פתרון

Y

שאלה 1 (38 נקודות)

במעדניה שכונתית יש 5 סוגים שונים של גבינות קשות: 2 סוגים של גבינות קשות צרפתיות, 3 סוגים של גבינות קשות איטלקיות. בנוסף יש במעדניה 4 סוגים שונים של גבינות רכות: 3 סוגים של גבינות רכות יווניות וסוג אחד של גבינה רכה צרפתית. הגבינות נמכרות באריזות של 250 גרם. בכל יום שישי אריאלה מגיעה למעדניה ובוחרת באקראי שני סוגים שונים (קונה שתי אריזות) של גבינות קשות וסוג אחד (קונה אריזה אחת) של גבינה רכה.

א. (8 נקודות) מהי ההסתברות ששני הסוגים של גבינות קשות שאריאלה תבחר במעדניה יהיו סוגים של גבינה צרפתית?

ב. (10 נקודות) ברי היא גבינה רכה צרפתית. ארוחת יום שישי שלא מכילה ברי נחשבת לארוחה משעממת. מהי ההסתברות שלפחות שלוש ארוחות בימי שישי בחודש הבא (ארבעה שבועות) **לא יהיו משעממות**?

ג. (10 נקודות) אריזה של גבינה צרפתית מסוג כלשהו עולה 10$, אריזה של גבינה איטלקית או יוונית מסוג כלשהו עולה 12$. מהי תוחלת התשלום עבור שלוש האריזות שאריאלה תקנה בשבוע כלשהו (כאמור, אריאלה קונה 2 סוגים של גבינות קשות וסוג אחד של גבינה רכה)?

ד. (5 נקודות) אריאלה מכינה לארוחה שלושה מאכלים. אחד מהם מכיל גבינה רכה. פלג, הבן של אריאלה, לא אוהב גבינה רכה. הוא טועם מאכלים אחד אחרי השני, עד שהוא טועם מאכל שמכיל גבינה רכה. מהי ההסתברות שפלג יטעם את כל שלושת מאכלים?

1. 4/27
2. 1/3!
3. 2/3
4. 1/3

ה. (5 נקודות) זמן הכנת פסטה בדקות מתפלג נורמלית עם סטיית תקן של דקה אחת. מהי התוחלת של זמן ההכנה אם ידוע ש 90% מהפסטות מוכנות תוך 10 דקות?

1. 8.718
2. 11.282
3. 12.82
4. 9.1003

פתרון:

א. נסמן ב- X מספר אריזות של גבינה צרפתית מתוך שתי אריזות גבינה קשה שאריאלה תבחר באקראי. מתקיים:
$X\sim HG(5,2,2)$. נחשב את ההסתברות המבוקשת:

$$P(X=2)=\frac{\binom{2}{2}}{\binom{5}{2}}=\frac{1}{20}$$

ב. ישנם ארבעה סוגים שונים של גבינה רכה. אריאלה בוחרת סוג אחד באקראי ולכן לכל סוג יש אותה הסתברות להיבחר. נסיק שסיכוי לבחור את ברי שווה ל-1/4. נסמן ב- Y מספר ארוחות מתוך ארבע שמכילות את ברי. מתקיים: $Y\sim Bin(4,1/4)$. לכן ההסתברות המבוקשת בשאלה היא:

$$P(Y\geq 3)=P(Y=3)+P(Y=4)=\binom{4}{3}\cdot\left(\frac{1}{4}\right)^3\cdot\frac{3}{4}+\left(\frac{1}{4}\right)^4=\frac{13}{256}$$

ג. נסמן ב - Z מספר אריזות גבינה צרפתית רכה מתוך אריזת גבינה רכה שאריאלה תבחר באקראי. נשתמש במשתנה מקרי X שהגדרנו בסעיף א. כמו כן נסמן ב- S תשלום עבור שלוש אריזות. מתקיים: $X\sim HG(5,2,2)$ ו- $Z\sim HG(4,1,1)$. קיים קשר בין המשתנים מקריים שהגדרנו:

$$S=10X+12(2-X)+10Z+12(1-Z)=36-2X-2Z$$

לכן נקבל:

$$E(S)=36-2E(X)-2E(Z)=36-2\cdot\frac{1}{4}-2\cdot 2\cdot\frac{2}{5}=33.9$$

ד. תשובה נכונה: 1/3

ה. נתון שעשירון העליון שווה ל-10. לכן מתקיים:

$$10=\mu+1\cdot z_{0.9}=\mu+1.282\Rightarrow\mu=8.718$$

שאלה 2 (36 נקודות)

זמן הנסיעה לעבודה של עורך דין, הוא משתנה מקרי בעל התפלגות נורמלית, עם תוחלת של 24 דקות, וסטיית תקן של 3.8 דקות.

א. (8 נקודות) עורך דין יוצא מהבית בשעה 8:30. הוא אמור להתחיל את יום העבודה בשעה 9:00. מהו אחוז הימים שבהם הוא יאחר לעבודה?

ב. (10 נקודות) בכל יום מגישים בעבודה קפה בין השעות 8:50 ל 9:00. אם הוא יוצא מהבית כל יום ב 8:30, מה הסיכוי לכך שלא יפספס את הקפה, בלפחות 90 מתוך 100 ימי העבודה הבאים?

ג. (8 נקודות) ידוע שהזמן שלוקח לעורך דין לטפל בתיק מתפלג מעריכית עם תוחלת של 10 ימי עבודה. מה ההסתברות שהוא יסיים לטפל בתיק שהוא עובד עליו כרגע בשלושה ימים הקרובים? נמקו.

ד. (5 נקודות) בימים מסוימים יש עבודות על הכביש, והזמן שלוקח לעורך דין להגיע לעבודה הוא משתנה מקרי עם פונקציית התפלגות מצטברת:

$$F_X(t) = \begin{cases} 0, & t < 30 \\ at^2 + b, & 30 \le t \le 60 \\ 1, & t > 60 \end{cases}$$

מצאו ערכי הפרמטרים a,b.

1. $a = \frac{1}{2700}, b = -\frac{1}{3}$
2. $a = \frac{1}{2700}$ ו-b יכול להיות כל מספר ממשי.
3. $a = \frac{1}{2700}, b = \frac{1}{3}$
4. אף תשובה אינה נכונה.

ה. (5 נקודות) נסמן ב- X זמן הנסיעה לעבודה ביום רגיל. מהו $x_{0.2}$ (אחוזון 20%):

1. 21.7987
2. 27.1996
3. 20.8004
4. 22.3055

פתרון:

א. נסמן ב- X זמן הנסיעה לעבודה. נחשב את ההסתברות לכך שזמן הנסיעה גדול מ-30 דקות:

$$P(X > 30) = 1 - \Phi\left(\frac{30-24}{3.8}\right) = 1 - \Phi(1.59) = 1 - 0.9441 = 0.0559 \Rightarrow 5.59\%$$

ב. הסיכוי שלא יפספס קפה ביום נתון הוא הסיכוי שהוא יגיע לעבודה תוך פחות מ-30 דקות: $p = P(X \le 30) = 0.9441$. נסמן ב-Y את מספר הימים בהם יצליח לא לפספס את הקפה מתוך 100 ימי העבודה הבאים. אז $Y \sim Bin(100, p)$. נבדוק את התנאים: $np > 5, nq > 5$. התנאים מתקיימים.

לפי קירוב נורמלי להתפלגות בינומית נקבל

$$P(Y \ge 90) = 1 - P(Y < 90) = 1 - \Phi\left(\frac{90 - 0.5 - 100 \cdot 0.9441}{\sqrt{100 \cdot 0.9441 \cdot (1 - 0.9441)}}\right) = \Phi(2.14) = 0.9838$$

ג. נסמן ב-T את הזמן שלוקח לעורך דין לטפל בתיק. מהנתון $T \sim \exp(0.1)$. מתכונת חוסר הזכרון מתקיים שההסתברות שווה להסתברות לסיים לטפל בתיק חדש תוך 3 ימים:

$$P(T \le 3) = 1 - P(T > 3) = 1 - e^{-0.1 \cdot 3} = 1 - e^{-0.3} = 0.2592$$

ד. מרציפות פונקציית ההתפלגות המצטברת נקבל שתי דרישות: $a \cdot 30^2 + b = 0$ ו-$a \cdot 60^2 + b = 1$ ולכן: $a = \frac{1}{2700}, b = -\frac{1}{3}$

ה. $x_{0.2} = 24 + 3.8 \cdot z_{0.2} = 24 - 3.8 \cdot z_{0.8} = 24 - 3.8 \cdot 0.842 = 20.8004$

שאלה 3 (26 נקודות)

א. (8 נקודות) ידוע כי תוחלת IQ של תלמידי שנה א' במוסדות להשכלה גבוהה בישראל בארץ היא 110, עם סטיית תקן 15. חוקר טוען כי תוחלת IQ של תלמידי שנה א' במכללת אפקה גבוהה יותר. לצורך בדיקת טענתו אסף החוקר מדגם מקרי של 100 תלמידי שנה א' באפקה ונמצא כי ממוצע ה- IQ שלהם הינו 112.8.

נסחו את השערות המחקר, הגדירו את סטטיסטי המבחן והכריעו האם החוקר צודק ברמת מובהקות

$\alpha = 0.05$ (*i*), $\alpha = 0.025$ (*ii*).

ב. (8 נקודות) הציון במבחן בסטטיסטיקה הוא משתנה מקרי שתוחלתו μ אינה ידועה. כדי לאמוד את התוחלת ילקח מדגם של 5 סטודנטים, ויבדקו ציוניהם. מוצעים שלושת האומדים לתוחלת על סמך המדגם. עבור כל אחד מהאומדים קבעו האם הוא חסר הטיה? נמקו.

$$T_1 = \frac{X_1 + ... + X_5}{5} \quad ; \quad T_2 = \frac{2X_1 - X_2 + 2X_3 - X_4 + X_5}{5} \quad ; \quad T_3 = \frac{2X_1 - X_2 + 2X_3 - X_4 + X_5}{3}$$

ג. (5 נקודות) שני חוקרים בדקו אותו מדגם לצורך בדיקת ההשערות $H_0: \mu = \mu_0$ כנגד $H_1: \mu > \mu_0$ כאשר השונות ידועה. חוקר א' עבד עם $\alpha = 0.05$ והגיע למסקנה כי יש לדחות את השערת האפס. איזו מבין הטענות הבאות נכונה:

1. אם חוקר ב' משתמש באותו מבחן ועובד עם $\alpha < 0.05$ אזי הוא בהכרח גם ידחה את השערת האפס.
2. אם חוקר ב' משתמש באותו מבחן ועובד עם $\alpha > 0.05$ אזי הוא בהכרח לא ידחה את השערת האפס.
3. אם חוקר ב' משתמש באותו מבחן ועובד עם $\alpha > 0.05$ אזי הוא בהכרח גם ידחה את השערת האפס.
4. ללא נתונים נוספים לא ניתן לדעת מה יחליט חוקר ב' ללא תלות בערכה של α לפיה בחר לעבוד.

ד. (5 נקודות) יהיו X_1, X_2 שתי תצפיות בלתי תלויות מהתפלגות אחידה בדידה על $1, 2, ..., N$. כדי לאמוד את הפרמטר N, הוצעו שני האומדים הבאים: $T_1 = X_1 + X_2 - 1$, $T_2 = 2X_1 - 1$. איזו מהטענות הבאות נכונה:

1. שני האומדים מוטים.
2. שני האומדים חסרי הטיה עם שונויות זהות.
3. אחד מהם מוטה והאחר חסר הטיה.
4. שני האומדים חסרי הטיה ושונות של T_1 קטנה יותר.

פתרון:

א. ההשערות הנבדקות הן H_0: $\mu = 110$ כנגד H_1: $\mu > 110$, סטטיסטי המבחן: $\bar{X}_{100}$ או $Z = \frac{\bar{X}_{100} - 110}{15/\sqrt{100}}$

נחשב את מובהקות התוצאה שהתקבלה בניסוי:

$$p - value = P(\bar{X}_{100} > 112.8) = 1 - \Phi\left(\frac{112.8 - 110}{15/10}\right) = 1 - \Phi(1.87) = 1 - 0.9694 = 0.031$$

לכן עבור $\alpha = 0.05$ נדחה את השערת האפס ונסיק כי תוחלת IQ של תלמידי אפקה גבוהה יותר. אבל, עבור $\alpha = 0.025$ לא נדחה את השערת האפס, כלומר לא ניתן להסיק כי תוחלת IQ של תלמידי אפקה גבוהה יותר.

ב. נחשב תוחלת של כל אחד מהאומדים:

$$E[T_1] = E\left[\frac{X_1 + ... + X_5}{5}\right] = \frac{5\mu}{5} = \mu$$

$$E[T_2] = E\left[\frac{2X_1 - X_2 + 2X_3 - X_4 + X_5}{5}\right] = \frac{3\mu}{5} \neq \mu$$

$$E[T_3] = E\left[\frac{2X_1 - X_2 + 2X_3 - X_4 + X_5}{3}\right] = \frac{3\mu}{3} = \mu$$

T_1 ו- T_3 הם אומדים חסרי הטיה ל- μ.

ג. אם חוקר ב' משתמש באותו מבחן ועובד עם $\alpha > 0.05$ אזי הוא בהכרח גם ידחה את השערת האפס.

ד. תוחלת של כל אחד משני האומדים שווה ל- N ושונות של אומד T_1 קטנה יותר ולכן תשובה הנכונה.