

סטודנטים יקרים

לפניכם ספר תרגילים בקורס ניתוח נתונים וסטטיסטיקה . הספר הוא חלק מקורס חדשני וראשון מסוגו בארץ בנושא זה, המועבר ברשת האינטרנט On-line.

הקורס באתר כולל פתרונות מלאים לספר התרגילים, וכן את התיאוריה הרלוונטית לכל נושא ונושא.

הקורס כולו מוגש בסרטוני וידאו המלווים בהסבר קולי, כך שאתם רואים את התהליכים בצורה מובנית, שיטתית ופשוטה, ממש כפי שנעשה בשיעור פרטי, לדוגמה לחצו כאן.

את הקורס בנה מר ברק קנדל, מרצה מבוקש במוסדות אקדמיים שונים ובעל ניסיון עתיר בהוראת המקצוע.

אז אם אתם עסוקים מידי בעבודה, סובלים מלקויות למידה, רוצים להצטיין או פשוט אוהבים ללמוד בשקט בבית, אנחנו מזמינים אתכם לחוויית לימודים יוצאת דופן וחדשה לחלוטין, היכנסו עכשיו לאתר www.gool.co.il.



אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

גול זה בול. בשבילך!

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

© כתב ופתר - ברק קנדל

תוכן

| | |
|-----|---|
| 3 | פרק 1 - סטטיסטיקה תיאורית - סיווג משתנים וסולמות מדידה |
| 6 | פרק 2 - סטטיסטיקה תיאורית- מדדי מיקום מרכזי |
| 11 | פרק 3 - סטטיסטיקה תיאורית - מדדי פיזור: הטוות, השונות וסטיית התקן |
| 16 | פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן |
| 19 | פרק 5 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים בטבלת שכירות בדידה |
| 22 | פרק 6 - סטטיסטיקה תיאורית - מקדם ההשתנות |
| 25 | פרק 7 - סטטיסטיקה תיאורית - שאלות אמריקאיות |
| 35 | פרק 8 - מדדי קשר - מדד הקשר של קרמר |
| 38 | פרק 9 - מדדי קשר - מדד הקשר פי |
| 40 | פרק 10 - מדדי קשר - מדד הקשר של ספירמן |
| 44 | פרק 11 - מדדי קשר - מדד הקשר הלינארי (פירסון) |
| 52 | פרק 12 - התפלגויות רציפות מיוחדות - התפלגות נורמלית |
| 61 | פרק 13 - הסקה סטטיסטית - הקדמה |
| 64 | פרק 14 - התפלגות הדגימה |
| 70 | פרק 15 - מושגים בסיסיים באמידה |
| 77 | פרק 16 - בדיקת השערות על פרמטרים |
| 82 | פרק 17 - בדיקת השערות על תוחלת (ממוצע) |
| 102 | פרק 18 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים |
| 113 | פרק 19 - בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים) |
| 125 | פרק 20-שאלות מסכמות בבדיקת השערות על פרמטרים |
| 140 | פרק 21- מבחני חי בריבוע |
| 162 | פרק 22 - ניתוח שונות חד כיוונית |

פרק 1 - סטטיסטיקה תיאורית - סיווג משתנים וסולמות מדידה

רקע:

סטטיסטיקה תיאורית הוא ענף בו לומדים כיצד לאסוף נתונים, להציג אותם ולנתח אותם. בסטטיסטיקה תיאורית אנו פונים לקבוצה מסוימת. באותה קבוצה אנו אוספים נתונים על הישויות באותה קבוצה.

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

חלוקה אחת של המשתנים הנמדדים היא לפי סולמות מדידה :

מיון משתנים לפי סולמות המדידה:

1. **סולם שמי** (נומינאלי) – משתנה שלערכיו יש משמעות רק מבחינת הזהות ואין עניין של יותר או פחות לדוגמה: מצב משפחתי רווק/נשוי/אלמן/גרש; אזור מגורים. משתנה דיכוטומי (הינו מסולם שמי) אותם משתנים שיש להם רק שני ערכים אפשריות זכר/נקבה. מעשך/לא מעשך.
2. **סולם סדר** (אורדינאלי) – כאשר לערכים של המשתנה בנוסף לשם ישנה גם משמעות לסדר אבל אין משמעות לגודל ההפרש. למשל, דרגה בצבא.
3. **סולם רווחים** (אינטרוואלי) – משתנה שלערכים שלו בנוסף לשם ולסדר בניהם יש משמעות לרווחים בין הערכים אבל אין משמעות ליחס בין הערכים. למשל, קומה בבניין. סולם לא כל כך פופולרי.
4. **סולם מנה/יחס** – משתנה שלערכיו בנוסף לשם, לסדר ולרווח יש משמעות גם ליחס בין הערכים. למשל, מספר מכוניות למשפחה, משקל אדם בק"ג. הדרך הקלה ביותר כדי לזהות עם הסולם הוא סולם מנה היא על ידי מבחן האפס. בסולם מנה האפס הוא מוחלט, אבסולוטי, ומייצג אין.

סוגי משתנים:

נבצע סיווג של המשתנים :

משתנה איכותי הוא משתנה שלערכיו אין משמעות של יותר או פחות , אין עניין כמותי לערכים המתקבלים.

כמו : מקום מגורים של אדם (רעננה, תל אביב, אשדוד..)

מין האדם (זכר, נקבה)

מצב משפחתי (רווק, נשוי, גרוש, אלמן)

משתנה כמותי הוא משתנה שערכיו הם מספרים להם יש משמעות כמותית כמו : גובה אדם בס"מ, ציון בבחינה וכדומה.

את המשתנה הכמותי נסווג לשני סוגים :

משתנה בדיד : משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים. כמו : מספר ילדים למשפחה (1,2,3..)

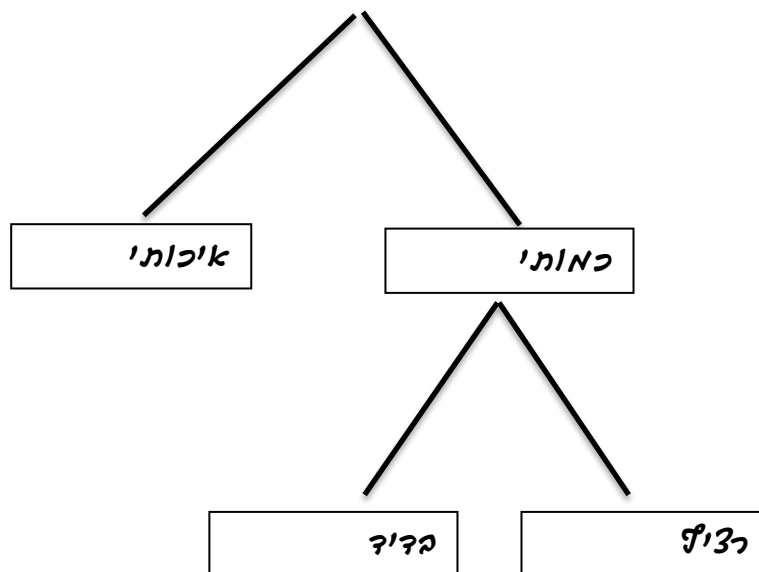
ציון בבחינה (מ 0 ועד 100 בקפיצות של 1)

משתנה רציף : משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים , הערכים מתקבלים ברצף וללא קפיצות של ערכים .

כמו : גובה בס"מ – אם למשל, הגובה הנמוך ביותר הוא 150 ועד 190 ס"מ בקבוצה הגבהים הם

ברצף. גם בין 160 ל 161 ס"מ יש רצף אינסופי של ערכים אפשריים לגובה (16.233 ס"מ הוא גם גובה אפשרי)

משקל בק"ג , מהירות בקמ"ש וכולי.



תרגילים:

1. לפניכם רשימה של משתנים:
- א. גובה אדם בס"מ.
 - ב. מספר ילדים למשפחה.
 - ג. מידת חרדה לפני מבחן.
 - ד. שביעות רצון משירות לקוחות בסקלה מ 1 עד 7 (1 כלל לא מרוצה עד 7 מרוצה מאד)
 - ה. השכלה.
 - ו. מספר אוטובוס.
 - ז. מקום מגורים.
 - ח. מין (1=גבר ו-2=אישה).
 - ט. מידת נעליים.

ציינו באיזה סולם מדידה המשתנה הנחקר (שמי, סדר, רווחים או מנה)

2. להלן התפלגות מספר האיחורים לעבודה בחודש של העובדים בחברת "סטאר".
בחברה 200 עובדים.

| מספר העובדים | מספר האיחורים |
|-----------------|---------------|
| 17 | 0 |
| 23 | 1 |
| 85 | 2 |
| 50 | 3 |
| 25 | 4 |

- א. מהו המשתנה הנחקר כאן?
- ב. האם מדובר במשתנה איכותי או כמותי? אם הוא כמותי האם הוא בדיד או רציף?
באיזה סולם מדידה המשתנה?
3. לפניכם רשימה של משתנים כמותיים. ציין ליד כל משתנה אם הוא רציף או בדיד.
- א. שכר עובד בש"ח.
 - ב. ציון בחינת בגרות.
 - ג. תוצאה בהטלת קובייה.
 - ד. מהירות ריצה בתחרות.
 - ה. שיעור התמיכה בממשלה.

פרק 2 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי

רקע:

המטרה במדדי המיקום המרכזי למדוד את מרכז ההתפלגות של התצפיות.

השכיח – MODE

השכיח הוא הערך הנפוץ ביותר בהתפלגות.

ברשימה : הערך החוזר על עצמו הכי הרבה פעמים .

בטבלת שכיחויות בדידה : הערך שהשכיחות שלו היא הגבוהה ביותר.

בדיאגרמת מקלות : שיעור ה- X של המקל הגבוה ביותר.

בעוגה : הערך של הפלח הגדול ביותר.

בטבלת שכיחויות במחלקות : המחלקה עם הצפיפות הגבוהה ביותר.

בהיסטוגרמה המחלקה הגבוהה ביותר.

יתכן שלהתפלגות יותר משכיח אחד.

השכיח הוא מדד הרלבנטי לכל סוגי המשתנים.

MEDIAN – החציון

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

ברשימה : נסדר את התצפיות בסדר עולה.

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

אם יש מספר זוגי של איברים החציון יהיה הממוצע של האיבר ה- $\frac{n}{2}$ והאיבר ה- $\frac{n}{2} + 1$

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

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

בטבלת שכיחויות בדידה : נעשה תהליך דומה אך נעזר בשכיחות המצטברת.

בטבלת שכיחויות במחלקות : המחלקה החציונית היא המחלקה שמיקומה $\frac{n}{2}$

החציון אינו רלבנטי למשתנה מסולם שמי ולא רלבנטי למשתנה איכותי.

הממוצע:

הנו מרכז הכובד של ההתפלגות.

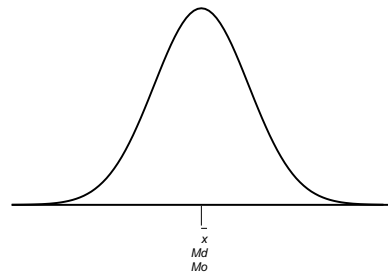
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} : \text{ברשימה}$$

$$\bar{x} = \frac{\sum x \cdot f}{n} : \text{בטבלת שכיחויות}$$

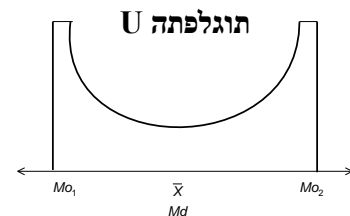
במחלקות: נשתמש באותה נוסחה רק נתייחס לאמצע המחלקה בתור ה X . הממוצע הזה יהיה ממוצע מקורב ולא אמיתי.
הממוצע רלבנטי רק למשתנה כמותי.

מדדי המיקום המרכזי בהתפלגויות המיוחדות:

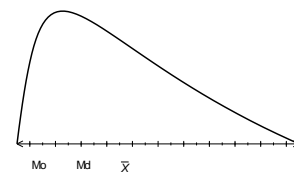
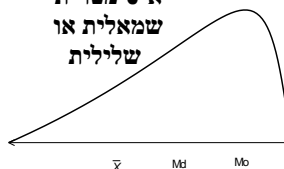
בהתפלגות סימטרית פעמונית כל מדדי המרכז שווים זה לזה:

התפלגות סימטרית

בהתפלגות סימטרית השכיח לא חייב להיות במרכז:



בהתפלגות אסימטרית

**התפלגות א-סימטרית
ימנית או חיובית****התפלגות
א-סימטרית
שמאלית או
שלילית**

תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:

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

חשב את החציון, השכיח, והממוצע של הציונים.

2. בדקו את מספר החדרים לדירה בבניין בן 5 דירות והתקבל ממוצע 3.8

לגבי 4 דירות נמצא מספר חדרים: 4, 3, 4, 5.

א. כמה חדרים יש בדירה החמישית?

ב. מהו השכיח ומהו החציון?

3. להלן התפלגות מספר מקלטי הטלוויזיה שנספרו עבור כל משפחה בישוב מסוים:

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

א. חשב את הממוצע, החציון והשכיח של ההתפלגות.

ב. הסבר ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא

כולן) שלא היה להם עד היום טלוויזיה היו רוכשים מקלט אחד.

4. להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן"

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

א. כמה משפחות יש בישוב?

ב. מה אחוז המשפחות בישוב עם לכל היותר 2 מכוניות?

ג. חשבו את הממוצע, החציון והשכיח.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

5. להלן התפלגות המשקל של קבוצה מסוימת בקייג:

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

המחלקה השכיחה והחציונית.

א. מהי

ב. חשב אומדן לממוצע.

ג. האם היה ניתן לדעת מהי התשובה לסעיף ב ללא חישוב? הסבר

פתרונות:**שאלה 1:**

החציון: 7

השכיח: 6

הממוצע: 6.9

שאלה 2:

א. 3

ב. שכיח: 3,4 חציון: 4

שאלה 3:

א. הממוצע: 1.7

החציון: 1.5

השכיח: 1

ב. הממוצע יגדל ויתר המדדים לא ישתנו.

שאלה 4:

א. 630

ב. 34.13%

ג. שכיח וחציון: 3 ממוצע: 2.952

פרק 3 - סטטיסטיקה תיאורית - מדדי פיזור: הטווח, השונות וסטיית התקן

רקע:

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

הטווחותחום RANGE:

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

שונות וסטיית תקן:

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

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

דוגמה : נחשב את השונות של סדרת המספרים הבאה : 5,4,9

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

להלן התפלגות הציונים בכיתה מסוימת בה ממוצע הציונים הוא 7.44

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

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

$$s = \sqrt{s_x^2} = \sqrt{1.8464} = 1.3588$$

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6
 חשבו את השונות, סטיית התקן והטווח של הציונים.

2. להלן התפלגות מספר המכוניות למשפחה ב"הגורן"

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

א. חשבו סטיית התקן.

ב. חשבו את הטווח של הנתונים.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

3. בחברה העוסקת בטלמרקטינג בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4 שנים וסטיית התקן היא שנתיים.

א. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים עם וותק של 4 שנים להתפלגות?

ב. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני

עובדים אשר אחד עם וותק של 0 שנים והשני עם וותק של 8 שנים להתפלגות?

4. נתונה רשימה של 5 תצפיות, אך רק עבור 4 מהן נרשמו הסטיות שלהן מהממוצע:

2, 3, 2, -1. חשב את השונות של חמש התצפיות.

5. בשכונה בדקו בכל דירה את מספר החדרים לדירה. בשכונה 200 דירות.

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

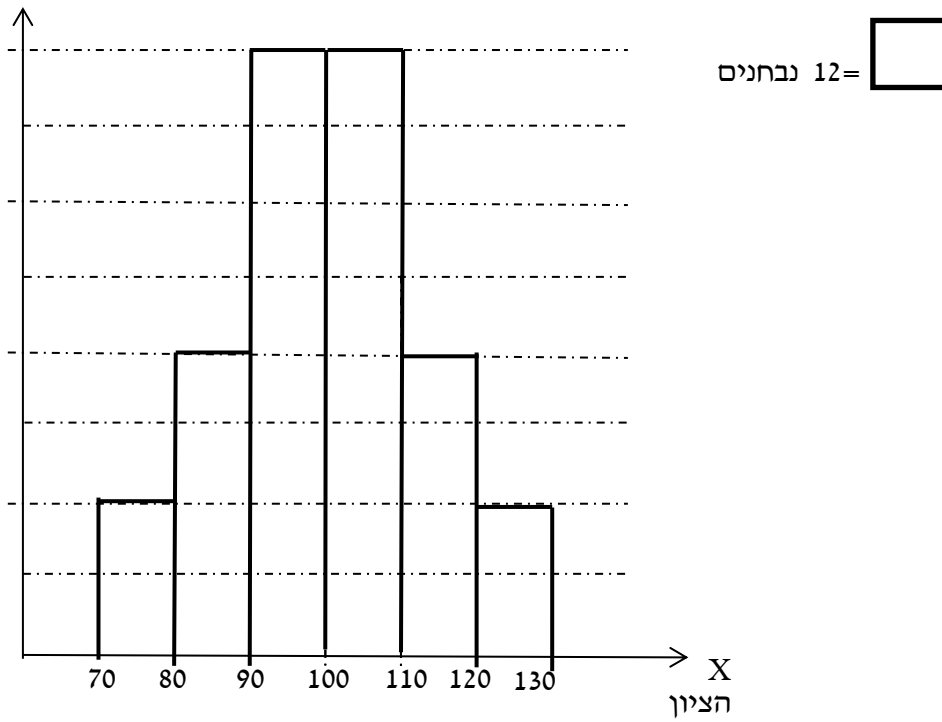
- א. מה הממוצע של מספר החדרים לשכונה בדירה?
 ב. חשבו את סטיית התקן של מספר החדרים לדירה.
 ג. חלק מבעלי הדירות בנות 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע (יקטין, יגדל, לא ישנה) כל מדד שחישבתם בסעיפים הקודמים.

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

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

מהי סטיית התקן של התפלגות המשקל?

7. להלן התפלגות הציונים במבחן אינטליגנציה:



- א. מה הממוצע ומה החציון של ההתפלגות?
- ב. חשבו את סטיית התקן של הציונים.
- ג. מסתבר שיש להוסיף 20 תצפיות לכל אחת משתי המחלקות 90-100 ו-100-110. כיצד הדבר ישתנה את כל אחד מהמדדים של הסעיפים הקודמים?

פתרונות :**שאלה 1:**

השונות : 2.19

סטיית תקן : 1.48

טווח : 6

שאלה 2:

א. סטיית תקן: 1.106

ב. טווח 4

שאלה 3:

א. ממוצע לא ישתנה, סטיית התקן תקטן.

ב. ממוצע לא ישתנה, סטיית התקן תגדל.

שאלה 4:

10.8

שאלה 5:

א. 3.05

ב. 1.16

שאלה 6:

7.73

שאלה 7:

א. 100

ב. 12.96

פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן

רקע:

המטרה למדוד איך תצפית ממוקמות יחסית לשאר התצפיות בהתפלגות.

ציון תקן:

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

ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע.

כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע.

ציון תקן חיובי אומר שהתצפית מעל הממוצע.

ציון תקן שלילי אומר שהתצפית מתחת לממוצע.

ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

דוגמה : (פתרון בהקלטה)

במקום עבודה מסוים ממוצע המשכורות 8 אלפי ₪ עם סטית תקן של 2 אלפי ₪ באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים עם סטית תקן של 1.5 שנים. ערך מרויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערך יותר באופן יחסי משכיל או משתכר ?

תרגילים

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

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

עודד קיבל: 68 בלשון ו70 במתמטיקה.

א. באיזה מקצוע עודד טוב יותר באופן יחסי לשכבה שלו?

ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שיהיה שקול לציונו בלשון?

2. במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום.

להלן טבלה המסכמת את האינפורמציה שנאספה על שני המשתנים:

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

באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים.

מה יותר חריג באותו היום יחסית לשאר הימים שנבדקו נתוני התפוקה או כמות הפועלים? בחר בתשובה הנכונה.

א. התפוקה.

ב. כמות הפועלים.

ג. חריגים באותה מידה.

ד. חסרים נתונים כדי לדעת זאת.

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

פתרונות:

שאלה 1:

א. לשון

ב. 72

שאלה 2:

תשובה ב

שאלה 3:

א. משקל

ב. 70

פרק 5 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים בטבלת שכיחויות בדידה

רקע:

האחוזון (המאון) p הוא הערך בנתונים המחלק את הנתונים בצורה כזאת שעד אליו כולל יש $p\%$ מהנתונים.

מסמנים את האחוזון ה- p ב- X_p .

חישוב האחוזון מתוך נתונים בטבלת שכיחויות בדידה:

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

דוגמה: (פתרון בהקלטה)

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

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

מצא את האחוזון ה-25.

מצא את הערך ש-20% מהמקרים מעליו.

תרגילים:

1. להלן התפלגות של משתנה כלשהו.

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

מצא להתפלגות את:
 האחוזון ה-60.
 המאון ה-40.
 העשירון העליון.
 הטווח בין הרבעונים.

2. להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן"

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

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

פתרונות:

שאלה 2

א. 1

ב. 2

ג. 4

ד. 4

פרק 6 - סטטיסטיקה תיאורית - מקדם ההשתנות

רקע:

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

$$CV = \frac{S(X)}{\bar{X}}$$

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

תרגילים:

1. להלן נתונים לגבי ציונים במבחן באנגלית ב-3 כיתות מתוך שכבה יי בתיכון:

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

א. חשבו את מקדם ההשתנות בכל כיתה.

ב. מהי הכיתה הכי הטרוגנית?

2. נתונות שתי קבוצות:

הממוצע בקבוצה א 100 והשונות 100.

הממוצע בקבוצה ב 500 והשונות 400.

באיזו קבוצה מידת הפיזור יחסית קטן יותר?

3. במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים

במאות) ואת מספר הפועלים שעבדו באותו היום.

להלן טבלה המסכמת את האינפורמציה שנאספה על שני המשתנים:

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

לפי קריטריון CV. בחר את התשובה הנכונה.

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

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

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

ד. אין מספיק נתונים כדי לחשב את CV.

פתרונות:

שאלה 2:

קבוצה 2

שאלה 1:

ב. כתה ב

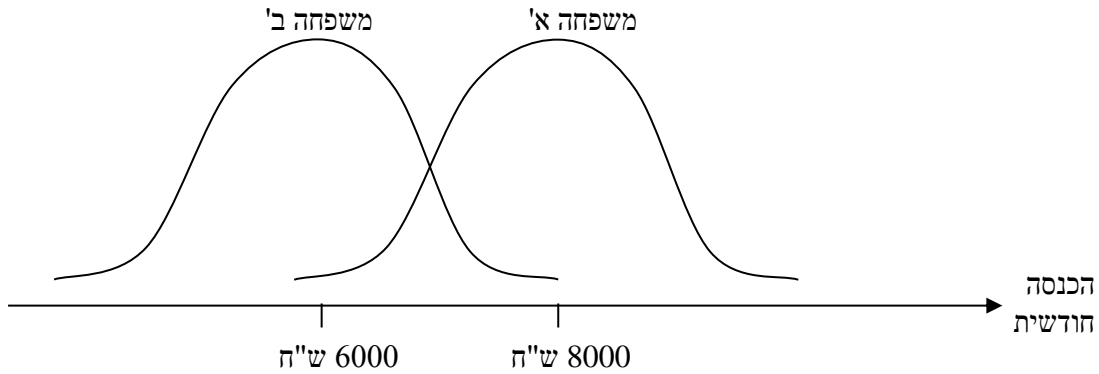
שאלה 3:

תשובה ב

פרק 7 - סטטיסטיקה תיאורית - שאלות אמריקאיות

שאלות 1-3 מתייחסות לקטע הבא:

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



שאלה 1

לאיזו משפחה הכנסה שכיחה גבוהה יותר?

- א. משפחה א'
- ב. משפחה ב'
- ג. לשתיהן אותה הכנסה שכיחה
- ד. לא ניתן לדעת – אין מספיק נתונים

שאלה 2

באיזו משפחה ההכנסה החציונית שווה להכנסה הממוצעת?

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

שאלה 3

באיזו משפחה סטית התקן של ההכנסה החודשית גבוהה יותר?

- א. משפחה א'
- ב. משפחה ב'
- ג. לשתיהן אותה סטית תקן
- ד. לא ניתן לדעת – אין מספיק נתונים

הנתונים הבאים מתייחסים לשאלות 4-6

להלן נתונים חלקיים של טבלת שכירויות:

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

כמו כן נתון הממוצע הוא 1.66

שאלה 4

השכיח של הנתונים הוא:

א. 0

ב. 15

ג. ישנם שני שכחים: 0 ו-3

ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של השכיח.

שאלה 5

חציון הנתונים הוא:

א. 2

ב. 1.5

ג. 25.5

ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

שאלה 6

הטווח של הנתונים

א. 11

ב. 3

ג. 4

ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

שאלה 7

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

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

שאלה 8

סדרת נתונים סטטיסטיים מונה 10 תצפיות. נתון כי סדרת הנתונים סימטרית סביב הממוצע. ממוצע הסדרה - 40 ושונות הסדרה - 100. בשלב מאוחר יותר נוספו שתי תצפיות נוספות לסדרה: 50 ו-30.

השונות של 12 התצפיות היא:

- א. תקטן
- ב. תגדל
- ג. לא תשתנה
- ד. לא ניתן לחשב את השונות ללא ידיעת התצפיות.

הנתונים הבאים מתייחסים לשאלות 9-10

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

שאלה 9

מהי המשכורת הממוצעת החדשה :

א. 2,300 .

ב. 6,900 .

ג. 4,650 .

ד. 4,600 .

ה. חסרים נתונים כדי לדעת.

שאלה 10

מהי סטיית התקן של המשכורת לאחר יישום המו"מ לגבי השכר ?

א. 200

ב. 300

ג. 675

ד. לא ניתן לדעת

שאלה 11

הוספת גודל קבוע לכל תצפיות סדרת נתונים.

א. תגדיל את סטיית התקן.

ב. תקטין את סטיית התקן.

ג. לא תשנה את סטיית התקן.

ד. לא ניתן לדעת.

הנתונים הבאים מתייחסים לשאלות 12-14

להלן נתונים על ציוני תלמידים שנבחנו במועדים שונים בסטטיסטיקה :

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

שאלה 12

התלמיד הטוב ביותר ביחס לנבחנים באותו מועד בו נבחן הוא :

א. מיטב.

ב. צבי.

ג. לובה.

ד. שרית.

ה. סטף.

שאלה 13

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

א. 80.55

ב. 65

ג. 80

ד. 81.66

שאלה 14

איזו כיתה היא ההומוגנית ביותר . הכיתה של :

א. מיטב.

ב. צבי.

ג. לובה.

ד. שרית.

ה. סטף.

הנתונים הבאים מתייחסים לשאלות 15-18

בבדיקת פתע של משרד הבריאות במפעל שוקולד נמצא ש :

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

שאלה 15

מהו החציון של מספר הפגומים בקופסא :

א. 1.

ב. 2.

ג. 4.

ד. לא ניתן לדעת

שאלה 16

מהו הרבעון התחתון של מספר הפגומים בקופסא ?

א. 1

ב. 2

ג. 3

ד. 4

ה. לא ניתן לדעת.

שאלה 17

מספר הפגומים בקופסא הוא משתנה :
א. סדר.

ב. שמי.

ג. כמותי בדיד

ד. כמותי רציף

שאלה 18

השכיח של מספר הפגומים בקופסא :
א. 63

ב. 1

ג. 200

ד. לא ניתן לדעת.

שאלה 19

ביחס לציר המספרים רוב הערכים בהתפלגות א-סימטרית ימנית נמצאים :
א. בערכים הגבוהים.

ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.

ג. בערכים הנמוכים.

ד. לא ניתן לדעת.

ה. אף לא תשובה מהני"ל נכונה.

שאלה 20

בוצע מחקר על מספר העובדים בחברות מזון לעומת חברות תקשורת.
החציון והממוצע בשתיהן שווה 8. איזה מהטענות הבאות היא הנכונה והמלאה ביותר :
א. השכיחות ב 2 החברות זהה אך שונה מ 8.

ב. השכיח ב 2 החברות זהה אך לא ניתן לדעת מהו.

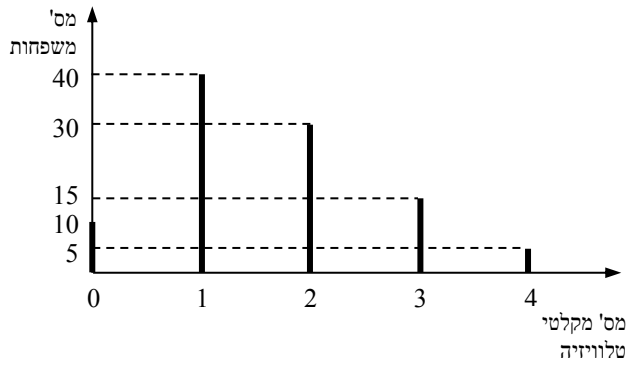
ג. השכיח בשתי חברות הינו בהכרח 8.

ד. שכיח בחברה אחת שונה מ 8 ובשנייה הוא 8.

ה. אף תשובה אינה נכונה.

הנתונים הבאים מתייחסים לשאלות 21 עד 25

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



שאלה 21

- המשתנה הנחקר כאן הוא :
- משתנה שמי.
 - משתנה מסולם סדר.
 - משתנה כמותי בדיד.
 - משתנה כמותי רציף.

שאלה 22

- הטווח של ההתפלגות הוא :
- 35
 - 4
 - 3
 - 2

שאלה 23

- ממוצע מספר מקלטי הטלוויזיה למשפחה הוא :
- 1.65
 - 1.5
 - 1
 - 2

שאלה 24

השכיח של התפלגות זו היא :

א. 40

ב. 1.5

ג. 1

ד. 2

שאלה 25

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

א. יקטין אותו.

ב. יגדיל אותו.

ג. לא ישנה אותו.

ד. אין לדעת

פתרונות:

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

פרק 8 - מדדי קשר - מדד הקשר של קרמר

רקע:

מתני 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/נמנע/נשאלו נושא. להלן טבלת השכיחויות המשותפת שהתקבלה.

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

בהקשר של קרמר הטבלה נקראת טבלת O (observed)

X - מין (גבר/אישה) – סולם שמי

Y - דעה (בעד/נמנע/נגד) – סולם שמי/סדר

שלבים בחישוב r_c :

שלב א': נבנה את טבלת E (Expected)

נעתיק את המסגרת של טבלת O ואז כל $E_i = (f_{(x)} * f_{(y)}) / n$

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

שלב ב': נחשב $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$

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

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

תרגילים:

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

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

האם קיים קשר בין מין להשכלה? נמק!

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. בנה טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.
- ב. האם קיים קשר בין פעילות גופנית למצב בריאותי? חשב לפי מדד הקשר של קרמר.

פתרונות :

שאלה 1

0.595

שאלה 2

ב. 0.19

פרק 9 - מדדי קשר - מדד הקשר פי

רקע:

מדד הקשר פי הינו דרך קיצור על מנת לחשב את מדד הקשר של קרמר. המדד רלבנטי רק כשטבלת השכיחות המשותפת היא מסוג 2/2 כלומר שני משתנים שהם דיכוטומיים.

$$\phi = \sqrt{\frac{(a \cdot d - b \cdot c)^2}{e \cdot f \cdot r \cdot k}} \quad \text{הנוסחה:}$$

| | | |
|---|---|--|
| | | |
| b | a | |
| d | c | |

מפעל עובד בשתי משמרות, משמרת יום ומשמרת ליל, דגמו 300 מוצרים ממשמרת היום ו-200 ממשמרת הלילה, מתוך המוצרים שנדגמו ביום 10 היו פגומים, מתוך המוצרים שנדגמו בלילה 150 היו תקינים. האם יש קשר בין סוג המשמרת לטיב המוצר?

תרגילים :

1. להלן תוצאות מחקר שבדק את הקשר בין מין לדעה מסוימת. לגבי כל נחקר נבדק המין שלו ודעתו האישית בדבר סוגיה מסוימת. הנחקרים היו צריכים לענות האם הם בעד, נמנעים או נגד הדעה שהוצגה להם. להלן התוצאות:

| נגד | נמנע | בעד | דעה |
|-----|------|-----|------|
| | | | מין |
| 20 | 40 | 120 | גבר |
| 80 | 20 | 20 | אישה |

האם אפשר לחשב במקרה זה את מדד הקשר פי? אם כן חשב.

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. האם ניתן לחשב את מדד הקשר של ϕ ? אם כן חשב והסבר את המשמעות.

פרק 10 - מדדי קשר - מדד הקשר של ספירמן

רקע:

מתי נשתמש במדד ספירמן ?

כאשר אחד המשתנים מסולם סדר והשני מסולם סדר ומעלה.

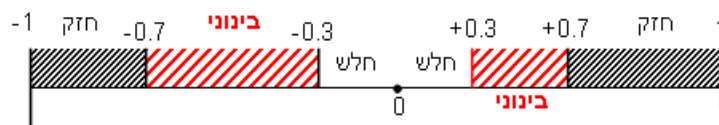
הקשר שהמדד בודק הוא קשר דירוגי.

מדד הקשר בודק :

1 כיוון של הקשר.

2 בודק את עצמת הקשר.

המדד מקבל ערכים בסקלה מ -1 ועד 1.



אם מדד הקשר של ספירמן יוצא 1 המשמעות היא שיש קשר דירוגי חיובי מלא : ככל המשתנה אחד עולה השני עולה ללא יוצא מן הכלל.

קשר דירוגי חיובי חלקי (שמקדם המתאם בין 0 ל-1) אומר שככל שמשתנה אחד עולה לשני יש נטייה לעלות אך לא באופן מוחלט.

אם מדד הקשר של ספירמן יוצא -1 המשמעות היא שיש קשר דירוגי שלילי מלא : ככל שהמשתנה אחד עולה השני יורד ללא יוצא מן הכלל.

קשר דירוגי שלילי חלקי (שמקדם המתאם הוא בין 0 ל- -1) אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אך לא באופן מוחלט.

על מנת לחשב את הקשר יש לבצע פעולת דירוג (RANK) נלמד את פעולת הדירוג דרך הדוגמה הבאה (פתרון בהקלטה)

| שם התלמיד | הערכה | דירוג R |
|-----------|--------|---------|
| ערן | בינוני | |
| מיכל | מצוין | |
| עודד | חלש | |
| רוני | טוב | |
| יעל | טוב | |

כאשר מדרגים אם יש כמה תצפיות שתופסות את אותו הערך אז הדירוג שלהם הוא הממוצע של המקומות שהן תופסות.

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

הנוסחה של מדד הקשר :

דוגמה : (פתרון בהקלטה)

בתחרות רוקדים עם כוכבים השתתפו 7 זוגות, 2 שופטים נתנו את ציוניהם לריקוד של כל זוג.
מהי מידת ההתאמה בין ציוני השופטים?

X - ציון שופט א (סולם סדר)

Y - ציון שופט ב (סולם סדר)

להלן התוצאות שהתקבלו :

| מספר הזוג | ציון שופט א | R_x | ציון שופט ב | R_y | $d=r_x-r_y$ | d^2 |
|-----------|-------------|-------|-------------|-------|-------------|-------|
| 1 | 4 | | 5 | | | |
| 2 | 5 | | 5 | | | |
| 3 | 6 | | 7 | | | |
| 4 | 5 | | 7 | | | |
| 5 | 8 | | 9 | | | |
| 6 | 7 | | 9 | | | |
| 7 | 3 | | 7 | | | |

תרגילים:

1. בתחרות יופי חילקו שני שופטים ציונים למועמדות:

| | | | | | | | |
|---|---|---|---|---|---|---|--------------|
| 7 | 6 | 5 | 4 | 3 | 2 | 1 | מספר מועמדת |
| 6 | 5 | 9 | 8 | 6 | 8 | 7 | ציון שופט א' |
| 7 | 5 | 9 | 8 | 7 | 8 | 8 | ציון שופט ב' |

האם קיים קשר בין שתי הערכות השופטים? נמק והסבר!

2. משרד רצה לבחון האם קיים קשר בין מידת המוטיבציה של העובדים שלו לבין מספר

החיסורים של העובדים בחודש עבודה. להלן התוצאות שהתקבלו:

| מספר חיסורים | מידת מוטיבציה |
|--------------|---------------|
| 0 | גבוהה |
| 4 | נמוכה |
| 2 | בינונית |
| 5 | נמוכה |
| 1 | גבוהה |

האם קיים קשר בין רמת המוטיבציה של העובד ומספר החיסורים שלו? חשב באמצעות מדד

הקשר המתאים והסבר.

3. אם $r_s = 1$ הדבר אומר שערכי X תמיד שווים לערכי Y. האם הטענה נכונה? הסבר.

פתרונות:**שאלה 1:**

0.973

שאלה 2:

-0.85

שאלה 3:

לא נכון

פרק 11 - מדדי קשר - מדד הקשר הלינארי (פירסון)

רקע:

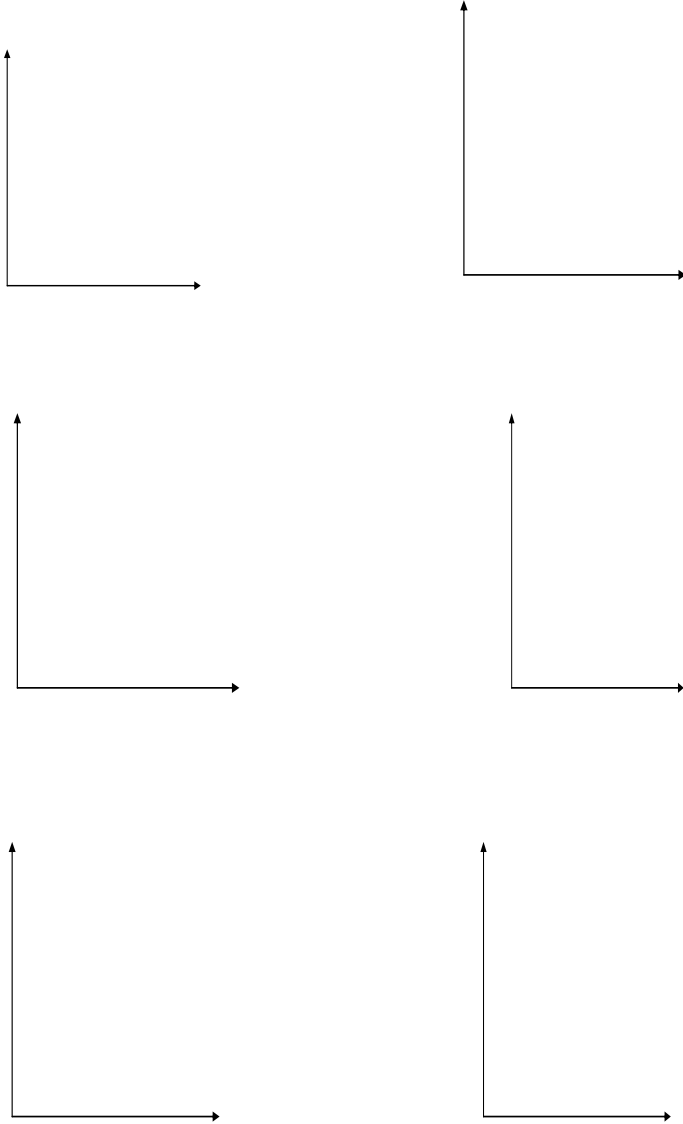
המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו Y הוא המשתנה המוסבר (התלוי). למשל, נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד X- מסבירה את ההכנסה שלו Y. במקרה זה שנות ההשכלה זהו המשתנה המסביר (או הבלתי תלוי) ואנחנו מעוניינים לבדוק כיצד שינויים בשנות ההשכלה של אדם יכולים להסביר את השינויים שלו בהכנסה, ולכן רמת ההכנסה זהו המשתנה המוסבר התלוי במשתנה המסביר אותו. בשלב הראשון, נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים. למשל, בבניין של 5 דירות בדקו את הנתונים הבאים: X - מס' חדרים בדירה. Y - מס' נפשות הגרות בדירה. להלן התוצאות שהתקבלו:

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

נשרטט מנתונים הללו דיאגרמת פיזור:



נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן :



בשלב השני, מחשבים את מקדם המתאם (מדד הקשר) שבודק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שנראה בשלב הראשון רק בעין.

המדד בודק את כיוון הקשר (חיובי או שלילי).

ואת עוצמת הקשר (חלש עד חזק).

מקדם מתאם זה מקבל ערכים בין -1 ל 1.

מקדם מתאם -1 או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי

$$y = bx + a$$

מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר לינארי מלא בו השיפוע b יהיה חיובי ואילו

מתאם שלילי מלא אומר שקיים קשר לינארי מלא בו השיפוע b שלילי (מקדם מתאם -1).

מתאם חיובי חלקי אומר שככל שמשנתה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת

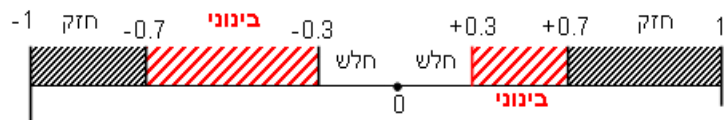
נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל

שמשנתה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y

באופן מוחלט.

ככל שערך מקדם המתאם קרוב לאפס נאמר שעוצמת הקשר חלשה יותר וככל שמקדם המתאם

רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר.



מקדם המתאם יסומן באות r .

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

$$COV(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} = \frac{\sum xy}{n} - \bar{x} \cdot \bar{y} : \text{שונות משותפת}$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 : \text{שונות של המשתנה X}$$

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2 : \text{שונות המשתנה Y}$$

$$r_{xy} = \frac{COV(x, y)}{s_x \cdot s_y} : \text{מקדם המתאם הלינארי}$$

תרגילים:

1. להלן נתונים לגבי שישה תלמידים שנגשו למבחן. בדקו לגבי כל תלמיד את הציון שלו בסוף הקורס וכמו כן את מספר החיסורים שלו מהקורס.

| מספר חיסורים | ציון |
|--------------|------|
| 2 | 80 |
| 1 | 90 |
| 0 | 90 |
| 2 | 70 |
| 3 | 70 |
| 4 | 50 |

- א. שרטט דיאגרמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר בין מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?
- ב. חשב את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?
- ג. הסבר ללא חישוב כיצד מקדם המתאם היה משתנה אם היה מתווסף תלמיד שהחסיר 4 פעמים וקיבל ציון 80?

2. במחקר רפואי רצו לבדוק האם קיים קשר בין רמת ההורמון X בדם החולה לרמת ההורמון Y שלו. לצורך כך מדדו את רמת ההורמונים ההלו עבור חמישה חולים. להלן התוצאות שהתקבלו:

| x | y |
|----|----|
| 10 | 12 |
| 14 | 15 |
| 15 | 15 |
| 18 | 17 |
| 20 | 21 |

- א. מה הממוצע של כל רמת הורמון?
- ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

3. נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

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

$$\sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 76 \quad \sum_{i=1}^{20} (X_i - \bar{X})^2 = 76$$

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

א. חשב את מדד הקשר הלינארי בין X ל- Y . מיהו המשתנה התלוי?
 ב. מה המשמעות של התוצאה שקיבלת בסעיף א?

4. נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

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

$$\sum_{i=1}^{20} Y_i^2 = 2080 \quad \sum_{i=1}^{20} X_i^2 = 2960$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

חשב את מדד הקשר הלינארי בין X ל- Y .

5. במוסד אקדמי ציון ההתאמה מחושב כך: מכפילים את הציון הממוצע בבגרות ב-3 ומפחיתים 2 נקודות. ידוע שעבור 40 מועמדים סטיית התקן של ממוצע הציון בבגרות הייתה 2. מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם?

6. להלן רשימת טענות, לגבי כל טענה קבע נכון/לא נכון ונמק:

- א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.
- ב. לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$ $S_x = S_y = 1$ לכן מדד הקשר של פירסון יהיה 1.
- ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות אמריקאיות:

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

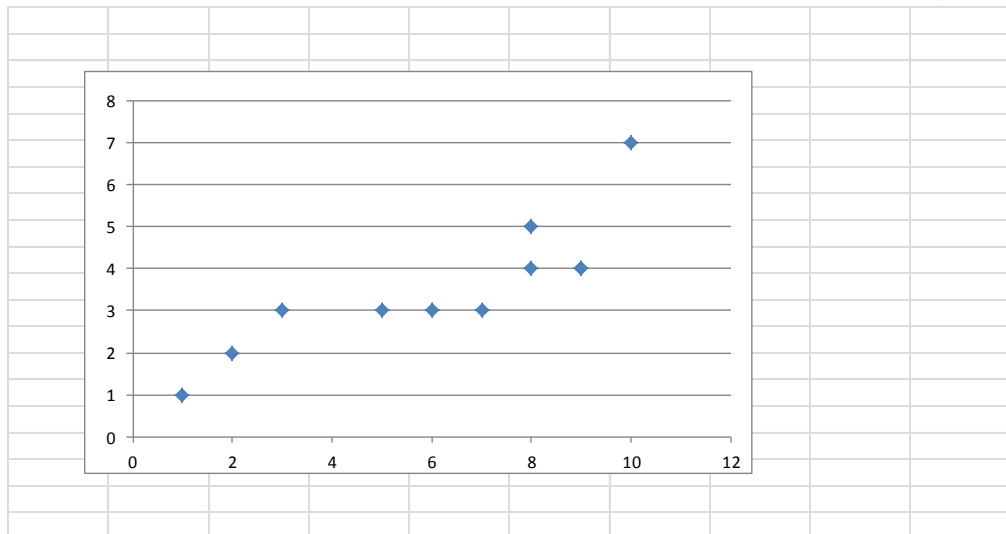
7. נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן :

- א. הדבר מעיד שהציונים בכתה היו שליליים.
- ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
- ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
- ד. אף אחת מהתשובות לא נכונה.

8. נלקחו 20 מוצרים וניבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו היום ערך הדולר היה - 4.2 ש"ח) מהו מקדם המתאם בין המחיר בדולר למחיר בש"ח?

- א. 1
- ב. 0
- ג. 4.2
- ד. לא ניתן לדעת.

9. להלן דיאגרמת פיזור :



מה יהיה מקדם המתאם בין שני המשתנים?

- א. 1
- ב. 0.85
- ג. 0.15
- ד. 0

פתרונות:**שאלה 1:**

א. בהקלטה

ב. -0.9325 **שאלה 2:**א. $\bar{y} = 16$ $\bar{x} = 15.4$ ב. $r_{xy} = 0.96$ **שאלה 3:**

א : 0.8

שאלה 4:

0.8

שאלה 5:

1

שאלה 6:

א. נכון

ב. לא נכון

ג. נכון

שאלה 7:

התשובה : ג

שאלה 8:

התשובה : א

שאלה 9:

התשובה : ב

פרק 12 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

רקע:

התפלגות נורמלית הינה התפלגות של משתנה רציף. ישנם משתנים רציפים מסוימים שנהוג להתייחס אליהם כנורמליים כמו: זמן ייצור, משקל תינוק ביום היולדו ועוד. פונקציית הצפיפות של ההתפלגות הנורמלית נראית כמו פעמון:



לעקומה זו קוראים גם עקומת גאוס ועקומה אחת נבדלת מהשנייה באמצעות הממוצע וסטיית התקן שלה. אלה הם הפרמטרים שמאפיינים את ההתפלגות.

$$X \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} : \text{נוסחת פונקציית הצפיפות}$$

כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלבנטיים שמתחת לעקומה. כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון.

התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת והיא תסומן באות Z .

$$Z \sim N(0, 1^2)$$

תהליך התקנון מבוצע על ידי הנוסחה הבאה:

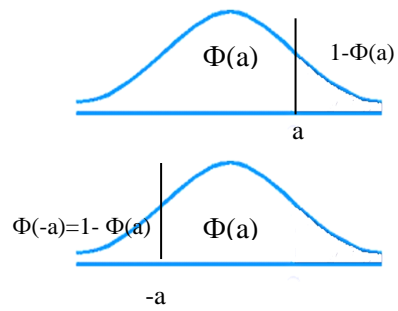
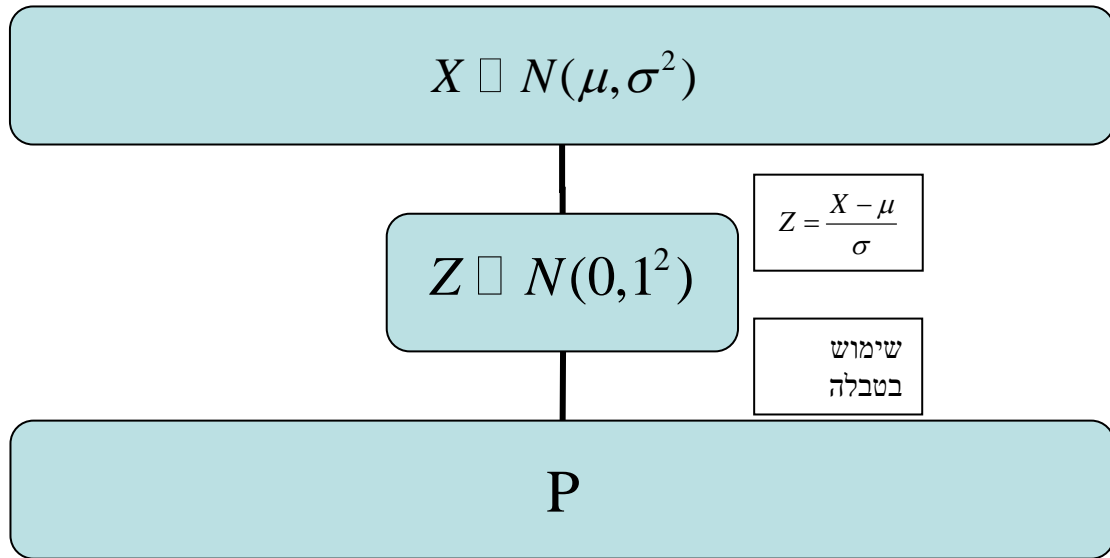
$$Z = \frac{X - \mu}{\sigma}$$

אחרי תקנון מקבלים ערך הנקרא ציון תקן.

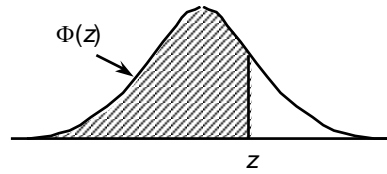
ציון התקן משמעו בכמה סטיות תקן הערך סוטה מהממוצע.

לאחר חישוב ציון התקן של ערך מסוים נעזרים בטבלה של ההתפלגות הנורמלית הסטנדרטית לחישוב השטח הרצוי.

ובאופן כללי נתאר את הסכמה הבאה :



טבלת ההתפלגות המצטברת הנורמלית סטנדרטית – ערכי $\Phi(z)$



| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |
| 2.1 | .9821 | .9826 | .9830 | .9834 | .9838 | .9842 | .9846 | .9850 | .9854 | .9857 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9875 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |
| 2.4 | .9918 | .9920 | .9922 | .9925 | .9927 | .9929 | .9931 | .9932 | .9934 | .9936 |
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3.0 | .9987 | .9987 | .9987 | .9988 | .9988 | .9989 | .9989 | .9989 | .9990 | .9990 |
| 3.1 | .9990 | .9991 | .9991 | .9991 | .9992 | .9992 | .9992 | .9992 | .9993 | .9993 |
| 3.2 | .9993 | .9993 | .9994 | .9994 | .9994 | .9994 | .9994 | .9995 | .9995 | .9995 |
| 3.3 | .9995 | .9995 | .9995 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9997 |
| 3.4 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9998 |

| z | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 | 3.891 | 4.417 |
|-----------|-------|-------|-------|-------|-------|-------|--------|---------|----------|
| $\Phi(z)$ | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 | 0.99995 | 0.999995 |

דוגמה: (הפתרון בהקלטה)

משקל חפיסות שוקולד המיוצרות בחברה מתפלג נורמלית עם ממוצע 100 גרם בסטיית תקן של 8 גרם.

- א. מה אחוז חפיסות השוקולד ששוקלות מתחת ל- 110 גרם?
- ב. מה אחוז חפיסות השוקולד ששוקלות מעל 110 גרם?
- ג. מה אחוז חפיסות השוקולד ששוקלות מתחת ל 92 גרם?
- ד. מהו המשקל ש-90% מהחפיסות בקו הייצור שוקלים פחות מהם?

תרגילים:

1. הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטית תקן של 10 ס"מ.

- א. מה אחוז האנשים שגובהם מתחת ל- 182.4 ס"מ?
- ב. מה אחוז האנשים שגובהם מעל 190 ס"מ?
- ג. מה אחוז האנשים שגובהם בדיוק 173.6 ס"מ?
- ד. מה אחוז האנשים שגובהם מתחת ל- 170 ס"מ?
- ה. מה אחוז האנשים שגובהם לכל היותר 170 ס"מ?

2. נתון שהזמן שלוקח לתרופה מסוימת להשפיע מתפלג נורמלית עם ממוצע של 30 דקות ושונות של 9 דקות רבועות .

- א. מהי פרופורציית המקרים בהן התרופה תעזור אחרי יותר משעה?
- ב. מה אחוז מהמקרים שבהן התרופה תעזור בין 35 ל-37 דקות?
- ג. מה הסיכוי שהתרופה תעזור בדיוק תוך 36 דקות?
- ד. מה שיעור המקרים שבהן ההשפעה של התרופה תסטה מ-30 דקות בפחות מ-3 דקות?

3. המשקל של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 60 ק"ג וסטית תקן של 8 ק"ג .

- א. מה אחוז האנשים שמשקלם נמוך מ- 55 ק"ג?
- ב. מהי פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
- ג. מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל- 70 ק"ג?
- ד. לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ- 4 ק"ג?
- ה. מה הסיכוי שאדם אקראי ישקול מתחת ל – 140 ק"ג?

4. משקל תינוקות ביום היוולדם מתפלג נורמלית עם ממוצע של 3300 גרם וסטית תקן 400 גרם .

- א. מצאו את העשירון העליון.
- ב. מצאו את האחוזון 95.
- ג. מצאו את העשירון התחתון.

5. ציוני מבחן אינטליגנציה מתפלג נורמלית עם ממוצע 100 ושונויות 225 .
- מה העשירון העליון של הציונים במבחן האינטליגנציה?
 - מה העשירון התחתון של ההתפלגות?
 - מהו הציון ש- 20% מהנבחנים מקבלים מעליו?
 - מהו האחוזון ה- 20?
 - מהו הציון ש- 5% מהנבחנים מקבלים מתחתיו?
6. נפח משקה בבקבוק מתפלג נורמלית עם סטיית תקן של 20 מ"ל, נתון ש-33% מהבקבוקים הם עם נפח שעולה על 508.8 מ"ל.
- מה ממוצע נפח משקה בבקבוק ?
 - 5% מהבקבוקים המיוצרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאיזה נפח שולחים בבקבוק לבדיקה?
 - 1% מהבקבוקים עם הנפח הקטן ביותר נתרמים לצדקה, מהו הנפח המקסימלי לצדקה?
7. אורך חיים של מכשיר מתפלג נורמלית . ידוע שמחצית מהמכשירים חיים פחות מ- 500 שעות, כמו כן ידוע ש- 67% מהמכשירים חיים פחות מ- 544 שעות.
- מהו ממוצע אורך חיי מכשיר?
 - מהי סטיית בתקן של אורך חיי מכשיר?
 - מה הסיכוי שמכשיר אקראי יחיה פחות מ- 460 שעות?
 - מהו המאון העליון של אורח חיי מכשיר?
 - 1% מהמכשירים בעלי אורך החיים הקצר ביותר נשלח למעבדה לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשליחת מכשיר למעבדה?

8. להלן שלוש התפלגויות נורמליות של שלוש קבוצות שונות ששורטטו באותה מערכת צירים. ההתפלגויות מוספרו כדי להבדיל ביניהן.



- א. לאיזו התפלגות הממוצע הגבוה ביותר?
 ב. במה מבין המדדים הבאים התפלגות 1 ו 2 זהות?
 א. בעשירון העליון.
 ב. בממוצע.
 ג. בשונות.
- ג. לאיזו התפלגות סטיית התקן הקטנה ביותר?

א. 1

ב. 2

ג. 3

ד. אין לדעת.

9. הזמן שלוקח לאדם להגיע לעבודתו מתפלג נורמלית עם ממוצע של 40 דקות וסטית תקן של 5 דקות.
- א. מה ההסתברות שמשך הנסיעה של האדם לעבודתו יהיה לפחות שלושת רבעי השעה?
 ב. אדם יצא לעבודתו בשעה 08:10 מביתו. הוא צריך להגיע לעבודתו בשעה 09:00. מה הסיכוי שיאחר לעבודתו?
 ג. אם ידוע שזמן נסיעתו לעבודה היה יותר משלושת רבעי השעה. מה ההסתברות שזמן הנסיעה הכולל יהיה פחות מ- 50 דקות?
 ד. מה הסיכוי שבשבוע (חמישה ימי עבודה) בדיוק פעם אחת יהיה זמן הנסיעה לפחות שלושת רבעי השעה?

10. ההוצאה החודשית לבית אב בעיר "טרירה" מתפלגת נורמלית עם ממוצע של 2000 דולר וסטית תקן של 300 דולר. בחרו באקראי 5 בתי אב. ההסתברות שלפחות אחד מהם מוציא בחודש מעל ל-T דולר היא 0.98976.

א. מה ערכו של T?

ב. מה הסיכוי שההוצאה החודשית של בית אב בעיר תהיה לפחות סטיית תקן אחת מעל T?

ג. מסתבר שנפלה טעות בנתונים, ויש להוסיף 100 דולר להוצאות החודשית של כל בתי האב בעיר. לאור זאת, מה ההסתברות שההוצאה החודשית של בית אב נמוכה מ-1800 דולר?

11. אורך שיר אקראי המשודר ברדיו מתפלג נורמלית עם תוחלת של 3.5 דקות וסטיית תקן של שלושים שניות.

א. מה ההסתברות שאורך של שיר אקראי המנוגן ברדיו יהיה בין 3 ל 2.5 דקות?

ב. מהו הטווח הבין רבעוני של אורך שיר המשודר ברדיו?

ג. ביום מסוים מנוגנים 200 שירים ברדיו. כמה שירים מתוכם תצפה שיהיו באורך הנמוך מ 3.5 דקות?

ד. בשעה מסוימת שודרו 8 שירים. מה ההסתברות שרבע מהם בדיוק היו ארוכים מ-4 דקות והיתר לא?

פתרונות :

| <u>שאלה 3</u> | <u>שאלה 1</u> |
|---------------|---------------|
| א. 26.43% | א. 89.25% |
| ב. 89.44% | ב. 2.28% |
| ג. 39.44% | ג. 0 |
| ד. 0.383 | ד. 50% |
| ה. 100% | |

| <u>שאלה 7</u> | <u>שאלה 5</u> |
|---------------|---------------|
| א. 500 | א. 119.2 |
| ב. 100 | ב. 80.8 |
| ג. 0.3446 | ג. 112.6 |
| ד. 733 | ד. 87.4 |
| ה. 267 | |

| <u>שאלה 9</u> | <u>שאלה 8</u> |
|---------------|---------------|
| א. 0.1587 | א. 3 |
| ב. 0.0228 | ב. בממוצע. |
| ג. 0.8563 | ג. 1 |
| ד. 0.3975 | |

| <u>שאלה 11</u> | <u>שאלה 10</u> |
|----------------|----------------|
| א. 0.1359 | א. 1925 |
| ב. 0.675 | ב. 0.2266 |
| ג. 100 | ג. 0.1587 |
| ד. 0.25 | |

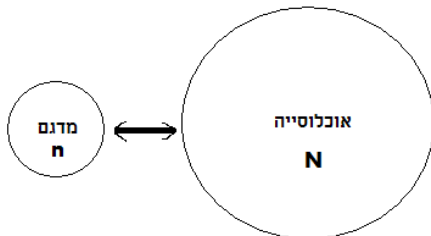
פרק 13 - הסקה סטטיסטית - הקדמה

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 תהייה דגימה מקרית הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכוי להיכלל במדגם.

סטטיסטי – גודל המחושב על המדגם.

פרמטר – גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים

למשל:

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

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

דוגמה (פתרון בהקלטה):

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

תרגילים :

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 מאוכלוסיה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.

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

א. מיהי האוכלוסייה ומהו המשתנה הנחקר?

ב. מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

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?

פרק 14 - התפלגות הדגימה

ממוצע המדגם ומשפט הגבול המרכזי

רקע:

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

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

גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים. להלן רשימה של פרמטרים החשובים לפרק זה:

ממוצע האוכלוסייה נסמן ב μ (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב- σ^2 .

סטיית תקן של אוכלוסייה: σ .

א. תכונות התפלגות

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה:

$$E(\bar{x}) = \mu_{\bar{x}} = \mu$$

שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב- n . תכונה זו נכונה רק במדגם מקרי:

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

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

$$\sigma(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

דוגמה: (פתרון בהקלטה)

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

א. מי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

ב. דגימה מהתפלגות נורמאלית

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע μ ושונות σ^2 ממוצע המדגם גם יתפלג נורמאלית:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

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

דוגמה: (פתרון בהקלטה)

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

ג. משפט הגבול המרכזי

אם אוכלוסייה מתפלגת כלשהו עם ממוצע μ ושונות σ^2 אזי עבור מדגם מספיק גדול ($n \geq 30$)

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

דוגמה: (פתרון בהקלטה)

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

תרגילים :

1. מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
 - מהו תוחלת ממוצע המדגם?
 - מהי טעות התקן?
4. משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
- מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?
- נתון כי ביום מסוים נולדו 4 תינוקות.
- מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
 - הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?
5. הגובה של המתגייסים לצה"ל מתפלג נורמאלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.
- מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?
 - מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?
 - מה ההסתברות שהגובה הממוצע שלהם יסטה מתחולת הגבהים בפחות מ-5 ס"מ?
 - מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?
6. הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך פתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.
- מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
 - מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?
 - מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
 - כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?

7. נפח היין בבקבוק מתפלג נורמאלית עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.
 א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ"ק?
 ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ"ק?
 ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ"ק?
 ד. בקבוקי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?

8. משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4.
 א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
 ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
 ג. הסבר את ההבדל בתשובות של שני הסעיפים.

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

12. אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ.
 א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל 78 ס"מ?
 ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש 100 המוטות יספיקו למלאכה?
 ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. העזר במשפט הגבול המרכזי.

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

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

(בחר בתשובה הנכונה)

א. μ ו- \bar{X} יהיו משתנים מקריים.

ב. μ יהיה משתנה מקרי ו \bar{X} קבוע.

ג. \bar{X} יהיה משתנה מקרי ו μ קבוע.

ד. μ ו \bar{X} יהיו קבועים.

פתרונות:**שאלה 4**

א. 0.8413

ב. 0.0013

ג. 0

ד. 0.1974

שאלה 6

א. 0.0465

ב. 27.71

ג. 0.2628

שאלה 7

א. 0

ב. 0.1587

ג. 0.1587

ד. 0.5

שאלה 8

א. 0.5468

ב. 0.6826

שאלה 10

0.0475

שאלה 12

א. 0.9772

ב. 0.0228

ג. 271

שאלה 16

התשובה ג

פרק 15 - מושגים בסיסיים 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: $|\hat{\theta} - \theta|$ - ההפרש בין האומד לאמת(הפרמטר).

דוגמה: (פתרון בהקלטה)

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

מה הפרמטר בדוגמה זו?

מהי טעות האמידה של ערוץ 10?

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

• טעות התקן של אומד היא סטיית התקן שלו, כלומר: $\sigma(\hat{\theta}) = S.E$

להלן 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 הטויה ל μ .

כמו כן טעות תקן: $SE = \frac{\sigma}{\sqrt{n}} = \sigma(\bar{x})$

פרופורציה באוכלוסייה: p

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

$E(\hat{p}) = p$ לכן \hat{p} הינו אומר חסר הטויה 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 הטויה 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:

2,1,3,2,1,4,5,2,1,3

אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים:

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

תרגילים:

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 האומדן 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 האומדן?
- 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. מצא אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
- ב. מצא אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
- ג. מצא אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.
4. נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים. להלן התוצאות שהתקבלו:
- $$\sum_{i=1}^{20} X_i^2 = 1502.2 \quad \sum_{i=1}^{20} X_i = 162$$
- 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 וחישובו את הממוצע שלהם. מהי טעות התקן?
 א. סטיית התקן של האוכלוסייה.
 ב. סטיית התקן של ממוצע האוכלוסייה.
 ג. סטיית התקן של המדגם.
 ד. סטיית התקן של ממוצע המדגם.

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

א. 3

ב. 2.5

ג. 1.581

ד. אין מספיק נתונים לדעת.

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

א. כאשר n קטן.

ב. כאשר תצפיות המדגם אינן בלתי תלויות.

ג. כאשר האוכלוסייה אינה מתפלגת נורמאלית.

ד. כאשר מעוניינים באומדן חסר הטיה לשונות האוכלוסייה ממנה הוצא המדגם.

ה. כאשר מעוניינים לחשב את שונות התפלגות הדגימה של ממוצע המדגם.

8. X_1, X_2, \dots, X_{16} מדגם מקרי מתוך אוכלוסייה בעלת ממוצע μ לא ידוע ושונות

$\sigma^2 = 64$. טעות התקן של האומדן 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.

פתרונות:**שאלה 3**

א. 177.9

ב. 64.1

ג. 0.4

שאלה 4

א. 8.1

ב. 3.16

שאלה 5

התשובה היא ד.

שאלה 6

התשובה היא ג.

שאלה 7

התשובה היא ד.

שאלה 8

התשובה היא ד.

שאלה 9

התשובה היא ג.

פרק 16 - בדיקת השערות על פרמטרים

הקדמה

רקע:

תהליך של בדיקת השערות הוא תהליך מאוד 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 שניקרא **כלל הכרעה**: הכלל יוצר אזור שניקרא **אזור דחייה** (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) ו**אזור קבלה** (קבלה של השערת האפס ודחייה של האלטרנטיבה). כלל ההכרעה מתבסס על איזשהו סטטיסטי.

אזור הדחייה מוכתב על ידי סיכון שלוקח החוקר מראש שניקרא רמת מובהקות ומסומן ב- α .

שלב ה:

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

שלב ו:

להסיק מסקנה בהתאם לתוצאות המדגם.

דוגמה: (פתרון בהקלטה)

משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום 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:

$$n = 20$$

$$\bar{X} = 3120$$

$$S = 280$$

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

תרגילים:

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

כתב ופתר - ברק קנדל ©

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.

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

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

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.

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

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.

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

טעויות בבדיקת השערות

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :

הכלל יוצר אזור שניקרא אזור דחייה (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) ואזור 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. נשנה את כלל ההכרעה אנחנו יכולים לקבל מסקנה אחרת . נבצע מדגם חדש אנחנו עלולים לקבל תוצאה אחרת.

לכן יתכנו טעויות במסקנות שלנו :

| | | הכרעה | |
|--------|----|-------------|-------------|
| | | H0 | H1 |
| מציאות | H0 | אין טעות | טעות מסוג 1 |
| | H1 | 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?

תרגילים:

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?

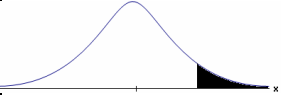


2. במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
 - א. האם ניתן לדעת אם בוצע טעות במחקר?
 - ב. מה סוג הטעות האפשרית?

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?

פרק 17 - בדיקת השערות על תוחלת (ממוצע)

כאשר שונות האוכלוסיה ידועה

רקע:

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

סטטיסטי המבחן :

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

חלופה אחרת לכלל הכרעה :

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

דוגמה : (פתרון בהקלטה)

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. בדוק את ההשערה ברמת מובהקות של 1%.

תרגילים:

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 תהייה 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 ושימוש 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? בחר בתשובה הנכונה.
- הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
 - הגדלת רמת המובהקות תשנה את מסקנת המחקר.
 - הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
 - הקטנת רמת המובהקות תשנה את מסקנת המחקר.
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: (בחר בתשובה הנכונה)
- השערת האפס הייתה נדחית.
 - השערת האפס הייתה לא נדחית.
 - 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.
- אם חוקר א' החליט לדחות את H_0 , מה יחליט חוקר ב'? נמקו.
 - אם חוקר א' יחליט לא לדחות את H_0 , מה יחליט חוקר ב'? נמקו.

פתרונות :**שאלה 1:** H_0 נקבל**שאלה 2:** H_0 נדחה**שאלה 3:** H_0 נדחה**שאלה 4:** H_0 נדחה**שאלה 5:** H_0 נקבל**שאלה 6:**

א

שאלה 7:

ג

שאלה 8:

א. אותה מסקנה

ב. לא ניתן לדעת.

מובהקות התוצאה (p-value) בבדיקת השערות על תוחלת עם שונות ידועה

רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה:
 באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v .
 את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות.
 המסקנה של המחקר תקבע לפי העיקרון הבא:

$$\text{אם } p_v \leq \alpha \text{ דוחים את } H_0$$

מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

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

אם ההשערה היא דו צדדית:

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

| $H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$ | $H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$ | $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$ | השערת האפס: השערה אלטרנטיבה: |
|--|--|--|---------------------------------|
| 3. σ ידועה | | | תנאים: |
| 4. $X \square N$ או מדגם מספיק גדול | | | |
| $P_{H_0}(\bar{X} \geq \bar{x})$ | $P_{H_0}(\bar{X} \leq \bar{x})$ | אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \leftarrow \bar{x} < \mu_0$ | p-value |

כאשר בהנחת השערת האפס: $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$

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

דוגמה: (פתרון בהקלטה)

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

א. מהי מובהקות התוצאה?

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

תרגילים:

1. לפניך השערות של מחקר :

$$H_0 : \mu = 70$$

$$H_1 : \mu > 70$$

המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות :

$$n = 100$$

$$\bar{x} = 74$$

מהי מובהקות התוצאה?

2. השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪. מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?

3. אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ – 100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטית התקן של משקל החטיף היא 12 גרם. האדם מתכוון לשקול 100 חפיסות חטיפים ולאחר מכן להגיע להחלטה. לאחר הבדיקה הוא קיבל משקל הממוצע של 98.5 גרם.

א. רשמו את השערות המחקר.

ב. מהי רמת המובהקות המינימלית עבורה דוחים את השערת האפס?

ג. מהי רמת המובהקות המקסימלית עבורה נקבל את השערת האפס?

ד. מה המסקנה ברמת מובהקות של 5%?

4. מכונה לחיתוך מוטות במפעל חותכת מוטות באורך שמתפלג נורמלית עם תוחלת אליה כוונה המכונה וסטיית תקן 2 ס"מ. ביום מסוים כוונה המכונה לחתוך מוטות באורך 80 ס"מ. אחראי האיכות מעוניין לבדוק האם המכונה מכוילת. לצורך כך נדגמו מקו הייצור 16 מוטות שנחתכו אורכן הממוצע היה 81.7 ס"מ.

א. מהי רמת המובהקות המינימלית עבורה נכריע שהמכונה לא מכוילת?

ב. אם נוסיף עוד תצפית שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?

ג. הכרע ברמת מובהקות של 5% האם המכונה מכוילת.

5. אם מקבלים בחישובים אלפא מינימלית (P value) קטנה מאוד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון? לא נכון? נמק.
6. בבדיקת השערות התקבל שה- $p\text{-value}=0.02$.
 מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? 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.
7. מובהקות התוצאה (PV) היא גם : (בחר בתשובה הנכונה)
 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. בבדיקת השערות מסוימת התקבל $p\text{ value}=0.0254$ לכן (בחר בתשובה הנכונה):
 א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .
 ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .
 ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .
 ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

פתרונות :**שאלה 1:**

0.0228

שאלה 2:

עבור כל רמת מובהקות סבירה.

שאלה 3:

ב. 0.1056

ג. 0.1056

ד. נכריע שיש עמידה בהתחייבות של החברה.

שאלה 4:

א. 0.0006

ב. יקטן.

ג. נכריע שאין כיוול.

שאלה 5:

נכון

שאלה 6:

תשובה: א

שאלה 7:

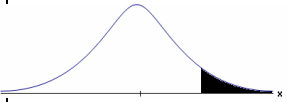
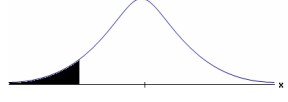
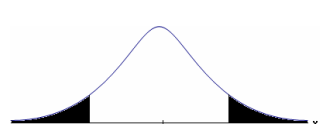
תשובה: א

שאלה 8:

תשובה: ג

בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה

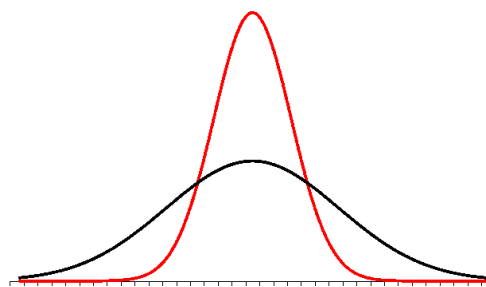
רקע:

| | | | |
|--|--|---|---|
| $H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$ | $H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$ | $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ | השערת האפס: השערה אלטרנטיבית: |
| 5. σ אינה ידועה | | | תנאים: |
| 6. $X \square N$ או מדגם מספיק גדול | | | |
| $t_{\bar{x}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ דוחים את H_0 | $t_{\bar{x}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ דוחים את H_0 | $t_{\bar{x}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{x}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ דוחים את H_0 | כלל ההכרעה: אזור הדחייה של H_0: |
| $\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$ | $\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$ | $\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ $\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ | חלופה לכלל הכרעה: נדחה H_0 אם מתקיים: |

סטטיסטי המבחן:

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

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

התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן $df=n-1$. ככל שדרגות החופש עלות 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%.

תרגילים:

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

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

$$n = 20$$

$$\bar{x} = 3120$$

$$S = 280$$

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

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

4. באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

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

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

נתון שההתפלגות היא נורמלית.

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

5. ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם.

ליאור השתמש בטבלה של התפלגות Z .

רוני השתמשה בטבלה של התפלגות t .

מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.

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 :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר.

החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את

קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15

תצפיות.

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.

פתרונות:

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

כתב ופטר - ברק קנדל ©

שאלה 1: H_0 נדחה**שאלה 2:** H_0 נדחה**שאלה 3:** H_0 נקבל**שאלה 4:** H_0 נקבל**שאלה 5:**

התשובה היא : ב

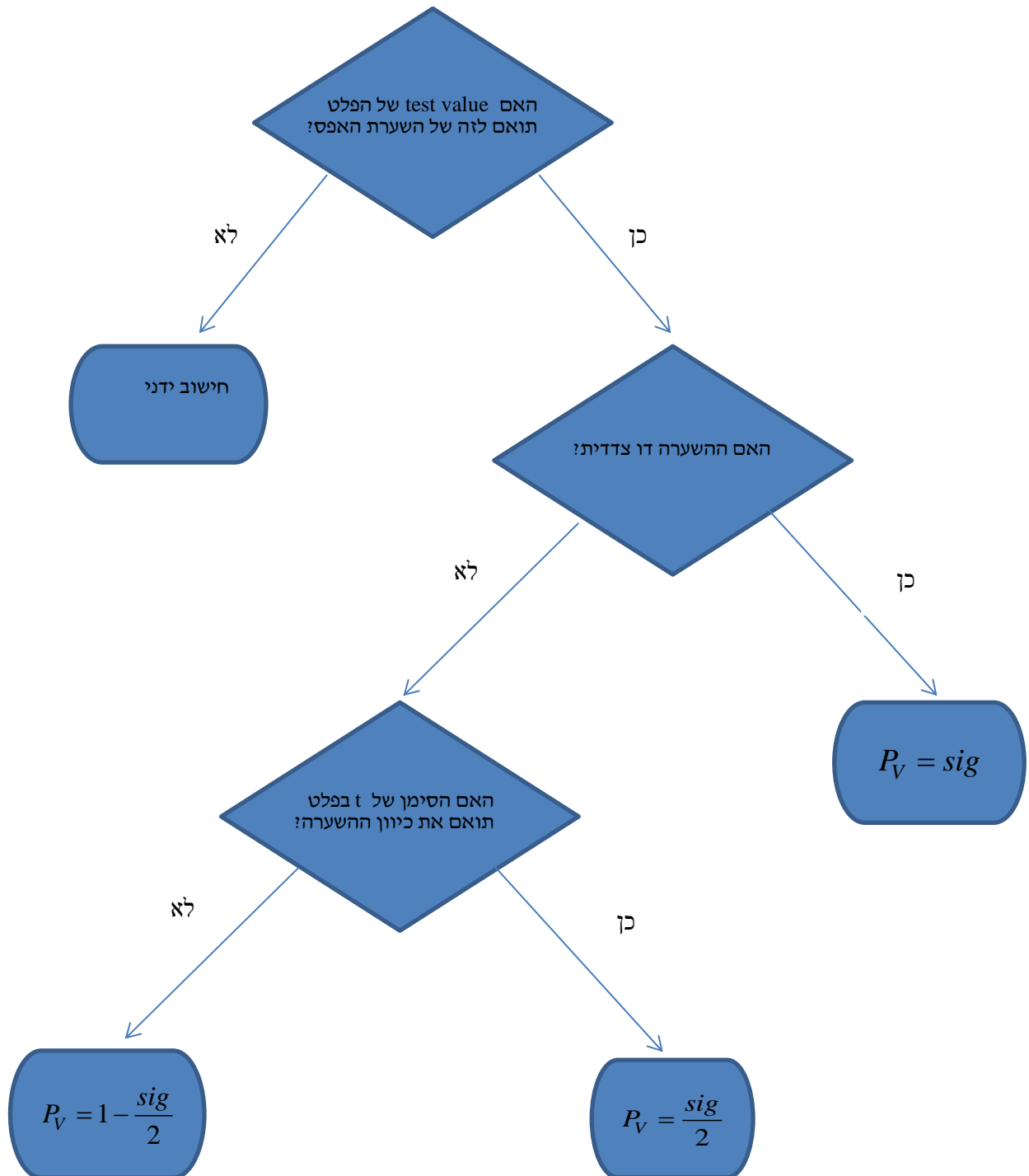
שאלה 6:

התשובה היא : ג

ניתוח פלטי SPSS

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS



דוגמה: (פתרון בהקלטה)

One-Sample Statistics

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|----|---------|----------------|-----------------|
| × | 25 | 87.6400 | 64.90434 | 12.98087 |

One-Sample Test

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

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

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

ב. מהו P_V לבדיקת יעילותה של שיטת הלימוד החדשה?

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

תרגילים :

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

One-Sample Statistics

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

One-Sample Test

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

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

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

One-Sample Test

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

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

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

One-Sample Statistics

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

One-Sample Test

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

א. השלם את הגדלים החסרים המסומנים בסמני שאלה בפלט.

- ב. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים שונה מ-75?
 ג. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים קטנה מ-75?
 ד. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים גדולה מ-75?

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

One-Sample Statistics

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

One-Sample Test

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

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

פתרונות :**שאלה 1:**

ב. 23

ג. (22.104, 24.771)

ד. נקבל H_0 **שאלה 2:**

א. 25

ב. (4.3056, 4.5104)

ד. נדחה H_0 **שאלה 3**א. $n = 44$

$$\bar{X} = 67.66$$

$$t = -2.48$$

$$upper = -1.3735$$

ב. 0.017

ג. 0.0085

ד. 0.9915

שאלה 4:

א. נכריע שיש אמת בפרסום.




ב. מסקנה לא תשתנה.

ג. נכריע שהתוחלת שונה מ-26.

פרק 18 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונויות האוכלוסיה לא ידועות ומניחים שהן שוות

רקע:

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

סטטיסטי המבחן:

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

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

חלופה אחרת לכלל הכרעה:

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

דוגמה : (פתרון בהקלטה)

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשמשים 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%.

תרגילים:

1. להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (מטרים רבועים):

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

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

2. נדגמו 15 ישראלים ו-15 אמריקאים.

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.

| 1-100W | 2-60W | הקבוצה |
|--------|-------|-----------|
| 956 | 1007 | \bar{x} |
| 72 | 80 | S |
| 15 | 13 | n |

א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות במוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפתרון.

ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקות במוצע יותר מאשר נורות מסוג W100?

ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ-1000 שעות. רשמו את כל ההנחות הדרושות.

פתרונות :**שאלה 1:**

לא נדחה H_0

שאלה 2:**שאלה 3:**

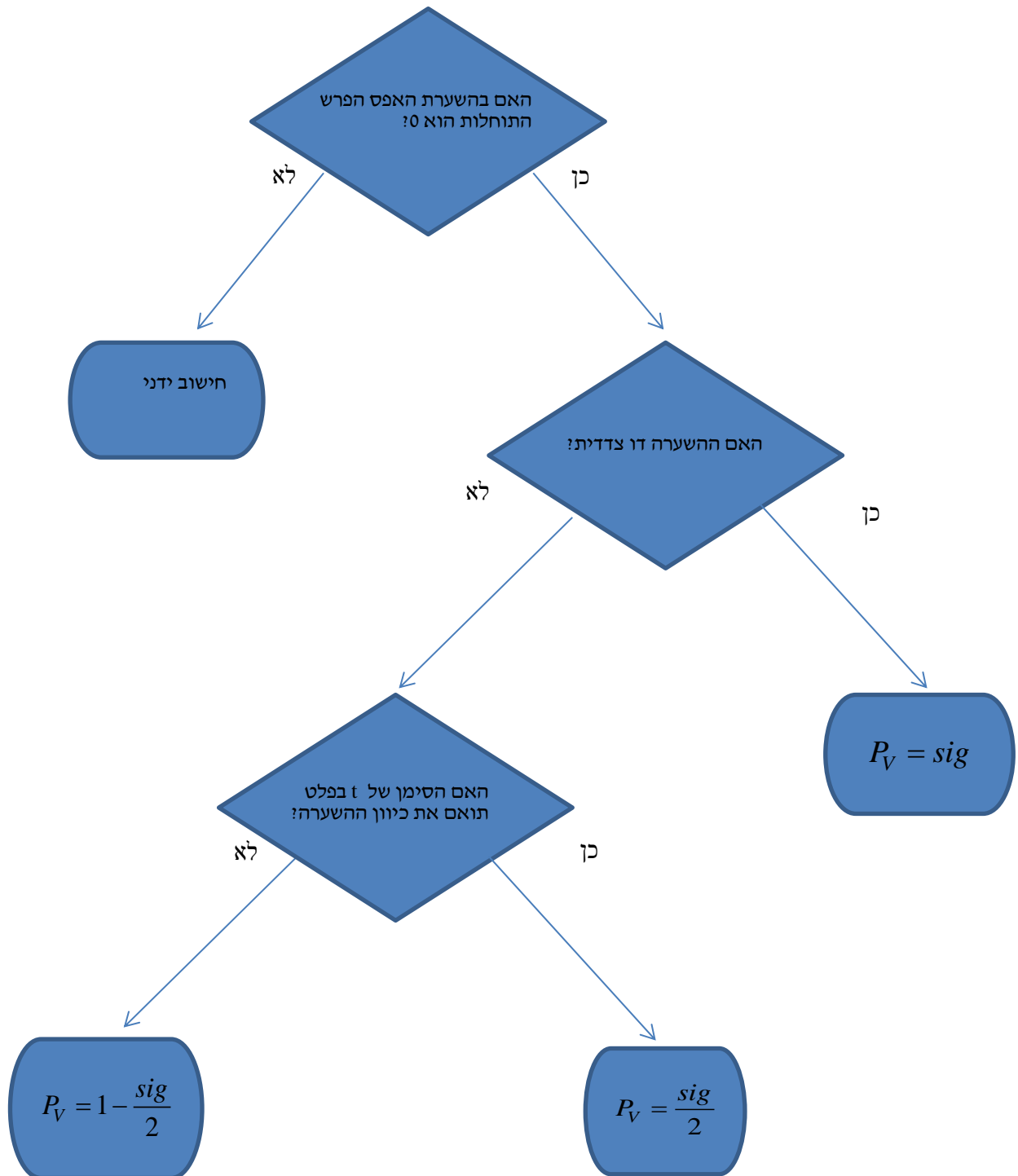
א. נדחה H_0

ב. רמות מובהקות של לפחות 5%

ג. לא נדחה H_0

ניתוח פלטים**רקע:**

מובהקות התוצאה על סמך הפלט:



דוגמה: (פתרון בהקלטה)

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

Group Statistics

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

Independent Samples Test

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

- א. מהו המבחן הסטטיסטי שנעשה כאן?
- ב. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?
- ג. בדוק האם קיים הבדל בין "South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.
- ד. מהי מובהקות התוצאה לבדיקת הטענה שהפרש הממוצע בין-"South East" לבין-"North East" חיובי ?

תרגילים :

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

Group Statistics

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

Independent Samples Test

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

- א. האם ניתן להניח ששוונות האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?
 ב. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?
 ג. מצא את הגבול העליון של רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.
 ד. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

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

T-Test

Group Statistic:

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

Independent Samples Test

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

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

- רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.
 - כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?
 - האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של $\alpha=0.05$.
 - האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha=0.01$.
3. כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל במוצג גיל הנישואין (הראשונים), בין נשים הגרות בערים מרכזיות לבין נשים הגרות בערים מרוחקות מהמרכז.
- לשם כך נדגמו 50 כלות מכל אחת משתי ערים עיר א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שלהלן:

T-Test

Group Statistics

| מקום המגורים | N | Mean | Std. Deviation | Std. Error Mean |
|--------------|-------|------|----------------|-----------------|
| גיל הנישואין | עיר א | 50 | 24.8072 | 1.38978 |
| | עיר ב | 50 | 23.0131 | 1.62070 |

Independent Samples Test

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

א. מהו המבחן הסטטיסטי שנעשה כאן?

ב. מצא רווח סמך ברמת סמך של 95% להפרש בין עיר א לעיר ב מבחינת גיל הנשים הממוצע בנישואין הראשונים.

ג. האם ניתן לומר ברמת מובהקות של 1% שנשים בערים מרכזיות מתחתנות בגיל מאוחר יותר מאשר נשים הגרות בערים מרוחקות?

4. להלן פלט של תוכנת SPSS :

T-Test

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

Independent Samples Test

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

א. השלם את סימני השאלה בטבלה.

ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של X לזה של Y?

ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X גדולה מהתוחלת של Y?

ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X קטנה מהתוחלת של Y?

פתרונות:**שאלה 1:**

- א. נכריע שוויון שוניות.
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.
 ג. 0.11076
 ד. $2.5665 \leq \mu \leq 2.6441$

שאלה 2:

- א. מבחן t למדגמים בלתי תלויים.
 ב. מבחן t למדגמים מזווגים.
 ג. נקבע שקיים שוויון שוניות.
 ד. לא נקבל את טענת הפסיכולוגים.

שאלה 3:

- א. מבחן t להשוואת תוחלת במדגמים בלתי תלויים.
 ב. $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$
 ג. כן.

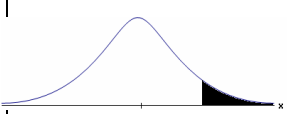

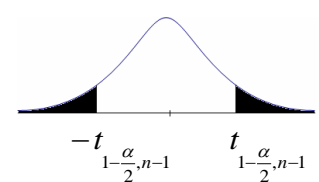
שאלה 4:

- א. -10.15, 4.69, 48
 ב. 0.04
 ג. 0.978
 ד. 0.022

**פרק 19 - בדיקת השערות על תוחלת הפרשים במדגמים מזווגים
(תלויים)**

בדיקת השערות למדגמים מזווגים

רקע:

| | | | |
|--|--|--|---|
| $H_0: \mu_D = C$ $H_1: \mu_D > C$ | $H_0: \mu_D = C$ $H_1: \mu_D < C$ | $H_0: \mu_D = C$ $H_1: \mu_D \neq C$ | השערת האפס : השערה אלטרנטיבית: |
| .4 σ_D אינה ידועה .5 $D \square N$ או מדגם מספיק גדול | | | תנאים: |
| $t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ - דוחים את H_0 | $t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ - דוחים את H_0 | $t_{\bar{D}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{D}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ - דוחים את H_0 | כלל הכרעה: אזור הדחייה של H_0 : |
| $\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ | $\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ | $\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ | חלופה לכלל הכרעה : נדחה H_0 אם מתקיים: |

סטטיסטי המבחן :

$$t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}$$

דוגמה : (פתרון בהקלטה)

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה נמוכים מהמחירים מרשת השיווק "שופרסל".

לצורך בדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות. להלן המחירים :

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

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

תרגילים:

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

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

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

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

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

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

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

- פורסם שתלמידים שמסיימים את סמסטר ב משפרים בממוצע את הציונים ב-5 נקודות לעומת סמסטר א'. הנח שהציונים מתפלגים נורמלית.
- א. מהי מובהקות התוצאה לבדיקת הטענה שהשיפור הוא יותר מ 5 נקודות?
- ב. על סמך הסעיף הקודם, מהי רמת המובהקות המינימלית להכרעה שהשיפור הוא יותר מ- 5 נקודות?
- ג. לאור זאת, מה המסקנה ברמת מובהקות של 10% ?

4. לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית.
- המבחן שיש לבצע כאן הוא:
- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים מזווגים.

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

| | | | | |
|---------------|-----|-----|-----|-----|
| משקל במכשיר 1 | 5.4 | 6.9 | 7.0 | 5.2 |
| משקל במכשיר 2 | 5.3 | 6.9 | 7.1 | 5.0 |

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

פתרונות:**שאלה 1:**לא נדחה H_0 **שאלה 2:**לא נדחה H_0 **שאלה 3:**א. $0.25 \leq p \leq 0.5$

ב. 0.5

ג. לא נדחה H_0 **שאלה 4:**

התשובה היא ד.

שאלה 5:

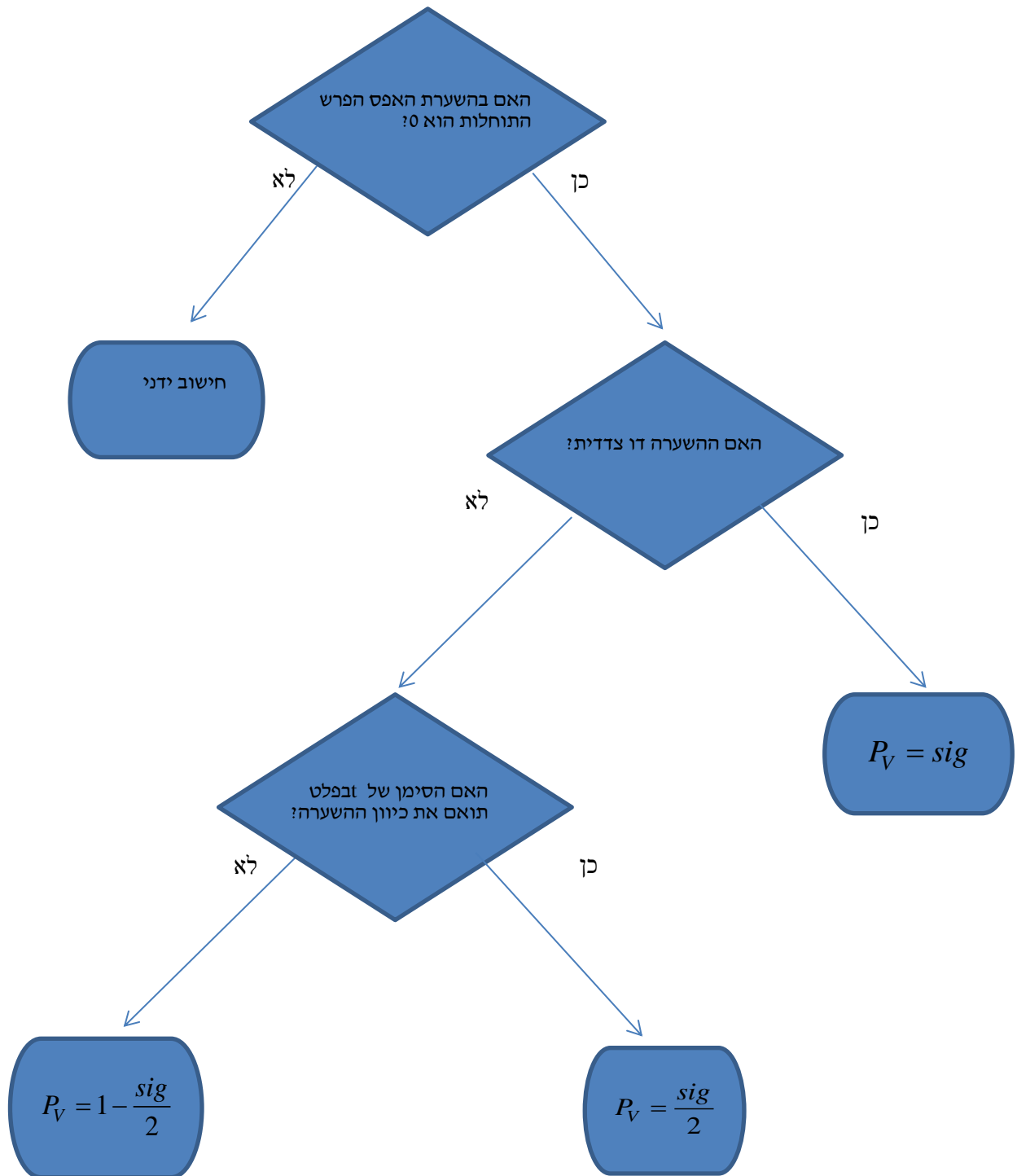
התשובה היא ד.

שאלה 6:

התשובה היא ג.

מדגמים מזווגים - ניתוח פלטים

רקע:



דוגמה : (פתרון בהקלטה)

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

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

Paired Samples Statistics

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

Paired Samples Test

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

בדוק ברמת מובהקות של 5% האם הקורס יעיל .

תרגילים :

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

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

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

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

Paired Samples Statistics

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

Paired Samples Test

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

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

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

Paired Samples Test

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

- א. רשום מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.
- ב. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.
- ג. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.
- ד. מצא רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

T-Test

Paired Samples Statistics

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

Paired Samples Test

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

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

פתרונות:

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

שאלה 1:א. -0.007 ב. 0.1

ג. אין הבדל מובהק

שאלה 2:א. $d.f = 299$ $n = 300$ $\bar{X} = 8.496$

Sig=0

ב. נדחה H_0 ג. $(0.46341, 0.38126)$ ד. $(0.4708, 0.3738)$ **שאלה 3:**ב. נדחה H_0 ג. לא נדחה את H_0 ד. $(1.592, 13.317)$ **שאלה 4:**א. $1.5876, 11.5715$ ב. 0.173

ג. יגדל

ד. 0.0865 ה. 0.9135 ו. $49.18 < \mu < 58.82$

פרק 20- שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות מסכמות בסגנון רב ברירה (אמריקאיות) על בדיקת השערות

1. בבדיקת השערה חד-צדדית ימנית ברמת מובהקות $\alpha=0.01$, נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות אותם נתונים ברמת מובהקות $\alpha=0.05$?
- השערת האפס הייתה נדחית.
 - השערת האפס לא הייתה נדחית.
 - ההשערה המחקרית הייתה נדחית.
 - בהעדר נתונים נוספים, לא ניתן לדעת.
2. לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמיתית.
- המבחן שיש לבצע כאן הוא:
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.
3. כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות:

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

מהן ההשערות הנבדקות? (ההפרש חושב X-Y)

א. $H_1 : \mu_d < 0, H_0 : \mu_d = 0$

ב. $H_1 : \mu_x - \mu_y < 0, H_0 : \mu_x - \mu_y = 0$

ג. $H_1 : \mu_x - \mu_y < 0, H_0 : \mu_x - \mu_y = 0$

ד. $H_1 : \mu_d > 0, H_0 : \mu_d = 0$

4. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :
- השערת האפס נכונה.
 - השערת האפס נדחתה.
 - השערת האפס לא נדחתה.
 - אף אחת מהתשובות לא נכונה בהכרח.
5. ידוע כי ילד בגיל שנתיים ישן בממוצע 9 שעות בלילה. במדגם של 20 תינוקות בני שנתיים המתגוררים בצפון נמצא, כי ממוצע שעות השינה בלילה הינו 10 עם סטיית תקן של 1.1 במדגם של 10 תינוקות בדרום נמצא, כי ממוצע שעות השינה בלילה הינו 7.9 עם סטיית תקן של 1.1. על מנת להשוות בין ממוצע שעות השינה של ילדים מהצפון לבין זה של כלל הילדים יש לערוך _____, ועל מנת להשוות בין ממוצע שעות השינה של ילדים מהדרום לזה של ילדים המתגוררים בצפון יש לערוך _____. יש להניח שההנחות הדרושות מתקיימות.
- מבחן Z למדגם יחיד ; מבחן T למדגם יחיד.
 - מבחן T למדגם יחיד ; מבחן T למדגמים תלויים.
 - מבחן T למדגם יחיד ; מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים בלתי תלויים ; מבחן T לממוצע יחיד.
6. מובהקות התוצאה ($P < 0.05$) היא גם :
- רמת המובהקות המינימאלית לדחות השערת האפס.
 - רמת המובהקות המקסימאלית לדחיית השערת האפס.
 - רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
 - רמת המובהקות המינימאלית לאי דחיית השערת האפס.
7. כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות :
- | | | | | |
|----|----|----|----|---------------|
| 68 | 82 | 93 | 69 | לפני הנישואין |
| 71 | 84 | 88 | 80 | לאחר הנישואין |
- באיזה התפלגות משתמשים לבדיקת ההשערות, ובכמה דרגות חופש :
- ההתפלגות Z ללא דרגות חופש.
 - ההתפלגות T ו-3 דרגות חופש.
 - ההתפלגות T ו-6 דרגות חופש.
 - ההתפלגות χ^2 ו-3 דרגות חופש.

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

11. שני חוקרים העוסקים בתחום מחקרי משותף החליטו להסתמך על נתונים של מדגם שפורסם על ידי הלשכה המרכזית לסטטיסטיקה. חוקר א' ניסח השערה דו צדדית ואילו חוקר ב' ניסח השערה חד צדדית. מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?
- אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב' באותה רמת מובהקות.
 - אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס באותה רמת מובהקות.
 - אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת האפס באותה רמת מובהקות.
 - אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת האפס בתנאי שרמת המובהקות כפולה בגודלה.

12. ידוע מנתוני העבר כי תוחלת הציונים בבחינה בפסיכולוגיה היא 79. הועלתה השערה כי תוחלת הציונים בקרב העולים החדשים נמוכה יותר. לצורך בדיקת הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75. מה משמעות הפרמטר בניסוח ההשערות?
- תוחלת ציוני העולים באוכלוסייה.
 - ממוצע ציוני העולים במדגם.
 - תוחלת ציוני האוכלוסייה מנתוני העבר.
 - ממוצע ציוני שאר האוכלוסייה במדגם.

13. חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?
- החוקר דחה את השערת H_0 כאשר היא הייתה נכונה.
 - החוקר דחה את השערת H_1 כאשר היא הייתה נכונה.
 - החוקר לא דחה את השערת H_0 כאשר היא הייתה לא נכונה.
 - המדגם של החוקר שייך בפועל להתפלגות הדגימה של H_1 .

14. חוקר ביקש לבחון האם תאומים זהים אשר הופרדו בילדותם שונים מתאומים זהים אשר גדלו יחדיו מבחינת מידת הפער בין התאומים בלחץ הדם. הוא דגם 20 זוגות תאומים מכל אוכלוסייה ומדד את הפרש בין לחץ הדם בכל זוג תאומים. מהו המבחן הסטטיסטי המתאים?
- מבחן t למדגמים בלתי תלויים עם 38 דרגות חופש.
 - מבחן t למדגמים מזווגים, עם 39 דרגות חופש.
 - מבחן t למדגמים בלתי תלויים עם 39 דרגות חופש.
 - מבחן t למדגמים מזווגים עם 38 דרגות חופש.

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

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

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

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:

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

22. נתון ש $X \sim N(\mu, \sigma^2)$ כמו כן נתונות ההשערות הבאות :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות.

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

23. אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי :

א. הסיכוי לטעות מסוג ראשון גדל.

ב. העוצמה של המבחן גדלה.

ג. הסיכוי לטעות מסוג שני גדל.

ד. תשובות א ו-ב נכונות.

24. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :

א. השערת האפס נכונה.

ב. השערת האפס נדחתה.

ג. השערת האפס לא נדחתה.

ד. אף אחת מהתשובות לא נכונה בהכרח.

25. מה המצב הרצוי לחוקר המבצע בדיקת השערה :

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

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- $p\text{-value}=0.02$. מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה.

ה. יקבל את השערת האפס בכל מקרה.

ו. ידחה את השערת האפס מקרה.

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

ח. לא ניתן לדעת כי אין מספיק נתונים.

29. מובהקות התוצאה (P-value) 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 $p\text{ value}=0.0254$ לכן:

א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .

ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .

ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .

ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

31. רמת המובהקות במחקר הייתה 2% לכן.
- בסיכוי של 2% נדחה את השערת האפס.
 - בסיכוי של 2% לא נדחה את השערת האפס.
 - בסיכוי של 2% השערת האפס לא נכונה.
 - אף תשובה לא נכונה.

32. נתון ש $X \sim N(\mu, \sigma^2)$ כמו כן נתונות ההשערות הבאות :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

- חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. 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:
- כעת בברור הוא ידחה את השערת האפס.
 - כעת הוא דווקא יקבל את השערת האפס.
 - כעת לא ניתן לדעת מה תהיה מסקנתו.

33. לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית.

מספר דרגות החופש במבחן הוא :

א. 9

ב. 19

ג. 18

ד. 8

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

| | | | | |
|---------------|-----|-----|-----|-----|
| משקל במכשיר 1 | 5.4 | 6.9 | 7.0 | 5.2 |
| משקל במכשיר 2 | 5.3 | 6.9 | 7.1 | 5.0 |

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:

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

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 מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו:

$-0.0293 < \mu_D < 0.2145$ רווח הסמך הוא ברמת סמך של 95%.

לכן מסקנת המחקר היא:

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

39. אם רמת מובהקות של מבחן סטטיסטי הינה 0, הכוונה היא:

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

40. חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות: $H_0: \mu = 10$ לצורך בדיקה הוא לקח $H_1: \mu \neq 10$

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

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 פרסם שהאיגוד האמריקאי מגיע למסקנותיו על סמך רמת מובהקות של 0.5%.

מה הסיכוי לבצע טעות מסוג ראשון במחקר?

- 0.005
- 0.03
- 0.0287
- 0.05

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:

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

43. איזה מהמשפטים הבאים נכון תמיד?

א. $POWER + \beta + \alpha = 1$

ב. $\beta - POWER = .05$

ג. $POWER + \alpha = 1$

ד. $\beta + \alpha = 1$

ה. הכול לא נכון

44. מה נכון לומר לגבי הנחת שיוויון השונויות במבחן T למדגמים בלתי תלויים?

א. היא אומרת שהשונויות המדגמיות שוות.

ב. בלעדיה אין שום דרך לבדוק השערה על הפרש בין תוחלות.

ג. היא חשובה הן עבור מדגמים מזווגים והן עבור מדגמים בלתי תלויים.

ד. אף תשובה אינה נכונה.

45. חוקר החליט לא לדחות השערה ברמת מובהקות של α . במידה וחוקר זה היה בודק השערה זו

ברמת מובהקות של 2α על סמך אותם נתונים האם ההשערה תדחה?

א. ההשערה תדחה.

ב. ההשערה לא תדחה.

ג. התשובה תלויה בעוצמת המבחן.

ד. לא ניתן לדעת בוודאות אם ההשערה תדחה או לא.

46. חוקרת שיערה, כי בגילאי הגן בנות יותר תקשורתיות מבנים. אם החוקרת תדגום אקראית 30 בנים ו-

30 בנות, ובמדגם יתקבל אותו ממוצע של ציון תקשורת. סטטיסטי המבחן יהיה:

א. אפס

ב. חיובי

ג. שלילי

ד. לא ניתן לדעת

47. עוצמה שווה ל-1 פרושה:

א. לעולם לא לדחות את השערת האפס כאשר היא נכונה.

ב. תמיד לדחות את השערת האפס כאשר היא נכונה.

ג. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

48. מה מהבאים נכון לגבי מבחן t מדגמים מזווגים?

- א. כל התצפיות במחקר אינן תלויות זו בזו.
- ב. כל התצפיות במחקר תלויות זו בזו.
- ג. כל הצמדים של תצפיות במחקר אינם תלויים זה בזה.
- ד. התצפיות בתוך כל צמד אינן תלויות זו בזו.

49. לבדיקת ההשערה החד צדדית על התוחלת של התפלגות נורמלית נלקח מדגם

$$H_0 : \mu \geq 10$$

$$H_1 : \mu < 10$$

והתקבלה רמת מובהקות מינימאלית לדחיית השערת האפס 0.058. לו רצינו לבדוק את ההשערה

הדו צדדית $H_0 : \mu = 10$, אז על סמך תוצאת אותו המדגם ברמת מובהקות 0.05:

$$H_1 : \mu \neq 10$$

- 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 נלקח מדגם מקרי בגודל n $H_0 : \mu = 55$
 $H_1 : \mu = 65$

מאוכלוסייה בעלת התפלגות נורמלית ושונות σ^2 . רמת המובהקות היא 5%. נמצא שהעוצמה היא 0.9. להלן 3 טענות:

(1) עבור מדגם בגודל n וברמת מובהקות 5% לבדיקת ההשערות: $H_0 : \mu = 55$
 $H_1 : \mu = 60$ תהיה גדולה מ- 0.9.

(2) עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות: $H_0 : \mu = 55$
 $H_1 : \mu = 65$ תהיה גדולה מ- 0.9.

(3) עבור מדגם בגודל n ורמת מובהקות 10% לבדיקת ההשערות: $H_0 : \mu = 55$
 $H_1 : \mu = 65$ 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.

פתרונות :

| | | | |
|---|----|---|----|
| א | 26 | א | 1 |
| ב | 27 | | |
| א | 28 | ד | 3 |
| א | 29 | א | 4 |
| ג | 30 | ג | 5 |
| ד | 31 | ג | 6 |
| א | 32 | א | 7 |
| א | 33 | ב | 8 |
| ד | 34 | ג | 9 |
| ג | 35 | א | 10 |
| א | 36 | א | 11 |
| ג | 37 | | |
| א | 38 | א | 13 |
| ג | 39 | א | 14 |
| א | 40 | א | 15 |
| א | 41 | א | 16 |
| א | 42 | | |
| ה | 43 | ג | 18 |
| ד | 44 | א | 19 |
| ד | 45 | ג | 20 |
| א | 46 | ב | 21 |
| ד | 47 | ג | 22 |
| ג | 48 | ד | 23 |
| ב | 49 | ג | 24 |
| | 5 | ג | 2 |

פרק 21 - מבחני חי בריבוע

מבחן טיב התאמה

רקע:

מבחן זה הוא מבחן הבא לבדוק האם אוכלוסייה מסוימת מתפלגת לפי התפלגות נתונה. המשתנה הנחקר מחולק למספר קטגוריות ויש לבדוק האם תוצאות המדגם תואמות להתפלגות הנתונה.

מבנה המבחן:

השערות:

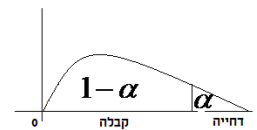
H_0 : המשתנה מתפלג לפי התפלגות מסוימת

H_1 : אחרת

כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש.

כאשר K - מספר הקטגוריות, $d.f = K - 1$.



הערך הקריטי הוא: $\chi^2_{1-\alpha, K-1}$, כלומר האחוזון ה- $1-\alpha$ בהתפלגות חי בריבוע שדרגות החופש הן $K-1$.

אם $\chi^2 > \chi^2_{1-\alpha, K-1}$, דוחים את השערת האפס.

סטטיסטי המבחן:

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

O_i - השכיחות שנצפתה במדגם בקטגוריה i .

p_i - הסתברות לקטגוריה i לפי השערת האפס.

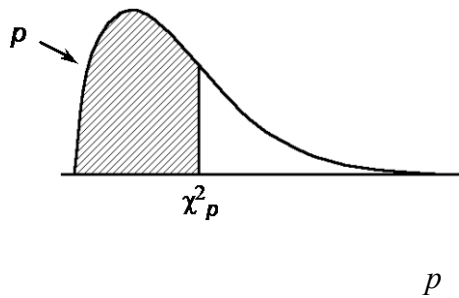
$E_i = np_i$ שכיחות צפויה במדגם לקטגוריה i בהנחת השערת האפס.

הערה: תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד

קטגוריות סמוכות עד שהתנאי יתקיים.

זוגמה: (פתרון בהקלטה)

במדינה מסוימת שלוש מפלגות. בפרלמנט הנוכחי התפלגות מספר המושבים היא 30% למפלגה A, 60% למפלגה B ו-10% למפלגה C. לקראת הבחירות המתוכננות בשבוע הבא נעשה סקר שכלל 300 אזרחים. בסקר התקבל ש-40% יצביעו למפלגה A, 50% למפלגה B ו-10% למפלגה C. האם תוצאות הסקר תואמות להתפלגות המושבים בפרלמנט הנוכחי? בדקו ברמת מובהקות של 5%.

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p 

| df | .005 | .01 | .025 | .05 | .10 | .25 | .50 | .75 | .90 | .95 | .975 | .99 | .995 |
|----|----------|----------|----------|----------|--------|-------|-------|------|------|------|------|------|------|
| 1 | 0.004393 | 0.005157 | 0.005982 | 0.006393 | 0.0158 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 0.211 | 0.575 | 1.39 | 2.77 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 |
| 3 | 0.0717 | 0.115 | 0.216 | 0.352 | 0.584 | 1.21 | 2.37 | 4.11 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.06 | 1.92 | 3.36 | 5.39 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 |
| 5 | 0.412 | 0.554 | 0.831 | 1.15 | 1.61 | 2.67 | 4.35 | 6.63 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 3.45 | 5.35 | 7.84 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 4.25 | 6.35 | 9.04 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 5.07 | 7.34 | 10.2 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 5.90 | 8.34 | 11.4 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 6.74 | 9.34 | 12.5 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 7.58 | 10.3 | 13.7 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 8.44 | 11.3 | 14.8 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 9.30 | 12.3 | 16.0 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 10.2 | 13.3 | 17.1 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 11.0 | 14.3 | 18.2 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 11.9 | 15.3 | 19.4 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 12.8 | 16.3 | 20.5 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 13.7 | 17.3 | 21.6 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 14.6 | 18.3 | 22.7 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 15.5 | 19.3 | 23.8 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 16.3 | 20.3 | 24.9 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 17.2 | 21.3 | 26.0 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 18.1 | 22.3 | 27.1 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 19.0 | 23.3 | 28.2 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 19.9 | 24.3 | 29.3 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 20.8 | 25.3 | 30.4 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 21.7 | 26.3 | 31.5 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 22.7 | 27.3 | 32.6 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 23.6 | 28.3 | 33.7 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 24.5 | 29.3 | 34.8 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 |

תרגילים:

1. במטרה לבדוק האם קובייה הוגנת, מטילים אותה 120 פעמים. התקבל 17 פעמים 1, 23 פעמים 2, 20 פעמים 3, 25 פעמים 4, 18 פעמים 5 ו- 17 פעמים 6. מה מסקנתכם ברמת מובהקות של 5%?
2. מפעל מייצר סוכריות בצבעים כחול, אדום, ירוק וכתום. מעוניינים לבדוק שפרופורציית הסוכריות הכחולות גדולה פי 2 מכל צבע אחר. לצורך כך נדגמו באקראי 200 סוכריות והתקבל: 70 כחולות, 50 אדומות, 40 ירוקות והיתר כתומות. מה מסקנתכם ברמת מובהקות של 5%?
3. 200 איש נתבקשו לבחור ספרה באקראי והנה התוצאות שהתקבלו :
 - 18 איש בחרו בספרה 0
 - 24 איש בחרו בספרה 1
 - 17 איש בחרו בספרה 2
 - 19 איש בחרו בספרה 3
 - 20 איש בחרו בספרה 4
 - 18 איש בחרו בספרה 5
 - 22 איש בחרו בספרה 6
 - היתר בחרו בספרות 7-9
 א. על סמך התוצאות הללו האם בחירת הספרות אקראית? רמת מובהקות של 2.5%.
 ב. תנו הערכה למובהקות התוצאה.
 ג. אם נגדיל את גודל המדגם פי 2 ונשמור על אותם יחסים של כמות האנשים במדגם שבחרו בספרות, כיצד הדבר ישפיע על ערכו של הסטטיסטי χ^2 ? מה תהיה המסקנה במקרה זה?

4. מעוניינים לבדוק האם קובייה היא הוגנת. הטילו את הקובייה פעמיים והתבוננו בסכום הוצאות. חזרו על התהליך 72 פעמים. להלן התוצאות שהתקבלו במדגם:

| מספר הטלות | סכום התוצאות |
|------------|--------------|
| 20 | 2-5 |
| 17 | 6-8 |
| 20 | 9-10 |
| 15 | 11-12 |

מה המסקנה ברמת מובהקות של 5%.

5. בפנס יש 4 סוללות. בבדיקה שנערכה ב-400 פנסים נמצאו סוללות פגומות לפי השכיחויות הבאות:

| מספר הסוללות הפגומות | 0 | 1 | 2 | 3 ומעלה |
|----------------------|-----|-----|----|---------|
| שכיחות | 276 | 104 | 12 | 8 |

מעוניינים לבדוק על סמך תוצאות מדגם אלה האם הסיכוי לסוללה פגומה הוא 20%. מהי רמת המובהקות המינימלית עבורה נכריע שהסיכוי לסוללה פגומה אינו 20%?

6. להלן השערות מחקר:

$$H_0 : X \sim N(40, 2^2)$$

$$H_1 : \text{else}$$

תוצאות המדגם הן:

| X | מתחת 36 | 36-40 | 40-44 | מעל 44 |
|--------------|---------|-------|-------|--------|
| מספר הדגימות | 3A | 50A | 45A | 2A |

מהו ערכו המקסימלי של A עבורו נקבל את H_0 ברמת מובהקות של 5%?

פתרונות:**שאלה 1 :**לא נדחה H_0 **שאלה 2 :**לא נדחה H_0 **שאלה 3 :**א. לא נדחה H_0 ב. בין 0.95 ל- 0.975 ג. יגדל פי 2. מסקנה לא תשתנה.**שאלה 4 :**

נכריע שהקובייה אינה הוגנת.

שאלה 5 :

0.005

שאלה 6 :

14

מבחן חי בריבוע לאי תלות בין משתנים

רקע:

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים. שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות.

מבנה המבחן:

השערות:

H_0 : אין תלות בין המשתנים

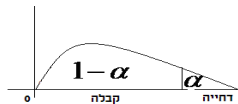
H_1 : יש תלות בין המשתנים

כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש.

$$d.f = (r-1)(c-1)$$

כאשר: r - מספר הקטגוריות של המשתנה שבשורות.
 c - מספר הקטגוריות של המשתנה שבעמודות.



הערך הקריטי הוא: $\chi^2_{1-\alpha, (r-1)(c-1)}$, כלומר האחוזון ה- $1 - \alpha$ בהתפלגות חי בריבוע שדרגות

החופש הן $(r-1)(c-1)$. אם $\chi^2 > \chi^2_{1-\alpha, (r-1)(c-1)}$ אז דוחים את השערת האפס.

סטטיסטי המבחן:

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

O_i - השכיחות נצפית במדגם בתא i .

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

$$E_i = \frac{f(x) \cdot f(y)}{n}$$

הערה: תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות

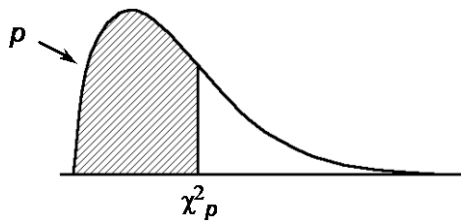
לאחד קטגוריות סמוכות עד שהתנאי יתקיים. תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ-20% מהתאים E קטן מ-5.

דוגמה: (הפתרון בהקלטה)

האם יש תלות בין המגדר לבין דעה מסוימת? יש לבדוק ברמת מובהקות של 5% על סמך תוצאות הסקר:

| סה"כ | נמנע | נגד | בעד | דעה המגדר |
|------|------|-----|-----|--------------|
| | 10 | 40 | 50 | גברים |
| | 20 | 60 | 20 | נשים |
| | | | | סה"כ |

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p



p

| df | .005 | .01 | .025 | .05 | .10 | .25 | .50 | .75 | .90 | .95 | .975 | .99 | .995 |
|----|----------------------|----------------------|----------------------|----------------------|--------|-------|-------|------|------|------|------|------|------|
| 1 | 0.0 ⁴ 393 | 0.0 ³ 157 | 0.0 ³ 982 | 0.0 ² 393 | 0.0158 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 0.211 | 0.575 | 1.39 | 2.77 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 |
| 3 | 0.0717 | 0.115 | 0.216 | 0.352 | 0.584 | 1.21 | 2.37 | 4.11 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.06 | 1.92 | 3.36 | 5.39 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 |
| 5 | 0.412 | 0.554 | 0.831 | 1.15 | 1.61 | 2.67 | 4.35 | 6.63 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 3.45 | 5.35 | 7.84 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 4.25 | 6.35 | 9.04 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 5.07 | 7.34 | 10.2 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 5.90 | 8.34 | 11.4 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 6.74 | 9.34 | 12.5 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 7.58 | 10.3 | 13.7 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 8.44 | 11.3 | 14.8 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 9.30 | 12.3 | 16.0 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 10.2 | 13.3 | 17.1 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 11.0 | 14.3 | 18.2 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 11.9 | 15.3 | 19.4 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 12.8 | 16.3 | 20.5 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 13.7 | 17.3 | 21.6 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 14.6 | 18.3 | 22.7 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 15.5 | 19.3 | 23.8 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 16.3 | 20.3 | 24.9 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 17.2 | 21.3 | 26.0 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 18.1 | 22.3 | 27.1 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 19.0 | 23.3 | 28.2 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 19.9 | 24.3 | 29.3 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 20.8 | 25.3 | 30.4 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 21.7 | 26.3 | 31.5 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 22.7 | 27.3 | 32.6 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 23.6 | 28.3 | 33.7 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 24.5 | 29.3 | 34.8 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 |

תרגילים :

1. נבדק התלות בין גודל הארגון לבין שביעות הרצון של העובדים. להלן התוצאות :

| שביעות רצון גודל המפעל | נמוכה | בינונית | גבוהה | סה"כ |
|---------------------------|-------|---------|-------|------|
| גדול | 182 | 203 | 215 | 600 |
| קטן | 154 | 110 | 136 | 400 |
| סה"כ | 336 | 313 | 351 | 1000 |

מה המסקנה ברמת מובהקות של 2.5%?

2. מפעל עובד בשלוש משמרות. להלן מספר המוצרים הפגומים והתקינים בכל אחת מן המשמרות לפי מדגם שנעשה :

| | יום | ערב | לילה |
|--------|-----|-----|------|
| פגומים | 50 | 60 | 70 |
| תקינים | 600 | 700 | 800 |

האם יש הבדל בין שיעורי הפגומים

במשמרות השונות? הסיקו עבור רמת מובהקות $\alpha = 0.05$.

3. נדגמו 50 מוצרים ממפעל מסוים מתוך 30 מוצרים שיוצרו ביום 17 נבחרו לייצוא מתוך המוצרים שיוצרו בלילה 10 נבחרו לייצוא. האם יש קשר בין היות מוצר לייצוא למועד שבו הוא יוצר ? בדוק ברמת בטחון של 95%.

פתרונות:

שאלה 1:

נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.

שאלה 2:

נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.

שאלה 3:

נסיק שאין קשר בין היות מוצר לייצוא למועד שבו הוא יוצר.

פלטטים על מבחן לאי תלות

רקע:

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים. שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות.

מבנה המבחן:

השערות:

H_0 : אין תלות בין המשתנים

H_1 : יש תלות בין המשתנים

דרגות חופש:

$$d.f = (r-1)(c-1)$$

כאשר r : מספר הקטגוריות של המשתנה שבשורות.
 c : מספר הקטגוריות של המשתנה שבעמודות.

סטטיסטי המבחן:

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

O_i - השכיחות נצפית במדגם בתא i .

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

$$E_i = \frac{f(x) \cdot f(y)}{n}$$

הערה: תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים. תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ-20% מהתאים E קטן מ-5.

דוגמה : (פתרון בהקלטה)

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

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

להלן הפלטים שהתקבלו :

Case Processing Summary

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

hair_color * eye_color Crosstabulation

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

Chi-Square Tests

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

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

תרגילים:

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

א. מה ההשערות הנבדקות ומהו סטטיסטי המבחן?

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

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

ד. מהו האומדן לאחוז התומכים במדינה פלסטינית מתוך הגברים?

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

להלן הפלטים :

Chi-Square Tests

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

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

b. minimum expected count is 7.25.

gender * opinion Crosstabulation

| | | | opinion | | | Total |
|--------|------|------------------|---------|-------|------------|--------|
| | | | yes | now | no opinion | |
| gender | Male | Count | 10 | 10 | 9 | 29 |
| | | Expected Count | 12.6 | 9.2 | 7.3 | 29.0 |
| | | % within gender | 34.5% | 34.5% | 31.0% | 100.0% |
| | | % within opinion | 38.5% | 52.6% | 60.0% | 48.3% |

| | | | | | |
|--------|------------------|--------|--------|--------|--------|
| | % of Total | 16.7% | 16.7% | 15.0% | 48.3% |
| female | Count | 16 | 9 | 6 | 31 |
| | Expected Count | 13.4 | 9.8 | 7.8 | 31.0 |
| | % within gender | 51.6% | 29.0% | 19.4% | 100.0% |
| | % within opinion | 61.5% | 47.4% | 40.0% | 51.7% |
| | % of Total | 26.7% | 15.0% | 10.0% | 51.7% |
| Total | Count | 26 | 19 | 15 | 60 |
| | Expected Count | 26.0 | 19.0 | 15.0 | 60.0 |
| | % within gender | 43.3% | 31.7% | 25.0% | 100.0% |
| | % within opinion | 100.0% | 100.0% | 100.0% | 100.0% |
| | % of Total | 43.3% | 31.7% | 25.0% | 100.0% |

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

Crosstab

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

Chi-Square Tests

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

a. 5 cells (55.6%) have expected count less than 5. The minimum expected count is 1.43.

- א. בין אלו משתנים נבדק הקשר? כמה קטגוריות לכל משתנה?
- ב. האם התנאים של המודל מתקיימים?
- ג. מה מספר דרגות החופש במבחן הנייל?
- ד. מה ההשערות של המבחן?

3. מחקר התעניין לבדוק את הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות. הפלטים מצורפים.
- א. השלימו את שלושת המספרים החסרים בטבלה (היכן שיש סימני שאלה).
- ב. מה ערכו של חי בריבוע הסטטיסטי.
- ג. תנו הערכה למובהקות התוצאה לבדיקת הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות.

Crosstabulation רמת_הכנסה * צרכן עגבניות

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

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

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

להלן הפלט:

Crosstabulation צבע עיניים * צבע שיער

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

5. במחקר מסוים רצו לבדוק האם יש קשר בין המגדר להוצאה על לבוש במשך שנה. דגמו באופן מקרי גברים ונשים ובדקו את רמת ההוצאה שלהם על לבוש בשנה האחרונה. חוקר א' בדק האם קיים הבדל בתוחלות ההוצאה בין גברים לנשים. חוקר ב' קיבץ את ההוצאה לקטגוריות ובאופן הזה בדק האם קיים הבדל בהתפלגות ההוצאה בין גברים לנשים. הקטגוריות חולקו לשלוש קבוצות הוצאה .
- א. איזה פלט מתאים לאיזה אחד מהחוקרים? נמקו.
- ב. מה מסקנתו של חוקר א'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ותנו מסקנה במונחי המשתנים)
- ג. איזו טעות יכולה להיות במסקנתו של חוקר א'? נסחו את הטעות במונחי השאלה.
- ד. מהי מסקנתו של חוקר ב'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ורשמו מסקנה במונחי המשתנים)
- ה. איזו טעות יכולה להיות במסקנתו של חוקר ב'? נסחו זאת במונחי השאלה
- ו. כיצד ניתן ליישב את מסקנות שני החוקרים?

להלן פלט ראשון :

T-Test
Group Statistics

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

dimensio
n1

Independent Samples Test

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

Crosstabs
Case Processing Summary

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

gender * category Crosstabulation

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

Chi-Square Tests

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

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

The minimum expected count is 8.00.

פתרונות:**שאלה 1:**

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

שאלה 2:

- א. בילוי מועדף ו- מוזיקה מועדפת עם 3 קטגוריות לכל משתנה.
 ב. לא.
 ג. 4

שאלה 3:

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

שאלה 4:

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

שאלה 5:

נקבל H_0

שאלה 6:

א

שאלה 7:

ג

שאלה 8:

- א. אותה מסקנה
 ב. לא ניתן לדעת.

שאלה 3:

- א. יש קשר
 ב. 0.394
 ג. 0.208

פרק 22 - ניתוח שונות חד כיוונית

רקע תיאורטי:

ניתוח שונות (חד כיוונית) הוא מבחן להשוואת תוחלות (μ_1, \dots, μ_k) של k אוכלוסיות שונות. ולכן בניתוח שונות השערות המחקר הן:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (\text{התוחלות של כל האוכלוסיות שוות})$$

$$H_1 : \quad \text{אחרת} \quad (\text{לפחות שתיים מהתוחלות שונות})$$

התנחות הדרושות לביצוע התהליך הן:

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.

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

\bar{X}_1 - ממוצע המדגם הראשון, \dots, \bar{X}_k - ממוצע המדגם ה-k-י.

\bar{X} - ממוצע כללי (של כל המדגמים).

$$SS_B = \sum_{i=1}^k n_i [\bar{X}_i - \bar{X}]^2 \quad \text{סכום ריבועים בין הקבוצות}$$

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

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

$$SST = SSB + SSW$$

www.GooL.co.il לפתרון מלא בסרטון וידאו היכנסו ל-

© כתב ופתר - ברק קנדל

יש למלא את טבלת ניתוח השונות הבאה :

טבלת ניתוח שונות

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

$$F = \frac{SS_B / (k - 1)}{SS_W / (n - k)} \sim F(k - 1, n - k)$$

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

תרגילים:

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 1 קיבלה "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 :

| בית הספר | "המתמיד" | "רביץ" | "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?
4. שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות.
- כל אופה נדרש לאפות בכל יום 4 עוגות.

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

האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות? בדקו ברמת מובהקות של 5%.

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.

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 | | | |

מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

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

ANOVA

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

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

ANOVA

pulse

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

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.

Homogeneous Subsets

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.

- א. בדוק 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. נמק!
- ד. לטבלת ה ANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "טוקיי". ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

8. בעיר מסוימת רצו לבדוק האם קיים הבדל ברמה של התלמידים בין בתי הספר השונים בעיר. ביצעו מדגם מכל בית ספר ונתנו מבחן זהה לכל הנידגמים. לאחר מכן ריכזו את הנתונים בתוכנה סטטיסטית והפעילו ניתוח שונות. מצורפים הפלטים שהתקבלו.
ענו על הסעיפים הבאים:
- א. כמה בתי ספר יש בעיר?
ב. כמה תלמידים השתתפו בסך הכול במחקר?
ג. האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%
ד. בביטחון של 95% אילו בתי ספר שונים זה מזה ברמת התלמידים? נמקו והסבירו.

Oneway

ANOVA

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

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

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

פתרונות סופיים חלקיים - ניתוח שונות חד כיוונית

2. אם חישוב נכון ה F הסטטיסטי יוצא: 0.58

3. נדחה את השערת האפס.

4. להלן טבלת הניתוח השונות המתקבלת:

| | Sum of Squares | df | Mean Square | F |
|----------------|----------------|----|-------------|------|
| Between Groups | 126.000 | 2 | 63.000 | .756 |
| Within Groups | 750.000 | 9 | 83.333 | |
| Total | 876.000 | 11 | | |

5. נקבל את השערת האפס.

6. א. 165.5 ב. 2 ג. 178.725 ד. 9.375 ה. 18.36

7. א. נדחה את השערת האפס. ב. לא משתנה. ג. כן

8. א. 5 ב. 25 ג. כן