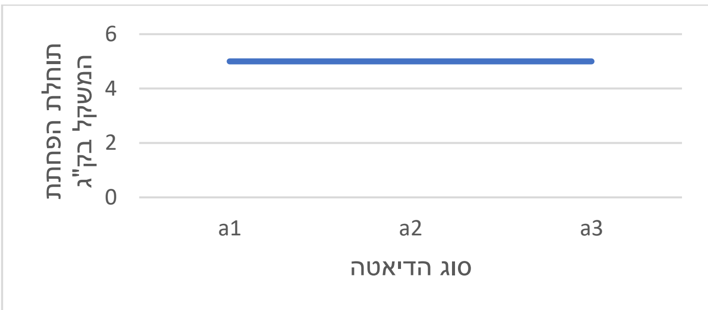
מצב ג

הטבלה הבאה מתארת את תוחלת הפחתת המשקל בק״ג לכל קבוצה:

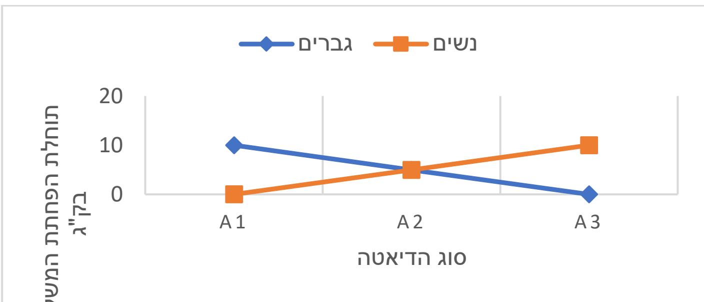
	גברים	נשים
a_1	10	0
a_2	5	5
a_3	0	10

תיאור גרפי לבדיקת אפקט למגדר

תיאור גרפי לבדיקת אפקט לסוג הדיאטה



גרף אפקטים פשוטים



ניתוח המצב: למגדר אין אפקט, לסוג הדיאטה אין אפקט, יש אפקט אינטראקציה.

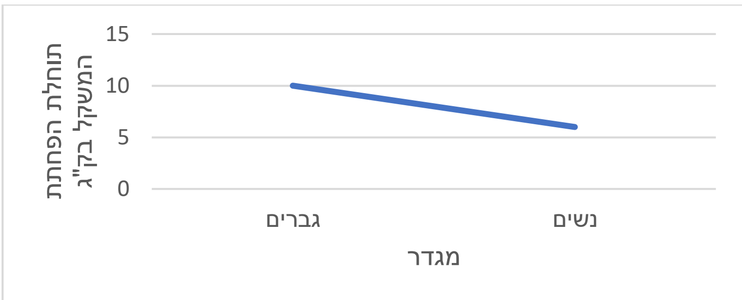
אינטראקציה דיסאורדינלית (נקראת גם "אינטראקציה מהותית"):
אפשר לזהות מצב של אינטראקציה כזו באמצעות גרף של אפקטים פשוטים, כאשר נוצרים קווים נחתכים שאחד מהם עולה והאחר יורד. המשמעות היא שגורם אחד משפיע על המשתנה התלוי ברמה מסוימת של הגורם השני באופן הפוך משהוא משפיע על המשתנה התלוי ברמה אחרת של הגורם השני.
במצב זה אין להתייחס לאפקטים עיקריים. יש להתייחס רק לאפקטים הפשוטים.

<u>מצב ה</u>

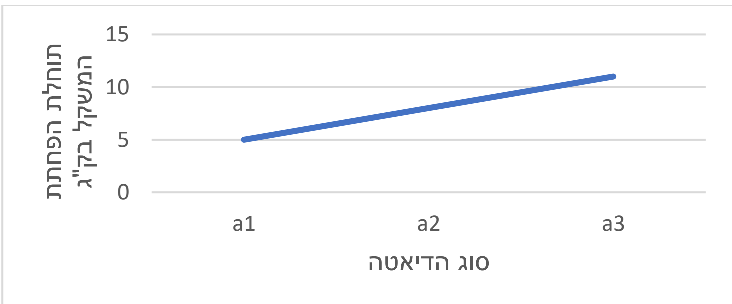
הטבלה הבאה מתארת את תוחלת הפחתת המשקל בק"ג לכל קבוצה:

	גברים	נשים
a_1	5	5
a_2	10	6
a_3	15	7

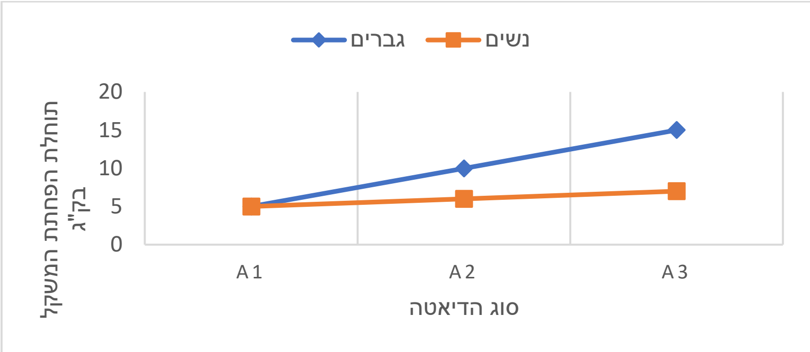
<u>תיאור גרפי לבדיקת אפקט למגדר</u>



<u>תיאור גרפי לבדיקת אפקט לסוג הדיאטה</u>



<u>גרף אפקטים פשוטים</u>



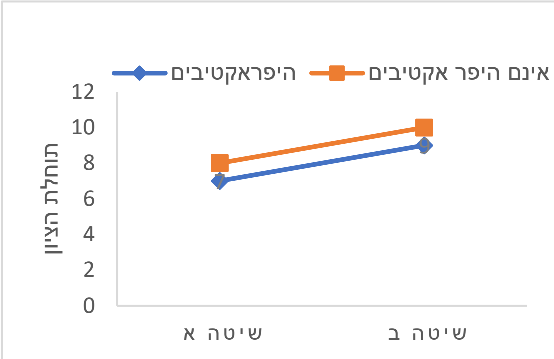
<u>ניתוח המצב</u>: למגדר יש אפקט, לסוג הדיאטה יש אפקט, יש אפקט אינטראקציה.

אינטראקציה אורדינלית (נקראת גם "אינטראקציה לא מהותית"): אפשר לזהות מצב של אינטראקציה כזו כאשר בגרף האפקטים הפשוטים נוצרים קווים נחתכים עם אותו הכיוון (כולם עולים או כולם יורדים אבל לא באותו השיפוע). המשמעות היא שבמעבר של גורם אחד מרמה אחת לרמה אחרת שלו הוא משפיע על המשתנה התלוי באותו אופן בכל רמה של המשתנה האחר אבל עם גודל אפקט שונה.

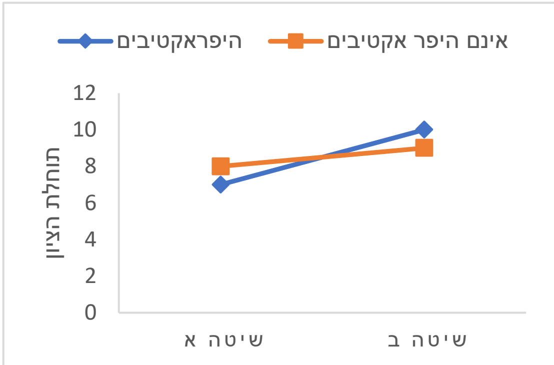
שאלות

1) בגני החובה יש שתי שיטות הוראה. שיטות אלו נוסו על ילדים היפראקטיביים וילדים שאינם היפראקטיביים. בתרשימים הבאים מיוצגים גרפים שמתארים את תוחלת הציון במבחן אוצר המילים שניתן לילדים בסוף השנה.
בכל אחד מהמקרים יש לקבוע האם קיימת אינטראקציה בין שני הגורמים. אם קיימת אינטראקציה, יש לקבוע האם היא אינטראקציה אורדינלית או דיסאורדינלית.

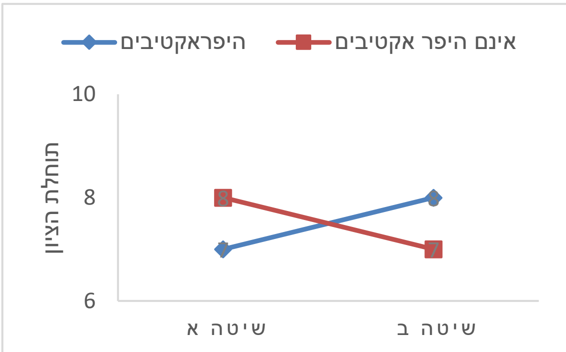
א.



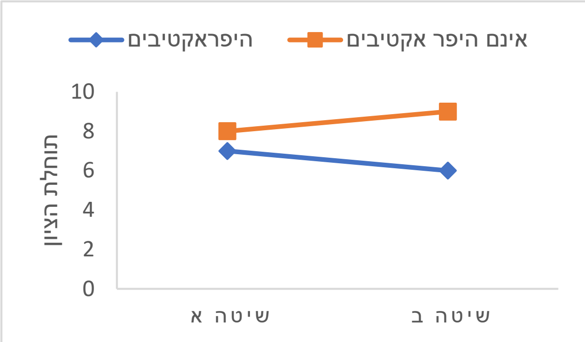
ב.



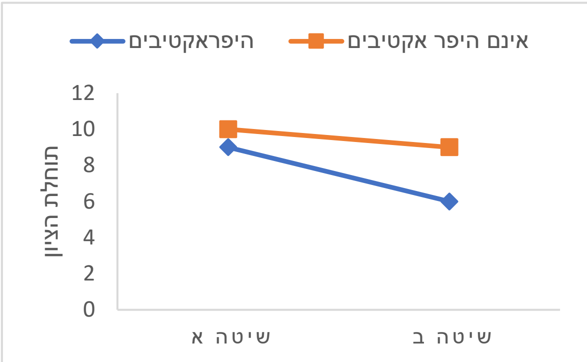
ג.



ד.



ה.



2) משרד האוצר פרסם נתונים על המחיר הממוצע של דירות גן ודירות גג של 4 חדרים ב-3 ערים בארץ. מחיר הדירות נמדד במיליוני שקלים.
להלן התוצאות שהתקבלו:

	דירות גן	דירות גג
הרצליה	4	3
אשדוד	2	1
חולון	3	2

א. מהו המשתנה התלוי ומה הם המשתנים הבלתי תלויים?
ב. האם קיים אפקט לעיר? היעזרו בגרף מתאים.
ג. האם קיים אפקט לסוג הדירה? היעזרו בגרף מתאים.
ד. האם קיימת אינטראקציה בין הגורמים? אם כן, מהו סוג האינטראקציה? היעזרו בגרף מתאים.
ה. האם יש אפקט פשוט לעיר עבור דירות גן?
ו. האם יש אפקט פשוט לעיר עבור דירות גג?
ז. האם יש אפקט פשוט לסוג הדירה בהרצליה?
ח. האם יש אפקט פשוט לסוג הדירה באשדוד?
ט. האם יש אפקט פשוט לסוג הדירה בחולון?

3) משרד החינוך פרסם נתונים על תוחלת הציונים בבחינת הבגרות באנגלית לפי עיר וסוג בית הספר (עיוני או מקצועי). להלן התוצאות שהתקבלו:

	עיוני	מקצועי
רעננה	85	70
תל אביב	75	75
פתח תקווה	70	85

א. תארו את הנתונים באמצעות גרף אפקטים פשוטים.
ב. האם קיימת אינטראקציה בין הגורמים? אם כן, מה סוג האינטראקציה?
ג. באילו ערים קיים אפקט פשוט לסוג בית הספר?

4) משרד התחבורה פרסם נתונים על תוחלת מספר עבירות התנועה לבעלי רישיון נהיגה לפי עיר ולפי מגדר. להלן התוצאות שהתקבלו:

	גבר	אישה
חיפה	2	1
אשקלון	2	1
רמת גן	2	1

א. האם קיים אפקט עיקרי לעיר?
ב. האם קיים אפקט עיקרי למגדר?
ג. האם יש אפקט פשוט לעיר אצל הגברים?
ד. האם קיימת אינטראקציה בין הגורמים? אם כן, מהו סוג האינטראקציה?

5) המשרד לאיכות הסביבה פרסם נתונים על תוחלת רמת זיהום האוויר בערים שונות בארץ בחורף ובקיץ. להלן התוצאות שהתקבלו:

	קיץ	חורף
חיפה	20	20
ירושלים	10	10
באר שבע	15	15

א. האם קיים אפקט עיקרי לעיר?
ב. האם קיים אפקט עיקרי לעונה?
ג. האם קיימת עיר שבה יש אפקט פשוט לעונה?
ד. האם קיימת אינטראקציה בין הגורמים? אם כן, מה סוג האינטראקציה?

6) המשרד לאיכות הסביבה פרסם נתונים על תוחלת רמת זיהום האוויר בערים שונות בארץ בחורף ובקיץ. להלן התוצאות שהתקבלו:

	קיץ	חורף
רמת גן	10	10
גבעתיים	10	10
בת ים	10	10

האם קיים אפקט עיקרי לגורם כלשהו? האם קיימת אינטראקציה?

בשאלות הבאות יש לבחור את התשובה הנכונה ביותר:

7) במחקר נדגמו 5 אנשים מכל אחת מ-4 הקבוצות הבאות: 1. מתעמלים באופן קבוע ושומרים על תזונה בריאה; 2. מתעמלים באופן קבוע ולא שומרים על תזונה בריאה; 3. לא מתעמלים באופן קבוע ושומרים על תזונה בריאה; 4. לא מתעמלים באופן קבוע ולא שומרים על תזונה בריאה. להלן טבלה המסכמת את ממוצע הטריגליצרידים בדם (מ״ג לדציליטר) שנמצא בכל מדגם:

	תזונה בריאה	לא תזונה בריאה
מתעמלים	90	100
לא מתעמלים	100	160

א. קיים אפקט עיקרי מובהק לגורם ההתעמלות.
ב. קיים אפקט עיקרי מובהק לגורם התזונה.
ג. קיים אפקט אינטראקציה מובהק בין שני הגורמים במחקר.
ד. אי אפשר לדעת אם קיים אפקט מובהק כלשהו על סמך תוצאות המדגם בלבד ללא ביצוע מבחן מתאים וללא קביעת רמת המובהקות של המחקר.

8) במחקר בדקו 3 טיפולים שונים לחולי פסוריאזיס. המחקר השווה גם בין גברים לנשים ובדק את זמן התגובה לטיפול. מסקנת המחקר הייתה שאצל גברים נמצאו הבדלים מובהקים בין הטיפולים השונים מבחינת תוחלת זמן התגובה. לאיזה סוג אפקט המסקנה מתייחסת?

א. אפקט אינטראקציה.
ב. אפקט עיקרי של גורם המין.
ג. אפקט עיקרי של גורם סוג הטיפול.
ד. אפקט פשוט.

9) במחקר בדקו 3 טיפולים שונים לחולי פסוריאזיס. המחקר השווה גם בין גברים לנשים ובדק את זמן התגובה לטיפול. במדגם היה ממוצע זמן התגובה של הגברים שונה מממוצע זמן התגובה של הנשים.

א. אפשר להגיד שבמדגם קיים אפקט עיקרי, אך אי אפשר לדעת אם האפקט העיקרי מובהק.
ב. אפשר להגיד שבמדגם קיימת אינטראקציה, אך אי אפשר לדעת אם האינטראקציה מובהקת.
ג. אפשר להגיד שקיים אפקט עיקרי מובהק.
ד. אפשר להגיד שקיימת אינטראקציה מובהקת.

10) במחקר בדקו 3 טיפולים שונים לחולי פסוריאזיס. המחקר השווה גם בין גברים לנשים ובדק את זמן התגובה לטיפול. אחת המסקנות של המחקר הייתה שהטיפולים השונים משפיעים במידה משמעותית יותר על זמן התגובה של הגברים מאשר על זה של הנשים, אם כי באותו האופן.

א. המסקנה היא שאין אינטראקציה בין הגורמים במחקר.
ב. המסקנה היא שיש אינטראקציה אורדינלית בין הגורמים במחקר.
ג. המסקנה היא שיש אינטראקציה דיסאורדינלית בין הגורמים במחקר.
ד. המסקנה היא שיש אפקט עיקרי של המגדר.

תשובות סופיות

1) א. אין אינטראקציה. ב. אינטראקציה אורדינלית.
ג. אינטראקציה דיסאורדנלית. ד. אינטראקציה דיסאורדנלית.
ה. אינטראקציה אורדינלית.

2) א. המשתנים הב"ת: העיר, סוג הדירה. המשתנה התלוי: מחיר.
ב. קיים. ג. קיים.
ד. לא קיים. ה. קיים.
ו. קיים. ז. קיים.
ח. קיים. ט. קיים.

3) א. עיין בסרטון הוידאו. ב. אינטראקציה דיסאורדינלית.
ג. רעננה ופתח תקווה.

4) א. לא. ב. כן.
ג. לא. ד. לא.

5) א. כן. ב. לא.
ג. לא. ד. לא.

6) לא, לא.

7) ד

8) ד

9) א

10) ב

תהליך ניתוח שונות דו כיווני – הליך מבחן

רקע

כפי שכבר ציינו, ניתוח שונות דו-כיווני נעשה כאשר יש שני גורמים מחקריים ומשתנה כמותי תלוי אחד. מטרת המחקר היא לבדוק האם הגורמים משפיעים על המשתנה התלוי. מערך מחקר זה נקרא "מערך מחקר פקטוריאלי", כיוון שאנו בונים את המחקר לפי גורמים. מערך דו-גורמי יסומן כמערך מסוג $A\times B$, כאשר A מייצג את מספר הרמות של גורם a, ו-B מייצג את מספר הרמות של גורם b.

במערך מחקרי תלת-גורמי נסמן את סוג המערך $A\times B\times C$, וכך הלאה.

<u>דוגמה</u>

נבדקו שלושה סוגי דיאטות על אנשים בעלי משקל עודף. נבחרו 18 מטופלים בעלי משקל עודף, 9 מהם גברים ו-9 נשים. המטופלים חולקו כך שבכל דיאטה השתתפו 3 גברים ו-3 נשים. כעבור שלושה חודשים מתחילת הדיאטה נשקלו כלל המטופלים ונבדק המשקל בק"ג שהם הפחיתו. הטבלה הבאה מסכמת את המשקל שכל מטופל במדגם הפחית כעבור שלושה חודשים.

סוג הדיאטה / מין	b_1	b_2	b_3	סה"כ
נשים	8 4 0	6 8 10	4 6 8	54
גברים	6 10 14	0 2 4	9 12 15	72
סה"כ	42	30	54	126

מטרת המחקר היא לבדוק האם יש השפעה של סוג הדיאטה, המין והשילוב ביניהם על ההפחתה במשקל.

- באיזה סוג מערך מחקרי מדובר?
- מהו המבחן הסטטיסטי המתאים לבדיקת ההשערות?
- מה הן השערות המחקר?

בדומה לניתוח שונות חד-כיווני גם התהליך של ניתוח שונות דו-כיווני דורש הנחות. ההנחות הן :

1. $A \times B$ הקבוצות שנוצרות בלתי תלויות זו בזו.
2. בכל $A \times B$ האוכלוסיות המשתנה התלוי מתפלג נורמלית.
3. בכל $A \times B$ האוכלוסיות אותה שונות, σ^2.

הערה : ניתוח שונות הוא מבחן רובסטי, כלומר יש לו רגישות נמוכה להנחות. התיאוריה הסטטיסטית שפותחה התבססה על ההנחות האלה, אבל הלכה למעשה השיטה תעבוד טוב גם אם ההנחות הללו לא יתקיימו במדויק במלואן. זו הסיבה שהשיטה הזו נפוצה כל כך בעולם הסטטיסטיקה.

בהמשך לדוגמה

רשמו את כל ההנחות הדרושות לביצוע ניתוח השונות.

הליך המבחן

בניית טבלת ממוצעים

נבנה טבלת ממוצעים לכל רמה ולכל תא:

$\bar{X}_{i.}$ – ממוצע המדגם ברמה i של גורם a

$\bar{X}_{.j}$ – ממוצע המדגם ברמה j של גורם b

$\bar{X}_{ij}$ – ממוצע המדגם בתא ij

בהמשך לדוגמה

- מלאו את טבלת הממוצעים הבאה:

סוג הדיאטה / מין	b_1	b_2	b_3	$\bar{X}_{i.}$
נשים				
גברים				
$\bar{X}_{.j}$				

- שרטטו גרפים מתאימים לבדיקת אפקטים עיקריים ולבדיקת אינטראקציה במדגם. האם אפשר להגיד שיש אפקט מובהק?

<u>בניית טבלת ריבועי הפרשים מהממוצעים</u>

נמלא את הטבלה הבאה. בתוך תא ij נחשב: $(\bar{X}_{ij}-\bar{X}_{i.}-\bar{X}_{.j}+\bar{X})^2$

<u>בהמשך לדוגמה</u>

- מלאו את טבלת הפרשי הממוצעים:

סוג הדיאטה / מין	b_1	b_2	b_3	$(\bar{X}_{i.}-\bar{X})^2$
נשים				
גברים				
$(\bar{X}_{.j}-\bar{X})^2$				

חישוב סכום ריבועי הסטיות מהממוצע

מתוך טבלת ריבועי הסטיות מהממוצע נחשב את סכום ריבועי הסטיות מהממוצע הבאים:

הסימון SS הוא ראשי התיבות של "sum of squares" (סכום הריבועים).

סכום ריבועי הסטיות מהממוצע של גורם a:

$$SS_a = m \cdot B \sum_{i=1}^{A} (\bar{X}_{i.} - \bar{X})^2$$

סכום ריבועי הסטיות מהממוצע של גורם b:

$$SS_b = m \cdot A \sum_{j=1}^{B} (\bar{X}_{.j} - \bar{X})^2$$

סכום ריבועי הסטיות של האינטראקציה:

$$SS_{ab} = m \sum_{i=1}^{A} \sum_{j=1}^{B} (\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X})^2$$

סכום ריבועי השגיאות (סכום ריבועי הסטיות של התצפיות בתא מהממוצע בתא):

$$SS_W = \sum_{i=1}^{A} \sum_{j=1}^{B} \sum_{k=1}^{m} (X_{ijk} - \bar{X}_{ij})^2 = (m-1) \sum_{i=1}^{A} \sum_{j=1}^{B} S_{ij}^2$$

סכום ריבועי הסטיות של כלל התצפיות מהממוצע הכללי:

$$SS_T = \sum_{i=1}^{A} \sum_{j=1}^{B} \sum_{k=1}^{m} (X_{ijk} - \bar{X})^2 = (n-1) \cdot S^2$$

הקשר המתמטי בין סכום הריבועים הללו הוא:

$$SS_T = SS_a + SS_b + SS_{ab} + SS_W$$

לכן אין אנו צריכים לחשב את כל חמשת המרכיבים הללו.

החלק הזה של הנוסחה מתייחס לשונות השיטתית: $SS_a + SS_b + SS_{ab}$. השונות השיטתית היא שונות שמקורה בגורמים עצמם.

החלק הזה של הנוסחה מתייחס לשונות המקרית: SS_W. השונות המקרית היא שונות שנקראת גם "שונות טעויות" או "שונות בתוך הקבוצות". זוהי שונות בין התצפיות שאינה נובעת מהגורמים הנחקרים. האות W מייצגת את המילה "Within", כלומר שונות **בתוך** התאים.

<u>בהמשך לדוגמה</u>

- חשבו את ריבועי הסטיות הבאים :

$SS_a =$

$SS_b =$

$SS_{ab} =$

$SS_T =$

$SS_W =$

חישוב ממוצע ריבועי הסטיות וסטטיסטי המבחן

MS הוא הסימון של ממוצע ריבועי הסטיות (Mean Square) שמהווה אומד לשונות של כל גורם. החישוב ייעשה על ידי חלוקת ה-SS המתאים בדרגות החופש המתאימות. לאחר מכן נחשב שלושה סטטיסטי מבחן, בהתאם לשלוש ההשערות הנבדקות.

נרכז את כלל החישובים הללו בטבלה הנקראת טבלת ניתוח שונות, ANOVA (Analysis of Variance).

מקור השונות **Source of Variation**	דרגות החופש **Degrees of Freedom**	סכום ריבועי הסטיות מהממוצע **Sum of Squares**	ממוצע ריבוע הסטייה **Mean Square**	***F***
a	$A-1$	SS_a	MS_a	$F_a = MS_a / MS_W$
b	$B-1$	SS_b	MS_b	$F_b = MS_b / MS_W$
ab	$(A-1)(B-1)$	SS_{ab}	MS_{ab}	$F_{ab} = MS_{ab} / MS_W$
Within	$AB(m-1)$	SS_W	MS_W	
Total	$n\text{-}1=ABm-1$	SS_T		

<u>בהמשך לדוגמה :</u> מלאו את טבלת ניתוח השונות

F	**ממוצע ריבועי הסטייה** Mean Square	**סכום ריבועי הסטיות מהממוצע** Sum of Squares	**דרגות החופש** Degrees of Freedom	**מקור השונות** Source of Variation
				a
				b
				ab
				Within
				Total

כללי ההכרעה לבדיקת ההשערות

הסטטיסטי F_a מייצג את היחס בין השונות המדגמית של גורם a ובין השונות המקרית. לכן ככל שהערכים שלו גבוהים יותר, נרצה להגיד שלגורם a יש השפעה גדולה יותר על המשתנה התלוי. F_a יקבל ערכים גבוהים אם השונות המדגמית של גורם A תגדל או אם השונות המדגמית המקרית תקטן.

הסטטיסטי מתפלג התפלגות F, ואזור הדחייה שלו יהיה בצד ימין.

- כלל ההכרעה לבדיקת המובהקות של גורם a :

  דחה את השערת H_0 ברמת מובהקות של α אם

  $$F_a > F_{1-\alpha}\left(df_a, df_W\right)$$

לפי אותו עיקרון שאר כללי ההכרעה יהיו :

- כלל ההכרעה לבדיקת המובהקות של גורם b :

  דחה את השערת H_0 ברמת מובהקות של α אם

  $$F_b > F_{1-\alpha}\left(df_b, df_W\right)$$

- כלל ההכרעה לבדיקת המובהקות של האינטראקציה :

  דחה את השערת H_0 ברמת מובהקות של α אם

  $$F_{ab} > F_{1-\alpha}\left(df_{ab}, df_W\right)$$

בהמשך לדוגמה

רשמו את כל כללי ההכרעה המתאימים והסיקו מסקנות מתאימות ברמת מובהקות של 5%.

הערות

1. אם מכריעים שקיימת אינטראקציה מובהקת, יש לבדוק האם היא אורדינלית או דיסאורדינלית. אם האינטראקציה דיסאורדינלית, יש לבדוק האם האפקטים העיקריים נמצאו מובהקים. אם לפחות אחד מהם נמצא מובהק נאמר שהוא אינו משמעותי כיוון שהוא נובע מהאינטראקציה בין הגורמים ולא מהגורם עצמו.
2. אם אחד מהאפקטים נמצא מובהק, אין זה אומר אילו רמות שונות זו מזו בתוחלת. למשל, אם נמצא הבדל מובהק בין סוגי הטיפולים, לא נוכל לדעת לפי זה איזה טיפול שונה מאחר באופן מובהק. לכן יש להמשיך בתהליך של השוואות מרובות כדי להסיק ממה נובע השוני.

בהמשך לדוגמה

האם יש סיבה לבצע השוואות מרובות במחקר?

שאלות

1) מחקר שיווקי בדק את השפעת גובה המדף בסופרמרקט והשפעת החומר שממנו עשוי הבקבוק (זכוכית או פלסטיק) על היקף המכירות של משקאות קלים. נבדקו שני סופרמרקטים. בכל סופרמרקט נבחן כל צירוף אפשרי של גובה המדף וחומר הבקבוק, ועבור כל צירוף כזה נבדק מספר בקבוקי המשקה הקל שנמכרו באותו סופרמרקט ביום מסוים. הנה התוצאות שהתקבלו :

סוג בקבוק / גובה המדף	זכוכית	פלסטיק
נמוך	23 32	59 63
בינוני	47 55	88 90
גבוה	40 48	51 56

בצעו ניתוח שונות דו-כיווני על נתוני מחקר זה ברמת מובהקות של 5%. סכמו את המסקנות מתוך ניתוח השונות שביצעתם. מה הן ההנחות הדרושות לביצוע המבחן?

2) במחקר בתחום החקלאות נדגמו 8 חלקות אדמה : 4 חלקות בנגב ו-4 בעמק יזרעאל. בכל חלקה ההשקיה הייתה או באמצעות ממטרות או באמצעות טפטפות. בדקו את יבול העגבניות (בטונה לדונם) בכל חלקה.
להלן התוצאות שהתקבלו :

מספר חלקה	מיקום החלקה	שיטת השקיה	יבול העגבניות
1	נגב	ממטרות	12
2	נגב	ממטרות	10
3	נגב	טפטפות	15
4	נגב	טפטפות	17
5	עמק יזרעאל	ממטרות	12
6	עמק יזרעאל	ממטרות	14
7	עמק יזרעאל	טפטפות	17
8	עמק יזרעאל	טפטפות	19

א. רשמו את כלל המשתנים במחקר וציינו לגבי כל אחד מהם האם הוא משתנה תלוי או בלתי תלוי.

ב. הציגו את נתוני המחקר באמצעות גרפים מתאימים. האם נראה שבמדגם יש אפקט עיקרי לכל גורם? האם יש אינטראקציה בין הגורמים במדגם? האם האפקטים מובהקים?

ג. בדקו ברמת מובהקות של 5% האם האפקט העיקרי של כל גורם הוא מובהק והאם האינטראקציה היא מובהקת. מה הן ההנחות הדרושות?

3) חברה לייצור מוצרי שיער פיתחה נוסחה חדשנית לצבע לשיער שאינו דורש תוספת חמצן בעת תהליך הצביעה. החברה השוותה את צבע השיער החדש לצבע השיער הרגיל מבחינת כושר הכיסוי וזאת על שלושה סוגי שיער: בהיר, כהה ושיבה. ציון רמת הכיסוי הוא משתנה שמתפלג נורמלית עם שונות קבועה לכל סוג שיער ולכל סוג צבע. לכל קבוצה של סוג צבע וסוג שיער נדגמו 4 צביעות שנוסו על אנשים שונים, וניתן ציון מספרי על רמת הכיסוי. להלן סיכום תוצאות המדגם שהתקבלו:

הקבוצה	ממוצע	שונות
צבע רגיל על שיער בהיר	62	40
צבע רגיל על שיער כהה	51	44
צבע רגיל על שיער שיבה	45	42
צבע חדש על שיער בהיר	60	46
צבע חדש על שיער כהה	54	40
צבע חדש על שיער שיבה	44	42

בצעו ניתוח שונות דו-כיווני על הנתונים ברמת מובהקות של 5%. סכמו את כל המסקנות המתקבלות.

4) בוצע ניתוח שונות על נתונים. במערך המחקרי לגורם a יש 4 רמות ולגורם b יש 3 רמות. נערכו 3 תצפיות לכל אחת מ-12 הקבוצות שנוצרו. להלן טבלת ניתוח שונות דו-גורמי שבוצע:

מקור השונות	df	SS	MS	F
a	?	318	?	?
b	?	?	?	?
אינטראקציה	?	190	?	?
W	?	156	?	
T	?	674		

א. מלאו את כל התאים בטבלה המסומנים בסימני שאלה.

ב. בצעו את הבדיקות הבאות ברמת מובהקות של 5%:

i. האם האינטראקציה מובהקת?

ii. האם גורם a משפיע על המשתנה התלוי הנחקר?

iii. האם לגורם b יש לפחות שתי רמות עם תוחלות שונות?

5) במחקר בדקו האם ארץ מוצא ומגדר של אדם משפיעים על שנות ההשכלה שלו. הנתונים סוכמו בטבלת ניתוח שונות:

מקור השונות	df	SS	MS	F
ארץ מוצא	4	34		
מגדר			2	
אינטראקציה		18	4.5	
W	10	12		
T				

א. כמה ארצות מוצא נבדקו במחקר זה?
ב. מהו גודל המדגם הכולל במחקר זה?
ג. חשבו את ערכי F הסטטיסטי עבור ארץ המוצא, המגדר והאינטראקציה.
ד. מה הם האפקטים המובהקים במחקר זה ברמת מובהקות של 5%?

6) בטבלה הבאה מסוכמים הממוצעים של מערך מחקרי דו-גורמי עם משתנה כמותי תלוי:

	b_1	b_2	b_3
a_1	8	14	11
a_2	6	13	16

מספר התצפיות בכל תא הוא 5.
הטבלה הבאה היא טבלה מסכמת של ניתוח השונות על סמך נתוני מחקר זה:

מקור השונות	df	SS	MS	F
a				
b		281.7		
ab		71.7		
W		190.1		
T				

א. מלאו את טבלת ניתוח השונות.
ב. הסיקו מסקנות ברמת מובהקות של 5%.
ג. שרטטו גרף אינטראקציות והסבירו את משמעות הממצאים.

תשובות סופיות

1) עיין בסרטון הוידאו.
2) א. משתנים ב"ת: מיקום החלקה, שיטת השקיה. משתנה תלוי: יבול בטונה לדונם.
ב. עיין בסרטון הוידאו. ג. עיין בסרטון הוידאו
3) עיין בסרטון הוידאו.
4) א. עיין בסרטון הוידאו. ב. i. כן. ii. כן. iii. לא.
5) א. 4 ב. 20 ג. עיין בסרטון הוידאו.
ד. עיין בסרטון הוידאו.
6) א. עיין בסרטון הוידאו. ב. עיין בסרטון הוידאו. ג. עיין בסרטון הוידאו.

ביוסטטיסטיקה

פרק 43 - תרגול שאלות אמריקאיות

תוכן העניינים

תרגול שאלות אמריקאיות:

שאלות:

הנתונים הבאים מתייחסים לשאלות 1-4:

פסיכולוגים צפו במשך שבוע שלם בהתנהגותם של 28 ילדים בגן חובה. לאחר מכן נאלצו לדווח על רמת הביטחון העצמי של כל ילד בסקלה של 1 עד 5. כאשר 5 נחשב לרמת בטחון עצמי גבוהה ו-1 לרמת בטחון עצמי נמוכה. להלן סיכום התוצאות:

בטחון עצמי	מספר הילדים
1	6
2	7
3	10
4	4
5	1

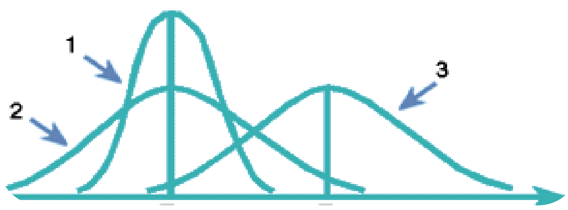
1) מהו סולם המדידה של המשתנה הנחקר?
א. שמי.
ב. סדר.
ג. רווח.
ד. מנה.

2) מהי הדרך הגרפית המתאימה ביותר כדי לתאר את הנתונים?
א. טבלת שכיחויות.
ב. דיאגרמת מקלות.
ג. היסטוגרמה.
ד. דיאגרמת עוגה.

3) מהו השכיח של התפלגות הנתונים שנאספו?
א. 2.
ב. 1.
ג. 3.
ד. 10.

4) התווסף עוד ילד עם רמת בטחון עצמי נמוכה לכן סטיית התקן של המשתנה הנחקר כתוצאה מההוספה:
א. תגדל.
ב. תקטן.
ג. לא תשתנה.
ד. אין לדעת.

הנתונים הבאים מתייחסים לשאלות 5-9:
להלן שלוש התפלגויות נורמליות של שלוש קבוצות שונות ששורטטו באותה מערכת צירים. ההתפלגויות מוספרו כדי להבדיל ביניהן.



5) לאיזו התפלגות הממוצע הגבוה ביותר?
א. 1.
ב. 2.
ג. 3.
ד. אין לדעת.

6) לאיזו התפלגות השכיח הגדול ביותר?
א. 1.
ב. 2.
ג. 3.
ד. אין לדעת.

7) במה התפלגות 1 ו-2 זהות?
א. בעשירון העליון.
ב. בממוצע.
ג. בשונות.
ד. אף אחת מהתשובות אינה נכונה.

8) איזה מהמשפטים הבאים נכון לגבי התפלגות מספר 3?
א. הממוצע שווה לחציון בהתפלגות.
ב. הטווח שווה לטווח הבין-רבעוני.
ג. העשירון התחתון שווה לעשירון העליון.
ד. סטיית התקן היא אפס.

9) לאיזו התפלגות סטיית התקן הקטנה ביותר?
א. 1.
ב. 2.
ג. 3.
ד. אין לדעת.

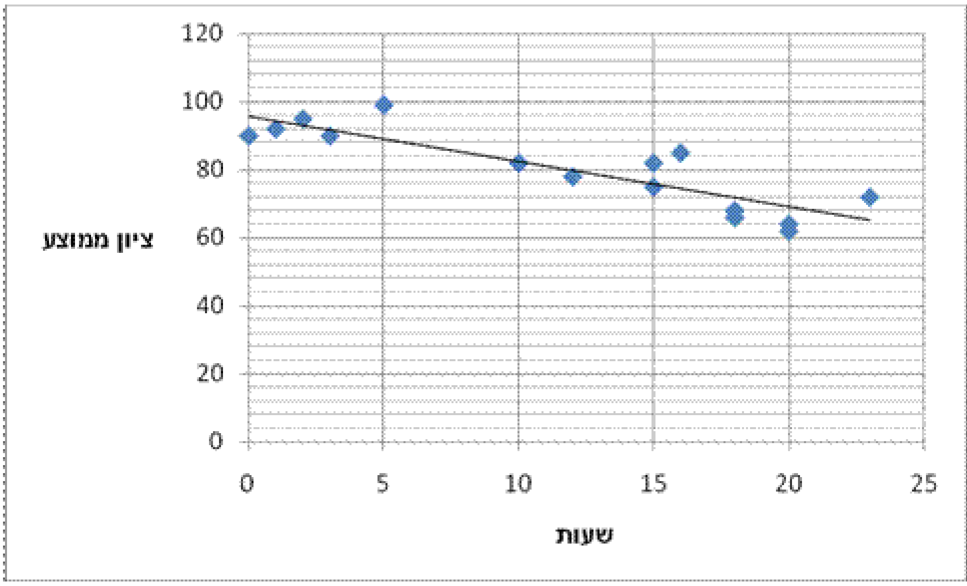
10) בהתפלגות אסימטרית ימנית סטיית התקן יותר גדולה מאשר בהתפלגות אסימטרית שמאלית.
א. הטענה תמיד נכונה.
ב. הטענה תמיד אינה נכונה בהכרח.
ג. אין מספיק נתונים כדי לדעת.

11) ביחס לציר המספרים, רוב הערכים בהתפלגות א-סימטרית ימנית נמצאים:
א. בערכים הגבוהים.
ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
ג. בערכים הנמוכים.
ד. לא ניתן לדעת.

12) הוספת גודל קבוע לכל תצפיות סדרת נתונים.
א. תגדיל את סטיית התקן.
ב. תקטין את סטיית התקן.
ג. לא תשנה את סטיית התקן.
ד. לא ניתן לדעת.

הנתונים הבאים מתייחסים לשאלות 19-21:

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 בעזרת האקסל דיאגרמת פיזור. החוקר אף הוסיף לדיאגרמה את קו המגמה המתאים לנתונים.



13) מיהו המשתנה הבלתי תלוי?
א. ציון ממוצע.
ב. מספר שעות לבילוי.
ג. מספר הסטודנטים.

14) מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר? (הסתמכו על הנתונים ולא על דעתכם האישית)
א. ככל שמבלים יותר הציון נוטה לרדת.
ב. אין קשר בין שעות הבילוי לציון.
ג. ככל שמבלים פחות הציון נוטה לרדת.
ד. ככל שהציון יורד הסטודנט מבלה פחות.

15) איזה מהמתאמים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?
א. 0.85.
ב. 0.15.
ג. ‎-0.85.
ד. ‎-0.15.

הנתונים הבאים מתייחסים לשאלות 16-17:

בכיתה 30 סטודנטים אותם 30 נבחנו במבחן באנגלית ובמבחן בסטטיסטיקה.
להלן פלט לגבי ציונים:

	אנגלית	סטטיסטיקה
ממוצע	90	80
שונות	121	100

16) יערה קיבלה 92 באנגלית ו-82 בסטטיסטיקה.
באיזה מקצוע היא יותר טובה יחסית לכיתתה?
א. אנגלית.
ב. סטטיסטיקה.
ג. אותו דבר יחסית.
ד. אין מספיק נתונים כדי לענות על השאלה.

17) עודד, שקיבל 80 בסטטיסטיקה, העתיק בבחינה. הוחלט לחשב מחדש את השונות של הציונים בסטטיסטיקה בלעדיו. השונות החדשה:
א. תקטן.
ב. תגדל.
ג. לא תשתנה.
ד. אין לדעת.

18) 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.

19) נתונה התפלגות של משתנה כלשהו.
א. הטווח של 20% התצפיות הגבוהות ביותר שווה לטווח של 20% התצפיות הנמוכות ביותר.
ב. הטווח של 50% התצפיות המרכזיות הינו הטווח הבין רבעוני.
ג. הרבעון העליון שווה לרבעון התחתון.
ד. הטווח הבין רבעוני הוא מחצית מהטווח.

הנתונים הבאים מתייחסים לשאלות 20-21:

חוקר רצה לחקור את הקשר הקווי שבין הציון במבחן הרשות בסטטיסטיקה ומימון 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.

20) על פי משוואת הרגרסיה, שעת הכנה נוספת משפרת את ציון המבחן ב:
א. 1.5 נקודות.
ב. 0.53 נקודות.
ג. 0.66 נקודות.
ד. 1.20 נקודות.
ה. 0.96 נקודה.

21) על פי משוואת הרגרסיה, תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון:
א. 29.
ב. 0.
ג. 33.
ד. 24.
ה. 26.

22) אם מקדם המתאם בין שני משתנים הוא שלילי אזי:
א. הערכים של המשתנים הם שליליים.
ב. ככל שמשתנה אחד עולה השני עולה.
ג. ככל שמשתנה אחד יורד השני יורד.
ד. קיימת טרנספורמציה לינארית שלילית בין שני המשתנים.
ה. אף טענה אינה נכונה.

23) בתיק 10 מניות. בהנחה שהמניות לא תלויות זו בזו והסיכוי שביום מסוים מניה תעלה 0.6. מה סטיית התקן של מספר המניות שייעלו ביום מסוים?
א. 6.
ב. 2.4.
ג. 1.55.
ד. 2.46.

24) הסטטיסטיקאית המפורסמת זהבה טוענת כי כאשר מאורעות E ו- F זרים, ניתן לומר כי הסתברות שמאורע E וגם מאורע F יתקיימו, שווה למכפלת ההסתברות כי מאורע E לבדו יתקיים בהסתברות כי מאורע F לבדו יתקיים (או בכתיב מתמטי: $\left(P\left(E\cap F\right)=P\left(E\right)\times P\left(E\right)\right)$. האם זהבה צודקת בטענתה?

א. לא ניתן לדעת.
ב. לא.
ג. כן.
ד. המונח "מאורעות זרים" לא קיים בסטטיסטיקה.
ה. אף תשובה אינה נכונה.

25) ככל שההתפלגות הנורמאלית חדה וצרה יותר במרכזה אזי:
א. השונות שלה יותר גבוהה.
ב. הממוצע שלה יותר גבוה.
ג. היא מייצגת אנשים גבוהים יותר.
ד. השונות שלה נמוכה יותר.
ה. החציון שלה גבוה יותר.

26) נתונה סדרה של N מדידות שלא כולן זהות.
נניח ששתי מדידות נוספות צורפו לסדרה ושתיהן זהות לממוצע הסדרה.
האם וכיצד תשנה הוספת שני הערכים החדשים את שונות הסדרה?
א. שונות הסדרה תקטן.
ב. שונות הסדרה תגדל.
ג. לא ניתן לדעת, זה תלוי במספר התצפיות.
ד. לא ניתן לדעת, זה תלוי בערכו של הממוצע.

27) שני סטודנטים עזבו את החוג לכלכלה. הציון של כל אחד מהם היה שווה לציון הממוצע. כיצד תשפיע עזיבתם על הממוצע ושונות ציוני התלמידים הנותרים? אם הממוצע לפני העזיבה היה 80 והשונות 100.
א. הממוצע לא ישתנה והשונות תגדל.
ב. הממוצע לא ישתנה והשונות תקטן.
ג. הממוצע לא ישתנה והשונות לא תשתנה.
ד. הממוצע יקטן והשונות תגדל.
ה. הממוצע יגדל והשונות תקטן.

28) החציון של סדרת נתונים מסוימת הוא 90.
הוסיפו שתי תצפיות נוספות: 100 ו-20, לכן החציון:
א. יקטן.
ב. יגדל.
ג. לא ישתנה.
ד. לא ניתן לדעת.

29) סטיית התקן של המשכורות בחברה הנה 3000 ₪ אם נוסיף לכל עובדי החברה 200 ₪ לשכר אז:
א. סטיית התקן תגדל אך אין לדעת בכמה.
ב. סטיית התקן תגדל בהכרח ב-200 ₪.
ג. סטיית התקן לא תשתנה.
ד. סטיית התקן תקטן.
ה. לא ניתן לדעת.

30) בתיק השקעות 5 מניות. נגדיר את המאורע: אף מניה לא תעלה מחר מבין מניות התיק. המאורע המשלים למאורע זה הוא (הנח שמניה יכולה או לעלות או לרדת בלבד).
א. לפחות מניה אחת תעלה.
ב. לפחות מניה אחת תרד.
ג. כל המניות יעלו.
ד. בדיוק מניה אחת תעלה.

31) ממוצע של סידרת נתונים הנה 50 וסטיית התקן 10.
אם נוסיף עוד שתי תצפיות שערכן 50 סטיית התקן:
א. תקטן.
ב. תגדל.
ג. לא תשתנה.
ד. אין לדעת.

32) בהתפלגות אסימטרית עם זנב ימני ציון התקן של הרבעון התחתון:
א. בהכרח שלילי.
ב. בהכרח חיובי.
ג. אפס.
ד. לא ניתן לדעת.

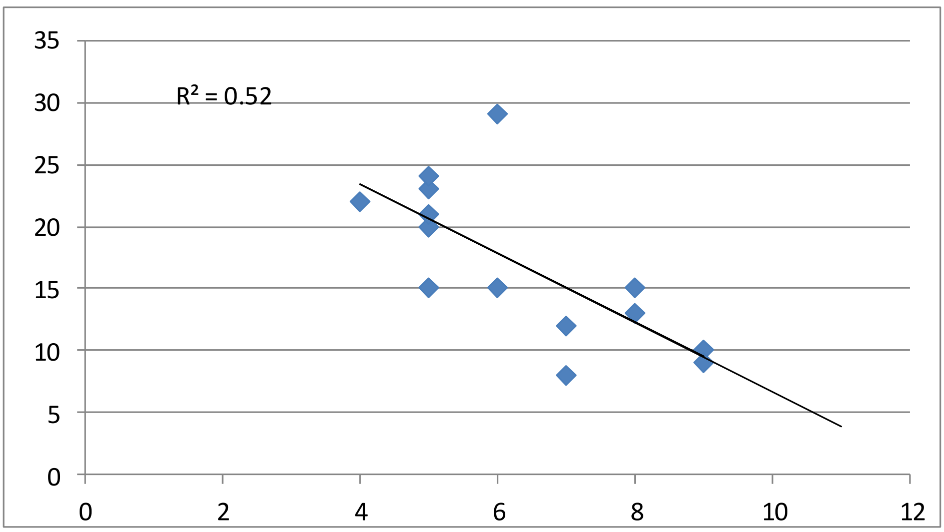
33) אם השונות של המשתנה שווה אפס. מה ניתן לומר על המשתנה?
א. עולה.
ב. יורד.
ג. קבוע.
ד. נורמלי.
ה. לא ניתן לדעת.

34) נמצא שקיים מקדם מתאם חיובי בין הציון בעברית לציון בחשבון בבחינה לכן:
א. הדבר מעיד שהציונים בכתה היו חיוביים.
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.

35) נתונים שני מאורעות המקיימים:
$P(A)=0.45$, $P(B)=0.5$, $P(A\cup B)=0.95$.
איזו טענה נכונה לגבי המאורעות הללו?
א. המאורעות בלתי תלויים.
ב. המאורעות זרים.
ג. המאורע B מכיל את המאורע A .
ד. המאורעות משלימים.

הנתונים הבאים מתייחסים לשאלות 36-38:

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים X (משתנה בלתי תלוי-בציר האופקי) ו- Y (משתנה תלוי), כמו כן הועבר קו הרגרסיה וחושב ריבוע מקדם המתאם.



36) לאור הנתונים המופיעים בדיאגרמה איזה מבין הערכים הבאים מתאים להיות התוצאה של מקדם המתאם שתופעל על הנתונים?

א. 0.52.
ב. -0.52.
ג. -0.72.
ד. 0.72.

37) מה תהיה התוצאה הכי מתאימה לפרמטר b ברגרסיה?

א. 0.52.
ב. 2.79.
ג. -2.79.
ד. -0.52.

38) מהו טווח התפלגות התצפיות של המשתנה הבלתי תלוי X ?

א. 5.
ב. 12.
ג. 6.5.
ד. 7.

הנתונים הבאים מתייחסים לשאלות 39-40:

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

	תפוקה	מספר פועלים
ממוצע	48	15
סטיית תקן	10	2

39) איזו טענה מהטענות הבאות נכונה?

א. המספר המקסימלי של העובדים במפעל הוא 17 עובדים.
ב. התפוקה הכוללת במשך 40 הימים הללו הייתה 192,000 מצברים.
ג. הטווח של התפלגות תפוקת המצברים הוא 20 מאות.
ד. אף אחת מהטענות לא נכונה.

40) 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. חריגים באותה מידה.
ב. כמות הפועלים.
ג. התפוקה.
ד. חסרים נתונים כדי לדעת זאת.

41) התפלגות הציונים במבחן מסוים היא סימטרית, לכן:

א. סטיית התקן של הציונים היא אפס.
ב. הציון החציוני שווה לציון הממוצע.
ג. העשירון העליון שווה לעשירון התחתון של הציונים.
ד. כל הטענות בשאר הסעיפים לא נכונות.

42) איזה מהמשפטים הבאים אינו נכון?

א. אם מוסיפים קבוע לתצפיות הדבר לא משפיע על פיזור הנתונים.
ב. בהתפלגות סימטרית הממוצע שווה לשכיח.
ג. אם כל התצפיות זהות סטיית התקן בהכרח אפס.
ד. הכפלה בקבוע משנה את סטיית התקן.

43) מעוניינים למצוא את הסיכוי לאיחוד שני מאורעות.
מותר לחבר הסתברויות אלה לשם כך, רק אם המאורעות:

א. זרים.
ב. לא זרים.
ג. תלויים.
ד. בלתי תלויים.

44) במכון לשטיפת מכוניות, זמן שטיפת המכונית מתפלג נורמלית עם תוחלת של 25 דקות וסטיית תקן של 5 דקות. מחיר שטיפת מכונית הוא 40 שקלים אם זמן שטיפת המכונית הוא עד 25 דקות. אם זמן שטיפת המכונית עובר את 25 הדקות משלמים 20 שקלים בלבד. עידן הכניס את המכונית לשטיפה. מהי תוחלת התשלום של השטיפה (ב-₪)?

א. 30.
ב. 32.5.
ג. 35.
ד. 25.
ה. לא ניתן לחשב ללא נתונים נוספים.

45) הכפלה בגודל קבוע לכל תצפיות סדרת נתונים:

א. תגדיל את סטיית התקן.
ב. תקטין את סטיית התקן.
ג. לא תשנה את סטיית התקן.
ד. לא ניתן לדעת.

46) בעיר "חולית", בקיץ, כמות הגשם היורד בחודש מתפלג נורמלית עם תוחלת 10 מ"מ וסטיית תקן 2, ובחורף עם תוחלת 10 מ"מ וסטיית התקן 3.
איפה יש יותר סיכוי שירד יותר מ 12 מ"מ גשם?

א. בקיץ.
ב. בחורף.
ג. סיכוי שווה.
ד. לא ניתן לדעת.

47) בהתפלגות שבה המאון ה-40 שווה לממוצע, ציון התקן של הממוצע יהיה:

א. חיובי.
ב. שלילי.
ג. אפס.
ד. לא ניתן לדעת.

תשובות סופיות:

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

ביוסטטיסטיקה

פרק 44 - מבחנים אפרמטריים למדגם יחיד

תוכן העניינים

מבחנים אפרמטריים למדגם יחיד

מבחן הבינום – רקע

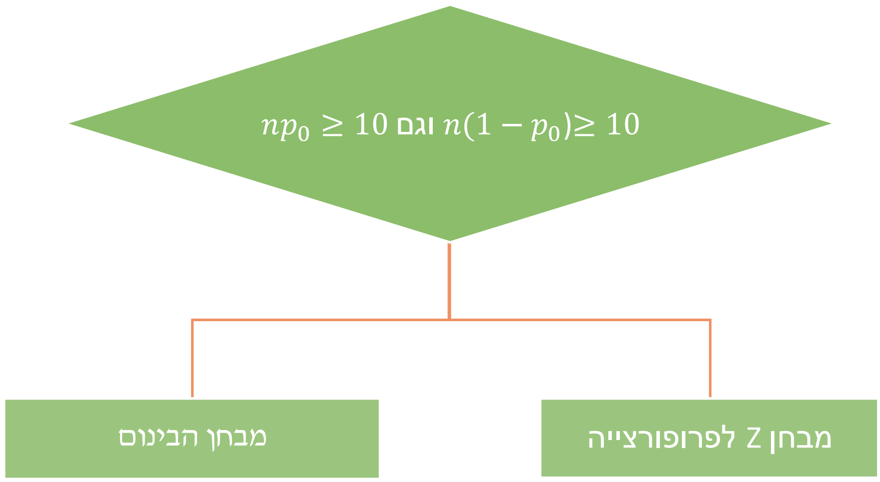
מבחן הבינום הינו מבחן סטטיסטי על הפרמטר p . הפרמטר p מייצג את פרופורציית ההצלחות באוכלוסייה כלומר, הסיכוי בניסוי בודד להצליח.
ההשערות האפשריות על הפרמטר הן :

השערת האפס: **השערה אלטרנטיבית:**	$H_0 : p = p_0$ $H_1 : p \neq p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p > p_0$

דוגמה:

במדינה אירופאית התנהל לפני 4 שנים משאל עם בדבר לגליזציה של הקנאביס.
במשאל העם 40% מהאזרחים היו בעד לגליזציה של הקנאביס ובשל כך חוק הלגליזציה לא עבר באותה המדינה. במדגם שנעשה כיום בו השתתפו 15 אזרחים מהמדינה 10 ענו שהם תומכים בלגליזציה של הקנאביס במדינה.
פרלמנטר מהמדינה חוקר האם כיום אחוז התומכים בלגליזציה של הקנאביס במדינה עלה יחסית לאחוז שהתקבל במשאל העם במדינה לפני 4 שנים.
רשמו את השערות המחקר.

במבחן הבינום מבצעים מדגם אקראי בגודל n ומתבוננים במספר ההצלחות שהתקבלו במדגם, אותן נסמן ב- Y. בהנחה והתצפיות במדגם בלתי תלויות זו בזו אנו אומרים ש: $Y \sim B(n, p)$.
מבחן הבינום נכנס לקטגוריה של מבחנים אפרמטרים והוא בא כחלופה למבחן הפרמטרי על פרופורציה אחת כאשר התנאים לקירוב הנורמלי אינם מתקיימים.



דוגמה:

מהו המבחן הסטטיסטי המתאים? נמקו.

הטכניקה הנוחה ביותר למבחן הבינום היא לחשב את מובהקות התוצאה ולדחות את השערת האפס אם $\alpha \geq PV$. מובהקות התוצאה היא הסיכוי לתוצאות של המדגם וקיצוני יותר בהנחת השערת האפס.
כזכור, פונקציית ההסתברות של ההתפלגות הבינומית היא:

$$k = 0,1,\ldots,n \qquad P(Y=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

כאשר התוחלת של ההתפלגות הבינומית היא: $E(Y) = np$.
בעמוד הבא מצורפת טבלה של התפלגות בינומית מצטברת.
זו טבלה שיכולה לעזור בתהליך החישוב.

דוגמה:

חשבו את מובהקות התוצאה. מה תהיה מסקנת המחקר ברמת מובהקות של 5%?

טבלת התפלגות בינומית מצטברת

		p									
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0474	0.0064	0.0002
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

שאלות

1) אחוז האנשים שאוכלים גלידות בחודשי החורף הינו 30%, קיים חשש בקרב מיצרי הגלידות כי השנה פחות אחוז אוכלי הגלידות בחודשי החורף. לשם כך נדגמו 12 אנשים אשר מתוכם 2 טענו שהם אוכלים גלידה בחודשי החורף.

א. רשמו את השערות המוצגות בשאלה זו וציינו מהו המבחן הסטטיסטי המתאים.

ב. מה תהיה המסקנה ברמת מובהקות של 10%?

ג. מה תהיה המסקנה ברמת מובהקות גדולה יותר מ- 10%?

2) מטבע הוטל 15 פעמים במטרה לבדוק האם המטבע סימטרי. בסך הכול התקבלו בהטלות 3 פעמים עץ.

א. בדקו ברמת מובהקות של 5% האם המטבע הוא סימטרי.

ב. עבור אילו רמות מובהקות נוכל להסיק שהמטבע אינו סימטרי?

3) אחוז המובטלים במשק לפני 5 שנים היה 10%. מעוניינים לבדוק האם כיום אחוז המובטלים קטן לעומת זה שהיה לפני 5 שנים. במדגם של 20 אנשים התקבל מובטל אחד ויחיד.

א. מהי רמת המובהקות הקטנה ביותר עבורה יוסק שכיום אחוז האבטלה נמוך מאשר לפני 5 שנים?

ב. מה תהיה המסקנה אם רמת המובהקות תהיה 5%?

4) נניח ש 40% מהאוכלוסייה מכירה את המוצרים של חברת "רמקס". החברה מתכננת לצאת בקמפיין פרסום שמטרתו להעלות את המודעות לקיומם של המוצרים של החברה באוכלוסייה. בזמן הקמפיין נדגמו 20 אנשים אקראיים מתוכם 70% טענו שהם מכירים את המוצרים של חברת "רמקס".

א. מהי מובהקות התוצאה?

ב. כיצד מובהקות התוצאה הייתה משתנה אם במדגם 80% היו טוענים שהם מכירים את המוצרים של חברת "רמקס"?

5) חברה לתוספי מזון דיווחה שנטילת מולטי ויטמין מקטינה את הסיכוי לחלות במחלות חורף במהלך החורף. לפי משרד הבריאות 70% מהאוכלוסייה חולים במחלות חורף במהלך החורף. במחקר השתתפו 20 אנשים אשר נטלו במהלך שנה מולטי ויטמין. במהלך החורף נמצא ש 12 מתוכם חלו במחלות חורף במהלך החורף. את המחקר יש לבצע ברמת מובהקות של 5%.
נסמן ב: PV את מובהקות התוצאה של המחקר.

א. האם המשפט הבא נכון?
"אם כלל האוכלוסייה הייתה נוטלת מולטי ויטמין אז בסיכוי של 5% ניתן להגיד ש-70% מהאוכלוסייה הייתה חולה במחלת חורף".

ב. האם המשפט הבא נכון?
"אם כלל האוכלוסייה הייתה נוטלת מולטי ויטמין אז בסיכוי של PV 70% מהאוכלוסייה הייתה חולה במחלת חורף".

ג. חשבו את PV.

תשובות סופיות

1) א. מבחן הבינום.
$H_0: p = 0.3$
$H_1: p < 0.3$
ב. לא נדחה את H_0.
ג. לא ניתן לדעת.

2) א. נדחה את H_0.
ב. לפחות 0.0352

3) א. 0.3917
ב. לא נדחה את H_0.

4) א. 0.0065
ב. תקטן.

5) א. לא נכון.
ב. לא נכון.
ג. 0.2277.

חישוב הסתברויות לטעויות ועוצמת המבחן במבחן הבינום – רקע

מבחן הבינום הינו מבחן סטטיסטי על הפרמטר p.

הפרמטר p מייצג את פרופורציית ההצלחות באוכלוסייה כלומר, הסיכוי בניסוי בודד להצליח. ההשערות האפשריות על הפרמטר הן:

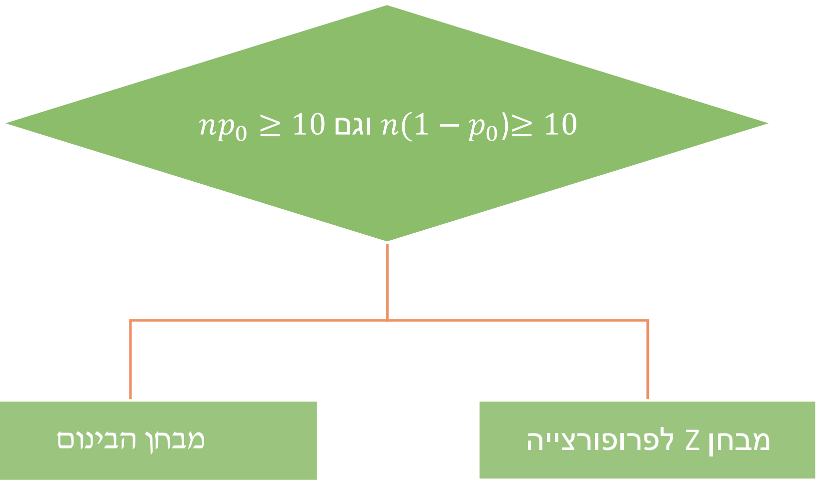
השערת האפס:	$H_0: p = p_0$	$H_0: p = p_0$	$H_0: p = p_0$
השערה אלטרנטיבית:	$H_1: p \neq p_0$	$H_1: p < p_0$	$H_1: p > p_0$

דוגמה:

בחברת פז 40% מעובדי החברה מאחרים לעבודה. מנכ״ל החברה מעוניין לצמצם את אחוז העובדים המאחרים בחברה. לצורך כך, בוחר המנכ״ל מחלקה אקראית במטרה לבדוק עליה שיטת תמריצים שאמורה לצמצם את אחוז המאחרים.
במחלקה שנבחרה 10 עובדים והופעלה עליה שיטת התמריצים הנבדקת.
הוסכם שאם מספר המאחרים במחלקה יהיה לכל היותר עובד אחד לאחר הפעלת שיטת התמריצים, תופעל שיטת התמריצים על כלל החברה.

- מהן השערות המחקר?

במבחן הבינום מבצעים מדגם אקראי בגודל n ומתבוננים במספר ההצלחות שהתקבלו במדגם, אותן נסמן ב- Y. בהנחה והתצפיות במדגם בלתי תלויות זו בזו אנו אומרים ש: $Y \sim B(n, p)$.
מבחן הבינום נכנס לקטגוריה של מבחנים אפרמטריים והוא בא כחלופה למבחן הפרמטרי על פרופורציה אחת כאשר התנאים לקירוב הנורמלי אינם מתקיימים.



דוגמה:

- מהו המבחן הסטטיסטי המתאים? נמקו.

כזכור, פונקציית ההסתברות של ההתפלגות הבינומית היא:

$k=0,1,\ldots,n \qquad P(Y=k)=\binom{n}{k}p^k(1-p)^{n-k}$.

בעמוד הבא מצורפת טבלה של התפלגות בינומית מצטברת. זו טבלה שיכולה לעזור בתהליך החישוב.
נזכיר את המצבים האפשריים בבדיקת השערות:

		הכרעה	
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

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

הסיכוי לבצע טעות מסוג 2:

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

רמת בטחון:

1-α=P(לקבל את H_0 | נכונה H_0)= P_{H_0}(לקבל H_0)

עוצמה:

π=1-β =P(לדחות את H_0 | נכונה H_1) = P_{H_1}(לדחות H_0)

דוגמה:

- מהי רמת המובהקות של המחקר המוצע?
- אם בשיטת התמריצים, הסיכוי לאחר הוא 10%, מה ההסתברות שהמחקר יגלה זאת?

טבלת הסתברות בינומית (מצטברת) עבור p -ים שונים

		p									
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0474	0.0064	0.0002
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

שאלות

1) במטרה לבדוק האם במוסד לימוד מסוים פרופורציית הבנים נמוכה מפרופורציית הבנות נדגמו באקראי 10 תלמידים. הוחלט שאם מספר הבנים במדגם יהיה לכל היותר 2 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות.

א. מהי רמת המובהקות במבחן המוצע?

ב. מהי העוצמה במבחן המוצע בהנחה ובמוסד 30% בנים?

ג. כיצד התשובות לסעיפים הקודמים ישתנו אם ישנו את המבחן כך שיוחלט שאם מספר הבנים במדגם יהיה לכל היותר 3 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות?
הסבירו ללא חישוב מחדש.

2) הסיכוי לזכות במשחק מסוים הינו 70%. מעוניינים לבדוק האם שיטת משחק מסוימת מעלה את סיכויי הזכייה לצורך כך מחליטים לשחק את המשחק 20 פעמים תוך שימוש בשיטה הנבדקת. מחליטים שאם מספר הזכיות במדגם יהיה לפחות 18 נקבל את הטענה שאכן השיטה עובדת והיא מעלה את סיכויי ההצלחה להיות 90%.
מה הסיכוי לבצע טעות מסוג ראשון ומה הסיכוי לבצע טעות מסוג שני במבחן שהוצע?

3) במטרה לבדוק האם מטבע הוא סימטרי הטילו אותו 8 פעמים. הוחלט שאם מספר העצים יהיה בין 1 ל 7 כולל, יוחלט שהמטבע הוגן, אחרת נחליט שהמטבע לא הוגן.

א. רשמו את השערות המחקר.

ב. מה ההסתברות להחליט שהמטבע אינו סימטרי למרות שבפועל המטבע הינו סימטרי?

ג. מהי ההסתברות שאם במטבע הסיכוי לעץ הוא 20%, אכן נגלה זאת?

ד. כיצד התשובה לסעיף הקודם תשתנה אם הסיכוי לעץ במטבע אף נמוך מ-20%?

4) נתון ש- $X\sim B(15,p)$ ההשערות של המחקר הן: $H_1: p>0.3 \quad H_0: p=0.3$.
כלל ההכרעה הוא: נדחה את השערת האפס אם $X\geq 8$.

א. מהי רמת המובהקות במחקר זה?

ב. מה יקרה לרמת המובהקות אם כלל ההכרעה יהיה: נדחה את השערת האפס אם $X\geq 7$? הסבירו ללא חישוב.

ג. מה יקרה לרמת המובהקות אם השערות המחקר יהיו $H_1: p\neq 0.3 \quad H_0: p=0.3$ אבל כלל ההכרעה לא ישתנה?
הסבירו ללא חישוב.

תשובות סופיות

1) א. 0.0547 ב. 0.3828 ג. תגדלנה.

2) $\beta = 0.3231$ $\alpha = 0.0355$.

3) א. $H_0: p = 0.5$
$H_1: p \neq 0.5$
ב. 0.0078 ג. 0.1678
ד. העוצמה תגדל.

4) א. 0.05 ב. תגדל. ג. לא תשתנה.

ביוסטטיסטיקה

פרק 45 - מבחנים אפרמטרים למדגמים מזווגים

תוכן העניינים

מבחן ווילקוקסון למדגמים מזווגים (על ידי שימוש בטבלה של ערכים קריטיים) – 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 על מדגם 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 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 Di 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 $(W-)$.
5) W יהיה $W+$ או $W-$, זה שאמור להיות יותר קטן לפי השערת המחקר או הקטן מביניהם אם ההשערה היא דו צדדית.

דוגמה (פתרון בהקלטה):

חשבון 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

כלל הכרעה:

במבחן ווילקוקסון זה כלל ההכרעה הוא: נדחה את H_0 אם $W \leq W_c$.

כאשר, W_c - הערך הקריטי; W - הסטטיסטי.

את הערכים הקריטיים נחלץ מתוך טבלה מתאימה:

n_1	חד-צדדי $\alpha=0.01$ דו-צדדי $\alpha=0.02$	חד-צדדי $\alpha=0.025$ דו-צדדי $\alpha=0.05$	חד-צדדי $\alpha=0.05$ דו-צדדי $\alpha=0.10$
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

דוגמה (פתרון בהקלטה):

א. רשמו את כלל ההכרעה המתאים ברמת מובהקות של 5%.

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 :

1-קושי קורס	4	3	2	4	3	2	1	5	4
2-קושי בחינה	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?

4) 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? בדקות ברמת מובהקות של 1%.

תשובות סופיות

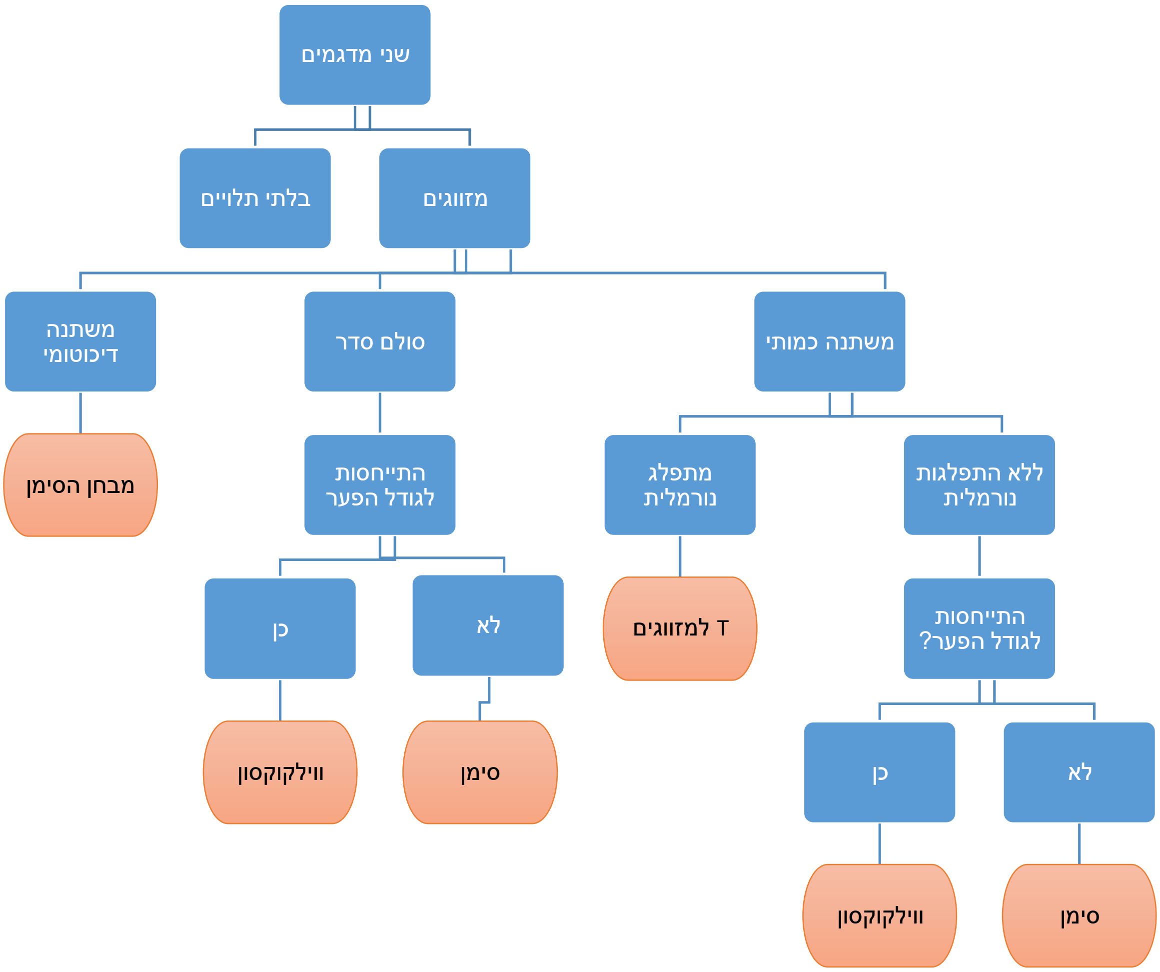
1) לא נדחה H_0.

2) לא נדחה H_0.

3) א. לא נדחה H_0. ב. לא משתנה. ג. לא משתנה.

4) לא נדחה H_0.

זיהוי מבחנים סטטיסטיים – רקע



שאלות

1) במטרה להשוות את רמת האפייה של שני קונדיטורים בחרו 9 מאפים שונים (קרואסון, בראוני וכדומה) ונתנו לכל אחד משני הקונדיטורים לאפות את 9 המאפים השונים. 18 המאפים שנאפו ניתנו למומחה שנתן ציון למאפים השונים. הציון שניתן הוא בין 1 ל-5 לפי ניסיונו וטעמו האישי של המומחה.
מהו המבחן המתאים ביותר במקרה זה?

א. מבחן ווילקוקוסון.
ב. מבחן הסימן.
ג. מבחן T למזווגים.
ד. מבחן למדגמים בלתי תלויים.

2) שני מוסיקאים מפורסמים נתנו ציון בסולם של 1-10 לקולם של 8 מתמודדים בתוכנית ראליטי ידועה. ציון 10 ניתן לקול שמצא חן ביותר בעיני המוסיקאי. מפיק התוכנית רצה לבדוק האם יש הבדל בין המוסיקאים מבחינת הטעם. בטבלה הבאה נתונים הציונים של כל אחד מהמוסיקאים את שמונת המתמודדים:

	1	2	3	4	5	6	7	8
מוסיקאי א'	6	5	7	4	3	1	1	4
מוסיקאי ב'	5	7	5	2	3	3	2	7

מהו המבחן המתאים ביותר במקרה זה?

א. מבחן ווילקוקוסון.
ב. מבחן הסימן.
ג. מבחן T למזווגים.
ד. מבחן למדגמים בלתי תלויים.

3) במחקר בדקו לאנשים את רמת הסוכר בבוקר ואת רמת הסוכר בערב.
מתוך 26 אנשים ל-3 רמת הסוכר הייתה זהה.
ל-14 רמת הסוכר הייתה גבוהה יותר בשעות הערב.
וליתר רמת הסוכר הייתה גבוהה יותר בשעות הבוקר.
רוצים לבדוק ברמת מובהקות של 6% האם קיים הבדל בין רמת הסוכר בבוקר לרמת הסוכר בערב אצל האנשים. מהו המבחן המתאים ביותר במקרה זה?

א. מבחן ווילקוקוסון.
ב. מבחן הסימן.
ג. מבחן T למזווגים.
ד. מבחן למדגמים בלתי תלויים.

4) חוקר מעוניין לבדוק את התפתחות היכולת לדחות סיפוקים מיידיים בקרב ילדים. לשם כך, הוא משתמש במבחן לבדיקה של דחיית סיפוקים, ומעביר אותו בו זמנית ל-2 קבוצות גיל. מבחן זה מודד כמה זמן (בשניות) מסוגל הילד לדחות קבלה של תגמול מיידי קטן על מנת לקבל תגמול גדול יותר בעתיד. התוצאות שמתקבלות הן הזמנים של הנחקרים בכל קבוצת גיל.
מהו המבחן המתאים ביותר במקרה זה?
א. מבחן ווילקוקוסון.
ב. מבחן הסימן.
ג. מבחן T למזווגים.
ד. מבחן למדגמים בלתי תלויים.

5) חברת משקאות יצאה בקמפיין שנוי במחלוקת. החברה מעוניינת לבדוק האם הקמפיין השפיע על הרגלי הצריכה. במחקר השתתפו נשאלים האם הם נהגו לרכוש את המשקה לפני הקמפיין והאם הם רכשו אותו לאחר הקמפיין.
מהו המבחן המתאים ביותר במקרה זה?
א. מבחן ווילקוקוסון.
ב. מבחן הסימן.
ג. מבחן T למזווגים.
ד. מבחן למדגמים בלתי תלויים.

6) מחקר התעניין בדפוסי שיחות הטלפון, שמנהל הפרט בעקבות פרידה מבן זוגו. במחקר השתתפו גברים ו-נשים (כולם נפרדו מבן זוגם). המשתתפים דיווחו על משך השיחות (בדקות; לפני ואחר הפרידה). שאלת המחקר בחנה האם פרידה מבן הזוג קשורה למשך השיחות (משך השיחות הוא משתנה שנהוג להתייחס אליו כמתפלג נורמלית). מהו המבחן המתאים ביותר במקרה זה?
א. מבחן ווילקוקוסון.
ב. מבחן הסימן.
ג. מבחן T למזווגים.
ד. מבחן למדגמים בלתי תלויים.

תשובות סופיות

1) א'.
2) א'.
3) ב'.
4) ד'.
5) ב'.
6) ג'.

ביוסטטיסטיקה

פרק 46 - מבחנים אפרמטריים למדגמים בלתי תלויים

תוכן העניינים

מבחנים אפרמטרים למדגמים בלתי תלויים

ניתוח פלטים במבחן מן-וויטני – רקע

מבחן מן-ויטני מיועד לבדוק האם לשתי אוכלוסיות התפלגות שווה.
המבחן בוחן באופן רוחבי את כל תחום הערכים ולא מתמקד בערך מרכזי אחד.
נשתמש במבחן זה כאשר יש שני מדגמים בלתי תלויים והמשתנה הכמותי הנחקר אינו מתפלג נורמלית או שמדובר במשתנה מסולם סדר.
המבחן מתבסס על דירוג כל התצפיות. בעצם, מבחן זה הוא המענה האפרמטרי למבחן הפרמטרי להפרש תוחלות במדגמים בלתי תלויים.

דוגמה (פתרון בהקלטה) :

להלן תוצאות הערכות שקיבלו שני מורים : ד"ר A ופרופסור B. סטודנטים נתנו משוב כללי על המורים בסקלה של 1 (גרוע) עד 5 (מצוין). הטענה היא שד"ר A הוא מרצה טוב יותר מאשר פרופסור B.

Mann-Whitney Test

Ranks

	teacher	N	Mean Rank	Sum of Ranks
grade	dr A	17	25.00	425.00
	prof B	20	13.90	278.00
	Total	37		

Test Statistics[a]

	grade
Mann-Whitney U	68.000
Wilcoxon W	278.000
Z	-3.249
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.001[b]

a. Grouping Variable: teacher
b. Not corrected for ties.

- הסבירו מדוע נעשה כאן מבחן מן-וויטני.
- מה המסקנה ברמת מובהקות של 1%?

שאלות

1) מחקרים טוענים שקיים הבדל בין שעות השינה של גברים לשעות השינה של נשים. כיוון שלא ניתן להוכיח ששעות שינה הינו משתנה המתפלג נורמלית ביצעו מבחן מן-וויטני בו לקחו נשים וגברים אקראיים ובדקו את שעות השינה שלהם.

Ranks

	gender	N	Mean Rank	Sum of Ranks
sleeptime	male	10	???	135.00
	female	???	7.50	75.00
	Total	20		

Mann-Whitney Test

Test Statistics[a]

	sleeptime
Mann-Whitney U	20.000
Wilcoxon W	75.000
Z	-2.319
Asymp. Sig. (2-tailed)	.020

א. השלימו את סימני השאלה החסרים בפלט.

ב. מהי המסקנה ברמת מובהקות של 5%?

ג. מה הייתה מובהקות התוצאה אם טענת המחקר הייתה שגברים ישנים יותר מנשים?

2) שני אנשים נתבקשו לבדוק את מספר תאונות הדרכים בשבוע בשני קטעי כביש שונים. כל אחד בחר את השבועות באופן אקראי ובלתי תלוי באחר וספר כמה תאונות היו בכל כביש בשבוע. הפלטים שהתקבלו:

מספר תאונות

	Frequency
.00	2
1.00	1
2.00	2
3.00	1
Total	6

a. road = 1.00

מספר תאונות

	Frequency
.00	2
1.00	1
2.00	1
3.00	1
4.00	1
5.00	1
Total	7

a. road = 2.00

Ranks

	road	N	Mean Rank	Sum of Ranks
מספר תאונות	1.00	A	B	C
	2.00	D	E	F
	Total	G		

Test Statistics[a]

	VAR00002
Asymp. Sig. (2-tailed)	.465

א. השלימו בטבלה השלישית את המספרים החסרים במקום האותיות.

ב. מהי מובהקות התוצאה לבדיקת ההשערה שאין הבדל בין הכבישים מבחינת התפלגות תאונות הדרכים?

ג. כיצד הייתה משתנה התשובה של הסעיף הקודם אם כל חוקר היה מוסיף נתונים על שבוע נוסף לכל כביש?

ד. מה המסקנה ברמת מובהקות של 5%?

ה. מהי מובהקות התוצאה לבדיקת ההשערה שכביש מספר 1 עם התפלגות תאונות גבוהה יותר מאשר כביש מספר 2?

תשובות סופיות

1) א. 10 ,13.5 ב. נדחה H_o. ג. 0.01.

2) א. 6, 6.17, 37, 7, 7.71, 54, 13 ב. 0.465.
ג. לא ניתן לדעת. ד. לא נדחה H_o. ה. 0.7675.

מבחן קרוסקל וואליס – ניתוח פלטים – רקע

כאשר לא ניתן לבצע מבחן ניתוח שונות חד כיווני (למשל, כאשר השונויות לא שוות, המשתנה התלוי אינו מתפלג נורמאלי או המשתנה התלוי נמדד בסולם סדר) נבחר במבחן לא-פרמטרי להשוואת k קבוצות בלתי תלויות.

מבחן Kruskal-Wallis הוא מבחן אפרמטרי להשוואת מספר אוכלוסיות בלתי תלויות. מבחן מן-וויטני הינו מקרה פרטי של מבחן Kruskal-Wallis כאשר יש שתי קבוצות בלתי תלויות.

השערות המבחן:

- השערת האפס: אין הבדל בין הקבוצות הנבדקות.
- השערת המחקר: קיים הבדל בין 2 קבוצות לפחות.

מבחן זה מתבסס על דירוג כלל התצפיות במחקר מכל המדגמים.

דוגמה (פתרון בהקלטה):

השוו את שביעות הרצון מהשירות של 4 חברות טלפון סלולרי שונות. שביעות הרצון הוא ציון שהלקוח היה צריך לתת לרמת השירות של החברה מ-1 ועד 10.

א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר זה?
ב. מהן השערות המחקר?
ג. מהו המבחן הסטטיסטי במקרה זה?
ד. מה מסקנות המחקר ברמת מובהקות של 5%?

Count

		company				Total
		1	2	3	4	
שביעות רצון	4.00	0	2	0	0	2
	5.00	0	1	0	0	1
	6.00	2	1	0	1	4
	7.00	2	2	1	0	5
	8.00	1	2	2	2	7
	9.00	1	1	3	2	7
	10.00	1	1	0	0	2
Total		7	10	6	5	28

Kruskal-Wallis Test

Ranks

	company	N	Mean Rank
שביעות רצון	1	7	13.93
	2	10	11.40
	3	6	18.50
	4	5	16.70
	Total	28	

Test Statistics[a,b]

	שביעות רצון
Chi-Square	3.363
df	3
Asymp. Sig.	.339

a. Kruskal Wallis Test

b. Grouping Variable: company

שאלות

1) מחקר פסיכולוגי רצה להשוות 3 טיפול שונים לטיפול בחרדה. נלקחו נבדקים הסובלים מחרדה. הם חולקו באקראי ל-3 קבוצות וכל קבוצה קיבלה שיטת טיפול אחרת לחרדה. להלן תוצאות המבחן הסטטיסטי שבוצע:

Kruskal-Wallis Test

Ranks

	group	N	Mean Rank
anxiety	1.00	8	15.88
	2.00	4	6.88
	3.00	8	6.94

Test Statistics[a,b]

	anxiety
Chi-Square	11.149
df	2
Asymp. Sig.	.004

א. כמה נבדקים השתתפו במחקר?
ב. באיזו שיטת טיפול רמת החרדה הייתה הנמוכה ביותר במדגם?
ג. האם, עבור ר״מ 0.05, יש הבדל ברמת החרדה המתקבלת בשיטות הטיפול השונות?

2) חברת תרופות לקחה 15 אנשים ברמת בריאות דומה. החברה חילקה את האנשים לשלוש קבוצות שוות בגודלן. לכל קבוצה ניתנה אותה תרופה במינון שונה (dosage). המינונים שניתנו הם: 10 מ"ג, 20 מ"ג ו-30 מ"ג. לאחר שעה מזמן לקיחת התרופה ניבדק קצב פעימות הלב של כל אדם (pulse).
הנתונים שלהלן הוזנו לתוכנה סטטיסטית, והתקבלו התוצאות הבאות:

מינון	10	10	10	10	10	20	20	20	20	20	30	30	30	30	30
דופק	87	87	83	81	79	79	76	82	82	82	68	67	75	73	72

Kruskal-Wallis Test

Ranks

	dosage	N	Mean Rank
pulse	10.00	A	E
	20.00	B	F
	30.00	C	G
	Total	D	

Test Statistics[a,b]

	pulse
Chi-Square	10.204
df	2
Asymp. Sig.	.006

א. השלימו את המספרים החסרים במקום האותיות A עד G בטבלת הדירוגים.
ב. בדוק ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת הדופק של האנשים?
ג. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.

תשובות סופיות

1) א. 20 ב. מדגם 2 ג. נדחה H_0.

2) א. 3 ,9.3 ,11.7 ,5,5,5,15 ב. נדחה H_0. ג. לא תשתנה.