

תאריך: 30.1.2018

מבחן ב: יסודות הסטטיסטיקה

מס' קורס: 102.1.0078

המרצה: ד"ר אורי ביבי

סמסטר: א', מועד א'

משך הבחינה: 3 שעות

חומר עזר: מחשב כיס ללא תכנות.

מבחן לא חסוי

במבחן זה שלושה חלקים. עליכם להשיב לכל השאלות. הניקוד עבור כל שאלה מופיע בצד השאלה.

חלק ראשון: שאלות תיאורטיות - 20 נקודות

בחלק זה 5 היגדים. בכל היגד ענו האם ההיגד נכון או לא נכון ונמקו.

שימו לב, הניקוד ניתן על הנימוק בלבד.

כל תשובה נכונה מזכה ב-4 נקודות.

שאלה מספר 1: (4 נקודות)

במאמר נכתב: "בבדיקת הפער בין הממוצעים נמצא כי: $[t(46) = 2.56, p < 0.035]$ ".

מכך ניתן להסיק שלא נמצא הבדל משמעותי בין הממוצעים.

לא נכון. אומנם לא נתונה לנו האלפא האפריורית אך כאשר ההבדל לא מובהק נהוג לכתוב:

$p > 0.05$.

שאלה מספר 2: (4 נקודות)

אם ידוע ש: $\sum (x_i - \bar{x}) = 0$, ניתן להסיק שהמשתנה מתפלג נורמאלית.

לא נכון. מרחק הסטיות מהממוצע תמיד שווה לאפס – זו הגדרת הממוצע.

שאלה מספר 3: (4 נקודות)

שכרו של עובד משויך למאון ה 74 בין עובדי החברה. העובד טוען עקב הוספת 500 ₪

לשכרם של כל העובדים, מעמדו היחסי בין עובדי החברה נפגע.

העובד טועה. נעשתה טרנספורמציה לינארית של +500 ₪ לכל עובד ולכן המיקום היחסי לא

השתנה כלל. מאון הוא מדד מיקום יחסי והיחס שלו לעומת שאר העובדים נשאר זהה.

שאלה מספר 4: (4 נקודות)

המעבר לבחינת השערות ממבחן Z למבחן t מוביל לכך שלחוקר קל יותר להגיע לתוצאה מובהקת.

לא נכון. במבחן t לא ידועה השונות באוכלוסייה ולכן מייצרים אומדן. באומדן יורדת ד"ח אחת מגודל המדגם ובכך מגדילה את ה-t הקריטי, דבר שמקשה על תוצאה מובהקת. כאשר המדגם מאוד גדול נוצר מצב של אינסוף ד"ח ומבחן t זהה למבחן Z.

שאלה מספר 5: (4 נקודות)

עבור כל שני משתנים בסולם רווח/ מנה, תמיד מתקיים: $r_p = r_s$

לא נכון. במתאם ספירמן ניבחת המונוטוניות של הקשר (כלומר הכיוון בלבד), ואילו במתאם פירסון נבדקת ליניאריות שמשמעותה כיוון + קצב.

כיוון שמתאם פירסון בוחן יותר מרכיבים ממתאם ספירמן, מתקבל למעשה: $r_p \leq r_s$

חלק שני: זיהוי המבחן הסטטיסטי – 16 נקודות

בחלק זה 2 שאלות.

בשאלות אלו, השיבו על הסעיפים הבאים בלבד:

- א. מהו המבחן הסטטיסטי.
- ב. נסחו השערות מתאימות למבחן הסטטיסטי.
- ג. האם התנאים הנדרשים לביצוע המבחן הסטטיסטי מתקיימים? לגבי כל תנאי, ציינו האם הוא מתקיים.
- ד. בהנחה וכל התנאים מתקיימים (כלומר, ללא קשר לתשובתכם בסעיף הקודם), האם יש טעם בביצוע המבחן הסטטיסטי? הסבירו.
- אין צורך לבצע מבחן סטטיסטי מלא על כל שלביו. כל תשובה נכונה מזכה ב-8 נקודות.

שאלה מספר 6: (8 נקודות)

על פי נתוני משרד התחבורה, בממוצע משלם נהג ישראלי 110 ₪ לשנה בגין קנסות. חוקר טוען שבקרוב נהגים צעירים הממוצע יהיה גדול יותר. לצורך בדיקת טענתו החוקר פונה ל- 36 נהגים צעירים שהוא מכיר ומוצא שבממוצע הם שילמו 95 ₪ לשנה בגין קנסות עם שונות 220. ידוע שהחוקר מעוניין לבחון את השערתו ברמת ביטחון של 95%.

א. מבחן t לממוצע בודד

ב. $H_0: \mu_x \leq 110$

$H_1: \mu_x > 110$

- ג. לא מתקיימת דגימה מקרית.
 מתקיימת התפלגות t בקירוב: $n \geq 30$.
 ד. לא. השערת החוקר היא בניגוד לכיוון הממצאים.

שאלה מספר 7: (8 נקודות)

חוקר מעוניין לבחון הבדלים בצריכת חלבונים בין מתאמנים בחדר כושר לבין כאלו שלא מתאמנים בחדר כושר. לצורך בדיקת הטענה נדגמו מקרית 49 מתאמנים מחדר הכושר ו 25 שלא מתאמנים בחדר הכושר. נמצא שצריכת החלבונים היומית של מתאמנים הייתה 175 גר' עם אומדן שונות 125 ואילו בקרב אלו שלא מתאמנים, הממוצע היה 125 גר' עם אומדן שונות 100.
 בבדיקה מקדימה לבחינת הומוגניות של שונות נמצא: $p=0.34$. ידוע שהחוקר מעוניין לבחון את השערתו ברמת ביטחון של 99%.

א. מבחן t להשוואת ממוצעים ב"ת כאשר השונות באוכל' הומוגניות

ב. $H_0; \mu_d = 0$

$H_1; \mu_d \neq 0$

ג. מתקיימת דגימה מקרית.

לא מתקיימת התפלגות נורמאלית או נורמלית בקירוב.

השונות באוכלוסיות שוות.

ד. כן. יש לנו שני ממוצעים שונים זה מזה וצריך לבדוק האם ההבדל מובהק או לא.

חלק שלישי: שאלות חישוב – 64 נקודות

בחלק זה 4 שאלות. בצעו פתרון חישובי מלא.

שאלה מספר 8: (24 נקודות)

בבדיקה עומסי צמתים של משרד התחבורה בוחנים את מספר הפקקים ביום, בצמתים שונים ברחבי הארץ, להלן הנתונים:

| מספר צמתים | מספר פקקים |
|------------|------------|
| 5 | 10 |
| 10 | 20 |
| 15 | 30 |
| 15 | 40 |

א. חשבו את מדדי המרכז האפשריים בהתאם לסולם המדידה של המשתנה. (9 נק')

| מספר פקקים (x) | מספר צמתים $f(x)$ | $F(x)$ | $X_i \cdot f(x)$ | $X_i^2 \cdot f(x)$ |
|-------------------|----------------------|--------|------------------|--------------------|
| 10 | 5 | 5 | 50 | 500 |
| 20 | 10 | 15 | 200 | 4000 |
| 30 | 15 | 30 | 450 | 13500 |
| 40 | 15 | 45 | 600 | 24000 |
| סה"כ | 45 | | 1300 | 42000 |

משתנה מסולם מנה ולכן נחשב את כל המדדים.

שכיח: 40;30 (דו שכיחית)

חציון: מיקום $\frac{45+1}{2} = 23$ ערך: 30

ממוצע: $\bar{X} = \frac{\sum X_i \cdot f(x)}{N} = \frac{1300}{45} = 28.89$

ב. חשבו את מדד הפיזור המתאים ביותר לסולם המדידה של המשתנה. הסבירו את

בחירתכם. (5 נק')

סטיית תקן, היות והיא מציגה את הפיזור סביב המשתנה במונחים של המשתנה. מתאימה לסולם מנה.

$$S^2 = \frac{\sum x_i^2 \cdot f(x)}{n} - \bar{x}^2 = \frac{42000}{45} - 28.89^2 = 98.7$$

$$S = \sqrt{S^2} = 9.93$$

ג. מהו אחוז הצמתים בהם יש יותר מ 22 פקקים ליום? (5 נק')

$$P_{22} = \frac{30}{45} * 100 = 66.67\% \rightarrow 66.67\%$$

ד. במדגם של עמותת אור ירוק נבדקו 200 צמתים אחרות ונמצא שמספר הפקקים

הממוצע היה 17. חשבו את הממוצע של הצמתים משני המחקרים יחד. (5 נק')

$$\bar{X} = \frac{(200 * 17) + (45 * 28.89)}{245} = 19.18$$

שאלה מספר 9: (10 נקודות)

ידוע שמספר כוסות הקפה שנמכרות בדוכן בכל יום, מתפלג נורמאלי עם ממוצע 1250 וסטיית תקן 250.

א. מנהל הדוכן הכריז כי "יום חלש הוא יום בו נמכרות 935 כוסות או פחות מכך". מהו

אחוז הימים שמוגדרים כ"ימים חלשים"? (5 נק')

$$(-1.26) = \frac{935 - 1250}{250} = \text{מצאתי בטבלת ההתפלגות הנורמלית } P = 0.1038$$

אחוז הימים המוגדרים חלשים הוא 10.38%.

ב. עובדי הדוכן מקבלים בונוס כאשר מספר הכוסות הנמכרות נמצא בחמשת האחוזים

העליונים של ההתפלגות. מאיזה מספר כוסות מתקבל הבונוס? (5 נק')

$$Z_p = 1.645 \leftarrow P = 0.95$$

$$1.645 = \frac{x_i - 1250}{250} \rightarrow 411.25 + 1250 = 1661.2$$

שאלה מספר 10: (14 נקודות)

הנהלת מלון רצתה לבדוק האם יש הבדל בשביעות הרצון (סקאלה 1-7) של אורחים,

השוהים בקומות שונות במלון. לבדיקת הטענה נדגמו מקרית 7 אורחים, ששהו בקומות

שונות. נרשמה שביעות רצונם, כשסיימו את חופשתם:

| קומת קרקע | קומה ראשונה | קומה שניה |
|-----------|-------------|-----------|
| 4 | 1 | 3 |
| 4 | 3 | 5 |
| 1 | | |

בהנחה שהשונויות הומוגניות והמשתנה מתפלג נורמאלי באוכלוסייה, האם קיימים הבדלים

בשביעות הרצון של אורחים אשר שהו בקומות שונות?

נרצה לבצע ניתוח שונות פשוט (F).

קודם כל נחשב את הסטטיסטי ע"מ לבדוק שהוא גדול מ-1 ושיש טעם למבחן.

ממוצעים (מהקומה הנמוכה לגבוהה): 3,2,4

$$3 = 7 / ((4 \cdot 2) + (2 \cdot 2) + (3 \cdot 3)) = \text{ממוצע משוקלל}$$

$$Ssb = 3(3-3)^2 + 2(2-3)^2 + 2(4-3)^2 = 4$$

$$Dfb = j - 1 = 3 - 1 = 2$$

$$Msb = 4/2 = 2$$

$$Ssw = (4-3)^2 + (4-3)^2 + (1-3)^2 + (1-2)^2 + (3-2)^2 + (3-4)^2 + (5-4)^2 = 10$$

$$Dfw = n - j = 7 - 3 = 4$$

$$Msw = 10/4 = 2.5$$

$$F = 2/2.5 = 0.8$$

| | SS | Df | Ms | F |
|---|----|---------------------|--------------|---------------|
| B | 4 | $j - 1 = 3 - 1 = 2$ | $4/2 = 2$ | $2/2.5 = 0.8$ |
| W | 10 | $N - j = 7 - 3 = 4$ | $10/4 = 2.5$ | |

היות ו-F קטן מאחד אין טעם לבצע את המבחן. אין הבדל בשביעות הרצון של אורחים ששהו בקומות שונות.

שאלה מספר 11: (16 נקודות)

למחקר נדגמו מקרית 120 נהגי מכוניות ו- 80 נהגי אופנוע. 40 מתוך נהגי האופנוע ו- 40 מתוך נהגי המכוניות נקנסו בגין מהירות, בשנתיים האחרונות.
א. חשבו את המתאם בין סוג הרכב (מכונית/אופנוע), לבין קנס בגין מהירות (כן/לא). (6 נק')

נחשב מתאם של קרמר היות ואלו משתנים מסולם שמי.

| Observed | מכונית | אופנוע | |
|-----------|--------|--------|-----|
| קנסות | 40 | 40 | 80 |
| ללא קנסות | 80 | 40 | 120 |
| | 120 | 80 | 200 |

| Expected | מכונית | אופנוע | |
|-----------|--------|--------|-----|
| קנסות | 48 | 32 | 80 |
| ללא קנסות | 72 | 48 | 120 |
| | 120 | 80 | 200 |

$$\chi^2 = \frac{(40-48)^2}{48} + \frac{(40-32)^2}{32} + \frac{(80-72)^2}{72} + \frac{(40-48)^2}{48} = 1.33 + 2 + 0.89 + 1.33 = 5.55$$

$$r_c = \sqrt{\frac{5.55}{200(2-1)}} = 0.167$$

ב. בצעו מבחן מובהקות לבדיקת קיומו של קשר בין המשתנים באוכלוסייה ($\alpha=0.05$). (10 נק')

מבחן חי בריבוע לאי תלות
1. השערות:

H_0 : אין קשר בין סוג כלי הרכב לקבלת קנס
 H_1 : יש קשר בין סוג כלי הרכב לקבלת קנס

2. תנאים:

דגימה מקרת – מתקיים

כל תצפית בקטגוריה אחת – מתקיים

המדגם גדול מספיק שכן השכיחות הצפויה בכל קטגוריה הינה לפחות 10.

3. קביעת רמת מובהקות ואזורי דחייה וקבלה: $\alpha = 0.05$, $df=1$, $\chi^2 = 3.84$

אזור קבלה $\chi^2 < 3.84$

אזור דחייה $\chi^2 \geq 3.84$

4. הסטטיסטי $\chi^2 = 5.55$