יסודות הסטטיסטיקה ב

תוכן העניינים

1. התפלגות נורמלית ה ..... (ללא ספר)
2. הסקה סטטיסטית - הקדמה 1
3. התפלגות הדגימה ומשפט הגבול המרכזי מנהל ..... (ללא ספר)
4. מושגי יסוד באמידה 4
5. רווח סמך לתוחלת (ממוצע) 9
6. מבוא לבדיקת השערות על פרמטרים 19
7. בדיקת השערות על תוחלת (ממוצע) 25
8. רווח סמך לפרופורציה 42
9. בדיקת השערות על פרופורציה 47
10. רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים 50
11. בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים 52
12. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 56
13. בדיקת השערות לתוחלת ההפרש במדגמים מזווגים 58
14. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות 62
15. ניתוח שונות חד כיוונית 65
16. מבחני חי בריבוע 74
17. גודל האפקט ..... (ללא ספר)
18. מבחנים אפרמטרים למדגמים מזווגים 76
19. שאלות מסכמות בבדיקת השערות 79
20. שאלות מסכמות על רווחי סמך 92
21. מדד הקשר של ספירמן ..... (ללא ספר)
22. מקדם המתאם (מדד קשר) הלינארי ומובהקותו 94
23. מדדי קשר - בחירת מדד מתאים 116

תוכן העניינים

יסודות הסטטיסטיקה ב

פרק 1 - התפלגות נורמלית ה

תוכן העניינים

1. כללי ..... (ללא ספר)

יסודות הסטטיסטיקה ב

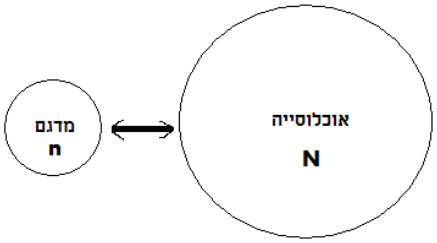
פרק 2 - הסקה סטטיסטית - הקדמה

תוכן העניינים

1. כללי 1

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.

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 3 - התפלגות הדגימה ומשפט הגבול המרכזי מנהל

תוכן העניינים

1. כללי ..... (ללא ספר)

יסודות הסטטיסטיקה ב

פרק 4 - מושגי יסוד באמידה

תוכן העניינים

1. כללי 4

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: $\sigma\left(\hat{\theta}\right)=S.E$

© כל הזכויות שמורות לברק קנדל

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 אומר 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 אומר 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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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'.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 5 - רווח סמך לתוחלת (ממוצע)

תוכן העניינים

1. רווח סמך כששונות האוכלוסיה ידועה. 9
2. קביעת גודל מדגם. 14
3. רווח סמך כששונות האוכלוסיה לא ידועה. 16

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: $P\left(A\le \mu \le B\right)=1-\alpha$.
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?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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'.

© כל הזכויות שמורות לברק קנדל

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).

© כל הזכויות שמורות לברק קנדל

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.

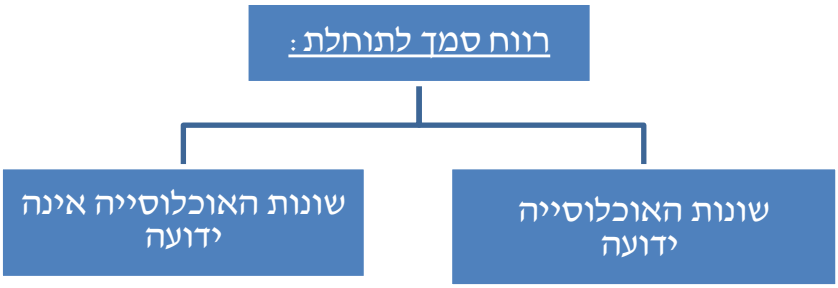
3) $n=62$.

© כל הזכויות שמורות לברק קנדל

רווח סמך כששונות האוכלוסייה לא ידועה:

רקע:

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

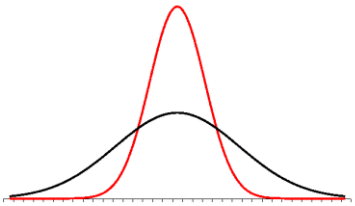


בפרק זה נעסוק במקרה ששונות האוכלוסייה $\left(\sigma^2\right)$ אינה ידועה לנו.

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

התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים.
התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן: $df = n-1$.
ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה.
כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

דוגמה (פתרון בהקלטה):

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי כיתות ח' נורמאלית.
במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח'. להלן התוצאות שהתקבלו בדקות: 5.3, 4.6, 5.2, 4.7.
בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

© כל הזכויות שמורות לברק קנדל

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 האמיתית של כלל הנבחנים.

ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.

5) נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X^2_i = 643.19$.
בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

© כל הזכויות שמורות לברק קנדל

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.

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'.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 6 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

1. הקדמה 19
2. סוגי טעויות 23

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.

© כל הזכויות שמורות לברק קנדל

דוגמה (פתרון בהקלטה):

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

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

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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. $\begin{matrix} H_0:p=0.08 \\ H_1:p\neq 0.08 \end{matrix}$.

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

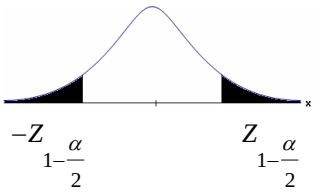
פרק 7 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

1. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה ידועה..... 25
2. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה לא ידועה..... 29
3. סיכוי לטעויות ועוצמה (ששונות האוכלוסיה ידועה)..... 33
4. קביעת גודל מדגם (ששונות האוכלוסיה ידועה)..... 39

בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה ידועה:

רקע:

השערת האפס: השערה אלטרנטיבה:	$H_0:\mu=\mu_0$ $H_1:\mu\neq\mu_0$	$H_0:\mu\geq\mu_0$ $H_1:\mu<\mu_0$	$H_0:\mu\leq\mu_0$ $H_1:\mu>\mu_0$
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 $\bar{X}<\mu_0-Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$	$\bar{X}<\mu_0-Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}$	$\bar{X}>\mu_0+Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}$

© כל הזכויות שמורות לברק קנדל

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.

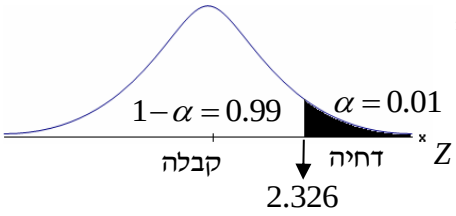
<u>הפרמטר</u>: μ = תוחלת היבול בשיטת הזיבול החדשה.

<u>השערות</u>: $H_0 : \mu = 10$
$H_1 : \mu > 10$

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

<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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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

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

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

3) נדחה H_0, ברי"מ של 5% נקבע שהמכונה לא מאופסת.

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

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

6) א'.

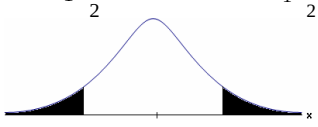
7) ג'.

8) א. לדחות. ב. לא ניתן לדעת.

© כל הזכויות שמורות לברק קנדל

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

רקע:

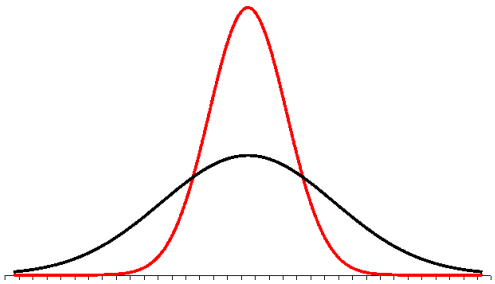
השערת האפס: השערה אלטרנטיבה:	$H_0:\mu=\mu_0$ $H_1:\mu\neq\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu<\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu>\mu_0$
תנאים:	1. σ אינה ידועה 2. $X\sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0:	$t_{\bar{x}}<-t^{(n-1)}_{1-\frac{\alpha}{2}}$ או $t_{\bar{x}}>t^{(n-1)}_{1-\frac{\alpha}{2}}$  $-t_{1-\frac{\alpha}{2},n-1}$ $t_{1-\frac{\alpha}{2},n-1}$ ■ - 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: $t_{\bar{x}}=\dfrac{\bar{x}-\mu_0}{\frac{S}{\sqrt{n}}}$.

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

© כל הזכויות שמורות לברק קנדל

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%.

© כל הזכויות שמורות לברק קנדל

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 .

© כל הזכויות שמורות לברק קנדל

5) ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה.
כמו כן הם התבססו על אותן תוצאות של מדגם.
ליאור השתמש בטבלה של התפלגות Z.
רוני השתמשה בטבלה של התפלגות t.
מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.

א. אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.

ב. אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.

ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.

ד. לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

6) נתון ש: $X \sim N(\mu,\sigma^2)$ כמו כן נתונות ההשערות הבאות: $\begin{matrix} H_0:\mu=\mu_0 \\ H_1:\mu<\mu_0 \end{matrix}$.
חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות.
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'.

© כל הזכויות שמורות לברק קנדל

סיכוי לטעויות ועוצמה (ששונות האוכלוסייה ידועה):

רקע:

	הכרעה		
		H_0	H_1
מציאות	H_0	אין טעות	טעות מסוג 1
	H_1	טעות מסוג 2	אין טעות

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

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

הסיכוי לבצע טעות מסוג 2:

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

רמת בטחון:

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

עוצמה:

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

© כל הזכויות שמורות לברק קנדל

התהליך לחישוב סיכוי לטעות מסוג שני:

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		
כלל ההכרעה: אזור הדחייה של H_0:	$\bar{X}>\mu_0+Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$ או $\bar{X}<\mu_0-Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$	$\bar{X}<\mu_0-Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}$	$\bar{X}>\mu_0+Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}$
חישוב β:	$P_{\mu_1}\left(\mu_0-Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}<\bar{X}<\mu_0+Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)$	$P_{\mu_1}\left(\bar{X}>\mu_0-Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right)$	$P_{\mu_1}\left(\bar{X}<\mu_0+Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right)$

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

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

דוגמה:

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

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

ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?

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

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

© כל הזכויות שמורות לברק קנדל

פתרון:

א. אוכלוסייה: משלמי חשבון טלפון סלולאר כיום.

המשתנה: X = חשבון הטלפון החודשי בשקלים.

הפרמטר: μ.

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

תנאים:

1. $\mu = 200$.

2. $n = 36$.

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

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

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

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

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

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

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

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

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

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

ד. הקטנת α מגדילה את β.

© כל הזכויות שמורות לברק קנדל

שאלות:

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

להלן השערות של חוקר לגבי הפרמטר μ: $H_1: \mu = 7$, $H_0: \mu = 5$.

מעוניינים ליצור כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.

א. עבור אילו ערכים של X שידגם נדחית השערת H_0?

ב. מה הסיכוי לבצע טעות מסוג שני?

ג. אם במדגם התקבל ש-$X = 6.9$ מה תהיה המסקנה ומה הטעות האפשרית?

2) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. מעוניינים לבדוק אם כיום ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות.
במדגם התקבל ממוצע 2.17 ילדים למשפחה.

א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 5%.

ב. בהמשך לסעיף א' מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?

ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העצמה של הכלל מסעיף א'?

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

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

א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.

ב. בהמשך לסעיף א', מהי העצמה אם התוחלת שווה ל-195?

ג. הסבירו, ללא חישוב, איך העצמה תשתנה אם רמת המובהקות תהיה 5%?

4) מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50 מ"מ וסטית תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומודדים את קוטרם, בכדי לבדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוילת כנדרש או שקוטר הצינורות קטן מהדרוש.

א. רשמו את ההשערות ואת כלל ההכרעה ברמת מובהקות של 5%.

ב. אם ביום כלשהו מכונת הייצור התקלקלה והיא מייצרת את הצינורות בקוטר שתוחלתו 48 מ"מ בלבד (סטית התקן לא השתנתה), מה ההסתברות שהתקלה לא תתגלה בביקורת האיכות? כיצד נקראת הסתברות זו?

ג. הסבירו ללא חישוב כיצד התשובה לסעיף ב תשתנה אם רמת המובהקות תגדל.

ד. הסבירו ללא חישוב כיצד התשובה לסעיף ב תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

© כל הזכויות שמורות לברק קנדל

5) להלן השערות של מחקר: $H_1:\mu=58$, $H_0:\mu=50$.
מעוניינים לדגום 100 תצפיות. ידוע שסטיית התקן של ההתפלגות הינה 20.

א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10%.
מהי רמת המובהקות?

ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו)?

i. סטיית התקן הייתה יותר גדולה.

ii. הסיכוי לטעות מסוג שני גדול יותר.

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

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

א. הסיכוי לטעות מסוג ראשון גדל.

ב. העוצמה של המבחן גדלה.

ג. הסיכוי לטעות מסוג שני גדל.

ד. תשובות א' ו-ב' נכונות.

7) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן:

א. השערת האפס נכונה.

ב. השערת האפס נדחתה.

ג. השערת האפס לא נדחתה.

ד. אף אחת מהתושבות לא נכונה בהכרח.

8) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

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

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

א. הן 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.

© כל הזכויות שמורות לברק קנדל

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

א. טעות מסוג ראשון.

ב. טעות מסוג שני.

ג. טעות מסוג שלישי.

ד. אין טעות במסקנתו.

תשובות סופיות:

1) א. מעל 6.645. ב. 0.3594.

ג. דחינו את H_0, תתכן טעות מסוג ראשון.

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

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

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

5) א. 0.0033. ב. i. רמת המובהקות הייתה קטנה.

ב. ii. רמת המובהקות הייתה גדלה.

6) ד'.

7) ג'.

8) ג'.

9) א'.

10) ב'.

© כל הזכויות שמורות לברק קנדל

קביעת גודל מדגם (ששונות האוכלוסייה ידועה):

רקע:

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

סטיית התקן של האוכלוסייה ידועה σ ומעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β.

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

דוגמה:

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

פתרון:

<u>האוכלוסייה</u>: ילדי גן חובה.

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

<u>הפרמטר</u>: μ.

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

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

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

© כל הזכויות שמורות לברק קנדל

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

$$\alpha = 0.05$$

$$1-\beta = 0.9$$

$$\mu_0 = 70$$

$$\mu_1 = 80$$

$$\sigma = 17$$

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

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

לכן, $n_{\min} = 25$.

© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

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

תשובות סופיות:

1) 41.

2) 78.

3) שאלת הוכחה.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 8 - רווח סמך לפרופורציה

תוכן העניינים

1. רווח הסמך לפרופורציה 42
2. קביעת גודל מדגם 45

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?

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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. $22.5\% < p < 30.9\%$. ב. $45.91\% < p < 60.72\%$.

6) 200.

© כל הזכויות שמורות לברק קנדל

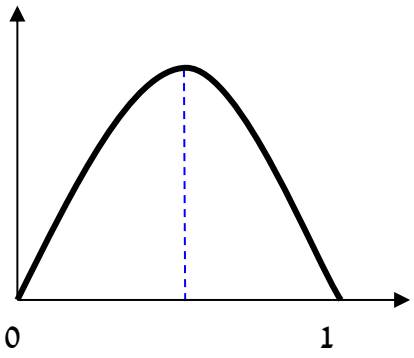
קביעת גודל מדגם:

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%.

© כל הזכויות שמורות לברק קנדל

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?

תשובות סופיות:

1) 1068.

2) א. 423. ב. 271.

3) א. 601. ב. 108,000 ₪.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

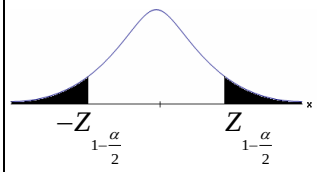
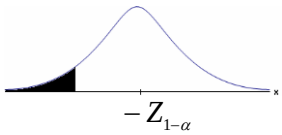
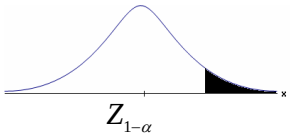
פרק 9 - בדיקת השערות על פרופורציה

תוכן העניינים

1. התהליך 47

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 $Z_{\hat{p}} > Z_{1-\frac{\alpha}{2}}$  H_0 דוחים את	$Z_{\hat{p}} < -Z_{1-\alpha}$  H_0 דוחים את	$Z_{\hat{p}} > Z_{1-\alpha}$  H_0 דוחים את

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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) נדחה H_0.

2) נדחה H_0.

3) נקבל H_0.

4) א. $\begin{matrix} H_0 : p = 0.2 \\ H_1 : p \neq 0.2 \end{matrix}$. ב. נקבל H_0. ג. המסקנה לא תשתנה.

5) א. נדחה H_0. ב. המסקנה לא תשתנה.

6) ג'.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 10 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסיה לא ידועות ובהנחת שוויון שונויות..... 50

כששונויות האוכלוסייה לא ידועות ובהנחת שוויון שונויות:

רקע:

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

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

התנאים לבניית רווח הסמך:

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

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

דרגות החופש: $d.f = n_1 + n_2 - 2$.

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

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

דוגמה (פתרון בהקלטה):

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

	תל אביב	באר שבע
מספר האקדמאים	20	10
ממוצע הכנסות של אקדמאים	11,000	9500
סטיית התקן של הכנסות אקדמאים	200	250

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

© כל הזכויות שמורות לברק קנדל

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

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

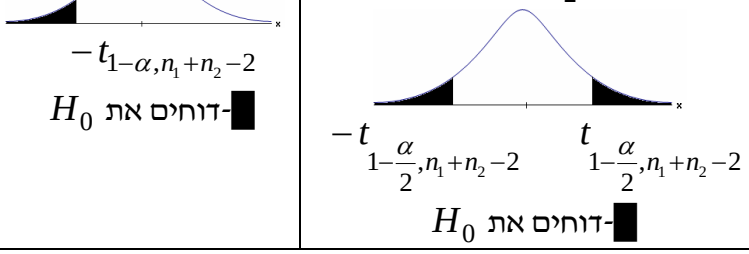
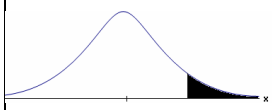
פרק 11 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסיה לא ידועות ומניחים שהן שוות..... 52

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 $t_{\bar{x}_1-\bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  H_0 דוחים את	$t_{\bar{x}_1-\bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$ H_0 דוחים את	$t_{\bar{x}_1-\bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  H_0 דוחים את

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:

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

© כל הזכויות שמורות לברק קנדל

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%.

© כל הזכויות שמורות לברק קנדל

שאלות

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

2012	120	94	90	130	95	112	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 60 דולקות 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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 12 - רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים

תוכן העניינים

1. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 56

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.

© כל הזכויות שמורות לברק קנדל

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) $-0.013 < \mu < 0.185$.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

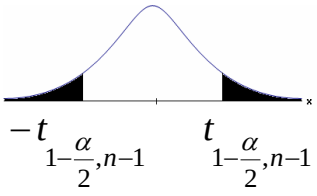
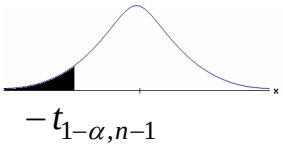
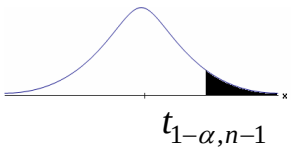
פרק 13 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

תוכן העניינים

1. בדיקת השערות למדגמים מזווגים 58

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 $t_{\bar{D}} < -t^{(n-1)}_{1-\frac{\alpha}{2}}$  ■ - דוחים את H_0	$t_{\bar{D}} < -t^{(n-1)}_{1-\alpha}$  ■ - דוחים את H_0	$t_{\bar{D}} > t^{(n-1)}_{1-\alpha}$  ■ - דוחים את H_0
חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:	$\bar{D} > C + t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{S_D}{\sqrt{n}}$ או $\bar{D} < C - t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t^{n-1}_{1-\alpha} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t^{n-1}_{1-\alpha} \cdot \frac{S_D}{\sqrt{n}}$

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

דוגמה (פתרון בהקלטה):

© כל הזכויות שמורות לברק קנדל

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".

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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

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

נניח שהמשקלים מתפלגים נורמלית, המבחן שיש לבצע כאן הוא:

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

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

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

תשובות סופיות

1) לא נדחה H_0.

2) לא נדחה H_0.

3) א. $0.25 \le pv \le 0.5$ ב. 0.5 ג. לא נדחה H_0.

4) ד'.

5) ד'.

6) ג'.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 14 - הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות

תוכן העניינים

1. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות..... 62

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

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

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

© כל הזכויות שמורות לברק קנדל

שאלות

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

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

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

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

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

מרצה X	82	90	68
מרצה Y	68	81	64

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

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

שאלות רב-ברירה:

3) סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים.
הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח $-2 < \mu_1 - \mu_2 < 4.5$.
אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות:
$H_0 : \mu_1 - \mu_2 = 0$; $H_1 : \mu_1 - \mu_2 \neq 0$ ברמת מובהקות 0.05, מסקנתו תהיה:

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

ג. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05 .

ד. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

© כל הזכויות שמורות לברק קנדל

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 מתפלגים נורמלית בנו רווח סמך למוצע 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.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 15 - ניתוח שונות חד כיוונית

תוכן העניינים

1. כללי 65

ניתוח שונות חד כיוונית

רקע תיאורטי

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות $\left(\mu_{1..},\mu_k\right)$ של k אוכלוסיות שונות.
לכן, בניתוח שונות, <u>השערות</u> 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:

1) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.

2) כל האוכלוסיות הן עם אותה שונות σ^2.

3) 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: $SS_W = \sum_{i=1}^{k} n_i \left[n_i - 1\right] \bullet \hat{S}_i^2$

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

$SST = SSB + SSW$

© כל הזכויות שמורות לברק קנדל

יש למלא את טבלת ניתוח השונות הבאה:

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

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

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

© כל הזכויות שמורות לברק קנדל

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 "אקמול" 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:

"המתמיד"	"רבין"	"הס"
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?

© כל הזכויות שמורות לברק קנדל

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			

© כל הזכויות שמורות לברק קנדל

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 בהשוואת מרובות. נמקו!

ד. לטבלת הANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת ״טוקי״. 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			

© כל הזכויות שמורות לברק קנדל

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

(I) school	(J) school	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	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	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.

© כל הזכויות שמורות לברק קנדל

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. שוויון 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 שונות.

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.

© כל הזכויות שמורות לברק קנדל

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 שונות.

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

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 16 - מבחני חי בריבוע

תוכן העניינים

1. מבחן טיב התאמה..... (ללא ספר)
2. מבחן לאי תלות..... (ללא ספר)
3. מקדם המתאם של קרמר..... 74

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.

© כל הזכויות שמורות לברק קנדל

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

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 17 - גודל האפקט

תוכן העניינים

1. גודל האפקט בהסקה על תוחלת שהשונות ידועה ..... (ללא ספר)
2. גודל האפקט בהסקה על תוחלת שהשונות לא ידועה ..... (ללא ספר)
3. גודל האפקט בהסקה על תוחלת ההפרש במדגמים מזווגים ..... (ללא ספר)
4. גודל האפקט בהסקה על הפרש תוחלות במדגמים בלתי תלויים ובהנחת שוויון שונויות . (ללא ספר)
5. גודל האפקט במבחן לניתוח שונות ..... (ללא ספר)

יסודות הסטטיסטיקה ב

פרק 18 - מבחנים אפרמטרים למדגמים מזווגים

תוכן העניינים

1. מבחן הסימן 76

מבחנים אפרמטרים למדגמים מזווגים

מבחן הסימן – רקע

מבחן הסימן הוא מבחן שמשתמשים בו כאשר לפנינו מדגם מזווג ולא ניתן להניח שהמשתנה הנחקר מתפלג נורמלית.

גם אם המשתנה הנחקר מתפלג נורמלית ניתן לבצע את מבחן הסימן אבל מבחן T למדגמים מזווגים יהיה מבחן עם עוצמה גבוהה יותר ולכן יש לבצע אותו.

מבחן הסימן נחשב למבחן אפרמטרי – מבחנים אפרמטרים הינם כל המבחנים שאינם דורשים שהמשתנה הנחקר יתפלג נורמלית.

מבחן הסימן נקרא כך כיוון שהוא דורש הגדרת סימן לכל תצפית:

$(+)$ – אם מצב X גבוה ממצב Y.

$(-)$ – אם מצב Y גבוה ממצב X.

(0) – אין הבדל בין המצבים.

במבחן הסימן נתעלם מהפרשים שהם 0, ולכן נסמן את מספר ההפרשים האפקטיביים (השונים מאפס) ב- n^*.

תחת השערת האפס נאמר שהסיכוי לקבל הפרש חיובי $(p+)$ שווה לסיכוי לקבל הפרש שלילי $(p-)$.

השערות המבחן:

$$H_0: P_+ = 0.5 \quad H_1: P_+ \neq,<,> 0.5 \quad \text{או} \quad H_0: P_- = 0.5 \quad H_1: P_- \neq,<,> 0.5$$

נסמן ב $n(+)$ או ב S_+ את מספר התצפיות שקיבלו את הערך (+), ובאופן דומה:

נסמן ב $n(-)$ או ב S_- את מספר התצפיות שקיבלו את הערך (–).

ניתן לומר שבהנחת השערת האפס: $n(+) \setminus n(-) \sim B(n^*, 0.5)$.

נחשב את PV על סמך תוצאות המדגם בעזרת ההתפלגות הבינומית כך שאם $PV \leq \alpha$ נדחה את השערת האפס.

במבחן הסימן אין התייחסות לגודל הפער בתצפיות אלא רק את כיוון ההבדל.

דוגמה (פתרון בהקלטה):

קופות החולים טוענות כי רכישת תרופות שאינן דורשות מרשם רופא, הינן זולות יותר אצלן מאשר מברשתות הפארם. דגמו 11 תרופות ובדקו את מחירן בבית המרקחת של קופות החולים וברשת הפארם. המחיר המוצג הינו עבור קפסולה בודדת:

בדקו ברמת מובהקות של 5% באמצעות מבחן הסימן.

© כל הזכויות שמורות לברק קנדל

שם התרופה	קופת חולים	פארם
אדוויל	1.2	1.5
אקמול	2.6	2.6
אופטלגין	0.9	1.4
פוסטינור	3.5	3.2
סטרפסיל	1.1	1.4
נורפן	1.7	1.8
לורסטין	0.8	1.1
קולדקס	1.5	2
אלרגיז	2	2.8
נוסידקס	2	2.5
קורמיר	3	3.3

שאלות

1) רוצים לבדוק את הטענה שהציונים במבחן בסטטיסטיקה ב גבוהים מאשר בסטטיסטיקה א. נלקחו 10 סטודנטים שסיימו את סטטיסטיקה ב.
עבור כל סטודנט נבדק מה הציון בסטטיסטיקה א ומה הציון בסטטיסטיקה ב.
להלן התוצאות שהתקבלו:

א	62	74	68	94	82	67	65	84	78	80
ב	70	80	70	90	77	67	80	86	79	82

א. בדקו ברמת מובהקות של 5% באמצעות מבחן הסימן.

ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם יוחלט לתת פקטור של 2 נקודות לכל הסטודנטים בשני המועדים?

2) מעוניינים לבדוק האם ההוצאות על "ג'אנק פוד" בקרב הסטודנטים רבות יותר בזמן הלימודים לעומת ימי החופשה.
נדגמו 15 סטודנטים מקריים, אצל 13 ההוצאות בתקופת הלימודים היו גבוהות יותר מימיי החופשה ואצל 2 נמוכות יותר.
מה מסקנתך בר"מ של 0.05?

3) מעוניינים לבדוק האם סם מסוים משפיע על לחץ הדם. נלקחו 24 אנשים אשר נמדד להם לחץ הדם לאחר מכן ניתן להם הסם ושוב מדדו להם את לחץ הדם.
לחמישה אנשים לחץ הדם לא השתנה ל 15 אנשים לחץ הדם עלה וליתר לחץ הדם ירד אחרי לקיחת הסם. מה מסקנתכם ברמת מובהקות של 5%?

© כל הזכויות שמורות לברק קנדל

4) במדגם שנעשה על 15 משפחות השוו את רמת הביטחון העצמי של הבכור במשפחה לעומת הצעיר שבמשפחה. תוצאות המדגם הראו שאצל 7 משפחות רמת הביטחון העצמי של הבכור הייתה גבוהה יותר, אצל 3 משפחות רמת הביטחון העצמי של הצעיר הייתה גבוהה יותר ואצל 5 משפחות לא נמצא הבדל בין האחים מבחינת רמת הביטחון העצמי. טענת החוקר הייתה שבמשפחות לבכור ביטחון עצמי גבוה מזה של הצעיר במשפחה.

א. מהי רמת המובהקות המינימלית עבורה יוחלט לקבל את טענת החוקר?

ב. רמת הביטחון הוערכה על ידי פסיכולוג זוטר. פסיכולוג בכיר ביצע הערכה מחודשת וקבע שלמשפחה אחת במדגם הייתה הערכה שגויה: הפסיכולוג הזוטר קבע שלצעיר במשפחה יש ביטחון עצמי יותר גבוה למרות שלאח הבכור יש ביטחון עצמי יותר גבוה במשפחה הזו. מה יקרה לרמת המובהקות המינימלית שחושבה בשאלה הקודמת?

5) איזה מהטענות הבאות נכונות?

א. $n(+)+n(-)=n^*$

ב. $n(+)+n(-)=n$

ג. $n(+)=n(-)$

ד. $n(+)-n(-)=n^*$

תשובות סופיות

1) א. לא נדחה את H_o. ב. לא תשתנה המסקנה.

2) נדחה את H_o.

3) נדחה את H_o.

4) א. 0.172 ב. תקטן.

5) א'.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 19 - שאלות מסכמות בבדיקת השערות

תוכן העניינים

1. שאלות רב ברירה (אמריקאיות) 79

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. $H_1 : \mu_d > 0, \quad H_0 : \mu_d = 0$

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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

© כל הזכויות שמורות לברק קנדל

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 מהתושבות 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 שה- $p-value=0.02$.

מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%?

בחר בתשובה הנכונה:

א. יקבל את השערת האפס בכל מקרה.

ב. ידחה את השערת האפס מקרה.

ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.

ד. לא ניתן לדעת כי אין מספיק נתונים.

© כל הזכויות שמורות לברק קנדל

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: $H_1: \mu < \mu_0, \quad H_0: \mu = \mu_0$.
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

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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 למוצע 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: $H_1: \mu \neq 10, \quad H_0: \mu = 10$.
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

© כל הזכויות שמורות לברק קנדל

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

א. יקטן הסיכוי לטעות מסוג שני (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

© כל הזכויות שמורות לברק קנדל

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 **<u>נכון</u>** 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.

© כל הזכויות שמורות לברק קנדל

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

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 20 - שאלות מסכמות על רווחי סמך

תוכן העניינים

1. שאלות מסכמות על רווחי סמך..... 92

שאלות מסכמות על רווחי סמך:

שאלות:

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

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

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

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

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

א. כמה אנשים במדגם טענו שכלל לא התקררו השנה?

ב. באיזו רמת סמך נבנה רווח הסמך?

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

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

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

א. אמדו ברמת ביטחון של 90% את ממוצע ציוני בחינת ה-IQ בישראל – מהי ההנחה הדרושה לפתרון?

ב. על סמך רווח הסמך של סעיף א' האם תקבלו את הטענה שבישראל ממוצע הציונים שונה מארה"ב?

ג. מה היה קורה לרווח הסמך אם היינו מגדילים את רמת הסמך שלו?

© כל הזכויות שמורות לברק קנדל

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.

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

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 21 - מדד הקשר של ספירמן

תוכן העניינים

1. הקדמה ..... (ללא ספר)
2. חישוב מדד הקשר של ספירמן ..... (ללא ספר)
3. בדיקת השערות על מדד ספירמן באמצעות התפלגות T ..... (ללא ספר)
4. ספירמן- ניתוח פלטי SPSS ..... (ללא ספר)

יסודות הסטטיסטיקה ב

פרק 22 - מקדם המתאם (מדד קשר) הלינארי ומובהקותו

תוכן העניינים

1. מקדם המתאם הלנארי (פירסון) 94
2. חישוב מקדם המתאם של פירסון (לנארי) במדגם ..... (ללא ספר)
3. בדיקת השערות על מקדם המתאם הלנארי 105
4. ניתוח פלטים על מקדם המתאם הלנארי 109

מקדם המתאם (מדד קשר) הלינארי ומובהקותו

מדד הקשר הלינארי (פירסון) – מבוא

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

דוגמה (פתרון בהקלטה) :

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

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

1) כמה תצפיות ישנן בדוגמה?

2) כמה משתנים ישנם בדוגמה, מי הם?

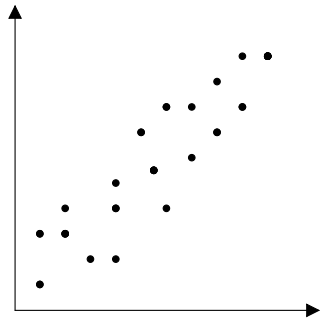
3) שרטטו לנתונים דיאגרמת פיזור.

4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

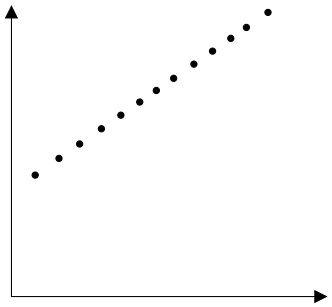
© כל הזכויות שמורות לברק קנדל

דיאגרמות פיזור לקשר בין משתנים וניתוחם

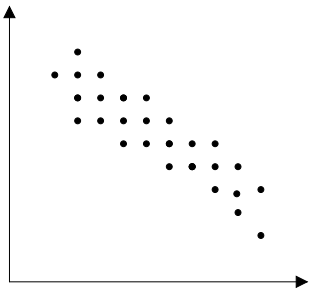
קשר לנארי חיובי חלקי



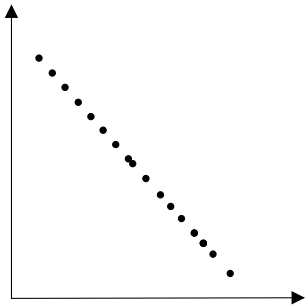
קשר לנארי חיובי מלא



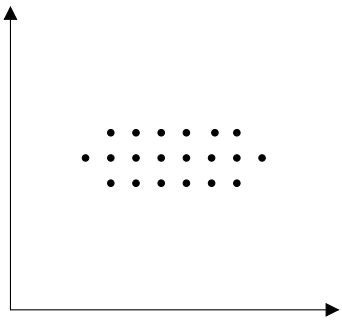
קשר לינארי שלילי חלקי



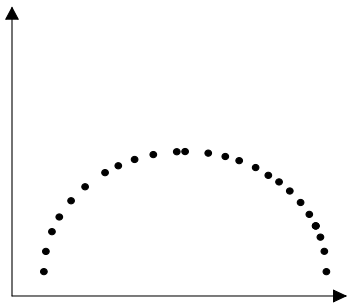
קשר לינארי שלילי מלא



אין קשר



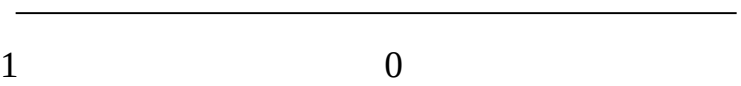
אין קשר לינארי



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.



© כל הזכויות שמורות לברק קנדל

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.

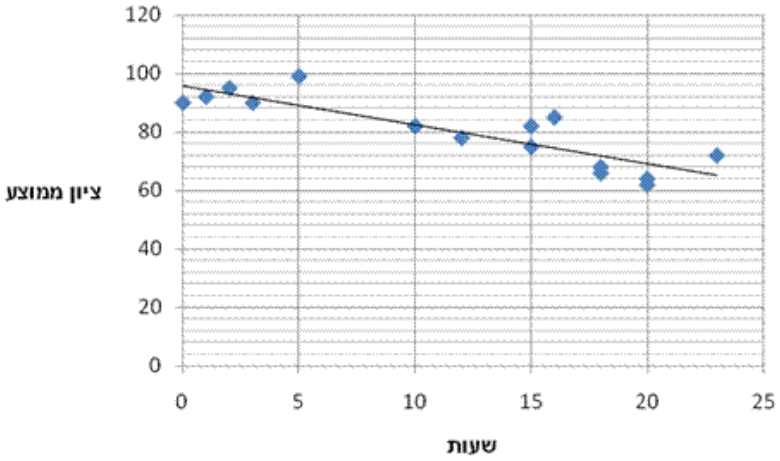
דוגמה (פתרון בהקלטה) :

- 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?

© כל הזכויות שמורות לברק קנדל

שאלות

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



א. מיהו המשתנה הבלתי תלוי?

ב. מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר? מה ניתן להגיד על עוצמת הקשר?

2) להלן טבלה המסכמת את מקדמי המתאם הלינארי בין ציוני מבחנים שונים שהתקבלו עבור תלמידים בכיתה מסוימת:

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

א. השלימו את מקדמי המתאם שמסומנים בסימן שאלה בטבלה.

ב. בין אילו שני ציוני מקצועות שונים קיים מתאם בעל העוצמה החזקה ביותר?

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

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

א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי בדוגמה זו?

ב. שרטטו דיאגרמת פיזור לנתונים.

ג. מה ניתן לומר על הקשר בין המשתנים במדגם?

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

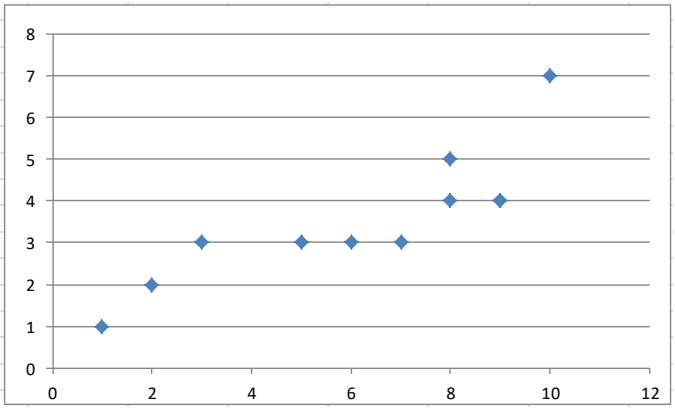
© כל הזכויות שמורות לברק קנדל

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 _____ (גדל\ 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.

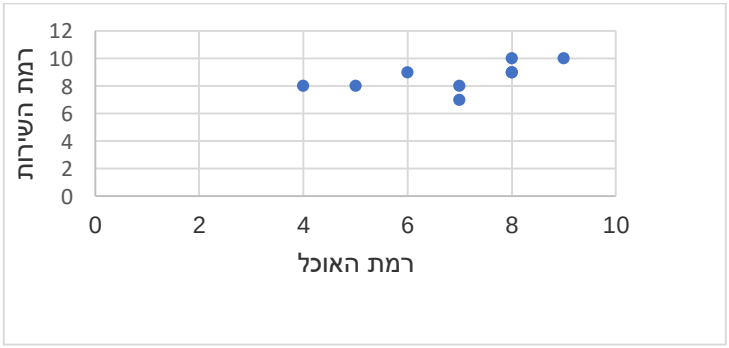
ב. במדגם, רוב הטמפרטורות בטורונטו היו שליליות.

ג. ההפרש בין הטמפרטורה בטורונטו לבין הטמפרטורה באוסטרליה, במדגם זה, הוא שלילי.

ד. במדגם יש נטייה שהטמפרטורה יורדת בטורונטו לטמפרטורה לעלות בסידני.

© כל הזכויות שמורות לברק קנדל

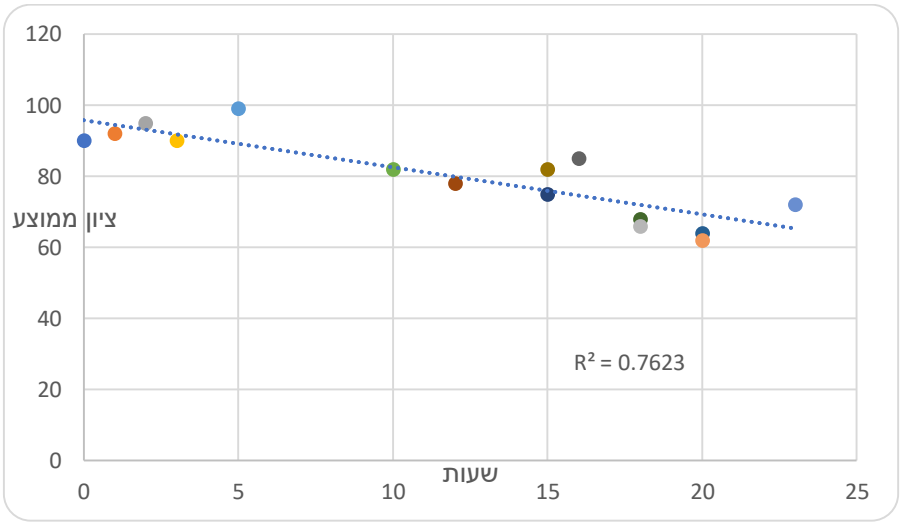
7) בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



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.

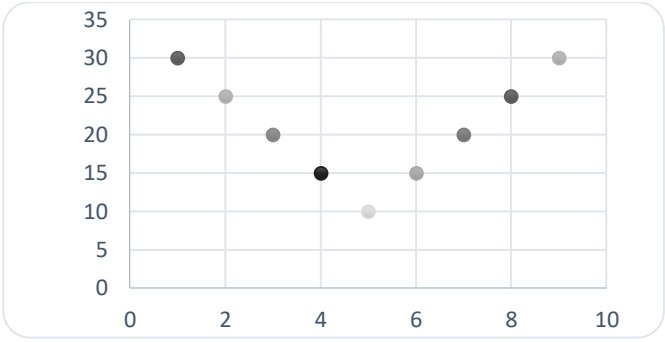


מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?

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.

© כל הזכויות שמורות לברק קנדל

9) התרשים הבא מתאר קשר בין שני משתנים, איזה מהמתאמים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?



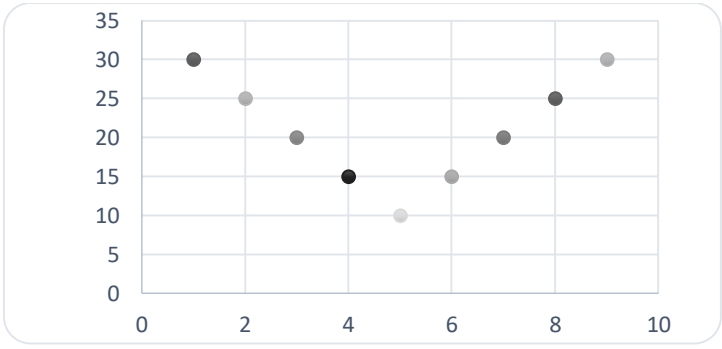
א. $r = 1$ היות ושני המשתנים יוצרים קוים ישרים.

ב. $r = 2$ היות ויש שני קוים בעלי קשר מושלם.

ג. $r = 0$ היות והקו יורד ואחר כך עולה באותו האופן.

ד. $r = \pm 1$ היות ויש קו עולה וגם קו יורד.

10) התרשים הבא מתאר דיאגרמת פיזור.



איזו טענה נכונה?

א. בתרשים מוצג הקשר בין שני משתנים.

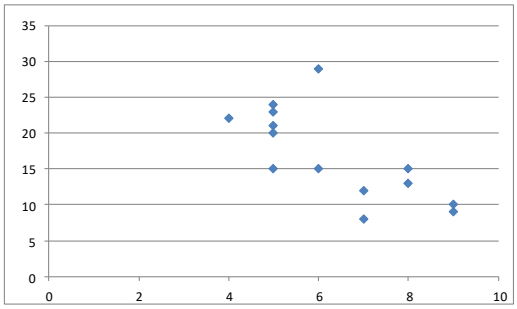
ב. בתרשים מוצג הקשר בין 9 משתנים.

ג. בתרשים מוצג הקשר בין 10 משתנים.

ד. אין לדעת כמה משתנים מוצגים בתרשים.

© כל הזכויות שמורות לברק קנדל

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים:



X - (משתנה בלתי תלוי בציר האופקי)
ו- Y (משתנה תלוי).
במדגם התקבל $r^2 = 0.52$.

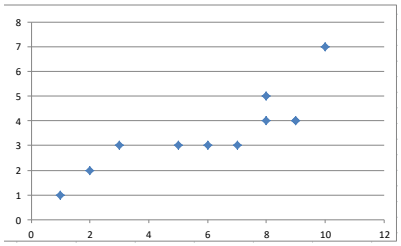
11) לאור הנתונים המופיעים בדיאגרמה, איזה מבין הערכים הבאים מתאים להיות התוצאה של r?

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.

© כל הזכויות שמורות לברק קנדל

15) לפי הפתגם "רחוק מהעין, רחוק מהלב", יש קשר ____ בין קרבה פיזית לקרבה נפשית.

א. חיובי

ב. שלילי

ג. אפסי

ד. לא ניתן לדעת.

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

		159
		170
		180
		185
		204
		224
		236
		212
		168
		189
		195
		163
		187
		206
		201
		223
		242
		203
		205
197.47	AVERAGE	
536.25	VARPA	

יש להוסיף עמודה נוספת לצד עמודת הציונים שתראה לכל תלמיד כמה נקודות חסרות לו כדי להשלים לציון המקסימלי בבחינה.

מה יהיה מקדם המתאם בין שתי העמודות (כלומר, מקדם המתאם בין הציון לבין הנקודות החסרות)?

א. -1

ב. 1

ג. -0.5

ד. 0.5

17) מקדם המתאם בין שטחי דירה למחיר שלהם חושב ונמצא 1.2. מה נובע מכך?

א. ככל שהדירה גדולה יותר בשטחה כך היא יקרה יותר.

ב. ככל שהדירה קטנה יותר בשטחה כך היא זולה יותר.

ג. לא קיים קשר בין שטח הדירה למחיר הדירה.

ד. מצב כזה שמתואר הנתונים לא אפשרי.

18) אם ניקח 10 אנשים ונרשום לכל אדם את הגובה במטר וכמו כן את הגובה בס"מ. מה יהיה מקדם המתאם בין גובה האדם במטר לגובה האדם בס"מ?

א. 1

ב. 0

ג. -1

ד. לא ניתן לדעת.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. שעות בילוי. ב. הקשר חלקי, כיוון הקשר שלילי .

2) א. להלן טבלה:

	ספורט	לשון	מתמטיקה
ספורט	1	-0.7	0.1
לשון	-0.7	1	0.6
מתמטיקה	-0.1	0.6	1

ב. ספורט ולשון.

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'.

© כל הזכויות שמורות לברק קנדל

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

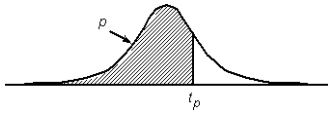
סטטיסטי זה מתפלג t עם $n-2$ דרגות חופש.

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	$t\geq t_{1-\alpha}$ או $t\leq -t_{1-\alpha}$	$t\leq -t_{1-\alpha}$	$t\geq t_{1-\alpha}$

© כל הזכויות שמורות לברק קנדל

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

© כל הזכויות שמורות לברק קנדל

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 -ווותק	13	18	2	8	16	9	5	28	17	24
Y -השכלה	12	14	17	8	11	12	13	8	12	15

מקדם המתאם חושב והתקבל: 0.31-.

א. האם קיים מתאם בין וותק העובד להשכלתו? בדקו ברמת מובהקות של 5%?

ב. אם הוותק של העובד היה נמדד בחודשים האם התשובה לסעיף א' הייתה משתנה?

2) מחקר התעניין לבדוק את הקשר בין גיל נשים בהריון לרמת ההמוגלובין שלהן בדם בזמן הריון. נדגמו 7 נשים והתקבלו התוצאות הבאות:

נחקרת	1	2	3	4	5	6	7
המוגלובין	14.7	13.5	9.7	12	10.8	13	10.3
גיל	39	34	30	29	28	26	23

במדגם חושב מדד הקשר של פירסון להיות 0.7.

א. האם ניתן לומר שבמדגם אם אישה היא יותר מבוגרת אזי בהכרח יש לה יותר המוגלובין בדם?

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

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

א. בדקו ברמת מובהקות של 2.5% האם קיים קשר לינארי שלילי בחודש ינואר בין הטמפרטורה במעלות צלזיוס לבין המשקעים במעלות צלזיוס.

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

ג. על סמך טבלת T המצורפת עבור אילו רמות מובהקות ניתן להחליט שקיים קשר לינארי שלילי מובהק?

4) מתווך דירות חישב את מקדם המתאם בין שטח דירה במרכז תל אביב לבין המחיר של הדירה עבור 17 דירות. מקדם המתאם שקיבל היה 0.6.

א. בדוק ברמת מובהקות של 5% האם ניתן להגיד שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה במרכז תל אביב?

ב. מהי מובהקות התוצאה לבדיקת ההשערה שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה בתל אביב.

© כל הזכויות שמורות לברק קנדל

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. $0.005 < P_V < 0.01$.

© כל הזכויות שמורות לברק קנדל

מדד הקשר הלינארי – ניתוח פלטים – רקע

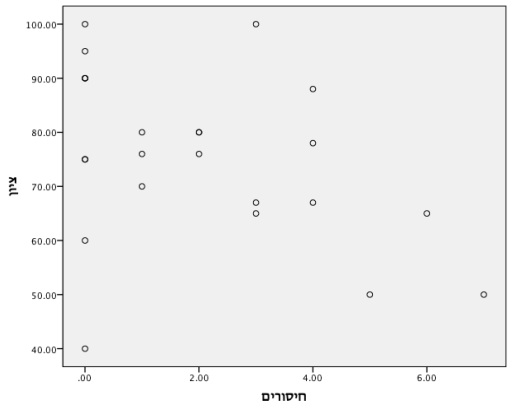
מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:
r - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר ρ.

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.

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 הלינארי 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 לינארי שלילי בין מספר החיסורים של הסטודנטים מהקורס לבין הציון של הסטודנטים בקורס?

© כל הזכויות שמורות לברק קנדל

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%?

© כל הזכויות שמורות לברק קנדל

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

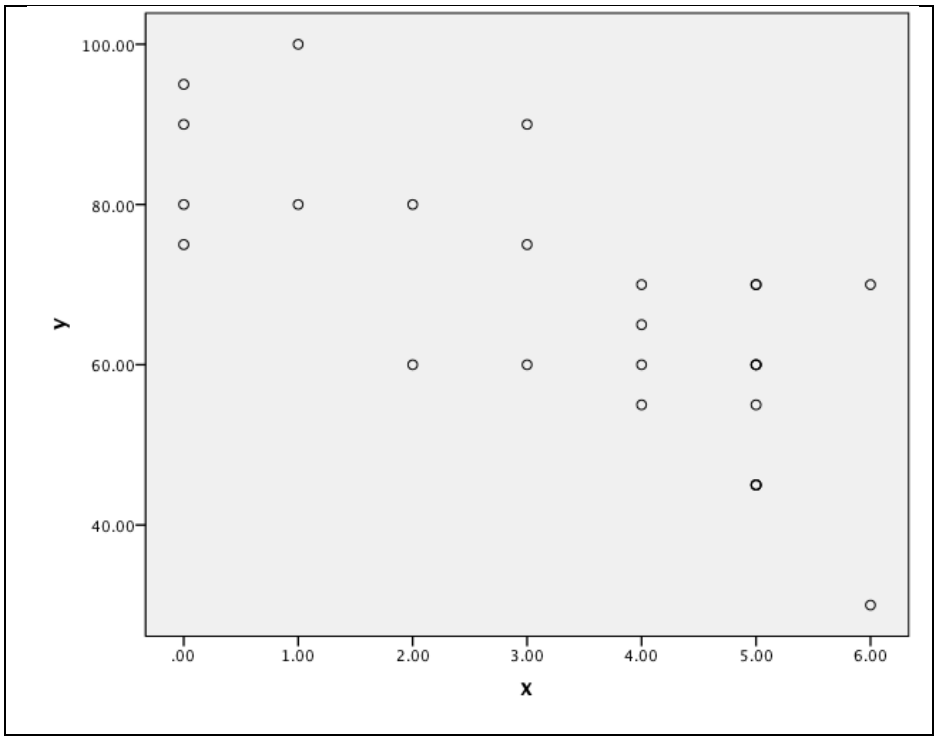
א. בין אילו שני משתנים שונים הקשר הלינארי במדגם נמצא עם העוצמה הכי חזקה?

ב. ברמת מובהקות של 5%, אילו שני משתנים בעל קשר לינארי מובהק?

ג. השלימו את המספרים המסומנים בפלט בסימני שאלה.

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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

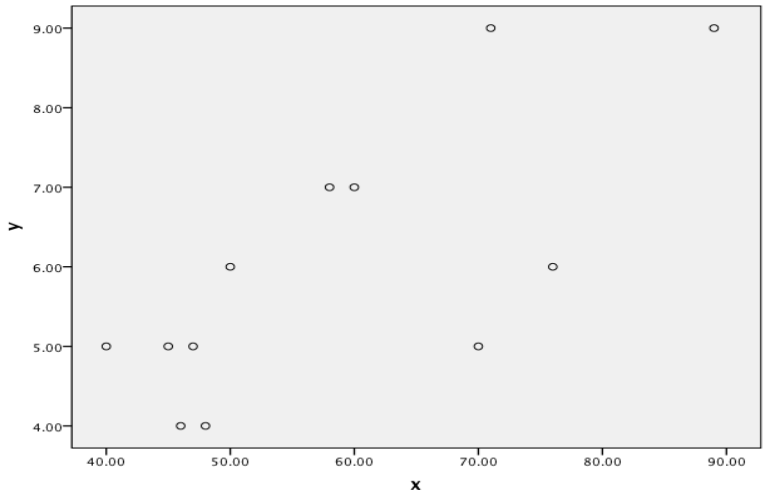
א. בדקו ברמת מובהקות של 5% האם קיים קשר לנארי חיובי בין המשקל והגובה.

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?

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

יסודות הסטטיסטיקה ב

פרק 23 - מדדי קשר - בחירת מדד מתאים

תוכן העניינים

1. בחירת מדד מתאים 116

מדדי קשר – בחירת מדד מתאים:

רקע:

בפרק זה נתרגל את התהליך של בחירת מדד הקשר (מקדם המתאם) המתאים.
נתרכז בשלושת מדדי הקשר הנפוצים ביותר:

- מדד הקשר של קרמר.
- מדד הקשר של ספירמן.
- מדד הקשר של פירסון (מדד הקשר הלינארי).

בחירת מדד הקשר נעשה לפי סולמות המדידה של שני המשתנים שאנחנו רוצים לבדוק את הקשר בינם. הנושא של סולמות מדידה נלמד כבר בפרק אחר, כמו כן כל מדד קשר נלמד בפרק נפרד. אנו מתרכזים ב 3 סולמות מדידה:

- סולם שמי/ זהות (nominal).
- סולם סדר (ordinal).
- סולם כמותי (scale): לכאן אנו מאחדים את סולם רווחים ומנה יחד.

שלושת מדדי הקשר שלעיל דנים בקשר בין שני משתנים. מדדי הקשר הם סימטריים, כלומר אין זה משנה איזה משתנה נגדיר בתור משתנה X ואיזה יוגדר בתור משתנה Y.

להלן טבלה שמסכמת את בחירת המדד המתאים:

X / Y	שמי	סדר	כמותי
שמי	קרמר	קרמר	קרמר
סדר	קרמר	ספירמן	ספירמן
כמותי	קרמר	ספירמן	פירסון

© כל הזכויות שמורות לברק קנדל

דוגמה (פתרון בהקלטה):

במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום.
איזה מדד קשר מתאים כדי לבדוק האם קיים קשר בין התפוקה היומית לכמות עובדים באותו היום במפעל?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. חסרים נתונים כדי לדעת זאת.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) בקרב תלמידי כיתות א' בבית הספר גבריאלי אשר בתל אביב בדקו לכל תלמיד את גובהו בס"מ ואת משקלו בק"ג. מהו מדד הקשר המתאים כדי לבדוק האם קיים קשר בין גובה התלמיד למשקלו?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

2) בסקר שנעשה על אזרחים במדינה בדקו לכל אזרח את השכלתו ואת שכרו. מהו מדד הקשר המתאים כדי לבדוק האם קיים קשר בין השכלה לשכר?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

3) בגן הילדים של שולה אספו נתונים על 25 הילדים שבגן. על כל ילד בדקו את רמת הביטחון העצמי שלו (X = מדד שמקבל ערכים בין 1 - נמוך ועד 5 - גבוה), ואת אוצר המילים שלו (Y = לפי מבחן שנעשה לכל ילד בו ספרו את מספר המילים שידע מתוך רשימה של 20 מילים). איסוף הנתונים נעשה על ידי איש מקצוע שצפה בילדים ובחן אותן.
מהו מקדם המתאם המתאים לבדיקת התלות בין X לבין Y?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

4) הועלתה השערה בבית הספר למדעי ההתנהגות שיש קשר בין המרצה להצלחת הסטודנט. לצורך בדיקת הטענה בדקו לגבי כל סטודנט שלמד סטטיסטיקה אצל איזה מרצה הוא למד (היו 3 מרצים שונים) והאם הוא עבר את הבחינה.
מהו מדד הקשר המתאים במקרה זה?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

© כל הזכויות שמורות לברק קנדל

5) בבורסה בתל אביב רצו לבדוק את הקשר בין גובה הריבית במשק בסוף החודש (באחוזים), לבין תשואת מניית אקטר (באחוזים) בסוף החודש.
מהו מדד הקשר המתאים?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

6) בעל מסעדה ביצע סקר על לקוחותיו, בין השאלות שנשאלו בסקר:

- מה מידת שביעות הרצון של הלקוח מאדיבות השירות של המלצר בסקלה של 1 עד 5.
- מה גילו של הלקוח בשנים.
- מה גובה התשר (טיפ) ב-₪ אשר נתן הלקוח למלצר בלכתו מהמסעדה.

מהו המדד המתאים כדי לבדוק האם קיים מתאם חיובי בין מידת שביעות הרצון של הלקוח מאדיבות השירות לבין גובה התשר שהוא נתן למלצר?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

7) בעל מסעדה ביצע סקר על לקוחותיו, בין השאלות שנשאלו בסקר:

- מה מידת שביעות הרצון של הלקוח מאדיבות השירות של המלצר בסקלה של 1 עד 5.
- מה גילו של הלקוח בשנים.
- מה גובה התשר (טיפ) ב-₪ אשר נתן הלקוח למלצר בלכתו מהמסעדה.

מהו המדד המתאים כדי לבדוק האם קיים מתאם בין גיל הלקוח לגובה התשר שהעניק לשירות?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

© כל הזכויות שמורות לברק קנדל

הנתונים הבאים מתאימים ל-3 השאלות הבאות:

חוקרים ערכו מדגם של ילדים מכיתות ב' ו-ג' מ-4 בתי ספר שונים. הועבר לילדים שאלון בו תואר מצב מסוים והילדים התבקשו לציין את רמת החרדה שלהם באשר לאותו מצב. המשתנים שלגביהם נאספו נתונים:

- מגדר (1 - בן, 2 - בת).
- כיתה (0 - ג', 1 - ב').
- בית ספר (A, B, C, D).
- רמת חרדה (ציון שהילד היה צריך לתת בסקלה של 1 עד 10).
- גיל התלמיד בחודשים.

8) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין גיל התלמיד לבין רמת החרדה שלו מהמצב?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

9) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין המגדר לבין רמת החרדה שלו מהמצב?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

10) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין המגדר לבין בית הספר?

א. פירסון.
ב. ספירמן.
ג. קרמר.
ד. אין מספיק נתונים כדי לדעת.

© כל הזכויות שמורות לברק קנדל

11) בטבלה שלהלן נתונות שכיחויות ההצלחה והכישלון של 150 חולים:

תוצאה/ התרופה	A	B	C
נרפא	35	13	45
לא נרפא	15	37	5

החולים קיבלו 3 תרופות שונות ובדקו עבור כל חולה אם התרופה הצליחה בריפוי. מהו מדד הקשר המתאים?

א. פירסון.

ב. ספירמן.

ג. קרמר.

ד. אין מספיק נתונים כדי לדעת.

12) שני מוסיקאים מפורסמים נתנו ציון בסולם של 1-10 לקולם של 8 מתמודדים בתוכנית ראליטי ידועה. ציון 10 ניתן לקול שמצא חן ביותר בעיני המוסיקאי. מפיק התוכנית רצה לבדוק האם יש קורלציה בין המוסיקאים מבחינת הטעם. בטבלה הבאה נתונים הציונים של כל אחד מהמוסיקאים את שמונת המתמודדים:

	1	2	3	4	5	6	7	8
מוסיקאי א'	6	5	7	4	3	1	1	4
מוסיקאי ב'	5	7	5	2	3	3	2	7

מהו מדד הקשר המתאים?

א. פירסון.

ב. ספירמן.

ג. קרמר.

ד. אין מספיק נתונים כדי לדעת.

13) להלן טבלה המסכמת את השכר באלפי ₪ של עובדים בחברה ואת רמת המוטיבציה שלהם מ-1 עד 5:

שכר	12	18	20	15	30
מוטיבציה	4	4	5	3	5

מהו מקדם המתאם המתאים לבדיקת רמת ההתאמה בין המוטיבציה לשכר של העובד?

א. פירסון.

ב. ספירמן.

ג. קרמר.

ד. אין מספיק נתונים כדי לדעת.

© כל הזכויות שמורות לברק קנדל

14) להלן טבלה על נתונים שנאספו על מספר תצפיות:

X	1	2	3	4	5
Y	12	14	17	17	20

אם מעוניינים לבדוק עד כמה קיים קשר לנארי בין שני המשתנים.
מהו המדד המתאים?

א. פירסון.

ב. ספירמן.

ג. קרמר.

ד. אין מספיק נתונים כדי לדעת.

תשובות סופיות:

1) א'	**2)** ד'	**3)** ב'	**4)** ג'	**5)** א'
6) ב'	**7)** א'	**8)** ב'	**9)** ג'	**10)** ג'
11) ג'	**12)** ב'	**13)** ב'	**14)** א'	

© כל הזכויות שמורות לברק קנדל