פתרון – תרגיל בית 2 – מבוא להסתברות וסטטיסטיקה

שאלה 1

חוקר א' בדק את ריכוז האבץ הממוצע במימי אגם. ריכוז האבץ (בגרם לליטר) נמדד במדגם שכלל 16 דגימות מים והתקבלו הנתונים: $\sum_{i=1}^{16} x_i^2 = 108.265$ $\sum_{i=1}^{16} x_i = 41.6$. ממדידות העבר, ידוע לחוקר א' שסטיית התקן של מדגמי אבץ באגם זה היא 0.3 גרם לליטר. מניחים שריכוז האבץ מתפלג נורמלית.

א. מצא רווח סמך, ברמת מובהקות 1%, לתוחלת רמת האבץ באותו אגם יקבל החוקר א'.

ב. חוקר א' מבקש לדעת מה גודל המדגם הנדרש, על מנת שברמת בטחון 95%, הסטייה באמידת התוחלת לא תעלה על 0.05?

ג. חוקר ב' איננו יודע את הנתון אודות סטיית התקן של ריכוז האבץ באגם. גם הוא מבקש לחשב רווח סמך לתוחלת רמת האבץ, ברמת מובהקות 1%, ולהשתמש בנתוני הדגימות של חוקר א'. מה יהיה רווח הסמך שיקבל?

ד. אילו חוקר ב' היה מקבל שסטיית התקן המדגמית היא 0.3 גרם לליטר. האם רווח הסמך שיחשב יהיה גדול או קטן מרווח הסמך של <u>חוקר א'</u> שחושב בסעיף א'? הסבר בקצרה (עד 5 שורות). אין צורך לחשב את רווח הסמך.

<u>פתרון:</u>

מאחר והמדגם גדול $(n \geq 30)$, נשתמש בטבלה הנורמלית.

א. רווח סמך ברמת מובהקות 1%:

מהטבלה הנורמלית $Z_{1-\frac{\alpha}{2}} = Z_{0.995} = 2.575$

הממוצע: $\bar{x} = \frac{41.6}{16} = 2.6$

$$2.6 - 2.575 \cdot \frac{0.3}{\sqrt{16}} < \mu < 2.6 + 2.575 \cdot \frac{0.3}{\sqrt{16}}$$

$$\boxed{2.41 < \mu < 2.79}$$

ב. נסמן את הסטייה המקסימלית: $C = 0.05$. מנוסחת רווח הסמך נחלץ את גודל המדגם-

$$Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = C \quad \Rightarrow n = \left(\frac{Z_{1-\frac{\alpha}{2}} \sigma}{C} \right)^2 \qquad \Rightarrow n = \left(\frac{1.96 \cdot 0.3}{0.05} \right)^2 = 138.3$$

לפיכך נדרש מדגם בגודל 139 לפחות.

ג. רווח סמך ברמת מובהקות 1%, כאשר השונות אינה ידועה:

הממוצע חושב בסעיף א': $\bar{x} = 2.6$

סטיית התקן המדגמית: $S = \sqrt{\frac{\sum_{i=1}^{16} x_i^2 - n \cdot \bar{x}^2}{n-1}} = \sqrt{\frac{108.265 - 16 \cdot 2.6^2}{15}} = 0.08367$

מטבלת t: $t_{(n-1);1-\frac{\alpha}{2}} = t_{(15);0.995} = 2.947$

רווח הסמך המבוקש-

$$2.6-2.947\cdot\frac{0.08366}{\sqrt{16}}<\mu<2.6+2.947\cdot\frac{0.08366}{\sqrt{16}}$$

$$\boxed{2.54<\mu<2.66}$$

ד. מכיוון שההבדל היחיד בין רווחי הסמך הם שחוקר א' משתמש בהתפלגות נורמלית (Z) ואילו חוקר ב' בהתפלגות t, הרי שרווח הסמך של חוקר ב' צפוי להיות גדול יותר, מאחר וערכי התפלגות t (בדרגות החופש שבשאלה = 15) הם גדולים יותר עבור אותה $1-\alpha/2$. בכתיב מתמטי: $Z_{1-\frac{\alpha}{2}}<t_{(15);1-\frac{\alpha}{2}}$

<u>שאלה 2</u>

מטילים מטבע 1,000 פעמים. יצא "H" 543 פעמים. רוצים לבדוק את הטענה שהמטבע מוטה לטובת "H", ברמת בטחון 99.5%. בצע את הבדיקה והסק המסקנות (נסח במפורש את השערות המבחן המתאימות).

<u>פתרון:</u>

<u>דרך I</u>

השערות המבחן החד-צדדי ימני לפרופורציה:

$$\begin{cases}H_0: p=0.5\\ H_1:\mu_1>0.5\end{cases}$$

הערך הקריטי של המבחן:

$$K_1=p_0+Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{p_0(1-p_0)}{n}}=0.5+2.575\sqrt{\frac{0.5\cdot0.5}{1000}}=0.5+0.041=0.541$$

כעת, $\bar{x}=543>K=541$ ולכן נדחה את H_0 ברמת מובהקות 0.005.

המסקנה מהמבחן: המטבע מוטה לטובת "H".

<u>דרך II</u>

נשתמש במבחן חי-בריבוע לטיב התאמה.

השערות המבחן: המשתנה מתאים להתפלגות: H_0

המשתנה לא מתאים להתפלגות: H_1

נשווה להנחה שצדדי במטבע שווי התפלגות, ואז: $E_i=p_i\cdot n=\frac{1}{2}\cdot1000=500\quad(i=1,2)$.

סטטיסטי המבחן:

$$\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}=\frac{(543-500)^2}{500}+\frac{(457-500)^2}{500}=7.396$$

הערך המתאים מטבלת חי-בריבוע: $\chi^2_{(1);0.995}=7.879$

מאחר ו- $\chi^2_{(1);0.995}=7.879$ $\chi^2=7.396\not>$ <= <u>לא</u> נדחה את H_0 ברמת מובהקות 0.005.

המסקנה מהמבחן: המטבע הוגן.

<u>הערה:</u> התוצאות הנ"ל אינן סותרות זו את זו. מבחן Z רגיש יותר ממבחן חי-בריבוע כאשר ישנן תצפיות מרובות (ובעוד מקרים). **<u>במבחן:</u>** אם לא נתבקשנו לבצע מבחן ספציפי, אפשר לבחור כל אחד מהשניים – חשובה הרבה יותר הדרך, פירוט החישובים והסקת המסקנות מהמבחן, מאשר התוצאה הסופית בלבד.

<u>שאלה 3</u>

חוקר ערך ניסוי ובדק את גובהם הממוצע של 16 ילדים. הוא סיכם את כל הגבהים וקיבל 22.79 מ'.
ידוע שסטיית התקן באוכלוסייה היא 0.1 מ' ושהאוכלוסייה מתפלגת נורמלית.
אם החוקר קיבל את רווח הסמך הבא: $1.3724 \le \mu \le 1.4765$, מהי רמת הביטחון שבה חושב רווח הסמך?

<u>פתרון:</u>

נסמן ב- X את הגובה , $X \sim N(\mu, 0.1^2)$, נתון: $n = 16$, $\sum_{i=1}^{16} x_i = 22.79$

הממוצע: $\overline{X} = \frac{22.79}{16} = 1.4244$. <u>אבל</u>, מכיוון שאנו מעוניינים רק בגודל רוול הסמך, אין לנו צורך בו.

רו"ס ברמת בטחון של $1-\alpha$ לתוחלת כאשר השונות ידועה הוא:

$$1.4244 - Z_{1-\frac{\alpha}{2}} \cdot \frac{0.1}{4} \le \mu \le 1.4244 + Z_{1-\frac{\alpha}{2}} \cdot \frac{0.1}{4}.$$

נתון ש- $1.3724 \le \mu \le 1.4765$.

כלומר אורך רו"ס הוא $L = 1.4765 - 1.3724 = 0.1041$.

מכאן- $\frac{L}{2} = Z_{1-\frac{\alpha}{2}} \cdot \frac{0.1}{\sqrt{16}} = \frac{0.1041}{2} \Leftarrow Z_{1-\frac{\alpha}{2}} = 2.08$

$$\Leftarrow 1 - \frac{\alpha}{2} = 0.9812 \Leftarrow \boxed{\alpha = 0.0376}.$$

לכן רמת הביטחון היא 96.24%.

הערה: מי שרוצה להשתמש בממוצע יכול לפתור כך- $1.4244 + Z_{1-\frac{\alpha}{2}} \cdot \frac{0.1}{4} = 1.4765 \Leftarrow Z_{1-\frac{\alpha}{2}} = 2.08$

וכמובן נקבל את אותה אלפא שקיבלנו לעיל.

<u>שאלה 4</u>

במפעל העובד ב-3 משמרות (יום, ערב, לילה) נבדקו כל המוצרים שיוצרו במשך שבוע. להלן מספר המוצרים הפגומים והתקינים בכל אחת מהמשמרות:

משמרת / איכות המוצר	יום	ערב	לילה	סה"כ
פגומים	30	55	75	160
תקינים	70	35	35	140
סה"כ	100	90	110	300

בדוק ברמת מובהקות של 2.5% האם קיים קשר באיכות המוצרים ובין המשמרות השונות בה יוצרו?

<u>*פתרון:*</u>

נבנה את טבלת השכיחות הצפויה $E_{i,j}$:

סה"כ	לילה	ערב	יום	משמרת / איכות המוצר
160	$\frac{160\cdot 110}{300}=58\frac{2}{3}$	$\frac{160\cdot 90}{300}=48$	$\frac{160\cdot 100}{300}=53\frac{1}{3}$	פגומים
140	$\frac{140\cdot 110}{300}=51\frac{1}{3}$	$\frac{140\cdot 90}{300}=42$	$\frac{140\cdot 100}{300}=46\frac{2}{3}$	תקינים
300	110	90	100	

$$\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}}=$$

$$=\frac{\left(30-53\frac{1}{3}\right)^2}{53\frac{1}{3}}+\frac{\left(70-46\frac{2}{3}\right)^2}{46\frac{2}{3}}+\frac{(55-48)^2}{48}+\frac{(35-42)^2}{42}+$$

$$\frac{\left(75-58\frac{2}{3}\right)^2}{58\frac{2}{3}}+\frac{\left(35-51\frac{1}{3}\right)^2}{51\frac{1}{3}}=33.8068$$

מטבלת חי-בריבוע: $\chi^2_{(2-1)(3-1):\,0.975}=7.378$

איזור הדחייה הוא $\chi^2>\chi^2_{(2-1)(3-1):\,1-\alpha}$, מאחר שמתקיים $33.8068>7.378$ דוחים את H_0 ,
כלומר ניתן להסיק כי יש קשר בין המשמרות השונות לאיכות המוצרים.

<u>שאלה 5</u>

הזמן הדרוש לביצוע 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 36 מנותחים, והתקבל ממוצע של $\bar{X}=6.8$ שעות.

א. בדוק את ההשערה ששיטת הניתוח המוצעת אמנם מקצרת את זמן הניתוח, ברמת מובהקות 0.01. (נסח את ההשערות והסק המסקנה).

ב. חשב את ערך P-value. בעזרתו שיטת הניתוח המוצעת אמנם מקצרת את זמן הניתוח, ברמת מובהקות 0.05.

<u>*פתרון:*</u>

א).

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

מבחן ההשערות: $\begin{cases}H_0:\mu=8\\ H_1:\mu<8\end{cases}$

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

בשאלה שלנו: $\alpha = 0.01 \Leftarrow Z_{1-\alpha} = Z_{0.99} = 2.325$.

הערך הקריטי של המבחן: $K = 8 - 2.325 \cdot \frac{4}{\sqrt{36}} = 6.45$.

מכיון ש- $\bar{X} = 6.8 \not< K = 6.45$ <= לא נדחה את H_0.

<u>המסקנה</u>: שיטת הטיפול החדשה אינה מקצרת את זמן הניתוח.

ב). נחשב P-value:

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

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

עבור: $P.V. = 0.0359 < 0.05$ ולכן <u>נדחה</u> את H_0.

<u>שאלה 6</u>

במפעל שתי מכונות: A ו- B. 10% מתוצרת המפעל מיוצרת במכונה A ו- 90% במכונה B.
ידוע ש- 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 כעת ההסתברות שמוצר המיוצר במכונה B יהיה פגום?

<u>*פתרון:*</u>

נגדיר את המאורעות הבאים:

A- המוצר הנבחר יוצר במכונה A. נתון: $P(A) = 0.1$

B- המוצר הנבחר יוצר במכונה B. נתון: $P(B) = 0.9$

C- המוצר שנבחר פגום.

מנתוני השאלה: $P(C \mid B) = 0.05$ $\quad P(C \mid A) = 0.01;$

א. בעזרת נוסחת ההסתברות השלמה:

$$P(C) = P(C \mid A)P(A) + P(C \mid B)P(B)$$
$$= 0.01 \cdot 0.1 + 0.05 \cdot 0.9 = 0.046$$

ב. בעזרת נוסחת בייס:

$$P(A \mid C) = \frac{P(A)P(C \mid A)}{P(C)}$$

נציב את הנתונים:

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

ג. שוב נעשה שימוש בנוסחת ההסתברות השלמה (נסמן ב- x את ההסתברות שאינה ידועה):

$$P(C)=P(C|A)P(A)+P(C|B)P(B)$$

$$0.019=0.01\cdot 0.1+x\cdot 0.9$$

$$\Rightarrow x=0.02$$

<u>שאלה 7</u>

מטילים שתי קוביות בזו אחר זו ומתייחסים למאורעות הבאים:

A = {בקובייה הראשונה התקבל "1" או "2" או "3"}

B = {בקובייה השנייה התקבל "4" או "5" או "6"}

C = {סכום שתי ההטלות הוא 7}

האם המאורעות A, B, C הם בלתי תלויים? הראה זאת.

<u>פתרון:</u>

בהטלה של שתי קוביות יש 36 תוצאות אפשריות, לכן גודל מרחב המדגם $|\Omega|=36$

המאורע C: $C=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\},\quad |C|=6$

ההסתברויות של המאורעות בשאלה:

$$P(A)=P(B)=\frac{1}{2}\qquad P(C)=\frac{6}{36}=\frac{1}{6}$$

נבדוק את קיום אי התלות לפי ההגדרה:

$$P(A\cap B)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}\ =\ P(A)P(B)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$$

$$P(A\cap C)=P(\{(1,6),(2,5),(3,4)\})=\frac{3}{36}=\frac{1}{12}\ =\ P(A)P(C)=\frac{1}{2}\cdot\frac{1}{6}=\frac{1}{12}$$

$$P(B\cap C)=P(\{(1,6),(2,5),(3,4)\})=\frac{3}{36}=\frac{1}{12}\ =\ P(A)P(C)=\frac{1}{2}\cdot\frac{1}{6}=\frac{1}{12}$$

השוויונים לעיל מתקיימים, אבל-

$$P(A\cap B\cap C)=\frac{3}{36}=\frac{1}{12}\ \neq\ P(A)P(B)P(C)=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{12}=\frac{1}{24}$$

<u>המסקנה</u>: המאורעות הם בלתי תלויים בזוגות, אך אינם בלתי תלויים באופן כללי.

<u>שאלה 8</u>

מטילים שלוש פעמים מטבע הוגן שעל צד אחד שלו "1", ועל צד שני "2". נסמן את סכום תוצאות ההטלות ב-k.

כעת מטילים מטבע הוגן אחר, שעל צד אחד שלו "H" ועל הצד שני "T", k פעמים. מה ההסתברות שלא יצא "H" כלל בהטלות?

<u>פתרון:</u>

נסמן: A_n – סכום תוצאות ההטלות $n = \{3, 4, 5, 6\}$.

בעזרת נוסחת ההסתברות השלמה-

$$P(B)=\sum_{n=3}^{6}P(B\,|\,A_n)\cdot P(A_n)$$

$$=\left(\frac{1}{2^3}\cdot\frac{1}{8}\right)+\left(\frac{1}{2^4}\cdot\frac{3}{8}\right)+\left(\frac{1}{2^5}\cdot\frac{3}{8}\right)+\left(\frac{1}{2^6}\cdot\frac{1}{8}\right)=\frac{15}{256}\approx 0.0586$$

<u>שאלה 9</u>

תצפיתן לאיתור מסתננים משקיף על עמק. לוקח בדיוק n שעות לחצות את העמק. לפני תום כל שעה התצפיתן עורך תצפית לגלות מסתננים. ההסתברות שיאתר מסתנן היא p_0 אם המסתנן לובש בגד הסוואה ו- p_1 אם אינו לובש בגד הסוואה. ההסתברות שמסתנן ילבש בגד הסוואה היא b. מסתנן מתחיל לחצות את העמק בדיוק בתחילת שעה מסויימת. חשב את ההסתברות שהתצפיתן יבחין במסתנן לפחות פעם אחת במהלך תצפיותיו.

<u>פתרון:</u>

ההסתברות שהתצפיתן יבחין במסתנן במהלך תצפית אחת היא-

$$bp_0+(1-b)\,p_1$$

ההסתברות שהמסתנן לא יאותר במהלך תצפית אחת היא-

$$1-bp_0-(1-b)\,p_1$$

ההסתברות שהמסתנן לא יאותר במהלך n תצפיות היא-

$$[1-bp_0-(1-b)\,p_1]^n$$

לכן, ההסתברות המבוקשת – שהתצפיתן יבחין במסתנן לפחות פעם אחת במהלך תצפיותיו היא:

$$1-[1-bp_0-(1-b)\,p_1]^n$$

<u>שאלה 10</u>

צוות של אח ואחות משתתף בחידון בו מוצגת בכל סיבוב שאלה על מסך. התשובות האפשריות לכל שאלה הן "נכון" או "לא נכון" בלבד.

מאחר ולאח ולאחות אותה רמת ידע, כל אחד מהם יודע לענות תשובה נכונה לשאלה בהסתברות קבועה a $(0 \le a \le 1)$ באופן בלתי תלוי האחד באחר. (כמובן שהם אינם יודעים אם ענו נכון על השאלה, אלא רק לאחר שהצביעו).

כללי החידון למתן תשובה הם כדלקמן: לאחר שהשאלה הופיעה על המסך, שני חברי הצוות דנים בה. אם הם מסכימים על התשובה, ישיבו את התשובה המוסכמת. אם הם חלוקים בדעותיהם, הם מטילים מטבע הוגן כדי לבחור את התשובה שישיבו.

א. מה ההסתברות שענו תשובה נכונה לשאלה המוצגת על המסך?

ב. בהינתן שהזוג לא הטיל מטבע לאחר הדיון המשותף, מה ההסתברות שענו תשובה נכונה לשאלה המוצגת על המסך?

<u>פתרון:</u>

א. נגדיר את המאורעות הבאים ונחשבם:

D_1 – שני חברי הצוות הסכימו על תשובה נכונה-

$$P(D_1) = P(A_1 \cap A_2) = P(A_1)P(A_2) = a^2$$

D_2 – שני חברי הצוות הסכימו על תשובה לא נכונה-

$$P(D_2) = P(A_1^c \cap A_2^c) = P(A_1^c)P(A_2^c) = (1-a)^2$$

D_3 – שני חברי הצוות לא הסכימו ביניהם-

$$P(D_3) = P\left[\left(A_1 \cap A_2^c\right) \cup \left(A_1^c \cap A_2\right)\right] = P\left(A_1 \cap A_2^c\right) + P\left(A_1^c \cap A_2\right)$$
$$= P(A_1)P(A_2^c) + P(A_1^c)P(A_2) = 2a(1-a)$$

נסמן ב- C את המאורע: "ענו תשובה נכונה".

את $P(C)$ נחשב בעזרת נוסחת ההסתברות השלמה: $P(C) = \sum_{i=1}^{3} P(C \mid D_i)P(D_i)$

ההסתברויות שנותר לנו לחשב כדי להשתמש בנוסחה:

$$P(C \mid D_1) = 1, \quad P(C \mid D_2) = 0, \quad P(C \mid D_3) = \frac{1}{2}$$

למצוא את ההסתברות המבוקשת, נציב בנוסחה ההסתברות השלמה-

$$P(C) = 1 \cdot a^2 + 0 \cdot (1-a)^2 + \frac{1}{2} \cdot 2a(1-a) = a$$

ב. חישוב ההסתברות המבוקשת:

$$P(C \mid D_1 \cup D_2) = \frac{P[C \cap (D_1 \cup D_2)]}{P(D_1 \cup D_2)} = \frac{P\left[(C \cap D_1) \cup (C \cap D_2)\right]}{P(D_1 \cup D_2)} = \frac{P(C \cap D_1)}{P(D_1 \cup D_2)}$$
$$= \frac{P(C \mid D_1)P(D_1)}{P(D_1 \cup D_2)} = \frac{1 \cdot a^2}{a^2 + (1-a)^2} = \frac{a^2}{a^2 + (1-a)^2}$$