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

תאריך הבחינה:	______03.02.16______
שם המרצה:	ד"ר יואב קסלר,
	ד"ר מעין קציר
שם הקורס:	__מבוא לסטטיסטיקה____
מספר הקורס:	__101-1-0049______
שנה: 2016	סמסטר: א' מועד: א'
משך הבחינה:	____שלוש שעות____

פתרון מרצה +
ציון : 80
(לאחר פקטור 95)

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

<u>חלק א</u>:

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. נגדיר את המשתנה המקרי T כסכום של שני משתנים מקריים רציפים, אשר ביניהם קשר שלילי.
<u>נכון/ לא-נכון</u>: השונות של T קטנה מסכום השונויות של שני המשתנים המקריים מהם הוא מורכב.

נכון. שונות של משתנה סכום הינה: $\mathrm{Var}(T) = \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$. Cov הינו חיובי כאשר הקשר בין המשתנים הינו חיובי, שלילי כאשר הקשר שלילי, ושווה ל-0 כאשר המשתנים בלתי תלויים. בשאלה זו נתון שהקשר שלילי, ולכן השונות של T תהיה שווה לסכום שונויות המשתנים **<u>פחות</u>** ה-cov. כלומר, שונות T תהיה קטנה מסכום שונויות המשתנים מה הוא מורכב.

2. X ו-Y הינם שני משתנים בסולם מנה במדגם אמפירי. <u>נכון/ לא-נכון</u>: מכיוון שממוצע ציוני תקן בכל אחת מההתפלגויות שווה ל-0, ממוצע מכפלות ציוני התקן של X ו-Y תמיד שווה ל-0.

לא נכון. החלק הראשון של המשפט הינו נכון: ממוצע ציוני התקן תמיד שווה ל-0. עם זאת, אין הדבר אומר שממוצע מכפלות ציוני התקן של שני משתנים תמיד שווה ל-0. חישוב זה מתבסס על הסדר בין התצפיות. ממוצע מכפלות ציוני תקן הינו ההגדרה של מתאם פירסון (במקרים בהם מדובר בתצפיות מזווגות), אשר ערכיו יכולים לנוע בין 1 ל-(1-). למעשה, ממוצע מכפלות ציוני התקן של שני משתנים שווה ל-0 רק כאשר אין קשר בין שני המשתנים.

3. הסביר/י מהו משתנה מקרי חסר זיכרון. תנ/י דוגמא להתפלגות חסרת זיכרון שנלמדה בקורס, והדגמ/י כיצד מתבטאת בו תכונת חוסר הזיכרון.

משתנה מקרי חסר זיכרון הינה משתנה שבו ההסתברות למאורע מסוים אינה מושפעת מהתרחשויות קודמות למאורע, לדוגמא, עבור משתנה בדיד:

$$P(X = n + k \mid X > k) = P(X = n)$$

בקורס נלמדו שתי התפלגויות חסרות זיכרון: גאומטרית (בדידה), ומעריכית (רציפה).
דוגמא:
נניח ש-p=0.3, n=6 ו-k=2:

$$P(X=6)=0.7^5\cdot 0.3$$

$$P(X=8 \mid X>2)=\frac{P((X=8)\cap(X>2))}{P(X>2)}=\frac{0.7^7\cdot 0.3}{0.7^2}=0.7^5\cdot 0.3$$

$$P(X=6)=P(X=8 \mid X>2)$$

4. נכון/ לא-נכון: עבור התפלגות סימטרית חד-שכיחית, הערכים המינימליים של שלושת פונקציות ההפסד שווים.

לא נכון. בהתפלגות סימטרית חד-שכיחית הממוצע, השכיח והחציון שווים. לכן, שלושת פונקציות ההפסד יהיו מינימליות **עבור אותו הערך**. עם זאת, הערך המינימלי עצמו לא יהיה זהה, מאחר וכל אחת מהן מחושבת בצורה שונה.
נקח לדוגמא את ההתפלגות: 11223334455.
הממוצע, השכיח והחציון הינם 3.
פונקציית הפסד ראשונה (מספר הטעויות): 8
פונקציית הפסד שנייה (סכום הסטיות המוחלטות): $\sum|x-3|=2+2+1+1+1+1+2+2=12$
פונקציית הפסד שלישית (סכום ריבועי הסטיות): $\sum(x-3)^2=4+4+1+1+1+1+4+4=20$

5. נתון: $\bar{A}\cap B=B$, ואף אחת מהקבוצות אינה הקבוצה הריקה. נכון/ לא-נכון: $P(A)\cdot P(B)=0$.

לא נכון. לפי הנתונים אין חיתוך בין הקבוצות: $A\cap B=\emptyset$ ולכן $P(A\cap B)=0$. מאורעות זרים הם בהכרח תלויים. כאשר מתקיימת אי-תלות, $P(A)\cdot P(B)=P(A\cap B)$. במקרה זה, אי-השיוויון לא מתקיים, ולכן מכפלת ההסתברויות למאורעות לא יכולה להיות שווה ל-0.

6. N זוגות גבר-אישה סועדים במסעדה עם שולחן עגול בו 2N מקומות. נכון/ לא-נכון: ההסתברות שבהושבה אקראית של כל הסועדים בשולחן כל הזוגות ישבו אחד ליד השני הינה: $\frac{2^N\cdot(N-1)!}{(2N-1)!}$.

נכון. כדי לסדר 2N אנשים במעגל, ישנם !(2N-1) סידורים אפשריים. נתייחס אליהם כאל כלל האפשרויות במרחב ההסתברות. כדי לסדר את הזוגות כך שבני זוג ישבו ביחד, נתייחס אליהם קודם כאל יחידה אחת (N זוגות), ונסדר אותה במעגל: !(N-1) אפשרויות. כעת, יש להתייחס לכל הסידורים הפנימיים בתוך הזוגות: !2 אפשרויות לזוג הראשון (לצורך הפשטות נרשום 2 מעתה), 2 לזוג השני, וכן הלאה - 2^N אפשרויות לסידור פנימי של הזוגות. מכאן שההסתברות שהסידור לפי זוגות במעגל יהיה מקרי הוא: $\frac{2^N\cdot(N-1)!}{(2N-1)!}$.

<u>חלק ב:</u>

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. רחל, פסיכולוגית ניסויית, חוקרת מצמוצי עיניים (ידוע כי הקצב שבו אדם ממצמץ קשור לפעילות מוחית המושפעת מדופמין). ידוע כי אדם ממצמץ 20 פעמים בדקה. אנא שמור/י על דיוק של 3 ספרות לאחר הנקודה בשאלה זו.

א. מהי ההסתברות שנבדק ימצמץ בין שלוש לחמש פעמים (כולל) בחצי הדקה הראשונה בניסוי? על אילו הנחות מבוססת ההסתברות שחישבת? (6 נק')

ביחידת זמן של חצי דקה, λ=10.
נשתמש בהתפלגות פואסון, תחת ההנחות הבאות:
1. ההסתברות ל-P שווה בכל שלב בניסוי.
2. ההסתברות ששני מאורעות יתרחשו בדיוק יחד היא 0.

$$P(X=k) \approx e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$
$$P(X=3) = e^{-10} \cdot \frac{10^3}{3!} = 0.007$$
$$P(X=4) = e^{-10} \cdot \frac{10^4}{4!} = 0.019$$

$$P(X=5) = e^{-10} \cdot \frac{10^5}{5!} = 0.038$$
ביחד: $P(3 \leq X \leq 5)=0.007+0.019+0.038=0.064$

ב. מהי ההסתברות שיעברו בין 5 ל-10 שניות עד המצמוץ הראשון? (6 נק')

מכיוון שאנו נדרשים למצוא את הזמן שנדרש עד להתרחשות המאורע הראשון, נשתמש בהתפלגות מעריכית (אקספוננציאלית):
ניתן להשתמש ב-λ הנתונה לפי חלקי דקה (כך שהזמנים יהיו 1/12 ושישית דקה), או להמיר את λ לשניות: קצב המצמוצים בשניות הינו שליש מצמוץ לשניה.

$$P(5<x<10) = F(10) - F(5)$$
$$F(x) = 1 - e^{-\lambda x}$$
$$F(5) = 1 - e^{-\frac{1}{3}5} = 0.811$$

$$F(10) = 1 - e^{-\frac{1}{3}10} = 0.964$$

$$P(5<x<10) = F(10) - F(5) = 0.964 - 0.811 = 0.153$$

(ניתן לפתור גם באמצעות (P(x>5 ו-P(x>15))

ג. במחקר משתתפים חמישה נבדקים. מהי ההסתברות שפחות משניים מהם לא ימצמצו ב-5 השניות הראשונות בניסוי? (10 נק')

ההתפלגות המתאימה לחישוב זה הינה בינומית.
ראשית נמצא את ההסתברות לא למצמץ ב-5 השניות הראשונות (המשלים של F(5) שחושב בסעיף הקודם):

$$P(x > 5) = e^{-\frac{1}{3}5} = 0.189$$

נגדיר הסתברות זו כ-p (ניתן להגדיר גם כ-q, ולבצע חישוב הפוך).
לכן, נשתמש בהתפלגות בינומית בה p=0.189, q=0.811, n=5

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

עבור k=0:

$$P(X = 0) = p(5,0,0.189) = \binom{5}{0} \cdot 0.189^0 \cdot 0.811^5 = 0.351$$

עבור k=1:

$$P(X = 1) = p(5,1,0.189) = \binom{5}{1} \cdot 0.189^1 \cdot 0.811^4 = 0.409$$

P(X<2)=0.76

(סה"כ לשאלה זו - 22 נקודות)

8. בגן שולה 100 ילדים שלומדים לשרוך שרוכים, לגזור בקו ישר ולספור עד 10. ההסתברות שילד בגן יכול לקשור את השרוכים לבד הינה 0.4. ההסתברות שילד יכול לגזור בקו ישר הינה 0.7 אם ידוע שהוא יודע לקשור את השרוכים לבד, ו-0.6 אם אינו יודע לקשור את השרוכים לבד. עבור ילדים שמצליחים לקשור את השרוכים לבד, הצלחה בספירה עד 10 אינה תלויה ביכולת הגזירה. ההסתברות שילד יצליח בשלושת הדברים הינה 0.14. כמו-כן, 8 מהילדים לא מצליחים באף אחד מהדברים, ו-46 מצליחים רק באחד.

א. מהי ההסתברות שילד יודע לשרוך שרוכים, אם ידוע שהוא יכול לגזור בקו ישר ויודע לספור עד 10? (10 נק')

נסמן:
A - קשירת שרוכים
B - גזירה בקו ישר
C - ספירה עד 10

אנחנו צריכים למצוא את $P(A/_{B\cap C})$
כדי למצוא זאת, נשתמש בנוסחא:

$$P(A/_B) = \frac{P(A\cap B)}{P(B)}$$

מהעץ ניתן לראות כי:

$$P(B\cap C)=P(A\cap B\cap C)+P(\bar{A}\cap B\cap C)=0.14+0.12=0.26$$

$$P\left({}^{A}\!/_{B\cap c}\right)=\frac{P(A\cap B\cap C)}{P(B\cap C)}=\frac{0.14}{0.26}=0.538$$

ב. חשב את המתאם בין היכולת לספור עד 10 ליכולת לשרוך שרוכים בגן שולה (10 נק').

מדובר בשני משתנים שמיים, ולכן נחשב מתאם קרמר.
לשם כך, עלינו לצור טבלת observed (אותה נחלץ מהעץ) וטבלת expected:

Observed:

	לא יכול לקשור שרוכים $\bar{A}$	יכול לקשור שרוכים A	
יכול לספור עד 10 C	12+16=28	14+6=20	**48**
לא יכול לספור עד 10 $\bar{C}$	24+8=32	14+6=20	**52**
	60	**40**	**100**

Expected:

	לא יכול לקשור שרוכים $\bar{A}$	יכול לקשור שרוכים A	
יכול לספור עד 10 C	28.8	19.2	**48**
לא יכול לספור עד 10 $\bar{C}$	31.2	20.8	**52**
	60	**40**	**100**

חישוב חי בריבוע:

$$\chi^2=\sum\frac{(O-E)^2}{E}=\frac{(28-28.8)^2}{28.8}+\frac{(20-19.2)^2}{19.2}+\frac{(32-31.2)^2}{31.2}+\frac{(20-20.8)^2}{20.8}$$
$$=0.022+0.033+0.021+0.031=0.107$$

נחשב קרמר:

$$r_c=\sqrt{\frac{1}{n(k-1)}\chi^2}=\sqrt{\frac{1}{100(2-1)}\cdot 0.107}=0.032$$

(סה"כ לשאלה זו - 20 נקודות)

9. פסיכולוגית חברתית מתעניינת בלקיחת סיכונים. היא מדדה לקיחת סיכונים באמצעות שאלון (ציונים נעים בין 1 ל-40), בשתי קבוצות: 10 מהמרים ו-10 סטודנטים להנדסה. להלן התוצאות שקיבלה:

	מהמרים	סטודנטים להנדסה
1	15	30
2	20	9
3	35	15
4	40	10
5	29	12
6	18	25
7	26	17
8	19	14
9	33	20
10	31	13

א. חשב/י את הממוצע, את סטיית התקן ואת AD בקבוצת הסטודנטים להנדסה (14 נק').

$$\bar{y} = \frac{\sum y_i}{N} = \frac{30+9+15+10+12+25+17+14+20+13}{10} = 16.5$$

$$S_y{}^2 = \frac{\sum (y-\bar{y})^2}{N}$$

$$= \frac{182.25+56.25+2.25+42.25+20.25+72.25+0.25+6.25+12.25+12.25}{10}$$

$$= \frac{406.5}{10} = 40.65 \rightarrow S_y = \sqrt{40.65} = 6.38$$

לפני חישוב ה-AD, יש למצוא את החציון:

נסדר את התצפיות לפי סדר: 9,10,12,13,14,15,17,20,25,30

החציון נמצא בין הערכים במיקום 5 ו-6: $\frac{14+15}{2} = 14.5$

$$AD = \frac{\sum |x - Md|}{N} = \frac{15.5+5.5+0.5+4.5+2.5+10.5+2.5+0.5+5.5+1.5}{10} = \frac{49}{10}$$

$$= 4.9$$

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

ב. חשב/י את הטווח הבין-רבעוני בקבוצת המהמרים (6 נק').

ראשית, נסדר את התצפיות לפי סדר: 15, 18, 19, 20, 26, 29, 31, 33, 35, 40

הערך במאון ה-25 (Q_1):

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

$$n_{p=25} = 0.25 \cdot 11 = 2.75$$

$$x = x_k + d(x_{k+1} - x_k)$$

$$x_{p=25} = x_2 + 0.75(x_3 - x_2) = 18 + 0.75 \cdot (19 - 18) = 18.75$$

הערך במאון ה-75 (Q_3):

$$n_{p=75} = 0.75 \cdot 11 = 8.25$$

$$x_{p=75} = x_8 + 0.25(x_9 - x_8) = 33 + 0.25 \cdot (35 - 33) = 33.5$$

הטווח הבין רבעוני: $Q_3 - Q_1 = X_{p=75} - X_{p=25} = 33.5 - 18.75 = 14.75$

הסעיף הבא בלתי תלוי בתוצאות הסעיפים הקודמים.

ג. ידוע שלקיחת סיכונים מתפלגת נורמלית בקרב מהמרים ובקרב סטודנטים להנדסה. תוחלת לקיחת הסיכונים בקרב מהמרים הינה 30 עם סטיית תקן 5.8, ובקרב סטודנטים להנדסה התוחלת הינה 18 והשונות 64.5. נבחרו מקרית מהמר אחד וסטודנט אחד להנדסה. מהי ההסתברות שמדד לקיחת הסיכונים של הסטודנט להנדסה יהיה נמוך מזה של המהמר ביותר מ-15 נקודות? (8 נק').

מכיוון שמדובר בשני משתנים שמתפלגים נורמלית עם יסוד סביר להנחה שהם בלתי-תלויים (הדגימה של סטודנט להנדסה הינה בלתי תלויה בדגימה של מהמר), נצור משתנה הפרש, X, המוגדר כציון לקיחת הסיכונים של סטודנט פחות ציון לקיחת הסיכונים של מהמהר.

תוחלת התפלגות זו תהיה: E[X]=18-30=-12

שונות המשתנה החדש תהיה: $Var[X]=5.8^2+64.5=98.14$

כדי למצוא את ההסתברות לקבל הפרש של יותר מ-15, ראשית נמצא את ציון התקן המתאים לערך 15-:

$$Z_{(x=-15)} = \frac{-15-(-12)}{\sqrt{98.14}} = \frac{-3}{9.91} = -0.30$$