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

פתרון מרצה

תאריך הבחינה: _08.02.17_

שם המרצה: ד"ר יואב קסלר

שם הקורס: _מבוא לסטטיסטיקה_

מספר הקורס: _101-1-0049_

שנה: 2017 סמסטר: א' מועד: א'

משך הבחינה: _שלוש שעות_

חומר עזר: _מחשבון מדעי_

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. הסבר/י מתי מתקיים $P(n,k)=P(n,n-k)$. נמק/י מדוע.

פרמוטציות הינן מספר אפשרויות הסידור של k איברים מובחנים, כאשר האיברים שנותרו (n-k) לא עוברים סידור. לכן, ככלל סידור של k איברים מובחנים יהיה שונה מסידור של n-k איברים מובחנים. הם יהיו שווים כאשר n-k=k, כלומר כאשר k=0.5n

$$\mathrm{P}(n,k)=\frac{n!}{(n-k)!}$$
$$\mathrm{P}(n,n-k)=\frac{n!}{(n-n+k)!}=\frac{n!}{k!}$$

המצב היחיד בו $\frac{n!}{k!}=\frac{n!}{(n-k)!}$ הינו כאשר n=2k (או k=0.5n), מכיוון שאז אנו מחלקים את n האיברים לשתי קבוצות שוות גודל, ומסדרים את האיברים של אחת מהן. מכיוון ששתיהן שוות, לא משנה איזו מהן סידרנו. כאשר הקבוצות אינן שוות גודל, מספר הסידורים יהיה שונה.

2. יעל דגמה שני משתנים אמפיריים במדגם בגודל N, וסידרה את הנתונים בשני וקטורים - X ו-Y. כל תצפית במדגם קיבלה את אותו אינדקס i בשני הווקטורים. יעל מצאה כי:

$$\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{x}_{\mathrm{i}}=\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{y}_{\mathrm{i}}=0;\qquad \mathrm{X}\cdot\mathrm{Y}=0$$

<u>נכון/ לא-נכון</u>: יעל יכולה לדעת האם ישנו קשר ליניארי בין X ל-Y על סמך נתונים אלו בלבד. נמק/י.

נכון. נתחיל מהנתון הראשון: לשני המשתנים ממוצע 0. אנו נשאלים לגבי קשר ליניארי, ולכן נבדוק כיצד נתון זה משפיע על מתאם פירסון:

$$\sum \frac{(x_i-\bar{x})(y_i-y)}{N\cdot s_x\cdot s_y} \Rightarrow \sum \frac{(x_i-0)(y_i-0)}{N\cdot s_x\cdot s_y} = \sum \frac{x_i\cdot y_i}{N\cdot s_x\cdot s_y}$$

הביטוי במונה הינו מכפלה פנימית (dot product) בין שני וקטורים, X ו-Y. הנתון השני הינו שהמכפלה הוקטורית בין X ל-Y שווה ל-0. מכאן שמתאם פירסון שווה 0.
(ניתן לענות גם דרך ציוני תקן).

3. גיא, סטודנט לסטטיסטיקה, יצר גרף היסטוגרמה של פונקציית מסת ההסתברות של משתנה המתפלג גיאומטרית. מהו היחס בין גובה כל עמודה בגרף של גיא לעמודה הקודמת לה? הסבר/י תשובתך.

היחס הינו q. כל עמודה בגרף פונקציית מסת ההסתברות מייצגת את ההסתברות לקבל ערך מסוים. ההסתברות לקבל כל ערך הינה מכפלה של p ב-q בחזקת מספר הניסיונות פחות 1:

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

בכל עמודה גדל מספר הניסיונות ב-1, ולכן כל עמודה שווה לעמודה הקודמת לה כפול q.

4. נתון מדגם בגודל N בו נמדדו שני משתנים אמפיריים, X ו-Y. <u>נכון/ לא-נכון</u>: יתכן שעבור כל תצפית במדגם מתקיים: $Z_x=\frac{Z_y}{2}$. נמק/י.

לא נכון. לא יתכן שמדובר בציוני תקן, מכיוון שבמצב הנתון סטיית התקן של לפחות אחד המשתנים המתוקננים אינה 1.
נניח ש-Z_y הינה אכן התפלגות סטנדרטית. אם כל ציון תקן ב-Z_x שווה לציון התקן ב-y חלקי 2, מדובר בטרנספורמציה ליניארית של הכפלה ב-1/2 על ציוני התקן של Z_y. משמעות הדבר שהשונות של Z_x הינה 0.25, ולכן סטיית התקן 0.5, והרי שאז אינה יכולה להיות התפלגות ציוני תקן.
שימו לב שאם משתנה *גלם* אחד הוא טרנספורמציה לינארית של השני (למשל: x=0.5y) הרי שציוני התקן שלהם זהים (Z_x=Z_y).

5. הסביר/י האם וכיצד טרנספורמציה ליניארית משפיעה על הטווח במדגם.

טרנספורמציה של הוספת קבוע (a) לא תשפיע על הטווח:

Range = X_{max}- X_{min}

Range'=(x_{max}+a)-(x_{min}+a)= X_{max}- X_{min}

עם זאת, טרנספורמציה של מכפלה בקבוע (b) כן תשפיע על הטווח - הוא יגדל פי b.

Range = X_{max}- X_{min}

Range'=bx_{max}-bx_{min}=b(x_{max}-x_{min})=b*Range

6. מאורעות A ו-B הינם זרים, ומאורעות B ו-C בלתי-תלויים. מה ניתן להסיק לגבי $P(A/_{C\cap B})$? נמק/י.

$P\left(\frac{A}{C\cap B}\right)=0$. אם A ו-B זרים, גם A והחיתוך של B עם C זרים. מכאן שאם התרחש חיתוך עם B ו-C, A לא התרחש.

<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. חוקרת הורות מתעניינת במסוגלות הורית של הורים ובמידה בה הם משתמשים בפורומי הורות באינטרנט לצורך התייעצות. לשם כך, ביקשה מ-10 הורים למלא שאלון מסוגלות הורית (ציונים נעים בין 0 ל-30), ומדדה את מספר הפעמים שהשתמשו בפורומי הורות בחודש האחרון לצורך התייעצות. להלן התוצאות שקיבלה:

	מסוגלות הורית	מספר התייעצויות בפורומי הורות
1	30	2
2	7	8
3	15	8
4	10	3
5	25	5
6	21	4
7	13	7
8	17	4
9	4	10
10	11	5

א. מהם החציון וה-AD במסוגלות הורית? (6 נק')

נסדר את הערכים לפי סדר: 4,7,10,11,13,15,17,21,25,30

$$M_d=\frac{13+15}{2}=14$$

$$AD=\frac{\sum|x-Md|}{N}=\frac{16+7+1+4+11+7+1+3+10+3}{10}=6.3$$

ב. מהו הקשר בין מסוגלות הורית למספר ההתייעצויות בפורומי הורות? חשב/י, הצג/י בדיאגרמת פיזור והסבר/י מה ניתן להסיק מהתוצאה שהתקבלה. (15 נק')

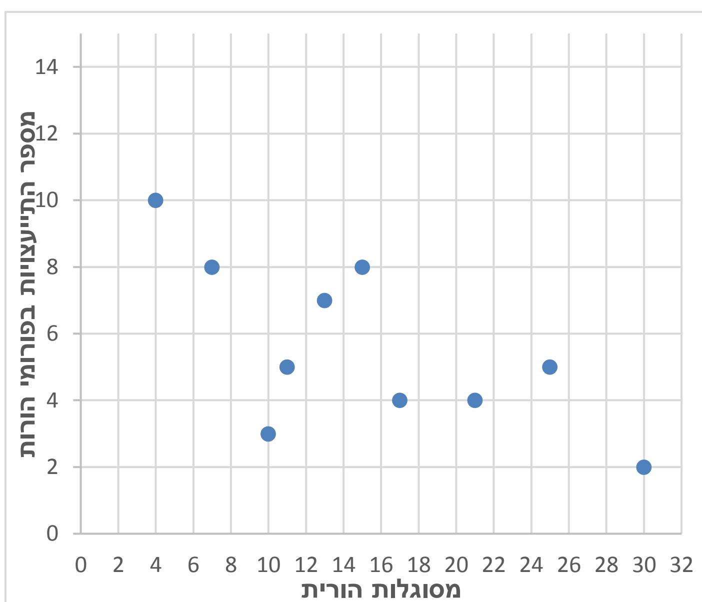
	x - מסוגלות	y - פורומים	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})\cdot(y-\bar{y})$
1	30	2	14.7	-3.6	216.09	12.96	-52.92
2	7	8	-8.3	2.4	68.89	5.76	-19.92
3	15	8	-0.3	2.4	0.09	5.76	-0.72
4	10	3	-5.3	-2.6	28.09	6.76	13.78
5	25	5	9.7	-0.6	94.09	0.36	-5.82
6	21	4	5.7	-1.6	32.49	2.56	-9.12
7	13	7	-2.3	1.4	5.29	1.96	-3.22
8	17	4	1.7	-1.6	2.89	2.56	-2.72
9	4	10	-11.3	4.4	127.69	19.36	-49.72
10	11	5	-4.3	-0.6	18.49	0.36	2.58
	$\bar{y} = 15.3$	$\bar{x} = 5.6$			$\sum$=594.1	$\sum$=58.4	$\sum$=-127.8

$$S_x = \sqrt{\frac{(x-\bar{x})^2}{N}} = \sqrt{\frac{594.1}{10}} = 7.71$$

$$S_y = \sqrt{\frac{(y-\bar{y})^2}{N}} = \sqrt{\frac{58.4}{10}} = 2.42$$

$$r_{xy} = \frac{\sum(x-\bar{x})\cdot(y-\bar{y})}{N\cdot S_x\cdot S_y} = \frac{-127.8}{10\cdot 7.71\cdot 2.42} = -0.69$$

דיאגרמת פיזור:



ניתן להסיק כי קיים קשר שלילי בין תחושת מסוגלות הורית למספר ההתייעצויות בפורומי הורות- ככל שהורים חשו מסוגלות גבוהה יותר, הם השתמשו פחות בפורומים להתייעצות. הקשר בינוני- גבוה.

ג. מהו טווח הערכים שכולל 30% מהערכים המרכזיים במספר התייעצויות בפורומי הורות? (6 נק')

30% הערכים המרכזיים בהתפלגות זמני השהייה במבחן נמצאים בין המאון ה-35 למאון ה-65.

נסדר את הערכים לפי סדר: 2,3,4,4,5,5,7,8,8,10

הערך במאון ה-35:

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

$$n_{p=35} = 0.35 \cdot 11 = 3.85$$

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

$$x_{p=35} = x_3 + 0.85(x_4 - x_3) = 4 + 0.85 \cdot (4-4) = 4$$

הערך במאון ה-65:

$$n_{p=65} = 0.65 \cdot 11 = 7.15$$

$$x_{p=65} = x_7 + 0.15(x_8 - x_7) = 7 + 0.15(8-7) = 7.15$$

35% הערכים המרכזיים בהתפלגות מספר ההתייעצויות בפורומי הורות נמצאים בטווח:

$Range=X_{p=65}-X_{p=7.15}=7.15-4=3.15$.

ד. בהנחה שמשתנה המסוגלות ההורית מתפלג נורמלית באוכלוסיה, כך שתוחלתו שווה לממוצע שהתקבל במדגם, וסטיית התקן במדגם שווה לסטיית התקן באוכלוסיה - מהו ציון המסוגלות ההורית של הורה הנמצא סטיית תקן וחצי מעל התוחלת? (6 נק')

סטיית תקן וחצי מעל התוחלת = ציון תקן 1.5:

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

$$1.5 = \frac{x - 15.3}{7.71} = 26.86$$

ה. נורית היא פסיכולוגית חדשה בצוות העונה על שאלות בפורום "הורים וילדים", ועליה לענות על שאלות בקצב של שאלה ביומיים בממוצע. מהו הטווח הבין-רבעוני של מספר הימים שצפויים לקחת לנורית לענות לראשונה לשאלה בפורום? הסבר/י את דרך החישוב, אין צורך לספק תשובה מספרית סופית. (6 נק').

ההתפלגות המתאימה לשאלה הינה אקספוננציאלית.

λ=0.5

כדי למצוא את הטווח הבין-רבעוני, נמצא את הערכים שעד אליהם נמצאים 25% ו-75% מהערכים. נעשה זאת באמצעות פונקציית ההתפלגות המצטברת.

$$F(x) = 1 - e^{-\lambda x}$$

פעם אחת נשווה את $1 - e^{-0.5x}$ ל-0.25, ופעם אחת נשווה ל-0.75. נחלץ את הערכים המתאימים לכל רבעון ונחשב את ההפרש ביניהם כדי להגיע לטווח הבין-רבעוני.

(התשובה הזו מספיקה לקבלת ניקוד מלא).

תשובה מספרית מלאה (לא חובה לקבלת ניקוד מלא):

1. Q_1:

$$F(x) = 1 - e^{-\lambda x}$$
$$1 - e^{-0.5x} = 0.25$$
$$e^{-0.5x} = 0.75$$

$$-0.5x = \ln(0.75)$$
$$x = 0.57$$

2. Q_3:

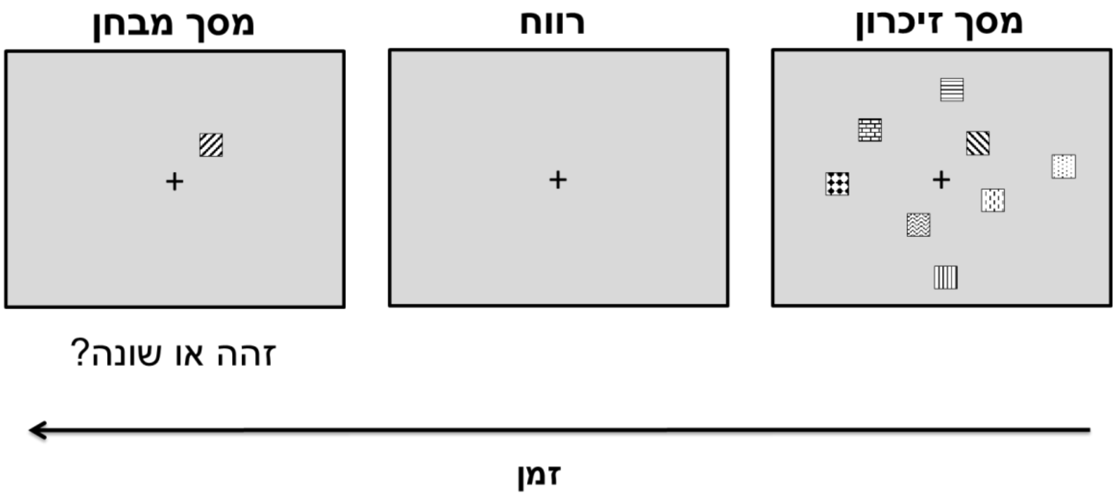
$$1 - e^{-0.5x} = 0.75$$
$$e^{-0.5x} = 0.25$$
$$-0.5x = \ln(0.25)$$
$$x = 2.77$$

הטווח הבין-רבעוני הוא:

$$Q_3 - Q_1 = 2.77 - 0.57 = 2.2\ days$$

(סה"כ: 39 נק')

8. טל מכינה מטלה קוגניטיבית ויזואלית, במסגרתה נבדקים מתבקשים לזכור צבעים של ריבועים צבעוניים. צעד בניסוי מתחיל בהצגה מהירה של מסך זיכרון, הכולל 8 ריבועים בצבעים שונים המפוזרים על המסך. לאחר מרווח זמן, מופיע מסך מבחן הכולל ריבוע אחד במיקום שהופיע בו אחד הריבועים במסך הזיכרון. על הנבדק לומר האם צבע הריבוע במסך המבחן זהה לצבע הריבוע שהופיע באותו המקום במסך הזיכרון או שונה ממנו. ב-50% מצעדי הניסוי הצבע זהה (מצורף תרשים של הניסוי).



א. בצעדים בהם צבע הריבוע במסך המבחן שונה מצבע הריבוע המתאים במסך הזיכרון, על הצבע במסך המבחן להיות חדש לחלוטין, כך שלא הופיע באף אחד מהריבועים במסך הזיכרון. טל תכנתה את המטלה כך שעל התוכנה להגריל צבע מבין 14 צבעים אפשריים. אם הצבע הופיע באחד מהריבועים במסך הזיכרון, היא מגרילה שוב. התוכנה ממשיכה להגריל צבעים עד שמתקבל צבע שלא הופיע במסך הזיכרון.

מהו הסיכוי שהתוכנה תגריל 4 צבעים או יותר לפני שיבחר צבע שלא הופיע במסך הזיכרון? על אילו הנחות מבוססת ההסתברות שחושבה? (9 נק').

ההתפלגות המתאימה היא התפלגות גיאומטרית - המשתנה מתייחס למספר ניסויים עד קבלת הצלחה בפעם הראשונה.

אנו מניחים שהניסוי הינו ניסוי ברנולי בו ההסתברות להצלחה (p) שווה בכל ניסוי (אי-תלות בין תוצאות הניסויים).

$$p = \frac{14-8}{14} = \frac{3}{7}$$

ישנן 2 דרכי פתרון:

- $P(X > 4) = q^4 = \left(\frac{4}{7}\right)^4 = 0.11$
- ניתן למצוא את ההסתברות המשלימה (עד 4 כישלונות), ולהחסיר מ-1:

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

$P(X=1)=\frac{3}{7}$

$P(X=2)=\frac{4}{7}\cdot\frac{3}{7}=\frac{12}{49}$

$P(X=3)=(\frac{4}{7})^2\cdot\frac{3}{7}=\frac{48}{343}$

$P(X=4)=\left(\frac{4}{7}\right)^3\cdot\frac{3}{7}=\frac{192}{2401}$

$P(X\leq 4)=\frac{3}{7}+\frac{12}{49}+\frac{48}{343}+\frac{192}{2401}=\frac{2145}{2405}\approx 0.89$

P(X>4)=1-P(X≤4)=1-0.89=0.11

הסיכוי שהתוכנה תדגום 4 צבעים או יותר לפני שיבחר צבע מתאים הינו 0.11.

ב. יוסי הוא נבדק בניסוי של טל. ליוסי קיבולת זיכרון עבודה ויזואלי של 3 צבעים בלבד (כלומר, הוא יכול לזכור בזמן נתון 3 ריבועים). כאשר מופיע ריבוע שאינו זוכר, הוא מנחש אקראית. מהי ההסתברות שיוסי יענה נכונה על 4 מתוך 5 צעדים בניסוי? (9 נק').

ההסתברות שיוסי זכר את הצבע במבחן הזיכרון היא $\frac{3}{8}$.

לכן, ההסתברות שיענה נכונה בצעד הינה ההסתברות שזכר את הצבע וההסתברות שלא זכר את הצבע אבל ניחש נכונה:

$$p=\frac{3}{8}+\frac{5}{8}\cdot\frac{1}{2}=\frac{11}{16}$$

כעת, נשתמש בהתפלגות בינומית כדי לחשב את ההסתברות:

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

$$P(X=4)=\binom{5}{4}\cdot\frac{11}{16}^4\cdot\frac{5}{16}=0.35$$

ההסתברות שיוסי יענה נכונה ב-4 מתוך 5 צעדים הינה 0.35.

ג. יוסי השתתף בניסוי נוסף של גל, חברה של טל. גל השתמשה באותה המטלה עם 80 צעדי ניסוי. בניסוי של גל ניתן להגיב: 'זהה', 'שונה', או 'לא יודע'. על כל תשובה נכונה מקבלים נקודה, על כל תשובה שגויה יורדות 3 נקודות, וכאשר מגיבים 'לא יודע' לא מקבלים ולא מפסידים נקודות.

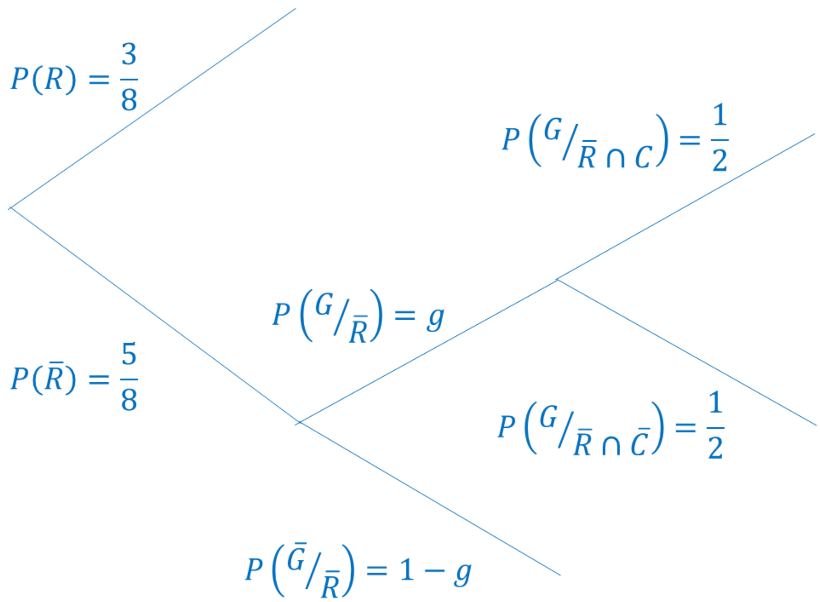
בהנחה שיוסי רוצה לסיים את הניסוי עם מספר נקודות חיובי, מהו המספר המקסימלי של צעדים בהם יוסי יכול להרשות לעצמו לנחש 'זהה' או 'שונה' אם אינו יודע את התשובה? (13 נק').

כדי להקל על החישוב, נסדר את הנתונים בעץ הסתברויות:

R - יוסי זכר את הצבע.

G - יוסי ניחש (המשלים = יוסי ענה 'לא יודע')

C - הניחוש היה נכון



מצב	x	P(x)	x·P(x)
תשובה נכונה	1	$\frac{3}{8}+\frac{5}{8}\cdot g\cdot\frac{1}{2}$	$\frac{3}{8}+\frac{5g}{16}$
תשובה שגויה	-3	$\frac{5}{8}\cdot\frac{g}{2}$	$\frac{-15g}{16}$
'לא יודע'	0	$\frac{5-5g}{8}$	0

$$E[X]=\frac{3}{8}+\frac{5g}{16}-\frac{15g}{16}$$

נמצא מתי התוחלת גדולה מ- 0:

$$\frac{3}{8}+\frac{5g}{16}>\frac{15g}{16}$$

$$6+5g>15g$$

$$g<0.6$$

יוסי יכול לנחש בפחות מ-60% מהמקרים בהם אינו יודע את התשובה ולסיים עם מספר נקודות חיובי. ישנם 80 צעדים, ויוסי לא ידע את התשובה בכ-50 צעדים מתוכם, לכן הוא צריך לנחש בפחות מ-30 צעדים כדי לסיים עם מספר נקודות חיובי.

(סה"כ 31 נק')

ב ה צ ל ח ה !