

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב



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



תוכן העניינים

1. מקדם המתאם (מדד קשר) הלינארי ומובהקותו..... 1
2. רגרסיה..... 19
3. מדדי קשר-רגרסיה -שונות מוסברת ושונות לא מוסברת..... 22
4. בדיקת השערות לתוחלת ההפרש במדגמים מזווגים..... 25
5. בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים..... 31
6. ניתוח שונות חד כיוונית..... 42
7. מבחנים אפרמטריים למדגמים מזווגים..... 51
8. מבחנים אפרמטריים למדגמים בלתי תלויים..... 58
9. גודל האפקט..... (ללא ספר).....
10. מבחני חי בריבוע..... 66
11. מדד הקשר של ספירמן..... (ללא ספר).....

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 1 - מקדם המתאם (מדד קשר) הלינארי ומובהקותו

תוכן העניינים

1. מקדם המתאם הלינארי (פירסון)..... 1

2. ניתוח פלטים על מקדם המתאם הלינארי..... 12

מקדם המתאם (מדד קשר) הלינארי ומובהקותו

מדד הקשר הלינארי (פירסון) – מבוא

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

דוגמה (פתרון בהקלטה) :

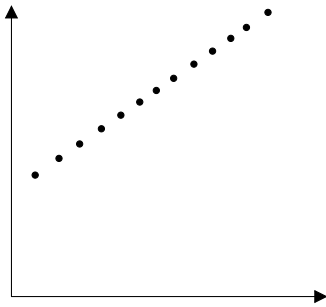
בבניין 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגרות בדירה. להלן התוצאות שהתקבלו :

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

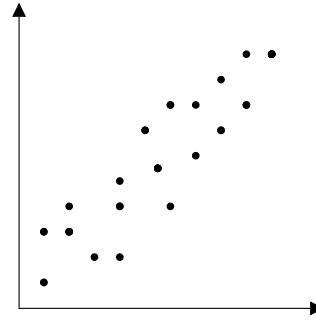
- (1) כמה תצפיות ישנן בדוגמה?
- (2) כמה משתנים ישנם בדוגמה, מי הם?
- (3) שרטטו לנתונים דיאגרמת פיזור.
- (4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

דיאגרמות פיזור לקשר בין משתנים וניתוחם

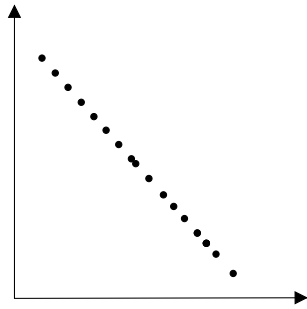
קשר לינארי חיובי מלא



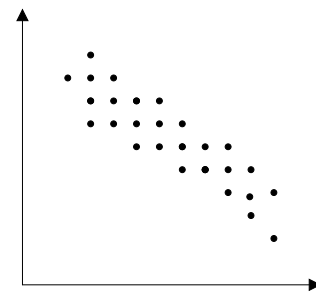
קשר לינארי חיובי חלקי



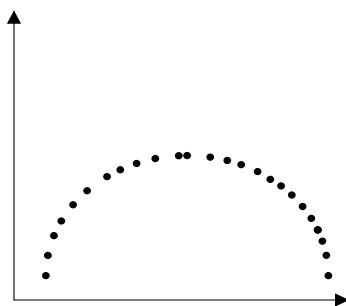
קשר לינארי שלילי מלא



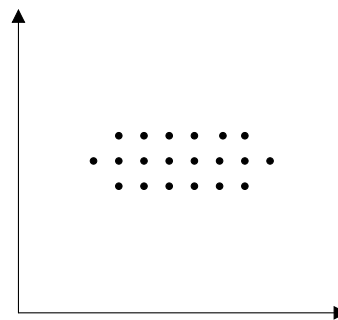
קשר לינארי שלילי חלקי



אין קשר לינארי



אין קשר



משמעות מקדם המתאם:

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

-1

0

1

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

מתאם חיובי מלא (מקדם מתאם 1):

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

מתאם חיובי חלקי:

ככל שמשנתנה 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 שמשנתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט. ככל שמקדם המתאם קרוב לאפס עוצמת הקשר יותר חלשה וככל שהמדד רחוק יותר מהאפס העוצמה יותר חזקה. לסיכום, מקדם המתאם בודק את עוצמת הקשר הלינארי, ואת כיוון הקשר.

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

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

r - מדד הקשר הלינארי במדגם שמהווה אומדן לפרמטר ρ .

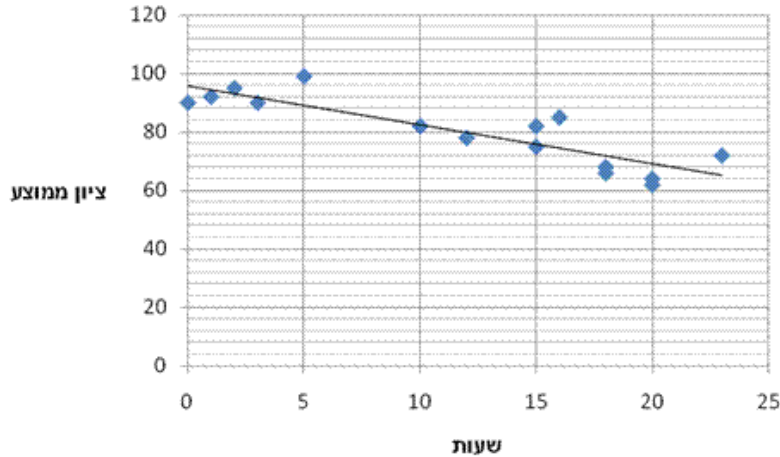
קיומו של מתאם בין שני משתנים אינו מצביע על סיבתיות בהכרח. למשל, אם נמצא מתאם חיובי בין כמות הסוכרזית שאדם אוכל לבין במשקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי, כלומר אם נחליף את X ב- Y התוצאה תהיה זהה.

דוגמה (פתרון בהקלטה):

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

שאלות

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



- א. מיהו המשתנה הבלתי תלוי?
- ב. מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר? מה ניתן להגיד על עוצמת הקשר?

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

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

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

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

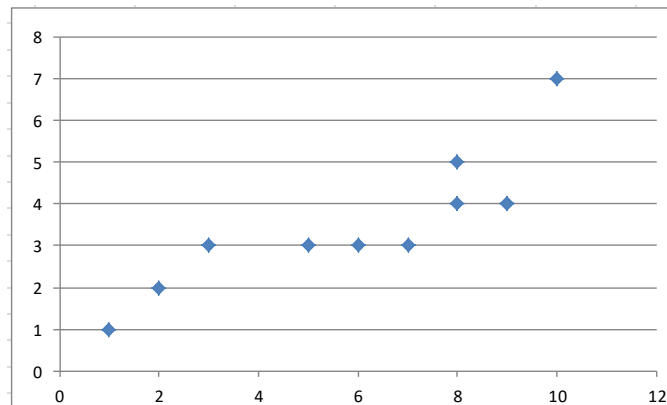
| שעות תרגול | ציון סופי |
|------------|-----------|
| 20 | 90 |
| 25 | 90 |
| 30 | 95 |
| 15 | 60 |
| 30 | 90 |
| 20 | 85 |
| 10 | 50 |

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

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

- א. השלימו את המשפט:
בחודש ינואר ככל שהטמפרטורה היומית נוטה לרדת, כך כמות המשקעים נוטה _____.
- ב. הוחלט להעביר את הטמפרטורה למעלות פרנהייט על מנת שיוכלו להשוות אותה לנתונים מארה"ב. נוסחת המעבר היא $F^0 = 32 + \frac{9}{5}C^0$.
כיצד הדבר ישפיע על מקדם המתאם בין הטמפרטורה במעלות פרנהייט לכמות המשקעים במ"מ?

5) להלן דיאגרמת פיזור המראה קשר בין שני משנים:

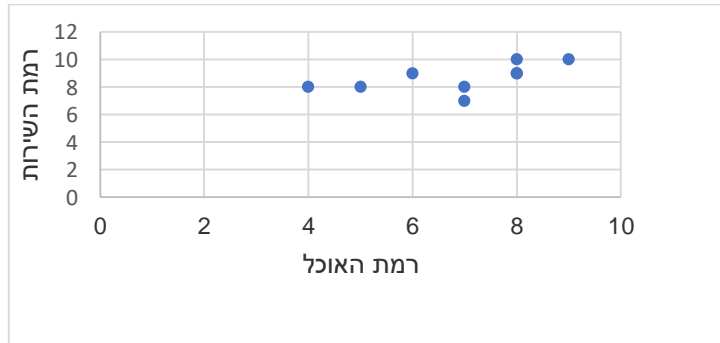


- א. השלימו: ניתן לראות שהקשר הוא לינארי _____ (מלאו חלקי) כיוון הקשר הוא (חיובי/שלילי).
- ב. השלימו: אם היינו מוסיפים תצפית שערך ה- X שלה הוא 4 וערך ה- Y שלה הוא 7, מקדם המתאם של פירסון היה _____ (גדלו קטן/לא משתנה).

שאלות רב ברירה (יש לבחור את התשובה הנכונה):

- 6) חוקר אקלים דגם כמה ימים בשנה ומדד את הטמפרטורה בטורונטו שבקנדה ואת הטמפרטורה בסידני שבאוסטרליה באותו היום. הוא חישב ומצא מקדם מתאם שלילי בין הטמפרטורה היומית בטורונטו לבין הטמפרטורה היומית בסידני. משמעות מקדם המתאם השלילי במדגם:
- א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני בימים שנדגמו.
ב. במדגם, רוב הטמפרטורות בטורונטו היו שליליות.
ג. ההפרש בין הטמפרטורה בטורונטו לבין הטמפרטורה באוסטרליה, במדגם זה, הוא שלילי.
ד. במדגם יש נטייה שהטמפרטורה יורדת בטורונטו לטמפרטורה לעלות בסידני.

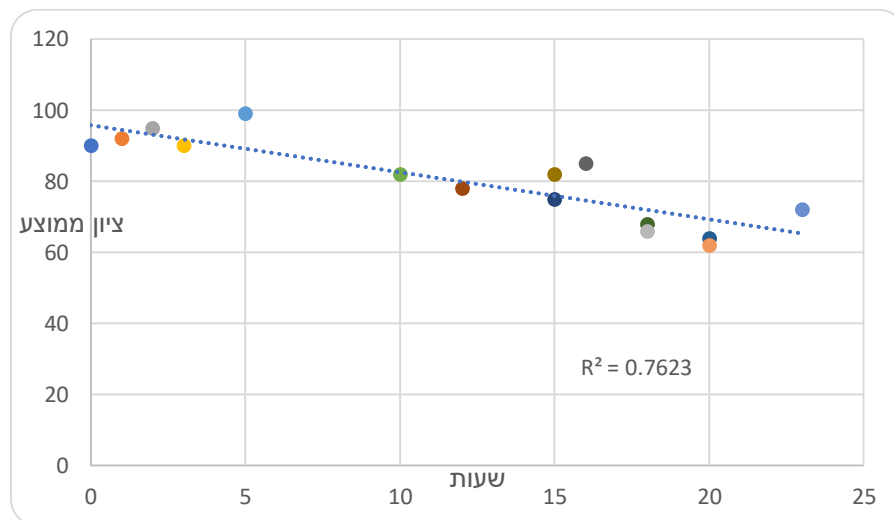
- 7) בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



מה יהיה ערכו של מקדם המתאם (r)?

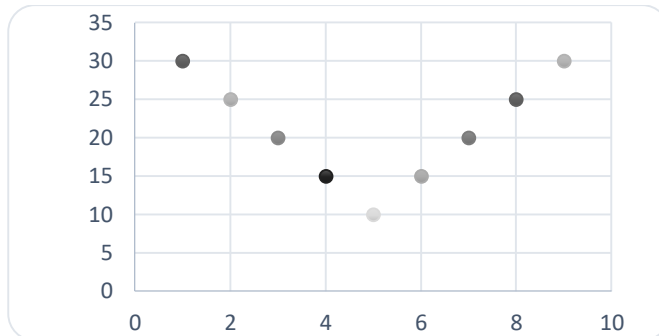
- א. $r = -0.3$
 ב. $r = 0$
 ג. $r = 1.125$
 ד. $r = 0.593$

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



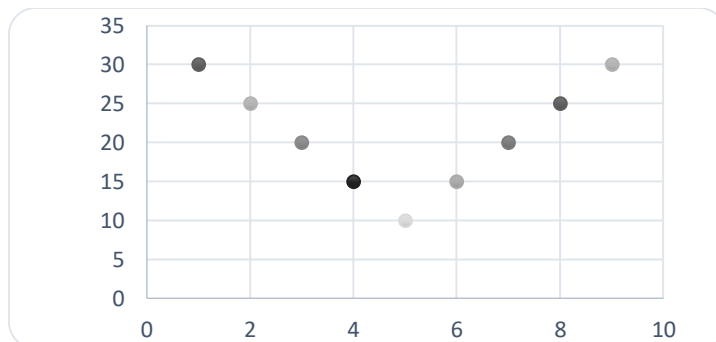
- מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?
- א. ככל שמבלים יותר הציון נוטה לרדת.
 ב. אין קשר בין שעות הבילוי לציון.
 ג. ככל שמבלים פחות הציון נוטה לרדת.
 ד. ככל שהציון נוטה לרדת הסטודנט מבלה פחות.

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



- א. $r = 1$ היות ושני המשתנים יוצרים קוים ישרים.
 ב. $r = 2$ היות ויש שני קוים בעלי קשר מושלם.
 ג. $r = 0$ היות והקו יורד ואחר כך עולה באותו האופן.
 ד. $r = \pm 1$ היות ויש קו עולה וגם קו יורד.

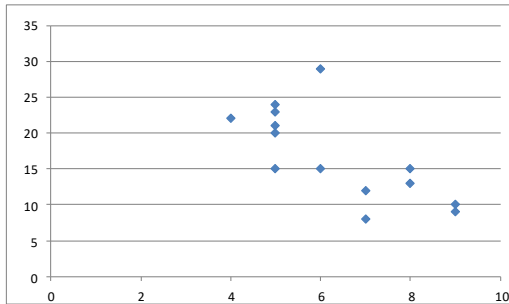
10) התרשים הבא מתאר דיאגרמת פיזור.



איזו טענה נכונה?

- א. בתרשים מוצג הקשר בין שני משתנים.
 ב. בתרשים מוצג הקשר בין 9 משתנים.
 ג. בתרשים מוצג הקשר בין 10 משתנים.
 ד. אין לדעת כמה משתנים מוצגים בתרשים.

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים:



X - (משתנה בלתי תלוי בציר האופקי)

Y - (משתנה תלוי).

במדגם התקבל $r^2 = 0.52$.

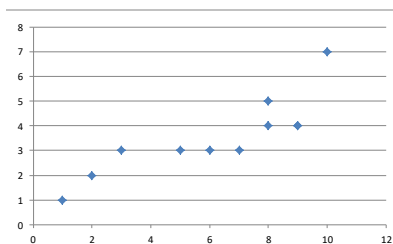
11) לאור הנתונים המופיעים בדיאגרמה, איזה מבין הערכים הבאים מתאים להיות התוצאה של r ?

- א. -0.52
- ב. 0.72
- ג. -0.72
- ד. 0.52

12) אם מקדם המתאם בין שני משתנים הוא 1, אזי:

- א. הערכים של המשתנים הם חיוביים.
- ב. עבור כל תצפית ערך של משתנה אחד שווה לערך של המשתנה השני.
- ג. הקשר הלינארי הוא בעוצמה חזקה.
- ד. אף אחת מהתשובות לא בהכרח נכונה.

13) להלן דיאגרמת פיזור:



מה יהיה מקדם המתאם בין שני המשתנים?

- א. 1
- ב. 0.85
- ג. 0.15
- ד. 0

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

- א. קיימת נוסחה לינארית הקושרת בין כל התצפיות.
- ב. לא קיים קשר בין שני המשתנים.
- ג. ככל שמשתנה אחד נוטה לרדת גם לשני יש נטייה לרדת.
- ד. קיים קשר בין שני המשתנים, אך לא ניתן לדעת מאיזה סוג.

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

- א. חיובי
- ב. שלילי
- ג. אפסי
- ד. לא ניתן לדעת.

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

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

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

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

- א. -1
- ב. 1
- ג. -0.5
- ד. 0.5

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

- א. ככל שהדירה גדולה יותר בשטחה כך היא יקרה יותר.
- ב. ככל שהדירה קטנה יותר בשטחה כך היא זולה יותר.
- ג. לא קיים קשר בין שטח הדירה למחיר הדירה.
- ד. מצב כזה שמתואר הנתונים לא אפשרי.

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

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

- 19) נמצא מתאם חיובי בעוצמה גבוהה בין X – ציון בבגרות בלשון ל Y – ציון בבגרות במתמטיקה. אילו מהמשפטים הבאים נכון?
- א. ניתן לומר שאחת מהסיבות להבדלים שיש לסטודנטים במתמטיקה נובעים מההבדלים שיש להם בלשון.
- ב. קיימת נוסחה של קו ישר שקושרת בין ציון בבגרות במתמטיקה לציון בבגרות בלשון.
- ג. ללא יוצא מן הכלל, ניתן להגיד שכל תלמיד שמצליח יותר מתלמיד אחר בלשון גם יצליח יותר מאותו תלמיד במתמטיקה.
- ד. אף אחד מהטענות שהוצגו אינה בהכרח נכונה.

- 20) עבור סדרה של תצפיות מדדו את X ואת Y . נמצא שעבור כל התצפיות שהערך של Y ירד הערך של X בהכרח ירד ללא יוצא מן הכלל. מקדם המתאם של פירסון יהיה בהכרח:
- א. 1
- ב. -1
- ג. 0
- ד. אף אחת מהתשובות.

תשובות סופיות

- (1) א. שעות בילוי.
 ב. הקשר חלקי, כיוון הקשר שלילי.
 (2) א. להלן טבלה:

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

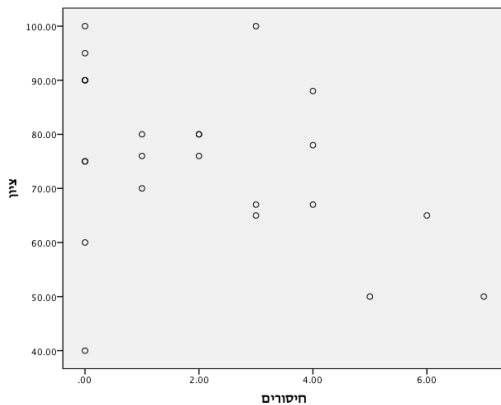
- (3) א. ב"ת- מס' שעות התרגול, תלוי- ציון.
 ג. קשר לינארי חיובי חלקי.
 (4) א. לעלות.
 (5) א. חלקי, חיובי.
 (6) ד'. (7) ד'. (8) א'. (9) ג'. (10) א'.
 (11) ג'. (12) ד'. (13) ב'. (14) א'. (15) א'.
 (16) א'. (17) ד'. (18) א'. (19) ד'. (20) ד'.

מדד הקשר הלינארי – ניתוח פלטים – רקע

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

השערת האפס: תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים: $H_0: \rho = 0$. ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית. **דוגמה (פתרון בהקלטה):**

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



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

Correlations

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

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

שאלות

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

Correlations

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

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

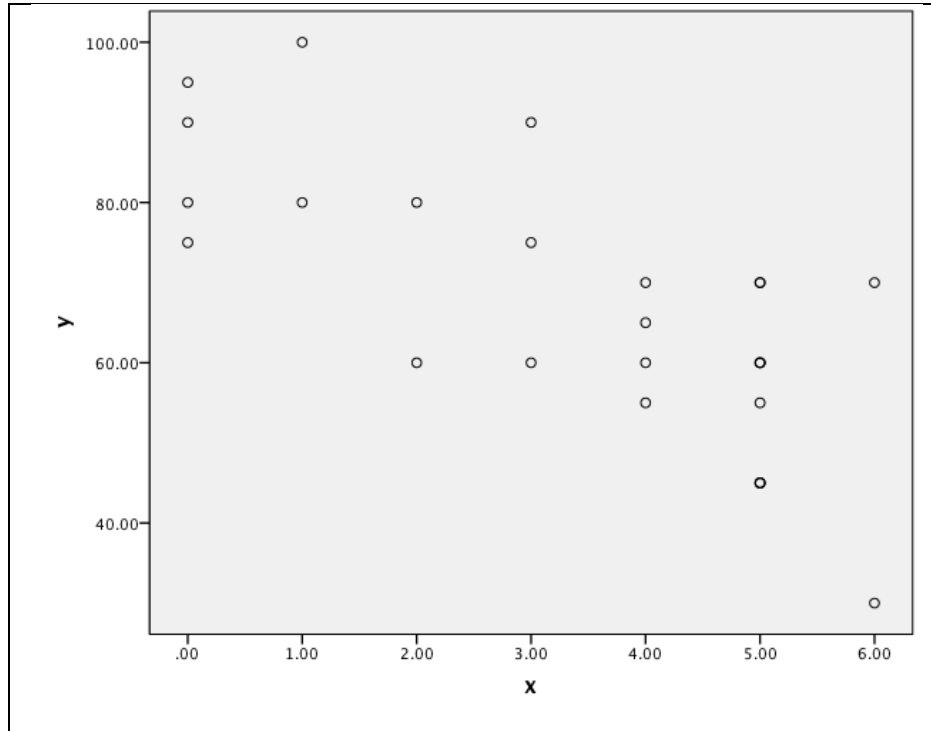
2) במדגם שנעשה נבדקו מספר משתנים על התצפיות שנדגמו. להלן פלט שהופק על המדגם:

Correlations

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

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

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



Correlations

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

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

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

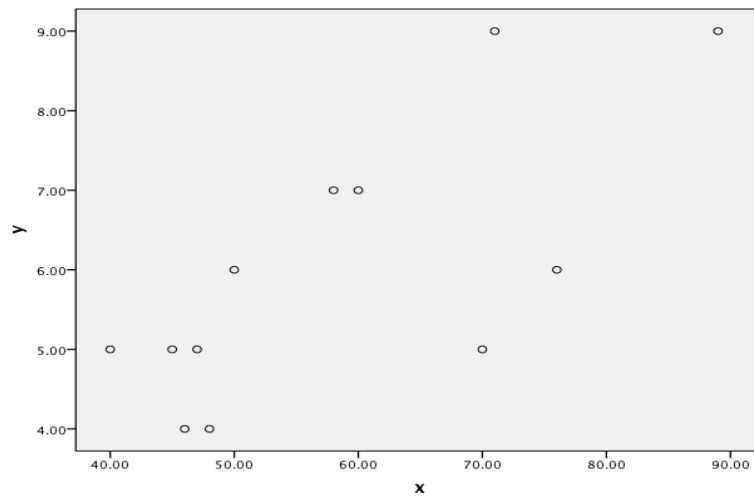
4) נדגמו ילדים בגיל 8 ונבדק עבור כל ילד גובהו בס"מ ומשקלו בק"ג. להלן הפלט שהתקבל עבור תוצאות המדגם:

Correlations

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

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

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



Correlations

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

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

- א. מה ניתן להגיד על סמך הפלט על הקשר שנימצא במדגם בין הטמפרטורה בתהליך לאחוז החומר?
- ב. האם הקשר בין הטמפרטורה בתהליך לבין אחוז החומר הוא קווי חיובי מובהק? בדקו ברמת מובהקות של 5%.
- ג. מה היה קורה למקדם המתאם במדגם ומובהקות התוצאה אם הייתה מתווספת תצפיות שבה הטמפרטורה היא 40 ואחוז החומר 9?

תשובות סופיות

- (1) א. נשים בהריון. ב. $H_0 : p = 0$
 $H_1 : p \neq 0$
- ג. משתנה תלוי – רמת ההמוגלובין, משתנה בלתי תלוי- גיל.
 ד. קיים קשר לינארי בין גיל האישה בהריון לרמת ההמוגלובין שלה בדם.
- (2) א. בין X ל- Y . ב. X ו- Y . כמו כן, W ו- Y . ג. ראה וידאו.
 ד. לא.
- (3) א. משתנה תלוי – ציון בבגרות במתמטיקה, משתנה בלתי תלוי- שעות צפייה.
 ב. כיוון שלילי ועוצמה של 0.741. ג. כן. ד. לא.
- (4) א. נדחה את H_0 . ב. לפחות 0.032. ג. לא ניתן לדעת.
- (5) א. קיים קשר לינארי חיובי וחלקי שעוצמתו: 0.732. ב. נדחה את H_0 .
 ג. מקדם המתאם קטן ומובהקות התוצאה גדלה.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 2 - רגרסיה

תוכן העניינים

1. כללי 19

מדדי קשר – רגרסיה ליניארית:

רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבוי. לבנות קו ניבויים הנקרא גם קו רגרסיה המנבא משתנה אחד על סמך האחר. מדובר בקו שמנבא את Y על סמך X . השיטה למציאת הקו הנ"ל נקראת שיטת הריבועים הפחותים והקו המתקבל נקרא קו הרגרסיה או קו הניבויים או קו הריבועים הפחותים. a - נותן את ערך Y כאשר X הנו אפס על גבי קו הניבויים. הוא נקרא החותך של הקו. b - הוא שיפוע הקו נותן בכמה בעצם Y משתנה כאשר X גדל ביחידה אחת על גבי קו הניבויים.

להלן המשוואות למציאת הפרמטרים של קו הרגרסיה: $Y = bX + a$, $b = r \frac{S_y}{S_x}$.

לצורך בניית קו ניבויים לניבוי X על סמך Y נצטרך לעדכן את הנוסחאות בהתאם.

שאלות:

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

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

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

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

- א. חשבו את מדד הקשר הלינארי בין X ל- Y . מיהו המשתנה התלוי?
 ב. מצאו את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבירו את משמעות הפרמטרים של קו הרגרסיה.
 ג. משפחת כהן הכניסה 15,000₪. מה ההוצאה הצפויה שלה?

- (2) נסמן ב- X את ההשכלה של אדם בשנות לימוד. נסמן ב- Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_x = 2, \quad S_y = 5, \quad \bar{X} = 14, \quad \bar{Y} = 8, \quad \text{COV}(X, Y) = 7.5$$

- א. חשבו את מדד הקשר של פירסון בין ההשכלה להכנסה.
 ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?
 ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000₪?

- (3) חוקר רצה לחקור את הקשר הקווי שבין הציון המבחן בסטטיסטיקה לבין מספר שעות ההכנה של הסטודנטים למבחן. במדגם של 100 סטודנטים שנבחנו בקורס נרשמו התוצאות הבאות: הציון הממוצע של הסטודנטים היה 65 עם סטיית תקן של 27. מספר שעות ההכנה הממוצע היה 30 עם סטיית תקן של 18. מקדם המתאם בין הציון לשעות ההכנה היה 0.8.

- א. על פי משוואת הרגרסיה, שעת הכנה נוספת משפרת את ציון המבחן ב-?
 ב. על פי משוואת הרגרסיה, תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון?
 ג. מהו קו הרגרסיה לניבוי הציון לפי שעות ההכנה?

- (4) נתונים 2 משתנים X ו- Y . כמו כן נתון: $S_x = S_y = 4, \bar{X} = 1.5$.
 וכן שקו הרגרסיה של Y על בסיס X הינו: $Y = -0.2X + 0.5$.
 חשבו מהו מקדם המתאם בין X ל- Y .

תשובות סופיות:

- | | | |
|----------------------|-----------------------|-------------|
| ג. 12.4 אלפי ₪. | ב. $Y = 0.8X + 0.4$. | א. 0.8 (1) |
| ג. 14.6 שנים. | ב. 4.25 אלפי ₪. | א. 0.75 (2) |
| ג. $Y = 1.2X + 29$. | ב. 29. | א. 1.2 (3) |
| | | א. -0.2 (4) |

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 3 - מדדי קשר-רגרסיה -שונות מוסברת ושונות לא מוסברת

תוכן העניינים

1. כללי 22

מדדי קשר – רגרסיה – שונות מוסברת ושונות לא מוסברת:

רקע:

המטרה ברגרסיה היא להסביר את השונות של המשתנה התלוי. למשל, להסביר את השונות של המשכורת באמצעות הוותק או להסביר את השוני בציונים באמצעות כמות החיסורים.

r^2 - החלק מהשונות של המשתנה התלוי מוסבר. השונות המוסברת נקראת גם שונות ניבויים. השונות הלא מוסברת נקראת גם שונות טעויות.

שאלות:

- (1) נמצא קשר חיובי בעוצמה של 0.7 בין שטח דירה למחירה. כמו כן, נתון שסטיית התקן של מחירי הדירות הינה 200.
- איזה אחוז מהשונות של מחירי הדירות מוסבר על ידי שטח הדירה?
 - איזה אחוז מהשונות של מחירי הדירות לא מוסבר על ידי שטח הדירה?
 - מהי השונות המוסברת ומהי השונות הלא מוסברת של מחירי הדירות?
- (2) להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו!
- אם שונות הטעויות שווה ל-0 (השונות הלא מוסברת) אז מקדם המתאם של פירסון יהיה 1.
 - אם מקדם המתאם של פירסון בין שני משתנים הוא 1 אזי שונות הטעויות (השונות הלא מוסברת) תהיה 0.
 - אם השונות המשותפת של X ושל Y היא 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות רב-ברירה:

- (3) בקשר בין שני משתנים התקבל: $r^2 = 0.64$, לכן:

- ללא יוצא מן הכלל ככל שערכי משתנה אחד עולה השני יעלה.
- 64% מהשונות של משתנה אחד מוסבר על ידי המשתנה השני.
- הקשר בין שני המשתנים הוא בעוצמה של 0.64.
- כל התשובות נכונות.

- (4) אם מגדילים את r^2 , ניתן לומר כי:

- אחוז השונות המוסברת יקטן.
- אחוז השונות המוסברת יגדל.
- אחוז השונות המוסברת יישאר ללא שינוי.
- סטיית התקן משתנה.
- לא ניתן לדעת.

- (5) בקורס מבוא לכלכלה ניתנו במשך השנה שני מבחנים : מבחן בסוף סמסטר א' X ומבחן בסוף סמסטר ב' Y . כאשר בנו את קו הרגרסיה של הציון במבחן סוף סמסטר ב' לפי הציון במבחן סוף סמסטר א' התקבלה שונות טעויות של 80, ושונות ניבויים של 20.
- לפי נתונים אלו, מקדם המתאם בין הציון במבחן סוף סמסטר א' לבין הציון במבחן סוף סמסטר ב' הוא :
- א. 0.44 .
 ב. - 0.44 .
 ג. עוצמת ההקשר הלינארי היא 0.44, אך אין אפשרות לדעת את סימנה.
 ד. אין אפשרות לחשב את מקדם המתאם.
 ה. 0.35 .

תשובות סופיות:

- (1) א. 49% . ב. 51% .
 ג. שונות מוסברת : 19,600, שונות לא מוסברת : 20,400 .
- (2) א. לא נכון . ב. נכון . ג. נכון .
- (3) ב' .
- (4) ב' .
- (5) ג' .

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

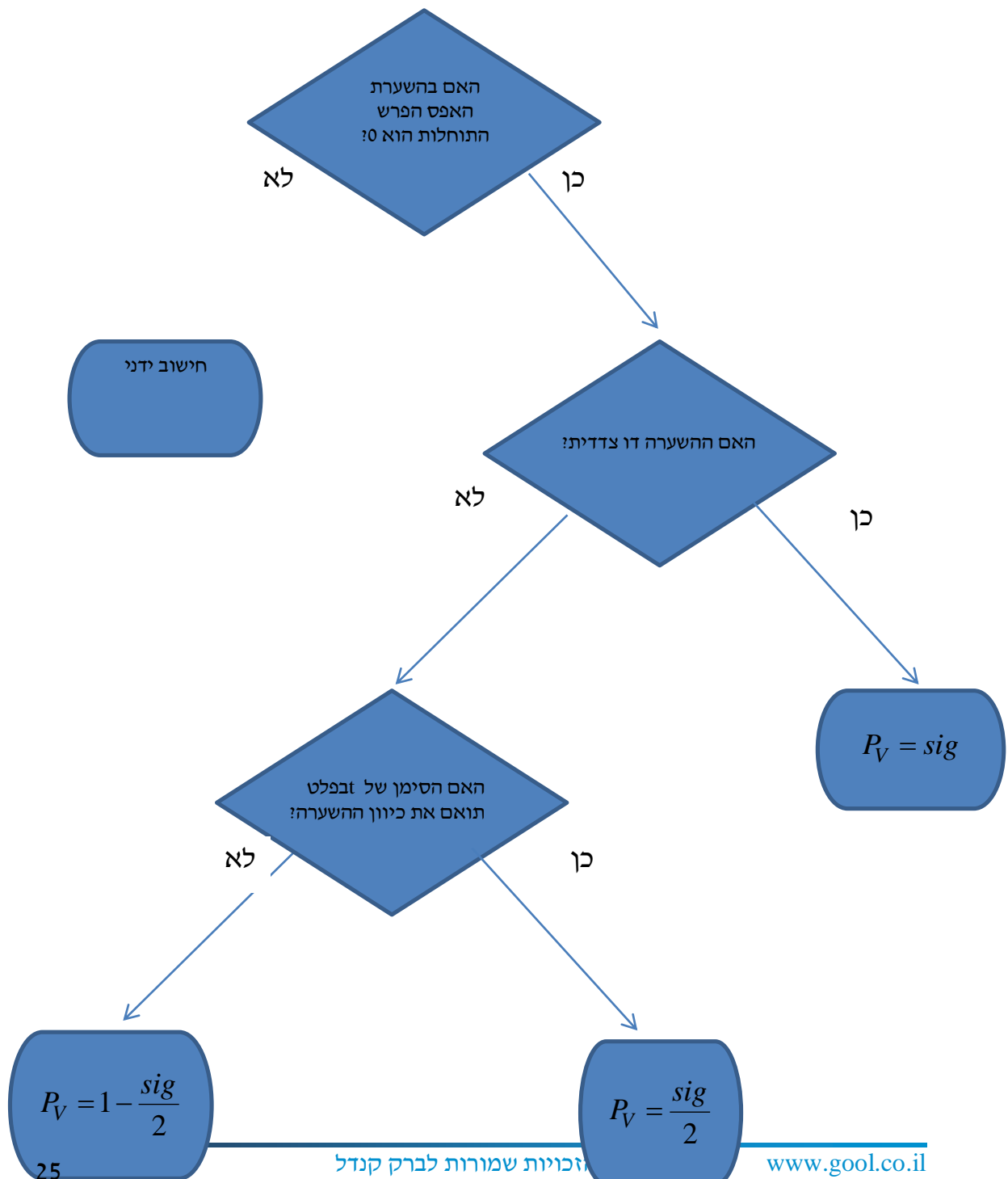
פרק 4 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

תוכן העניינים

1. ניתוח פלטים.....25

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

מדגמים מזווגים – ניתוח פלטים – רקע



דוגמה (פתרון בהקלטה) :

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

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

Paired Samples Statistics

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

Paired Samples Test

| | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|-----------------------|--------------------|----------------|-----------------|---|----------|-------|----|-----------------|
| | 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 |

בדקו ברמת מובהקות של 5% האם הקורס יעיל.

שאלות

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

| Paired Samples Test | | | | | | | | | |
|---------------------|---|--------------------|----------------|-----------------|---|-------|-------|-----|-----------------|
| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| | | 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. תנו אומדן להפרש הממוצעים.
 ב. תנו אומדן לטעות התקן של הפרש הממוצעים.
 ג. האם קיים הבדל מובהק בין השכלת האבות להשכלת האימהות ברמת מובהקות של 5%?

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

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

| Paired Samples Test | | | | | | | | | |
|---------------------|---------------------|--------------------|----------------|-----------------|---|---------|---------|-----|-----------------|
| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| | | 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. בדקו את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.
 ג. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים?
 ד. בנו את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.

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

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

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

(4) להלן פלט של תכנת SPSS:

T-Test

Paired Samples Statistics

| | Mean | N | Std. Deviation | Std. Error Mean |
|----------|---------|---|----------------|-----------------|
| Pair 1 x | 54.0000 | 6 | 5.86515 | 2.39444 |
| Pair 1 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%.

תשובות סופיות

- (1) א. -0.007 ב. 0.1 ג. אין הבדל מובהק.
- (2) א. $d.f = 299$ ב. $n = 300$ ג. $\bar{X} = 8.496$, $\text{Sig} = 0$.
- (3) א. ראה וידאו. ב. נדחה את H_0 . ג. לא נדחה את H_0 .
- ד. (1.592, 13.317).
- (4) א. 1.5876, 11.5715 ב. 0.173 ג. יגדל.
- ד. 0.0865 ה. 0.9135 ו. $49.18 < \mu < 58.82$.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

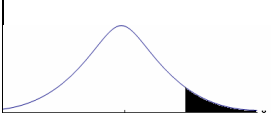
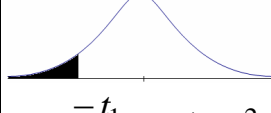
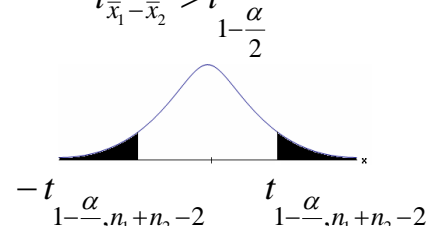
פרק 5 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסיה לא ידועות ומניחים שהן שוות..... 31
2. ניתוח פלטים..... 35

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

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

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

סטטיסטי המבחן:

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

השונויות המשוקללת:

חלופה אחרת לכלל הכרעה:

| נדחה H_0 אם מתקיים: | |
|---|---|
| $\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ | $\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ <p style="text-align: center;">או</p> $\bar{x}_1 - \bar{x}_2 < c - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ |
| $\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ | |

דוגמה (פתרון בהקלטה):

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשמשים בה כיום לבניית בניינים. לצורך בדיקת טענת המחקר נדגמו 10 יחידות של מתכות מהסוג הישן ו-12 יחידות של מתכות טמפרטורת ההתכה הממוצעת במתכת הישנה 1170 מעלות עם אומד חסר הטיה לשונות $S^2 = 200$.

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

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

שאלות

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

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

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

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

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

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

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

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

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

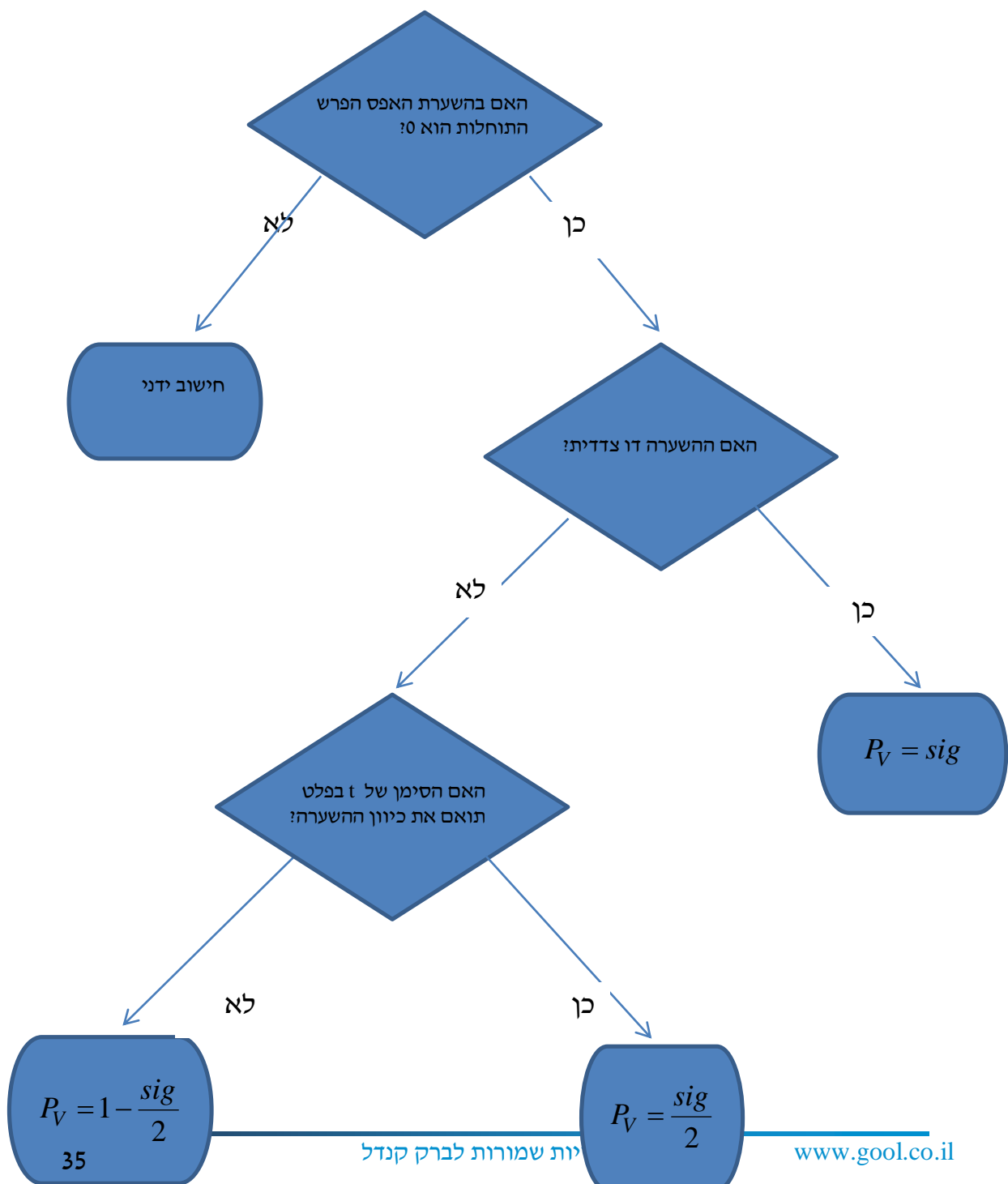
תשובות סופיות

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

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

ניתוח פלטים – רקע

מובהקות התוצאה על סמך הפלט:



דוגמה (פתרון בהקלטה) :

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

Group Statistics

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

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|--------------------------------|-----------------------------|---|------|------------------------------|---------|-----------------|-----------------|-----------------------|---|-------|
| | | 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?

שאלות

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

Group Statistics

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

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|----------|-----------------------------|---|------|------------------------------|----------|-----------------|-----------------|-----------------------|---|--------|
| | | 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:

Group Statistics

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

| | | 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 |

T-Test

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

- א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.
- ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?
- ג. האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של $\alpha = 0.05$.
- ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha = 0.01$.

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

T-Test

Group Statistics

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

Independent Samples Test

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

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

4) להלן פלט של תוכנת SPSS:

T-Test

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

Independent Samples Test

| | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|---------|
| | 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. השלימו את סימני השאלה בטבלה.
- ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של X לזה של Y ?
- ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X גדולה מהתוחלת של Y ?
- ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X קטנה מהתוחלת של Y ?

תשובות סופיות

- (1) א. נקבל את H_0 ונכריע שיש שוויון שוניות.
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.
 ג. 0.11076
 ד. $2.5665 \leq \mu \leq 2.6441$.
- (2) א. מבחן T להפרש ממוצעים במדגמים בלתי תלויים.
 ב. מבחן T למדגמים מזווגים.
 ג. נקבל את H_0 , נקבע לקיום שוויון שוניות.
 ד. נקבל את H_0 , לא נקבל את טענת הפסיכולוגים.
- (3) א. מבחן T להשוואת תוחלת במדגמים בלתי תלויים.
 ב. $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$ ג. כן.
- (4) א. 10.15, 4.69, -48
 ב. 0.04
 ג. 0.978
 ד. 0.022

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 6 - ניתוח שונות חד כיוונית

תוכן העניינים

1. כללי 42

ניתוח שונות חד כיוונית

רקע תיאורטי

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

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

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

ההנחות הדרושות לביצוע התהליך:

(2) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.

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

(4) המדגמים בלתי תלויים זה בזה.

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

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

\bar{X}_1 - ממוצע המדגם הראשון, \dots, \bar{X}_k - ממוצע המדגם ה- k .
 \bar{X} - ממוצע כללי (של כל המדגמים).

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

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

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

$$SST = SSB + SSW$$

יש למלא את טבלת ניתוח השונות הבאה:

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

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

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

שאלות

- (1) מחקר מעוניין להשוות בין שלוש תרופות לשיכוך כאבים במטרה לבדוק האם קיים הבדל בין התרופות מבחינת הזמן בדקות שלוקח עד שהתרופה משפיעה. לצורך הבדיקה נלקחו 15 אנשים שסובלים מכאבי ראש. אנשים אלה חולקו באקראי לשלוש קבוצה: קבוצה 1 קיבלה "אקמול" קבוצה 2 קיבלה "אופטלגין" קבוצה 3 קיבלה "נורופן". כל אדם במחקר מסר את מספר הדקות עד שהתרופה השפיעה עליו.
- מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר?
 - מהו המבחן הסטטיסטי המתאים כאן? רשמו את ההשערות.
 - 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:

| "הס" | "רבין" | "המתמיד" |
|------|--------|----------|
| 85 | 98 | 78 |
| 83 | 62 | 65 |
| 74 | 55 | 70 |
| 85 | 80 | 90 |
| 75 | | 56 |

- מהו המבחן הסטטיסטי המתאים? רשמו את ההשערות ואת ההנחות של המבחן.
- מהו גודל המדגם? מהו המשתנה הבלתי תלוי (factor) כמה רמות יש לו?
- חשבו את הממוצע ואת סטיית התקן של הציונים בכל אחד מהמדגמים.
- מלאו את טבלת ANOVA.
- רשמו את כלל ההכרעה למבחן שהוצע בסעיף א ברמת מובהקות של 5%.
- 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.

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

- רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.
- מה מסקנת המחקר ברמת מובהקות של 5%?
- האם יש צורך לבצע השוואות מרובות?

4) שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות.

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

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

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

ANOVA

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

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

ANOVA

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

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

| ANOVA | | | | | | pulse | | | |
|----------------|----------------|----|-------------|--------|------|------------------------|---|-------------------------|---------|
| pulse | | | | | | Tukey HSD ^a | | | |
| | Sum of Squares | df | Mean Square | F | Sig. | dosage | N | Subset for alpha = 0.05 | |
| | | | | | | | | 1 | 2 |
| Between Groups | 414.400 | 2 | 207.200 | 19.733 | .000 | 30.00 | 5 | 71.0000 | |
| Within Groups | 126.000 | 12 | 10.500 | | | 20.00 | 5 | | 80.2000 |
| Total | 540.400 | 14 | | | | 10.00 | 5 | | 83.4000 |
| | | | | | | Sig. | | 1.000 | .299 |

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

Post Hoc Tests

Multiple Comparisons

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

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

- א. בדקו ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.
- ב. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.
- ג. האם יש צורך במחקר בהשוואת מרובות. נמקו!
- ד. לטבלת ANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "טוקי". ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

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

תשובות סופיות

(1) א. משתנה בלתי תלוי : סוג התרופה. ב. ניתוח שונות חד כיווני

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

$$H_1 : otherwise$$

משתנה תלוי : הזמן עד להשפעת התרופה בדקות.

ג. 1. מדגמים בלתי תלויים.

2. שווין שונויות.

3. משתנים מתפלגים נורמלית.

(2) א. המבחן לניתוח שונות חד כיוונית.

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

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. משתנים מתפלגים נורמלית.

3. שווין שונויות.

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

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

ד. להלן טבלה :

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

ה. $F > 3.98$.

ו. נקבל את H_0 .

(3) א. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. ב. נדחה את H_0 . ג. כן.

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. שווין שונויות.

3. משתנים מתפלגים נורמלית.

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

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad (5)$$

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. שוויון שונות.

3. משתנים מתפלגים נורמלית.

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

6) להלן טבלה :

ANOVA

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

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad (7)$$

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. משתנים מתפלגים נורמלית.

3. שוויון שונות.

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

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

$$(8) \quad \text{א. 5 ב. 25}$$

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

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

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 7 - מבחנים אפרמטרים למדגמים מזווגים

תוכן העניינים

| | |
|----|--------------------------------|
| 51 | 1. מבחן ווילקוקסון- פלטים..... |
| 55 | 2. תרגול בזיהוי מבחנים..... |

ניתוח פלטי SPSS במבחן ווילקוקסון למדגמים מזווגים – רקע

מבחן זה לא דורש הנחה של התפלגות נורמלית, אולם דורש ערכים מספריים המאפשרים חישוב הפרש בין ערכי X לערכי Y . מבחן זה הוא הגרסה הלא פרמטרית למבחן t למדגם מזווג. נשתמש במבחן זה שיש משתנה כמותי שאינו מתפלג נורמלית או שיש משתנה מסולם סדר. נראה איך מנתחים פלט של תכנת SPSS במבחן זה על ידי דוגמה.

דוגמה (פתרון בהקלטה) :

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

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

להלן פלט שמתקבל מהרצת מבחן ווילקוקסון על הנתונים הללו:

Wilcoxon Signed Ranks Test

| | | Ranks | | |
|-----------|----------------|----------------|-----------|--------------|
| | | N | Mean Rank | Sum of Ranks |
| HMO-PHARM | Negative Ranks | 9 ^a | ??? | 51.00 |
| | Positive Ranks | 1 ^b | 4.00 | 4.00 |
| | Ties | 1 ^c | | |
| | Total | 11 | | |

a. HMO < PHARM

b. HMO > PHARM

c. HMO = PHARM

Test Statistics^a

| | HMO-PHARM |
|------------------------|---------------------|
| Z | -2.434 ^b |
| Asymp. Sig. (2-tailed) | .015 |
| Exact Sig. (2-tailed) | .014 |
| Exact Sig. (1-tailed) | .007 |

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

א. הסבירו מדוע מבחן ווילקוקסון למדגמים מזווגים מתאים למקרה זה?

ב. השלם את הערך החסר שמסומן בסימני שאלה בטבלה.

ג. מה תהיה מסקנת המחקר ברמת מובהקות של 5%?

שאלות

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

Wilcoxon Signed Ranks Test

| | | Ranks | | |
|--------|----------------|----------------|-----------|--------------|
| | | N | Mean Rank | Sum of Ranks |
| STAS2- | Negative Ranks | 2 ^a | 6.50 | 13.00 |
| STAS1 | Positive Ranks | 8 ^b | 5.25 | 42.00 |
| | Ties | 0 ^c | | |
| | Total | 10 | | |

a. STAS2 < STAS1

b. STAS2 > STAS1

c. STAS2 = STAS1

Test Statistics^a

| | STAS2- STAS1 |
|------------------------|---------------------|
| Z | -1.483 ^b |
| Asymp. Sig. (2-tailed) | .138 |

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

א. בדקו ברמת מובהקות של 5% את הטענה.
 ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם יוחלט לתת פקטור של 2 נקודות לכל הסטודנטים בשני המועדים?

2) מחקר בדק את רמת שביעות הרצון משירות לקוחות אחרי רפורמה שבוצעה בחברה. להלן תוצאות שהתקבלו כאשר שביעות הרצון הייתה בסקלה מ-1 ככל לא מרוצה ועד 5 מרוצה מאד.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---------------|
| 4 | 1 | 1 | 4 | 5 | 4 | 2 | 5 | 4 | 5 | 4 | 2 | Before |
| 5 | 5 | 5 | 4 | 3 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | After |

Wilcoxon Signed Ranks Test

| | | Ranks | | |
|--------------|----------------|--------|-----------|--------------|
| | | N | Mean Rank | Sum of Ranks |
| after-before | Negative Ranks | ?????a | 3.25 | ?????? |
| | Positive Ranks | ?????b | 4.92 | ?????? |
| | Ties | ?????c | | |
| | Total | 12 | | |

- a. after < before
- b. after > before
- c. after = before

Test Statistics^a

| | after-before |
|------------------------|--------------|
| Z | -1.622 |
| Asymp. Sig. (2-tailed) | .105 |

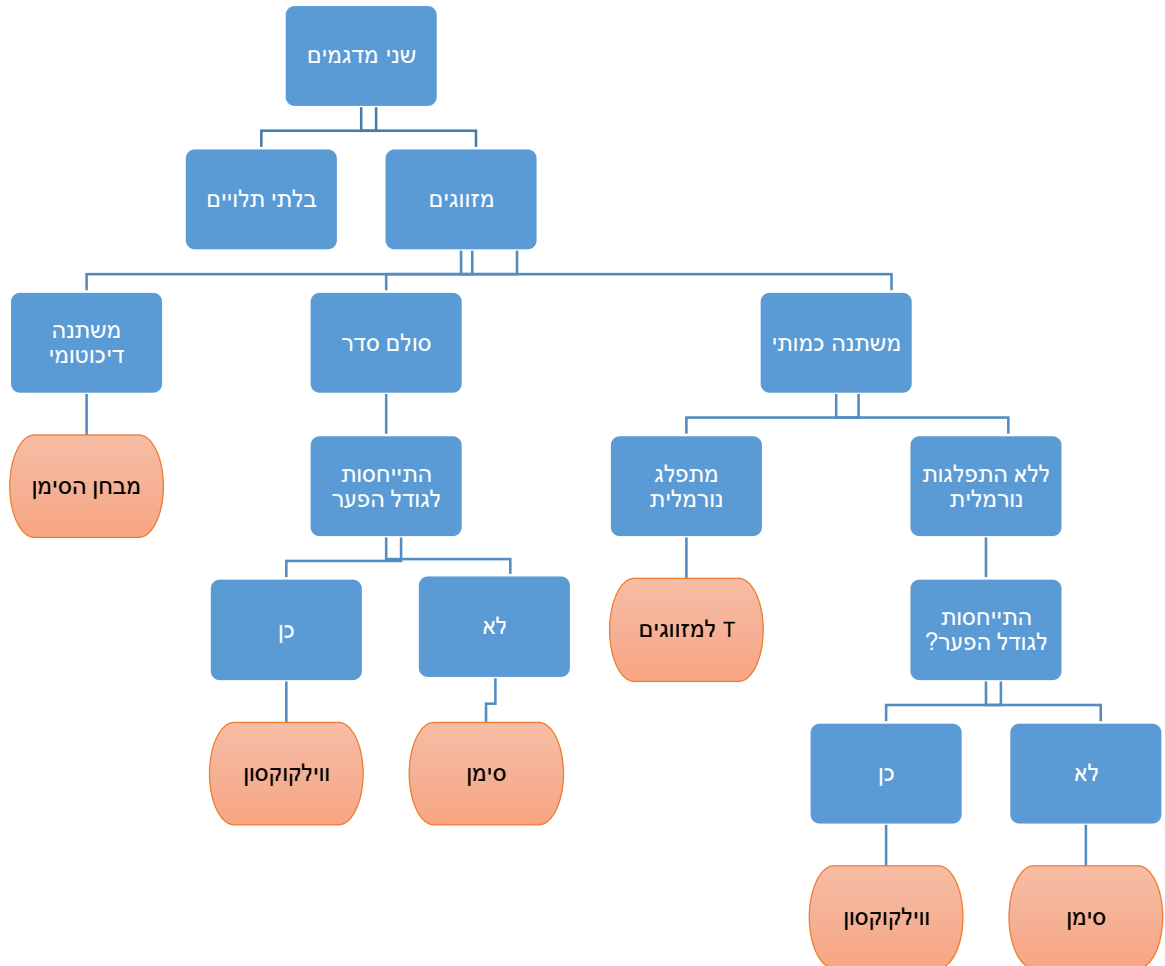
- א. השלימו את הערכים עם סימני השאלה אשר בטבלה.
- ב. בדקו ברמת מובהקות של 5% האם הרפורמה הייתה יעילה.
- ג. כיצד הייתה משתנה מובהקות התוצאה אם היו מוסיפים עוד 2 תצפיות שעבורן הפרש שביעות הרצון היה אפס?
- ד. כיצד הייתה משתנה מובהקות התוצאה של הפלט אם הינו מוסיפים תצפית כזו:

| before | after |
|--------|-------|
| 5 | 4 |

תשובות סופיות

- 1) א. לא נדחה H_0 . ב. לא תשתנה המסקנה.
- 2) א. 2, 6, 4, 6.5, 29.5 ב. לא נדחה H_0 . ג. לא משתנה. ד. גדלה.

זיהוי מבחנים סטטיסטיים – רקע



שאלות

(1) במטרה להשוות את רמת האפיייה של שני קונדיטורים בחרו 9 מאפים שונים (קרואסון, בראוני וכדומה) ונתנו לכל אחד משני הקונדיטורים לאפות את 9 המאפים השונים. 18 המאפים שנאפו ניתנו למומחה שנתן ציון למאפים השונים. הציון שניתן הוא בין 1 ל-5 לפי ניסיונו וטעמו האישי של המומחה. מהו המבחן המתאים ביותר במקרה זה?

- מבחן ווילקוקוסון.
- מבחן הסימן.
- מבחן T למזווגים.
- מבחן למדגמים בלתי תלויים.

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

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

מהו המבחן המתאים ביותר במקרה זה?

- מבחן ווילקוקוסון.
- מבחן הסימן.
- מבחן T למזווגים.
- מבחן למדגמים בלתי תלויים.

(3) במחקר בדקו לאנשים את רמת הסוכר בבוקר ואת רמת הסוכר בערב. מתוך 26 אנשים ל-3 רמת הסוכר הייתה זהה. ל-14 רמת הסוכר הייתה גבוהה יותר בשעות הערב. וליתר רמת הסוכר הייתה גבוהה יותר בשעות הבוקר. רוצים לבדוק ברמת מובהקות של 6% האם קיים הבדל בין רמת הסוכר בבוקר לרמת הסוכר בערב אצל האנשים. מהו המבחן המתאים ביותר במקרה זה?

- מבחן ווילקוקוסון.
- מבחן הסימן.
- מבחן T למזווגים.
- מבחן למדגמים בלתי תלויים.

- (4) חוקר מעוניין לבדוק את התפתחות היכולת לדחות סיפוקים מיידיים בקרב ילדים. לשם כך, הוא משתמש במבחן לבדיקה של דחיית סיפוקים, ומעביר אותו בו זמנית ל-2 קבוצות גיל. מבחן זה מודד כמה זמן (בשניות) מסוגל הילד לדחות קבלה של תגמול מיידי קטן על מנת לקבל תגמול גדול יותר בעתיד. התוצאות שמתקבלות הן הזמנים של הנחקרים בכל קבוצת גיל. מהו המבחן המתאים ביותר במקרה זה?
- מבחן ווילקוקוסון.
 - מבחן הסימן.
 - מבחן T למזווגים.
 - מבחן למדגמים בלתי תלויים.
- (5) חברת משקאות יצאה בקמפיין שנוי במחלוקת. החברה מעוניינת לבדוק האם הקמפיין השפיע על הרגלי הצריכה. במחקר השתתפו נשאלים האם הם נהגו לרכוש את המשקה לפני הקמפיין והאם הם רכשו אותו לאחר הקמפיין. מהו המבחן המתאים ביותר במקרה זה?
- מבחן ווילקוקוסון.
 - מבחן הסימן.
 - מבחן T למזווגים.
 - מבחן למדגמים בלתי תלויים.
- (6) מחקר התעניין בדפוסי שיחות הטלפון, שמנהל הפרט בעקבות פרידה מבן זוגו. במחקר השתתפו גברים ו-נשים (כולם נפרדו מבן זוגם). המשתתפים דיווחו על משך השיחות (בדקות; לפני ואחר הפרידה). שאלת המחקר בחנה האם פרידה מבן הזוג קשורה למשך השיחות (משך השיחות הוא משתנה שנהוג להתייחס אליו כמתפלג נורמלית). מהו המבחן המתאים ביותר במקרה זה?
- מבחן ווילקוקוסון.
 - מבחן הסימן.
 - מבחן T למזווגים.
 - מבחן למדגמים בלתי תלויים.

תשובות סופיות

- (1) א'
- (2) א'
- (3) ב'
- (4) ד'
- (5) ב'
- (6) ג'

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 8 - מבחנים אפרמטריים למדגמים בלתי תלויים

תוכן העניינים

1. מבחן מאן וויטני - שימוש בפלטים 58
2. מבחן קרוסקל ווליס - ניתוח פלטים 62

מבחנים אפרמטריים למדגמים בלתי תלויים

ניתוח פלטים במבחן מן-וויטני – רקע

מבחן מן-וויטני מיועד לבדוק האם לשתי אוכלוסיות התפלגות שווה. המבחן בוחן באופן רוחבי את כל תחום הערכים ולא מתמקד בערך מרכזי אחד. נשתמש במבחן זה כאשר יש שני מדגמים בלתי תלויים והמשתנה הכמותי הנחקר אינו מתפלג נורמלית או שמדובר במשתנה מסולם סדר. המבחן מתבסס על דירוג כל התצפיות. בעצם, מבחן זה הוא המענה האפרמטרי למבחן הפרמטרי להפרש תוחלות במדגמים בלתי תלויים.

דוגמה (פתרון בהקלטה) :

להלן תוצאות הערכות שקיבלו שני מורים: ד"ר A ופרופסור B. סטודנטים נתנו משוב כללי על המורים בסקלה של 1 (גרוע) עד 5 (מצוין). הטענה היא שד"ר A הוא מרצה טוב יותר מאשר פרופסור B.

Mann-Whitney Test

Ranks

| | teacher | N | Mean Rank | Sum of Ranks |
|-------|---------|----|-----------|--------------|
| grade | dr A | 17 | 25.00 | 425.00 |
| | prof B | 20 | 13.90 | 278.00 |
| | Total | 37 | | |

Test Statistics^a

| | grade |
|--------------------------------|-------------------|
| Mann-Whitney U | 68.000 |
| Wilcoxon W | 278.000 |
| Z | -3.249 |
| Asymp. Sig. (2-tailed) | .001 |
| Exact Sig. [2*(1-tailed Sig.)] | .001 ^b |

a. Grouping Variable: teacher

b. Not corrected for ties.

- הסבירו מדוע נעשה כאן מבחן מן-וויטני.
- מה המסקנה ברמת מובהקות של 1%?

שאלות

1) מחקרים טוענים שקיים הבדל בין שעות השינה של גברים לשעות השינה של נשים. כיוון שלא ניתן להוכיח ששעות שינה הינו משתנה המתפלג נורמלית ביצעו מבחן מן-וויטני בו לקחו נשים וגברים אקראיים ובדקו את שעות השינה שלהם.

| Ranks | | | | |
|-----------|--------|-----|-----------|--------------|
| | gender | N | Mean Rank | Sum of Ranks |
| sleeptime | male | 10 | ??? | 135.00 |
| | female | ??? | 7.50 | 75.00 |
| | Total | 20 | | |

Mann-Whitney Test

| Test Statistics ^a | |
|------------------------------|-----------|
| | sleeptime |
| Mann-Whitney U | 20.000 |
| Wilcoxon W | 75.000 |
| Z | -2.319 |
| Asymp. Sig. (2-tailed) | .020 |

- א. השלימו את סימני השאלה החסרים בפלט.
- ב. מהי המסקנה ברמת מובהקות של 5%?
- ג. מה הייתה מובהקות התוצאה אם טענת המחקר הייתה שגברים ישנים יותר מנשים?

2) שני אנשים נתבקשו לבדוק את מספר תאונות הדרכים בשבוע בשני קטעי כביש שונים. כל אחד בחר את השבועות באופן אקראי ובלתי תלוי באחר וספר כמה תאונות היו בכל כביש בשבוע. הפלטים שהתקבלו:

מספר תאונות

| | Frequency |
|-------|-----------|
| .00 | 2 |
| 1.00 | 1 |
| 2.00 | 2 |
| 3.00 | 1 |
| Total | 6 |

a. road = 1.00

מספר תאונות

| | Frequency |
|-------|-----------|
| .00 | 2 |
| 1.00 | 1 |
| 2.00 | 1 |
| 3.00 | 1 |
| 4.00 | 1 |
| 5.00 | 1 |
| Total | 7 |

a. road = 2.00

Ranks

| | road | N | Mean Rank | Sum of Ranks |
|-------------|-------|---|-----------|--------------|
| מספר תאונות | 1.00 | A | B | C |
| | 2.00 | D | E | F |
| | Total | G | | |

Test Statistics^a

| | VAR00002 |
|------------------------|----------|
| Asymp. Sig. (2-tailed) | .465 |

- א. השלימו בטבלה השלישית את המספרים החסרים במקום האותיות.
- ב. מהי מובהקות התוצאה לבדיקת ההשערה שאין הבדל בין הכבישים מבחינת התפלגות תאונות הדרכים?
- ג. כיצד הייתה משתנה התשובה של הסעיף הקודם אם כל חוקר היה מוסיף נתונים על שבוע נוסף לכל כביש?
- ד. מה המסקנה ברמת מובהקות של 5%?
- ה. מהי מובהקות התוצאה לבדיקת ההשערה שכביש מספר 1 עם התפלגות תאונות גבוהה יותר מאשר כביש מספר 2?

תשובות סופיות

- (1) א. 10, 13.5 ב. נדחה H_0 ג. 0.01
- (2) א. 6, 6.17, 7, 7.71, 13, 54 ב. 0.465
- ג. לא ניתן לדעת. ד. לא נדחה H_0 ה. 0.7675

מבחן קרוסקל וואליס – ניתוח פלטים – רקע

כאשר לא ניתן לבצע מבחן ניתוח שונות חד כיווני (למשל, כאשר השונויות לא שוות, המשתנה התלוי אינו מתפלג נורמאלי או המשתנה התלוי נמדד בסולם סדר) נבחר במבחן לא-פרמטרי להשוואת k קבוצות בלתי תלויות.

מבחן Kruskal-Wallis הוא מבחן אפרמטרי להשוואת מספר אוכלוסיות בלתי תלויות. מבחן מן-וויטני הינו מקרה פרטי של מבחן Kruskal-Wallis כאשר יש שתי קבוצות בלתי תלויות.

השערות המבחן:

- השערת האפס: אין הבדל בין הקבוצות הנבדקות.
- השערת המחקר: קיים הבדל בין 2 קבוצות לפחות.

מבחן זה מתבסס על דירוג כלל התצפיות במחקר מכל המדגמים.

דוגמה (פתרון בהקלטה):

השוו את שביעות הרצון מהשירות של 4 חברות טלפון סלולרי שונות. שביעות הרצון הוא ציון שהלקוח היה צריך לתת לרמת השירות של החברה מ-1 ועד 10.

- מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר זה?
- מהן השערות המחקר?
- מהו המבחן הסטטיסטי במקרה זה?
- מה מסקנות המחקר ברמת מובהקות של 5%?

Count

| | | company | | | | Total |
|-------------|-------|---------|----|---|---|-------|
| | | 1 | 2 | 3 | 4 | |
| שביעות רצון | 4.00 | 0 | 2 | 0 | 0 | 2 |
| | 5.00 | 0 | 1 | 0 | 0 | 1 |
| | 6.00 | 2 | 1 | 0 | 1 | 4 |
| | 7.00 | 2 | 2 | 1 | 0 | 5 |
| | 8.00 | 1 | 2 | 2 | 2 | 7 |
| | 9.00 | 1 | 1 | 3 | 2 | 7 |
| | 10.00 | 1 | 1 | 0 | 0 | 2 |
| Total | | 7 | 10 | 6 | 5 | 28 |

Kruskal-Wallis Test

Ranks

| | company | N | Mean Rank |
|-------------|---------|----|-----------|
| שביעות רצון | 1 | 7 | 13.93 |
| | 2 | 10 | 11.40 |
| | 3 | 6 | 18.50 |
| | 4 | 5 | 16.70 |
| | Total | 28 | |

Test Statistics^{a,b}

| | שביעות רצון |
|-------------|-------------|
| Chi-Square | 3.363 |
| df | 3 |
| Asymp. Sig. | .339 |

a. Kruskal Wallis Test

b. Grouping Variable: company

שאלות

1) מחקר פסיכולוגי רצה להשוות 3 טיפול שונים לטיפול בחרדה. נלקחו נבדקים הסובלים מחרדה. הם חולקו באקראי ל-3 קבוצות וכל קבוצה קיבלה שיטת טיפול אחרת לחרדה. להלן תוצאות המבחן הסטטיסטי שבוצע:

Kruskal-Wallis Test

Ranks

| | group | N | Mean Rank |
|---------|-------|---|-----------|
| anxiety | 1.00 | 8 | 15.88 |
| | 2.00 | 4 | 6.88 |
| | 3.00 | 8 | 6.94 |

Test Statistics^{a,b}

| | anxiety |
|-------------|---------|
| Chi-Square | 11.149 |
| df | 2 |
| Asymp. Sig. | .004 |

- א. כמה נבדקים השתתפו במחקר?
- ב. באיזו שיטת טיפול רמת החרדה הייתה הנמוכה ביותר במדגם?
- ג. האם, עבור ר"מ 0.05, יש הבדל ברמת החרדה המתקבלת בשיטות הטיפול השונות?

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

| | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-------|
| 30 | 30 | 30 | 30 | 30 | 20 | 20 | 20 | 20 | 20 | 10 | 10 | 10 | 10 | 10 | מינון |
| 72 | 73 | 75 | 67 | 68 | 82 | 82 | 82 | 76 | 79 | 79 | 81 | 83 | 87 | 87 | דופק |

Kruskal-Wallis Test

| Ranks | | | | Test Statistics ^{a,b} | |
|-------|--------|---|-----------|--------------------------------|--------|
| | dosage | N | Mean Rank | | pulse |
| pulse | 10.00 | A | E | Chi-Square | 10.204 |
| | 20.00 | B | F | df | 2 |
| | 30.00 | C | G | Asymp. Sig. | .006 |
| Total | | D | | | |

- א. השלימו את המספרים החסרים במקום האותיות A עד G בטבלת הדירוגים.
- ב. בדוק ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת הדופק של האנשים?
- ג. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.

תשובות סופיות

- 1) א. 20 ב. מדגם 2 ג. נדחה H_0 .
- 2) א. 3, 9.3, 11.7, 5, 5, 5, 15 ב. נדחה H_0 . ג. לא תשתנה.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 9 - גודל האפקט

תוכן העניינים

1. גודל האפקט בהסקה על תוחלת ההפרש במדגמים מזווגים (ללא ספר)
2. גודל האפקט בהסקה על הפרש תוחלות במדגמים בלתי תלויים ובהנחת שוויון שוניות . (ללא ספר)
3. גודל האפקט במבחן לניתוח שונות (ללא ספר)

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 10 - מבחני חי בריבוע

תוכן העניינים

1. ניתוח פלטים במבחן אי תלות.....66

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

r : מספר הקטגוריות של המשתנה שבשורות.

c : מספר הקטגוריות של המשתנה שבעמודות.

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

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 צ i . במידה ותנאי זה לא מתקיים יש

אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.

תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ 20% מהתאים E קטן מ-5.

דוגמה (פתרון בהקלטה) :

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

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

להלן הפלטים שהתקבלו :

Case Processing Summary

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

hair_color * eye_color Crosstabulation

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

Chi-Square Tests

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

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

שאלות

- 1) בסקר שנעשה על ידי משרד ראש הממשלה נדגמו 60 אזרחים. כל אזרח נשאל על מגדרו והאם הוא בעד הקמת מדינה פלסטינית.
- מה ההשערות הנבדקות ומהו סטטיסטי המבחן?
 - אם סטטיסטי המבחן היה גדל כיצד הדבר היה משפיע על SIG שבפלט.
 - האם קיים קשר בין מגדר ודעה ברמת מובהקות של 5%?
 - מהו האומדן לאחוז התומכים במדינה פלסטינית מתוך הגברים?
 - איזה אחוז מהנשאלים שהיו בעד מדינה פלשתינית הם גברים?

להלן הפלטים:

Chi-Square Tests

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

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

b. The minimum expected count is 7.25.

gender * opinion Crosstabulation

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

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

Crosstab

Count

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

Chi-Square Tests

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

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

b. The minimum expected count is 1.43.

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

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

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.
- מה תהיה מסקנת המחקר ברמת מובהקות של 1%?
 - מה ערכו של E עבור עיניים כחולות וצבע שיער כהה.
 - מה יהיה בקירוב ערכו של מקדם המתאם של קרמר?
 - 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% | 100.0% |
| | | % within | 21.6% | 53.3% | 30.8% |
| | | % of Total | 15.4% | 15.4% | 30.8% |
| חום | | Count | | | |
| | | % within | 83.3% | 16.7% | 100.0% |
| | | % within | 27.0% | 13.3% | 23.1% |
| | | % of Total | 19.2% | 3.8% | 23.1% |
| ירוק | | Count | | | |
| | | % within | 79.2% | 20.8% | 100.0% |
| | | % within | 51.4% | 33.3% | 46.2% |
| | | % of Total | 36.5% | 9.6% | 46.2% |
| Total | | Count | | | |
| | | % within | 71.2% | 28.8% | 100.0% |
| | | % within | 100.0% | 100.0% | 100.0% |
| | | % of Total | 71.2% | 28.8% | 100.0% |

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

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 |

להלן פלט שני:

Crosstabs

Case Processing Summary

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

gender * category Crosstabulation

| | | | category | | | Total |
|--------|--------|-------------------|----------|--------|--------|--------|
| | | | a | b | c | |
| gender | Female | Count | 2 | 30 | 8 | 40 |
| | | Expected Count | 11.0 | 21.0 | 8.0 | 40.0 |
| | | % within gender | 5.0% | 75.0% | 20.0% | 100.0% |
| | | % within category | 9.1% | 71.4% | 50.0% | 50.0% |
| Male | Count | 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.

תשובות סופיות

- (1) א. 1.973 ב. קטן . ג. לא נדחה H_0 .
 ד. 34.5% ה. 38.5%
- (2) א. בילוי מועדף ומוזיקה מועדפת עם 3 קטגוריות לכל משתנה.
 ב. לא ג. 4.
 ד. H_0 אין תלות בין בילוי מועדף למוזיקה מועדפת.
 H_1 יש תלות בין בילוי מועדף למוזיקה מועדפת.
- (3) א. 6, 21.6%, 71.2% ב. 15.8 ג. קטן מ-0.005.
 ד. 20.8%
- (4) א. לא נדחה H_0 . ב. 11.4 ג. 0.394
- (5) א. פלא א' - חוקר א', פלא ב' - חוקר ב'. ב. נקבל את H_0 .
 ג. טעות מסוג שני- הכרענו שאין הבדל בין גברים לנשים למרות שיש במציאות הבדל.
 ד. נקבל את H_1 .
 ה. טעות מסוג ראשון- הכרענו שיש קשר בין מין להוצאה למרות שבמציאות אין קשר.
 ו. כל חוקר פעל בשיטה סטטיסטית שונה ובמצב כזה יתכן מסקנות סותרות.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר ב

פרק 11 - מדד הקשר של ספירמן

תוכן העניינים

1. הקדמה..... (ללא ספר)

2. ספירמן- ניתוח פלטי SPSS..... (ללא ספר)