תרגילי חזרה + פתרונות – 2012

רווחי סמך ובדיקת השערות

1. להפרש תוחלות של שני מדגמים תלויים (מזווגים)

שאלה 1:

נלקח מדגם מקרי של 10 מעשנים אשר השתתפו בתוכנית לגמילה מעישון באמצאות היפנוזה. להלן הנתונים על מספר הסיגריות היומי שעישן כל נבדק לפני ההיפנוזה וחודש לאחר ההיפנוזה.

לפני	25	30	35	25	15	40	20	30	30	40
אחרי	25	10	5	10	20	15	5	10	20	20

ידוע כי מספר הסיגריות מתפלג נורמלית בכל אחת מהאוכלוסיות.

א. האם ההיפנוזה הפחיתה את כמות העישון? בדוק ברמת מובהקות $\alpha = 0.025$

ב. חשב רווח בר-סמך לתוחלת כמות ההפחתה בעישון בעקבות ההיפנוזה ברמת סמך 0.95.

שאלה 2

במספר חנויות בדקו מחירים של סלסילות שי לחגים עם תכולה דומה, והשוו למחירי רכיבי התכולה כשהם נקנים בנפרד.

חנות	א	ב	ג	ד	ה
מחיר סלסלה	109	99	99	119	109
מחיר התכולה בנפרד	82	75	70	85	84

א. מהי רמת המובהקות המינימאלית עבורה ניתן להסיק כי מחיר הסלסילה גדול באופן מובהק ממחיר מרכיביה ביותר מ-22 ₪.

ב. מצא רווח סמך של 98% לתוחלת ההבדלים במחירים בין סלסלה מוכנה ומרכיביה.

שאלה 3

ציונים במבחן מסוים מתפלגים נורמלית. ישנם נתונים לגבי ציונים של 6 סטודנטים שעשו שני מועדים:
מועד א : 51,40,38,59,55,46; מועד ב: 91,69,70,81,83,79.

בנה רווח סמך להפרש התוחלות עם רמת סמך 90%.

פתרונות

שאלה 1:

א. נמצא את הפרש מספר הסיגריות:

לפני	25	30	35	25	15	40	20	30	30	40
אחרי	25	10	5	10	20	15	5	10	20	20
הפרש (d)	0	20	30	15	-5	25	15	20	10	20

ברצוננו לבדוק האם יש ההיפנוזה הפחיתה את צריכת הסיגריות:
$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 > 0$$

נמצא P value: $\bar{d} = 15$, $S_d = 10.80$

$$p\_value: P\left(\bar{d} > 15 \mid H_0\right) = P\left(t_9 > \frac{15-0}{10.80/\sqrt{10}}\right) = P\left(t_9 > 4.39\right) \Rightarrow 0.0005 < p\_v < 0.001$$

לכן עבור רמת מובהקות של 0.025 נדחה את השערת האפס.

ב. נחשב רווח סמך כאשר השונות אינה ידועה: $\bar{d} - t_{n-1,1-\alpha/2} \cdot \frac{S_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{n-1,1-\alpha/2} \cdot \frac{S_d}{\sqrt{n}}$

$$t_{9,0.975} = 2.262$$

נציב את הנתונים ונקבל: $15 - 2.262 \cdot \frac{10.8}{\sqrt{10}} < \mu_d < 15 + 2.262 \cdot \frac{10.8}{\sqrt{10}}$

$$7.27 < \mu_d < 22.72$$

מכיוון שהתוחלת עפ"י השערת האפס (0) אינה נמצאת ברווח נדחה את השערת האפס.

שאלה 2

א. נמצא את הפרש המחירים:

חנות	א	ב	ג	ד	ה
מחיר סלסלה	109	99	99	119	109
מחיר התכולה בנפרד	82	75	70	85	84
הפרש (d)	27	24	29	34	25

$$\bar{d} = 27.8$$
$$S_d = 3.96$$
$$H_0 : \mu_d = 22$$
$$H_1 : \mu_d > 22$$
$$p.v. = P(\bar{d} \geq 27.8 \mid H_0) = P(t_4 \geq \frac{27.8-22}{3.96/\sqrt{5}}) = P(t_4 \geq 3.28) = 1 - P(t_4 \leq 3.28)$$
$$\Rightarrow 0.01 < p.v. < 0.025$$

כלומר רמת המובהקות המינימלית היא 0.025

ב. רווח סמך של 98% ל- μ_d : $\alpha = 0.02 \Leftarrow$ $t_{n-1,1-\frac{\alpha}{2}} = t_{4,0.99} = 3.75$

$$27.8 - 3.75 \cdot \frac{3.96}{\sqrt{5}} \leq \mu_d \leq 27.8 + 3.75 \cdot \frac{3.96}{\sqrt{5}}$$

$$\Leftarrow 21.16 \leq \mu_d \leq 34.44$$

ג. 22 הוא בתוך רו"ס ולכן לא דוחים את השערת האפס בר"מ 0.02 או בר"מ 0.01 במבחן חד-צדדי כמו בסעיף א' ולכן יש התאמה בין ההכרעה בסעיף א' (לא לדחות את השערת האפס עבור ר"מ 0.01)

לבין רו"ס.

<u>שאלה 3</u>

$$D : x_b - x_a : -33, -28, -22, -32, -29, -40.$$
$$\bar{D} = -30\frac{2}{3}$$

$$\overline{s_d^{\,2}} = \frac{\sum (D_i - \bar{D})^2}{n-1} = 35.86$$

$$t_{5,0.95} = 2.445 \quad 30\frac{2}{3} - \frac{5.989}{\sqrt{6}} * t_{5,0.95} < \mu_d < 30\frac{2}{3} + \frac{5.989}{\sqrt{6}} * t_{5,0.95}$$

$$25.72 < \mu_d < 35.6$$

2. *לפרופורציה במדגם אחד*

שאלה 1

בכדי לאמוד את אחוז הגירושין בארה"ב נלקח מדגם מקרי של 625 איש ונמצא ש-125 מהם גרושים.

א. בנו רו"ס לאחוז הגירושין ברמת ביטחון של 95%.

ב. איך ישתנה אורך רו"ס עבור רמת ביטחון 90% ? 99% ? (ענו ללא חישוב נוסף)

ג. אם נגדיל את המדגם פי 4 כיצד יושפע אורך רו"ס?

ד. מהו גודל המדגם הדרוש כדי להקטין את אורך הרווח פי 2 ?

ה. אמדו את אחוז האנשים חסרי השכלה גבוהה באוכלוסיה ברמת סמך 95%.

שאלה 2

במאפיה נשרפים 60% מכיכרות הלחם בכל יום. לאחר קורס הכשרה מיוחד לאופים נלקח מדגם של 100 כיכרות ונמצא כי 53 כיכרות נשרפו. האם הקורס עזר? בדקו עבור $\alpha = 0.05$.

למרות שרווח הסמך נהיה צר יותר מכיוון שהמדגם גדול יותר, לא נדחה את H_0.

שאלה 3

ידוע כי 80% מהאוכלוסייה עובדים. בעקבות תוכנית ויסקונסין יש הטוענים כי אחוז העובדים באוכלוסיה עלה ויש הטוענים כי אחוז העובדים לא השתנה. כדי לבדוק זאת נלקח מדגם מקרי של 400 איש ונמצא כי 336 מתוכם עובדים. מי צודק? האם ייתכן כי שני הצדדים צודקים?

פתרונות

שאלה 1

א. נתון: $\hat{p} = \frac{125}{625} = 0.2$, $n = 625$

נוסחא לרו"ס לפרופורציה: $\hat{p} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} \le p \le \hat{p} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$

95% $\Leftarrow$ $Z_{1-\frac{\alpha}{2}} = 1.96$

$$0.17 \le p \le 0.23 \quad \Leftarrow \quad p \in 0.2 \pm 1.96\sqrt{\frac{0.2\cdot 0.8}{625}}$$

ב. אם נקטין את רמת הביטחון ל-90% גם רווח הסמך יקטן, ולחילופין אם נגדיל את רמת הביטחון ל-99% רווח הסמך יגדל.

ג. מכיוון שהפרופורציה עצמה תלויה בגודל n ולא רק רווח הסמך, אם נגדיל את גודל המדגם לא ברור כיצד רווח הסמך ישתנה. הדבר שונה מרווח הסמך עבור תוחלת שבו אם נגדיל את n רווח הסמך יקטן.

ד. מכיוון שכפי שציינו לעיל שינוי גודל המדגם משפיע גם על הפרופורציה, ואין לנו שום מידע לגבי הפרופורציה באוכלוסיה נציב $\hat{p} = 0.5$ עבורו הביטוי $\hat{p}(1-\hat{p})$ מקבל ערך מקסימאלי כך שנקבל חסם עליון ל-n.

כדי להקטין את אורך הרווח פי 2 אפשר פשוט להקטין את הסטייה מ- $\hat{p}$ פי 2 , כלומר:

$$1.96 \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}} = 0.5 \cdot 0.03$$

$\sqrt{\frac{0.5 \cdot 0.5}{n}} = \frac{0.015}{1.96} \Leftarrow n = 4268.4$. נעגל ונקבל כי יש לקחת מדגם בגודל של 4269 כדי להקטין את הרווח פי 2 .

ה. נתון: $\hat{p} = \frac{500}{625} = 0.8$, $n = 625$

נוסחא לרו״ס לפרופורציה: $\hat{p} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \le p \le \hat{p} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$

95% $\Leftarrow$ $Z_{1-\frac{\alpha}{2}} = 1.96$

$p \in 0.8 \pm 1.96\sqrt{\frac{0.2 \cdot 0.8}{625}} \Leftarrow 0.77 \le p \le 0.83$. נשים לב כי רווח סמך זה הוא המשלים של רווח הסמך מסעיף א׳ מכיוון הוא רווח סמך להסתברות המשלימה, וכן מכיוון שהסטייה מהפרופורציה שווה.

<u>שאלה 2</u>

ברצוננו לבדוק את ההשערות: $\begin{matrix} H_0 : p \ge 0.6 \\ H_1 : p < 0.6 \end{matrix}$.

נשתמש במבחן דו צדדי: דחה H_0 אם $\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$. כאשר $\hat{p} = \frac{53}{100} = 0.53$.

נבדוק האם $\hat{p} < 0.6 - Z_{0.95} \cdot \sqrt{\frac{0.6 \cdot 0.4}{100}} = 0.6 - 1.645 \cdot 0.049 = 0.52$.

אי השוויון לא מתקיים ולכן לא נדחה את השערת האפס ונסיק כי הקורס לא עזר.

<u>שאלה 3</u>

ברצוננו לבדוק את ההשערות: $\begin{matrix} H_0: p \le 0.8 \\ H_1: p > 0.8 \end{matrix}$.

$$\hat{p} = \frac{336}{400} = 0.84 \ , \ p = 0.8 \quad \hat{p} \sim N(p, \frac{p(1-p)}{n})$$

נחשב .P.V:

$$P.V. = P(\hat{p} \ge 0.84 \mid H_0) = P(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \ge \frac{0.84-0.8}{0.02}) = P(Z \ge 2) = 1 - P(Z < 2) = 0.0228$$

לכן אם ר"מ היא 0.05 נדחה את השערת האפס ואם ר"מ היא 0.01 לא נדחה. ז"א ששני הצדדים צודקים (תלוי איזו רמת מובהקות בוחרים).

<u>הסתברות מותנה ואי תלות, נוסחת הכפל, נוסחת הסתברות השלמה ובייס.</u>

- **הסתברות מותנה:** A ו-B שני מאורעות: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- **אי תלות:** $P(A \mid B) = P(A), \quad P(A \cap B) = P(A)P(B)$
- **חוק הכפל:**

$$P(\bigcap_{k=1}^{n} A_k) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1})$$

כאשר: $P(\bigcap_{i=1}^{k} A_i) > 0, \ \forall k = 1 \ldots n-1$, A_k מאורעות.

- **נוסחת ההסתברות השלמה:** $P(A) = \sum_{k=1}^{n} P(A \mid B_k)P(B_k)$

כאשר: $i \ne j$: $\sum_{k=1}^{n} P(B_k) = 1, \quad B_i \cap B_j = \phi$

- **נוסחת בייס:** $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$

<u>שאלה 11</u>

באוכלוסיה של 48% זכרים ו-52% נקבות, 5% מהזכרים עוורי צבעים ו 0.25% מהנקבות עוורות צבעים.

א. אם בוחרים אדם מהאוכלוסייה מה ההסתברות שהוא (היא) עיוורי צבעים?

ב. אם בוחרים אדם (באקראי) עיוור צבעים, מה ההסתברות שנשחר זכר?

שאלה 12

במאפייה מסוימת עובדים 3 אופים A,B,C כאשר האופים מכינים את " עוגת הבית " העוגה אינה תופחת בהסתברות : 0.02,0.03, 0.05 בהתאמה. A מכין 50% מהעוגות, B מכין 30% מהעוגות וC מכין את השאר. איזה % מכישלונות העוגה נגרם ע"י A?

שאלה 13

אם במבחן סטודנט קורא את ההוראות הוא מקבל ציונים A,B,C,D בהסתברות $\frac{1}{8},\frac{1}{4},\frac{3}{8},\frac{1}{4}$ בהתאמה.

אם הוא לא קורא את ההוראות, דבר שקורא ב 75% מהמקרים ההסתברויות הם: $\frac{1}{10},\frac{1}{5},\frac{2}{5},\frac{3}{10}$

א) מה ההסתברות שהסטודנט יקבל ציון B?
ב) מה ההסתברות שסטודנט שקיבל ציון B קרא את ההוראות?

שאלה 14

אדם ניגש לבצע 3 הגרלות. לפניו 2 מכונות, במכונה A הסיכוי לזכות : 0.6. במכונה B הסיכוי לזכות : 0.3. להגרלה הראשונה האדם בוחר את המכונה באופן מקרי. בכל שלב , אם הוא זוכה הוא מגריל שוב מאותה מכונה, ולא הוא עובר למכונה הבאה.

א. מה הסיכוי שההגרלה השנייה תבוצע במכונה A.
ב. כנ"ל, אם ידוע שההגרלה השלישית בוצעה במכונה A.

שאלה 15:

במפעל פעולות שתי מכונות A ו-B . 10% מתוצרת המפעל מיוצרת במכונה A ו-90% במכונה B. 1% מהמוצרים המיוצרים במכונה A ו 5% מהמוצרים המיוצרים במכונה B הם פגומים.

א. נבחר מוצר אקראי, מה ההסתברות שהוא פגום?
ב. נמצא מוצר שהוא פגום, מה ההסתברות שיוצר במכונה A?
ג. אחרי ביקור של טכנאי שמטפל במכונה B , מוצאים ש 1.9% ממוצרי המפעל הם פגומים. מה עכשיו ההסתברות שמוצר המיוצר במכונה B יהיה פגום?

<u>שאלה 16</u>

באוכלוסיה מסוימת 30% מעשנים, 20% בעלי עודף משקל, 20% סובלים מבעיות לב.

מתוך אלו הסובלים מבעיות לב, 40% מעשנים ו40% בעלי עודף משקל.

מצא:

א. את ההסתברות שמעשן יסבול מבעיות לב ואת ההסתברות שמישהו בעל עודף משקל יסבול מבעיות לב.
ב. מתוך המעשנים באכלוסיה הכללית, 20% בעלי עודף משקל, אך מתוך המעשנים הסובלים מבעיות לב 50% בעלי עודף משקל.
ג. מצא את ההסתברות שמישהו שגם מעשן וגם בעל עודף משקל, יסבול מבעיות לב.
ד. בנוסף, מצא את ההסתברות שמעשן בעל משקל רגיל יסבול מבעיות לב.

<u>שאלה 17:</u>

הוכח או הפרך:

אם A ו-B בת"ל, אזי גם A^c ו B^c בת"ל.

אם A לא תלוי ב B ,אזי B לא תלוי בA.

<u>שאלה 18</u>

בכד א שלושה כדורים לבנים ואחד שחור. בכד ב שלושה כדורים שחורים ואחד לבן.

בוחרים כד אקראי ומוצאים ממנו כדור אחד. אם הכדור הוא לבן, מוצאים מאותו כד כדור נוסף.(ללא החזרה) אם הכדור השני הוא שחור, עוברים לכד שני ומוצאים ממנו כדור אחד.

אם בסופו של דבר שני הכדורים שבידנו באותו הצבע, מה ההסתברות שהכד שבחרנו בהתחלה היה כד א ?

<u>שאלה 19</u>

באחת מחברות הביטוח הישיר, בשל עומס פניות, רק 60% מהפניות נענות באופן מיידי.

שאר הפונים מתבקשים להשאיר את מס הטל" . ב-75% מהמקרים חוזר נציג חברת הביטוח לפונה באותו יום ובשאר ביום למחרת.

הסיכוי שפונה ירכוש בחברה הוא: 0.8 אם נענה מיד, 0.6 אם חזרו אליו באותו יום, ו-0.4 אם חזרו אליו למחרת.

מה ההסתברות שאדם הפונה לחברת ביטוח ירכוש בה ביטוח?

ידוע כי אדם רכש ביטוח בחברה. מה ההסתברות שהשאיר את מס הטל" וחזרו אליו באותו יום?

האם רכישת ביטוח והשארת מספר לטלפון בת"ל –נמק !

<u>שאלה 20</u>

הוכח:

אם A ו- B בת"ל, וכן A ו- C בת"ל, ובנוסף B ו- C זרים (כלומר: $B\cap C=\varnothing$) אזי מתקיים:

A ו- $B\cup C$ בת"ל.

רמז: למעשה, עלינו להוכיח כי- $P\left[A\cap(B\cup C)\right]=P(A)P(B\cup C)$

פתרונות

שאלה 11

נתון: A= אדם עיוור צבעים.

B= זכר   P(B)=0.48

B^c=נקבה $P(B^c)=0.52$

$P(A|B^c)=0.0025$ $P(A|B)=0.05$

א) לפי נוסחת ההסתברות השלמה $P(A)=(0.05)\cdot(0.48)+(0.0025)\cdot(0.52)=0.0253$

ב) לפי בייס $P(B|A)=\dfrac{(0.05)(0.48)}{(0.0253)}=0.9486$

שאלה 12

יהי D המאורע המסמן כשלון.

נתונים:

$$P(A)=0.5 \quad P(B)=0.3 \quad P(C)=0.2$$
$$P(D|A)=0.02 \quad P(D|B)=0.03 \quad P(D|C)=0.05$$

צ"ל : P(A|D).   לפי נוסחת בייס ולפי נוסחת ההסתברות השלמה:

$$\mathrm{P}(\mathrm{A}\,|\,D)=\frac{(0.02)(0.5)}{(0.02)(0.5)+(0.03)(0.3)+(0.05)(0.2)}=0.3448$$

כלומר 34.48% מהכישלונות נגרמים ע"י A.

שאלה 13

נתונים: K= קרא את ההוראות

$$P(B|K)=\frac{1}{4}$$
$$P(B|K^c)=\frac{1}{5}$$

א) לפי נוסחת ההסתברות השלמה $P(B)=\frac{1}{4}(0.25)+\frac{1}{5}(0.75)=0.2125$

ב) לפי בייס $$P(K \mid B) = \frac{P(B \mid K)P(K)}{P(B)} = \frac{\frac{1}{4}(0.25)}{0.2125} = 0.294$$

<u>שאלה 14</u>

א. ההסתברות לבחור מכונה : 0.5.

$$P(A_2) = P(A_1 \cap A_2) + P(B_1 \cap A_2) =$$

=0.5*0.6+0.5(1-0.3)=0.65

ב.

$$P(A_2 / A_3) = \frac{P(A_3 / A_2) * P(A_2)}{P(A_3)} = \frac{0.6 * 0.65}{0.635} = 0.614$$

<u>שאלה 15:</u>

נגדיר את המאורעות הבאים:

C - המוצר שנבחר פגום.

A - המוצר הנבחר יוצר במכונה A

B - המוצר הנבחר יוצר במכונה B

עפ"י נוסחת ההסתברות השלמה מתקיים:

$$p(c) = p(c \mid a)\,p(a) + p(c \mid b)\,p(b) =$$

0.9*0.05 +0.1*0.01=0.046

ב. עפ"י נוסחת בייס:

$$p(A \mid C) = \frac{p(A)\,p(C \mid A)}{p(C)}$$

ולכן : $$P(A \mid C) = \frac{0.1 \cdot 0.01}{0.046} = 0.0217$$

ג.שוב נעשה שימוש בנוסחת ההסתברות השלמה ונקבל:

$$0.019 = 0.1 \cdot 0.01 + 0.9 \cdot X$$
$$\Downarrow$$
$$X = 0.02$$

<u>שאלה 16:</u>

נסמן את המאורעות:

אדם סובל מבעיות לב, B-אדם מעשן, C-אדם בעל עודף משקל.

עפ"י נוסחת בייס:

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)} = \frac{0.2 \cdot 0.4}{0.3} = 0.2666$$
$$p(A|C) = \frac{p(A)p(C|A)}{p(C)} = \frac{0.2 \cdot 0.4}{0.2} = 0.4$$

ב. בהסתמך על הכללת נוסחת הסתברות חיתוך המאורעות:

$$p(A|B \cap C) = \frac{p(A \cap B \cap C)}{p(B \cap C)} = \frac{p(A)p(B|A)p(C|B \cap A)}{p(B \cap C)} =$$
$$\frac{0.2 \cdot 0.4 \cdot 0.5}{0.3 \cdot 0.2} = 0.6666$$

מצא את ההסתברות שמעשן בעל משקל רגיל יסבול מבעיות לב:

עפ"י נוסחת ההסתברות השלמה וסעיף א':

$$p(A|B) = p((A|B)|C)p(C) + p((A|B)|C^c) =$$
$$0.2666 = 0.666 \cdot 0.2 + X \cdot 0.8$$
$$\Downarrow$$
$$X = 0.1666$$

שאלה 17:

הוכחה :

$$p(A^c \mid B^c) = \frac{p(A^c \cap B^c)}{p(B^c)} = \frac{p[(A \cup B)^c]}{p(B^c)} =$$
$$\frac{1-p(A \cup B)}{p(B^c)} = \frac{1-p(A)-p(B)+p(A \cap B)}{1-p(B)} = \frac{1-p(A)-p(B)+p(A)p(B)}{1-p(B)} =$$
$$\frac{1-p(A)-p(B)(1-p(A))}{1-p(B)} = \frac{(1-p(A))(1-p(B))}{1-p(B)} = 1-p(A) = p(A^c)$$

ב. הוכחה:

$$p(B \mid A) = \frac{p(B)p(A \mid B)}{p(A)} = \frac{p(B)p(A)}{p(A)} = p(B)$$

שימו לב לאופן השימוש בנוסחת בייס ולאופן השימוש בעובדה ש A לא תלוי ב- B.

שאלה 18

נגדיר את המאורעות הבאים:

A-הכדור הראשון הוצא מכד A

B – הכדור הראשון הוצא מכד B

C-הכדורים שבידנו באותו צבע.

נשתמש בנוסחת ההסתברות השלמה ונקבל:

$$p(C) = p(C \cap A) + p(C \cap B) =$$
$$\frac{3}{4}\cdot\frac{2}{3}+\frac{1}{4}\cdot\frac{3}{4}+\frac{3}{4}\cdot\frac{1}{4} = 0.875$$

המחובר הראשון מציין את ההסתברות שהכד הראשון שנבחר היה כד א' וכן שבידנו שני כדורים לבנים. מחובר שני מציין את ההסתברות שכד א' נבחר ראשון וכן בידנו שני שחורים, מחובר אחרון מציין את ההסתברות שכד ב' נבחר ראשון ובידנו שני שחורים.

כעת נשתמש בנוסחה הבאה: $p(A|C)=\frac{p(A\cap C)}{p(C)}$

ונקבל:

$$p(A\cap C)=0.6875;$$
$$p(A|C)=\frac{0.6875}{0.875}=0.785$$

שאלה 19

נגדיר את המאורעות :

A- אדם רכש ביטוח.

B – אדם השאיר את מס הטל' שלו.

C– אדם נענה באופן מיידי, D- אדם השאיר טל' וחזרו אליו באותו היום.

E – אדם השאיר טל' וחזרו אליו יום למחרת.

עפ"י נוסחת ההסתברות השלמה נקבל:

$$p(A)=p(A|C)p(C)+p(A|D)p(D)+p(A|E)p(E)=$$
$$0.8\cdot 0.6+0.4\cdot 0.75\cdot 0.6+0.4\cdot 0.25\cdot 0.4=0.7$$

$$p(D|A)=\frac{p(D)p(A|D)}{p(A)}=\frac{0.3\cdot 0.6}{0.7}=0.257$$

נבדוק עפ"י הגדרה:

$$p(A)=0.7$$
$$p(B)=0.4$$
$$p(A\cap B)=(0.4\cdot 0.25\cdot 0.4)+(0.4\cdot 0.75\cdot 0.6)=0.22$$
$$\Downarrow$$
$$p(A\cap B)=0.22\neq p(A)p(B)=0.28$$

ולכן הם תלויים.

<u>שאלה 20</u>

הוכחה

$$
\begin{aligned}
&p(A\cap(B\cup C)) = p((A\cap B)\cup(A\cap C)) = \\
&p(A\cap B) + p(A\cap C) - p((A\cap B)\cap(A\cap C)) = \\
&p(A\cap B) + p(A\cap C) - p(A\cap B\cap C) = \\
&p(A\cap B) + p(A\cap C) - p(\varnothing) = p(A\cap B) + p(A\cap C) \\
&= p(A)p(B) + p(A)p(C) = p(A)(p(B) + p(C)) = \\
&p(A)p(B\cup C)
\end{aligned}
$$

הערה: בהוכחה הסתמכנו על אסוציאטיביות החיתוך, חוק הפילוג לגבי החיתוך, וכן על אקסיומות הסתברות.

משתנים מקריים בדידים

- פונקצית ההתפלגות המצטברת F של משתנה מקרי X מוגדרת לכל מספר ממשי $b\in(-\infty,\infty)$: $F(b)=P(X\le 5)$.

תכונות: א. F היא פונקציה לא יורדת, כלומר : $a<b\Rightarrow F(a)\le F(b)$

ב. $F(b)\xrightarrow[b\to\infty]{}1$

ג. $F(b)\xrightarrow[b\to-\infty]{}1$

ב. F רציפה מימין. כלומר לכל b ולכל סדרה יורדת b_n, כאשר $n\ge 1$, המתכנסת ל-b, מתקיים $\lim\limits_{b\to\infty}F(b_n)=F(b)$.

- $E(g(x))=\sum_{k=-\infty}^{\infty}g(k)*P(X=k)$

  $E(C+K*X)=C+K*E(X)$

  $E(E(X))=E(X)$

- $Var(X)=E((X-E(X))^2)=E(X^2)-E(X)^2$

  $Var(C+K*X)=K^2Var(X)$

שאלה 1

מטילים שתי קוביות סימטריות. יהי Y – סכום ההטלות. מצאו התפלגות Y.

שאלה 2

נתונה ההתפלגות של X ע"י:

7	6	5	k
5/15	2/15	8/15	$P(X=k)$

א. מצא את התוחלת והשונות של X.

ב. נגדיר משתנה חדש: $Y=3X+120$. מצא את התוחלת והשונות של Y.

שאלה 3

בכובע נמצאים שישה תלושי הגרלה. על שלושה רשום מספר 0. על שניים רשום מס' 20. ועל אחד רשום מס' 40. כל מהמר משקיע $20 על מנת להשתתף בהגרלה. עליו למשוך שני תלושים מן הכובע ומקבל סכום דולרים אשר שווה לממוצע של שני המספרים. חשב:

א. את ההתפלגות הסכום שהמהמר מקבל.

ב. את תוחלת הרווח ושונות הרווח.

שאלה 4

משתנה מקרי X מקבל את הערכים ...0,1,2 בהסתברות $i=0,1,2...$ $P(X=i)=\dfrac{C}{3^i},$

א. מצא את הערך של הקבוע C.

ב. את התוחלת של X.

ג. את ההסתברות של X>5

ד. את ההסתברות ש-X לא זוגי.

שאלה 5:

פונקצית ההתפלגות המצטברת של משתנה מקרי X נתונה ע"י:

$$F(X)=\begin{cases} 0 & x<0 \\ x/2 & 0\le x<1 \\ 2/3 & 1\le x\le 2 \\ 11/12 & 2\le x<3 \\ 1 & 3\le x \end{cases}$$

חשב: $P(x<3),\ P(x=1),\ P(x>1/2),\ P(2<x<4)$

שאלה 6:

משתנה מקרי X מקבל את הערכים 0,2,4,6,8 בהסתברויות:

i	0	2	4	6	8
$P(X=i)$	1/9	2/9	3/9	2/9	1/9

מצא את התפלגותם ואת תוחלתם של המשתנים המקריים

$Y=(X-2)/(X+2),\ Z=(X-2)^2$

שאלה 7:

בכד יש ארבעה כדורים שחורים, שלושה כדורים לבנים ושני כדורים אדומים. מוציאים מתוך הכד מדגם של ארבעה כדורים באופן מקרי וללא החזרה. נגדיר: X- מספר הצבעים השונים המופיעים במדגם.

א. מצאו התפלגות X.

ב. חשבו תוחלת ושונות של X, ותוחלת של X^3+X .

ג. נגדיר Y – מספר הכדורים האדומים במדגם. חשבו תוחלת ושונות X.

ד. מסדרים באקראי את ארבעת הכדורים שבמדגם בשורה. מהי ההסתברות שבשורה יש שני כדורים אדומים הנמצאים אחד ליד השני?

סוגים של התפלגויות בדידות

התפלגות בינומית $X \sim Bin(n,p)$

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

מ"מ X = מספר ההצלחות ב-n ניסויים ברנוליים (0 או 1). כאשר ההסתברות להצלחה בניסוי בודד =P.

P(X=k) = ההסתברות לk הצלחות מתוך n. (k=0,...n)

$$E(X)=n\cdot p \qquad V(X)=n\cdot p\cdot(1-p)$$

שאלה 8

בניתוח מסויים מצליחים לרפא 70% מהחולים. מבצעים 5 ניתוחים. חשבו את ההסתברויות הבאות:

א) שבדיוק 3 ניתוחים הצליחו.
ב) שלפחות 3 ניתוחים הצליחו.
ג) שלכל היותר 3 ניתוחים הצליחו.

שאלה 9

מתוך סקר שנערך בין תושבי עיר מסוימת התברר כי 45% מהתושבים היו בעד פתיחת הקניון בעיר, 30% נגד ו- 25% אין דעה בעניין. בוחרים באקראי 4 אנשים תושבי העיר. מה ההסתברות ש:

א) ארבעתם בעד פתיחת הקניון ?
ב) שלפחות אחד יהיה נגד הפתיחה ?
ג) שלארבעתם תהיה דעה מסוימת ?
ד) שלפחות לאחד מהם לא תהיה דעה בנידון?

שאלה 10

ידוע כי 45% מתושבי מדינה מסוימת מעשנים.

א) בוחרים באקראי 6 אנשים. מה ההסתברות שבניהם מס' המעשנים= מס ' הלא מעשנים?
ב) בוחרים 8 אנשים מה ההסתברות שבניהם מספר המעשנים שונה ממספר הלא מעשנים?

שאלה 11

רוני מקבל דמי כיס פעם בחודש ואז הוא מחליט אם לקנות לעצמו ספר חדש או לא. ההסתברות שרוני יקנה ספר חדש היא 0.6 . נניח שבהתחלה לא היה לרוני אף ספר.

א) מהי ההסתברות שאחרי 10 חודשים, יש לרוני 6 ספרים?
ב) מהי תוחלת ושונות מספר הספרים שיש לרוני בתום שנה?
ג) מה ההסתברות שלרוני יהיו 4 ספרים בדיוק לאחר חצי שנה?

התפלגות גיאומטרית $X \sim Geo(p)$

$$P(X=k)=(1-P)^{k-1}\cdot P \qquad k=1,..,n$$

מ"מ X = מספר הניסויים עד להצלחה הראשונה (כולל ההצלחה הראשונה).

$$E(X)=\frac{1}{p} \qquad V(X)=\frac{1-p}{p^2}$$

שאלה 12

סטודנט חייב במשך הסמסטר לעבור 5 מבחנים כדי להבחן במבחן הסופי. ההסתברות לעבור כל מבחן, בלי קשר למבחנים האחרים, היא 0.7 . אם הוא לא עובר איזשהו מבחן הוא לא ממשיך הלאה. מה ההסתברות

א) שהוא יכשל במבחן ה-5 ?
ב) שהוא יעבור פחות מ-3 מבחנים?

שאלה 13

בסוף כל דקה של שיחת טלפון מחליטים אם להמשיך או לא. מה ההסתברות שאדם שמדבר עכשיו כבר n דקות ידבר עוד 10 דקות בלבד, כאשר ההסתברות שימשיך כל דקה היא 0.7

שאלה 14

אדם מנסה לחייג לאוניברסיטה- מניסיונות קודמים ידוע כי ההסתברות לקבל מענה בכל ניסיון חיוג הוא 0.25 . אדם שמחייג לאוניברסיטה ממשיך לחייג עד שמתקבל מענה- מה ההסתברות

א) שיקבל מענה בניסיון ה-9?
ב) שיקבל מענה לאחר יותר מ 4 ניסיונות חיוג?
ג) חשבו את התוחלת של משתנה מקרי זה.

התפלגות פואסון $X \sim Poi(\lambda)$

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

$$E(X)=V(X)=\lambda$$

שאלה 15
ידוע שמספר הפניות לדקה ל 144 מתפלג פואסון עם $\lambda=5$ מה ההסתברות
א) שבין 10:00 ל 10:01 לא תתקבל אף פניה ?
ב) שבדקה זו תתקבלנה לכל היותר 3 פניות?
ג) שבין 10:00 ל 10:05 תתקבלנה 4 פניות?

<u>שאלה 16</u>

לחברת דיוור יש 10,000 לקוחות. אם ידוע שבממוצע 20 אנשים מבקשים לעזוב את השרות כל חודש, מה ההסתברות שבחודש אחד יעזבו 30 אנשים?

<u>שאלה 17</u>

בליל קיץ ניתן לצפות בכוכב נופל אחד כל 10 דקות. מה ההסתברות לצפות ב2 כוכבים נופלים בטווח של 15 דקות?

<u>התפלגות בינומית שלילית</u> $X \sim NB(r, p)$

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \qquad , k = r, r+1....$$

$$E(X) = \frac{r}{p} \qquad V(X) = \frac{r(1-p)}{p^2}$$

מ"מ X הוא מספר הניסויים שיש לבצע על מנת לקבל בדיוק r הצלחות.

<u>שאלה 18</u>

חשבו את התוחלת ואת השונות של מספר הפעמים שיש להטיל קובייה עד שמקבלים 4 פעמים את התוצאה 1.

<u>שאלה 19</u>

מבקרת מהמרת בקזינו ב 5$ על אדום ברולטה, שוב ושוב עד לזכייתה הרביעית. בכל הימור היא מרוויחה 5$ בהסתברות $\frac{18}{38}$ או מפסידה 5$ בהסתברות $\frac{20}{38}$

א) מהי ההסתברות שהיא תהמר 9 פעמים בסה"כ?

ב) מהי תוחלת הרווח של המהמרת בהפסיקה לשחק?

<u>התפלגות היפר-גיאומטרית</u> $\underline{X \sim HG(m,N,n)}$

$$P(X=k)=\frac{\binom{m}{k}\binom{N-m}{n-i}}{\binom{N}{n}} \qquad k=0,1,..,n$$

בוחרים באקראי מדגם מגודל n , מתוך כד המכיל N כדורים – m לבנים ו N –m שחורים.

מ"מ X = מספר הכדורים הלבנים שנבחרו.

$$E(X)=\frac{nm}{N} \qquad V(X)=\frac{N-n}{N-1}np(1-p)$$

<u>שאלה 20</u>

קמעוני רוכש רכיבים חשמליים בחבילות של 10 יחידות. הרכישה של כל חבילה מתבצעת רק לאחר שהוא בדק באקראי 3 רכיבים מתוכה, ומצא שהם תקינים. אם ב 30% מהחבילות יש 4 רכיבים פגומים וב- 70% יש רכיב 1 פגום, איזה % מהחבילות שבדק אין הקמעוני רוכש?

<u>שאלה 21</u>

כד מכיל 4 כדורים לבנים ו- 4 כדורים שחורים. מוציאים באקראי 4 כדורים ללא החזרה. אם 2 מהם לבנים ו2 שחורים הניסוי מסתיים. אם לאו, מחזירים את הכדורים לכד ושוב מוציאים באקראי 4 כדורים וממשיכים כקודם עד שיש בדיוק 2 לבנים בין ה4 שהוצאו. מה ההסתברות לחזור על הניסוי בדיוק n פעמים?

<u>פתרונות</u>

<u>שאלה 1:</u>

הערכים האפשריים של Y הם 2,3,,,,12 וההסתברויות נתונות בטבלה הבאה:

k	2	3	4	5	6	7	8	9	10	11	12
$P(X=k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

לדוגמא: $P(X=4)=3/36$, משום ש- X=4 מתרחש כאשר מתקבלת אחת מן התוצאות: (1,3),(2,2),(1,3).

<u>שאלה 2:</u>

א.

$$E(X)=5*8/15+6*2/15+7*5/15=87/15=5.8$$
$$E(X^2)=5^2*8/15+6^2*2/15+7^2*5/15=34.4666$$
$$Var(X)=E(X^2)-E(X)^2=34.466-5.8^2=0.826$$

ב.

$$E(Y)=E(3X+120)=3E(X)+120=3*5.8+120=137.4$$
$$Var(Y)=Var(3X+120)=3^2Var(X)=9*0.826=7.434$$

שאלה 3:

נסמן ב-X את הסכום שהמהמר מקבל. Y – הרווח שלו. $Y = X - 20$.

התפלגות	המקרה	ממוצע
$\binom{3}{2}\Big/\binom{6}{2} = 3/15$	0,0	0
$\binom{2}{1}\binom{3}{1}\Big/\binom{6}{2} = 6/15$	0,20	10
$\binom{2}{2}\Big/\binom{6}{2} + \binom{3}{1}\binom{1}{1}\Big/\binom{6}{2} = 4/15$	20,20 או 40,0	20
$\binom{1}{1}\binom{2}{1}\Big/\binom{6}{2} = 2/15$	40,20	30

$$E(X) = 0*3/15 + 10*6/15 + 20*4/15 + 30*2/15 = 40/3$$

$$E(Y) = E(X-20) = E(X) - 20 = -20/3$$

$$Var(Y) = Var(X-20) = Var(X)$$

$$Var(x) = E(x^2) - E(X)^2 = 0*3/15 + 10^2*6/15 + 20^2*4/15 + 30^2*2/15 - (40/3)^2 = 800/9$$

שאלה 4:

$$\sum_{i=0}^{\infty} \frac{C}{3^i} = 1 = C * \sum_{i=0}^{\infty} \left(\frac{1}{3}\right)^i = C * \frac{1}{1-1/3} \Rightarrow C = 2/3$$

חישוב התוחלת:

$$E(X) = \sum_{i=0}^{\infty} i * P(X=i) = \sum_{i=0}^{\infty} i * 2/3 * \frac{1}{3^i} = 2/3 * \sum_{i=0}^{\infty} \frac{i}{3^i} \underset{*\text{הערה}}{=} 2/3 * 3/4 = 1/2$$

הערה:

$$\sum_{i=0}^{\infty} \frac{i}{3^i} = 1/3*1*(1/3) + 2*(1/3)^2 + 3*(1/3)^3 + ... = 1/3*\left(1 + 2*1/3 + 3*(1/3)^2 + ...\right)$$

$$\left(\frac{1}{1-q}\right)' = \frac{1}{(1-q)^2}, \quad \Rightarrow \sum_{i=0}^{\infty} \frac{i}{3^i} = 1/3 * \frac{1}{(1-1/3)^2}$$

ההסתברות של ערכים הגדולים מ- 5:

$$P(X>5) = 1 - P(\le 5) = 1 - 2/3 * \left(\left(\frac{1}{3}\right)^0 + \left(\frac{1}{3}\right)^1 + \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^3 + \left(\frac{1}{3}\right)^4 + \left(\frac{1}{3}\right)^5\right) = 0.00137$$

הסתברות עבור ערכים אי זוגיים:

$$P(X\ odd) = 1 - P(X\ even) = 1 - \sum_{k=0}^{\infty} P(X=2k) = 1 - 2/3 * \sum_{k=0}^{\infty} \frac{1}{3^{2k}} = 1 - 2/3 * \sum_{k=0}^{\infty} \left(\frac{1}{9}\right)^k =$$

$$1 - 2/3 * \frac{1}{1-1/9} = 1 - 3/4 = 1/4$$

שאלה 5:

שימו לב לרציפות מימין ולא משמאל:

$$P(x<3)=\lim_{n\to\infty}P(x\le 3-1/n)=\lim_{n\to\infty}F(3-1/n)=11/12$$
$$P(x=1)=P(x\le 1)-P(x<1)=F(1)-\lim_{n\to\infty}F(1-1/n)=2/3-1/2=1/6$$
$$P(x>1/2)=1-P(x\le 1/2)=1-F(1/2)=1-\frac{1/2}{2}=3/4$$
$$P(2<x\le 4)=F(4)-F(2)=1-11/12=1/12$$

שאלה 6:

X	0	2	4	6	8
Y	-1	0	1/3	1/2	3/5
Z	4	0	4	16	36
$P(X=i)$	**1/9**	2/9	**3/9**	2/9	1/9

$$E(Y)=-1*1/9+0*2/9+1/3*3/9+1/2*2/9+3/5*1/9=8/45$$
$$E(Z)=0*2/9+4*4/9+16*2/9+36*1/9=28/3$$

שאלה 7:

א. מרחב המדגם מונה $\binom{9}{4}=126$ אפשרויות. הערכים האפשריים ל –X הם: 1,2,3.

$$P(X=1)=P(four\ black)=\frac{\binom{4}{4}}{\binom{9}{4}}=1/126=0.007$$
$$P(X=3)=\frac{\binom{4}{1}\binom{3}{1}\binom{2}{2}+\binom{4}{2}\binom{3}{1}\binom{2}{1}+\binom{4}{1}\binom{3}{2}\binom{2}{1}}{\binom{9}{4}}=72/126=0.571$$
$$P(X=2)=1-P(X=1)-P(X=3)=53/126=0.42$$

ב.

$$E(X)=1*1/126+2*53/125+3*72/126=2.563$$
$$E(X^2)=1^2*1/126+2^2*53/125+3^2*72/126=6.833$$
$$Var(x)=E(X^2)-E^2(X)=0.264$$
$$E(X^3)=1^3*1/126+2^3*53/125+3^3*72/126=18.801$$
$$E(X^3+X)=E(X^3)+E(X)=21.364$$

ג.

Y משתנה היפרגאומטרי : באוכלוסיה יש N פריטים (במקרה שלנו = 9). מהם m מיוחדים (במקרה שלנו = 2), לוקחים מן אוכלוסיה מדגם של n פריטים באופן מקרי וללא החזרה (במקרה שלנו = 4). Y הוא מספר הפריטים המיוחדים במדגם.

$$Y \sim HG(9,2,4)$$

$$P(X=k)=\frac{\binom{2}{k}\binom{9-2}{4-k}}{\binom{9}{4}}$$

$$E(X)=4*2/9$$

$$Var(X)=4*2/9*(1-2/9)*\left(\frac{9-4}{9-1}\right)=0.4321$$

ד. לפי נוסחת הכפל:

$$P(Y=2\cap two\,red)=P(Y=2)*P(two\,red\,|\,Y=2)$$

לפי סעיף ג':

$$P(Y=2)=\frac{\binom{2}{2}\binom{7}{2}}{\binom{9}{4}}=21/126=1/6$$

$$P(two\,red\,|\,Y=2)=\frac{2*3!}{4!}=1/2$$

לכן: 1/6*1/2=1/12.

<u>שאלה 8</u>

$$X \sim Bin(5,0.7)$$

א) $P(X=3)=\binom{5}{3}(0.7)^3(0.3)^2=0.3087$

ב)

$$P(X\geq 3)=P(X=3)+P(X=4)+P(X=5)$$
$$=\binom{5}{3}(0.7)^3(0.3)^2+\binom{5}{4}(0.7)^4(0.3)+\binom{5}{5}(0.7)^5(0.3)^0=0.8369$$

ג)

$$P(X\leq 3)=P(X=0)+P(X=1)+P(X=2)+P(X=3)$$

או

$$1-P(X>3)=1-\left(P(X=4)+P(X=5)\right)=0.4717$$

א) נסמן ב- X את האנשים שבעד פתיחת הקניון $X \sim Bin(4,0.45)$

$$P(X=4)=\binom{4}{4}(0.45)^4=0.041$$

ב) כעת נסמן ב-Y את האנשים שנגד פתיחת הקניון. $Y \sim Bin(4,0.3)$

P(לפחות אחד נגד הפתיחה)= 1-P(אף אחד לא נגד)= $1-(0.7)^4=0.2599$

שאלה 9

ג) כעת נסמן ב- Z את האנשים עם דעה מסוימת $Z \sim Bin(4,0.75)$

$$P(X=4)=\binom{4}{4}(0.75)^4=0.3164$$

ד) כעת נסמן ב- W את האנשים שאין להם דעה בנידון. $W \sim Bin(4,0.25)$

P(לפחות לאחד אין דעה)= -P(לאף אחד אין דעה)= $1-(0.75)^4=0.6836$

שאלה 10

נסמן ב X את מספר המעשנים.

א) $X \sim Bin(6,0.45)$. $P(X=3)=\binom{6}{3}(0.45)^3(0.55)^3=0.303$

ב) $X \sim Bin(8,0.45)$ $1-P(X=4)=1-\binom{8}{4}(0.45)^4(0.55)^4=0.732$

שאלה 11

X= מספר הספרים שיש לרוני לאחר תקופה מסוימת.

א)

$$X \sim Bin(10,0.6)$$
$$P(X=6)=\binom{10}{6}(0.6)^6(0.4)^4=0.2580$$

ב)

$$X \sim Bin(12,0.6)$$
$$E(X)=np=12\cdot(0.6)=7.2$$
$$V(X)=npq=12\cdot(0.6)\cdot(0.4)=2.88$$

ג)בחודש השישי הוא קונה ספר רביעי, אך לא ידוע מתי קנה את 3 הספרים הראשונים.

$$X \sim Bin(5,0.6)$$

$$P(X=3)\cdot(0.6)=\left[\binom{5}{3}(0.6)^3(0.4)^2\right]\cdot(0.6)=0.20736$$

// הערה- זוהי בדיוק התפלגות בינומית שלילית.

<u>שאלה 12</u>

א) "הצלחה" בשאלה זו = הסתברות לכשלון במבחן שהיא 0.3 כלומר $X \sim Geo(0.3)$

$$P(X=5)=(0.7)^4(0.3)=0.07203$$

ב) יעבור פחות מ3 מבחנים= יכשל ב 1 או יכשל ב 2 או יכשל ב 3 מבחנים

$$P(X<3)=P(X=1)+P(X=2)+P(X=3)$$
$$=0.3+(0.7)(0.3)+(0.7)^2(0.3)=0.657$$

הערה- להתפלגות גיאומטרית יש תכונת <u>"חוסר זיכרון"</u> – כלומר לא משנה היסטוריית המקרה על ההתפלגות. $P(X=s+t \mid X=t)=P(X=s)$

<u>שאלה 13</u>

לא משנה כמה דקות הוא דיבר בעבר(היסטורית המקרה) ולכן

$$P(X=n+10 \mid X=n)=P(X=10)=(0.7)^9(0.3)=0.1434$$

<u>שאלה 14</u>

א) $P(X=9)=(0.75)^8(0.25)=0.02502$

ב)

$$P(X>4)=1-\left[P(X=4)+P(X=3)+P(X=2)+P(X=1)\right]$$
$$=1-(0.75)^3(0.25)-(0.75)^2(0.25)-(0.75)(0.25)-(0.25)=0.5664$$

ג) $E(X)=\frac{1}{p}=\frac{1}{0.25}=4$

שאלה 15

$X \sim Poi(5)$ יש לשים ♥ שהיחידות הן <u>בדקה</u>.

א) $P(X=0)=\frac{5^0e^{-5}}{0!}=e^{-5}=0.0067$

ב)

$$P(X\le 3)=P(X=0)+P(X=1)+P(X=2)+P(X=3)=e^{-5}+\frac{5e^{-5}}{1!}+\frac{5^2e^{-5}}{2!}+\frac{5^3e^{-5}}{3!}=0.265$$

ג)

$$X\sim Poi(\lambda)\Rightarrow Xt\sim Poi(\lambda t)$$

ולכן עבור יחידת זמן של 5 דקות $X\sim Poi(5\cdot 5)$, כלומר 5 פניות בדקה ו25 פניות ב5 דקות.

ולכן $P(Xt=4)=P(5X=4)=\frac{25^4e^{-25}}{4!}=2.26\times10^{-7}$

שאלה 16

כאשר n גדול ו P קטן אזי ניתן לקרב את ההתפלגות הבינומית להתפלגות פואסון כאשר $\lambda\to np$.

במקרה זה יש לנו 10,000 ניסויים בודדים(n), בכל אחד ההסתברות להצלחה $p=\frac{20}{10000}$ כלומר

$X\sim Bin\left(10000,\frac{20}{10000}\right)$ וכיוון שזה דיי מטורף לחשב את כל העניין פשוט נשתמש בקירוב:

$$\lambda\cong np=10000\cdot\frac{20}{10000}=20$$

נסמן בX את מספר הפורשים משירות בחודש ולכן :

$$P(X=30)=\frac{e^{-20}20^{30}}{30!}\approx 0.008$$

שאלה 17

נסמן ב -X את מספר הכוכבים הנופלים.

$X\sim Poi(1)$ ז"א כוכב אחד נופל כל 10 דקות (כלומר 10 דקות = יחידת הזמן שלנו).

$$P(1.5X=2)=\frac{e^{-1\times1.5}(1\times1.5)^2}{2!}=0.25$$

שאלה 18:

שאלה כזו יכולה להישמע קצת מסובך אך זה בסה"כ מ"מ בינומי שלילי עם הפרמטרים

$p=\frac{1}{6}$ $r=4$ ולכן:

$$E[X]=\frac{4}{1/6}=24 \qquad V(X)=\frac{4\cdot(5/6)}{(1/6)^2}=120$$

שאלה 19:

א)מספר ההימורים עד לזכייתה הרביעית של המהמרת הוא מ"מ בינומי שלילי עם הפרמטרים 4 ו $\frac{18}{38}$ ולכן ההסתברות שתהמר 9 פעמים בסה"כ היא

$$P(X=9)=\binom{8}{3}\left(\frac{18}{38}\right)^4\left(\frac{20}{38}\right)^5\approx 0.1139$$

ב) יהי W הרווח של המהמרת ויהי X מספר ההימורים שהשתתפה בהם עד לזכייתה הרביעית. מכיוון שהמהמרת זוכה 4 פעמים בסה"כ לכן מפסידה $X-4$ בסה"כ.

$$W=4\cdot 5-(X-4)\cdot 5=40-5X$$

לכן $E(W)=40-5E(X)=40-5\cdot\frac{4}{18/38}=-\frac{20}{9}$

מסקנה לחיים: חשבו פעמיים לפני שאתם נכנסים לקזינו...

שאלה 20:

נסמן בA את המאורע שהקמעוני רוכש חבילה. לכן לפי נוסחת ההסתברות השלמה:

P(A)=P{A| 4 damage components }*0.3+ P{A| 1 damage component }*0.7

$$=\frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}\cdot\frac{3}{10}+\frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}}\cdot\frac{7}{10}=\frac{54}{100}$$

כלומר הקמעוני אינו רוכש 46% מהחבילות שבדק.

שאלה 21:

שימו ♥ שמבצעים חזרות ב"ת של הניסוי המקרי: הוצאת מדגם של 4 כדורים מתוך כד המכיל 4 כדורים לבנים ו4 שחורים. הצלחה בניסוי היא הוצאת 4 כדורים שבדיוק 2 מהם לבנים. מספר הכדורים המוצאים במדגם הוא מ"מ היפרגאומטרי עם הפרמטרים n=4 N=8 m=4 ולכן ההסתברות להצלחה בכל אחת מהחזרות היא: $\frac{\binom{4}{2}\binom{4}{2}}{\binom{8}{4}}$. כעת נסמן ב-X את מספר החזרות על הניסוי עד לקבלת ההצלחה הראשונה. X הוא מ"מ גאומטרי עם הפרמטר $\frac{18}{35}$ ולכן:

$$P(X=n)=\left(1-\frac{18}{35}\right)^{n-1}\cdot\frac{18}{35}=\left(\frac{17}{35}\right)^{n-1}\cdot\frac{18}{35}$$

מ"מ רציפים, התפלגויות רציפות, קירובים ואי שיוויונים

- מ"מ X נקרא מ"מ רציף אם קיימת פונ' ממשית אי-שלילית f המכונה פונקצית הצפיפות של X, כך שלכל קבוצה B מתקיים $P\{X\in B\}=\int_B f(x)dx$.
- אם X מ"מ רציף, אזי פונ' ההתפלגות המצטברת שלו, F, מוגדרת לכל X ממשי ע"י $F(x)=P\{X\le x\}$.
- פונ' ההתפלגות המצטברת גזירה ומתקיים $F'(x)=f(x)$ כלומר הנגזרת של פונקצית ההתפלגות המצטברת שווה לפונקצית הצפיפות.
- התוחלת של X מ"מ רציף $E(X)=\int_{-\infty}^{\infty}x\cdot f(x)dx$
- השונות של X מ"מ רציף $E(X^2)-[E(x)]^2$

משתנים רציפים:

עד כה עסקנו במשתנים מקריים בדידים, כלומר שקבוצת הערכים האפשריים שלהם סופית או אינסופית בת מניה. במקרים בהם הקבוצה אינה בת מניה נשתמש במשתנים מקריים רציפים. X מ"מ רציף אם קיימת פונקציה אי שלילית f המוגדרת לכל X ממשי ומתקיים:

1) $f(x)\ge 0\quad \forall x\in\Re$

2) $\int_{-\infty}^{\infty}f(x)dx=1$

3) $f(x)dx\approx P(x<X<x+dx)$

4) $\int_a^b f(x)dx=P(a<X<b)=P(X<b)-P(X<a)=F(b)-F(a)$

5) $P(Y>y)=\int_{-\infty}^{y}f_Y(x)dx$

6) $E(X)=\int_{-\infty}^{\infty}x\cdot f(x)dx$

7) $E[g(X)]=\int_{-\infty}^{\infty}g(x)\cdot f(x)dx$

שאלה 1:

נתונה פונקצית הצפיפות של מ"מ רציף X:

$$f(x)=\begin{cases}1 & 0\le x\le 1\\ 0 & other\end{cases}$$

מצא את התוחלת של e^X .

<u>התפלגות אחידה רציפה</u> $X\sim U(\alpha,\beta)$

בוחרים באופן מקרי מספר X בין α ל β משיקולי סימטרייה בין כל תוצאות הניסוי, פונקציית הצפיפות:

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \alpha\le x\le\beta\\ 0 & else\end{cases}$$

$$E(X)=\frac{\beta+\alpha}{2}\qquad V(X)=\frac{(\beta-\alpha)^2}{12}$$

<u>התפלגות מעריכית</u> $X\sim Exp(\lambda)$

נתון זרם ארועים פואסוני בעל קצב λ . יהי X משך הזמן הדרוש עד לאירוע הראשון.

פונ' הצפיפות: $f(x)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0\\ 0 & x<0\end{cases}$

בהתפלגות זו (בנוסף להתפלגות גאומטרית) קיימת תכונת "חוסר הזכרון":

שאלה 2:

יהי X משתנה רציף בעל פונקצית הצפיפות:

$$f(X)=\begin{cases}c\cdot(4x-2x^2) & 0\le x\le 2\\ 0 & otherwise\end{cases}$$

א. מצאו את c .

ב. מצאו את פונקצית ההתפלגות של X .

ג. מצאו את $\mathrm{P}(X>1.5\mid X>1)$.

שאלה 3:

המ"מ X הוא בעל פונקצית הצפיפות

$$f(x)=\begin{cases} c\cdot\cos(x) & |x|\le\frac{\pi}{2} \\ 0 & otherwise \end{cases}$$

א. מצאו את c .

ב. מצאו את פונקצית ההתפלגות של X .

שאלה 4:

אוטובוסים מגיעים לתחנה מסוימת ברווחי זמן של 15 דקות, החל מהשעה 7:00. אם נוסע מגיע לתחנה בזמן שמתפלג אחידה מ-7:00 ל-7:30, מה ההסתברות –

א. שיחכה פחות מ-5 דקות לאוטובוס?

ב. שיחכה יותר מ-10 דקות?

שאלה 5:

אורך שיחה בטלפון מתפלג מעריכית עם λ=0.1 . מה ההסתברות –

א. ששיחה אורכת יותר מ-10 דקות?

ב. ששיחה אורכת בין 10 ל-20 דקות?

שאלה 6:

הזמן שדרוש לתיקון מכונה הוא מ"מ המתפלג מעריכית עם $\lambda=2$ מה ההסתברות ש
א. זמן התיקון יעלה על חצי שעה?
ב. שתיקון יקח לפחות 12.5 שעות בהנחה שהוא כבר נמשך יותר מ12 שעות?

שאלה 7:

זמן החיים (בשנים) T של מנוע מכונית הוא מ"מ רציף בעל פונקציית צפיפות:

$$f(X)=\begin{cases} cx^2\ (10-x) & 0<x<10 \\ 0 & else \end{cases}$$ כאשר c קבוע.

א) מצא את c.
ב) מצא את $E(T),V(T)$
ג) מה ההסתברות שמנוע בן 7 שנים יחזיק מעמד עוד שנתיים?

התפלגות נורמלית (ותקנון)

$$X \sim N(\mu,\sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$$

<u>שאלה 8:</u>

הגובה הממוצע של 500 תלמידים בבי"ס הוא 151 ס"מ עם סטיית תקן 15 ס"מ. אם נניח שהתפלגות הגבהים היא נורמלית, מצאו כמה תלמידים הם בגובה:

א. בין 120 ל-155 ס"מ.
ב. יותר מ-180 ס"מ.

<u>שאלה 9:</u>

נתון שקבוצת תצפיות מתפלגת נורמלית עם ממוצע m וסטיית תקן s. מצא קבוע a כך ש:

א. אחוז התצפיות בתוך הטווח m±as הוא 75%.
ב. אחוז התצפיות הקטנות מ- m-as הוא 25%.

<u>שאלה 10:</u>

במבחן בסטטיסטיקה הציונים התפלגו נורמלית עם ממוצע 80 וסטיית תקן 10.

א) מה הסיכוי שממוצע הציונים של 5 סטודנטים לא יעלה על 85?
ב) מה ההסתברות שסכום ציונם של 10 סטודנטים גבוה מ 790?

<u>שאלה 11:</u>

מספר הבדיקות המגיעות למעבדה מסוימת ביום מתפלג נורמאלית עם תוחלת 120 וסטיית תקן 20.

א) מצאו את המאיון העליון של מספר הבדיקות היומי?

ב) מה ההסתברות שבסה"כ ב-25 ימי עבודה בחודש מגיעות לפחות 3150 בדיקות?

קירובים להתפלגות נורמאלית

קירוב הנורמאלי להתפלגות בינומית

אם $X \sim Bin(n,p) \quad and \quad n > 30$

אז $X \sim N(np, \left(\sqrt{npq}\right)^2)$

שאלה 12:

ל- 40% מהסטודנטים באוניברסיטה יש מכונית.

א. מה ההסתברות שמתוך 6 סטודנטים שנבחרו באקראי:

1) ל- 4 יהיו מכוניות?

2) לפחות לאחד תהיה מכונית?

ב. מהי ההסתברות שבקבוצה של 100 סטודנטים לפחות ל50 תהיה מכונית?

שאלה 13:

שולחים הזמנות לאירוע ל300 איש. ההסתברות שאורח יבוא היא 0.8 . כמה מנות צריך להכין כדי שבהסתברות של לפחות 0.95 לכל אורח תהיה מנה?

קירוב נורמאלי לפואסון

אם $X \sim Poi(\lambda) \quad and \quad \lambda > 5$

אז $X \sim N(\lambda, \left(\sqrt{\lambda}\right)^2)$

שאלה 14:

למ"מ X יש התפלגות פואסון עם פרמטר $\lambda = 200$. מצאו בקירוב את:

$P(180 \le X \le 210)$

$P(X < 150)$

<u>משפטי גבול ואי שיוויונים</u>

<u>א"ש מרקוב</u>

אם X הוא מ"מ המקבל ערכים אי שלילים בלבד, אז לכל ערך חיובי a מתקיים- $P\{X \geq a\} \leq \frac{E(X)}{a}$

<u>א"ש צבישב</u>

אם X מ"מ שתוחלתו μ ושונותו σ^2 הן סופיות, אז לכל ערך חיובי k מתקיים

$$P\{| X-\mu | \geq k\} \leq \frac{\sigma^2}{k^2}$$

<u>משפט הגבול המרכזי</u>

יהיו $X_{1,}, X_{2....}$ מ"מ ב"ת ש"ה בעלי תוחלת(כל אחד מבינהם) סופית μ ושונות סופית σ^2 אזי ההתפלגות של $\frac{X_1+...+X_n}{\sigma\sqrt{n}}$ שואפת להתפלגות נורמלית סטנדרטית כאשר $n \to \infty$

<u>שאלה 15:</u>

אסטרונום מעוניין לחשב בשנות אור את המרחק מתחנת החלל שלו לכוכב מרוחק. למרות שיש לאסטרונום טכניקות למדידה הוא יודע שבגלל שינויים בתנאי האטמוספרות וטעויות נורמליות שבכל פעם שמודדים את המרחק לא מקבלים את המרחק המדויק אלא הערכה בלבד. כתוצאה מכך האסטרונום מתכוון לעשות סדרת מדידות ואז להשתמש בערך הממוצע של המדידות הללו כהערכה למרחק האמיתי. אם האסטרונום מאמין שהערכים של המדידות ב"ת ושווי התפלגות עם תוחלת d ושונות 4 שנות אור, כמה מדידות צריך לעשות כדי שיהיה מספיק בטוח שהערכי המרק שלו יהיו מדויקים עד כדי ± 0.5 שנות אור?

שאלה 16:

מצאו את ההסתברויות הבאות:

א) $P(X_1+...+X_{100} > 400)$ כאשר $X_1+...+X_{100}$ מ"מ ב"ת בעלי התפלגות פואסון עם $\lambda = 4$.

ב) $P(75 < X_1+...+X_{40} < 40)$ כאשר $X_1+...+X_{40}$ מ"מ ב"ת בעלי התפלגות מעריכית עם $\lambda = 0.5$.

<u>שאלה 17:</u>

מספר הפריטים הנשלחים כל יום לתיבת דואר הוא מ"מ עם תוחלת 50 ושונות 5. מה ניתן לומר על ההסתברות שמספר הפריטים הוא בין 40 ל-60?

שאלה 18:

מטילים מטבע מאוזנת n פעמים. נסמן ב-X את מספר ההצלחות וב-Y=n-X את מספר הכשלונות.

הוכיחו בעזרת אי שוויון צ'בישב כי לכל a>0 מתקיים: $P(\frac{X}{Y} > 1 + \frac{a}{\sqrt{n}}) < \frac{5}{a^2}$ עבור n מאוד גדול.

שאלה 19:

תהא $\left\{X_n\right\}_{n=1}^{\infty}$ סדרה של מ"א ב"ת כך ש- $EX^2 < \infty$. נגדיר $S_n = X_1 + X_2 + ...X_n$. הוכיחו כי

עבור כל $\varepsilon > 0$ מתקיים: $\lim_{n\to\infty} P\left(\left|\frac{S_n}{E(S_n} - 1\right| \geq \varepsilon\right) = 0$

פתרונות

שאלה 1:

$$Y=\exp(X)=e^X,$$
$$Y=\{y\in\Re \mid 1\le y\le e\}$$
$$F_Y(y)=P(Y\le y)=P(e^X\le y)=P(X\le \ln y)=\int_0^{\ln y} f(x)dx=$$
$$\int_0^{\ln y} 1*dx=\ln y-0=\ln y$$
$$F_Y'(y)=f_Y(y)=(\ln y)'=1/y$$
$$E(\exp(X))=E(Y)=\int_{-\infty}^{\infty} y*f_Y(y)dy=\int_1^e dy=e-1$$

שאלה 2:

א) לכל מ"מ רציף מתקיים: $\int_{-\infty}^{\infty} f_X(x)dx=1$

$$\int_{-\infty}^{\infty} c(4x-2x^2)dx=\int_0^2 \left(4cx-2cx^2\right)dx=4c\int_0^2 xdx-2c\int_0^2 x^2dx=4c\frac{x^2}{2}\Big|_0^2-2c\frac{x^3}{3}\Big|_0^2=8c-\frac{16}{3}c=\frac{8}{3}c=1$$
$$\Rightarrow c=\frac{3}{8}$$

ב)

$$F(x)=\int_{-\infty}^{x} f(u)du=\int_{-\infty}^{x}\frac{3}{8}\left(4u-2u^2\right)du=\frac{2}{3}\int_0^x udu-\frac{3}{4}\int_0^x u^2=\frac{2}{3}\frac{u^2}{2}\Big|_0^x-\frac{3}{4}\frac{u^3}{3}\Big|_0^x=\frac{3}{4}x^2-\frac{1}{4}x^3$$

ולכן:

$$F(X)=\begin{cases}1 & ,x>2\\ \frac{3}{4}x^2-\frac{1}{4}x^3 & ,0\le x\le 2\\ 0 & ,x<0\end{cases}$$

שאלה 3:

א)

$$\int_{-\infty}^{\infty} c\cdot\cos x dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} c\cdot\cos x dx = c\cdot\sin x\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = c\left[\sin\frac{\pi}{2}-\sin\left(-\frac{\pi}{2}\right)\right] = c[1-(-1)] = 2c$$

$$\Rightarrow c = \frac{1}{2}$$

ב) $F(x)=\int_{-\infty}^{x} f_X u du = \int_{-\frac{\pi}{2}}^{x}\frac{1}{2}\cos u du = \frac{1}{2}\sin u\Big|_{-\frac{\pi}{2}}^{x} = \frac{1}{2}\left[\sin x-\sin\left(-\frac{\pi}{2}\right)\right] = \frac{1}{2}[\sin x+1]$

ולכן

$$F(x)=\begin{cases}0 & x<-\frac{\pi}{2}\\ \frac{1}{2}[\sin x+1] & -\frac{\pi}{2}\le x\le\frac{\pi}{2}\\ 1 & x>\frac{\pi}{2}\end{cases}$$

שאלה 4:

נסמן ב-X את מספר הדקות אחרי השעה 7 שהנוסע הגיע ולכן $X \sim U(0,30)$

א) הנוסע יחכה פחות מ5 דקות אם יגיע בין 7:10 ל 7:15 או בין 7:25 ל 7:30 ולכן

$$P(10<x<15)+P(25<x<30)=\int_{10}^{15}\frac{1}{30}dx+\int_{25}^{30}\frac{1}{30}dx=\frac{1}{3}$$

ב) הנוסע יחכה יותר מ 10 דקות אם יגיע בין 7:00 ל7:05 או 7:15 ל7:20 ולכן

$$P(0<x<5)+P(15<x<20)=\int_{0}^{5}\frac{1}{30}dx+\int_{15}^{20}\frac{1}{30}dx=\frac{1}{3}$$

שאלה 5:

א)

$$P(X>10)=1-P(X\le 10)=1-\int_{0}^{10}0.1e^{-0.1x}dx=1-\left[(0.1)\cdot\frac{e^{-0.1x}}{-0.1}\right]_{0}^{10}=1+\left[e^{-0.1x}\right]_{0}^{10}=1+e^{-1}-1=\frac{1}{e}$$

ב) $P(10<x<20)=\int_{10}^{20}0.1e^{-0.1x} = \frac{1}{e}-\frac{1}{e^2}$

שאלה 6:

$X \sim Exp(2)$

א. $P(X > 0.5) = 1 - P(X \le 0.5) = 1 - \int_0^{0.5} 2e^{-2x}dx = 1 - 2\frac{e^{-2x}}{-2}\Big|_0^{0.5} = \frac{1}{e}$

ב. $P(x > 12.5 \mid x > 12) = P(X > 0.5) = \frac{1}{e}$

שאלה 7:

א. $\int_0^{10} cx^2(10-x)dx = c\left(\frac{10x^3}{3} - \frac{x^4}{4}\Big|_0^{10}\right) = c \cdot \frac{10^4}{12} = 1 \quad \Rightarrow c = \frac{3}{2500}$

ב.

$$E(T) = \int_{-\infty}^{\infty} x \cdot f(x) = \int_0^{10} x\frac{3}{2500}x^2(10-x)dx = \ldots = 6$$

$$E(T^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) = \int_0^{10} x^2\frac{3}{2500}x^2(10-x)dx = \ldots = 40$$

$$\Downarrow$$

$$V(T) = E(T^2) - E^2(T) = 40 - 6^2 = 4$$

ג.

$$P(X \ge 9 \mid X \ge 7) = \frac{P(X \ge 9) \cap P(X \ge 7)}{P(X \ge 7)} = \frac{P(X \ge 9)}{P(X \ge 7)} = \frac{\int_0^9 \frac{3}{2500}x^2(10-x)dx}{\int_0^7 \frac{3}{2500}x^2(10-x)dx} = .. = 0.15$$

שאלה 8:

$X \sim N(151, 15^2) \Rightarrow Z = \frac{X-151}{15} \sim N(0,1)$

א.

$$P(120 \le X \le 155) = P\left(\frac{120-151}{15} \le \frac{X-151}{15} \le \frac{155-151}{15}\right) =$$

$$P(-2.067 \le Z \le 0.267) = P(Z \le 0.267) - P(Z \le -0.267) = \Phi(0.267) - (1-\Phi(0.267)) = 0.585$$

ב.

$$P(X > 180) = 1 - P(X \le 180) = 1 - P\left(\frac{X-151}{15} \le \frac{180-151}{15}\right) = 1 - P(Z \le 1.933) = 1 - \Phi(1.933) = 0.0268$$

שאלה 9:

א.

$$P(m-as \le X \le m+as) = 0.75$$
$$\Rightarrow P\left(\frac{m-as-m}{s} \le \frac{X-m}{s} \le \frac{m+as-m}{s}\right) = 0.75$$
$$\Rightarrow P(-a \le Z \le a) = 0.75$$
$$\Rightarrow \Phi(a) - (1-\Phi(a)) = 0.75$$
$$\Rightarrow 2\Phi(a) - 1 = 0.75$$
$$\Phi(a) = 0.875$$
$$a = \Phi^{-1}(0.875) = 1.15$$

ב.

$$P(X \le m-as) = 0.25$$
$$P\left(\frac{x-m}{s} \le \frac{m-as-m}{s}\right) = 0.25$$
$$\Rightarrow P(Z \le -a) = \Phi(-a) = 1-\Phi(a) = 0.25$$
$$\Phi(a) = 0.75$$
$$a = \Phi^{-1}(0.75) = 0.675$$

קירובים להתפלגות נורמלית בעזרת משפט הגבול המרכזי:

$$X_i \sim N(\mu,\sigma^2) \qquad \forall i = 1..n$$
$$\Downarrow$$
$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$
$$\Downarrow$$
$$\frac{\sum_{i=1}^{n} X_i}{n} = \overline{X} \sim N(\mu, \frac{\sigma}{n})$$

<u>שאלה 10:</u>

נתון: $X_i \sim N(80,10^2)$

א.

$$\frac{\sum_{i=1}^{5} X_i}{5} = \overline{X}_5 \sim N\left(80,\frac{10^2}{5}\right)$$

$$P\left(\overline{X}_5 \le 85\right) = P\left(Z \le \frac{85-80}{\frac{10}{\sqrt{5}}}\right) = \Phi(1.12) = 0.8686$$

ב.

$$\sum_{i=1}^{10} X_i \sim N\left(10\cdot 80, 10\cdot 10^2\right)$$

$$P\left(\sum_{i=1}^{10} X_i > 790\right) = P\left(Z > \frac{790-800}{\sqrt{10}\cdot 10}\right) = P(Z > -0.316) = P(Z < 0.316) = \Phi(0.316) = 0.62$$

<u>שאלה 11:</u>

$$X \sim N(120,20^2)$$

א.

$$P(X \le a) = 0.99$$

$$\Downarrow$$

$$P\left(Z \le \frac{a-120}{20}\right) = 0.99$$

$$\Downarrow \ \textit{from the Z table}:$$

$$\frac{a-120}{20} = 2.326 \Rightarrow a = 166.52$$

ב.

$$\sum_{i=1}^{25} X_i \sim N\left(25\cdot 120, 25\cdot 20^2\right)$$

$$P(\sum_{i=1}^{25} X_i \ge 3150) = P\left(Z \ge \frac{3150-25\cdot 120}{\sqrt{25}\cdot 20}\right) = P(Z \ge 1.5) = 1 - P(Z \le 1.5) = 1 - \Phi(1.5) = 0.0668$$

שאלה 12:

א.

$$X \sim N(6,0.4)$$

$$P(X=4)=\binom{6}{4}(0.4)^4(0.6)^2=0.1382$$

$$P(X\ge 1)=1-P(X<1)=1-P(x=0)=1-\binom{6}{0}(0.4)^0(0.6)^6=0.95$$

ב. נעבור לקירוב נורמאלי:

$$X \sim Bin(100,0.4)$$

$$\mu=np=100\cdot 0.4=40$$

$$\sigma=\sqrt{npq}=\sqrt{100\cdot 0.4\cdot 0.6}=4.9$$

$$\Downarrow$$

$$X \sim N\left(40,(4.9)^2\right)$$

$$P\left(X\ge 50\right)=1-P(X<50)=1-P\left(Z<\frac{50-0.5-40}{4.9}\right)=1-P(Z<1.94)=1-\Phi(1.94)=0.02$$

שאלה 13:

נסמן ב- a את מספר המנות וב- X את מספר האורחים.

אנחנו רוצים שמספר האורחים יהיה קטן או שווה למספר המנות בהסתברות של 0.95 לפחות. כעת,

$$X \sim Bin(300,0.8)$$

$$\Downarrow$$

$$X \sim N\left(240,\left(\sqrt{48}\right)^2\right)$$

צ"ל: $P(X\le a)\ge 0.95$

$$P\left(Z\le\frac{a+0.5-240}{\sqrt{48}}\right)\ge 0.95$$

$$\Phi\left(\frac{a+0.5-240}{\sqrt{48}}\right)\ge 0.95$$

$$\Downarrow \textit{look at the table}$$

$$\frac{a+0.5-240}{\sqrt{48}}=1.645$$

$$\Downarrow$$

$$a=251$$

כלומר יש להכין יותר מ251 כדי שבהסתרות של לפחות 95% יהיה מנות לכל האורחים.

שאלה 14:

$$X \sim Poi(200)$$
$$\Downarrow$$
$$X \sim N\left(200, \left(\sqrt{200}\right)^2\right)$$
$$P(180 \le X \le 210) = P\left(\frac{179.5-200}{\sqrt{200}} \le Z \le \frac{210.5-200}{\sqrt{200}}\right) = \Phi(0.74) - \Phi(-1.45)$$
$$= \Phi(0.74) - (1-\Phi(-1.45)) = 0.7704 - (1-0.9265) = 0.697$$
$$P(X < 150) = P\left(Z < \frac{149.5-200}{\sqrt{200}}\right) = P(Z < -3.57) = 1-\Phi(3.57) = 1-0.9999 \approx 0$$

שאלה 15:

נניח שהאסטרונום מבצע n מדידות, נסמן: $X_{1,}X_{2,}.....X_n$ אזי לפי משפט הגבול המרכזי:

$$Z_n = \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sum X_i - nd}{2\sqrt{n}} \sim N(0,1)$$
$$\frac{\sum X_i}{n} - d \to \frac{\sum X_i - nd}{n} \to \frac{\sum X_i - nd}{2\sqrt{n}}\frac{2}{\sqrt{n}}$$
$$P\left\{-0.5 \le \frac{\sum X_i}{n} - d \le 0.5\right\} = P\left\{-0.5\cdot\frac{2}{\sqrt{n}} \le \frac{\sum X_i - nd}{2\sqrt{n}} \le 0.5\cdot\frac{2}{\sqrt{n}}\right\} = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

שאלה 16:

א)

$$\mu = E(x_i) = \lambda = 4$$
$$\sigma^2 = V(x_i) = \lambda = 4$$
$$\sum X_i \sim N(n\mu, n\sigma^2) \Rightarrow \sum X_i \sim N(100\cdot 4, 100\cdot 4) \Rightarrow \sum X_i \sim N(400, 400)$$
$$P\left(\sum_{i=1}^{100} X_i > 440\right) = 1 - P\left(\sum_{i=1}^{100} X_i \le 440\right) = 1 - P\left(Z \le \frac{440+0.5-400}{20}\right) = 20.0215$$

ב)

$$\mu = E(x_i) = \frac{1}{\lambda} = 2$$

$$\sigma^2 = V(x_i) = \frac{1}{\lambda^2} = 4$$

$$\sum X_i \sim N\left(n\mu, n\sigma^2\right) \Rightarrow \sum X_i \sim N(40\cdot 2, 40\cdot 4) \Rightarrow \sum X_i \sim N(80,160)$$

$$P\left(75 < \sum_{i=1}^{100} X_i < 40\right) = P\left(\frac{75-80}{\sqrt{160}} \le Z \le \frac{85-80}{\sqrt{160}}\right) = 0.3034$$

שאלה 17:

נסמן בX את מספר הפריטים בדואר. לכן

$$P(40 \le X \le 60) = P\left(-10 \le X - 50 \le 10\right) = P\left(|X-50| \le 10\right)$$

נשתמש באי שיווין צבישב $P\left(|X-\mu| > k\right) \le \frac{\sigma^2}{k^2}$

$$P(|X-50| \le 10) = 1 - P(|X-50| > 10) = 1 - \frac{5}{100} = 95\%$$

שאלה 18:

הרעיון הוא להפעיל את אי שוויון צ'יבישב על המשתנה הבינומי X:

$$E(X) = \sum_{j=1}^{n} E(I_j) = np = n/2$$

$$V(X) = \sum_{j=1}^{n} V(I_j) = npq = n/4$$

כאשר מ"מ $\left\{I_j\right\}_{j=1}^{\infty}$ ב"ת מפולגים לפי: $\begin{matrix} P(I_j = 1) = p = 1/2 \\ P(I_j = 0) = q = 1/2 \end{matrix}$, מכאן:

$$P\left(\frac{X}{Y} > 1 + \frac{a}{\sqrt{n}}\right) = P\left(X > n - X + a\sqrt{n} - \frac{aX}{\sqrt{n}}\right)$$

$$= P\left(X\left(2 + \frac{a}{\sqrt{n}}\right) > n + a\sqrt{n}\right) = P\left(X - n/2 > \frac{n + a\sqrt{n}}{2 + a\sqrt{n}} - n/2\right) =$$

$$= P\left(X - n/2 > \frac{a\sqrt{n}}{4 + 2a/\sqrt{n}}\right) \le \frac{n/4}{\left(\frac{a\sqrt{n}}{4 + 2a/\sqrt{n}}\right)^2} \le \frac{(2 + a/\sqrt{n})^2}{a^2} \le (\text{for large n}) = 5/a^2$$

מאחר ואי שוויון צ'בישב תקף לכל התפלגות של משתנה מקרי X, איננו יכולים לצפות שתמיד יתקבל ממנו חסם להסתברות, שערכו קרוב מאוד להסתברות בפועל. תרגילים 3,4 ימחיש זאת:

יהי X משתנה מקרי אחיד בקטע (0,10), אז E(X)=5, ו-V(X)=25/3. ולכן מקבלים מאי

שוויון צ'בישב: $P(|X-5|>4)\le\frac{25/3}{16}\approx 0.521$ בעוד שהתוצאה המדוייקת היא:

$$P(|X-5|>4)=P(X<1)+P(X>9)=0.20$$

<u>שאלה 19:</u>

לפי אי שוויון של צ'בישוב:

$$P\left(\left|\frac{S_n}{E(S_n)}-1\right|\ge\varepsilon\right)=P\left(\left|S_n-E(S_n)\right|\ge\varepsilon\cdot E(S_n)\right)\le\frac{VAR(S_n)}{\left(\varepsilon\cdot E(S_n)\right)^2}$$

$$E(S_n)=E(X_1)+...+E(X_n)=nE(X_1)$$

$$VAR(S_n)=VAR\left(\sum_{k=1}^{n}X_k\right)=\sum_{k=1}^{n}VAR(Y_k)=nVAR(X_1)$$

מאי השוויון של צ'בישב נובע כי: $P\left(\left|\frac{S_n}{E(S_n)}-1\right|>\varepsilon\right)\le\frac{VAR(X_1)}{n\left(\varepsilon\cdot E(X_1)\right)^2}\to 0$

מ"מ דו מימדי, התפלגות משותפת ומקדם המתאם

<u>נוסחאות:</u>

$E(\sum x)=\sum E(x)$

$x,y-independable$

$V(x+y)=V(x)+V(y)$

$\mathrm{cov}(x,y)=E(xy)-E(x)E(y)$

$x,y-dependable$

$\mathrm{var}(x+y)=v(x)+v(y)+2\,\mathrm{cov}(x,y)$

<u>שאלה 1:</u>

מרכיבים באקראי מס' דו סיפרתי מהספרות 1,2,3,4.

יהי X מס' הספרות השונות המופיעות במס' ו Y מס' הפעמים שהספרה 1 מופיעה.

מצא:

א. ההתפלגות המשותפת של הזוג (X,Y).

ב. האם X ו- Y בת"ל?

ג. מצא את השונות המשותפת – COV(X,Y).

<u>שאלה 2:</u>

נתונים שני כובעים,בכל אחד ארבעה פתקים ממוספרים מ 1-4.

מוצאים באקראי פתק מכל כובע.

נסמן ב- X תוצאה מקסימלית וב-Y –תוצאה מינימלית.

מצא:

א. ההתפלגות המשותפת של X ו Y
ב. COV(X,Y)

<u>שאלה 3:</u>

בכל הטלה של קובייה לא הוגנת, כל תוצאה אי זוגית מתקבלת בהסתברות C(1,3,5), וכל תוצאה זוגית בהסתברות 2C .

א. חשב את C
ב. נניח שמטילים את הקובייה פעם אחת, ונגדיר את המשתנים המקריים הבאים:

$$X=\begin{cases}1\text{ - even}\\0\text{ - odd}\end{cases}$$ אחרת

$$Y=\begin{cases}1-(x>3)\\0-(x\le 3)\end{cases}$$ אחרת

מצא את פונק' ההסתברות המשותפת של X וY.

שאלה 4:

יהיו X ו- Y מ"מ בת"ל בעלי אותה התפלגות: $P(X=r)=P(Y=r)$

r	1	2	3	4
P(X=r)	1/8	3/8	3/8	1/8

א. מצא את ההסתברות ש Y=X ואת ההסתברות ש Y<X

ב. מצא את ההתפלגות של Y+X, ‏E(X+Y) ואת V(X+Y).

שאלה 5:

מטילים מטבע לא הוגן אינסוף פעמים. הסתברות להצלחה בהטלה בודדת שווה ל-p. יהי X_k מספר ההטלה בה התקבלה הצלחה – k ית. למשל, עבור סדרת התוצאות משמאל לימין

$$X_1=1, X_2=4, X_3=5, X_4=10 \Rightarrow 1,0,0,1,1,0,0,0,0,1$$

מצאו את מקדם המתאם של $X_m X_n$.

שאלה 6:

לנשף של קיסר רומאי הוזמנו n זוגות. לפני ריקוד הפתיחה הקיסר הרכיב באופן אקראי לגמרי n זוגות הרוקדים. תניחו כי הקיסר דאג לכך שבנים רוקדים רק עם בנות.

א) יהי I_k המשתנה המקרי המוגדר על ידי: אם גבר k רוקד עם בת זוגתו $I_k=1$. אחרת $I_k=0$.
חישבו את $E(I_k)$ ו- $E(I_j I_k)$ עבור $k \neq j$.

ב) בהשתמש בהנ"ל, או אחרת, הסיקו כי $\{I_k\}$ משתנים תלויים.

ג) יהי S_n מספר הגברים שבכל זאת ירקדו עם בת זוגם. חישבו את $E(S_n)$ בעזרת הנוסחה

$$S_n=\sum_{k=1}^{n} I_k.$$

ד) חישבו את $VAR(S_n)$ בעזרת הנוסחה $VAR(S_n)=\sum_{k=1}^{n} VAR(I_k)+2\sum_{i<j} COV(I_i,I_j)$.

שאלה 7:

יהי (X,Y) וקטור אקראי בעל צפיפות:

$$f_{X,Y}(x,y)=\begin{cases} 1/4(1+xy) & |x|\le 1, |y|\le 1 \\ 0 & else \end{cases}$$

א) הוכיחו כי המשתנים המקריים X ו-Y אינם בלתי תלויים.

ב) הוכיחו כי המשתנים המקריים X^2 ו- Y^2 בלתי תלויים.

ג) חשבו את $COV(X^2, X^2+Y^2)$.

שאלה 8:

גבר ואישה מחליטים להיפגש במקום מסוים, בין השעות 12:00 ל- 13:00. כל אחד מהם מגיע למקום שנקבע באופן בלתי בתלוי באחר, ברגע שהתפלגותו אחידה בפרק הזמן דלעיל. מצא את ההסתברות שהמגיע ראשון חייב להמתין למעלה מ-10 דקות.

פונקציה יוצרת מומנטים

פונקציה יוצרת מומנטים:

$$M(t) = E(e^{tX})$$

אם X ו-Y בלתי תלויים אזי:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

שאלה 9:

יהי X משתנה מקרי בינומי עם פרמטרים n ו-p. מצא את פונקצית יוצרת המומנטים שלו ע"י 2 דרכים שונות.

שאלה 10:

יהיו $X \sim Bin(n,p),\ \ Y \sim Bin(m,p)$ מהי ההתפלגות של הסכום X+Y:

שאלה 11:

שתי קוביות משחק נזרקות. יהי X התוצאה הגבוהה ביותר ו-Y מספר הקוביות עם תוצאה זוגית.

א. מצא את טבלת ההסתברות של (X,Y).
ב. חשב את ההסתברויות $P(2Y < X),\ \ P(X+Y \le 6),\ \ P(X \ge 4, Y = 2),\ \ P(X \ge 4, Y = 1)$.
ג. חשב את ההתפלגויות השוליות של X ו-Y.
ד. מצא את $Var(Y),\ E[Y],\ Var(X),\ E[X]$.
ה. האם X ו-Y בלתי תלויים? נמק.
ו. מצא את התפלגות של $Z = X + Y$.
ז. מהי התוחלת והשונות של Z.
ח. מהי פונקצית יוצרת מומנטים של X
ט. מהי פונקציה יוצרת מומנטים של Y
י. מהי פונקציה יוצרת מומנטים של Z
יא. מצא באמצעות סעיפים ט,ח,ג את התוחלת והשונות והשווה לתוצאה שקבלת בסעיף ד.
יב. מצא באמצעות סעיף ט,ו את התוחלת והשונות של Z.

מקדם המתאם

שאלה 12:

לשליש מהאוכלוסייה יש שתי מכוניות, לשליש יש רק מכונית אחת וליתר אין מכונית. בוחרים אדם באופן מקרי ושואלים אותו למספר המכוניות שיש לו. הוא עונה תשובה נכונה בהסתברות 0.5 ומשקר בהסתברות 0.5, כאשר לכל אחת משתי התשובות השקריות העומדות לרשותו יש אותה הסתברות.

יהי X – מספר המכוניות שיש לאדם.

Y – מספר המכוניות שענה שיש לו.

א. בנה את פונקצית ההסתברות המשותפת של X ו- Y.

ב. חשב את מקדם המתאם בין X ל- Y.

שאלה 13:

בכד חמישה כדורים הממוספרים מ-1 עד 5. כדורים 1 ו- 2 אדומים, כדורים 3 ו- 4 ירוקים וכדור 5 הוא לבן. מוציאים מהכד, באופן מקרי וללא החזרה, שני כדורים. יהיו:

X – מספר הכדורים הירוקים שיצאו.

Y – מספר הכדורים, שעליהם מספרים זוגיים, שיצאו.

א. מצא את פונקצית ההסתברות המשותפת של X ו- Y.
ב. האם X ו- Y בלתי מתואמים?
ג. חשב את מקדם המתאם בין X ו- Y.
ד. ידוע כי יצא לפחות כדור אחד אדום. מה ההסתברות שיצא בדיוק כדור אחד שעליו מספר זוגי?

שאלה 14:

ההסתברות שקוביית שוקולד אגוזים אכן תכיל אגוז היא 0.5 (בהסתברות 0.5 הקובייה אינה מכילה אגוז). בצלחת מס' 1 שתי קוביות שוקולד ובצלחת מס' 2 שלוש קוביות. אני בוחר צלחת באקראי. יהי N מספר הצלחת ויהי X מספר האגוזים שאקבל. חשב את מקדם המתאם בין N ל- X.

פתרונות

שאלה 1:

סה"כ ישנן 16 אפשרויות ליצור מס' דו סיפרתי מארבע הספרות

ישנן $2!\binom{4}{2}$ אפשרויות למס' בעל שתי ספרות שונות ועוד ארבע למס' ששתי ספרותיו זהות

א. ההתפלגות הדו מימדית:

X / Y	1	2
0	$3/16$	$6/16$
1	0	$6/16$
2	$1/16$	0

ב. X ו-Y תלויים כיוון ש-0 מופיע בטבלה או ניתן לבדוק כי

$$p(Y=1 \mid X=1)=0 \neq p(Y=1)=\frac{6}{16}$$

ג. עפ"י ההגדרה:

$$Cov(X,Y)=E(X\cdot Y)-E(X)E(Y)$$

$$E(X\cdot Y)=\sum_{\substack{1\le i\le 2\\ 0\le j\le 2}} i\cdot j\cdot p(X=i,Y=j)=\frac{14}{16}$$

$$E(X)=\frac{7}{4},E(Y)=0.5$$

$$Cov(X,Y)=0$$

שאלה 2:

סה"כ ישנם 16 צירופים אפשריים כאשר מוצאים פתק מכל כובע.

X / Y	1	2	3	4
1	$\frac{1}{16}$	0	0	0
2	$\frac{2}{16}$	$\frac{1}{16}$	0	0
3	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	0
4	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

א. X ו Y תלויים, קיים אפס בטבלה או ניתן לבדוק כי $p(X=3|Y=2)\neq p(X=3)$

ב. שוב נחשב שונות משותפת עפ"י ההגדרה:

$$Cov(X,Y)=E(X\cdot Y)-E(X)E(Y)$$
$$E(X\cdot Y)=\sum_{\substack{1\le i\le 4\\ 1\le j\le 4}} i\cdot j\cdot p(X=i,Y=j)=6.25$$

$$E(X)=\frac{30}{16},E(Y)=\frac{50}{16}$$
$$Cov(X,Y)=0.4$$

שאלה 3:

א. $1=\sum_{i=1}^{6} p(X=i)=9c\Rightarrow c=\frac{1}{9}$

ב.

X / Y	0	1
0	$\frac{2}{9}$	$\frac{2}{9}$
1	$\frac{1}{9}$	$\frac{4}{9}$

שאלה 4:

מאחר ו -Y X בת"ל הרי ש $p(X=i,Y=j)=p(X=i)p(y=j)$

X / Y	1	2	3	4
1	$1/64$	$3/64$	$3/64$	$1/64$
2	$3/64$	$9/64$	$9/64$	$3/64$
3	$3/64$	$9/64$	$9/64$	$3/64$
4	$1/64$	$3/64$	$3/64$	$1/64$

ולכן ע"ס הטבלה

$$p(X=Y)=\sum_{\substack{i=j\\1\le i\le 4}} a_{i,i}=20/64$$

וכן $p(X>Y)=\sum_{i<j} a_{i,j}=22/64$

Y+X	2	3	4	5	6	7	8
P(X+Y=k)	$1/64$	$6/64$	$15/64$	$20/64$	$15/64$	$6/64$	$1/64$

תמיד מתקיים: $E(X+Y)=E(X)+E(Y)$, ובמקרה זו $E(Y)=E(X)=20/8$

לכן $E(X+Y)=5$. מאחר ו- X , Y בעלי אותה התפלגות הרי שמתקיים-

$$V(X+Y)=2V(X)$$

נמצא את השונות של X:

$$V(X)=E(X^2)-(E(X))^2$$

$$E(X^2)=7, E(X)=\frac{20}{8}$$

$$V(X)=7-6.25=0.75$$

$$V(X+Y)=2V(X)=2\cdot 0.75=1.5$$

שאלה 5 :

נניח כי $m<n$ לפי ההגדרה: $\rho(X_m,X_n)=\dfrac{COV(X_m,X_n)}{\sqrt{VAR(X_m)VAR(X_n)}}$ נגדיר

$Y_1=X_1,\ Y_m=X_m-X_{m-1}$. אזי $\{Y_k\}_{k=1}^{\infty}$ הם משתנים ב"ת מפולגים באופן זהה: $Y_k\sim Geom(p)$.

$$VAR(X_n)=VAR(\sum_{k=1}^{n}Y_k)=\sum_{k=1}^{n}VAR(Y_k)=nVAR(Y_1)$$

מאחר ו- $X_m,\ X_n-X_m$ ב"ת נקבל:

$$COV(X_m,X_n)=COV(X_m,X_m+X_n-X_m)=COV(X_m,X_m)+COV(X_m,X_n-X_m)=$$
$$=COV(X_m,X_m)=VAR(X_m)$$

לכן:

$$\rho(X_m,X_n)=\frac{mVAR(X_1)}{\sqrt{mVAR(X_1)nVAR(X_1)}}=\sqrt{\frac{m}{n}}$$

שאלה 6:

א)

$$E(I_k)=P(I_k=1)=\frac{1}{n}$$
$$E(I_kI_m)=P(I_kI_m=1)=\frac{(n-2)!}{n!}=\frac{1}{n(n-1)},\ k\neq m$$

ב) אילו I_k ו- I_m היו ב"ת היה מתקיים $E(I_kI_m)=E(I_k)E(I_m)$. זה נוגד את התוצאה שקיבלנו בסעיף הקודם.

ג) $E(S_n)=E(I_1)+...+E(I_n)=\dfrac{n}{n}=1$

ד)

$$VAR(S_n)=\sum_{i=1}^{n}VAR(I_i)+\sum_{i\neq j}COV(I_i,I_j)=$$
$$=\sum_{i=1}^{n}\left(E(I^2{}_i)-E(I_i)^2\right)+\sum_{i\neq j}\left(E(I_iI_j)-E(I_i)E(I_j)\right)=$$
$$=n\left(\frac{1}{n}-\frac{1}{n^2}\right)+n(n-1)\left(\frac{1}{n(n-1)}-\frac{1}{n^2}\right)=$$
$$=\frac{n-1}{n}-\frac{1}{n}=1$$

שאלה 7:

א) התנאי לתלות: $f_{X,Y}(x,y) = f_X(x) f_Y(y),\ (x,y) \in \Re^2$. במקרה שלנו:

$$f_X(x) = \int_{-1}^{1} 1/4(1+xy)dy = 1/2,\ \ -1 \le x \le 1$$

$$f_Y(y) = 1/2,\ \ -1 \le y \le 1$$

$$\Rightarrow\ f_{X,Y}(x,y) \ne f_X(x) f_Y(y)$$

ב) עבור $s \le 1$ נקבל:

$$P(X^2 \le s) = P(-\sqrt{s} \le x \le \sqrt{s}) = \int_{-\sqrt{s}}^{\sqrt{s}} f_X(x)dx = \sqrt{s}, \quad 0 \le s \le 1$$

$$P(Y^2 \le t) = \sqrt{t}, \quad 0 \le t \le 1$$

$$\Rightarrow P(X^2 \le s, Y^2 \le t) = \int_{-\sqrt{t}}^{\sqrt{t}} \int_{-\sqrt{s}}^{\sqrt{s}} f_{X,Y}(x)dxdy = \sqrt{s}\sqrt{t}$$

ג)

$$COV(X^2, X^2+Y^2) = COV(X^2, X^2) + COV(X^2, Y^2) =$$

$$VAR(X^2) = E(X^4) - (E(X^2))^2 = 1/5 - (1/3)^2 = 4/45$$

שאלה 8:

נסמן ב-X וב-Y בהתאמה, את הזמנים (בדקות) לאחר השעה 12:00 שבהם הגבר והאישה מגיעים לפגישה. X ו-Y הם משתמשים מקריים בלתי תלויים, שלכל אחד מהם התפלגות אחידה בקטע [0,60]. לכן, ההסתברות המבוקשת היא- $P(X+10<Y)+P(Y+10<X)$. מטעמי סימטריה הסתברות זו שווה ל- $2P(X+10<Y)$, והיא מתקבלת כך-

$$2P(X+10<Y) = 2\iint_{x+10<y} f(x,y)dxdy = 2\iint_{x+10<y} f_X(x), f_Y(y)dxdy =$$

$$= 2\int_{10}^{60}\int_{0}^{y-10} (1/60)^2 dxdy = 2/60\int_{10}^{60}(y-10)dy = 25/36$$

שאלה 9:

גישה ישירה למומנטים:

$$M(t) = E(e^{tX}) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} = (pe^t + 1 - p)^n$$

גישה ב':

$$E(x^k)=\sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

$$use: \quad i\binom{n}{i}=n\binom{n-1}{i-1}$$

$$\Rightarrow E(x^k)=np\sum_{i=1}^{n} i^{k-1}\binom{n-1}{i-1}p^{i-1}(1-p)^{n-i}$$

$$=np\sum_{j=0}^{n-1}(j+1)^{k-1}\binom{n-1}{j}p^j(1-p)^{n-1-j}$$

$$=npE((Y+1)^{k-1})$$

כאשר Y הוא משתנה מקרי בינומי עם פרמטרים: (n-1,p). הצבת k=1 במשוואה האחרונה מניבה:

$$E(X)=E(X^1)=np$$

עבור k=2 נקבל: $E(X^2)=npE(Y+1)=np(E(Y)+1)=np((n-1)p+1)$

ולכן מאחר ש- $E(X)=np$ נקבל

שונות:
$$V(X)=E(X^2)-E^2(X)=np((n-1)p+1)-(np)^2$$
$$=np(1-p)$$

כשאר השוויון האחרון נובע מנוסחת הבינום. ומכאן:

$$M'(t)=n(pe^t+1-p)^{n-1}pe^t$$
$$\Rightarrow E(X)=M'(0)=np$$

וכן עבור השונות:
$$M''(t)=n(n-1)(pe^t+1-p)^{n-2}(pe^t)^2+n(pe^t+1-p)^{n-1}pe^t$$
$$\Rightarrow E(X^2)=M''(0)=n(n-1)p^2+np$$

ומכאן השונות של X נתונה ע"י:
$$V(X)=E(X^2)+E^2(X)=$$
$$=n(n-1)p^2+np-n^2p^2=np(1-p)$$

<u>שאלה 10:</u>

$$M_{X+Y}(t)=M_X(t)M_Y(t)=(pe^t+1-p)^n(pe^t+1-p)^m=$$
$$=(pe^t+1-p)^{n+m}$$

$(pe^t+1-p)^{n+m}$ היא פונקציה יוצרת מומנטים של התפלגות בינומית עם הפרמטרים m+n ו- p, ולכן זוהי בהכרח ההתפלגות של X+Y.

שאלה 11:

X- התוצאה הגבוהה ביותר (i), Y- מספר הקוביות עם תוצאה זוגית (j).

א)

X Y	1	2	3	4	5	6	
0	1	0	3	0	5	0	x1/36
1	0	2	2	4	4	6	
2	0	1	0	3	0	5	

ב)

$$P(X \ge 4, Y=1) = \frac{1}{36}(4+4+6) = \frac{14}{36} = \frac{7}{18}$$

$$P(X \ge 4, Y=2) = \frac{1}{36}(3+0+5) = \frac{8}{36} = \frac{2}{9}$$

$$P(X+Y \le 6) = 1 - P(X+Y>6) = 1 - \frac{1}{36}(6+0+5) = \frac{25}{36}$$

$$P(2Y < X) = \frac{1}{36}(1+0+3+0+5+0+2+4+4+6+0+5) = \frac{30}{36}$$

ג)

i	1	2	3	4	5	6	
P(X=i)	1	3	5	7	9	11	1/36x

i	0	1	2	
P(Y=j)	9	18	9	1/36x

ד)

$$E(X) = \sum_{i=1}^{6} iP(X=i) = \frac{1}{36}(1*1+2*3+3*5+4*7+5*9+6*11) = 4\frac{17}{36}$$

$$E(Y) = \sum_{j=0}^{2} jP(Y=j) = \frac{1}{36}(0*9+1*18+2*9) = 1$$

$$E(X^2) = \sum_{i=1}^{6} i^2 P(X=i) = 21\frac{35}{36}$$

$$V(X) = E(X^2) - E^2(X) = 1.9715$$

$$V(Y) = 0.5$$

ה) X,Y לא בלתי תלויים מכיוון: $P(X=3, Y=0) = \frac{3}{36} \ne P(X=3)P(Y=0) = \frac{1.25}{36}$

ו) עבור Z=X+Y

s	1	2	3	4	5	6	7	8	
P(Z=s)	1	0	5	3	9	7	6	5	1/36x

לכן:

s	1	3	4	5	6	7	8	
P(Z=s)	1	5	3	9	7	6	5	1/36x

ז) תוחלת ושונות של Z:

$$E(Z)=E(X+Y)=E(X)+E(Y)=4\frac{17}{36}+1$$

$$V(Z)=V(X+Y)=V(X)+V(Y)+2COV(X,Y)=$$

$$=1.9715+0.5+2COV(X,Y)$$

$$COV(X,Y)=E(XY)-E(X)E(Y)$$

$$E(XY)=\sum_{j=0}^{2}\sum_{i=1}^{6} jiP(X=i,Y=j)$$

גישה נוספת: $E(Z)=\sum_{s=1}^{8} sP(Z=s)$

ח) פונקציה יוצרת מומטים של X:

$$M_X(t)=E(e^{tx})=\sum_{i=1}^{6} e^{ti}P(X=i)=$$

$$=e^{1t}\frac{1}{36}+e^{2t}\frac{3}{36}+e^{3t}\frac{5}{36}+e^{4t}\frac{7}{36}+e^{5t}\frac{9}{36}+e^{6t}\frac{11}{36}$$

ט) פונקציה יוצרת מומטים של Y:

$$M_Y(t)=E(e^{ty})=\sum_{j=0}^{2} e^{tj}P(Y=j)=$$

$$=e^{0t}\frac{9}{36}+e^{1t}\frac{18}{36}+e^{2t}\frac{9}{36}$$

י) פונקציה יוצרת מומטים של Z:

$$M_Z(t)=E(e^{tz})=\sum_{s=1}^{8} e^{ts}P(Z=s)=$$

$$=e^{1t}\frac{1}{36}+e^{2t}\frac{0}{36}+e^{3t}\frac{5}{36}+e^{4t}\frac{3}{36}+e^{5t}\frac{9}{36}+e^{6t}\frac{7}{36}+e^{7t}\frac{6}{36}+e^{8t}\frac{5}{36}$$

יא) תוחלת ושונות של X באמצעות פונקציה יוצרת מומנטים:

$$M'_X(t)=e^{1t}\frac{1}{36}+2e^{2t}\frac{3}{36}+3e^{3t}\frac{5}{36}+4e^{4t}\frac{7}{36}+5e^{5t}\frac{9}{36}+6e^{6t}\frac{11}{36}$$

$$M'_X(0)=E(X)=\frac{1}{36}+\frac{6}{36}+\frac{15}{36}+\frac{28}{36}+\frac{45}{36}+\frac{66}{36}=\frac{161}{36}=4.472$$

$$M''_X(t)=e^{1t}\frac{1}{36}+4e^{2t}\frac{3}{36}+9e^{3t}\frac{5}{36}+16e^{4t}\frac{7}{36}+25e^{5t}\frac{9}{36}+36e^{6t}\frac{11}{36}$$

$$M''_X(0)=E(X^2)=\frac{1}{36}+\frac{12}{36}+\frac{45}{36}+\frac{112}{36}+\frac{225}{36}+\frac{396}{36}=\frac{791}{36}=$$

$$V(X)=E(X^2)-E^2(X)=21.972-(4.472)^2=1.974$$

וכך עבור Y ו-Z.

:שאלה 12

	0	1	2	Py
0	(1/3)*(1/2)	(1/3)*(1/4)	(1/3)*(1/4)	4/12
1	(1/3)*(1/4)	(1/3)*(1/2)	(1/3)*(1/4)	4/12
2	(1/3)*(1/4)	(1/3)*(1/4)	(1/3)*(1/2)	4/12
Px	4/12	4/12	4/12	1

E(X)=1 , E(X)=1 , E(XY)=7/6

COV(X,X)=E(XX)-E(X)E(X)=1/6

V(X)=2/3 , V(X)=2/3

$$\rho(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma(x)\sigma(y)}=0.25$$

:שאלה 13

א.

	0	1	2	Py
0	1/10	2/10	0	3/10
1	2/10	3/10	1/10	6/10
2	0	1/10	0	1/10
Px	3/10	6/10	1/10	1

ב. Cov(x,y)=E(xy)-E(x)E(y)=0.06 - הם מתואמים.

ג. $\rho(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma(x)\sigma(y)}=\frac{0.06}{0.6\cdot 0.6}=1/6$

ד. $P(y=1\mid x\geq 1)=\frac{P(x=1,y=1)+p(x=2,y=1)}{p(x=1)+p(x=2)}=0.571$

:שאלה 14

	1	2	P(x)
0	1/8	1/16	3/16
1	1/4	3/16	7/16
2	1/8	3/16	5/16
3	0	1/16	1/16
P(N)	0.5	0.5	1

יש לפתור לפי התפלגות בינומית. לדוגמה:

$$P(X=0,N=1)=1/2\binom{2}{0}0.5^{0}0.5^{2}$$

E(X)=20/16 , E(N)=1.5 , E(x*N)=2

Cov(X,N)=0.125

$$\rho(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma(x)\sigma(y)}=\frac{0.125}{0.789\cdot 0.5}=0.3$$

רווחי סמך ובדיקת השערות

3. כאשר השונות באוכלוסיה ידועה

שאלה 1

מכונה מייצרת ברגים אשר אורכם מתפלג נורמלית עם תוחלת 4 ס"מ וסטיית תקן 0.2 ס"מ.

עקב תקלה במכונה הועלה חשד כי המכונה אינה מייצרת ברגים באורך הנדרש.

לשם בדיקה נלקח מדגם מקרי של 25 ברגים אשר יוצרו במכונה לאחר גילוי התקלה ונמצא כי האורך הממוצע שלהם הוא 3.9 ס"מ .

מהו רו"ס לתוחלת אורך הברגים לאחר התקלה בר"מ 2% ?

שאלה 2

משקל של ילדים מתפלג נורמלית עם סטית תקן 1.25 ק"ג.

ידוע שממוצע מדגם בגודל n הוא 52 ק"ג . בונים רווח סמך לתוחלת ברמת בטחון של 95% .

מה צריך להיות n ע"מ שאורכו של רווח הסמך יהיה 1 ?

שאלה 3

קיים ויכוח בין מרצה של קורס מסוים לסטודנטים. המרצה טוען שממוצע הציונים בקורס הוא 67.5 ואילו הסטודנטים טוענים שהוא נמוך יותר.

כדי להכריע בסוגיה, לקחו מדגם של 100 סטודנטים, בדקו את ציוניהם והתקבל שהממוצע הוא 65. ממבחני העבר ידוע שסטיית התקן היא 3.

א. מי צודק ? יש להגיע להכרעה באמצעות חישוב רו"ס ברמת בטחון של 95%.
ב. מהו גודל המדגם המינימלי שיש לקחת כך שבהסתברות של 95% הסטייה בין ממוצע המדגם לתוחלת לא תעלה על 2 נקודות ?

שאלה 4:

הזמן הדרוש לבצע ניתוח מסוים מתפלג נורמלית עם תוחלת של 8 שעות וסטיית תקן של 4 שעות.

רופא מציע שיטת ניתוח חדשה לקיצור התהליך.

השיטה נבדקת על מדגם של $n = 36$ מנותחים , ובמדגם נמצא ממוצע של $\bar{X} = 6.8$ שעות.

א. האם שיטת הניתוח המוצעת אמנם מקצרת את זמן הניתוח, ברמת מובהקות 0.01 ?
ב. מצא רו"ס לתוחלת זמן הניתוח לפי השיטה החדשה ברמת סמך של 98% .
ג. מה הקשר בין תשובות א' ל- ב' ?

שאלה 5:

חוקר מבצע מבחן לבדיקת ההשערות $H_0: \mu \le 100$ כנגד $H_0: \mu > 100$ בעזרת מדגם בגודל $n = 225$. נתון כי השונות היא $\sigma^2 = 2025$. החוקר החליט כי ידחה את השערת האפס אם $\bar{X} > 105.25$ ואחרת, לא ידחה את השערת האפס.

רמת המובהקות של המבחן היא (בקירוב):

1. 0.01
2. 0.04
3. 0.05
4. לא ניתן לדעת ללא נתונים נוספים

שאלה 6:

מכונה למילוי בקבוקים בתרופה מסוימת מכוונת למלא בממוצע 200 מ"ג תרופה לבקבוק עם סטיית תקן של 10 מ"ג. (עפ"י התפלגות נורמלית)

חושדים שהמכונה התקלקלה ואינה עוד עקבית במילוי תקין של כמות התרופה, לצורך בדיקה , נלקח מדגם מקרי של 25 בקבוקים , ובמדגם התקבל ממוצע $\overline{X} = 205$ מ"ג. האם החשד מוצדק ברמת מובהקות $\alpha = 0.05$?

שאלה 7

מהו גודל המדגם המינימלי (n) לבדיקת : $H_0 : \mu = 3500$ מול $H_1 : \mu = 3600$ כאשר

$\sigma = 200$, $\alpha = 5\%$, ו - $\beta \le 1\%$?

שאלה 8

בבדיקת השערות, בהשוואה בין בחירה של רמת מובהקות $\alpha = 0.05$ לעומת $\alpha = 0.01$, נכון לומר כי:

1. בבחירת $\alpha = 0.05$ אנו מוכנים לטעות מסוג II גדולה יותר.
2. בבחירת $\alpha = 0.05$ אנו מוכנים לטעות מסוג I קטנה יותר.
3. בבחירת $\alpha = 0.05$ אנו מוכנים לטעות מסוג I גדולה יותר.
4. מבחן עם $\alpha = 0.05$ נחשב למבחן שמרני יותר.

שאלה 9

שני חוקרים בודקים השערות ברמת מובהקות 0.05 על סמך אותו מדגם. חוקר א' בודק את ההשערות

$H_0 : \mu = \mu_0$ כנגד $H_1 : \mu \ne \mu_0$ ומסיק כי יש לדחות את השערת האפס. חוקר ב' בודק את ההשערות

$H_0 : \mu = \mu_0$ כנגד $H_0 : \mu > \mu_0$. מהי מסקנתו של חוקר ב'?

1. יש לדחות את השערת האפס.
2. אין לדחות את השערת האפס.
3. בהכרח ה- Pvalue של חוקר א' שווה לכפלים ה-Pvalue של חוקר ב'.
4. ללא נתונים נוספים לא ניתן לדעת מהי מסקנתו.

<u>שאלה 10</u>

להלן 2 טענות:

א. אם השערת האפס נדחתה ברמת מובהקות 0.05 , אזי ניתן להסיק כי גם עבור כל רמת מובהקות <u>הקטנה</u> מ- 0.05 נדחה את השערת האפס.

ב. אם בבדיקת השערות $H_0: \mu = \mu_0$ כנגד $H_1: \mu \neq \mu_0$ נדחה את השערת האפס ברמת מובהקות α, אזי בהכרח גם בבדיקת השערות $H_0: \mu = \mu_0$ כנגד $H_0: \mu > \mu_0$ נדחה את השערת האפס ברמת מובהקות α, על סמך אותו המדגם.

1. שתי הטענות נכונות.
2. שתי הטענות שגויות.
3. טענה א' נכונה וטענה ב' שגויה.
4. טענה א' שגויה וטענה ב' נכונה.

<u>*פתרונות*</u>

<u>שאלה 1:</u>

נסמן ב- X את אורך הברגים לאחר התקלה , $X \sim N(\mu, 0.2^2)$.

כמו כן נתון : $n = 25$ ו- $\bar{X} = 3.9$

רמת מובהקות של 2% $\Leftarrow$ $\alpha = 0.02$ $\Leftarrow$ $Z_{1-\frac{\alpha}{2}} = 2.325$

רו"ס ברמת מובהקות של α לתוחלת כאשר השונות ידועה:

$\bar{X} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ נציב את הנתונים ונקבל:

$3.9 - 2.235 \cdot \frac{0.2}{5} \leq \mu \leq 3.9 + 2.235 \cdot \frac{0.2}{5}$ $\Leftarrow$ $3.807 \leq \mu \leq 3.993$.

<u>שאלה 2:</u>

נשתמש בנוסחא רו"ס ברמת מובהקות של α לתוחלת כאשר השונות ידועה:

$$\bar{X}-Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{X}+Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$$

במקרה שלנו, רמת ביטחון של 95% $\Rightarrow$ $\alpha=0.05$ $\Rightarrow$ $Z_{1-\frac{\alpha}{2}}=1.96$. נציב ונקבל:

$52-1.96\cdot\frac{1.25}{\sqrt{n}}\le\mu\le 52+1.96\cdot\frac{1.25}{\sqrt{n}}$. נרצה שאורך רווח הסמך יהיה 1, ז"א שכל סטייה מהממוצע שווה ל-0.5.

$1.96\cdot\frac{1.25}{\sqrt{n}}=0.5\Rightarrow\sqrt{n}=4.9\Rightarrow n\approx 24$. ז"א שיש צורך במדגם בגודל 24 ילדים.

<u>שאלה 3:</u>

א.

נסמן ב- X את הציון בקורס , $X\sim N(\mu,3^2)$.

כמו כן נתון : $n=100$ ו- $\bar{X}=65$

רמת ביטחון של 95% $\Rightarrow$ $\alpha=0.05$ $\Rightarrow$ $Z_{1-\frac{\alpha}{2}}=1.96$.

רו"ס ברמת מובהקות של α לתוחלת כאשר השונות ידועה:

$\bar{X}-Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{X}+Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$ נציב את הנתונים ונקבל:

$65-1.96\cdot\frac{3}{10}\le\mu\le 65+1.96\cdot\frac{3}{10}$ $\Rightarrow$ $64.412\le\mu\le 65.588$. לכן הסטודנטים צודקים.

ב.

נשתמש בנוסחא רו"ס ברמת מובהקות של α לתוחלת כאשר השונות ידועה:

$$\bar{X}-Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{X}+Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$$

במקרה שלנו, רמת ביטחון של 95% $\Leftarrow \alpha = 0.05 \Leftarrow Z_{1-\frac{\alpha}{2}} = 1.96$. נציב ונקבל:

$65 - 1.96 \cdot \frac{3}{\sqrt{n}} \le \mu \le 65 + 1.96 \cdot \frac{3}{\sqrt{n}}$. נרצה שאורך כל סטייה מהממוצע יהיה קטן מ-2.

$1.96 \cdot \frac{3}{\sqrt{n}} \ge 2 \Rightarrow \sqrt{n} \ge 2.94 \Rightarrow n \ge 8.6436 \Rightarrow n \ge 9$. ז"א שיש צורך בלפחות 9 סטודנטים.

<u>שאלה 4</u>

א.

נתון: $X \sim N(8, 4^2)$, $n = 36$, $\bar{X} = 6.8$.

ברצוננו לבדוק את ההשערות: $\begin{matrix} H_0 : \mu = 8 \\ H_1 : \mu < 8 \end{matrix}$.

לשם כך נשתמש במבחן חד צדדי שמאלי: דחה H_0 אם $\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$.

אצלנו $\alpha = 0.01 \Rightarrow Z_{1-\alpha} = Z_{0.99} = 2.325$.

$\bar{X} < 8 - 2.325 \cdot \frac{4}{\sqrt{36}} = 6.45$. במקרה שלנו $\bar{X} = 6.8 > 6.45$, ולכן לא נדחה את השערת האפס ונסיק ששיטת הטיפול החדשה אינה מקצרת את זמן הניתוח.

נחשב P_value:

תחת H_0 $\bar{X} \sim N(8, \frac{4^2}{36})$.

$$P.V. = P(\bar{X} \le 6.8 \mid H_0) = P(Z \le \frac{6.8 - 8}{4/6}) = P(Z \le -1.8) = \Phi(-1.8) = 1 - \Phi(1.8) =$$
$$= 1 - 0.9641 = 0.0359$$

$P.V. = 0.0359 > 0.01 = \alpha$ ולכן בהתאם לתוצאות לעיל לא נדחה את H_0.

ב. נשתמש בנוסחא למציאת רו"ס ל- μ: $\bar{X}-Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{X}+Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}$.

רמת הביטחון המבוקשת היא 98% $\Leftarrow \alpha=0.02 \Leftarrow Z_{1-\frac{\alpha}{2}}=Z_{0.99}=2.325$.

נציב ונקבל $6.8-2.325\cdot\frac{4}{\sqrt{36}}\le\mu\le 6.8+2.325\cdot\frac{4}{\sqrt{36}} \Leftarrow 5.25\le\mu\le 8.35$.

התוחלת לפי השערת האפס (8) מוכלת ברו"ס ולכן ברמת ביטחון של 98% נסיק שהשיטה החדשה לא מקצרת את זמן הניתוח באופן מובהק.

ג. מבחן חד צדדי עם $\alpha=0.01$ לבדיקת μ אמרו לתת את אותה מסקנה של רו"ס ל- μ עם $\alpha=0.02$ מכיוון שרו"ס הוא תמיד דו צדדי ובכל צד יש רק חצי מרמת המובהקות.

__שאלה 5:__

החוקר בשאלה מעוניין לבצע מבחן חד צדדי ימני: $\begin{matrix}H_0:\mu\le 100\\ H_1:\mu>100\end{matrix}$.

לשם כך נשתמש במבחן: דחה H_0 אם $\bar{X}>\mu_0+Z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}$.

עפ"י הנתון בשאלה $n=225$, $\sigma^2=2025$. החוקר החליט שיידחה את H_0 אם $\bar{X}>105.25$.

נשווה בין שני הביטויים שמתארים את הנקודה הקריטית של המבחן:

$$Z_{1-\alpha}=1.75 \Leftarrow 3\cdot Z_{1-\alpha}=5.25 \Leftarrow 100+Z_{1-\alpha}\cdot\frac{45}{\sqrt{225}}=105.25$$

$$\alpha=0.0401 \Leftarrow 1-\alpha=0.9599 \Leftarrow$$

לכן נבחר בתשובה 2.

שאלה 6:

מעוניינים לבדוק את ההשערות: $\begin{matrix} H_0 : \mu = 200 \\ H_1 : \mu \neq 200 \end{matrix}$.

נתון: $n = 25$, $\bar{X} = 205$

לשם כך נשתמש במבחן דו צדדי כאשר השונות ידועה ונדחה H_0 בר"מ α אם

$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$.

במקרה שלנו $\alpha = 0.05 \Rightarrow Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$.

נציב את הנתונים ונקבל $\bar{X} > 200 + 1.96 \cdot \frac{10}{\sqrt{25}} = 203.92$ או $\bar{X} < 200 - 1.96 \cdot \frac{10}{\sqrt{25}} = 196.08$.

$\bar{X} = 205 > 203.92$ ולכן נדחה את H_0 ונסיק שהמכונה התקלקלה.

נחשב P_value:

תחת H_0: $\bar{X} \sim N(200, \frac{10^2}{25})$.

$$P.V. = P(\bar{X} \geq 205 \mid H_0) = P(Z \geq \frac{205-200}{10/5}) = P(Z \geq 2.5) = 1 - \Phi(2.5) =$$
$$= 1 - 0.9938 = 0.0124$$

$P.V. = 0.0124 < 0.05 = \alpha$ ולכן בהתאם לתוצאות לעיל נדחה את H_0 .

שאלה 7

גודל המדגם המינימלי (n) לבדיקת : $H_0 : \mu = 3500$ מול $H_1 : \mu = 3600$

כאשר נתון : $\sigma = 200$

$\alpha = 5\%$

ו- $\beta \leq 1\%$

זהו מבחן חד-צדדי ימני ולכן יש לדחות את H_0 אם :

$$\overline{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} = 3500 + 1.645 \cdot \frac{200}{\sqrt{n}}$$

נשתמש בנתון על β ע"מ לבנות אי-שוויון :

$$\beta = P(\overline{X} \le 3500 + 1.645 \cdot \frac{200}{\sqrt{n}} \mid H_1)$$

תחת H_1 : $\overline{X} \sim N(3600, \frac{200^2}{n})$

ולכן : $$\beta = P(Z \le \frac{3500 + 1.645 \cdot \frac{200}{\sqrt{n}} - 3600}{\frac{200}{\sqrt{n}}}) = \Phi(1.645 - \frac{\sqrt{n}}{2})$$

נתון : $\beta \le 1\%$

$\Leftarrow$ $\Phi(1.645 - 0.5\sqrt{n}) \le 0.01$

$\Leftarrow$ $1.645 - 0.5\sqrt{n} \le Z_{0.01}$

$\Leftarrow$ $0.5\sqrt{n} - 1.645 \ge -Z_{0.01} = Z_{0.99} = 2.326$

$\Leftarrow$ $0.5\sqrt{n} \ge 3.971$

$\Leftarrow$ $\sqrt{n} \ge 7.942$

$\Leftarrow$ $n \ge 63.07$

n תמיד מספר טבעי ולכן נעגל כלפי מעלה $\Leftarrow$ $n \ge 64$

<u>שאלה 8</u>

תשובה ג'.

הסבר : α היא ההסתברות לטעות מסוג ראשון.

ככל שניקח α גדולה יותר , ההסתברות לטעות טעות מסוג ראשון גדולה יותר , ז"א אנו מוכנים לטעות מסוג ראשון גדולה יותר.

שאלה 9

תשובה ד'.

הסבר: ייתכן כי חוקר א' דחה את השערת האפס מהכיוון השמאלי שלה , וכשמחליטים על מבחן דו-צדדי ימני על סמך אותו מדגם מסיקים לא לדחות את השערת האפס.

ולכן אם לא נתון מאיזה צד דחה את השערת האפס – לא נוכל לדעת מה המסקנה לגבי מבחן חד-צדדי.

שאלה 10

תשובה ב'.

הסבר : טענה א' לא נכונה :

השערת האפס נדחית אם $p.v < \alpha$.

אם השערת האפס נדחתה עבור $\alpha = 0.05$, ניתן להסיק ש - $p.v < 0.05$, אך לא ידוע בדיוק מהו ערכו של ה- $p.v.$ וייתכן שעבור ערכים מסוימים של α הקטנים מ- 0.05 , ה- $p.v.$ הספציפי הזה יהיה גדול מהם ולכן לא נדחה את השערת האפס.

טענה ב' לא נכונה : ראה הסבר לתשובה מספר 3

רווחי סמך ובדיקת השערות

4. כאשר השונות באוכלוסיה אינה ידועה

שאלה 1

במדגם של 9 גברים העובדים בתחום מסוים נמצא כי בממוצע עבדו 43 שעות בשבוע, עם סטיית תקן של 3 שעות. בהנחה כי מספר שעות העבודה השבועיות מתפלג נורמלית:

א. מצא רווח סמך ברמת סמך של 90% לתוחלת מספר שעות העבודה השבועיות בתחום זה.

ב. אם באוכלוסייה מקובל כי גברים עובדים בממוצע 45 שעות בשבוע, באיזו רמת מובהקות נוכל להסיק כי בתחום הנ"ל עובדים פחות?

האם תשובתך מתאימה לרווח הסמך מסעיף א'?

שאלה 2

באזור מסוים היו עד כה דירות עם שטח ממוצע של 90 מ"ר. במדגם של דירות אשר נבנו לאחרונה נמצאו השטחים הבאים: 95, 125, 100, 140, 80, 120 מ"ר.

א. מצא רווח סמך לתוחלת שטח דירה חדשה ברמת סמך של 98%.
ב. מהי רמת המובהקות המינימלית עבורה נוכל להסיק כי תוחלת שטח דירה חדשה באזור שונה מתוחלת שטח דירה ישנה?

השווה לתשובתך בסעיף א'.

שאלה 3

בכדי לאמוד את אחוז הגירושין בארה"ב נלקח מדגם מקרי של 625 איש ונמצא ש-125 מהם גרושים.

ו. בנו רו"ס לאחוז הגירושין ברמת ביטחון של 95%.
ז. איך ישתנה אורך רו"ס עבור רמת ביטחון 90% ? 99% ? (ענו ללא חישוב נוסף)
ח. אם נגדיל את המדגם פי 4 כיצד יושפע אורך רו"ס?
ט. מהו גודל המדגם הדרוש כדי להקטין את אורך הרווח פי 2 ?
י. אמדו את אחוז האנשים חסרי השכלה גבוהה באוכלוסיה ברמת סמך 95%.

שאלה 4:

זקני צפת טוענים כי בירושלים יורדים בממוצע 800 מ"ל ליטר גשם בשנה.

במדידות שבוצעו על פני 9 השנים האחרונות נבדקו כמויות הגשמים והתוצאות הן:

810 , 750 , 789 , 800 , 790 , 795 , 797 , 799 , 802

א. מצא את ה- p.value
ב. האם כמות הגשמים השתנתה בשנים האחרונות ? מהי המסקנה עבור $\alpha = 0.01$, $\alpha = 0.05$?
ג. בנה רו"ס ברמת מובהקות 0.05 .
ד. כיצד ישתנה רווח הסמך כאשר $n = 25$?

שאלה 5:

רמה של חומר מסוים בדמו של אדם בריא מתפלגת נורמלית עם תוחלת 20 יחידות.

קיימת השערה שבגלל זיהום האוויר באזור ת"א עולה רמת החומר בדמם של תושבי ת"א מעל הרמה הרגילה.

לצורך בדיקת השערה זו, נלקח מדגם מקרי של 9 תושבי ת"א ונמצא שרמת החומר הממוצע בדמם הייתה 22 יחידות עם סטיית תקן 3.

האם ההשערה נכונה עבור $\alpha = 0.05$? $\alpha = 0.01$?

שאלה 6:

חוקר ביצע מבחן דו צדדי לבדיקת השערת האפס $H_0 : \mu = \mu_0$ על סמך מדגם מקרי בגודל 5, מתוך אוכלוסייה המתפלגת נורמלית והשונות אינה ידועה.
סטטיסטי המבחן שהתקבל היה: $t_{\overline{X}} = -2.611$. לכן המסקנה היא:

1. התוצאה מובהקת ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
2. התוצאה מובהקת ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025.
3. התוצאה מובהקת ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01.
4. התוצאה אינה מובהקת ברמת מובהקות 0.1.

שאלה 7:

במבחן השערה חד צדדי ברמת מובהקות של 5% תוכנן, לפי מהותו ואופיו, להשתמש בהתפלגות Z (התפלגות נורמלית). משה סיגמא המתמחה טעה והשתמש בטבלת התפלגות t (התפלגות סטודנט) עם 9 דרגות חופש (בתוך טבלה זו הוא בחר נכון את ערכי t, ז"א התייחס לעובדה שמדובר במבחן חד צדדי ורמת המובהקות שצוינה). משה סיגמא דיווח את תוצאת המבחן ולפיה נדחתה השערת האפס והתקבלה ההשערה האלטרנטיבית.

ציין את המשפט הנכון המתייחס לאירוע שלעיל:

1. בהיעדר נתונים נוספים אי אפשר לקבוע אם אכן היה מגיעה לאותה תוצאה או לתוצאה אחרת.
2. השימוש בהתפלגות Z והתפלגות t הוא חלופי ולכן אין זה משנה באיזה מהן הוא בחר, תמיד התוצאות תהיינה זהות.
3. לאותה תוצאת מבחן היה מגיע משה סיגמה אם היה משתמש (כפי שהיה צריך לעשות) בהתפלגות נורמלית.
4. לו היה משתמש בהתפלגות הנורמלית (כפי שהיה צריך לעשות) היה מגיע לתוצאה הפוכה.

שאלה 8:

שני חוקרים בודקים השערות ברמת מובהקות 0.05 על סמך אותו מדגם. חוקר א' בודק את ההשערות

$H_0: \mu = \mu_0$ כנגד $H_1: \mu \neq \mu_0$ ומסיק כי יש לדחות את השערת האפס. חוקר ב' בודק את ההשערות

$H_0: \mu = \mu_0$ כנגד $H_0: \mu > \mu_0$. מהי מסקנתו של חוקר ב'?

5. יש לדחות את השערת האפס.
6. אין לדחות את השערת האפס.
7. בהכרח ה- Pvalue של חוקר א' שווה לכפלים ה-Pvalue של חוקר ב'.
8. ללא נתונים נוספים לא ניתן לדעת מהי מסקנתו.

שאלה 9:

במחקר מסוים מעוניין החוקר לבנות רווח סמך ברמת סמך 90% לתוחלת של משתנה המפולג נורמלית. לשם כך אסף מדגם בגודל $n = 7$ וחישב את ממוצע המדגם $\bar{X}$. מאחר ולא ידע את השונות האמיתית, חישב את האומד לשונות S^2 מתוך המדגם. בטעות חישב החוקר את רווח הסמך לפי טבלת Z בעזרת הנוסחא: $\left(\bar{X} - 1.645\frac{S}{\sqrt{n}}, \bar{X} + 1.645\frac{S}{\sqrt{n}}\right)$.

אם היה החוקר מחשב את רווח הסמך על סמך ההתפלגות הנכונה:

1. רווח הסמך השגוי יתקבל להיות רחב יותר מרווח הסמך הנכון.
2. רווח הסמך הנכון יתקבל להיות רחב יותר מרווח הסמך השגוי.
3. במקרה זה שני רווחי הסמך יתקבלו להיות זהים.
4. אין מספיק נתונים בכדי להשוות בין שני רווחי הסמך.

שאלה 10

להלן רשומים משקלם של עכברים אשר הואכלו במשך 28 יום בשמן דגים: 24, 23.5, 24, 24, 25, 22.5, 20, 23.5, 28.5, 18, 20, 26. בדרך כלל עכבר מסוג זה גדל למשקל של 22 גרם.

בצע מבחן דו צדדי והחלט האם שמן דגים משפיע על משקל העכבר עבור α=0.05. ציין את ההשערות.

פתרונות

שאלה 1:

א. נמצא רווח סמך לתוחלת כאשר השונות לא ידועה: $\left(\bar{X}-t_{(n-1)}^{1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{X}+t_{(n-1)}^{1-\alpha/2}\frac{s}{\sqrt{n}}\right)$

עפ"י נתוני השאלה: $\bar{X}=43$, $n=9$, $\alpha=0.1$, $s=3$ $\Leftarrow t_{(n-1)}^{1-\alpha/2}=t_{(8)}^{0.95}=1.86$.

נציב את הנתונים ונקבל: $43\pm\frac{3}{\sqrt{9}}t_{8,0.95}=43\pm1.860$

$$41.14\le\mu\le44.86$$

ב. נמצא P_value עבור ההשערות: $\begin{matrix}H_0:\mu=45\\ H_1:\mu<45\end{matrix}$

$$P\_value=P(\bar{X}<43)=P(t<\frac{43-45}{3/\sqrt{9}})=P(t<-2)=1-P(t<2)$$

נחפש את הערך 2 בטבלת t בשורה 8 $(n-1)$: $1.860\le2\le2.306$

$\Leftarrow 0.950\le1-P(t<2)\le0.975$. נחסר 1 מכל אחד מהאגפים ונכפיל פי 1-:

$$0.025\le P\_value\le0.05$$

לכן לכל רמת מובהקות הגדולה מ-0.05 נדחה את ההנחה הראשונית ונקבע כי בתחום הנ"ל עובדים פחות שעות מהממוצע.

ואכן 45 (התוחלת תחת השערת האפס) אינו נמצא ברווח הסמך של סעיף א'.

שאלה 2:

א. נמצא את הממוצע וסטית התקן עפ"י המדגם:

$$\bar{X}=110\ ,\ s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=490=22.136^2\ .\ \alpha=0.02$$

נמצא רווח סמך לתוחלת כאשר השונות לא ידועה: $\left(\bar{X}-t_{(n-1)}^{1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{X}+t_{(n-1)}^{1-\alpha/2}\frac{s}{\sqrt{n}}\right)$

$t_{(n-1)}^{1-\alpha/2}=t_{(5)}^{0.99}=3.365$. נציב את הנתונים ונקבל:

$$79.5704\le\mu\le140.4196 \Leftarrow 110\pm t_{5,0.99}\frac{22.136}{\sqrt{6}}=110\pm3.365\cdot9.04$$

ב. נמצא $P-value$ עבור ההשערות:

$$H_0:\mu=90$$
$$H_1:\mu\ne90$$

$$P\_value=P(\bar{X}>110)=P(t>\frac{110-90}{22.136/\sqrt{6}})=P(t>2.213)=1-P(t<2.213)$$

נחפש את הערך 2.213 בטבלת t בשורה 5 $(n-1)$: $2.015<2.213<2.571$

$\Leftarrow 0.950\le1-P(t<2.213)\le0.975$. נחסר 1 מכל אחד מהאגפים ונכפיל פי 2- מכיוון שההשערה היא דו צדדית: $0.05\le P\_value\le0.1$

לכן עבור רמת מובהקות של 10% ומעלה נדחה את השערת האפס. בסעיף א' רמת המובהקות הייתה 2% לכן לא דחינו את השערת האפס, ואכן התוחלת 90 הייתה ברווח הסמך.

שאלה 3:

ג. נתון: $n=625$, $\hat{p}=\frac{125}{625}=0.2$

נוסחא לרו"ס לפרופורציה: $\hat{p}-Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}\le p\le\hat{p}+Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$

95% $\Leftarrow$ $Z_{1-\frac{\alpha}{2}}=1.96$

$$0.17\le p\le0.23 \quad\Leftarrow\quad p\in0.2\pm1.96\sqrt{\frac{0.2\cdot0.8}{625}}$$

ד. אם נקטין את רמת הביטחון ל-90% גם רווח הסמך יקטן, ולחילופין אם נגדיל את רמת הביטחון ל-99% רווח הסמך יגדל.

ה. מכיוון שהפרופורציה עצמה תלויה בגודל n ולא רק רווח הסמך, אם נגדיל את גודל המדגם לא ברור כיצד רווח הסמך ישתנה. הדבר שונה מרווח הסמך עבור תוחלת שבו אם נגדיל את n רווח הסמך יקטן.

ו. מכיוון שכפי שציינו לעיל שינוי גודל המדגם משפיע גם על הפרופורציה, ואין לנו שום מידע לגבי הפרופורציה באוכלוסיה נציב $\hat{p}=0.5$ עבורו הביטוי $\hat{p}(1-\hat{p})$ מקבל ערך מקסימאלי כך שנקבל חסם עליון ל-n.

כדי להקטין את אורך הרווח פי 2 אפשר פשוט להקטין את הסטייה מ- $\hat{p}$ פי 2 , כלומר:

$$1.96\cdot\sqrt{\frac{0.5\cdot0.5}{n}}=0.5\cdot0.03$$

$\sqrt{\frac{0.5\cdot0.5}{n}}=\frac{0.015}{1.96}\Leftarrow n=4268.4$. נעגל ונקבל כי יש לקחת מדגם בגודל של 4269 כדי להקטין את הרווח פי 2 .

ז. נתון: $n=625$, $\hat{p}=\frac{500}{625}=0.8$

נוסחא לרו"ס לפרופורציה: $\hat{p}-Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}\le p\le\hat{p}+Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$

95% $\Leftarrow$ $Z_{1-\frac{\alpha}{2}}=1.96$

$p\in0.8\pm1.96\sqrt{\frac{0.2\cdot0.8}{625}}$ $\Leftarrow$ $0.77\le p\le0.83$. נשים לב כי רווח סמך זה הוא המשלים של רווח הסמך מסעיף א' מכיוון הוא רווח סמך להסתברות המשלימה, וכן מכיוון שהסטייה מהפרופורציה שווה.

שאלה 4:

א+ב).

ברצוננו לבדוק את ההשערות: $\begin{matrix}H_0:\mu=800\\H_1:\mu\ne800\end{matrix}$.

במבחן לבדיקת $\begin{matrix}H_0:\mu=\mu_0\\H_1:\mu\ne\mu_0\end{matrix}$ כאשר השונות לא ידועה מתקיים: $\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{n-1}$.

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

נחשב את ממוצע ושונות המדגם:

$$\bar{X}=\frac{802+799+...+810}{9}=792.4$$

$$S^2=\frac{(802-792.4)^2+...+(810-792.4)^2}{8}=17.125^2$$

$$\frac{P.V.}{2}=P(\bar{X}\le 792.4\,|\,H_0)=P(t_8\le\frac{792.4-800}{17.125/\sqrt{9}})=P(t_8\le -1.33)=1-P(t_8\le 1.33)$$

$$0.2<P.V.<0.4 \Leftarrow 0.1<1-P(t_8\le 1.33)<0.2 \Leftarrow 0.8<P(t_8\le 1.33)<0.9$$

לכן גם עבור $\alpha=0.05$ וגם עבור $\alpha=0.01$ לא נדחה את השערת האפס. $0.2<P.V.$.

ג). רו"ס ל- μ בר"מ α כאשר השונות לא ידועה:

$$\bar{X}-t_{n-1,1-\alpha/2}\cdot\frac{S}{\sqrt{n}}\le\mu\le\bar{X}+t_{n-1,1-\alpha/2}\cdot\frac{S}{\sqrt{n}}.$$

נציב $\alpha=0.05$ ונחשב: $t_{24,0.975}=2.064$.

$$792.4-2.306\cdot\frac{17.125}{\sqrt{9}}\le\mu\le 792.4+2.306\cdot\frac{17.125}{\sqrt{9}}$$

$$779.26\le\mu\le 805.4.$$

$\mu=800$ מוכל ברווח ולכן לא נדחה את השערת האפס.

ד). $n=25$ ולכן $t_{8,0.975}=2.306$. נציב את הנתונים ברווח הסמך ונקבל:

$$792.4-2.064\cdot\frac{17.125}{\sqrt{24}}\le\mu\le 792.4+2.064\cdot\frac{17.125}{\sqrt{24}}$$

$$784.5\le\mu\le 800.3.$$

למרות שרווח הסמך נהיה קטן יותר מכיוון שהמדגם גדול יותר, לא נדחה את H_0.

שאלה 5:

ברצוננו לבדוק את ההשערות: $\begin{matrix} H_0: \mu = 20 \\ H_1: \mu > 20 \end{matrix}$.

נמצא .P.V ונשווה אותו ל-α-ות הנתונות.

במבחן לבדיקת $\begin{matrix} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{matrix}$ כאשר השונות לא ידועה מתקיים: $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$.

מנתוני השאלה: $n=9$, $\bar{X}=22$ ו-$S=3$.

$$P.V. = P(\bar{X} \geq 22 \mid H_0) = P(t_8 \geq \frac{22-20}{3/\sqrt{9}}) = P(t_8 \geq 2) = 1 - P(t_8 < 2)$$

$$0.025 < 1 - P(t_8 < 2) < 0.05 \Leftarrow 0.95 < P(t_8 < 2) < 0.975$$

עבור $\alpha = 0.05$ נדחה את השערת האפס מכיוון ש-$P.V. < 0.05$ אך עבור $\alpha = 0.01$ לא נדחה את H_0 מכיוון ש-$P.V. > 0.01$.

שאלה 6:

הכוונה בסימון היא ש - $t_{\bar{X}} = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$

ולפי הנתון: $\frac{\bar{X}-\mu_0}{S/\sqrt{n}} = -2.611 \Rightarrow \bar{X}-\mu_0 < 0 \Rightarrow \bar{X} < \mu_0$ ולכן מתמקדים בצד השמאלי של המבחן הדו-צדדי.

כמו"כ מכיוון שהמדגם בגודל 5 , $\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_4$

וכשיודעים את התפלגות הסטטיסטי אפשר לחשב את ה- $p.v.$:

$$\frac{p.v.}{2} = P(t_4 \leq -2.611 \mid H_0) = 1 - P(t_4 \leq 2.611)$$
$$\Rightarrow 0.025 < \frac{p.v.}{2} < 0.05 \Rightarrow 0.05 < p.v. < 0.1$$

כלומר תשובה 1 נכונה.

שאלה 7:

$$t_{9,0.95}=1.833$$
$$Z_{0.95}=1.645$$

ערך t גדול יותר מערך ה- Z שבו היה צריך להשתמש, מכאן שידידנו משה סיגמא הרחיק את הנקודה הקריטית ימינה יותר אם המבחן שהשתמש בו היה ימני, או שמאלה יותר אם המבחן שהשתמש בו היה שמאלי. במילים אחרות מבחן שמציבים בו t במקום Z הוא מבחן מחמיר יותר ולכן אם דחה עבור t כמובן שידחה עבור Z ומחינת ההכרעה במבחן יוצא שלאותה תוצאת מבחן היה מגיע משה סיגמא אם היה משתמש ב- Z כפי שהיה צריך.

כלומר תשובה 3 נכונה.

שאלה 8:

לא נתון לנו מאיזה צד של המבחן התקבלה המובהקות כלומר ייתכן שהמדגם נתן "ערך נצפה" שקטן מהנקודה הקריטית השמאלית ואז ברור שלא נדחה במבחן החד צדדי הימני וכמובן יכול להיות שהמדגם נתן "ערך נצפה" גדול מהנקודה הקריטית הימנית ואז דווקא כן נדחה במבחן החד צדדי הימני ומכאן שאין מספיק נתונים בשאלה ותשובה 4 נכונה.

בנוסף תשובה 3 נראית נכונה אך היא לא כי בתרחיש הראשון שהוצג ה- $p.v.$ ייצא קרוב ל- 1 ...

שאלה 9:

במקום להשתמש באחוזון ה- 95 של התפלגות Z היה על החוקר להשתמש באחוזון ה- 95 של התפלגות t_6 מכיוון שהשונות אינה ידועה.

לפי טבלת ההתפלגות: $t_{6,0.95}=1.943$, ערך זה גדול יותר מ- 1.645 ולכן רו"ס ייצא רחב יותר.

ולכן תשובה 2 נכונה.

שאלה 10

$$\overline{X}=23.25$$
$$\hat{S}^2=\frac{4\cdot0.75^2+2\cdot0.25^2+1.75^2+2\cdot3.25^2+2\cdot5.25^2+2.75^2}{11}=8.11$$

ההשערות: $\begin{matrix} H_0:\mu=22 \\ H_1:\mu\neq22 \end{matrix}$. נערוך מבחן דו צדדי ע"פ נקודות קריטיות:

$$C_1 = 22 - t_{11,0.975}\frac{\hat{s}}{\sqrt{n}} = 22 - 2.201\cdot\frac{8.11}{\sqrt{12}} = 16.847$$

$C_2 = 27.153$ ולכן נקבל את השערת האפס.

$$C_1 < \overline{X} = 23.25 < C_2$$

5. להפרש תוחלות של שני מדגמים ב"ת כאשר השונות באוכלוסיה ידועה

שאלה 1

חוקר בדק את הטענה כי שכרן של נשים מועסקות נמוך משכרם של גברים מועסקים. נבדק מדגם מקרי של 20 מועסקות ונמצא כי שכרן הממוצע 1300 שקלים. במדגם מקרי של 40 מועסקים נמצא כי שכרם הממוצע 1600 שקלים. נניח כי סטיית התקן של משכורות ידועה והיא 400 שקלים הן למועסקים והן למועסקות.

א. האם ניתן להסיק כי שכרן של המועסקות נמוך משכרם של המועסקים? בדוק ל- $\alpha = 0.01$.

ב. האם ניתן לדעת ללא חישוב נוסף מהי המסקנה ל- $\alpha = 0.05$?

שאלה 2

חוקר מעוניין לאמוד את ההפרש הממוצע בין משקלם של בנים בלידה לעומת משקלן של בנות בלידה. נבדק מדגם מקרי של 100 יילודים בנים ושל 100 יילודים בנות והתקבלו התוצאות: משקלם הממוצע של הבנים בלידה הוא 3120 גרם ומשקלן הממוצע של הבנות בלידה הוא 2960 גרם. $\sigma = 320$ של אוכלוסיית הבנים ו- $\sigma = 300$ של אוכלוסיית הבנות.

א. מהו רווח הסמך להפרש הממוצע בין משקלם של הבנים בלידה לעומת משקלן של הבנות בלידה, ל- $\alpha = 0.05$?

ב. בדוק האם ניתן להסיק שמשקלם של הבנים בלידה גדול ממשקלן של הבנות בלידה?

פתרונות

שאלה 1

א.
$$H_0 : \mu(male) - \mu(female) = 0$$
$$H_1 : \mu(male) - \mu(female) > 0$$

$$C = 0 + \sqrt{\frac{400^2}{40} + \frac{400^2}{20}} Z_{0.99} = 109.54 \cdot 2.326 = 254.8$$
$$\overline{Xmale} - \overline{Xfemale} = 300 > 254.8$$

ולכן נדחה את ההשערה הראשונית ונקבע ששכר הנשים נמוך יותר.

ב. ברמת בטחון של 1% דחינו את השערת האפס לכן קל וחומר שנדחה אותה ברמת בטחון של 5%.

שאלה 2

א.
$$\mu_{male} - \mu_{female} \in (3120 - 2960) \pm Z_{1-0.025} \sqrt{\frac{320^2}{100} + \frac{300^2}{100}}$$
$$.\ 160 \pm 1.96 \cdot 43.86$$
$$74.0344 \le \mu \le 245.9656$$

ב. נבדוק מה הPval :

$$H_0 : \mu_{male} - \mu_{female} = 0$$
$$H_1 : \mu_{male} - \mu_{female} > 0 \quad Z_{a'} = \frac{|\overline{X}_1 - \overline{X}_2 - \mu_0|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{160}{43.86} = 3.65$$

$$Z_{a'} = \frac{\overline{X}_1 - \overline{X}_2 - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{160}{43.86} = 3.65$$

ולכן לכל רמת בטחון הקטנה מ-99.99% אפשר

$$a' = 0.9999 \qquad Pval < 0.0001$$

לקבוע שמשקל הבנים גדול ממשקל הבנות. (קל לראות זאת גם לפי רווח הסמך)

6. להפרש תוחלות של שני מדגמים ב"ת כאשר השונות באוכלוסיה אינה ידועה

שאלה 1

חוקר ערך ניסוי כדי לבדוק האם יש הבדל בין דיאטת הרזיה מסוג א' ודיאטת הרזיה מסוג ב'. הניסוי נערך על שני מדגמים מקריים בלתי תלויים של נבדקים. למדגם א' ניתנה דיאטה מסוג א' ולמדגם ב' ניתנה דיאטה מסוג ב'. לאחר חודש נשקלו הנבדקים שנית. להלן התוצאות של מספר הקילוגרמים שהפחית כל נבדק:

מדגם א': 3,9,4,6,8.

מדגם ב': 4,1,3,2,1,1.

נניח כי סטיות התקן במדגמים הן סטיות התקן באוכלוסיה.

ג. מהי מסקנת הניסוי ל- $\alpha = 0.05$?

חשב רווח סמך להפרש הממוצע של ההפחתה במשקל בין שתי הדיאטות ברמת סמך 0.99.

שאלה 2

הגובה הממוצע של 12 עצי לימון סוג A היה 13.8 רגל .

ל – 15 עצים מסוג B היה גובה ממוצע של 12.9 רגל .

א. בהנחה שהגבהים מתפלגים נורמלית עם אותה שונות האם יש הבדל מובהק בתוחלת גובה העצים משני הסוגים.

מצא רווח סמך , ברמת סמך 95% להבדל תוחלת גובה העצים משני הסוגים. השווה לסעיף א

פתרונות

שאלה 1

א.

$$Sa^2 = \frac{8^2+6^2+4^2+9^2+3^2}{5} - 6^2 = 5.2 \qquad Sa = 2.28$$

$$Sb^2 = \frac{1^2+1^2+2^2+3^2+1^2+4^2}{6} - 2^2 = 9\frac{1}{3} \qquad Sb = 1.15$$

$$S_{\mu_A-\mu_B}{}^2 = \frac{5.2}{5} + \frac{9.33}{6} = 2.6 \qquad S_{\mu_A-\mu_B} = 1.61$$

$$H_0: \mu_A - \mu_B = 0$$
$$H_1: \mu_A - \mu_B \neq 0$$

כעת, זהו מבחן דו כיווני:

$$C_{1,2}=0\pm Z_{1-0.025}1.61$$
$$C_1=1.96\cdot1.61=3.1556$$
$$C_2=-1.96\cdot1.61=-3.1556$$

אמנם, $\widehat{X_A}-\widehat{X_B}=4>3.1556$ ולכן דיאטה א' יעילה יותר מדיאטה ב'.

ב.

$$P(\ (\overline{X_A}-\overline{X_B})-Z_{1-\frac{a}{2}}\sqrt{\frac{\sigma_1^{\ 2}}{n_1}+\frac{\sigma_2^{\ 2}}{n_2}}\le\mu\le(\overline{X_A}-\overline{X_B})+Z_{1-\frac{a}{2}}\sqrt{\frac{\sigma_1^{\ 2}}{n_1}+\frac{\sigma_2^{\ 2}}{n_2}})=1-a$$

$$P(\ (6-2)-Z_{0.995}1.61\le\mu\le(6-2)+Z_{0.995}1.61)=0.99$$
$$P(4-2.575\cdot1.61\le\mu\le4+2.575\cdot1.61)=0.99$$
$$P(-0.14575\le\mu\le8.14575)=0.99$$

<u>שאלה 2</u>

1. נתון: $n_A=12\ \overline{X}_A=138\ \hat{S}_A=1.2$
$n_B=15\ \overline{X}_B=129\ \hat{S}_B=1.5$

לכן $S_p^2=\frac{11*(1.2)^2}{25}+\frac{14*(1.5)^2}{25}=1.8936$

א.

$$H_o:\mu_A-\mu_B=0$$
$$H_1:\mu_A-\mu_B\neq0$$

$$\frac{PV}{2}=P\left(\overline{X}_A-\overline{X}_B\ge13.8-12.9\right)=P\left(t_{25}\ge\frac{0.9-0}{\sqrt{1.8939*\left(\frac{1}{12}+\frac{1}{15}\right)}}\right)=P\left(t_{25}\ge1.69\right)<0.1$$

$$p\_value<0.2$$

לכל רמת מובהקות גדולה או שווה ל20%, נדחה את H_0 ונסיק כי יש הבדל מובהק בין שני סוגי עצי הלימון. לרוב לוקחים ר"מ 0.05 ולכן לרוב לא נדחה את H_0 - כי $0.1 < p.v. < 0.2$ ונסיק כי אין הבדל בתוחלות גובה העצים.

ב. $0.9 \pm \sqrt{1.8936 * \left(\frac{1}{12} + \frac{1}{15}\right)} * t_{25,0.975} = 0.9 \pm 1.1 \Rightarrow -0.2 \le \mu \le 2.0$

תיקון קל : $-0.2 \le \mu_A - \mu_B \le 2.0$

מתוך הרווח לא ניתן להסיק כי תוחלת הגובה של אחד הסוגים גבוהה יותר. כי 0 נכלל ברו"ס.

מתאים למסקנה בסעיף א.

רווחי סמך ובדיקת השערות

7. להפרש תוחלות של שני מדגמים תלויים (מזווגים)

שאלה 1:

נלקח מדגם מקרי של 10 מעשנים אשר השתתפו בתוכנית לגמילה מעישון באמצאות היפנוזה. להלן הנתונים על מספר הסיגריות היומי שעישן כל נבדק לפני ההיפנוזה וחודש לאחר ההיפנוזה.

לפני	25	30	35	25	15	40	20	30	30	40
אחרי	25	10	5	10	20	15	5	10	20	20

ידוע כי מספר הסיגריות מתפלג נורמלית בכל אחת מהאוכלוסיות.

א. האם ההיפנוזה הפחיתה את כמות העישון? בדוק ברמת מובהקות $\alpha = 0.025$

ב. חשב רווח בר-סמך לתוחלת כמות ההפחתה בעישון בעקבות ההיפנוזה ברמת סמך 0.95.

שאלה 2

במספר חנויות בדקו מחירים של סלסילות שי לחגים עם תכולה דומה, והשוו למחירי רכיבי התכולה כשהם נקנים בנפרד.

חנות	א	ב	ג	ד	ה
מחיר סלסלה	109	99	99	119	109
מחיר התכולה בנפרד	82	75	70	85	84

א. מהי רמת המובהקות המינימאלית עבורה ניתן להסיק כי מחיר הסלסילה גדול באופן מובהק ממחיר מרכיביה ביותר מ-22 ₪.

ב. מצא רווח סמך של 98% לתוחלת ההבדלים במחירים בין סלסלה מוכנה ומרכיביה.

שאלה 3

ציונים במבחן מסוים מתפלגים נורמלית. התקבלו נתונים לגבי ציונים של 6 סטודנטים שעשו שני מועדים: מועד א : 51,40,38,59,55,46; מועד ב: 91,69,70,81,83,79.

בנה רווח סמך להפרש התוחלות עם רמת סמך 90%.

פתרונות

שאלה 1:

א. נמצא את הפרש מספר הסיגריות:

לפני	25	30	35	25	15	40	20	30	30	40
אחרי	25	10	5	10	20	15	5	10	20	20
הפרש (d)	0	20	30	15	-5	25	15	20	10	20

ברצוננו לבדוק האם יש ההיפנוזה הפחיתה את צריכת הסיגריות:

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 > 0$$

נמצא P value: $\bar{d} = 15$, $S_d = 10.80$

$$p\_value : P\left(\bar{d} > 15 \mid H_0\right) = P\left(t_9 > \frac{15-0}{10.80/\sqrt{10}}\right) = P\left(t_9 > 4.39\right) \Rightarrow 0.0005 < p\_v < 0.001$$

לכן עבור רמת מובהקות של 0.025 נדחה את השערת האפס.

ב. נחשב רווח סמך כאשר השונות אינה ידועה: $\bar{d} - t_{n-1,1-\alpha/2} \cdot \frac{S_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{n-1,1-\alpha/2} \cdot \frac{S_d}{\sqrt{n}}$

$$t_{9,0.975} = 2.262$$

נציב את הנתונים ונקבל: $15 - 2.262 \cdot \frac{10.8}{\sqrt{10}} < \mu_d < 15 + 2.262 \cdot \frac{10.8}{\sqrt{10}}$

$$7.27 < \mu_d < 22.72$$

מכיוון שהתוחלת עפ"י השערת האפס (0) אינה נמצאת ברווח נדחה את השערת האפס.

שאלה 2

ב. נמצא את הפרש המחירים:

חנות	א	ב	ג	ד	ה
מחיר סלסלה	109	99	99	119	109
מחיר התכולה בנפרד	82	75	70	85	84
הפרש (d)	27	24	29	34	25

$\overline{d} = 27.8$

$S_d = 3.96$

$H_0 : \mu_d = 22$

$H_1 : \mu_d > 22$

$$p.v. = P(\overline{d} \geq 27.8 \mid H_0) = P(t_4 \geq \frac{27.8 - 22}{3.96/\sqrt{5}}) = P(t_4 \geq 3.28) = 1 - P(t_4 \leq 3.28)$$

$\Rightarrow 0.01 < p.v. < 0.025$

כלומר רמת המובהקות המינימלית היא 0.025

ב. רווח סמך של 98% ל- μ_d : $\Leftarrow \alpha = 0.02$ $\quad t_{n-1,1-\frac{\alpha}{2}} = t_{4,0.99} = 3.75$

$$27.8 - 3.75 \cdot \frac{3.96}{\sqrt{5}} \leq \mu_d \leq 27.8 + 3.75 \cdot \frac{3.96}{\sqrt{5}}$$

$\Leftarrow 21.16 \leq \mu_d \leq 34.44$

ג. 22 הוא בתוך רו"ס ולכן לא דוחים את השערת האפס בר"מ 0.02 או בר"מ 0.01 במבחן חד-צדדי כמו בסעיף א' ולכן יש התאמה בין ההכרעה בסעיף א' (לא לדחות את השערת האפס עבור ר"מ 0.01)לבין רו"ס.

<u>שאלה 3</u>

$$D : x_b - x_a : -33, -28, -22, -32, -29, -40.$$

$$\overline{D} = -30\frac{2}{3}$$

$$\overline{s_d^{\,2}} = \frac{\sum (D_i - \overline{D})^2}{n-1} = 35.86$$

$$t_{5,0.95} = 2.445 \quad 30\frac{2}{3} - \frac{5.989}{\sqrt{6}} * t_{5,0.95} < \mu_d < 30\frac{2}{3} + \frac{5.989}{\sqrt{6}} * t_{5,0.95}$$

$$25.72 < \mu_d < 35.6$$

8. לפרופורציה במדגם אחד

שאלה 1

בכדי לאמוד את אחוז הגירושין בארה"ב נלקח מדגם מקרי של 625 איש ונמצא ש-125 מהם גרושים.

יא. בנו רו"ס לאחוז הגירושין ברמת ביטחון של 95%.

יב. איך ישתנה אורך רו"ס עבור רמת ביטחון 90% ? 99% ? (ענו ללא חישוב נוסף)

יג. אם נגדיל את המדגם פי 4 כיצד יושפע אורך רו"ס?

יד. מהו גודל המדגם הדרוש כדי להקטין את אורך הרווח פי 2 ?

טו. אמדו את אחוז האנשים חסרי השכלה גבוהה באוכלוסיה ברמת סמך 95%.

שאלה 2

במאפיה נשרפים 60% מכיכרות הלחם בכל יום. לאחר קורס הכשרה מיוחד לאופים נלקח מדגם של 100 כיכרות ונמצא כי 53 כיכרות נשרפו. האם הקורס עזר? בדקו עבור $\alpha = 0.05$.

למרות שרווח הסמך נהיה צר יותר מכיוון שהמדגם גדול יותר, לא נדחה את H_0.

שאלה 3

ידוע כי 80% מהאוכלוסייה עובדים. בעקבות תוכנית ויסקונסין יש הטוענים כי אחוז העובדים באוכלוסיה עלה ויש הטוענים כי אחוז העובדים לא השתנה. כדי לבדוק זאת נלקח מדגם מקרי של 400 איש ונמצא כי 336 מתוכם עובדים. מי צודק? האם ייתכן כי שני הצדדים צודקים?

פתרונות

שאלה 1

ח. נתון: $n = 625$, $\hat{p} = \frac{125}{625} = 0.2$

נוסחא לרו"ס לפרופורציה: $\hat{p} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} \le p \le \hat{p} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$

95% $\Leftarrow$ $Z_{1-\frac{\alpha}{2}} = 1.96$

$$0.17 \le p \le 0.23 \;\Leftarrow\; p \in 0.2 \pm 1.96\sqrt{\frac{0.2\cdot 0.8}{625}}$$

ט. אם נקטין את רמת הביטחון ל-90% גם רווח הסמך יקטן, ולחילופין אם נגדיל את רמת הביטחון ל-99% רווח הסמך יגדל.

י. מכיוון שהפרופורציה עצמה תלויה בגודל n ולא רק רווח הסמך, אם נגדיל את גודל המדגם לא ברור כיצד רווח הסמך ישתנה. הדבר שונה מרווח הסמך עבור תוחלת שבו אם נגדיל את n רווח הסמך יקטן.

יא. מכיוון שכפי שציינו לעיל שינוי גודל המדגם משפיע גם על הפרופורציה, ואין לנו שום מידע לגבי הפרופורציה באוכלוסיה נציב $\hat{p} = 0.5$ עבורו הביטוי $\hat{p}(1-\hat{p})$ מקבל ערך מקסימאלי כך שנקבל חסם עליון ל-n.

כדי להקטין את אורך הרווח פי 2 אפשר פשוט להקטין את הסטייה מ- $\hat{p}$ פי 2 , כלומר:

$$1.96\cdot\sqrt{\frac{0.5\cdot 0.5}{n}} = 0.5\cdot 0.03$$

$\sqrt{\frac{0.5\cdot 0.5}{n}} = \frac{0.015}{1.96} \Rightarrow n = 4268.4$. נעגל ונקבל כי יש לקחת מדגם בגודל של 4269 כדי להקטין את הרווח פי 2 .

יב. נתון: $n = 625$, $\hat{p} = \frac{500}{625} = 0.8$

נוסחא לרו"ס לפרופורציה: $\hat{p} - Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} \le p \le \hat{p} + Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$

95% $\Rightarrow Z_{1-\frac{\alpha}{2}} = 1.96$

$p \in 0.8 \pm 1.96\sqrt{\frac{0.2\cdot 0.8}{625}} \Rightarrow 0.77 \le p \le 0.83$. נשים לב כי רווח סמך זה הוא המשלים של רווח הסמך מסעיף א' מכיוון הוא רווח סמך להסתברות המשלימה, וכן מכיוון שהסטייה מהפרופורציה שווה.

<u>שאלה 2</u>

ברצוננו לבדוק את ההשערות : $H_0: p \ge 0.6$, $H_1: p < 0.6$.

נשתמש במבחן דו צדדי : דחה H_0 אם $\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$. כאשר $\hat{p} = \frac{53}{100} = 0.53$.

נבדוק האם $\hat{p} < 0.6 - Z_{0.95} \cdot \sqrt{\frac{0.6 \cdot 0.4}{100}} = 0.6 - 1.645 \cdot 0.049 = 0.52$.

אי השוויון לא מתקיים ולכן לא נדחה את השערת האפס ונסיק כי הקורס לא עזר.

<u>שאלה 3</u>

ברצוננו לבדוק את ההשערות: $H_0: p \le 0.8$, $H_1: p > 0.8$.

$$\hat{p} \sim N(p, \frac{p(1-p)}{n}) \quad p = 0.8 \ , \ \hat{p} = \frac{336}{400} = 0.84$$

נחשב .P.V:

$$P.V. = P(\hat{p} \ge 0.84 \mid H_0) = P(\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \ge \frac{0.84 - 0.8}{0.02}) = P(Z \ge 2) = 1 - P(Z < 2) = 0.0228$$

לכן אם ר"מ היא 0.05 נדחה את השערת האפס ואם ר"מ היא 0.01 לא נדחה. ז"א ששני הצדדים צודקים (תלוי איזו רמת מובהקות בוחרים).

מבחן χ^2 לאי תלות ורגרסיה

9. מבחן χ^2 לאי תלות

שאלה 1:

נתונה הטבלה הבאה :

סה"כ	מדעי הרוח	מדעי הטבע	מדעי החברה	
153	50	47	56	א'
27	5	14	8	ב'
180	55	61	64	סה"כ

X – טיפוס אישיות

Y – תחום לימוד

בדוק האם יש תלות בין טיפוס אישיות X , אליו משתייך נבדק מסוים לבין תחום הלימוד שלו ברמת מובהקות 0.05 .

שאלה 2

אוניברסיטה פרסמה נתונים על הרישום לפקולטות שונות לפי מין :

סה"כ	משפטים	מדעי הטבע	מדעי החברה	מדעי הרוח	
1410	200	210	700	300	בנים
1590	100	90	700	700	בנות
3000	300	300	1400	1000	סה"כ

האם יש קשר בין מין לפקולטה בה נוטים ללמוד בר"מ 0.05 ?

שאלה 3:

מדגם אקראי של 100 חברי איגוד מקצועי נתבקשו להתייחס לשתי שאלות:
(1). האם אתה מרוצה מהסטאטוס הפיננסי שלך?
(2). האם אתה מסכים עם המדיניות הכלכלית של הממשלה?

התוצאות מובאות בטבלה הבאה:

שאלה מספר 2		שאלה מספר 1		
		כן	לא	סה"כ
	כן	22	48	70
	לא	12	18	30
	סה"כ	34	66	100

מבחן חי בריבוע נערך כדי לבחון את השערת האפס לפיה התשובה לשאלה (1) אינה תלויה בתשובה לשאלה (2) ברמת מובהקות של 5%. הערך הצפוי עבור התא (כן,כן) והערך הקריטי של חי בריבוע במבחן זה יהיו בהתאמה:

1. 23.8 ו- 7.81
2. 23.8 ו- 3.84
3. 23.8 ו- 1.96
4. 10.2 ו- 3.84

<u>פתרונות</u>

<u>שאלה 1:</u>

סה"כ	מדעי הרוח	מדעי הטבע	מדעי החברה	
153	50	47	56	א'
27	5	14	8	ב'
180	55	61	64	סה"כ

נחשב את הערכים הצפויים (Expected values):

מדעי הרוח	מדעי הטבע	מדעי החברה	
$\frac{55\cdot153}{180}=46.75$	$\frac{61\cdot153}{180}=51.85$	$\frac{64\cdot153}{180}=54.4$	א'
$\frac{55\cdot27}{180}=8.25$	$\frac{61\cdot27}{180}=9.15$	$\frac{64\cdot27}{180}=9.6$	ב'

על מנת לבדוק את השערת האפס על פיה אין תלות בין הנבדקים נחשב את הסטטסטי:

$\chi^2=\sum_{i=1}^{R\times C}\frac{\left(O_i-E_i\right)^2}{O_i}$ כאשר R מספר השורות ו-C מספר העמודות.

$$\chi^2=\frac{(56-54.4)^2}{56}+\frac{(47-51.85)^2}{47}+\frac{(50-46.75)^2}{50}+\frac{(8-9.6)^2}{8}+\frac{(14-9.15)^2}{14}+\frac{(5-8.25)^2}{5}=4.87$$

נמצא את הנקודה הקריטית: $\chi^2_{(R-1)\cdot(C-1)}=\chi^2_{(2-1)\cdot(3-1)}=\chi^2_2=5.99$ (עבור רמת מובהקות 0.05)

מכיוון ש-הערך המחושב (4.87) קטן מ- 5.99 לא נדחה את השערת האפס ונסיק שאין תלות בין המשתנים.

<u>שאלה 2:</u>

סה"כ	משפטים	מדעי הטבע	מדעי החברה	מדעי הרוח	
1410	200	210	700	300	בנים
1590	100	90	700	700	בנות
3000	300	300	1400	1000	סה"כ

נחשב את הערכים הצפויים (Expected values):

משפטים	מדעי הטבע	מדעי החברה	מדעי הרוח	
$\frac{300\cdot1410}{3000}=141$	$\frac{300\cdot1410}{3000}=141$	$\frac{1400\cdot1410}{3000}=658$	$\frac{1000\cdot1410}{3000}=470$	בנים
$\frac{300\cdot1590}{3000}=159$	$\frac{300\cdot1590}{3000}=159$	$\frac{1400\cdot1590}{3000}=742$	$\frac{1000\cdot1590}{3000}=530$	בנות

על מנת לבדוק את השערת האפס על פיה אין תלות בין הנבדקים נחשב את הסטטיסטי:

$\chi^2=\sum_{i=1}^{R\times C}\frac{\left(O_i-E_i\right)^2}{O_i}$ כאשר R מספר השורות ו-C מספר העמודות.

נמצא את הנקודה הקריטית: $\chi^2_{(R-1)\cdot(C-1)}=\chi^2_{(2-1)\cdot(4-1)}=\chi^2_3=7.82$ (עבור רמת מובהקות 0.05)

מכיוון ש-הערך המחושב (270.44) גדול מ-7.82 נדחה את השערת האפס ונסיק שיש תלות בין המשתנים.

<u>שאלה 3:</u>

נחשב את הערכים הצפויים (Expected values):

		שאלה 1	
		כן	**לא**
שאלה 2	**כן**	$\frac{34\cdot70}{100}=23.8$	$\frac{66\cdot70}{100}=46.2$
	לא	$\frac{34\cdot30}{100}=10.2$	$\frac{66\cdot30}{100}=19.8$

על מנת לבדוק את השערת האפס על פיה אין תלות בין הנבדקים נחשב את הסטטסטי:

נמצא את הנקודה הקריטית: $\chi^2_{(R-1)\cdot(C-1)}=\chi^2_{(2-1)\cdot(2-1)}=\chi^2_1=3.841$ (עבור רמת מובהקות 0.05)

לכן התשובה הנכונה היא 2.

10. רגרסיה

שאלה 1

לפניך נתונים על גובה באינצ'ים של צמחי קקטוס שהורכבו בתנאי סביבה מבוקרים:

X=מספר שבועות אחרי ההרכבה	1	2	4	5	6	8
Y=גובה באינצ'ים	2.0	2.4	5.1	7.3	9.4	18.3

א. האם ניתן לדבר על קשר ליניארי בין גובה הצמח למספר השבועות שעברו מאז ההרכבה?

ב. מהו גובה הצמח הצפוי לאחר 7 שבועות?

ג. באיזה שבוע יגיע הגובה ל-30 אינץ'?

שאלה 2

לפניך נתונים על שאריות כלור בבריכת שחייה, מספר שעות לאחר הכנסתו לבריכה. (שני המשתנים בטבלה מקבלים ערכים ממשיים אי-שליליים):

X=מספר שעות	2	4	6	8	10	12
Y=שאריות כלור	1.8	1.5	1.4	1.1	1.1	0.9

(כמות שאריות הכלור הן ביחס למיליון יחידות)

א. חשב את מקדם המתאם בין מספר השעות שעברו מאז הכנסת הכלור, וכמות הכלור שנותרה בבריכה.

ב. מהי כמות הכלור הצפויה אחרי 15 שעות?

ג. אחרי כמה שעות תרד כמות הכלור ל-0.1?

שאלה 3

נתון:

$$\sum y_i = 50 \qquad \sum x_i = 100$$
$$\sum y_i^2 = 134.84 \qquad \sum x_i^2 = 509.12$$
$$n = 20 \qquad \sum x_i y_i = 257.66$$

התאם קו רגרסיה של y על x.

שאלה 4

סטודנט ערך חישובים לקביעת הקשר בין שני משתנים x ו- y . הוא מצא כי משוואת הרגרסיה היא $\hat{Y}_i = 0.38x_i = 0.42$ וכי מקדם המתאם $r = -0.23$. לאור ממצאיו ניתן לקבוע כי :

1. קיים קשר שלילי חזק בין המשתנים .
2. קיים קשר שלילי חלש בין המשתנים.
3. קיים קשר חיובי חזק בין המשתנים.
4. הסטודנט טעה בחישוביו.

פתרונות

שאלה 1

א. נוסחת מקדם המתאם: $r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$.

למציאת המתאם נבנה טבלת חישובים:

x	y	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
1	2.0	-3.3	-5.4	17.82	10.89	29.16
2	2.4	-2.3	-5	11.5	5.29	25
4	5.1	-0.3	-2.3	0.69	0.09	5.29
5	7.3	0.7	-0.1	-0.07	0.49	0.01
6	9.4	1.7	2	3.4	2.89	4
8	18.3	3.7	10.9	40.33	13.69	118.81
26	**44.5**			**73.67**	**33.34**	**182.27**

כאשר: $\bar{x}=\frac{26}{6}=4.3$, $\bar{y}=\frac{44.5}{6}=7.4$.

נציב את הנתונים ונקבל: $r=\frac{73.64}{\sqrt{33.34\cdot182.27}}=0.945$.

ז"א שקיים קשר ליניארי חזק מאוד בין הזמן שחלף מאז ההרכבה לגובה הצמח.

ב. על מנת לנבא את גובה הצמח נבנה קו רגרסיה: $\hat{Y}_i=a+bx_i$. כאשר:

$$b=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{73.67}{33.34}=2.2\ ,\ a=\bar{y}-b\bar{x}=7.4-2.2\cdot3.4=-2.06.$$

$\Leftarrow$ $\hat{Y}_i=-2.06+2.2x_i$.

על מנת לנבא את גובה הצמח לאחר 7 שבועות נציב במשוואת הרגרסיה $x_i=7$.

$\hat{Y}=-2.06+2.2\cdot7=13.34$. ז"א שגובהו הצפוי של הצמח לאחר 7 שבועות יהיה 13.34 אינצ'ים.

ג. על מנת לנבא את מספר השבועות נבנה קו רגרסיה: $\hat{X}_i = a + by_i$. כאשר:

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \frac{73.67}{182.27} = 0.4 \text{ , } a = \overline{x} - b\overline{y} = 4.3 - 0.4 \cdot 7.4 = 1.34 \text{ .}$$

$$\Leftarrow \hat{X}_i = 1.34 + 0.4y_i \text{ .}$$

על מנת לנבא את מספר השבועות עד שצמח יגיע לגובה של 30 אינצ'ים נציב במשוואת הרגרסיה $y_i = 30$. $\hat{X} = 1.34 + 0.4 \cdot 30 = 13.34$. ז"א שגובהו הצפוי של הצמח לאחר 13.34שבועות יהיה 30 אינצ'ים.

שאלה 2

א. לחישוב מקדם המתאם (r), נבנה טבלת חישובים:

x	y	$(x_i - \overline{x})$	$(y_i - \overline{y})$	$(x_i - \overline{x})(y_i - \overline{y})$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$
2	1.8	-5	0.5	-2.5	25	0.25
4	1.5	-3	0.2	-0.6	9	0.04
6	1.4	-1	0.1	-0.1	1	0.01
8	1.1	1	-0.2	-0.2	1	0.04
10	1.1	3	-0.2	-0.6	9	0.04
12	0.9	5	-0.4	-2	25	0.16
42	**7.8**			**-6**	**70**	**0.54**

כאשר חישבנו: $\overline{x} = \frac{42}{6} = 7$, $\overline{y} = \frac{7.8}{6} = 1.3$.

נציב את הנתונים ונקבל: $r = \frac{-6}{\sqrt{70 \cdot 0.54}} = -0.976$

נסיק כי קיים קשר ליניארי חזק מאוד בין X ל-Y.

ב. על מנת לנבא את כמות הכלור נבנה קו רגרסיה: $\hat{Y}_i = a + bx_i$. כאשר:

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{-6}{70} = -0.086 \text{ , } a = \overline{y} - b\overline{x} = 1.3 - 7 \cdot 0.086 = 1.902 \text{ .}$$

$$\Leftarrow \hat{Y}_i = 1.902 - 0.086x_i$$

נציב במשוואת הרגרסיה $x_i = 15$: $\hat{Y} = 1.902 - 0.086 \cdot 15 = \boxed{0.612}$.

ז"א שכמות הכלור הצפויה בבריכה אחרי 15 שעות היא 0.612.

ב. על מנת לנבא את מספר השעות נבנה קו רגרסיה: $\hat{X}_i = a + by_i$. כאשר:

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \frac{-6}{0.54} = -11.11 \;,\; a = \overline{x} - b\overline{y} = 7 + 11.11 \cdot 1.3 = 21.443 \;.$$

$\Leftarrow$ $\hat{X}_i = 21.443 + -11.11 y_i$.

על מנת לנבא את מספר השעות עד שכמות הכלור תרד ל-0.1 נציב במשוואת הרגרסיה

$y_i = 0.1$. $\hat{X} = 21.443 - 11.11 \cdot 0.1 = 20.332$. ז"א שמספר השעות הצפוי עד אשר

כמות הכלור תרד ל-0.1 הוא 20.332.

שאלה 3

נתאים את קו הרגרסיה:

$$\overline{x} = \frac{100}{20} = 5 \;,\; \overline{y} = \frac{50}{20} = 2.5 \;,\qquad a = \frac{257.66 - \frac{50 \cdot 100}{20}}{509.12 - \frac{10000}{20}} = \frac{7.66}{9.12} \approx 0.84$$

$$b = 2.5 - 0.84 \cdot 5 = -1.7 \quad \Rightarrow \quad \boxed{y_i = -1.7 + 0.84 x_i}$$

שאלה 4

לא ייתכן כי המתאם יהיה שלילי ושיפוע ישר הרגרסיה יהיה חיובי.
לכן התשובה הנכונה היא 4, הסטודנט טעה בחישוביו.