אוניברסיטת בן-גוריון בנגב
Ben-Gurion University of the Negev

פתרון מרצה

תאריך הבחינה: ______24.02.17_______

שם המרצה: ד"ר יואב קסלר

שם הקורס: __מבוא לסטטיסטיקה____

מספר הקורס: __101-1-0049_______

שנה: 2017 סמסטר: א' מועד: ב'

משך הבחינה: ____שלוש שעות____

חומר עזר: __מחשבון מדעי_____

המלצה: התחילו מהשאלות הקלות ביותר בעיניכם, והשאירו את השאלות הקשות לסוף.

חלק א:

ענו על הסעיפים הבאים ונמקו את תשובותיכם. תשובה ללא נימוק לא תזכה בנקודות. אין לנמק ביותר מחמש שורות, ניקוד יורד על חריגה. כל תשובה נכונה מלאה מזכה ב-5 נקודות. סה"כ לחלק זה 30 נקודות.

1. נכון/ לא-נכון: ניתן לחשב את המאון ה-20 בהתפלגות מדגם סופי של תצפיות באמצעות מציאת הערך המתאים לציון התקן $Z=-0.84$.

לא נכון. אמנם ערך זה הינו הערך המתאים למאון ה-20 בהתפלגות נורמלית, אך לא נתון שהמדגם מתפלג נורמלית ולכן לא (בהכרח) ניתן להשתמש בטבלת Z על מנת לחשב בו מאונים.

(הערה: עקרונית, מדגם סופי אינו יכול להתפלג נורמלית, מכיוון שהתפלגות מדגם אינה התפלגות רציפה. בסמסטר הבא נלמד שניתן לבדוק אם מדגם הוא נורמלי בקירוב).

2. עוזר מחקר התבקש לחשב מתאם פירסון בין שני משתנים מדגם. הוא סידר את ערכי כל אחד מהמשתנים לפי סדר, מהגדול לקטן. לאחר מכן, חישב את ממוצע מכפלות ציוני התקן לפי הסדר שקיבל. חווה/י דעתך על פעולות עוזר המחקר.

עוזר המחקר "ערבב" את הערכים של שני המשתנים, כך שכבר אינם מצומדים לפי התצפיות אליהם היו שייכים. יש לחשב מתאם פירסון בהתאם לזיווג התצפיות, ולפיכך עוזר המחקר לא חישב מתאם פירסון כראוי, ואין משמעות לערך שיקבל. סביר להניח שיתקבל מתאם גבוה יותר מכפי שקיים בפועל, מכיוון שעוזר המחקר יצר מצב בו התצפיות יופיעו ביחד לפי סדר גודל, כך שכאשר X עולה גם Y עולה (כלומר, הוא מיקסם את המונוטוניות במדגם). יוצאי דופן הם מקרים בהם התצפיות מלכתחילה היו מזווגות לפי סדר גודל (כך שהמניפולציה שביצע על הנתונים לא שינתה אותם), או שאחד המשתנים הינו קבוע.

3. בגלגל הענק בלונה פארק 5 תאים זהים בעלי 4 מושבים כל אחד. בערב יום האהבה, מקפידים הסדרנים שבכל תא ישבו שני זוגות (סדר הישיבה בתוך התא אינו חשוב לסדרנים). נכון/ לא-נכון: ישנן 22,680 דרכים בהם הסדרן יכול לסדר 10 זוגות בתאי הגלגל הענק בערב ביום האהבה.

נכון.

נתחיל מלסדר את 10 הזוגות ב-5 רביעיות (רביעייה = 2 זוגות) לפי סדר. יש לנו 10! דרכים לסדר את הזוגות בשורה, וכדי לצור מתוך זה רביעיות נחלק בסידורים הפנימיים של כל רביעיה ע"י חלוקה ב-2^5.

דרכים נוספות לחשוב על זה:

נעבור לפי הקרונות:

- קרון ראשון: $\frac{10\cdot 9}{2}$ (החלוקה ב-2 הינה לביטול סידור פנימי בין הזוגות בתוך הקרון).
- קרון שני: $\frac{8\cdot 7}{2}$
- קרון שלישי: $\frac{6\cdot 5}{2}$
- קרון רביעי: $\frac{4\cdot 3}{2}$
- קרון חמישי: $\frac{2\cdot 1}{2}$
- סה"כ: $\frac{10!}{2^5}$

נשתמש בהרחבה של נוסחת הקומבינציה ליותר מ-2 קבוצות לא מובחנות:

$$\frac{10!}{2!\,2!\,2!\,2!\,2!} = \frac{10!}{2^5} = 113{,}400$$

עד כאן הזוגות סודרו ב-5 קרונות. עם זאת, מכיוון שהגלגל הענק הוא בצורה מעגל (ולא שורה של קרונות, כמו רכבת הרים למשל), יש לבטל את הסדרים שחוזרים על עצמם. מתוך 5 הקרונות, חוזרים על עצמם 5 סידורים לכן, סה"כ הסידורים הינו:

$$\frac{113{,}400}{5} = 22{,}680$$

4. נתונה מטריצה Y שאיבריה מוגדרים: $y_{ij} = i + j$. באמצעות מטריצה Y נוצרו איברי מטריצה Z, שמוגדרים: $z_{ij} = \frac{y_{ij}}{j}$. נכון/ לא-נכון: איברי מטריצה Z הינם טרנספורמציה ליניארית של איברי מטריצה Y.

לא נכון. טרנספורמציה ליניארית מתקבלת בהכפלה בקבוע או בהוספת/החסרת קבוע מכל התצפיות באותו האופן. במקרה המדובר, כל ערך חולק במספר הטור שלו במטריצה. במצב זה, לא כל איברי Z עברו את אותה הטרנספורמציה (איברי כל טור בלבד עברו את אותה הטרנספורמציה).

5. אבנר הוא שחקן פינג-פונג מקצועי, היודע את הקצב הממוצע בו הוא משיג נקודות במשחק. אבנר רצה לדעת מה ההסתברות שישיג את הנקודה הראשונה במשחק בין דקה לשתי דקות לאחר שהתחיל המשחק. באמצעות פונקציית הצפיפות של התפלגות מעריכית, הוא חישב את ההסתברות שישיג את הנקודה הראשונה לאחר דקה או יותר ואת ההסתברות שישיג אותה תוך שתי דקות, והכפיל בין שתי ההסתברויות:

$$P(1 < x < 2) = P\big((X > 1) \cap (X < 2)\big) = P(X > 1) \cdot P(X < 2)$$

האם החישוב של אבנר נכון? נמק/י.

החישוב של אבנר אינו נכון. ניתן לחשב חיתוך בין מאורעות ע"י הכפלת ההסתברות לכל מאורע רק אם אין תלות בין המאורעות. עם זאת, המאורעות במקרה הזה תלויים (לדוגמא: אם X<1, כך שהתקיים המשלים למאורע הראשון, ההסתברות ש-X<2 שווה 1). היה עליו לחשב את ההסתברות לקבל ערך עד X=2 ולחסר ממנה את ההסתברות לקבל ערך עד x=1:

$$P(1 < x < 2) = P(X < 2) - P(X < 1) = F(2) - F(1) = 1 - e^{-2\lambda} - 1 + e^{-\lambda} = e^{-\lambda} - e^{-2\lambda}$$

6. <u>נכון/ לא-נכון</u>: $(A - B) - C = (\overline{B \cup C \cup \bar{A}})$.

נכון.

$$(A - B) - C = A \cap \bar{B} \cap \bar{C}$$

לפי דה-מורגן:

$$A \cap \bar{B} \cap \bar{C} = (\overline{\bar{A} \cup B \cup C})$$

איחוד הוא סימטרי, ולכן:

$$\big(\overline{\bar{A} \cup B \cup C}\big) = (\overline{B \cup C \cup \bar{A}})$$

חלק ב:

בחלק זה שאלות חישוביות במספר סעיפים. עליכם לענות על כל הסעיפים תוך מתן הפירוט המרבי. הניקוד עבור כל סעיף ניתן לצידו. סה"כ לחלק זה 70 נקודות.

7. בחוות הסוסים "עין הסערה" עורכים מרוצי סוסים. מסלול המרוצים נמשך מספר ימים, ומפוזרים בו דגלונים, אשר המרחקים ביניהם מתפלגים מעריכית. כל פעם שסוס חולף על פני דגלון, הוא מקבל נקודה. הסוס 'רעם' מקבל 8 נקודות ביום בממוצע. מנהלי החווה רכשו סוס חדש, 'ברק', אשר משיג 9 נקודות ביום בממוצע, וערכו סדרת מרוצים בינו לבינו רעם. חברת הסבונים "נקה 13" הינה ספונסר של המרוץ. החברה החליטה להעניק פרס יומי לסוס שירוויח בדיוק 13 נקודות באותו היום (שני הסוסים יכולים לזכות באותו היום). בשאלה זו נתייחס לפרס זה בתור "הפרס היומי".

הניחו כי קצב הדהירה של כל סוס הינו קבוע, וכי קצבי הדהירה של הסוסים בלתי-תלויים אחד בשני.

א. באיזה אופן מתפלג משתנה הנקודות היומיות של כל סוס? (3 נק')

מכיוון שהמרחק בין הנקודות מתפלג מעריכית וקצב הדהירה של כל סוס הינו קבוע, קצב צבירת הנקודות מהווה זרם פואסון, כאשר צבירת נקודה מהווה הצלחה אחת. לפיכך, נשתנה הנקודות היומיות של כל סוס מתפלג פואסון.

ב. מהי ההסתברות שרעם יקבל את הפרס היומי, וגם שיעקוף את ברק ב-2 נקודות לכל היותר? (8 נק').

עלינו למצוא חיתוך בין שני מאורעות בלתי-תלויים.

1. ההסתברות שרעם ישיג 13 נקודות:

λ=8

$$P(X=k) \approx e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

$$P(X=13) \approx e^{-8} \cdot \frac{8^{13}}{13!} = 0.03$$

2. ההסתברות שברק יקבל נקודה או שתיים פחות:

λ=9

$$P(X=12) \approx e^{-9} \cdot \frac{9^{12}}{12!} = 0.07$$

$$P(X=11) \approx e^{-9} \cdot \frac{9^{11}}{11!} = 0.10$$

$$P\big((X=12) \cup P(X=11)\big) = 0.07 + 0.10 = 0.17$$

3. הסתברות החיתוך ביניהם:

$$P = 0.03 \cdot 0.17 = 0.005$$

ג. מהי תוחלת מספר תחרויות "הפרס היומי" שיעברו עד שרעם יקבל אותו לראשונה? (5 נק')

ההתפלגות המתאימה הינה גיאומטרית.

מסעיף שעבר -

p=0.03

תוחלת משתנה גיאומטרי:

$$E(X) = \frac{1}{p}$$

לכן:

$$E(X) = \frac{1}{p} = \frac{1}{0.03} = 33.33$$

בתוחלת, יעברו 33.33 ימי מרוץ לפני שרעם יקבל הצטיינות יומית.

ד. בסוף כל יום מרוץ מתבצע משחק שש-בש בין 2 מהמרים שהימרו במרוץ. אם המנצח בשש-בש הימר על סוס שלא זכה בפרס היומי יוכל לקבל את הסכום עליו הימר בחזרה. ניצן מהמרת תמיד על רעם. ההסתברות שניצן תנצח בשש-בש היא 0.4. אם ניצן ניצחה בשש-בש, ההסתברות שרעם לא זכה בפרס היומי הינה 0.95. מהי ההסתברות שגם ניצן וגם רעם יפסידו באותו היום? (9 נק')

ישנן מספר דרכים לפתור את השאלה, להלן דרך אחת.

נגדיר:

A - רעם ניצח בתחרות

B - ניצן נצחה בשש-בש

נתון:

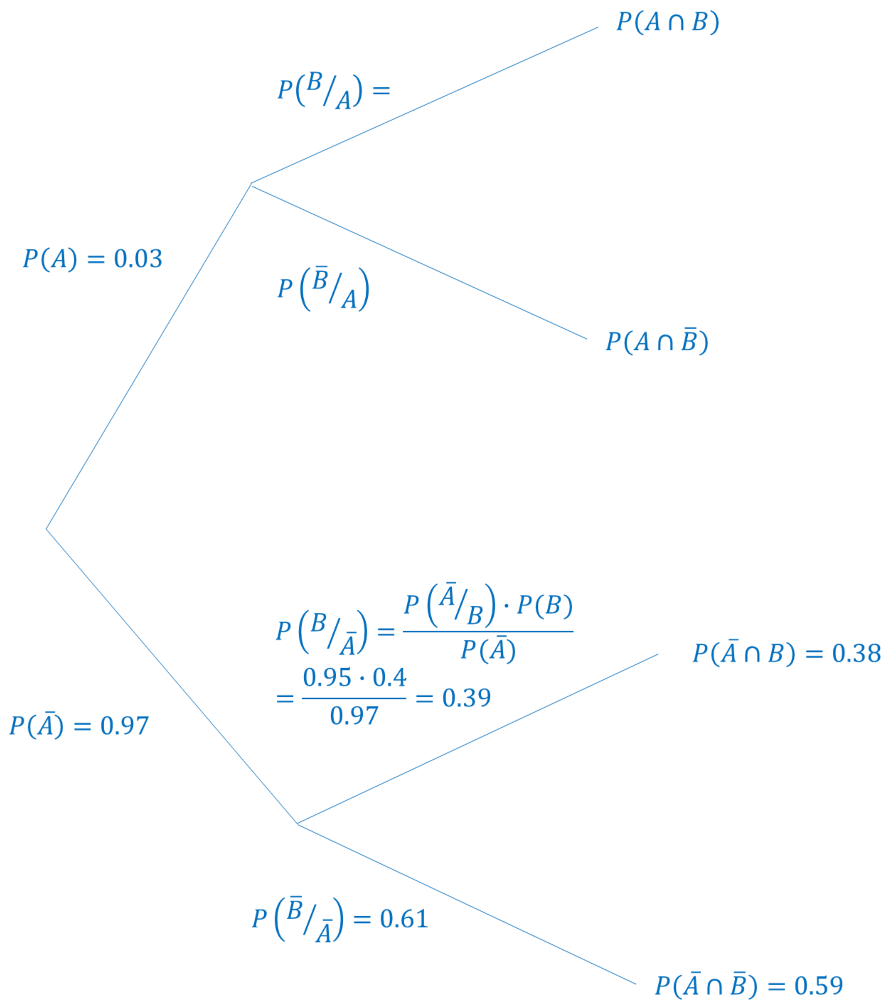
$$P(B) = 0.4$$

$$P(\bar{A}/B) = 0.95$$

מסעיף א' ידוע:

$$P(A) = 0.03$$

צ"ל: $P(\bar{A} \cap \bar{B})$



$$P(\bar{A} \cap \bar{B}) = 0.59$$

ה. בהמשך לנתוני סעיף ד', מהי תוחלת ושונות המקרים בהם רעם זוכה בפרס היומי או ניצן מנצחת בשש-בש (אבל לא שניהם) ביום מרוצים טיפוסי, אם ידוע שמרוץ נמשך 6 ימים? (10 נק')

מדובר במשתנה בינומי בו n=6. נמצא קודם את p:

$$p = P(A \cap \bar{B}) + P(\bar{A} \cap B)$$

את $P(\bar{A} \cap B)$ חישבנו בסעיף הקודם, וכעת חסר את $P(A \cap \bar{B})$:

$$P(A \cap B) = P(B) - P(\bar{A} \cap B) = 0.4 - 0.38 = 0.02$$

$$P(A \cap \bar{B}) = P(B) - P(A \cap B) = 0.03 - 0.02 = 0.01$$

p=0.01+0.38=0.39

$$E[X] = n \cdot p = 0.39 \cdot 6 = 2.34$$

$$Var[X] = n \cdot p \cdot q = 2.34 \cdot 0.61 = 1.43$$

(35 נקודות)

8. חוקר רגשות ערך מחקר ובו מדד אמפתיה, יכולת ויסות עצמי וחרדה בקרב סטודנטים בתואר ראשון. כל המשתנים נמדדו בסקאלה של 0-10.

מספר נבדק	אמפתיה	ויסות עצמי	חרדה
1	4	9	3
2	3.5	5	3
3	4	1	3
4	5	2	3
5	9	6.5	3
6	8	3	3
7	1.5	7.5	3

א. על-פי נתוני המדגם, האם קיים קשר בין ויסות עצמי לחרדה? ענה/י <u>מבלי לחשב</u> ונמק/י תשובתך. (5 נק')

ניתן לראות שחרדה הינה <u>קבועה</u>. לא יכול להיות מתאם בין משתנה בעל שונות לבין קבוע (שאין לו שונות) - רמות של ויסות עצמי לא ישתנו כפונקציה של חרדה, ולא להיפך, מכיוון שחרדה תמיד תהיה עם אותו הערך (3). לפיכך, לא קיים קשר בין ויסות עצמי לחרדה.

ב. חשב/י את הממוצע, את סטיית התקן ואת הטווח הבין-רבעוני באמפתיה. (13 נק').

$$\bar{x} = \frac{\sum x_i}{N} = \frac{4+3.5+4+5+9+8+1.5}{7} = 5$$

$$S_x^{\ 2} = \frac{\sum(x-x)^2}{N} = \frac{1+2.25+1+0+16+9+12.5}{7} = \frac{41.5}{7} = 5.93$$

$$S_x = \sqrt{5.93} = 2.43$$

טווח בין רבעוני:

ראשית, נסדר את התצפיות לפי סדר: 1.5,3.5,4,4,5,8,9

הערך במאון ה-25 (Q_1):

$$n = \frac{p}{100} \cdot (N+1)$$

$$n_{p=25} = 0.25 \cdot 8 = 2$$

$$x_{p=25} = 3.5$$

הערך במאון ה-**75** (Q_3):

$$n_{p=75} = 0.75 \cdot 8 = 6$$

$$x_{p=75} = 8$$

הטווח הבין רבעוני: $Q_3 - Q_1 = X_{p=75} - X_{p=25} = 8 - 3.5 = 2.5$

ג. Skewness (צידוד) הינו מדד סטטיסטיקה תיאורית שלא נלמד בקורס ומלמד על צורת ההתפלגות של משתנה. אחת הדרכים לחשב אותו הינה באמצעות חישוב ההפרש בין הממוצע לחציון, הכפלתו ב-3, וחלוקה בסטיית התקן:

$$Sk = \frac{3 \cdot (\bar{x} - M_d)}{s}$$

ג.1. הסביר/י כיצד ערך Skewness יכול ללמד אודות צורת התפלגותו של משתנה. (5 נק').

ה-Skewness נותן מדד לגבי סימטריות ההתפלגות. ההתפלגות הינה סימטרית כאשר הממוצע והחציון שווים. עם זאת, כאשר אינם שווים ההתפלגות מוטה. המדד מספק ערך סטנדרטי למדידת סטייה זו.

כאשר ההתפלגות סימטרית, נקבל Sk=0 - מכיוון שהמונה יתאפס (החציון והממוצע שווים).

כאשר ההתפלגות א-סימטרית חיובית (זנב ימני) - הממוצע יהיה (לרוב) גבוה מהחציון, ונקבל Sk חיובי.

כאשר ההתפלגות א-סימטרית שלילית (זנב שמאלי) - הממוצע יהיה (לרוב) נמוך מהחציון, ונקבל Sk שלילי.

ג.2. חשב/י את ה-Skewness של התפלגות האמפתיה במדגם, וציינ/י מה ניתן ללמוד מהערך על צורת ההתפלגות של אמפתיה במדגם (5 נק').

$$M_d = 4$$

$$Sk = \frac{3 \cdot (\bar{x} - M_d)}{s} = \frac{3(5-4)}{2.43} = 1.23$$

במדגם זה, ה-Skewness הינו חיובי, כך שניתן ללמוד מערך זה שההתפלגות הינה א-סימטרית חיובית, עם זנב ימני.

ד. אדם נחשב למאופק אם הוא נמצא בעשירון העליון באוכלוסיה בויסות עצמי. ידוע שויסות עצמי מתפלג נורמלית באוכלוסיה, בה התוחלת הינה 4.5 והשונות הינה 4. אילו נבדקים במדגם של החוקר נחשבים למאופקים? (7 נק').

נתון:

μ=4.5

$\sigma^2=4 \rightarrow \sigma=2$

נשתמש בטבלת Z כדי לאתר את ציון התקן הנמצא במאון ה-90:

$$Z_{P(z\to\infty)=.10} = 1.28$$

נבדקים שציון התקן שלהם גבוה מ-1.28 נמצאים בעשירון העליון ונחשבים למאופקים. נמצא את צין הויסות העצמי המתאים לציון Z=1.28:

$$Z_x = \frac{x-\mu_x}{\sigma_x}$$

$$1.28 = \frac{x-4.5}{2}$$

$$x = 7.06$$

נחפש אילו נבדקים הם בעלי ציוני ויסות עצמי גבוהים מ-7.06: **נבדקים 1 ו-7 במדגם החוקר נחשבים מאופקים**.

(35 נקודות)

ב ה צ ל ח ה !