סטטיסטיקה לאקטואריה

תוכן העניינים

סטטיסטיקה לאקטואריה

פרק 1 - פונקציה יוצרת מומנטים

תוכן העניינים

פונקציה יוצרת מומנטים:

רקע:

פונקציה יוצרת מומנטים של משתנה מקרי כלשהו מוגדרת להיות: $M_X(t)=E\left(e^{tx}\right)$.

אם מדובר במשתנה מקרי בדיד, הפונקציה יוצרת המומנטים תהיה:

$$M_X(t)=E\left(e^{tX}\right)=\sum_k e^{tk}\cdot P(X=k).$$

אם מדובר במשתנה מקרי רציף, פונקציית יוצרת המומנטים תהיה:

$$M_X(t)=E\left(e^{tx}\right)=\int_x e^{tx}\cdot f(x)\,dx.$$

המומנט מסדר n מוגדר להיות: $E\left(X^n\right)$.

מומנט מסדר n של משתנה מקרי X מתקבל מהנגזרת ה-n-ית לפי t של פונקציית יוצרת המומנטים $M_X(t)$ בנקודה שבה $t=0$. כלומר: $M_x^{(n)}(t)|_{t=0}=E\left(X^n\right)$.

משפט:

קיימת התאמה חד-חד-ערכית בין משתנה מקרי לבין פונקציית יוצרת המומנטים שלו.

תזכורת מתמטית לנגזרות:

$$\left(x^n\right)'=nx^{n-1}$$

$$\left(k\cdot f(x)\right)'=k\cdot f'(x)$$

$$\left(a^x\right)'=a^x\cdot\ln a$$

$$\left(e^x\right)'=e^x$$

$$\left(\log_a{}^x\right)'=\frac{1}{x\ln a}$$

$$(\ln x)'=\frac{1}{x}$$

$$\left(\frac{1}{x}\right)' = -\frac{1}{x^2}$$

$$(k)' = 0$$

$$(f(x) \pm g(x))' = f'(x) \pm g'(x)$$

$$(f(x) \cdot g(x))' = f'(x)g(x) + g'(x)f(x)$$

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - g'(x)f(x)}{(g(x))^2}$$

כלל שרשרת - $[f(g(x))]' = f'(g(x)) \cdot g'(x)$

דוגמה (פתרון בהקלטה):

הראו שהפונקציה יוצרת המומנטים של ההתפלגות המעריכית: $X \sim \exp(\lambda)$,

היא: $\frac{\lambda}{\lambda - t}$. מצאו את המומנט הראשון והמומנט השני של ההתפלגות.

שאלות:

1) נתונה פונקציה ההסתברות הבאה למשתנה מקרי בדיד.
   א. מצאו את פונקציית יוצרת המומנטים.
   ב. מצאו את התוחלת על סמך סעיף א'.

X	1	2	3
$P(X)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

2) מצאו את פונקציית יוצרת המומנטים של התפלגות הבינומית: $X \sim B(n,p)$,
ומצאו את המומנט הראשון והשני של הפונקציה.

3) מצאו את פונקציית יוצרת המומנטים של ההתפלגות הגיאומטרית: $X \sim G(P)$,
וחשבו את תוחלת של ההתפלגות מתוך פונקציית יוצרת המומנטים.

4) מצאו את פונקציית יוצרת מומנטים של התפלגות הפואסונית: $x \sim p(\lambda)$.
מצאו את המומנט הראשון והשני של ההתפלגות.

5) יהי X משתנה מקרי בעל פונקציית הצפיפות הבאה:

$$f(x)=\begin{cases} Ae^{-x} & 0 \le x \le 7 \\ 0 & \text{אחר ת} \end{cases}.$$

   א. מצאו את ערכו של A.
   ב. מצאו את הפונקציה יוצרת המומנטים של X.

6) יהי X משתנה מקרי עם תוחלת 5 ושונות 16, ותהי $m_x(t)$ פונקציית יוצרת המומנטים של X. Y הינו משתנה מקרי עם פונקציית יוצרת מומנטים $m_y(t)$,
ונתון: $m_y(t)=t\cdot m_x(t)$.
חשבו את התוחלת והשונות של Y.

תשובות סופיות:

1) א. פונקציה יוצרת מומנטים: $\frac{1}{2}e^{t}+\frac{1}{3}e^{2t}+\frac{1}{6}e^{3t}$. ב. $1\frac{2}{3}$.

2) פונקציה יוצרת מומנטים: $\left(e^{t}\cdot p+1-p\right)^{n}$.

3) פונקציה יוצרת מומנטים: $\frac{e^{t}p}{1-e^{t}\cdot(1-p)}$.

4) פונקציה יוצרת מומנטים: $e^{\lambda\left(e^{t}-1\right)}$.

5) א. $\frac{1}{1-e^{-7}}$. ב. פונקציה יוצרת מומנטים: $\frac{e^{7(t-1)}-1}{t-1}\cdot\frac{e^{7}}{e^{7}-1}$.

6) תוחלת: 1, שונות: 9.

נספחים:

התפלגות	פונקציית צפיפות $f(X)t$	פונקציית התפלגות מצטברת $f(X)t$
אחיד $U(a,b)$	$f_x(t)=\begin{cases} \frac{1}{(b-a)} & a \le t \le b \\ 0 & else \end{cases}$	$f_x(t)=\begin{cases} 0 & t<0 \\ \frac{t-a}{b-a} & a \le t \le b \\ 1 & t<b \end{cases}$
מעריכי $\exp(\lambda)$	$f_x(t)=\begin{cases} \lambda e^{-\lambda t} & t \ge 0 \end{cases}$	$f_x(t)=\begin{cases} 1-e^{-\lambda t} & t \ge 0 \\ 0 & else \end{cases}$
נורמלית $N(\mu,\sigma^2)$	$f_x(t)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	$\phi\left(\frac{t-\mu}{\sigma}\right)$

התפלגות	$E(X)$	$VAR(X)$	$M_X(t)$
אחיד $U(a,b)$	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{t\,b}-e^{t\,a}}{t(b-a)}$
מעריכי $\exp(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$
נורמלית $N(\mu,\sigma^2)$	μ	σ^2	$e^{\mu t+\frac{\sigma^2 t^2}{2}}$

משתנה מקרי	משמעות	$P_X(x)$	$E(x)$	$Var(x)$	$M_X(t)$
בינומי $Bin(n,p)$	חוזרים באופן בלתי תלוי על אותו ניסוי ברנולי n פעמים: P ההסתברות להצלחה. $1-P=q$ ההסתברות לכישלון X - מספר ההצלחות	$\binom{n}{x}p^xq^{n-x}$ $x=0,1,\ldots,n$	$n\cdot p$	$n\cdot p\cdot q$	$\left[pe^t+q\right]^n$
גיאומטרי $G(P)$	חוזרים באופן בלתי תלוי על אותו ניסוי ברנולי עד ההצלחה הראשונה. X - מספר ניסויים עד הצלחה ראשונה.	pq^{x-1} $x=1,2,\ldots,\infty$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1-qe^t}$
פואסוני $Pois(\lambda)$	X - מספר ההופעות בילידת זמן. מ"מ המקבל ערכים $0,1,\ldots,\infty$.	$e^{-\lambda}\frac{\lambda^x}{x!}$	λ	λ	$e^{\lambda(e^t-1)}$

סטטיסטיקה לאקטואריה

פרק 2 - תכונות של פונקציית יוצרת מומנטים

תוכן העניינים

תכונות של פונקציה יוצרת מומנטים:

רקע:

להלן מספר תכונות שפונקציית יוצרת מומנטים מקיימת:

- קיימת התאמה חד-חד-ערכית בין משתנה מקרי לבין פונקציית יוצרת המומנטים שלו.
- השפעת טרנספורמציה לינארית על פונקציית יוצרת מומנטים:
  $M_{aX+b}(t)=e^{bt}M_X(at)$.
- אם X ו- Y משתנים בלתי תלויים מתקיים ש:
  $M_{X+Y}(t)=E(e^{tx})\cdot E(e^{ty})=M_X(t)\cdot M_Y(t)$.

תזכורת:

התפלגות	פונקציית צפיפות $f_X(t)$	פונקציית התפלגות מצטברת $F_X(t)$
אחיד $U(a,b)$	$f_x(t)=\begin{cases}\frac{1}{(b-a)} & a\le t\le b\\ 0 & else\end{cases}$	$f_x(t)=\begin{cases}0 & t<0\\ \frac{t-a}{b-a} & a\le t\le b\\ 1 & t<b\end{cases}$
מעריכי $\exp(\lambda)$	$f_x(t)=\{\lambda e^{-\lambda t} \quad t\ge 0$	$f_x(t)=\begin{cases}1-e^{-\lambda t} & t\ge 0\\ 0 & else\end{cases}$
נורמלית $N(\mu,\sigma^2)$	$f_x(t)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	$\phi\left(\frac{t-\mu}{\sigma}\right)$

התפלגות	$E(X)$	$VAR(X)$	$M_X(t)$
אחיד $U(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
מעריכי $\exp(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$
נורמלית $N(\mu,\sigma^2)$	μ	σ^2	$e^{\mu t+\frac{\sigma^2t^2}{2}}$

משתנה מקרי	משמעות	$P_X(x)$	$E(x)$	$Var(x)$	$M_X(t)$
בינומי $Bin(n,p)$	חוזרים באופן בלתי תלוי על אותו ניסוי ברנולי n פעמים: P ההסתברות להצלחה $1-P=q$ ההסתברות לכישלון x: מספר ההצלחות	$\binom{n}{x}p^x q^{n-x}$ $x=0,1,\dots,n$	$n\cdot p$	$n\cdot p\cdot q$	$\left[pe^t+q\right]^n$
גיאומטרי $G(p)$	חוזרים באופן בלתי תלוי על אותו ניסוי ברנולי עד ההצלחה הראשונה. x: מספר ניסויים עד הצלחה ראשונה	pq^{x-1} $x=1,2,\dots,\infty$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1-qe^t}$
פואסוני $Pois(\lambda)$	x: מספר ההופעות בילידת זמן. מ"מ המקבל ערכים $0,1,\dots,\infty$	$e^{-\lambda}\frac{\lambda^x}{x!}$	λ	λ	$e^{\lambda(e^t-1)}$

דוגמה (פתרון בהקלטה):

נתון: $X\sim P(\lambda=4)$ $Y\sim P(\lambda=2)$.

X ו- Y הינם בלתי תלויים.

א. מהי פונקציית יוצרת המומנטים של: $5X-3$?

ב. נגדיר את: $T=X+Y$. מה ההתפלגות של T?

שאלות:

1) נתון ש- $X_i \sim p(\lambda)$ בלתי תלויים.

א. מצאו את פונקציית יוצרת מומנטים של $\sum_{i=1}^{n} X_i$.

ב. הוכיחו ש- $\sum_{i=1}^{n} X_i \sim P(n \cdot \lambda)$.

2) נתון: $X \sim P(\lambda = 10)$, $Y \sim P(\lambda = 2)$.

X ו- Y הינם בלתי תלויים. נגדיר את: $T = X + Y$.

א. מצאו את פונקציית יוצרת המומנטים של T.

ב. הוכיחו ש- $T \sim P(\lambda = 12)$.

ג. הוכיחו ש- $X / T = 8 \sim B\left(8, \frac{5}{6}\right)$. כלומר, ההתפלגות של X, בהינתן ש- $T = 8$ היא בינומית עם הפרמטרים: $n = 8$ ו- $p = \frac{5}{6}$.

3) יהי: $X_i \sim \exp(1)$, $i = 1, 2 \ldots n$ והמשתנים הם בלתי תלויים.

נגדיר את $T = \sum_{i=1}^{n} X_i$.

א. מצאו את פונקציית יוצרת המומנטים של T.

ב. חשבו את התוחלת והשונות של T.

ג. יהי: $Z = \frac{T - E(T)}{\sigma(T)}$ כלומר התקנון של T.

מצאו את פונקציית יוצרת המומנטים של Z.

4) נתון שפונקציית יוצרת מומנטים של ההתפלגות הנורמלית נתונה על ידי הנוסחה הבאה: $M_X(T) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$ לכל t, כאשר: $X \sim N(\mu, \sigma^2)$.

א. הוכיחו שאם $Y = 2X$ אזי $Y \sim N(2\mu, 4\sigma^2)$.

ב. הוכיחו שאם $T = X_1 + X_2$ ו- X_1 ו- X_2 בלתי תלויים מאותה התפלגות נורמלית אז מתקיים ש: $T \sim N(2\mu, 2\sigma^2)$.

תשובות סופיות:

1) א. פונקציה יוצרת מומנטים: $e^{(n\lambda)\cdot\left(e^t-1\right)}$.   ב. שאלת הוכחה.

2) א. פונקציה יוצרת מומנטים: $e^{12\left(e^t-1\right)}$.   ב. שאלת הוכחה.

ג. שאלת הוכחה.

3) א. פונקציה יוצרת מומנטים: $\left(\frac{1}{1-t}\right)^n$.   ב. תוחלת: n, שונות: n.

ג. פונקציה יוצרת מומנטים: $\left(\frac{1}{1-\left(n^{-\frac{1}{2}}t\right)}\right)^n \cdot e^{-n^{\frac{1}{2}}\cdot t}$.

4) א. שאלת הוכחה.   ב. שאלת הוכחה.

סטטיסטיקה לאקטואריה

פרק 3 - משתנה דו-מימדי בדיד - פונקציית הסתברות משותפת

תוכן העניינים

משתנה דו מימדי בדיד – פונקציית הסתברות משותפת:

רקע:

התפלגות דו ממדית הינה התפלגות שדנה בשני משתנים. נרצה כעת לבנות פונקציית הסתברות דו ממדית, בה יש התפלגות של שני משתנים בו זמנית: X ו- Y.

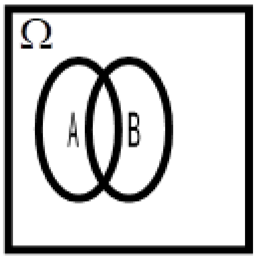
דוגמה:

תלמיד ניגש בסמסטר לשני מבחנים: מבחן בכלכלה ומבחן בסטטיסטיקה. כמו כן, נתון שהסיכוי לעבור את המבחן בכלכלה הנו 0.8, הסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9 והסיכוי לעבור את שני המבחנים הנו 0.75.

יהי X מספר הקורסים שהסטודנט עבר. ויהי Y משתנה אינדיקטור המקבל את הערך 1 אם הסטודנט עבר את הבחינה בכלכלה, ו-0 אחרת.

בנו את פונקציית ההסתברות המשותפת של X ו- Y.

נחשב את כל ההסתברויות המשותפות:



y/x	0	1	2
0	0.05	0.15	0
1	0	0.05	0.75

$$p(x=o,y=0)=0.05$$
$$p(x=o,y=1)=0$$
$$p(x=1,y=0)=0.15$$
$$p(x=1,y=1)=0.05$$
$$p(x=2,y=0)=0$$
$$p(x=2,y=1)=0.75$$

שימו לב שסכום כל ההסתברויות בפונקציית ההסתברות המשותפת הוא 1.

כעת נסכם את השורות ואת העמודות ונקבל את פונקציות הסתברות שוליות:

Y/X	0	1	2	P_Y
0	0.05	0.15	0	0.2
1	0	0.05	0.75	0.8
P_X	0.05	0.2	0.75	1

משתנים בלתי תלויים:

X ו- Y יהיו משתנים בלתי תלויים, אם עבור כל X ו- Y אפשריים התקיים הדבר הבא: $p(x=k, y=l)=p(x=k)\cdot p(y=l)$.

מספיק פעם אחת שהמשתנים אינם מקיימים תנאי זה אזי הם תלויים.

דוגמה:

$p(x=2, y=1)=0.75 \neq p(x=2)\cdot p(y=1)=0.75\cdot 0.8=0.6$.

ככלל, אם יש אפס בתוך פונקציית ההסתברות המשותפת ניתן להבין באופן מידי שהמשתנים תלויים, שאז הרי התנאי לא מתקיים. אך אם אין אפס בטבלה, אין זה אומר שהמשתנים בלתי תלויים ויש לבדוק זאת.

שאלות:

1) אדם נכנס לקזינו עם 75 דולר. הוא ישחק במכונת מזל בה יש סיכוי של 0.3 לנצח. במקרה של ניצחון במשחק הוא יקבל מהקזינו 25 דולר ובמקרה של הפסד הוא ישלם 25 דולר. אותו אדם החליט שיפסיק לשחק ברגע שיהיה לו 100 דולר, אך בכל מקרה לא ישחק יותר מ-3 משחקים.
נגדיר את X להיות הכסף שברשות האדם בצאתו מהקזינו ואת Y כמספר המשחקים שהאדם שיחק.

א. בנו את פונקציית ההסתברות המשותפת והשוליות.
ב. מה תוחלת מספר המשחקים שישחק האדם?
ג. אם האדם יצא מהקזינו שברשותו 100 דולר, מה התוחלת ומהי השונות של מספר המשחקים ששיחק?

2) להלן פונקציית ההסתברות המשותפת והשוליות של שני משתנים מקריים בדידים:

Y/X	0	1	2	$P(Y)$
2		0.08	0.12	0.4
3	0.1	0.05		
4				0.45
$P(X)$		0.4	0.2	

א. השלימו את ההסתברויות החסרות בטבלה.
ב. האם X ו- Y תלויים?
ג. מצאו את הסתברות ש-$Y=3$, אם ידוע ש-$X=1$.

3) מפעל משווק מוצר הנארז בחבילות בגדלים שונים. ישנו מספר שווה של חבילות בנות שני מוצרים ושלושה מוצרים. ההסתברות שמוצר מסוים יהיה פגום היא $\frac{1}{10}$. מהנדס הייצור בוחר באקראי חבילת מוצרים לשם בקרת איכות.
יהי X מספר המוצרים בחבילה, ו- Y מספר המוצרים הפגומים בחבילה.

א. מה ההתפלגות של המשתנה Y בהינתן $x=3$.
ב. מה ההתפלגות של המשתנה Y בהינתן X הינו K כלשהו.
ג. מהי תוחלת מספר המוצרים הפגומים בחבילות בנות 3 מוצרים? נמקו.
ד. בנו את פונקציית ההסתברות המשותפת.

4) מתוך כד עם 3 כדורים ממוספרים במספרים 2, 4, 8 שולפים באקראי שניים ללא החזרה. יהי X המספר הקטן מבין השניים ו- Y הגדול מביניהם.

א. חשבו את ההתפלגות של (X,Y).

ב. אם המספר המינימאלי שנבחר הוא 2, מה הסיכוי שהמקסימאלי הוא 8?

ג. חשבו את ההתפלגות המותנית של X בהינתן $Y=4$. מצאו: $E(X/Y=4)$.

5) ביישוב שני סניפי בנק. סניף פועלים וסניף לאומי. להלן הנתונים לגבי האוכלוסייה הבוגרת המתגוררת ביישוב: ל-60% יש חשבון בסניף פועלים של היישוב, ל-50% יש חשבון בסניף לאומי של היישוב ול-95% יש חשבון בלפחות אחד מהסניפים.
יהי X מספר הסניפים בישוב אשר לתושב בוגר יש בהם חשבון, ויהי Y משתנה אינדיקטור:
1 – אם יש לתושב חשבון בסניף פועלים.
0 – אחרת.

א. בנו את פונקציית ההסתברות המשותפת של X ו- Y.

ב. הוסיפו את פונקציית ההסתברות השולית.

ג. ידוע שלתושב בוגר חשבון בבנק פועלים, מה ההסתברות שיש לו חשבון בנק בסניף אחד בלבד?

תשובות סופיות:

1) א. להלן טבלה: ב. 2.4. ג. תוחלת: 1.348, שונות: 0.575.

$x\backslash y$	0	50	100	$P(y)$
1	0	0	0.3	0.3
3	0.343	0.294	0.063	0.7
$P(x)$	0.343	0.294	0.363	1

2) א. להלן טבלה: ב. תלויים. ג. 0.125.

$x\backslash y$	0	1	2	$P(y)$
2	0.2	0.08	0.12	0.4
3	0.1	0.05	0	0.15
4	0.1	0.27	0.08	0.45
$P(x)$	0.4	0.4	0.2	1

3) א. $y/x=3 \sim B\left(n=3,\ p=\frac{1}{10}\right)$. ב. $y/x=k \sim B\left(n=k,\, p=\frac{1}{10}\right)$.

ג. 0.3. ד. להלן טבלה:

$x\backslash y$	2	3	$P(y)$
0	0.405	0.3645	
1	0.09	0.1215	
2	0.005	0.0135	
3	0	0.0005	
$P(x)$	0.5	0.5	1

4) א. להלן טבלה: ב. 0.5. תוחלת: 2.

$x\backslash y$	2	4	$P(y)$
4	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
8	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
$P(x)$	$\frac{2}{3}$	$\frac{1}{3}$	1

5) א+ב. להלן טבלה: ג. 0.75.

$x\backslash y$	**0**	**1**	**2**	$P(y)$
0	0.05	0.35	0	0.4
1	0	0.45	0.15	0.6
$P(x)$	0.05	0.8	0.15	1

סטטיסטיקה לאקטואריה

פרק 4 - משתנה דו מימדי בדיד - מתאם בין משתנים

תוכן העניינים

משתנה דו מימדי בדיד – מתאם בין משתנים:

רקע:

נרצה לבדוק את מידת ההתאמה הלינארית בין שני המשתנים על ידי מקדם המתאם הלינארי שמסומן ב- ρ.
מקדם מתאם זה מקבל ערכים בין 1- ל-1.

-1 0 1

מקדם מתאם 1- או 1 אומר שקיים קשר ליניארי מוחלט ומלא בין המשתנים, שניתן לבטאו על ידי הנוסחה: $y = ax + b$.
מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר ליניארי מלא בו השיפוע a יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר ליניארי מלא בו השיפוע a שלילי (מקדם מתאם 1-).
מתאם חיובי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

חישוב מקדם המתאם:

הנוסחה של מקדם המתאם היא: $\rho = \dfrac{\text{cov}(x,y)}{\sigma_x \cdot \sigma_y}$.

השונות המשותפת: $\text{cov}(x,y) = E\left[(x-\mu_x)(y-\mu_y)\right] = E(x \cdot y) - E(x) \cdot E(y)$.

תכונות של השונות המשותפת:

1) $\text{cov}(X,Y) = \text{cov}(Y,X)$.

2) $\text{cov}(X,X) = Var(X)$.

משתנים בלתי מתואמים:

משתנים בלתי מתואמים הם משתנים שמקדם המתאם שלהם אפס, וכדי שדבר כזה יקרה השונות המשותפת צריכה להתאפס. משתנים בלתי מתואמים הם משתנים שכלל אין ביניהם התאמה לינארית.
משתנים בלתי תלויים הם משתנים שאין ביניהם קשר ולכן גם הם בלתי מתואמים, אך משתנים בלתי מתואמים אינם בהכרח בלתי תלויים.

השפעת טרנספורמצייה לינארית על מקדם המתאם:

$$\rho\left[(aX+b),(cY+d)\right]=\begin{cases}\rho(X,Y)\_\_if\_\_a\cdot c>0\\-\rho(X,Y)\_\_if\_\_a\cdot c<0\end{cases}$$

כלומר, טרנספורמציה לינארית על שני משתנים לא משנה את עוצמת הקשר ביניהם היא עלולה לשנות רק את כיוון הקשר.

דוגמה (פתרון בהקלטה):

נחזור לדוגמה שהוצגה בפרק הקודם:
תלמיד ניגש בסמסטר לשני מבחנים מבחן בכלכלה ומבחן בסטטיסטיקה. כמו כן, נתון שהסיכוי לעבור את המבחן בכלכלה הנו 0.8, הסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9 והסיכוי לעבור את שני המבחנים הנו 0.75.

יהי X מספר הקורסים שהסטודנט עבר, ויהי Y משתנה אינדיקטור המקבל את הערך 1, אם הסטודנט עבר את הבחינה בכלכלה, ו-0 אחרת.
נחשב את מקדם המתאם:

X/Y	0	1	2	P_Y
0	0.05	0.15	0	0.2
1	0	0.05	0.75	0.8
P_X	0.05	0.2	0.75	1

X	0	1	2
P_X	0.05	0.2	0.75

$$E(X)=\sum_i x_iP(x_i)=\mu=0\cdot 0.05+1\cdot 0.2+2\cdot 0.75=1.7$$

$$V(X)=\sum_i(x_i-\mu)^2P(x_i)=\sum_i x_i^2P(x_i)-\mu^2=0^2\cdot 0.05+1^2\cdot 0.2+2^2\cdot 0.75-1.7^2=0.31=\sigma^2$$

y	P_Y
0	0.2
1	0.8

$$\sigma_x=\sqrt{V(X)}=\sqrt{0.31}=0.557$$

$$E(y)=\sum_i y_iP(y_i)=0+0.8=0.8$$

$$V(y)=\sum_i(y_i-\mu_y)^2P(y_i)=\sum_i y_i^2P(y_i)-\mu_y^{\,2}=0+0.8-0.8^2=0.16=\sigma_y^{\,2}$$

$$\sigma_y=\sqrt{0.16}=0.4$$

$$E(xy)=0\cdot0\cdot0.05+0\cdot1\cdot0+1\cdot0\cdot0.15+1\cdot1\cdot0.05+2\cdot0\cdot0+2\cdot1\cdot0.75=1.55$$

$$\operatorname{cov}(x,y)=E(x\cdot y)-E(x)\cdot E(y)=1.55-1.7\cdot0.8=0.19$$

$$\rho=\frac{\operatorname{cov}(x,y)}{\sigma_x\cdot\sigma_y}=\frac{0.19}{0.557\cdot0.4}=0.853$$

כל קורס שהסטודנט מסיים מזכה אותו ב-3 נקודות אקדמאיות.
מה יהיה מקדם המתאם בין נקודות הזכות שיצבור למשתנה Y?

שאלות:

1) הסיכוי שסטודנט יעבור את המבחן במועד א' בסטטיסטיקה הוא 0.8. אם הוא נכשל במועד א' הוא ניגש למועד ב' שם הסיכוי לעבור את המבחן מוערך ב-0.9 (סטודנט שעובר את א' לא ניגש לב'). במידה והסטודנט נכשל במועד ב' הוא מגיש בקשה למועד ג' אותה מאשרים בסיכוי של 0.2, והסיכוי שלו לעבור את מועד ג' הוא 0.7.
נגדיר את X להיות מספר המבחנים אליהם ניגש הסטודנט, ונגדיר את Y להיות מספר המבחנים שנכשל בהם.

א. בנו את פונקצית ההסתברות המשותפת ואת פונ' ההסתברות השולית.
ב. האם המשתנים הינם בלתי תלויים?
ג. ידוע שהסטודנט ניגש ליותר ממבחן אחד, מה ההסתברות שהוא נכשל בפחות משלושה מבחנים?
ד. האם המתאם בין X ל- Y מלא או חלקי? חיובי או שלילי? הסבירו ללא חישוב.
ה. חשבו את מקדם המתאם בין X לבין Y.
ו. האם המשתנים הם בלתי מתאומים?

2) נטיל מטבע שלוש פעמים. נגדיר את X להיות מספר העצים המתקבלים בשתי ההטלות הראשונות, ואת Y להיות מספר העצים המתקבלים בשתי ההטלות האחרונות.

א. בנו את פונקציית ההסתברות המשותפת של X ו- Y ואת פונקציות ההסתברות השוליות.
ב. האם X ו- Y הם משתנים בלתי תלויים?
ג. מהו מקדם המתאם בין X ל- Y. האם המשתנים מתואמים?
ד. אם בשתי ההטלות הראשונות יצא בדיוק עץ אחד, מה ההסתברות שבשתי ההטלות האחרונות יצאו שני עצים?
ה. אם בשתי ההטלות האחרונות יצא לפחות פעם אחת עץ, מה ההסתברות שבשתי ההטלות הראשונות יצא עץ אחד?

3) נפזר שלושה כדורים שונים בשלושה תאים. נגדיר את המשתנים הבאים:
X - מספר הכדורים בתא הראשון.
Y - מספר הכדורים בתא השני.

א. בנו את פונקציית ההסתברות המשותפת.
ב. האם המשתנים בלתי מתואמים?

4) קובייה הוגנת הוטלה פעמיים.
יהי X ההטלה הגדולה מבין שתי התוצאות, ויהי Y מס' ההטלות בהן יצאה תוצאה זוגית.

א. מצאו את פונקציית ההסתברות המשותפת של X ו- Y.

ב. חשבו את מקדם המתאם של X ו- Y.

ג. מצאו את ההתפלגות של Y בהינתן ש- $X = 2$.

5) בבניין שלנו 5 דירות. דירות מספר אחת ושלוש הן דירות משופצות והשאר אינן.
הוחלט לבחור שתי דירות באקראי מבין הדירות בבניין.
נגדיר את המשתנים הבאים:
X - מספר הדירות המשופצות שנבחרו.
Y - מספר הדירות האי זוגיות שנדגמו.

א. בנו את פונקצית ההסתברות המשותפת ואת פונקציות ההסתברות השולית.

ב. האם המשתנים מתואמים?

ג. מה מקדם המתאם בין X לבין Y?

ד. מה יהיה מקדם המתאם:

i. בין מספר הדירות המשופצות למספר הדירות הזוגיות שנדגמו.

ii. בין מספר הדירות הזוגיות לדירות האי זוגיות שנדגמו.

ה. כל דירה משופצת עולה 2 מיליון ₪ וכל דירה לא משופצת עולה 1.5 מיליון ₪. מה המתאם בין עלות הדירות שנדגמו למספר הדירות הזוגיות?

תשובות סופיות:

1) א. להלן טבלה: ב. תלויים. ג. 0.994 ד. חלקי חיובי.

$x\backslash y$	1	2	3	$P(y)$
0	0.8	0	0	0.8
1	0	0.18	0	0.18
2	0	0.016	0.0028	0.0188
3	0	0	0.0012	0.0012
$P(x)$	0.8	0.196	0.004	1

ה. 0.963. ו. מתואמים.

2) א. להלן טבלה: ב. תלויים. ג. מקדם המתאם: 0.5, מתואמים.

$x\backslash y$	0	1	2	$P(y)$
0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{2}{8}$
1	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{4}{8}$
2	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$
$P(x)$	$\frac{2}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	1

ד. 0.25. ה. 0.5.

3) א. להלן טבלה: ב. מתואמים.

$x\backslash y$	0	1	2	3
0	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{3}{27}$	$\frac{1}{27}$
1	$\frac{3}{27}$	$\frac{6}{27}$	$\frac{3}{27}$	0
2	$\frac{3}{27}$	$\frac{3}{27}$	0	0
3	$\frac{1}{27}$	$\frac{6}{27}$	0	0

4) א. להלן טבלה: ב. 0.252.

$x\backslash y$	1	2	3	4	5	6
0	$\frac{1}{36}$	0	$\frac{3}{36}$	0	$\frac{5}{36}$	0
1	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{4}{36}$	$\frac{6}{36}$
2	0	$\frac{1}{36}$	0	$\frac{3}{36}$	0	$\frac{5}{36}$

5) א. להלן טבלה: ב. X ו- Y מתואמים.   ג. $\frac{2}{3}$.

$x\backslash y$	0	1	2	$P(y)$
0	0.1	0	0	0.1
1	0.2	0.4	0	0.6
2	0	0.2	0.1	0.3
$P(x)$	0.3	0.6	0.1	1

ד. i. $-\frac{2}{3}$.   ii. -1.   ה. $-\frac{2}{3}$.

סטטיסטיקה לאקטואריה

פרק 5 - המשתנה המקרי הדו ממדי - קומבינציות ליניאריות

תוכן העניינים

המשתנה המקרי הדו מימדי – קומבינציות לינאריות:

רקע:

יהיו שני משתנים מקריים X ו- Y.

התוחלת והשונות של סכומם היא:

$$E(X+Y)=E(X)+E(Y)$$
$$V(X+Y)=V(X)+V(Y)+2\cdot\operatorname{cov}(X,Y)$$

התוחלת והשונות של הפרשם היא:

$$E(X-Y)=E(X)-E(Y)$$
$$V(X-Y)=V(X)+V(Y)-2\cdot\operatorname{cov}(X,Y)$$

קומבינציה ליניארית:

יוצרים משתנה חדש שהוא קומבינציה לינארית של שני משתנים אחרים:
$W=(aX+b)+(cY+d)$. אזי:

$$\operatorname{cov}[(aX+b),(cY+d)]=a\cdot c\cdot\operatorname{cov}(X,Y)$$
$$E(W)=E((aX+b)+(cY+d))=aE(X)+b+cE(Y)+d$$
$$V(W)=V((aX+b)+(cY+d))=a^2V(X)+c^2V(Y)+2\cdot a\cdot c\cdot\operatorname{cov}(X,Y)$$

דוגמה (פתרון בהקלטה):

נתונים שני משתנים מקריים X ו- Y המקיימים:
$\mu_X=80$, $\sigma_X=15$, $\mu_Y=70$, $\sigma_Y=20$, $\operatorname{cov}(X,Y)=200$

א. מצאו את התוחלת והשונות של סכום המשתנים.

ב. מצאו את התוחלת והשונות של X ו- Y.

ג. מצאו את השונות ומה התוחלת של המשתנה $W=2X+3Y$.

שאלות:

1) נתונה פונקציית ההסתברות המשותפת הבאה:

Y / X	1	2	3	$P(X)$
2		0.1	0.3	0.6
3	0.2		0.1	
$P(X)$				

א. השלימו את ההסתברויות החסרות.
ב. האם המשתנים תלויים?
ג. האם המשתנים בלתי מתואמים?
ד. חשבו את השונות המשותפת.
ה. חשבו את התוחלת והשונות של סכום המשתנים.
ו. חשבו את התוחלת והשונות של הפרש המשתנים.

2) מבחן בנוי מחלק כמותי וחלק מילולי. תוחלת הציון בחלק הכמותי היא 100, עם סטיית תקן 20. תוחלת הציונים בחלק המילולי היא 90 עם סטיית תקן 15. מקדם המתאם בין הציון הכמותי לציון המילולי הוא 0.8.
א. חשבו את השונות המשותפת בין הציון הכמותי לציון המילולי.
ב. חשבו את התוחלת והשונות של סכום הציונים בחלק הכמותי ובחלק המילולי.
ג. חשבו את התוחלת והשונות של הפרש הציונים בין החלק הכמותי לחלק המילולי.
ד. עלות הבחינה 2000 שקלים. הוחלט לזכות שקל עבור כל נקודה שנצברה בחלק המילולי ושני שקלים עבור כל נקודה שנצברה בחלק הכמותי. מהי התוחלת ומהי השונות של עלות הבחינה נטו (העלות לאחר הזיכוי)?

3) נתון: $\operatorname{var}(X-2Y)=2$, $\operatorname{var}(X+2Y)=3$.
חשבו: $\operatorname{cov}(X,Y)$.

4) מטילים קובייה n פעמים. נגדיר את המשתנים הבאים:
X = מספר הפעמים שהתקבלה התוצאה 6.
Y = מספר הפעמים שהתקבלה התוצאה 5
בטאו את השונות המשותפת באמצעות n.

תשובות סופיות:

1) א. להלן טבלה: ב. תלויים. ג. מתואמים. ד. 0.1-.

$x \backslash y$	1	2	3	$P(y)$
2	0.2	0.1	0.3	0.6
3	0.2	0.1	0.1	0.4
$P(x)$	0.4	0.2	0.4	1

ה. תוחלת: 4.4, שונות: 0.84. ו. תוחלת: 0.4-, שונות: 1.24.

2) א. 240. ב. תוחלת: 190, שונות: 1105.
ג. תוחלת: 10 , שונות: 145 ד. תוחלת: 1710, שונות: 2785.

3) 0.125-.

4) $-\frac{n}{36}$.

סטטיסטיקה לאקטואריה

פרק 6 - המשתנה המקרי הדו ממדי הבדיד - שאלות מסכמות

תוכן העניינים

המשתנה המקרי הדו ממדי הבדיד – שאלות מסכמות:

רקע:

משתנים בלתי תלויים:

יהיו משתנים X ו- Y. הם יהיו משתנים בלתי תלויים אם עבור כל X ו- Y אפשריים מתקיים: $p(x=k, y=l) = p(x=k)\cdot p(y=l)$.

מקדם המתאם:

מגדירים את מקדם המתאם: $\rho = \dfrac{\operatorname{cov}(x,y)}{\sigma_x \cdot \sigma_y}$.

שונות משותפת:

מגדירים את השונות המשותפת:

$\operatorname{cov}(x,y) = E\left[(x-\mu_x)(y-\mu_y)\right] = E(x\cdot y) - E(x)\cdot E(y)$.

תכונות של השונות המשותפת:

1. $\operatorname{cov}(X,Y) = \operatorname{cov}(Y,X)$.
2. $\operatorname{cov}(X,X) = \operatorname{var}(X)$.
3. $\operatorname{cov}\left[(aX+b),(cY+d)\right] = a\cdot c\cdot \operatorname{cov}(X,Y)$.

משתנים בלתי מתואמים:

משתנים בלתי מתואמים הם משתנים שמקדם המתאם שלהם אפס וכדי שדבר כזה יקרה השונות המשותפת צריכה להתאפס.

השפעת טרנספורמציה לינארית על מקדם המתאם:

$$\rho\left[(aX+b),(cY+d)\right]=\begin{cases}\rho(X,Y) & \text{if} \quad a\cdot c>0\\ -\rho(X,Y) & \text{if} \quad a\cdot c<0\end{cases}$$

תוחלת ושונות של סכום משתנים:

$$E(X+Y)=E(X)+E(Y)\ V(X+Y)=V(X)+V(Y)+2\cdot\operatorname{cov}(X,Y)$$

קומבינציות לינאריות:

נגדיר קומבינציה ליניארית כללית באופן הבא: $W=(aX+b)+(cY+d)$.

אזי מתקיים:

$$E(W)=E\big((aX+b)+(cY+d)\big)=aE(X)+b+cE(Y)+d$$

$$V(W)=V\big((aX+b)+(cY+d)\big)=a^2V(X)+c^2V(Y)+2\cdot a\cdot c\cdot\operatorname{cov}(X,Y)$$

שאלות:

1) יש ליצור סיסמה בת 3 תווים. כל תו יכול להיבחר רק מתוך כלל התווים הבאים: A , B , C , 1 , 2 . יהי X מספר הפעמים שהספרה 1 מופיעה בסיסמה, ויהי Y מספר הפעמים שהספרה 1 מופיעה בקצה הסיסמה (שני הקצוות).

א. זהו את ההתפלגויות השוליות של X ו- Y כהתפלגויות מיוחדות.

ב. מצאו את ההתפלגות המשותפת של X ושל Y.

ג. מצאו את מקדם המתאם בין X ל- Y.

ד. מהו המתאם בין $2X$ ל- $3Y+5$?

2) במסיבת סוף שנה ישנו ארגז קרח ובתוכו 7 בקבוקי בירה:
4 "מכבי", 2 "גולדסטאר" ו-1 "טובורג".
קרן לקחה 3 בקבוקי בירה באקראי מתוך ארגז הקרח.
נסמן ב- X את מספר בקבוקי "מכבי " שנלקחו על ידי קרן,
ונסמן ב- Y את מספר בקבוקי "טובורג" שנלקחו על ידי קרן.

א. בנו את פונקציית ההסתברות המשותפת של X ושל Y.

ב. חשבו את התוחלת והשונות של X ושל Y .

ג. מצאו את השונות המשותפת של X ושל Y.

ד. נגדיר את W כמספר בקבוקי ה"גולדסטאר" שנלקחו על ידי קרן.
בטאו את W באמצעות X ו- Y, וחשבו את התוחלת והשונות של W על סמך התוצאות שהתקבלו בשני הסעיפים הקודמים בלבד.

ה. מהו מקדם המתאם בין מספר בקבוקי "מכבי" שנלקחו על ידי קרן, למספר בקבוקים שאינם "מכבי" שנלקחו על ידי קרן?

3) במגירה 6 זוגות נעליים. יהודה הוציא מהמגירה 4 נעליים (לא בהכרח זוגות באקראי. נסמן ב- W את מספר זוגות הנעליים שהוציא יהודה, ונסמן ב- R את מספר הנעליים השמאליות שהוציא יהודה.

א. מצא את ההתפלגות המשותפת של המשתנים שהוצגו.

ב. האם המשתנים שהוצגו בלתי תלויים?

ג. מצא את התפלגות מספר הנעליים השמאליות שהוצאו אם בסך הכול הוצא זוג נעלים יחיד על ידי יהודה.

ד. אם ידוע שהוצאו לפחות 3 נעליים שמאליות מה הסיכוי שהוצא לכל היותר זוג אחד?

4) בכד 5 כדורים כחולים, 4 כדורים לבנים ו-3 כדורים ירוקים. בוחרים באקראי וללא החזרה 3 כדורים. נגדיר את המשתנים הבאים:

X - מקבל את הערך 1 אם נבחר לפחות כדור אחד כחול, ו-0 אחרת.

Y - מספר הכדורים הלבנים שנבחרו.

א. חשבו את $P(X=1)$.

ב. בנו את פונקציית ההסתברות המשותפת של X ו- Y.

ג. מה התוחלת של Y, אם ידוע שלא הוצאו כדורים כחולים?

ד. מה השונות של X, אם ידוע שהוצא לכל היותר כדור לבן אחד?

5) ביום ההולדת הרביעי של טל הוא מחלק שלושה פרסים שונים באקראי ל-5 ילדים. בכל פעם שטל מחלק פרס הוא בוחר באקראי ילד מתוך ה-5 באופן אקראי ובלתי תלוי בבחירות הקודמות. נגדיר את המשתנים הבאים:

X - מספר הפרסים שקיבלה יוליה.

Y - מספר הילדים שלא קיבלו פרס.

א. בנו את פונקציית ההסתברות המשותפת והשוליות של X ו- Y.

ב. האם X ו- Y הם משתנים בלתי מתואמים?

ג. מצאו את התוחלת של $X \cdot Y^2$.

ד. מה מקדם המתאם בין מספר הפרסים שקיבלה יוליה, למספר הילדים שקיבלו פרס?

6) קבעו אילו מהטענות הבאות נכונות. נמקו.

א. אם שני משתנים הם מתואמים, אזי הם תלויים.

ב. אם שני משתנים הם תלויים, אזי הם מתואמים.

ג. אם שני משתנים הם בלתי תלויים, אזי הם בלתי מתואמים.

ד. אם שני משתנים הם בלתי מתואמים, אזי הם בלתי תלויים.

7) במקום עבודה 50 עובדים מתוכם 25 גברים ו-25 נשים. כל עובד נתבקש לבחור מתנה לחג. לכל עובד מוצגות 5 אופציות, מתוכן הוא צריך לבחור אחת. העובדים בוחרים מתנה באקראי ובאופן בלתי תלוי זה בזה.

נסמן X_i - מספר הגברים שבחרו במתנה i.

נסמן Y_i - מספר הנשים שבחרו במתנה i.

א. האם X_1 ו- Y_1 הם משתנים בלתי תלויים? אין צורך לחשב רק להסביר.

ב. האם X_1 ו- X_2 הם משתנים בלתי תלויים? אין צורך לחשב רק להסביר.

ג. מהי ההתפלגות של X_1+X_2?

ד. האם המתאם בין X_1 ו- X_2 מלא או חלקי? חיובי או שלילי? אין צורך לחשב רק להסביר.

8) הוכיחו את הזהות הבאה עבור שלושת המשתנים: X, Y ו- Z:
$\operatorname{cov}(X+Y,Z)=\operatorname{cov}(X,Z)+\operatorname{cov}(Y,Z)$.

9) מספר העלים שנושרים בסתיו מהעץ בגינה מתפלג פואסונית עם תוחלת של 50 עלים בדקה. נסמן ב- Y את מספר העלים שנושרים מהעץ בין 12:00 ל-12:10, ונסמן ב- Q את מספר העלים שנושרים בין 12:05 ל-12:30.

א. חשבו את: $\operatorname{cov}(4Y,Q+6)$.

ב. מה המתאם בין Y ל- Q?

10) בסל יש 20 כדורים אדומים, 20 ירוקים ו-20 כחולים. מוציאים באקראי מהסל 20 כדורים. מצאו את מקדם המתאם בין מספר הכדורים האדומים שהוצאו למספר הכדורים הירוקים שהוצאו.

11) נתון ש: $Y\sim B(1,p)$ כאשר $0<P<1$.
הוכיחו שאם מתקיים: $P(X=x|Y=0)=P(X=x|Y=1)$ לכל X,
אז X ו- Y הם משתנים בלתי תלויים.

12) נתון ש- $X\sim B(n,p)$ וכן: $Y\sim B(m,p)$, שאינם תלויים זה בזה.
הוכיחו שמתקיים: $X\,|\,X+Y=k\sim HG(n+m,n,k)$.

תשובות סופיות:

1) א. $X\sim B\left(3,\frac{1}{5}\right)$, $Y\sim B\left(2,\frac{1}{5}\right)$.

ב. להלן טבלה: ג. 0.816. ד. 0.816.

X/Y	0	1	2	3	P_Y
0	$\frac{64}{125}$	$\frac{16}{125}$	0	0	$\frac{80}{125}$
1	0	$\frac{32}{125}$	$\frac{8}{125}$	0	$\frac{40}{125}$
2	0	0	$\frac{4}{125}$	$\frac{1}{125}$	$\frac{5}{125}$
P_X	$\frac{64}{125}$	$\frac{48}{125}$	$\frac{12}{125}$	$\frac{1}{125}$	1

2) א. להלן טבלה: ב. $E(X)=\frac{12}{7}$, $V(X)=\frac{24}{49}$, $E(Y)=\frac{3}{7}$, $V(Y)=\frac{12}{49}$.

X/Y	0	1	2	3	P_Y
0	0	$\frac{3}{35}$	$\frac{12}{35}$	$\frac{4}{35}$	$\frac{20}{35}$
1	$\frac{1}{35}$	$\frac{8}{35}$	$\frac{6}{35}$	0	$\frac{15}{35}$
P_X	$\frac{1}{35}$	$\frac{12}{35}$	$\frac{18}{35}$	$\frac{4}{35}$	1

ג. $-\frac{8}{49}$. ד. $E(W)=\frac{6}{7}$, $V(W)=\frac{20}{49}$. ה. ‎-1.

3) א. להלן טבלה: ב. המשתנים תלויים.

R/W	0	1	2	P_R
0	$\frac{15}{495}$	0	0	$\frac{15}{495}$
1	$\frac{60}{495}$	$\frac{60}{495}$	0	$\frac{120}{495}$
2	$\frac{90}{495}$	$\frac{120}{495}$	$\frac{15}{495}$	$\frac{225}{495}$
3	$\frac{60}{495}$	$\frac{60}{495}$	0	$\frac{120}{495}$
4	$\frac{15}{495}$	0	0	$\frac{15}{495}$
P_W	$\frac{240}{495}$	$\frac{240}{495}$	$\frac{15}{495}$	1

ג. להלן טבלה:

$R/w=1$	1	2	3
$P(R/w=1)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

ד. 1.

4) א. $\frac{185}{220}$.   ב. להלן טבלה: ג. 1.714.   ד. 0.071.

X/Y	0	1	P_Y
0	$\frac{1}{220}$	$\frac{55}{220}$	$\frac{56}{220}$
1	$\frac{12}{220}$	$\frac{100}{220}$	$\frac{112}{220}$
2	$\frac{18}{220}$	$\frac{30}{220}$	$\frac{48}{220}$
3	$\frac{4}{220}$	0	$\frac{4}{220}$
P_X	$\frac{35}{220}$	$\frac{185}{220}$	1

5) א. להלן טבלה: ב. X ו-Y בלתי מתואמים.

X/Y	0	1	2	3	P_Y
2	$\frac{24}{125}$	$\frac{36}{125}$	0	0	$\frac{60}{125}$
3	$\frac{36}{125}$	$\frac{12}{125}$	$\frac{12}{125}$	0	$\frac{60}{125}$
4	$\frac{4}{125}$	0	0	$\frac{1}{125}$	$\frac{5}{125}$
P_X	$\frac{64}{125}$	$\frac{48}{125}$	$\frac{12}{125}$	$\frac{1}{125}$	1

ג. 4.128.   ד. 0.

6) א. נכון.   ב. לא נכון.   ג. נכון.   ד. לא נכון.

7) א. בלתי תלויים. ב. תלויים. ג. $x_1+x_2 \sim B\left(n=25, p=\frac{2}{5}\right)$.

ד. חלקי ושלילי.

8) שאלת הוכחה.

9) א. 1000. ב. 0.316.

10) -0.5.

11) שאלת הוכחה.

12) שאלת הוכחה.

סטטיסטיקה לאקטואריה

פרק 7 - המשתנה המקרי הדו ממדי הרציף

תוכן העניינים

משתנה מקרי דו ממדי רציף:

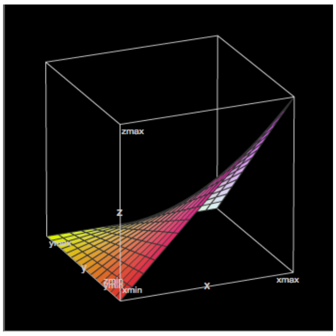
רקע:

יהיו X ו- Y משתנים מקריים רציפים המוגדרים בתחום R מסוים.
פונקציית הצפיפות המשותפת שלהם תסומן על ידי: $f(x,y)$.
פונקציית צפיפות משותפת צריכה לקיים את שני התנאים הבאים:

1) $f(x,y)\geq 0$ לכל $(x,y)\in R$.

2) $\iint\limits_R f(x,y)dxdy=1$.

דוגמה (פתרון בהקלטה):



נתונה הפונקציה: $f(x,y)=\begin{cases}4x(1-y) & 0\leq x,y\leq 1\\ 0 & else\end{cases}$.

הראו שפונקציה זו יכולה להיות פונקציית צפיפות משותפת.

פונקציית צפיפות שולית:

פונקציית הצפיפות השולית של X תתקבל באופן הבא: $f(x)=\int\limits_{-\infty}^{\infty} f(x,y)dy$.

פונקציית הצפיפות השולית של Y תתקבל באופן הבא: $f(y)=\int\limits_{-\infty}^{\infty} f(x,y)dx$.

דוגמה (פתרון בהקלטה):

מצאו לפונקציית הצפיפות: $f(x,y)=\begin{cases}4x(1-y) & 0\leq x,y\leq 1\\ 0 & else\end{cases}$

את פונקציית הצפיפות השולית של X, וחשבו את $E(X)$ דרכה.

אי-תלות בין משתנים רציפים:

X ו- Y יהיו משתנים מקרים בלתי תלויים, אם עבור כל X ו- Y בתחום ההגדרה R מתקיים ש: $f(x,y)=f(x)\cdot f(y)$.

דוגמה (פתרון בהקלטה):

האם X ו- Y, המתפלגים לפי פונקציית הצפיפות המשותפת:

$$f(x,y)=\begin{cases}4x(1-y) & 0\le x,y\le 1\\ 0 & else\end{cases},$$

הם משתנים בלתי תלויים?

חישוב הסתברויות עבור משתנה מקרי רציף דו ממדי:

הנפח הכלוא מתחת למשטח $f(x,y)$ בתחום מסוים ייתן את ההסתברות ש- X ו- Y יהיו בתחום הזה: $P[(x,y)\in A]=\iint\limits_A f(x,y)dxdy$.

דוגמה (פתרון בהקלטה):

משתנים מתפלגים לפי פונקציית הצפיפות: $f(x,y)=\begin{cases}4x(1-y) & 0\le x,y\le 1\\ 0 & else\end{cases}$.

חשבו את הסיכוי: $P(X<0.5\cap Y<0.5)$.

פונקציית התפלגות מצטברת משותפת:

פונקציית התפלגות מצטברת משותפת הינה פונקציה של שני משתנים רציפים המחזירה את הסיכוי שהמשתנים יהיו קטנים מערכים מסוימים:

$$F(s,t)=P(X\le s\cap Y\le t)=\int_{-\infty}^{t}\int_{-\infty}^{s} f(x,y)dxdy.$$

דוגמה (פתרון בהקלטה):

משתנים מתפלגים לפי פונקציית הצפיפות: $f(x,y)=\begin{cases}4x(1-y) & 0\le x,y\le 1\\ 0 & else\end{cases}$.

מצאו את פונקציית ההתפלגות המצטברת המשותפת

ועל פיה חשבו את הסיכוי: $P(X<0.5\cap Y<0.5)$.

פונקציית צפיפות מותנית:

אם ל- X ול- Y ישנה פונקציית צפיפות משותפת $f(x,y)$, אז מגדירים את פונקציית הצפיפות המותנית של X , בהינתן ש- $Y=y$ לכל ערכי y המקיימים: $f(y)>0$ על ידי: $f(x|y)=\dfrac{f(x,y)}{f(y)}$.

באופן דומה, פונקציית הצפיפות המותנית של Y בהינתן ש- $X=x$ לכל ערכי x המקיימים: $f(x)>0$ על ידי: $f(y|x)=\dfrac{f(x,y)}{f(x)}$.

דוגמה (פתרון בהקלטה):

משתנים מתפלגים לפי פונקציית הצפיפות: $f(x,y)=\begin{cases}4x(1-y) & 0\le x,y\le 1\\ 0 & else\end{cases}$.

מצאו את: $f(x|y)$.

תוחלת מותנית:

ל- X ול- Y ישנה פונקציית צפיפות משותפת $f(x,y)$.

התוחלת של X בהינתן ש- $Y=y$ תהיה: $E(X|Y=y)=\int_{-\infty}^{-\infty} x\cdot f(x|y)dx$.

ובאופן דומה, התוחלת של Y בהינתן ש $X=x$ תהיה:

$E(Y|X=x)=\int_{-\infty}^{-\infty} y\cdot f(y|x)dy$.

דוגמה (פתרון בהקלטה):

משתנים מתפלגים לפי פונקציית הצפיפות: $f(x,y)=\begin{cases}4x(1-y) & 0\le x,y\le 1\\ 0 & else\end{cases}$.

מצאו את: $E(X|Y)$.

שאלות:

1) נתונה פונקציית הצפיפות הבאה: $f(x,y)=x+y$, המוגדרת בתחום שבו: $0\le x\le 1$ וגם: $0\le y\le 1$. הוכיחו שמדובר בפונקציית צפיפות.

2) נתונה פונקציית הצפיפות הבאה: $f(x,y)=Ax(x-y)$, המוגדרת בתחום שבו: $0\le x\le 2$ וגם: $-x\le y\le x$. מצאו את ערכו של הפרמטר A.

3) נתונה פונקציית הצפיפות הבאה: $f(x,y)=\frac{(x\cdot y)^3+x\cdot y}{C}$, המוגדרת בתחום שבו: $0\le x\le 1$ וגם: $0\le y\le 1$.

א. מצאו את ערכו של C.
ב. מצאו את $f(y)$.
ג. האם X ו- Y הינם משתנים בלתי תלויים?

4) משתנה מקרי דו ממדי מתפלג לפי פונקציית הצפיפות הבאה: $f(x,y)=\frac{1}{800}$, המוגדרת בתחום שבו: $60\le x\le y$ וגם: $60\le y\le 100$.

א. הראו שפונקציה זו מקיימת את התנאים של פונקציית צפיפות.
ב. מצאו את פונקציית הצפיפות השולית של Y.
ג. חשבו את $V(X)$, $E(X)$.
ד. האם X ו- Y הם משתנים בלתי תלויים?
ה. חשבו את מקדם המתאם בין X ל- Y.
ו. חשבו את הסיכוי: $P(Y>X+10)$.

5) משתנה מקרי דו ממדי מתפלג לפי פונקציית הצפיפות הבאה: $f(x,y)=\lambda\mu\cdot e^{-(\lambda x+\mu y)}$, המוגדרת בתחום שבו: $x,y>0$.

א. מצאו את פונקציית הצפיפות של X ואת פונקציית הצפיפות של Y.
ב. האם X ו- Y הם משתנים תלויים?
ג. מהו מקדם המתאם בין X ל- Y?
ד. חשבו את הסיכוי: $P(Y>X)$.

6) Y הינו משתנה מקרי אחיד רציף המתפלג בקטע: $[2,4]$.

בנוסף, נתון ש- X הינו משתנה מקרי רציף המקיים: $0 \le x \le y$, $f(x|y)=\frac{2x}{y^2}$.

מצאו את השונות המשותפת של X ו- Y.

7) נתונים שני משתנים מקרים רציפים X ו- Y. פונקציות הצפיפות המשותפות שלהם היא: $f(x,y)=\begin{cases} x & 0<y<1 \quad 1-y \le x \le 1+y \\ 0 & else \end{cases}$.

א. מצאו את $f(x)$.

ב. מצאו את $f(y|x)$.

ג. מצאו את $E(Y|X)$.

8) יהיו X ו- Y משתנים רציפים המתפלגים אחיד בתוך משולש שקדקודיו: $(0,1)$, $(-1,0)$, $(-1,2)$.

א. רשמו את פונקציית הצפיפות המשותפת.

ב. מצאו את פונקציית הצפיפות השולית של X ו- Y.

ג. חשבו את התוחלת של X ו- Y.

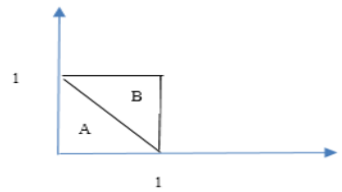
ד. האם X ו- Y משתנים בלתי מתואמים?

ה. האם X ו- Y משתנים בלתי תלויים?

9) פונקציית צפיפות משותפת מקבלת ערך אחיד באופן הבא:
הצפיפות על פני משולש A הינה 1.5 והצפיפות על פני משולש B היא 0.5.
האם פונקציית הצפיפות המשותפת היא לגיטימית?

א. מצאו את $f(x)$.

ב. מצאו את $f(x|y)$.



10) נתונה פונקציית הצפיפות המשותפת: $f(x,y)=cx$. פונקציה זו מוגדרת בתחום שבו: $0 \le x \le 1$ וכן: $0 \le y \le x^2$.

א. מצאו את הקבוע C.

ב. חשבו את ההסתברות ש- $6Y<1-X$.

11) נתונים X ו- Y שני משתנים מקריים רציפים כך ש: $Y \sim U(0,1)$
ו-$X|Y=y \sim U\left(0,\sqrt{y}\right)$. חשבו את: $E(Y|X=0.5)$.

12) נתונה פונקציית הצפיפות: $f(x,y)=2e^{-x}\cdot e^{-2y}$ בתחום שבו: $x,y\geq 0$.
חשבו את הסיכוי: $P(X<Y)$.

13) נתונה פונקציית הצפיפות המשותפת: $f(x,y)=\dfrac{e^{-y}\cdot e^{-\frac{x}{y}}}{y}$, המוגדרת לרביע הראשון. חשבו את: $P(X>1|Y=2)$.

14) יוסי וערן עובדים באותו המשרד. הם מגיעים לעבודה בכל יום בין 8:00 ל-9:00.
נניח שזמן ההגעה של כל אחד מתפלג אחיד ובאופן בלתי תלוי זה בזה.
מה הסיכוי שיוסי יצטרך לחכות לערן יותר מ-10 דקות?

15) נתונים שני משתנים מקרים רציפים: $X\sim N(Y,1)$ ו-$Y\sim U(0,2)$.
א. מצאו את פונקציית הצפיפות המשותפת של X ו- Y.
ב. מצאו את $E\left(X^2|Y\right)$.
ג. מצאו את $E(X)$.

16) פונקציית הצפיפות המשותפת של X ו- Y היא: $f(x,y)=1$.
פונקציה זו מוגדרת בתחומי: $0\leq x,y\leq 1$.
הוכיחו ש: $E\left(|X-Y|^n\right)=\dfrac{2}{(n+1)(n+2)}$.

17) $X\sim \exp(1)$ וכן: $Y\sim \exp(1)$, הינם משתנים מקרים בלתי תלויים.
נגדיר את: $Z=\dfrac{X}{X+Y}$.
הוכיחו: $Z\sim U(0,1)$.

תשובות סופיות:

1) שאלת הוכחה.

2) $A=\frac{1}{8}$.

3) א. $\frac{5}{16}$. ב. $f(y)=0.8y^3+1.6y$. ג. תלויים.

4) א. שאלת הוכחה. ב. $f(y)=\frac{y-60}{800}$. ג. $V(X)=88\frac{8}{9}$, $E(X)=73\frac{1}{3}$.
ד. לא. ה. 0.5. ו. 0.5625.

5) א. $f(x)=\lambda e^{-\lambda x}$, $f(y)=\mu e^{-\mu y}$. ב. לא.
ג. 0. ד. $\frac{\lambda}{\lambda+\mu}$.

6) $\frac{2}{9}$.

7) א. $f(x)=\begin{cases} x^2 & 0\le x<1 \\ 2x-x^2 & 1\le x\le 2 \\ 0 & else \end{cases}$.

ב. $f(y|x)=\begin{cases} \frac{1}{x} & 0\le x<1 \quad 1-x<y<1 \\ \frac{1}{2-x} & 1\le x\le 2 \quad x-1<y<1 \\ 0 & else \end{cases}$. ג. $E(y|x)=\begin{cases} 1-\frac{x}{2} & 0\le x<1 \\ \frac{x}{2} & 1\le x\le 2 \\ 0 & else \end{cases}$.

8) א. $f(x,y)=\begin{cases} 1 & 1+x<y<1-x \quad -1<x<0 \\ 0 & else \end{cases}$.

ב. $f(x)=\begin{cases} -2x & -1<x<0 \\ 0 & else \end{cases}$, $f(y)=\begin{cases} y & 0\le y<1 \\ 2-y & 1\le y\le 2 \\ 0 & else \end{cases}$.

ג. $E(Y)=1$, $E(X)=-\frac{2}{3}$. ד. כן.
ה. לא.

9) א. כן.
ב. $f(x)=\begin{cases}1.5-x & 1<x<0\\ 0 & else\end{cases}$.

ג. $f(x|y)=\begin{cases}\frac{1.5}{1.5-y} & 0\le x<1-y\\ \frac{0.5}{1.5-y} & 1-y\le x\le 1\\ 0 & else\end{cases}$.

10) א. 4.
ב. 0.0947.

11) $\frac{7}{12}$.

12) $\frac{1}{3}$.

13) $e^{-\frac{1}{2}}$.

14) $\frac{25}{72}$.

15) א. $f(x,y)=\begin{cases}\frac{1}{2\sqrt{2\pi}}e^{-\frac{(x-y)^2}{2}} & 0<y<2\\ 0 & else\end{cases}$.
ב. y^2+1.
ג. 1.

16) שאלת הוכחה.

17) שאלת הוכחה.

סטטיסטיקה לאקטואריה

פרק 8 - נוסחת התוחלת השלמה

תוכן העניינים

נוסחת התוחלת השלמה:

רקע:

כאשר התפלגות של משתנה X תלויה במשתנה אחר Y, מתקיים: $E(X)=E\left[E(X/Y)\right]$.

עבור משתנה Y בדיד כלשהו: $E(X)=\sum_{y}E(X|Y=y)\cdot P(Y=y)$.

עבור משתנה Y רציף כלשהו: $E(X)=\int_{-\infty}^{\infty}E(X|Y=y)\cdot f(y)dy$.

דוגמה (הפתרון בהקלטה):

מרצה מלמד שתי כיתות. הוא מעוניין לבדוק 5 מחברות בחינה.
בכיתה מספר 1: 10 בנים ו-20 בנות.
בכיתה מספר 2: 15 בנים ו-15 בנות.
המרצה בוחר כיתה באופן מקרי וממנה בוחר 5 מחברות בחינה באקראי.
יהי X מספר מחברות הבחינה של בנים שנבחרו.
יש לחשב את $E(X)$.

שאלות:

1) בכד 2 כדורים ירוקים, 4 כדורים כחולים ו-4 אדומים. שולפים כדור באקראי.
אם הכדור ירוק מטילים קובייה עד אשר מתקבלת התוצאה 1.
אם הכדור כחול מטילים קובייה עד אשר מתקבלת התוצאה 5.
אם הכדור אדום מטילים קובייה עד אשר מתקבלת תוצאה זוגית.
חשבו את תוחלת מספר הפעמים שהקובייה הוטלה.

2) מטילים n מטבעות ומוציאים מהמשחק את כל המטבעות שהראו ראש.
כעת מטילים את כל המטבעות שנותרו. בטאו באמצעות n את תוחלת מספר הראשים שהתקבלו בסבב השני של ההטלות.

3) בהגרלה מבצעים את התהליך הבא: בסיכוי 0.25 מגרילים מספר ממכונה A בסיכוי 0.75 מגרילים מספר ממכונה B. במכונה A המספר המוגרל מתפלג פואסונית עם תוחלת 6. במכונה B המספר המוגרל מתפלג פואסונית עם תוחלת 2. אם הוגרל המספר 0 זוכים ב15₪. אחרת זוכים ב-50₪.
חשבו את תוחלת סכום הזכיה.

4) נתון ש- $Y/X \sim U(0,X)$, כאשר פונקציית הצפיפות של X
הינה:

$$f(x)=\begin{cases}0.1 & 2\le x\le 6\\ 0.2 & 6<x\le 9\\ 0 & \text{אחרת}\end{cases}.$$

חשבו את $E(Y)$.

5) בתכנית הריאליטי "המרוץ למיליון" מגיעים לנקודה שבה שלוש אפשרויות בפני המתמודדים: נקודה A - שבה חוזרים אחרי 1 שעות לנקודת המוצא. נקודה B - שבה חוזרים אחרי 2 שעות לנקודת המוצא ונקודה C - המובילה תוך 2 שעות לנקודת הסיום. המתמודדים בוחרים בכל פעם את הנקודה באופן מקרי. נסמן ב- X את זמן ההגעה לנקודת הסיום.
חשבו את $E(Y)$.

תשובות סופיות:

1) 4.4.

2) $\frac{n}{4}$.

3) 46.4.

4) 3.05.

5) 5.

סטטיסטיקה לאקטואריה

פרק 9 - נוסחת השונות השלמה (שונות של משתנה מותנה)

תוכן העניינים

נוסחת השונות השלמה (המותנית):

רקע:

כאשר התפלגות של משתנה X תלויה במשתנה אחר Y, מתקיים: $E(X)=E\left[E(X|Y)\right]$.

כמו כן, מתקיים לגבי השונות: $\text{Var}(X)=\text{Var}\left(E[X|Y]\right)+E\left[\text{Var}(X|Y)\right]$.

דוגמה (הפתרון בהקלטה):

מרצה מלמד שתי כיתות. הוא מעוניין לבדוק 5 מחברות בחינה.
בכיתה מספר 1: 10 בנים ו-20 בנות.
בכיתה מספר 2: 15 בנים ו-15 בנות.
המרצה בוחר כיתה באופן מקרי וממנה בוחר 5 מחברות בחינה באקראי.
יהי X מספר מחברות הבחינה של בנים שנבחרו.
יש לחשב את $V(X)$.

שאלות:

1) בכד 2 כדורים ירוקים, 4 כדורים כחולים ו-4 אדומים. שולפים כדור באקראי.
אם הכדור ירוק מטילים קובייה עד אשר מתקבלת התוצאה 1.
אם הכדור כחול מטילים קובייה עד אשר מתקבלת התוצאה 5.
אם הכדור אדום מטילים קובייה עד אשר מתקבלת תוצאה זוגית.
חשבו את התוחלת והשונות של מספר הפעמים שהקובייה הוטלה.

2) נטיל קובייה: $Y+4$ פעמים. נתון ש- $Y\sim P(4)$.
נגדיר את X כמספר הפעמים שהתקבלה התוצאה 2.
א. מצאו את התוחלת של X.
ב. מצאו את השונות של X.

3) נתון ש- $Y|X\sim U(0,X)$, כאשר פונקציית הצפיפות של X
הינה: $f(x)=\begin{cases}0.1 & 2\le x\le 6\\ 0.2 & 6<x\le 9\\ 0 & \text{אחרת}\end{cases}$.
חשבו את $V(Y)$.

תשובות סופיות:

1) תוחלת: 4.4, שונות: 22.64.

2) תוחלת: $\frac{4}{3}$ שונות: $\frac{11}{9}$.

3) 4.4.

סטטיסטיקה לאקטואריה

פרק 10 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

התפלגות ממוצע המדגם ומשפט הגבול המרכזי:

רקע:

בפרק זה נדון בהתפלגות של ממוצע המדגם: $\bar{x}=\frac{\sum x_i}{n}$.
מכיוון שממדגם למדגם אנו יכולים לקבל ממוצע מדגם שונה, אזי ממוצע המדגם הוא משתנה מקרי ויש לו התפלגות.
גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים.
להלן רשימה של פרמטרים החשובים לפרק זה:
ממוצע האוכלוסייה נסמן ב- μ (נקרא גם תוחלת).
שונות אוכלוסייה נסמן ב- σ^2.
סטיית תקן של אוכלוסייה: σ.

תכונות התפלגות:

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה: $E(\bar{x})=\mu_{\bar{x}}=\mu$.
שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב- n.
תכונה זו נכונה רק במדגם מקרי: $V(\bar{x})=\sigma_{\bar{x}}^2=\frac{\sigma^2}{n}$.
יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.
אם נוציא שורש לשונות נקבל סטיית תקן של ממוצע המדגם שנקראת גם
טעות תקן: $\sigma(\bar{x})=\sqrt{\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}}$.

דוגמה (פתרון בהקלטה):

השכר הממוצע במשק הינו 9000 ₪ עם סטיית תקן של 4000. דגמו באקראי 25 עובדים.

א. מיהי אוכלוסיית המחקר? מהו המשתנה הנחקר?
ב. מהם הפרמטרים של האוכלוסייה?
ג. מה התוחלת ומהי סטית התקן של ממוצע המדגם?

דגימה מהתפלגות נורמאלית:

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע μ ושונות σ^2.

ממוצע המדגם גם יתפלג נורמאלית: $Z_{\bar{x}} = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$, $\bar{x} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

דוגמה (פתרון בהקלטה):

משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.

מה ההסתברות שבמדגם של 4 תינוקות אקראיים בעת הולדתם המשקל הממוצע של התינוקות יהיה מתחת ל-3.5 ק"ג?

משפט הגבול המרכזי:

אם אוכלוסייה מתפלגת כלשהו עם ממוצע μ ושונות σ^2 אזי עבור מדגם מספיק גדול $(n \geq 30)$ ממוצע המדגם מתפלג בקירוב נורמאלי: $\bar{x} \sim\succ N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

דוגמה (פתרון בהקלטה):

משקל חפיסת שוקולד בקו ייצור מתפלג עם ממוצע 100 גרם וסטיית תקן של 4 גרם.
דגמו מקו הייצור 36 חפיסות שוקולד אקראיות.
מה ההסתברות שהמשקל הממוצע של חפיסות השוקולד שנדגמו יהיה מתחת ל-102 גרם?

שאלות:

1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.

א. מיהי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו תוחלת ממוצע המדגם?

ו. מהי טעות התקן?

2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב מסוים:

מספר המשפחות	מספר מקלטים
500	0
2500	1
3500	2
3000	3
500	4
סך הכול $N=10000$	

א. בנו את פונקציית ההסתברות של X.

ב. חשבו את התוחלת, השונות וסטיית התקן של X.

ג. אם נדגום 4 משפחות מהישוב עם החזרה מה תהיה התוחלת, מהי השונות ומהי סטיית התקן של ממוצע המדגם?

3) אם נטיל קובייה פעמיים ונתבונן בממוצע התוצאות שיתקבלו, מה תהיה התוחלת ומה תהיה סטיית התקן של ממוצע זה?

4) משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.

א. מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?

נתון כי ביום מסוים נולדו 4 תינוקות.

ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק״ג?

ג. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק״ג?

ד. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?

ה. הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?

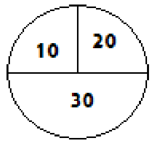
5) הגובה של המתגייסים לצה״ל מתפלג נורמאלית עם תוחלת של 175 ס״מ וסטיית תקן של 10 ס״מ. ביום מסוים התגייסו 16 חיילים.
   א. מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס״מ?
   ב. מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס״מ?
   ג. מה ההסתברות שהגובה הממוצע שלהם יסטה מתחולת הגבהים בפחות מ-5 ס״מ?
   ד. מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

6) הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים.
   לצורך הפתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.
   א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
   ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?
   ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
   ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?

7) נפח היין בבקבוק מתפלג נורמאלית עם תוחלת של 750 סמ״ק וסטיית תקן של 10 סמ״ק.
   א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ״ק?
   ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ״ק?
   ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ״ק?
   ד. בקבוקיי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?

8) משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4.
   א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
   ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
   ג. הסבר את ההבדל בתשובות של שני הסעיפים.

9) בקזינו ישנה רולטה. על הרולטה רשומים המס׳ הבאים כמוראה בשרטוט:



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.

א. בנו את פונקציית ההסתברות של סכום הזכיה במשחק בודד.

ב. מה התוחלת ומה השונות של סכום הזכייה?

ג. אם האדם ישחק את המשחק 5 פעמים מה התוחלת ומה השונות של ממוצע סכום הזכייה בחמשת המשחקים?

ד. אם האדם משחק את המשחק 50 פעם מה ההסתברות שבסה״כ יזכה ב-1050 ₪ ומעלה?

10) לפי הערכות הלשכה המרכזית לסטטיסטיקה השכר הממוצע במשק הוא 8000 ₪ עם סטיית תקן של 3000 ₪.
מה ההסתברות שבמדגם מקרי של 100 עובדים השכר הממוצע יהיה יותר מ-8500 ₪?

11) קובייה הוטלה 50 פעמים.
מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72?

12) אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס״מ וסטיית תקן של 10 ס״מ.

א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל 78 ס״מ?

ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס״מ. מה ההסתברות ש-100 המוטות יספיקו למלאכה?

ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס״מ. היעזרו במשפט הגבול המרכזי.

13) נתון משתנה מקרי בדיד בעל פונקציית ההסתברות הבאה:

X	8	6	4	2
$P(X)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

מתוך התפלגות זו נלקח מדגם מקרי בגודל 50.
מה הסיכוי שממוצע המדגם יהיה קטן מ-5?

14) נתון ש- $X \sim N\left(\mu,\sigma^2\right)$. דגמו 5 תצפיות מאותה התפלגות והתבוננו בממוצע המדגם $\bar{X}$. לכן: $P\left(\bar{X} > \mu\right)$ יהיה (בחרו בתשובה הנכונה):

א. 0.
ב. 0.5.
ג. 1.
ד. לא ניתן לדעת.

15) נתון ש- X מתפלג כלשהו עם תוחלת: μ ושונות σ^2.
החליטו לבצע מדגם בגודל 200 מתוך ההפלגות הנתונה לפי משפט הגבול המרכזי מתקיים (בחרו בתשובה הנכונה):

א. $X \sim N\left(\mu,\frac{\sigma^2}{200}\right)$.
ב. $\mu \sim N\left(\mu,\frac{\sigma^2}{200}\right)$.
ג. $\bar{X} \sim N\left(\mu,\sigma^2\right)$.
ד. $\bar{X} \sim N\left(\mu,\frac{\sigma^2}{200}\right)$.

16) נתון ש- $X \sim N\left(\mu,\sigma^2\right)$. אם נדגום n תצפיות מתוך ההתפלגות ונגדיר: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$,
אזי (בחרו בתשובה הנכונה):

א. μ ו- $\bar{X}$ יהיו משתנים מקריים.
ב. μ יהיה משתנה מקרי ו- $\bar{X}$ קבוע.
ג. $\bar{X}$ יהיה משתנה מקרי ו- μ קבוע.
ד. μ ו- $\bar{X}$ יהיו קבועים.

17) משקל חפיסת שוקולד בקו ייצור מתפלג עם ממוצע 100 גרם. החפיסות נארזות בקרטון המכיל 36 חפיסות שוקולד אקראיות. ההסתברות שהמשקל הממוצע של חפיסות השוקולד בקרטון יהיה מעל 99 גרם הוא 0.9932.

א. מהי סטיית התקן של משקל חפיסת שוקולד בודדת?
ב. מה הסיכוי שמתוך 4 קרטונים בדיוק קרטון אחד יהיה עם משקל ממוצע לחפיסה הנמוך מ-100 גרם?

18) משתנה מקרי כלשהו מתפלג עם סטיית תקן של 20.
מה הסיכוי שאם נדגום 100 תצפיות בלתי תלויות מאותה התפלגות אזי ממוצע המדגם יסטה מתוחלתו בפחות מ-2?

19) מספר המכוניות הנכנסות לחניון "בציר" במשך היום מתפלג פואסונית עם קצב של מכונית אחת לדקה. שומר מסר נתונים על מספר המכוניות שנכנסות בכל שעה לגבי 40 שעות שאסף נתונים.
מה ההסתברות שממוצע מספר המכוניות שנכנסו לחניון לשעה בשעות אלה יהיה לפחות 63?

20) הוכיחו שאם משתנה מתפלג כלשהו עם תוחלת μ ושונות σ^2, ומבצעים מדגם בגודל n של תצפיות בלתי תלויות מהמשתנה, אזי מתקיימות התכונות הבאות לגבי ממוצע המדגם: $E(\overline{x})=\mu$ ו- $V(\overline{x})=\frac{\sigma^2}{n}$.

תשובות סופיות:

1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון.
ג. ממוצע: 78, סטיית תקן: 15. ד. 2.
ה. 78. ו. 10.6.

2) א. להלן טבלה:

X	0	1	2	3	4
$P(X)$	0.05	0.25	0.35	0.3	0.05

ב. $\mu = 2.05$, $\sigma^2 = 0.9475$, $\sigma = 0.973$

ג. $\mu_{\bar{x}} = 2.05$, $\sigma_{\bar{x}}^2 = 0.2369$, $\sigma(\overline{X}) = 0.486$

3) $\mu_{\bar{x}} = 3.5$, $\sigma(\overline{X}) = 1.21$.

4) א. 0.8413. ב. 0.0013. ג. 0. ד. 0.1974.

5) א. 0. ב. 0. ג. 0.9544. ד. 178.205.

6) א. 0.0465. ב. 27.71. ג. 0.2628. ד. התשובה הייתה קטנה.

7) א. 0. ב. 0.1587. ג. 0.1587. ד. 0.5.

8) א. 0.5468. ב. 0.6826.

9) א. להלן טבלה:

X	10	20	30
$P(X)$	0.25	0.25	0.5

ב. התוחלת: 22.5, השונות: 68.75.

ג. התוחלת: 22.5, השונות: 13.75. ד. 0.8997.

10) 0.0475.

11) 0.1814.

12) א. 0.9772. ב. 0.0228. ג. 271.

13) 0.5.

14) ב'.

15) ד'.

16) ג'.

17) א. 2.429. ב. 0.25.

18) 0.6826.

19) 0.0071.

20) שאלת הוכחה.

התפלגות סכום תצפיות בלתי תלויות ומשפט הגבול המרכזי:

רקע:

כעת נדון בסטטיסטי המבטא את סכום התצפיות במדגם: $T=\sum_{i=1}^{n} X_i$.

כאשר כל התצפיות נדגמו באקראי מאותה אוכלוסייה, כלומר, היו: $X_1,...,X_n$ -

משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה μ ושונותה σ^2 אזי:

התוחלת והשונות של סכום התצפיות: $E(T)=n\mu\ ,\ V(T)=n\sigma^2$.

דגימה מתוך התפלגות נורמלית:

אם: $X\sim N(\mu,\sigma^2)$, אזי: $Z=\frac{T-n\mu}{\sqrt{n\sigma^2}}\ ,\ T\sim N(n\mu,n\sigma^2)$.

משפט הגבול המרכזי:

אם X מתפלג כלשהו וידוע כי: $E(X)=\mu\ ,V(X)=\sigma^2$,

אזי עבור מדגם מספיק גדול (לפחות 30): $T\ \tilde{\succ}\ N(n\mu,n\sigma^2)$.

דוגמה (פתרון בהקלטה):

בעיר מסוימת המשכורת הממוצעת של עובד הינה 8000 ₪. עם סטיית תקן של 2000 ₪. נדגמו 100 עובדים מהעיר שמפקידים את משכורותיהם לסניף בנק.

1. מה התוחלת וסטיית התקן של סך המשכורות שיופקדו לסניף הבנק על ידי העובדים הללו?
2. מה ההסתברות שלסניף יופקד פחות מ-780 אלף ₪ ע"י אותם עובדים?

שאלות:

1) המשקל באוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת של 60 ק״ג וסטיית תקן של 10 ק״ג.

א. מה הסיכוי שאדם אקראי מהאוכלוסייה ישקול מתחת ל-65 ק״ג?

ב. מה הסיכוי שהמשקל הממוצע של 4 אנשים אקראיים יהיה מתחת ל-65 ק״ג?

ג. מה הסיכוי שהמשקל הכולל של 4 אנשים אקראיים יהיה מתחת ל-240 ק״ג?

2) נפח יין בבקבוק מתפלג נורמאלית עם תוחלת של 750 מ״ל וסטיית תקן של 20 מ״ל. אדם קנה מארז של 4 בקבוקי יין.

א. מהי התוחלת ומהי סטיית התקן של נפח היין במארז?

ב. את היין שבמארז האדם מזג לכלי שקיבולתו 3.1 ליטר. מה ההסתברות שהיין יגלוש מהכלי?

ג. אם לא היה נתון שנפח היין מתפלג נורמאלית. האם התשובה לסעיף א׳ הייתה משתנה? האם התשובה לסעיף ב׳ הייתה משתנה?

3) בספר כלשהו 500 עמודים. קצב הקריאה הממוצע הוא עמוד אחד ב-4 דקות עם סטיית תקן של 1 דקות.

א. מה ההסתברות לסיים את הפרק הראשון (40 עמודים) תוך שעתיים וחצי?

ב. מהו האחוזון ה-95 לזמן סיום קריאת הספר?

4) במגדל נבנו 40 יחידות דיור. כמו כן נבנו 135 מקומות חנייה לבניין. להלן פונקציית ההסתברות של מספר המכוניות ליחידת דיור:

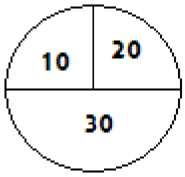
x	1	2	3	4	5
$P(X=x)$	0.1	0.2	0.3	0.25	0.15

נניח שמספר המכוניות ליחידת דיור בלתי תלויות זו בזו ועם אותה פונקציית הסתברות לכל יחידת דיור (אין צורך בתיקון רציפות).

א. מהי ההסתברות שיהיה מקום בחניון המגדל לכל מכוניות הבניין?

ב. בהינתן ויש מקום במגדל לכל המכוניות, מה הסיכוי שבפועל מספר המכוניות נמוך מ-130?

5) בקזינו ישנה רולטה עליה מסומנים המספרים הבאים:



אדם מסובב את הרולטה וזוכה בסכום הרשום.

א. אם האדם משחק 50 פעמים, מה ההסתברות שבסך הכול יזכה בסכום של 1050 ₪ ומעלה?

ב. האדם מגיע בכל יום לקזינו ומשחק 50 פעם, עד אשר מגיע היום בו הוא יזכה ב-1050 ₪ ומעלה. מה התוחלת ומהי השונות של מספר הימים שיבלה בקזינו?

6) נתון ש- $X_i \sim \exp(\lambda = 1)$, כאשר: $i = 1, 2...100$.

חשבו את הסיכוי: $P\left(\sum_i X_i \geq 115\right)$.

7) אורך חיי סוללה (בשעות) הוא בעל פונקציית הצפיפות הבאה: $0 < x < 1$, $f(x) = 2x$. ברגע שסוללה מתרוקנת מחליפים אותה במיידית בסוללה אחרת. כמה סוללות יש להחזיק במלאי אם רוצים שבסיכוי של 90% לפחות המלאי יספיק עבור 35 שעות לפחות?

תשובות סופיות:

1) א. 0.6915. ב. 0.8413. ג. 0.5.

2) א. תוחלת: 3000 מ"ל, סטיית תקן: 40 מ"ל. ב. 0.0062.
ג. סעיף א' - לא משתנה, סעיף ב' - לא פתיר, התבסס על התפלגות נורמלית.

3) א. 0.0571. ב. 2036.8.

4) א. 0.883. ב. 0.7949.

5) א. 0.8997. ב. תוחלת 1.111, שונות 0.1239.

6) 0.0668.

7) 56.

התפלגות מספר ההצלחות במדגם – קירוב נורמלי להתפלגות הבינומית:

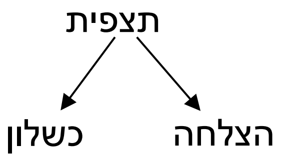
רקע:

תזכורת על התפלגות בינומית:

בפרק זה נדון בהתפלגות מספר ההצלחות במדגם אקראי (תצפיות בלתי תלויות זו בזו).

את מספר ההצלחות במדגם נסמן ב- Y.

מחלקים כל תצפית במדגם להצלחה או כישלון.



כעת מה שמשתנה מתצפית לתצפית הוא משתנה דיכוטומי (משתנה שיש לו שני ערכים).

הסיכוי להצלחה יסומן עם הפרמטר p וכישלון יסומן ע"י הפרמטר: $q=1-p$.

מבצעים מדגם אקראי בגודל n : $Y\sim B(n,p)$.

פונקציית ההסתברות של ההתפלגות הבינומית היא: $p(y=k)=\binom{n}{k}p^k q^{n-k}$,

תוחלת: $E(y)=np$.

שונות: $V(y)=npq$.

קירוב נורמלי עבור התפלגות בינומית:

אם לפנינו התפלגות בינומית: $Y\sim B(n,p)$, ומתקיים ש:

1. $n\cdot p\ge 5$.
2. $n\cdot(1-p)\ge 5$.

אז:

$$y \sim\succ N(np,npq)$$

$$Z_y=\frac{y-np}{\sqrt{npq}}.$$

תיקון רציפות:

כאשר משתמשים בקירוב הנורמלי להתפלגות הבינומית יש לבצע תיקון רציפות.
הסיבה שעוברים כאן מהתפלגות בדידה להתפלגות נורמלית שהיא התפלגות רציפה.
על פי הכללים הבאים:

1. $p(Y=a)\cong p\left(a-\frac{1}{2}\le Y\le a+\frac{1}{2}\right)$.
2. $P(Y\le a)\cong P(Y\le a+0.5)$.
3. $P(Y\ge a)\cong P(Y\ge a-0.5)$.

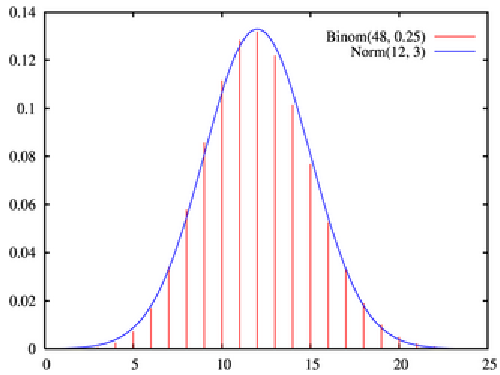
הערות:

- התנאים למעבר מבינומי לנורמלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:
  1. $n\cdot p\ge 5$.
  2. $n\cdot(1-p)\ge 5$.
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
  1. $n\cdot p\ge 10$.
  2. $n\cdot(1-p)\ge 10$.
- וישנם מרצים שהתנאי שהם נותנים הוא: $(n\ge 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור מהתפלגות בינומית לנורמלית.
- הערה נוספת היא לגבי תיקון רציפות. ישנם מרצים שלא מחייבים לבצע תיקון רציפות שהמדגמים גדולים (בדרך כלל מעל 100 תצפיות) בפתרונות שאציג תמיד אבצע תיקון רציפות במעבר מבינומי לנורמלי כיוון שכך הפתרון יהיה יותר מדויק (בכל מקרה שהמדגמים גדולים העניין זניח).

דוגמה (הפתרון בהקלטה):

נתון שבקרב אוכלוסיית הנוער 25% זקוקים למשקפיים. נדגמו באקראי 48 בני נוער.

1. מה הסיכוי שבדיוק 14 מתוכם יהיו זקוקים למשקפיים?
2. מה הסיכוי שלכל היותר 13 מתוכם זקוקים למשקפיים?



שאלות:

1) נתון ש-20% מאוכלוסייה מסוימת אקדמאית. נבחרו באקראי 10 אנשים באותה אוכלוסייה.
   א. מה ההסתברות ששלושה מהם אקדמאים?
   ב. מה ההסתברות שלכל היותר אחד מהם אקדמאי?
   ג. מה התוחלת ומהי סטיית התקן של מספר האקדמאים במדגם?

2) במפעל 10% מהמוצרים פגומים. נלקחו 100 מוצרים באקראי מקו הייצור.
   א. מה ההסתברות שנדגמו לפחות 6 מוצרים פגומים?
   ב. מה ההסתברות שמספר המוצרים הפגומים יהיה לכל היותר 11 במדגם?

3) ציוני פסיכומטרי בקרב הנרשמים למוסד מסוים מתפלגים נורמאלית עם ממוצע 500 וסטיית תקן 100. למוסד מסוים הוחלט לקבל אך ורק סטודנטים שקיבלנו מעל 600 בפסיכומטרי. 100 סטודנטים אקראיים נרשמו למוסד. מה ההסתברות שלפחות 20 יתקבלו?

4) מטילים מטבע 50 פעמים.
   א. מה ההסתברות לקבל לכל היותר 30 עצים?
   ב. מה ההסתברות לקבל 28 עצים לפי התפלגות הבינומית ולפי הקירוב הנורמאלי?

5) במטוס מקום ל-400 נוסעים. נרשמו לטיסה 430 אנשים (overbooking). מנתונים סטטיסטיים ידוע שהסיכוי שאדם שנרשם לטיסה אכן יגיע הוא 0.9.
   א. מה ההסתברות שלא יהיו מקומות ישיבה לכל האנשים שהגיעו לטיסה?
   ב. מה צריך להיות גודל המטוס כדי שבסיכוי שלפחות 95% המטוס יספיק לכמות הנרשמים?

6) מפעל לייצור ארטיקים טוען שהסיכוי שארטיק שהוא מייצר יהיה פגום הוא 0.01. מוכר הזמין 1000 ארטיקים מהמפעל. מה ההסתברות שהמוכר יקבל לפחות 980 ארטיקים תקינים אם טענת המפעל מוצדקת?

7) מהמר מטיל קובייה הוגנת 100 פעמים. בכל הטלה, אם מתקבל תוצאה זוגית בקובייה המהמר זוכה בשקל. אחרת, המהמר משלם שקל. המהמר הטיל את הקובייה 100 פעמים מה הסיכוי שהרווח של המהמר יהיה לכל היותר 10?

תשובות סופיות:

1) א. 0.201. ב. 0.3758. ג. התוחלת: 2, סטיית התקן: 1.2649.

2) א. 0.9332. ב. 0.6915.

3) 0.1611.

4) א. 0.9406. ב. בינומית - 0.0788, קירוב לנורמלית - 0.0778.

5) א. 0.015. ב. 398.

6) 0.9996.

7) 0.8643.

סטטיסטיקה לאקטואריה

פרק 11 - אמידה נקודתית

תוכן העניינים

אומד חסר הטייה:

רקע:

$\hat{\theta}$ יהיה אומד חסר הטיה ל-θ, אם התוחלת של $\hat{\theta}$ תהיה שווה ל-θ: $E\left(\hat{\theta}\right)=\theta$.

דוגמה (פתרון בהקלטה):

המשתנה X הוא בעל פונקציית ההסתברות הבאה:

X	1	2	3
הסתברות	2θ	$1-60\theta$	4θ

מעוניינים לאמוד את θ על סמך שתי תצפיות מההתפלגות: X_1 ו-X_2.

א. הראו שהאומד: $T_1=\frac{2X_1+X_2}{2}$, הוא אומד מוטה ל-θ.

הטיה של אומד היא: $E\left(\hat{\theta}\right)-\theta$. כמובן שלאומד חסר הטיה אין הטיה.

ב. מהי ההטיה של האומד T_1?

ג. תקנו את T_1, כך שיהיה אומד חסר הטיה.

אם יש שני אומדים חסרי הטיה עדיף זה עם השונות היותר קטנה.

ד. מוצא האומד הבא: $T_3=1.5X_1-X_2-1$.

האם הוא עדיף על האומד שהצעת בסעיף ג'?

אם $\hat{\theta}$ אומד חסר הטיה ל-θ, אז $g\left(\hat{\theta}\right)$ יהיה אומד חסר הטיה עבור $g\left(\theta\right)$, רק אם g תהיה לינארית.

ה. מצאו אומד חסר הטיה ל: $P\left(X=3\right)$.

אומד חסר הטיה לשונות האוכלוסייה σ^2: $S^2=\frac{\sum\left(x_i-\bar{x}\right)^2}{n-1}=\frac{\sum x_i^2-n\bar{x}^2}{n-1}$.

ו. מצאו אומד חסר הטיה לשונות של X.

תזכורות חשובות:

אם: $Y=aX+b$, אזי: $E(Y)=aE(X)+b$, $V(Y)=a^2\cdot V(X)$, $\sigma_Y=|a|\sigma_x$.

אם: $X_1,X_2,.......,X_n$ משתנים מקריים, אזי:

$E(T)=E(X_1+X_2+......+X_n)=E(X_1)+E(X_2)+......+E(X_n)$.

אם: $X_1,X_2,.......,X_n$ משתנים מקריים בלתי תלויים בזוגות, אזי:

$V(T)=V(X_1+X_2+......+X_n)=V(X_1)+V(X_2)+......+V(X_n)$.

שאלות:

1) הציון במבחן מסוים של תלמידי כתה ח' הנו משתנה מקרי בעל תוחלת μ וסטיית תקן 10. כדי לאמוד את התוחלת μ, נלקח מדגם של 5 ציונים: $X_1,\ldots,X_5$.

שלושה חוקרים הציעו אומדים לתוחלת על סמך מדגם זה:

חוקר א' הציע: $T_1=\dfrac{X_1+\ldots+X_5}{5}$.

חוקר ב' הציע: $T_2=\dfrac{2X_1-X_3+X_4}{2}$.

חוקר ג' הציע: $T_3=\dfrac{2X_1+X_3}{2}$.

א. איזה מן האומדים הוא חסר הטיה?

ב. הציעו תיקון לאומד המוטה כך שיהיה חסר הטיה.

ג. במדגם התקבלו הציונים הבאים: 65, 78, 58, 82, 100.
חשבו את האומדנים המתקבלים עבור האומדים חסרי ההטיה.

ד. איזה מבין שני האומדים חסרי ההטיה עדיף? נמקו.

2) כדי לאמוד את המשקל הממוצע של הנשים בארה"ב, נבחר מדגם של $2n$ נשים. נסמן את שונות הגובה ב- σ^2. הוצעו שני אומדים לממוצע המשקל על סמך מדגם זה: $T_1=\dfrac{1}{n}\sum_{i=1}^{n}X_i$, $T_2=\dfrac{1}{2n}\sum_{i=1}^{2n}X_i$.

א. בדקו לגבי כל אומד אם הוא בלתי מוטה.

ב. איזה אומד עדיף? נמקו.

3) $X\sim B(n,p)$. כלומר, X הינו משתנה מקרי המתפלג בינומית עם פרמטר P (סיכוי להצלחה בניסיון בודד) במדגם בגודל n.

א. פתחו אומד חסר הטיה ל- P.

ב. מהו אומד חסר הטיה לסיכוי לכישלון בניסיון בודד?

ג. מהו אומד חסר הטיה ל- $E(X)$?

ד. מצאו אומד חסר הטיה ל- $E\left(X^2\right)$.

4) בתיק מניות שתי מניות. מספר המניות שיעלו ביום מסוים הוא משתנה מקרי התלוי בפרמטר לא ידוע: θ, $0 \le \theta \le 2$.

פונקציית ההסתברות של X - מספר המניות שיעלו ביום מסוים:

$$P(X=2)=\frac{\theta}{6}\ ,\ P(X=1)=\frac{\theta}{3}\ ,\ P(X=0)=1-\frac{\theta}{2}\ .$$

א. מִצאו אומד בלתי מוטה ל-θ, שמתבסס על מספר המניות שיעלו ביום מסוים.

ב. מצאו אומד בלתי מוטה ל-θ, שמתבסס על מספר המניות שעלו ביום, במשך שלושה ימים - X_1, X_2, X_3 (לכל אחד מהם אותה התפלגות כנ"ל והם בלתי תלויים).

5) בקרב המטפלות בת"א, מספר התינוקות שבטיפולן הוא משתנה מקרי בעל התפלגות התלויה בפרמטר θ באופן הבא:

הסיכוי שמטפלת תטפל בתינוק אחד בלבד הוא 3θ,

הסיכוי שמטפלת תטפל ב-2 תינוקות הוא $1-4\theta$,

הסיכוי שמטפלת תטפל ב-3 תינוקות הוא θ.

במדגם מיקרי של 4 מטפלות מת"א, נמצא כי שתים מהם מטפלות בתינוק אחד בלבד, אחת מהן בשנים ואחת השלושה תינוקות.

א. מצאו אומד חסר הטיה לפרמטר θ על סמך תצפית בודדת.

ב. מצאו אומד חסר הטיה לפרמטר θ על סמך 4 תצפיות.

ג. מהו האומדן לפרמטר θ על סמך תוצאות המדגם.

ד. מצאו אומד חסר הטיה לסיכוי שלמטפלת בת"א תטפל בתינוק בודד אחד.

ה. מצאו אומדים חסרי הטיה לתוחלת ולשונות של מספר התינוקות בטיפול אצל מטפלת מת"א. חשבו אומדנים.

6) קבעו אילו מהטענות הבאות נכונות:

א. אם T הוא אומד בלתי מוטה עבור פרמטר θ, אז $5T$ אומד בלתי מוטה עבור הפרמטר 5θ.

ב. אם T הוא אומד בלתי מוטה עבור פרמטר θ, אז T^2 אומד בלתי מוטה עבור הפרמטר θ^2.

7) במפעל שתי מכונות המייצרות מוצרים. במכונה הראשונה ההסתברות שמכשיר תקין היא p, ובמכונה השנייה ההסתברות שמכשיר תקין היא $2p$.
דוגמים 20 מכשירים מהייצור של כל מכונה.
נסמן ב- X את מספר המכשירים התקינים שיוצרו על ידי המכונה הראשונה,
וב- Y את מספר המכשירים התקינים שיוצרו על ידי המכונה השנייה.
איזה מבין האומדים הבאים אינו אומד חסר הטיה ל- p ?

א. $\frac{X}{20}$.

ב. $\frac{Y}{20}$.

ג. $\frac{X+Y}{60}$.

ד. $\frac{2X+Y}{80}$.

8) יהיו T_1 ו- T_2 אומדים חסרי הטיה ובלתי תלויים לפרמטר θ.

א. מצאו אומד חסר הטיה ל- θ^2, המתבסס על T_1 ו- T_2.

ב. מצאו אומד חסר הטיה ל- $\theta(1-\theta)$, המתבסס על T_1 ו- T_2.

9) נתון ש- X הינו משתנה מקרי עם תוחלת μ ושונות σ^2.
נדגמו n תצפיות בלתי תלויים מאותה אוכלוסיה.

א. הראו ש- $\sum_{i=1}^{n} p_i x_i$ אומד חסר הטייה ל- μ, כאשר: $\sum_{i=1}^{n} p_i = 1$.

ב. נתבונן במכפלת שתי התצפיות הראשונות: $X_1 \cdot X_2$.
הראו שהוא אומד חסרי הטיה ל- μ^2.

10) $X_i \sim N(\mu,1)$, כאשר: $i = 1,2...,n$.
נתון שהתצפיות הינן בלתי תלויות זו בזו.
מצאו אומד חסר הטיה ל- μ^2.

11) נתונות n תצפיות בלתי תלויות מתוך התפלגות בעלת הצפיפות הבאה:

$$f(x)=\begin{cases}\dfrac{1+\beta x}{2} & -1<x<1\\ 0 & else\end{cases}.$$

א. הראו כי האומד $3\overline{X}$ הנו אומד בלתי מוטה ל-β.

ב. מצאו את השונות של האומד מהסעיף הקודם.

12) $X_1,X_2,...,X_n$ הינם משתנים מקריים רציפים בלתי תלויים בעלי פונקציית הצפיפות הבאה:

$$f(x)=\begin{cases}X\cdot A & 0\le x\le\theta\\ 0 & \text{אחר}\end{cases}.$$

א. בטאו את ערכו של A באמצעות θ, כדי שפונקציית הצפיפות תהיה לגיטימית.

ב. מצאו אומד חסר הטיה ל-θ, על סמך n התצפיות.

תשובות סופיות:

1) א. T_1 ו- T_2. ב. $\frac{2}{3}T_3$. ג. $T_1 = 76.6$, $T_2 = 110$ ד. T_1.

2) א. ראו בווידאו. ב. T_2.

3) א. $\frac{x}{n}$. ב. $1-\frac{x}{n}$ ג. X. ד. θ.

4) א. $\frac{3x}{2}$. ב. $\frac{3\overline{x}}{2}$.

5) א. $1-\frac{x}{2}$. ב. $1-\frac{1}{2}\overline{x}$. ג. 0.125. ד. $3\left(1-\frac{1}{2}\overline{x}\right)$.
ה. לשונות 0.917.

6) א. נכון. ב. לא נכון.

7) ב'.

8) א. $T_1 \cdot T_2$. ב. $T_1 - T_1 \cdot T_2$.

9) א. שאלת הוכחה. ב. שאלת הוכחה.

10) $\overline{X}^2 - \frac{1}{n}$.

11) א. שאלת הוכחה. ב. $V\left(3\overline{X}\right) = \frac{3-\beta^2}{n}$.

12) א. $A = \frac{2}{\theta^2}$. ב. $\theta = \frac{3}{2}\overline{X}$.

אומד נראות מקסימלית:

רקע:

להלן נלמד את שיטת הנראות המקסימלית למציאת אומדים.

נניח ש- X משתנה מקרי בדיד עם פונקציית הסתברות $P(x,\theta)$,
כאשר θ הפרמטר הבלתי ידוע.

יהיו: $X_1, X_2, ..., X_n$ תוצאות מדגם מקרי בגודל n הנלקח מאוכלוסייה זו.

נבנה את פונקציית ההסתברות המשותפת (פונקציית הדגימה).

אם אנו יודעים את תוצאות המדגם, ולא את הפרמטר, קוראים לפונקציית הנראות שהיא פונקציה של הפרמטר.

נגדיר את פונקציית הנראות:

$$L(\theta)=P(x_1,\theta)\cdot P(x_2,\theta)\cdot\ldots\cdot P(x_n,\theta)=\prod_{i=1}^{n}P(x_i,\theta).$$

פונקציית הנראות היא ההסתברות לקבל את התצפית הראשונה (כפונקציה של θ), כפול ההסתברות לקבל את התצפית השנייה, וכולי. כלומר, המשמעות של פונקציית הנראות היא ההסתברות לקבל את המדגם שהתקבל, כפונקציה של הפרמטר המבוקש θ.

אם מדובר במשתנה רציף, נכפיל את פונקציות הצפיפות ולא את פונקציות ההסתברות:

$$L(\theta)=f(x_1,\theta)\cdot f(x_2,\theta)\cdot\ldots\cdot f(x_n,\theta)=\prod_{i=1}^{n}f(x_i,\theta).$$

דוגמה (פתרון בהקלטה):

הסיכוי של שחקן כדורסל לקלוע לסל הוא p (לא ידוע). השחקן זורק כדורים לסל עד שהוא קולע בפעם הראשונה. נניח כי הזריקות בלתי תלויות זו בזו.
הכדור נכנס לסל לראשונה בניסיון השלישי.
השחקן חוזר על התהליך שוב, והפעם הכדור נכנס לסל בניסיון החמישי.
מצאו את פונקציית הנראות של p.

אומד נראות מקסימלית עבור θ הוא האומד $\hat{\theta}$, שממקסם את פונקציית הנראות $L(\theta)$.
כלומר, אנו מחפשים את האומד שיגרום לכך שהמדגם המקרי שקיבלנו יהיה כמה שיותר סביר.

שלבים למציאת אומד נראות מקסימלית:

- לוקחים את פונקציית ההסתברות המשותפת של המדגם (או צפיפות משותפת אם המשתנה רציף).
- מציבים את תוצאות המדגם ומקבלים את פונקציית הנראות (פונקציה של הפרמטר הנחקר).
- מוצאים מקסימום לפונקציית הנראות (לעיתים כדאי להוסיף ln כדי להקל על המלאכה).

המשך דוגמה:

חשבו את אומדן הנראות המקסימלית עבור p.

משפט: אם $\hat{\theta}$ הוא אומד נראות מקסימלית עבור θ, אזי $g\left(\hat{\theta}\right)$ הוא אומד נראות מקסימלית עבור $g\left(\hat{\theta}\right)$, בהנחה והפונקציה היא חד-חד ערכית (אינווריאנטיות).

המשך דוגמה:

מצאו אומדן נראות מקסימלית לסיכוי של שחקן הכדורסל לקלוע לסל פעמיים ברצף.

שאלות:

1) הסיכוי של שחקן לנצח במשחק הוא p (לא ידוע).
השחקן משחק במשחק עד אשר הוא מנצח בפעם הראשונה.
נתון שהשחקן ניצח לראשונה רק במשחק השני.

א. חשבו את פונקציית הנראות של p, וציירו גרף שלה.

ב. מצאו אומדן נראות מקסימלית עבור p.

ג. מצאו אומדן נראות מקסימלית ל- p, אם ביום אחד הוא נאלץ לשחק 4 פעמים וביום אחר הוא נאלץ לשחק 5 פעמים, עד אשר ניצח.

2) מספר הלקוחות שנכנסים לחנות מסוימת, מתפלג פואסונית עם תוחלת של λ לקוחות ביום.

א. מצאו אומד נראות מקסימלית ל- λ, על סמך מספר הלקוחות שנכנסים ביום מסוים.

ב. מצאו אומד נראות מקסימלית ל- λ, על סמך מספר הלקוחות שנכנסים ב- n ימים מסוימים.

3) הזמן שלוקח לאדם לחכות בתור מתפלג מעריכית עם פרמטר λ.
דגמו 4 אנשים מקריים שחיכו בתור ומדדו את זמני ההמתנה שלהם.
התוצאות שהתקבלו בדקות הן: 3, 7,5 ו-3.

א. פתחו אומד נראות מקסימלית לפרמטר זה על סמך n תצפיות כלשהן.

ב. מהו האומדן לפרמטר?

4) משך זמן הכנת שיעורי הבית (בשעות) של בני נוער, ביום אחד, מתפלג אחיד: $U(0,q)$.

כדי לאמוד את θ, נשאלו ביום מסוים מספר בני נוער כמה שעות הם הכינו שיעורי-בית באותו יום.

א. אלעד הכין ביום מסוים שעורי בית במשך שעה שלמה. חשבו את פונקציית הנראות של θ המתבססת על תצפית זו, וציירו את הגרף שלה.

ב. מצאו אומדן נראות מקסימלית ל- θ על סמך התצפית.

ג. משכי הכנת שיעורי בית (שעות) של 3 בני נוער היו $1,3,1.5$.
מצאו אומדן נראות מקסימלית ל- θ על סמך המדגם הזה.

ד. מצאו באופן כללי אומד נראות מקסימלית ל- θ, על סמך מדגם של n בני נוער – $X_1,\ldots,X_n$.

5) הגובה של אוכלוסייה מסוימת מתפלג נורמלית עם תוחלת ידועה של 170 ס״מ ושונות σ^2 לא ידועה.

א. מצאו אומד נראות מקסימלית עבור השונות על סמך מדגם $X_1,\ldots,X_n$ מתצפיות מהאוכלוסייה.

ב. נדגמו 5 אנשים בלתי תלויים בעלי הגבהים: 170, 182, 165, 174, 174. מהו האומדן לשונות הגבהים באוכלוסייה?

6) פתחו אומד נראות מקסימלית לפרמטר p בהתפלגות הבינומית, על סמך מדגם בגודל n, בו X הוא מספר ההצלחות במדגם.

7) X הוא משתנה מקרי בעל פונקציית הצפיפות: $f(x)=\begin{cases} 2\theta x e^{-\theta x^2} & x>0 \\ 0 & x\le 0 \end{cases}$.

א. מצאו אומד נראות מקסימלית ל-θ על סמך n תצפיות בלתי תלויות: $X_1,\ldots,X_n$.

ב. מצאו אומד נראות מקסימלית ל-θ^2.

8) בכד א' 10 כדורים שחורים ו-10 לבנים ובכד ב' 5 כדורים שחורים ו-15 לבנים. דוגמים באקראי כדור, בלי לדעת מאיזה כד.

א. מצא אומד נראות מקסימלית לכד שממנו הוצא הכדור על סמך הצבע של הכדור.

ב. מהו האומדן אם הצבע הוא שחור?

9) הזמן שלוקח ליוסי לפתור תשבץ מתפלג מעריכית עם תוחלת לא ידועה. נתנו ליוסי לפתור חמישה תשבצים ובממוצע לקח לו 32 דקות לפתור אותם.

א. מה אומדן הנראות המקסימלית לתוחלת זמן הפתרון של תשבץ על ידי יוסי (אין חובה לפתח).

ב. מה אומדן הנראות המקסימלית לסיכוי שייקח לו לפחות חצי שעה לפתור את התשבץ הבא?

10) מספר הלקוחות הממתינים בתור במוקד טלפוני הוא משתנה מקרי X, בעל התפלגות התלויה בפרמטר θ, באופן הבא:

X	0	1	2
$P(X)$	$4\theta^2$	$4\theta-8\theta^2$	$1-4\theta+4\theta^2$

בחמישה זמנים שונים שנבחרו באקראי נמצאו: 0, 0, 0, 1, 0 לקוחות ממתינים בתור.

א. מצאו אומדן בשיטת הנראות המקסימלית עבור הפרמטר θ, על-סמך המדגם הנתון.

ב. מצאו אומדן בשיטת הנראות המקסימלית לסיכוי שלא יהיו לקוחות בתור.

11) אדם מחזיק בידו שני מטבעות: מטבע הוגן ומטבע שאינו הוגן – שהסיכוי לקבל בו תוצאה של עץ הוא 0.2. האדם מטיל את אחד המטבעות פעמיים ומודיע לך כמה פעמים הוא קיבל עץ. אתה צריך לנחש איזה מטבע הוא הטיל: את ההוגן או את זה שאינו הוגן.

א. מצא אומד בשיטת הנראות המקסימלית לסוג המטבע שהוטל.

ב. מהו האומדן אם האדם קיבל פעמיים עץ?

12) מעוניינים לאמוד את אחוז המובטלים באוכלוסייה.
דוגמים 50 אנשים אקראיים ומתקבל ש-4 מהם מובטלים.

א. מצא אומדן נראות מקסימלית לשיעור המובטלים באוכלוסייה.

ב. מצא אומדן לשיעור העובדים באוכלוסייה.

ג. מצא אומדן ליחס בין שיעור העובדים לשיעור המובטלים באוכלוסייה.

13) במשחק מחשב שלוש רמות משחק:
ברמה 1 הסיכוי של יוסי לסיים את המשחק הוא 0.9.
ברמה 2 הסיכוי של יוסי לסיים את המשחק הוא 0.7.
ברמה 3 הסיכוי של יוסי לסיים את המשחק הוא 0.4.
יוסי בחר ברמה מסוימת, אך אינו יודע באיזו רמה הוא בחר.
הוא משחק במשחק ברמה שבחר פעמיים.

א. הציעו א.נ.מ. לרמה של המשחק שיוסי שיחק, על סמך מספר הפעמים שסיים את המשחק.

ב. אם יוסי סיים את שני המשחקים, מה יהיה האומדן לרמה?

ג. מהו א.נ.מ. לסיכוי, שמתוך שני משחקים הוא יצליח בדיוק משחק אחד?

14) $X_1,X_2,...,X_n$ מתפלגים אחיד בקטע: $[-\theta,\theta]$.
מצא אומד נראות מקסימלית לפרמטר θ.

15) $X_1,X_2,...,X_n$ מתפלגים בדיד לפי פונקציית ההסתברות הבאה:

$$K=1,2 \qquad P(X=k)=\frac{\binom{2}{k}\cdot P^k\cdot(1-P)^{2-k}}{1-(1-P)^2}.$$

הוכח שא.נ.מ ל- P, הינו: $2-\frac{2}{\overline{X}}$.

16) במכשיר חשמלי יש 2 סוללות שפועלות באופן ב"ת זו בזו, והוא מפסיק לפעול ברגע שאחת הסוללות מפסיקה לעבוד. הסיכוי של סוללה לתפקד לפחות חודש הוא P. כאשר המכשיר מפסיק לפעול מחליפים את שתי הסוללות שלו.
בתחילת הניסוי נלקחו 80 מכשירים כאלה עם סוללות חדשות ולאחר חודש נמצא ש-30 מהם עדיין פועלים.

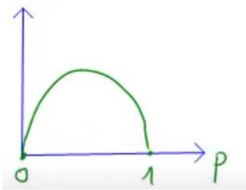
א. מצא אומדן נראות מקסימלית עבור P.
ב. רשמו את האומד שבו השתמשתם בחלק א' באופן כללי, עבור מדגם של n מכשירים שמתוכם נמצאו Y מכשירים שעדיין פועלים לאחר חודש אחד.
ג. בהנחה שאורך החיים (בחודשים) של סוללה בודדת הוא מעריכי,
עם פי צפיפות: $f(t)=\theta e^{-\theta t}$ עבור $t>0$.
מצא א.נ.מ. עבור θ, המבוסס על Y.
מהו האומדן המתאים מן המדגם הנתון?

17) חיוג אוטומטי של מכשיר טלפון משדר אות אחת לשתי דקות. אם לאחר 20 דקות (10 אותות חיוג) המספר שאליו מטלפנים עדיין תפוס, החיוג האוטומטי נפסק.

א. רשמו את פונקציית ההסתברות של המשתנה X – מספר הפעמים שהחייגן האוטומטי מחייג למספר הטלפון המבוקש, אם ההסתברות לקבלת צליל "פנוי" בשידור אחד של אות חיוג הוא P.
ב. מתוך 12 ניסיונות חיוג אוטומטי למשרד הרישוי בזמנים שונים במשך 5 ימים, התקבלו התוצאות הבאות: בשני ניסיונות הופסק החיוג האוטומטי ובשאר הניסיונות שבהם הצליח המטלפן להשיג את המספר המבוקש, מספר החיוגים האוטומטיים עד לקבל צליל "פנוי" היו: 1, 6, 2, 7, 3, 8, 2, 2, 1, 5.
מצאו אומדן נראות מקסימלית עבור P, על סמך התוצאות שהתקבלו.

תשובות סופיות:

1) א. $L(p)=(1-p)\cdot p$. להלן גרף:



ב. 0.5. ג. $\frac{2}{9}$.

2) א. X. ב. $\bar{X}$.

3) א. $\frac{1}{\bar{X}}$. ב. $\frac{2}{9}$.

4) א. 1. ב. 1. ג. 3. ד. $\hat{\theta}=\max\{X_1,\dots,X_n\}$.

5) א. $\frac{\sum_{i=1}^{n}(x_i-170)^2}{n}$. ב. 40.2.

6) $\frac{x}{n}$.

7) א. $\frac{n}{\sum X_i^2}$. ב. $\left(\frac{n}{\sum X_i^2}\right)^2$.

8) א. ראה סרטון. ב. כד א'.

9) א. 32. ב. 0.3916.

10) א. 0.45. ב. 0.81.

11) א.ראה סרטון. ב. הוגן.

12) א. 0.08. ב.0.92. ג. 11.5.

13) א. $\hat{\theta}=\begin{cases}3 & X=0,1\\ 1 & X=2\end{cases}$. ב. 1. ג. $\hat{p}=\begin{cases}2\cdot 0.4\cdot 0.6 & X=0,1\\ 2\cdot 0.9\cdot 0.1 & X=2\end{cases}$.

14) $\max|X_i|$.

15) שאלת הוכחה.

16) א. 0.6124 ב. $\hat{p}=\sqrt{\frac{y}{n}}$ ג.0.49.

17) א. $P(x)=\begin{cases}(1-p)^{x-1}p & 1\le x\le 9\\ (1-p)^9 & x=10\end{cases}$. ב. 0.1818.

נספח:

התפלגויות רציפות

ההתפלגות	פונקציית הצפיפות	פונקציית ההתפלגות המצטברת	תוחלת	שונות	הערות	אנ"מ
$X \sim U(a,b)$	$f(x)=\frac{1}{b-a}$ $a \le x \le b$	$\frac{t-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$		$b=\max(X_i)$ $a=\min(X_i)$
$X \sim \exp(\lambda)$	$f(x)=\lambda e^{-\lambda x}$	$1-e^{-\lambda t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	הזמן עד להתרחשות מאורע מסוים. λ - הוא ממוצע האירועים ביחידת זמן.	$\hat{\lambda}=\frac{1}{\bar{X}}$
$X \sim N(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\Phi(t)$	μ	σ^2	$Z=\frac{X-\mu}{\sigma}$	$\hat{\mu}=\bar{X}$ $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

התפלגויות בדידות

ההתפלגות	פונקציית ההסתברות $P(X=k)$	תוחלת	שונות	הערות	אנ"מ
בינומית $B(n,p)$ $0 \le p \le 1$	$\binom{n}{k}p^k(1-p)^{n-k}$ $k=0,1,\dots,n$	np	$np(1-p)$	(1)	$\hat{P}=\frac{Y}{n}$
גיאומטרית $G(p)$ $0 < p \le 1$	$(1-p)^{k-1}p$ $k=1,2,\dots,\infty$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	(2)	$\hat{P}=\frac{1}{\bar{X}}$
אחידה $U(a,b)$	$\frac{1}{b-a+1}$ $K=a,\dots,b$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$	(3)	$b=\max(X_i)$ $a=\min(X_i)$
פואסונית $P(\lambda)$ $\lambda > 0$	$\frac{e^{-\lambda}\lambda^k}{k!}$ $k=0,1,\dots,\infty$	λ	λ	(4)	$\hat{\lambda}=\bar{X}$

(1) מספר ההצלחות ב- n ניסויי ברנולי ב"ת. p - ההסתברות להצלחה.

(2) מספר הניסויים עד להצלחה הראשונה בסדרת ניסויי ברנולי ב"ת, p - ההסתברות להצלחה.

(3) בחירה אקראית של מספר בין a ו- b.

(4) מספר אירועים ביחידת זמן, λ - קצב האירועים.

קריטריון MSE – תוחלת ריבוע הטעות:

רקע:

הקריטריון הנפוץ ביותר כדי לבדוק את טיב האומד הוא קריטריון MSE:
Mean Squared Error – תוחלת ריבוע טעות האמידה.

$$MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = V(\hat{\theta}) + \left(E(\hat{\theta})-\theta\right)^2 .$$

כאשר: $V(\hat{\theta})$ - הינה שונות האומד.

$E(\hat{\theta})-\theta$ - הינה ההטיה של האומד.

אם T_1 ו- T_2 הינם אומדים לפרמטר θ, האומד העדיף יהיה זה עם MSE קטן יותר.
כלומר, אם: $MSE(T_1) > MSE(T_2)$, אז T_2 עדיף על T_1.

דוגמה (הפתרון בהקלטה):

נתון משתנה X המתפלג אחיד רציף באופן הבא: $X \sim U(3,\theta)$.

מוצעים שני אומדים לפרמטר θ על סמך תצפית בודדת: $T_1 = 2X-3$ ו- $T_2 = \frac{3X-3}{2}$.

איזה אומד עדיף לאמידת הפרמטר θ?

שאלות:

1) מעוניינים לאמוד את התוחלת של התפלגות מסוימת.
מוצעים שני אומדים אפשריים ממוצע של שתי תצפיות וממוצע של שלוש תצפיות.
לפי קריטריון תוחלת ריבוע הטעות (MSE), איזה אומד עדיף? הסבירו.

2) בעיר מסוימת בשוויץ בכל θ דקות רכבת מגיעה לתחנה מסוימת.
דוד מגיע לתחנה בזמן אקראי ומודד את זמן ההמתנה לרכבת - X.

א. הצע אומד חסר הטיה ל-θ, על סמך X.

ב. סטטיסטיקאי הציע לאמוד את θ על סמך האומד: $1.5X$.
האם האומד הנ"ל מוטה?

ג. איזה אומד מבין האומדים בסעיפים א' ו-ב' עדיף?

3) חוקר מעוניין לאמוד את הסיכוי לחלות במחלת השפעת בחורף (להלן: הפרמטר P). הוא דוגם חמישה אנשים בריאים, ומתבונן בסטטיסטי X - מספר האנשים שחלו בשפעת בחורף.
הוא מתלבט בין שני אומדים: $T_1=\frac{X}{5}$ ו- $T_2=\frac{X+1}{7}$.

א. מי מבין האומדים הללו הוא חסר הטיה?

ב. מי מבין האומדים עדיף אם $P=0.5$?

ג. מי מבין האומדים עדיף אם $P=0.1$?

4) מספר השריפות המתרחשות בארץ בחודש אוקטובר מתפלג פואסונית עם תוחלת λ. נלקח מדגם של 10 חודשי אוקטובר. להלן שני אומדים אפשריים:

$$\hat{\lambda}_1=\frac{\sum_{i=1}^{10}X_i}{10} \quad \text{ו-} \quad \hat{\lambda}_2=\frac{2\cdot\sum_{i=1}^{5}X_i+2\cdot\sum_{i=6}^{10}X_i}{10}.$$

כאשר: X_i = מספר השריפות בחודש אוקטובר ה-i.
איזה מהאומדים עדיף, לצורך אמידת הפרמטר λ?

5) הוכח ש: $E\left(\hat{\theta}-\theta\right)^2=V\left(\hat{\theta}\right)+\left(E\left(\hat{\theta}\right)-\theta\right)^2$.

תשובות סופיות:

1) שלוש תצפיות.
2) א. $2x$. ב. אומד מוטה. ג. סעיף ב.
3) א. T_1. ב. T_2. ג. T_1.
4) $\hat{\lambda}_1$.
5) שאלת הוכחה.

שיטת המומנטים:

רקע:

מומנט מסדר ראשון של משתנה X מוגדר להיות: $E(X)$.

מומנט מסדר שני של משתנה X מוגדר להיות: $E(X^2)$.

באופן כלל, מומנט מסדר r מוגדר להיות: $E(X^r)$.

מומנט מסדר ראשון של n תצפיות בלתי תלויות מאותה התפלגות מוגדר להיות: $\frac{\sum X_i}{n}$ - זהו מומנט מסדר ראשון של המדגם.

מומנט מסדר שני של n תצפיות בלתי תלויות מאותה התפלגות מוגדר להיות: $\frac{\sum {X_i}^2}{n}$ - זהו המומנט מסדר שני של המדגם.

באופן כללי, מומנט מסדר r של n תצפיות בלתי תלויות מאותה התפלגות מוגדר להיות: $\frac{\sum {X_i}^r}{n}$ - זהו מומנט ה-r של המדגם.

השיטה: משווים את המומנט המתאים של ההתפלגות לפי המומנט המתאים של המדגם.

דוגמה (פתרון בהקלטה):

נגיד שמספר הפעמים שאדם מתעטש ביום מתפלג פואסונית על ידי פרמטר λ (קצב ההתעטשויות ביום). רוצים לאמוד את λ בשיטת המומנטים.

שאלות:

1) X מתפלג אחיד רציף מהערך המינימלי a לערך המכסימלי 20. מצא אומד לערך מינימלי a לפי שיטת המומנטים על סמך n תצפיות מההתפלגות.

2) דוגמים n תצפיות בלתי תלויות מתוך התפלגות נורמאלית אשר תוחלתה היא μ והשונות שלה היא σ^2. מצא אומדים לפרמטרים אלה לפי שיטת המומנטים.

3) אדם מטיל מטבע רגיל n פעמים. יש לאמוד את מספר הפעמים שהוא מטיל את המטבע וזאת על סמך X – מספר העצים שהוא קיבל.
א. מצא אומד בשיטת המומנטים ל-n על סמך X בודד.
ב. מצא אומד בשיטת המומנטים ל-n על סמך חזרה של m פעמים על אותו תהליך בו מטילים את המטבע ההוגן n פעמים.
ג. מהו האומדן אם האדם חזר על התהליך שלוש פעמים:
פעם אחת קיבל 5 עצים, בפעם השנייה הוא קיבל 4 עצים ובפעם השלישית הוא קיבל 7 עצים.

4) נתון ש- $X_i \sim \exp(\lambda)$. מצא אומד בשיטת המומנטים לפרמטר λ על סמך מדגם של n תצפיות.

5) נתונה פונקציית הצפיפות הבאה: $f(x)=\begin{cases}\theta\cdot X^{\theta-1} & 0\le x\le 1\\ 0 & else\end{cases}$.
א. בטא את $E(X)$ כפונקציה של הפרמטר θ.
ב. מצא אומד ל-θ על פי שיטת המומנטים.

6) הזמן בדקות להכנת לחם במאפייה מתפלג באופן הבא: $X_i \sim N(10,\sigma^2)$.
במדגם של הכנת ארבעה לחמים התקבלו התוצאות הבאות: 5, 10, 6, 4.
א. אמוד את σ^2 בשיטת המומנטים על סמך מדגם בגודל n.
ב. מצא את האומדן ל- σ^2. מה הבעייתיות בתשובה?

תשובות סופיות:

1) $\hat{a}=2\overline{X}-20$.

2) $\overline{X}=\hat{\mu}$, $\hat{\sigma}^2=\frac{\sum X_i^2}{n}-\overline{X}^2$.

3) א. $\hat{n}=2X$. ב. $\hat{n}=2\overline{X}_m$. ג. $\hat{n}=10\frac{2}{3}$.

4) $\hat{\lambda}=\frac{1}{\overline{X}}$.

5) א. $\frac{\theta}{\theta+1}$. ב. $\hat{\theta}=\frac{\overline{X}}{1-\overline{X}}$.

6) א. $\hat{\sigma}^2=\frac{\sum X_i^2}{n}-100$. ב. ‎-55.75.

אומד חסר הטיה בעל שונות מינימלית:

אומד חסר הטיה יעיל ביותר – MVUE (Minimum-variance unbiased estimator).

רקע:

T יהיה MVUE, אם מתקיים ש- T אומד חסר הטיה ל- θ, ובנוסף מתקיים ש:
$V(T) \le V(\hat{\theta})$, לכל $\hat{\theta}$ חסר הטיה אחר.

דוגמה (פתרון בהקלטה):

לרשת חנויות ישנם שני סניפים. מספר הלקוחות הנכנסים לכל סניף ביום מתפלג פואסונית עם קצב של λ בסניף A וקצב של 2λ בסניף B.
נדגמו n ימים מכל סניף, ונבדק בכל יום:
X_i - מספר הלקוחות שנכנסו לסניף A ביום i.
Y_j - מספר הלקוחות שנכנסו לסניף B ביום j.
על מנת לאמוד את λ, מוצע האומד: $\alpha\bar{X}+\beta\bar{Y}$.

א. מה התנאי, שצריך להתקיים על α ו- β, כדי שהאומד יהיה חסר הטיה?

ב. מה צריכים להיות α ו- β כדי שהאומד יהיה גם בעל שונות מינימלית?

שאלות:

1) T_1 ו- T_2 הינם אומדים חסרי הטיה ובלתי תלויים לפרמטר θ.

כמו כן, נגדיר: $T = aT_1 + bT_2$.

א. מה צריך להיות התנאי על a ו- b, כדי ש- T יהיה אומד חסר הטיה?

ב. σ_1^2 ו- σ_2^2 הם השונויות של T_1 ו- T_2, בהתאמה.

מצאו a ו- b, כך ש- T יהיה אומד חסר הטיה ל- θ, ובעל שונות מינימלית.

2) במפעל 3 מכונות המייצרות את אותו חלק.

תוחלת הקוטר של החלקים המיוצרים בכל מכונה זהה.

השונויות של כל מכונה שונות, ומקיימות: $\sigma_2^2 = 2\sigma_1^2$, $\sigma_3^2 = 3\sigma_1^2$.

הוחלט לדגום n חלקים מכל מכונה, ולחשב את ממוצע הקוטר המתקבל.

$\bar{X}_i$ - יהיה הממוצע המתקבל במכונה i.

יהי: $W = \sum_{i=1}^{3} a_i \bar{X}_i$ האומד לתוחלת קוטר החלקים המיוצרים על ידי מכונה כלשהי.

א. מה התנאי שצריך להתקיים על המשקלים a_i, כדי שהאומד המוצע יהיה בלתי-מוטה?

ב. נניח ש- $a_1 = a_2$.

מה במקרה זה המשקלים המביאים את האומד להיות MVUE?

תשובות סופיות:

1) א. $a + b = 1$. ב. $a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$, $b = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$.

2) א. $\sum_{i=1}^{3} a_i = 1$. ב. $a_1 = a_2 = 0.4$, $a_3 = 0.2$.

שאלות מסכמות:

שאלות:

1) במפעל מייצרים מוצרים בשלוש מכונות שונות ובלתי תלויות.
במכונה הראשונה הסיכוי שמוצר יהיה תקין הוא P , במכונה השנייה
ההסתברות שמוצר יהיה תקין הוא P^2 ובמכונה השלישית הסיכוי הוא $2P$.
דוגמים 20 מוצרים מכל מכונה.
נסמן ב- X את מספר המוצרים התקינים שיוצרו במכונה הראשונה,
ב- Y את מספר המוצרים התקינים שיוצרו במכונה השנייה
וב- Z את מספר המוצרים התקינים שיוצרו במכונה השלישית.

א. מהם הערכים האפשריים של הפרמטר P ?

ב. מצאו אומד בלתי מוטה עבור הפרמטר P , על סמך X ו- Z .

ג. אם התקבל ש- $Y=3$, $X=6$, מהו אומדן נראות מקסימלית ל- P ?

2) מספר תאונות הדרכים בקטע כביש א' מתפלג פואסונית עם קצב של λ
תאונות בחודש, ומספר תאונות הדרכים בקטע כביש ב' מתפלג פואסונית
עם קצב של 2λ תאונות בחודש.
הוחלט לספור את כמות התאונות בחודש בכל אחד מקטעי הכביש.
נסמן ב- X את מספר התאונות בחודש בקטע א' וב- Y בקטע ב'.

א. מצאו אומד נראות מקסימלית לפרמטר λ, על סמך X ו- Y .

ב. מצאו אומד נראות מקסימלית, לסיכוי שבקטע כביש א תהיה לפחות תאונה אחת בחודש.

ג. האם האומד שמצאת בסעיף א הוא חסר הטיה ל- λ ?

3) זמן הייצור של מוצר מסוים בתהליך ייצור מתפלג נורמאלית, עם תוחלת
ושונות שאינן ידועות.

א. הציעו אומדים חסרי הטיה לתוחלת והשונות של זמן הייצור של המוצר.

ב. הציעו אומדי נראות מקסימלית לתוחלת ולשונות של זמן הייצור של המוצר.

ג. הציעו אומד נראות מקסימלית לריבוע התוחלת של זמן הייצור.

ד. האם האומד מהסעיף הקודם הוא גם חסר הטיה?

4) בקזינו משחק, ובו 4 תאים ממוספרים מ-1 עד 4.
מפעיל המשחק שם כסף באחד מארבעת התאים והאדם המשתתף צריך לנחש באיזה תא הכסף מוחבא. מפעיל הקזינו מודיע שהסיכוי להחביא את הכסף בכל אחד משלושת התאים הראשונים שווה, אך לא בהכרח שווה לסיכוי להחביא אותו בתא הרביעי.
יש לאמוד את הסיכוי להחביא את הכסף בתא הראשון: P.

א. מצא את תחום ההגדרה של הפרמטר P.

יעל שיחקה את המשחק 3 פעמים וקיבלה שפעם אחת הכסף הוחבא בתא מספר 1 ובפעמים האחרות בתא מספר 2.

ב. מצאו אומדן ל- P על סמך התוצאות הללו בשיטת הנראות המקסימלית.

ג. מצאו אומד חסר הטיה ל- P. מהו האומדן לפי התוצאות של יעל?

ד. מצאו אומדן חסר הטיה ונראות מקסימלית לסיכוי שהכסף יוחבא בתא מספר 4 על סמך התוצאות של יעל.

5) יהי: $X_1, X_2, ... X_n$ מדגם מקרי מתוך ההתפלגות הבאה:

$$f(x)=\begin{cases}\dfrac{\theta}{\lambda}\left(\dfrac{x}{\lambda}\right)^{\theta-1} & 0<x<\lambda,\ \theta>0\\ 0 & else\end{cases}$$

א. מצא אח"ה ל- λ (כאשר θ קבוע ידוע).

ב. מצא אנ"מ ל- θ (כאשר λ קבוע ידוע).

ג. מצא אנ"מ ל- λ (כאשר θ קבוע ידוע).

6) X - משך זמן הפרסומות בערוץ 2 מתפלג אחיד רציף בתחום $(0,\theta)$.

Y - משך זמן הפרסומות בערוץ 10 מתפלג אחיד רציף בתחום $(0,2\theta)$.

א. מצא אומד חסר הטיה ל- θ, המשתמש במשך זמן אקראי של פרסומת בודדת בערוץ 2 ופרסומת בודדת בערוץ 10.

ב. מוצע האומד: $T_2 = X + 0.5Y$. האם האומד הנ"ל הוא חסר הטיה?

ג. איזה אומד יותר עדיף זה של סעיף א או זה של סעיף ב'?

ד. מצא אומד נראות מקסימלית ל- θ על סמך X ו- Y.

7) נדגמו 2 תצפיות, X_1, X_2, בלתי תלויות מהתפלגויות אחידות רציפות התלויות בפרמטר θ.
ידוע כי: $X_1 \sim U(0,\theta)$, $X_2 \sim U(0,a\theta)$ (כאשר a קבוע ידוע וחיובי).

א. מצא אנ"מ ל- θ, על סמך 2 התצפיות הנ"ל.

ב. חשב את תוחלת ושונות האנ"מ מסעיף א'. האם האנ"מ מוטה?

ג. מצא אח"ה ל- θ על סמך סכומן של 2 התצפיות הנ"ל. מהי שונותו?

תשובות סופיות:

1) א. $0 \le P \le 0.5$. ב. $\hat{p} = \frac{x+z}{60}$. ג. 0.345.

2) א. $\frac{x+y}{3}$. ב. $1 - e^{-\frac{x+y}{3}}$. ג. כן.

3) א. מאחר ולא התבקשתם לפתח, הרי שהאומד זהה לנוסחה הכללית (ראו נספח).
ב. כנ"ל. ג. כנ"ל (ראו הפרק על אומד נראות מקסימלי). ד. לא.

4) א. $0 \le P \le \frac{1}{3}$. ב. $\frac{1}{3}$. ג. 0.389. ד. ‎-0.167.

5) א. אחי"ה יהיה: $\hat{\lambda} = \frac{\theta+1}{\theta}\bar{x}$. ב. $\hat{\theta} = \frac{n}{n \ln \lambda - \sum_{i=1}^{n} \ln x_i}$.
ג. $\hat{\lambda} = X_{\max}$.

6) א. $T_1 = (x+y)\frac{2}{3}$. ב. כן. ג. ב'. ד. $\hat{\theta} = \max\left\{x, \frac{1}{2}y\right\}$.

7) א. $\hat{\theta} = \max\left(X_1, \frac{X_2}{a}\right)$. ב. $E(\hat{\theta}) = \frac{2}{3}\theta$, $V(\hat{\theta}) = \frac{1}{18}\theta^2$.
ג. $\tilde{\theta} = \left(\frac{2}{1+a}\right)(X_1 + X_2)$.

נספח: אומדי נראות מקסימלית ואומדים חסרי הטיה בהתפלגויות השונות:

מודל בינומי

נתון מדגם של משתנה בינומי: $X \sim B(n,p)$.

א.נ.מ עבור p הוא: $\hat{p}=\frac{X}{n}$, והוא גם א.ח.ה.

מודל אחיד (בדיד)

נתון מדגם: $X_1,X_2,...,X_n$ של משתנים אחידים: $X_i \sim U(1,N)$ בלתי-תלויים בזוגות.

א.נ.מ עבור N הוא: $\hat{N}=\max\{X_1,...,X_n\}$, ואינו א.ח.ה.

מודל פואסוני

נתון מדגם: $X_1,X_2,...,X_n$ של משתנים פואסוניים: $X_i \sim P(\lambda)$ בלתי-תלויים בזוגות.

א.נ.מ עבור λ הוא: $\hat{\lambda}=\overline{X}$ וגם א.ח.ה.

מודל גיאומטרי

נתון מדגם: $X_1,X_2,...,X_n$ של משתנים גיאומטריים: $X_i \sim G(p)$ בלתי-תלויים בזוגות.

א.נ.מ עבור p הוא $\hat{p}=\frac{1}{\overline{X}}$, אינו א.ח.ה. א.נ.מ עבור התוחלת $\frac{1}{p}$, הוא $\overline{X}$ והינו א.ח.ה.

מודל נורמלי

נתון מדגם: $X_1,X_2,...,X_n$ של משתנים נורמליים: $X_i \sim N(\mu,\sigma^2)$ בלתי-תלויים בזוגות.

א.נ.מ עבור μ, הוא: $\hat{\mu}=\overline{X}$.

כאשר μ ידוע, א.נ.מ עבור σ^2 הוא: $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ (אומד חסר-הטיה).

כאשר μ לא-ידוע, א.נ.מ עבור σ^2 הוא: $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ (אומד מוטה!!!).

אומד חסר-הטיה עבור σ^2:

כאשר μ ידוע: $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$.

כאשר μ לא-ידוע: $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$.

מודל מעריכי

נתון מדגם: $X_1, X_2, ..., X_n$ של משתנים מעריכיים: $X_i \sim \exp(\theta)$ בלתי-תלויים בזוגות.

א.נ.מ עבור θ הוא: $\hat{\theta} = \frac{1}{\overline{X}}$ - מהווה אומד מוטה, וא.נ.מ עבור התוחלת $\frac{1}{\theta}$ הוא $\overline{X}$ א.ח.ה.

מודל אחיד (רציף)

נתון מדגם: $X_1, X_2, ..., X_n$ של משתנים אחידים: $X_i \sim U(0,\theta)$ בלתי-תלויים בזוגות.

א.נ.מ עבור θ הוא: $\hat{\theta} = \max\{X_1, ..., X_n\}$ אינו א.ח.ה.

בכל התפלגות:

א.ח.ה עבור μ הוא: $\hat{\mu} = \overline{X}$.

אומד חסר-הטיה עבור σ^2:

כאשר μ ידוע: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$.

כאשר μ לא-ידוע: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

סטטיסטיקה לאקטואריה

פרק 12 - בדיקת השערות כללית (סיכוי לטעויות ועוצמת מבחן)

תוכן העניינים

בדיקת השערות כללית (סיכוי לטעויות ועוצמת מבחן):

רקע:

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטיקה.
בתהליך זה ישנן שתי השערות שנבדקות:

1. השערת האפס: המסומנות ב- H_0.
2. השערה אלטרנטיבית (השערת המחקר): המסומנת ב- H_1.

בדרך כלל השערת האפס מסמנת את אשר היה מקובל עד עכשיו, את השגרה הנורמה ואילו ההשערה האלטרנטיבית את החדשנות בעצם ההשערה האלטרנטיבית מדברת על הסיבה שהמחקר נעשה.

דוגמה:

ישנה תרופה קיימת למחלה A אשר גורמת ל-10% מהמשתמשים בה לתופעות לוואי. חברת תרופות טוענת שפיתחה תרופה שיעילה באותה מידה, אך מקטינה את הסיכוי לתופעות הלוואי. לכן יש לבצע מחקר שעל סמך תוצאותיו ננסה להכריע איזה השערה נקבל:

H_0: התרופה החדשה הנה קונבנציונאלית וגורמת ל-10% תופעות לוואי.

H_1: התרופה החדשה מקטינה את אחוז הסובלים מתופעות לוואי מתחת ל-10%.

בתהליך של בדיקת השערות יוצרים כלל שנקרא כלל הכרעה.
הכלל יוצר אזורים:

1. אזור דחייה: דחייה של השערת האפס כלומר קבלה של האלטרנטיבה).
2. אזור קבלה: קבלה של השערת האפס ודחייה של האלטרנטיבה.

כלל ההכרעה מתבסס על איזשהו סטטיסטי.
בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה.
המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם.
אם נשנה את כלל ההכרעה אז אנחנו יכולים לקבל מסקנה אחרת, אם נבצע מדגם חדש אז אנחנו עלולים לקבל תוצאה אחרת.

לכן יתכנו טעויות במסקנות שלנו:

	הכרעה		
		H_0	H_1
מציאות	H_0	אין טעות	טעות מסוג 1
	H_1	טעות מסוג 2	אין טעות

הגדרת הטעויות:

טעות מסוג ראשון: להכריע לדחות את H_0 למרות שבמציאות H_0 נכונה.

טעות מסוג שני: להכריע לקבל את H_0 למרות שבמציאות H_1 נכונה.

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

(לדחות H_0) P_{H_0} = (H_0 נכונה | לדחות את H_0) $P = \alpha$.

הסיכוי לבצע טעות מסוג 2:

(לקבל H_0) P_{H_1} = (H_1 נכונה | לקבל את H_0) $P = \beta$.

רמת בטחון:

(לקבל H_0) P_{H_0} = (H_0 נכונה | לקבל את H_0) $P = (\alpha-1)$.

עוצמה:

(לקבל H_0) P_{H_1} = (H_1 נכונה | לדחות את H_0) $P = \pi = (\beta-1)$.

דוגמה (פתרון בהקלטה):

בכד יש 10 כדורים. יתכן ש-5 מהם לבנים והיתר שחורים (כד א' – השערת האפס) או ש-7 מהם לבנים והיתר שחורים (כד ב' – השערה אלטרנטיבית).
כדי להחליט איזה מהכדים ברשותנו, הוחלט להוציא כדור ולהשתמש בכלל ההחלטה הבא: אם הכדור שהוצא הוא לבן שזהו כד ב' H_1.

א. חשבו את רמת המובהקות ואת רמת הביטחון של המבחן המוצע.
ב. חשבו את הסיכוי לטעות מסוג שני והעוצמה של המבחן המוצע.

שאלות:

1) אדם חשוד בביצוע פשע. מהן הטעויות האפשריות בהכרעת הדין?

2) ילד קנה שקית סוכריות אטומה שבה ציפה ל-10 סוכריות תות ו-5 לימון.
ישנה שקית אחרת אותה הוא לא רצה בה 6 סוכריות תות ו-9 לימון.
הוא החליט להוציא באקראי סוכרייה, אם היא תהיה לימון הוא יחזיר את השקית לחנות. מה הסיכויים לכל סוג של טעות בהכרעתו?

3) יהי X מספר שלם הנבחר באקראי מבין המספרים השלמים. הסיכוי ש- X יקבל ערך כלשהו נתון על ידי הנוסחה: $p(X=k)=\frac{1}{n}$ עבור: $k=1,2,......,n$.
נתונות ההשערות הבאות לגבי התפלגות של X: $H_0: n=4$, $H_1: n=6$.
כמו כן נתון כלל ההכרעה הבא: נדחה את השערת האפס אם: $X>3$.
חשבו את הסיכוי לטעות מסוג ראשון וטעות מסוג שני ואת העוצמה?

4) איכות של מוצר מסווגת ל-4 רמות איכות: מצוין, טוב, בינוני וירוד.
להלן התפלגות טיב המוצר בשני מפעלים:

מפעל / איכות	מצוין	טוב	בינוני	ירוד
"היוצר"	0.6	0.2	0.2	0
"שמשון"	0.1	0.2	0.3	0.4

בוחרים ממשלוח מוצר באקראי, אך לא יודעים מאיזה מפעל המשלוח הגיע.
על סמך בדיקת האיכות מנסים להכריע האם מדובר במפעל "היוצר" (השערת האפס) או במפעל "שמשון" (השערה אלטרנטיבית).

א. להלן כלל החלטה: אם מדובר במוצר שטיבו "טוב" נכריע שהמוצר בא ממפעל "שמשון", מהן ההסתברויות לסוגי הטעויות השונים?
ב. להלן כלל החלטה: אם מדובר במוצר שטיבו "בנוני" או גרוע מכך נכריע שהמוצר בא ממפעל "שמשון", מה מהן ההסתברויות לסוגי הטעויות השונים?
ג. איזה כלל החלטה עדיף? נמקו!

5) במטרה לבדוק האם מטבע תקין הטילו אותו 8 פעמים. הוחלט שאם מספר העצים יהיה בין 1 ל-7 כולל יוחלט שהמטבע תקין, אחרת נחליט שהמטבע מזויף.

א. רשמו את השערות המחקר.
ב. מה ההסתברות לטעות מסוג ראשון?
ג. מהי עצמת המבחן אם במציאות אכן המטבע אינו תקין כי הסיכוי לעץ בו הוא 20%.

6) להלן השערות:

$H_0: X \sim t(5)$ - התפלגות T עם חמש דרגות חופש.

$H_1: X \sim Z$ - התפלגות נורמלית.

כלל החלטה: נדחה את השערת האפס אם X גדול מ-2.015.

א. מהי רמת המובהקות של כלל ההחלטה?
ב. מהי העוצמה של כלל ההחלטה?

7) במפעל מסוים נפלטים לאוויר חומרים רעילים. במצב שיגרה העוצמה הממוצעת של החומר הרעיל אמורה להיות 6,000 יחידות עם סטיית תקן 900. במצב חירום העוצמה הממוצעת היא 7,000 עם סטיית תקן 900. במפעל מערכת התראה נתמכת על ידי 9 חיישנים. אם ממוצע העוצמה של החומר הרעיל לפי תשעת החיישנים עולה על 6,600 יחידות מופעלת מערכת ההתראה. נתון שעוצמת הזיהום מתפלגת נורמאלית.

א. מה הסיכוי להתראת שווא? (באיזה סוג טעות מדובר)?
ב. מה הסיכוי שבמצב חירום מערכת ההתראה לא תפעל? (באיזה סוג טעות מדובר)?
ג. מה ההסתברות שאם המצב הוא מצב חירום מערכת ההתראה תפעל? (איך קוראים להסתברות זו)?
ד. בסעיפים הבאים נשנה בכל סעיף נתון מסוים. כל סעיף עומד בפני עצמו, כיצד השינוי ישנה את הסיכוי לטעות מסוג ראשון ושני?
   i. המפעל יקנה עוד 4 חיישנים.
   ii. מצב חרום מוגדר כעת בתוחלת של 7,500 יחידות.
   iii. מערכת ההתראה תופעל אם ממוצע של תשעת החיישנים יהיה מעל 6,700.

8) במטרה לבדוק האם במקום עבודה מסוים פרופורציית הבנים נמוכה מפרופורציית הבנות נדגמו באקראי 10 עובדים. הוחלט שאם מספר הבנים במדגם יהיה לכל היותר 2 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות.

א. מה רמת המובהקות של כלל ההכרעה הנ"ל?
ב. מהי העצמה בהנחה ובחברה 30% בנים?

9) זמן ההשפעה של משכך הכאבים "אופטלנוס" מתפלג נורמאלית עם תוחלת של 40 דקות וסטיית תקן של 12 דקות. חברת התרופות המייצרת את התרופה מנסה לשפר את התרופה כך שתוחלת הזמן עד להשפעה תתקצר. לצורך כך, דגמו 25 מטופלים שיקבלו את התרופה "אופטלנוס פורטה", ממוצע זמן התגובה של המטופלים היה 34.5 דקות. חברת התרופות החליטה מראש שאם ממוצע הזמן עד להשפעה יהיה נמוך מ-35 דקות, היא תמשיך בתהליך שיווק "אופטלנוס פורטה".

א. מהי רמת המובהקות של המבחן המוצע?

ב. על סמך תוצאות המדגם, מהי המסקנה ומהי הטעות האפשרית במסקנה?

ג. מהי עצמת המבחן המוצע אם במציאות התרופה "אופטלנוס פורטה" מפחיתה את התוחלת לכדי 32 דקות?

ד. כיצד תשתנה התשובה לסעיף ג' אם החברה הייתה מחליטה שהיא תמשיך בתהליך שיווק התרופה החדשה כאשר ממוצע המדגם יהיה נמוך מ-36 דקות?

10) ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 120.
מכון טוען שלימודים אצלו מעלים את ממוצע הציונים ביותר מ-30 נקודות. נלקחו 20 שלמדו במכון ו-20 שניגשו לבחינה בלמידה עצמית. הוחלט במשרד פרסום לקבל את טענת המכון רק אם במדגם ממוצע הציונים של אלה שלמדו במכון יהיה גבוהה בלפחות 50 נקודות מאלה שלא היו.

א. מהי רמת המובהקות של המחקר?

ב. מה הסיכוי לעשות טעות מסוג שני II בהנחה שהמכון מעלה את ממוצע הציונים ב-60 נקודות?

ג. כיצד התשובות לסעיף א ו ב' היו משתנות אם מסתבר שסטיית התקן בציוני הפסיכומטרי הינה 100. הסבירו ללא חישוב.

11) קו ייצור נחשב תקין אם יש בו לכל היותר 4% פגומים, ונחשב שאינו תקין אחרת. מנהל האיכות דוגם בכל יום מקו הייצור 500 מוצרים. אם במדגם יהיה לפחות 30 מוצרים פגומים יפסיקו באותו היום את קו הייצור.

א. מה ההסתברות להפסיק את קו הייצור כשהוא תקין. איך קוראים להסתברות זאת?

ב. מה ההסתברות להמשיך ביום מסוים את קו הייצור למרות שאינו תקין כי היו 8% פגומים בקו הייצור. איך קוראים להסתברות זאת?

12) מעוניינים לבדוק האם בפקולטה מסוימת ישנה העדפה לגברים. הוחלט לדגום 200 מתקבלים ועל סמך מספר הבנים לקבוע אם טענת המחקר מתקבלת. חוקר א' קבע רמת מובהקות של 5% וחוקר ב' החליט לקבל את טענת המחקר אם במדגם יהיו לפחות 120 בנים. למי מבין החוקרים רמת מובהקות גדולה יותר?

13) מספר המכוניות הנכנסות לחניון "עזרים" מתפלג פואסונית. בשנה שעברה המכוניות נכנסו לחניון בקצב של 2 מכוניות לדקה. בעקבות תלונות על עומס יתר בכניסה לחניון מעוניין מנהל החניון לבדוק האם קצב כניסת המכוניות לחניון גדל השנה. מנהל החניון החליט לספור את מספר המכוניות שיכנסו לחניון בדקה אקראית. אם מספר המכוניות שיספרו יהיה לפחות 4 יפתח מנהל החניון שער נוסף לחניון.

א. רשמו את השערות מנהל החניון ואת כלל ההחלטה שלו. האם כלל ההכרעה הגיוני?

ב. מהי רמת המובהקות של כלל ההכרעה?

ג. מהי העוצמה של כלל ההחלטה, אם כיום קצב כניסת המכוניות לחניון גדל ל-4 מכוניות בדקה?

14) עודד עובד במפעל שבו מתחילים לעבוד בשעה 8:00. עודד בדרך כלל מאחר לעבודה והמנהל החליט לרשום את שעת הגעתו. המנהל טוען שמשך האיחור של עודד (בדקות), X , הוא משתנה אחיד $U\left(0,60\right)$. עודד טוען שהוא לא מגיע באיחור כה גדול, אלא שהתפלגות X היא בעלת התפלגות מעריכית עם תוחלת איחור של 20 דקות.

לבדיקת טענת המנהל (H_0) כנגד טענת עודד (H_1), המבוסס על משך האיחור של חגי ביום אחד. מוצאים שני כללי הכרעה:

כלל 1: דחה את השערת האפס אם משך האיחור יהיה לפחות 40 דקות.

כלל 2: דחה את השערת האפס אם משך האיחור יהיה לכל היותר 20 דקות.

חשבו את הסיכוי לטעות מסוג ראשון ושני לכל אחת מכללי ההכרעה. מי עדיף?

תשובות סופיות:

1) ראה סרטון וידאו.

2) $\alpha=\frac{1}{3}$, $\beta=\frac{2}{5}$.

3) $\alpha=0.25$, $\beta=0.5$.

4) א. $\alpha=0.2$, $\beta=0.8$. ב. $\alpha=0.2$, $\beta=0.3$. ג. כלל ב'.

5) א. השערות: H_0 -מטבע תקין.
H_1 -מטבע לא תקין.
ב. 0.00781250. ג. 0.1678.

6) א. 0.05. ב.0.022.

7) א. 0.0228. ב. 0.0918. ג. 0.9082.
ד. i. α,β יקטנו. ד. ii. α לא משתנה, β קטנה.
ד. iii. α קטנה, β גדלה.

8) א. 0.055. ב. 0.383.

9) א. 0.0188. ב. טעות מסוג I. ג. 0.8944.
ד. העוצמה תגדל.

10) א. 0.2981. ב. 0.3974. ג. קטן.

11) א. 0.0113. ב. 0.0495.

12) חוקר א'.

13) א. ראה סרטון וידאו. ב. 0.1428. ג. 0.566.

14) להלן טבלת טעויות, ממנה ניתן להסיק שכלל 2 עדיף.

כלל	α	β
1	$\frac{1}{3}$	0.865
2	$\frac{1}{3}$	0.368

סטטיסטיקה לאקטואריה

פרק 13 - הלמה של ניימן פירסון

תוכן העניינים

הלמה של ניימן פירסון:

רקע:

שיטה זו עוזרת לנו לבנות מבחנים בעלי עוצמה מקסימאלית עבור α נתונה.
הלמה של ניימן פירסון אומרת שמבחן בעל עוצמה מקסימלית מתקבל כאשר אזור הדחיה שלו כולל את התוצאות שעבורן יחס הנראות הוא הגבוה ביותר.

נגדיר את יחס הנראות: $x_1, x_2, ..., x_n$ - תוצאות הניסוי.

ההשערות: $H_1 : x_i \sim p_1$, $H_0 : x_i \sim p_0$.

$$\lambda(x_1, x_2, ..., x_n) = \frac{p_{H1}(x_1, x_2, ..., x_n)}{p_{H0}(x_1, x_2, ..., x_n)}$$

המשמעות של יחס הנראות היא פי כמה H_1 יותר סבירה מ- H_0.

עבור משתנים שמתפלגים בדיד:

שלב א: עבור כל תוצאות המדגם האפשריים מחשבים את הסיכויים בהנחת השערת האפס ובהנחת ההשערה האלטרנטיבית.

שלב ב: מחלקים את הסיכויים באופן הבא: $\lambda(x_1, x_2, ..., x_n) = \frac{p_{H1}(x_1, x_2, ..., x_n)}{p_{H0}(x_1, x_2, ..., x_n)}$.

ומקבלים את יחס הנראות.

שלב ג: מסדרים את תוצאות הניסוי על פי סדר יורד מהתוצאה שמניבה את ערך יחס הנראות הגבוה ביותר עד התוצאה שמניבה את ערך יחס הנראות הנמוך ביותר.

שלב ד: מכניסים את התוצאה שמניבה את יחס הנראות הגבוה ביותר לאזור הדחיה ובודקים תחת השערת האפס מהי רמת המובהקות המתקבלת . צוברים את התוצאות לפי העיקרון שהוצג עד שרמת המובהקות לא תעלה על ה- α הרצויה.

דוגמה (הפתרון בהקלטה):

איכות של מוצר מסווגת ל-4 רמות איכות: מצוין, טוב, בינוני וירוד.
להלן התפלגות טיב המוצר בשני מפעלים:

מפעל / איכות	מצוין	טוב	בינוני	ירוד
"היוצר"	0.6	0.15	0.25	0
"שמשון"	0.1	0.2	0.3	0.4

בוחרים ממשלוח מוצר באקראי, אך לא יודעים מאיזה מפעל המשלוח הגיע. על סמך בדיקת האיכות מנסים להכריע האם מדובר במפעל "היוצר" (השערת האפס) או במפעל "שמשון" (השערה אלטרנטיבית). צרו כלל הכרעה לפי הלמה של ניימן פירסון ברמת מובהקות שלא תעלה על 20%.

עבור משתנים שמתפלגים רציף:

שלב א: בונים את פונקציית הצפיפות המשותפת בהנחת השערת האפס ובהנחת ההשערה האלטרנטיבית.

שלב ב: מחלקים את שלב א באופן הבא: $\lambda(x_1,x_2,...,x_n)=\dfrac{f_{H1}(x_1,x_2,...,x_n)}{f_{H0}(x_1,x_2,...,x_n)}$.

ומקבלים את פונקציית יחס הנראות.

שלב ג: מזהים את האזור עבורו יחס הנראות הוא הגבוה ביותר.

שלב ד: לפי ההתפלגות של השערת האפס מוצאים את הערכים הקריטיים באזור שנקבע בסעיף הקודם כך שרמת המובהקות תהיה ה- α שנקבעה מראש.

דוגמה (פתרון בהקלטה):

נתון: $X \sim \exp(\lambda)$.

ההשערות הן: $H_1:\lambda=2$, $H_0:\lambda=1$.

מצאו מבחן בעל עוצמה מקסימלית ברמת מובהקות של 5% על סמך תצפית בודדת.

שאלות:

1) איכות של מוצר מסווגת ל-5 רמות איכות: מצוין, טוב, בינוני, ירוד ופסול.
להלן התפלגות טיב המוצר בשני מפעלים:

מפעל / איכות	מצוין	טוב	בינוני	ירוד	פסול
"היוצר"	0.4	0.2	0.1	0.1	0.2
"שמשון"	0.1	0.2	0.3	0.4	0

בוחרים משלוח מוצר באקראי, אך לא יודעים מאיזה מפעל המשלוח הגיע.
על סמך בדיקת האיכות מנסים להכריע האם מדובר במפעל "היוצר" (השערת האפס) או במפעל "שמשון" (השערה אלטרנטיבית).

א. חשבו את יחס הנראות עבור כל תוצאות המדגם האפשריים.

ב. צרו כלל הכרעה לפי הלמה של ניימן פירסון ברמת מובהקות שלא תעלה על 25%.

ג. מהי עוצמת המבחן שיצרת בסעיף הקודם?

2) מטבע הוטל 3 פעמים ומתבוננים במספר הפעמים שהתקבלה התוצאה ראש.
נסמן ב- p את הסיכוי בהטלה בודדת לקבל את התוצאה ראש.
ההשערות הן: $H_0: p=0.5\ ,\ \ H_1: p=0.25$.
מצאו מבחן ברמת מובהקות שלא תעלה על 30% עם עוצמה מקסימלית.
מהי העוצמה?

3) בכד א' 7 כדורים לבנים ו-8 שחורים. בכד ב' 8 כדורים לבנים ו-7 שחורים.
אדם בוחר כד וממנו מוציא באקראי 4 כדורים ללא החזרה. הוא מתבוננן במספר הכדורים הלבנים שהוצאו ומודיע לך את המספר המתקבל.
יש לבנות כלל הכרעה על סמך המספר המתקבל שיכריע האם מדובר בהוצאה מכד א' (השערת האפס) או מכד ב' (השערה אלטרנטיבית).

א. בנו כלל הכרעה בעל עוצמה מקסימלית ברמת מובהקות שלא תעלה על 10%.

ב. מהי רמת המובהקות של כלל ההכרעה שבנית בסעיף הקודם?

ג. מה הסיכוי לטעות מסוג שני של כלל ההכרעה שבנית?

4) בצרור מפתחות 5 מפתחות שרק אחד פותח את הדלת. על סמך מספר הניסיונות לפתיחת הדלת יש להחליט האם הניסיונות נעשו ללא החזרה (השערת האפס) או עם החזרה (השערה אלטרנטיבית) של המפתחות לצרור.
מצאו מבחן לפי הלמה של ניימן פירסון ברמת מובהקות שלא עולה על 0.25.

5) מספר תאונות הדרכים בכביש 4 מתפלג פואסונית עם קצב של תאונה ביממה. לאחרונה התעורר החשד שתוחלת מספר התאונות בכביש עלתה לקצב של שתי תאונות ביממה. דגמו 4 ימים אקראיים וקיבלו את מספר התאונות הבאות ליממה: 1, 0, 3, 3.

א. נסחו את הבעיה ובנו מבחן MP עם: $\alpha \le 0.1$.

ב. מהי מסקנתך ומהי הטעות האפשרית במסקנה?

ג. מה הסיכוי להכריע שכיום קצב תאונות הדרכים בכביש מספר 4 עלה לשתי תאונות ביממה שאכן כך הדבר באמת?
פתרו על סמך המבחן של סעיף א'.

6) התפלגות זמן ההמתנה לקופה בסופרמרקט מתפלג מעריכית. בעל הסופרמרקט טוען שתוחלת זמן ההמתנה היא 5 דקות אך הלקוחות חושדים שהתוחלת גבוהה יותר ושווה ל-10 דקות.

א. רשמו את השערות המחקר וחשבו את פונקציית יחס הנראות על סמך זמן המתנה של לקוח אקראי לקופה.

ב. מה כיוון אזור הדחייה של השערת האפס?

ג. מצאו את אזור הדחייה עבור רמת מובהקות של 5%.

ד. בהמשך לסעיף הקודם, מה הסיכוי להכריע לטובת בעל הסופרמרקט בטעות?

7) יהי X תצפית בודדת מפונקציית הצפיפות הבאה, כאשר θ פרמטר חיובי:

$$f_\theta(x)=\begin{cases}2\theta X+1-\theta & 0\le x\le 1\\ 0 & otherwise\end{cases} \qquad \frac{n!}{r!(n-r)!}$$

ההשערות הן: $H_0:\theta=0 \;,\; H_1:\theta=1$.

א. הוכיחו שמבחן MP (עוצמה מקסימלית) עם רמת מובהקות α יהיה: $C=\{X>1-\alpha\}$.

ב. הוכיחו שהעוצמה של המבחן שנמצאה היא: $2\alpha-\alpha^2$.

ג. מצאו את כלל ההכרעה והעוצמה עבור: $\alpha=0.05$.

8) מחשב חניון "אתרים" רושם את זמן כניסת כל מכונית לחניון. ישנו חשד שעקב תקלה המחשב מבצע את הרישום לכל מכונית שניה. נסמן ב- X את הזמן בדקות בין רישום לרישום.

אם הרישום הוא תקין, ההתפלגות היא מעריכית: $f_0(x)=\lambda\cdot e^{-\lambda x}$.

אם הרישום הוא לפי החשד, פונקציית הצפיפות היא: $f_1(x)=\lambda^2\cdot x\cdot e^{-\lambda x}$.

כאשר מדובר באותו פרמטר λ.

א. רשמו את ההשערות.

ב. מצאו מבחן בעל עוצמה מקסימלית כדי לבדוק את ההשערות.

ג. פתרו עבור רמת מובהקות של 5% ו- $\lambda=1$.

ד. רשמו את איזו הדחייה עבור שתי תצפיות אקראיות של X.

9) יהי $X_1,\ldots,X_n$ מדגם מקרי מהתפלגות בעלת פונקציית הצפיפות הבאה, כאשר θ פרמטר חיובי: $f_\theta(x)=\begin{cases}\theta e^{1-\theta x} & x\ge\dfrac{1}{\theta}\\ 0 & otherwise\end{cases}$.

א. מצאו מבחן בעל עוצמה מקסימאלית לבדיקת ההשערות: $H_0:\theta=1$ כנגד: $H_1:\theta=2$, על סמך תצפית בודדת (ברמת מובהקות 5%).

ב. מהי עוצמת המבחן שמצאת?

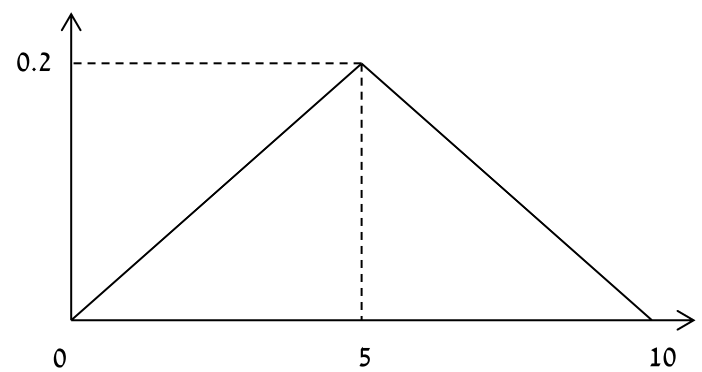
ג. כעת נשנה את הערך תחת האלטרנטיבה ל-3 במקום 2.
בחרו בתשובה הנכונה ונמקו:

i. אפשר לומר ללא חישוב נוסף שהעוצמה תגדל.

ii. אפשר לומר ללא חישוב נוסף שהעוצמה תקטן.

iii. יש לחשב כדי להחליט.

10) נתונה פונקציית הצפיפות הבאה שנסמנה ב- $f_1(x)$:



כמו כן נתון ש : $f_0(x)=\begin{cases}0.1 & 0\le x\le 10\\ 0 & else\end{cases}$.

א. עבור השערות : $H_0 : f = f_0$ כנגד : $H_1 : f = f_1$, מצאו את צורת אזור הדחיה של מבחן בעל עצמה מקסימלית, על סמך תצפית בודדת.

ב. בהינתן : $\alpha = 0.05$, מצאו את כלל הכרעה מתאים בעל עוצמה מקסימלית.

(11 X הוא משתנה רציף המוגדר בין 0 ל-20.

להלן השערות מחקר : $H_1 : f(X)=\dfrac{e^{-\frac{X}{4}}}{4\left(1-e^{-5}\right)}$, $H_0 : X \sim U(0,20)$.

יש לבנות מבחן בעל עוצמה מקסימלית ברמת מובהקות של 10% על סמך הממוצע 100 תצפיות אקראיות.

תשובות סופיות:

1) א. להלן טבלה: ב. ראה סרטון. ג. 0.7.

המפעל	ירוד	בינוני	טוב	מצוין	פסול
H_0	0.1	0.1	0.2	0.4	0.2
H_1	0.4	0.3	0.2	0.1	0
$\gamma(x)$	4	3	1	0.25	0

2) $C=\{X=0\}$, עוצמה $\frac{27}{64}$.

3) א. $C=\{X=4\}$. ב. 0.0256. ג. 0.9487.

4) $C=\{X=1 \ or \ X\geq 6\}$.

5) א. נדחה את השערת האפס אם מספר התאונות הכולל ב-4 הימים יהיה לפחות 8 תאונות. ב. נקבל את השערת האפס (טעות מסוג שני). ג. 0.5471.

6) א. H_0: התוחלת של זמן ההמתנה 5 דקות.
H_1: התוחלת של זמן המתנה 10 דקות.
פונקציית יחס הנראות: $0.5e^{0.1x}$.
ב. $c=\{x\geq k\}$. ג. $c=\{x\geq 14.98\}$. ד. 0.776.

7) א. שאלת הוכחה. ב. שאלת הוכחה. ג. $c=\{x>0.95\}$, 0.0975.

8) א. השערת האפס: שומר הלילה רושם כל מבקר אשר ניכנס לבניין.
ההשערה האלטרנטיבית: שומר הלילה מדלג ברישום על מבקר, רושם אחד כן ואחד לא.
ב. אזור הדחייה יהיה היכן שיחס הנראות הינו גבוה לכן הוא יהיה מאינסוף ועד ערך k מסוים.
ג. נדחה את השערת האפס אם: $x>2.996$. ד. $x_1\cdot x_2>k$.

9) א. נדחה את H_0 עבור: $\frac{1}{2}\leq x\leq 1.0513$. ב. חישוב עוצמת המבחן: 0.668.
ג. i. אפשר לומר ללא חישוב נוסף שהעוצמה תגדל.

10) א. $c=\{|x-5|\leq a\}$. ב. $c=\{4.75\leq x\leq 5.25\}$.

11) כלל ההכרעה הוא נדחה את השערת האפס אם: $\bar{D}\leq 9.26$.

סטטיסטיקה לאקטואריה

פרק 14 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

הקדמה:

רקע:

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטיקה.
בבדיקת השערות על פרמטרים נעבוד לפי השלבים הבאים:

שלב א: נזהה את הפרמטר הנחקר.

שלב ב: נרשום את השערות המחקר.
השערת האפס המסומנות ב- H_0.
בדרך כלל השערת האפס מסמלת את אשר היה מקובל עד עכשיו, את השגרה, הנורמה.
השערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1.
ההשערה האלטרנטיבית מסמלת את החדשנות בעצם ההשערה האלטרנטיבית מדברת על הסיבה שהמחקר נעשה היא שאלת המחקר.

שלב ג: נבדוק האם התנאים לביצוע התהליך מתקיימים ונניח הנחות במידת הצורך.

שלב ד: נרשום את כלל ההכרעה. בתהליך של בדיקת השערות יוצרים כלל שנקרא כלל הכרעה. הכלל יוצר אזורים שנקראים:

1. **אזור דחייה:**
דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.
2. **אזור קבלה:**
קבלה של השערת האפס ודחייה של האלטרנטיבה. כלל ההכרעה מתבסס על איזשהו סטטיסטי. אזור הדחיה מוכתב על ידי סיכון שלוקח החוקר מראש שנקרא רמת מובהקות ומסומן ב- α.

שלב ה: בתהליך יש ללכת לתוצאות המדגם ולחשב את הסטטיסטי המתאים ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה.

שלב ו: להסיק מסקנה בהתאם לתוצאות המדגם.

דוגמה (פתרון בהקלטה):

משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום לידתם בישראל 3300 גרם.
משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון.
להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$S = 280$, $\overline{X} = 3120$, $n = 20$.

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

שאלות:

בשאלות הבאות, ענו על הסעיפים הבאים:

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

1) ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות.
מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים.
משרד החינוך החליט לתת למורה 36 תלמידים אקראיים.
ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5.

2) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.

3) במשך שנים אחוז המועמדים שהתקבל לפקולטה למשפטים היה 25%.
השנה מתוך מדגם של 120 מועמדים התקבלו 22.
מחקר מעוניין לבדוק האם השנה מקשים על הקבלה לפקולטה למשפטים.

4) בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. רוצים לבדוק ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

תשובות סופיות:

1) א. נבחנים בבגרות באנגלית.
ב. ציון.
ג. ממוצע הציונים בשיטת לימוד חדשה.
ד. $H_0: \mu = 72$, $H_1: \mu > 72$.

2) א. משקאות בבקבוק של חברה מסוימת.
ב. נפח משקה בסמ״ק.
ג. ממוצע נפח המשקה בבקבוק.
ד. $H_0: \mu = 500$, $H_1: \mu < 500$.

3) א. מועמדים לפקולטה למשפטים.
ב. משתנה דיכוטומי (התקבל, לא התקבל).
ג. אחוז הקבלה.
ד. $H_0: p = 0.25$, $H_1: p < 0.25$.

4) א. אזרחים בוגרים במשק.
ב. משתנה דיכוטומי (מובטל, עובד).
ג. אחוז האבטלה כיום.
ד. $H_0: p = 0.08$, $H_1: p \neq 0.08$.

סוגי טעויות:

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה.
הכלל יוצר אזורים שנקראים:

1. אזור דחייה – דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.
2. אזור קבלה – קבלה של השערת האפס ודחייה של האלטרנטיבה.

כלל ההכרעה מתבסס על איזשהו סטטיסטי .
בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה – המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם. אם נשנה את כלל ההכרעה אז אנחנו יכולים לקבל מסקנה אחרת. אם נבצע מדגם חדש אז אנחנו עלולים לקבל תוצאה אחרת.
לכן יתכנו טעויות במסקנות שלנו:

		הכרעה	
		H_0	H_1
מציאות	H_0	אין טעות	טעות מסוג 1
	H_1	טעות מסוג 2	אין טעות

הגדרת הטעויות:

טעות מסוג ראשון: להכריע לדחות את H_0 למרות שבמציאות H_0 נכונה.

טעות מסוג שני: להכריע לקבל את H_0 למרות שבמציאות H_1 נכונה.

דוגמה (פתרון בהקלטה):

אדם חשוד בביצוע עבירה ונתבע בבית המשפט.
אילו סוגי טעויות אפשריות בהכרעת הדין?

שאלות:

1) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25. בסופו של דבר הוחלט להכריע לטובת חברת המשקאות.

א. רשמו את השערות המחקר.

ב. מה מסקנת המחקר?

ג. איזו סוג טעות יתכן וביצעו במחקר?

2) במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.

א. האם ניתן לדעת אם בוצע טעות במחקר?

ב. מה סוג הטעות האפשרית?

3) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. ישנה טענה שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מה השערות המחקר?

ה. מה מסקנת המחקר?

ו. מהי סוג הטעות האפשרית במחקר?

תשובות סופיות:

1) א. $\begin{matrix} H_0: \mu = 500 \\ H_1: \mu < 500 \end{matrix}$. ב. לא דחינו את H_0. ג. טעות מסוג שני.

2) א. לא ניתן לדעת. ב. טעות מסוג ראשון.

3) א. משפחות כיום. ב. מס' הילדים.

ג. תוחלת מספר הילדים למשפחה כיום. ד. $\begin{matrix} H_0: \mu = 2.3 \\ H_1: \mu < 2.3 \end{matrix}$.

ה. לא לדחות את H_0. ו. טעות מסוג שני.