

סטטיסטיקה והסתברות



תוכן העניינים

1.	יסודות ההסתברות	1
2.	פעולות בין מאורעות (חיתוך ואיחוד) - מאורעות זרים ומכילים	5
3.	קומבינטוריקה - כלל המכפלה	13
4.	קומבינטוריקה - תמורה - סידור עצמים בשורה	17
5.	קומבינטוריקה - דגימה סידורית ללא החזרה ועם החזרה	20
6.	קומבינטוריקה - דגימה ללא סדר וללא החזרה	22
7.	קומבינטוריקה - שאלות מסכמות	25
8.	הסתברות מותנית-במרחב מדגם אחיד	30
9.	הסתברות מותנית - מרחב לא אחיד	33
10.	דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה	37
11.	תלות ואי תלות בין מאורעות	42
12.	שאלות מסכמות בהסתברות	45
13.	המשתנה המקרי הבדיד - פונקציית ההסתברות	48
14.	המשתנה המקרי הבדיד - תוחלת - שונות וסטיית תקן	52
15.	המשתנה המקרי הבדיד - טרנספורמציה לינארית	56
16.	תוחלת ושונות של סכום משתנים מקריים	59
17.	התפלגויות בדידות מיוחדות - התפלגות בינומית	62
18.	התפלגויות בדידות מיוחדות - התפלגות אחידה	66
19.	התפלגויות בדידות מיוחדות - התפלגות פואסונית	69
20.	המשתנה המקרי הרציף - התפלגויות כלליות (שימוש באינטגרלים)	72
21.	התפלגויות רציפות מיוחדות - התפלגות נורמלית	81
22.	הסקה סטטיסטית - הקדמה	89
23.	התפלגות הדגימה ומשפט הגבול המרכזי	92

תוכן העניינים

24.	מושגי יסוד באמידה	99
25.	רווח סמך לתוחלת (ממוצע)	104
26.	רווח סמך לפרופורציה	114
27.	רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים	120
28.	רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים	126
29.	מבוא לבדיקת השערות על פרמטרים	128
30.	בדיקת השערות על תוחלת (ממוצע)	134
31.	בדיקת השערות על פרופורציה	166
32.	בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים	179
33.	בדיקת השערות לתוחלת ההפרש במדגמים מזווגים	194
34.	הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות	204
35.	שאלות מסכמות בבדיקת השערות	207
36.	מבחני חי בריבוע	223
37.	משתנה דו-מימדי בדיד - פונקציית הסתברות משותפת	228
38.	משתנה דו מימדי בדיד - מתאם בין משתנים	234
39.	המשתנה המקרי הדו ממדי - קומבינציות ליניאריות	241
40.	המשתנה המקרי הדו ממדי הבדיד - שאלות מסכמות	244
41.	המשתנה המקרי הדו ממדי הרציף	252

סטטיסטיקה והסתברות

פרק 1 - יסודות ההסתברות

תוכן העניינים

1. כללי 1

הגדרות יסודיות:

רקע:

ניסוי מקרי: תהליך לו כמה תוצאות אפשריות. התוצאה המתקבלת נודעת רק לאחר ביצוע התהליך. למשל: תוצאה בהטלת קובייה, מזג האוויר בעוד שבועיים.

מרחב מדגם: כלל התוצאות האפשריות בניסוי המקרי. לדוגמה, בהטלת קובייה: $\{1, 2, 3, 4, 5, 6\}$, או: מזג האוויר בעוד שבועיים: $\{\text{נאה, שרבי, מושלג, גשום, מעונן, חלקית, אביד}\}$.

מאורע: תת קבוצה מתוך מרחב במדגם. מסומן באותיות: A, B, C . בהטלת קובייה למשל, המאורע 'לקבל לפחות 5' יסומן: $A = \{5, 6\}$. המאורע 'לקבל תוצאה זוגית' יסומן: $B = \{2, 4, 6\}$.

גודל מרחב המדגם: מספר התוצאות האפשריות במרחב המדגם. בהטלת קובייה למשל נקבל: $|\Omega| = 6$.

גודל המאורע: מספר התוצאות האפשריות במאורע עצמו. למשל, בהטלת הקובייה האירועים הקודמים יסומנו: $|A| = 2, |B| = 3$.

מאורע משלים: מאורע המכיל את כל התוצאות האפשריות במרחב המדגם פרט לתוצאות במאורע אותו הוא משלים. למשל, בהטלת הקובייה: $\bar{A} = \{1, 2, 3, 4\}$, $\bar{B} = \{1, 3, 5\}$.

מרחב מדגם אחיד (סימטרי): מרחב מדגם בו לכל התוצאות במרחב המדגם יש את אותה עדיפות, אותה סבירות למשל, קובייה הוגנת, אך לא כמו מזג האוויר בשבוע הבא.

הסתברות במרחב מדגם אחיד: במרחב מדגם אחיד הסיכוי למאורע יהיה: $P(A) = \frac{|A|}{|\Omega|}$.

דוגמה: מה הסיכוי בהטלת קובייה לקבל לפחות 5? $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6}$.

דוגמה: מה הסיכוי בהטלת קובייה לקבל תוצאה זוגית? $P(B) = \frac{|B|}{|\Omega|} = \frac{3}{6}$.

הסתברות במרחב לא אחיד: תחושב לפי השכיחות היחסית: $\frac{f}{n}$.

דוגמה:

להלן התפלגות הציונים בכיתה מסוימת:

הציון x	מספר התלמידים – השכיחות f
5	2
6	4
7	8
8	5
9	4
10	2

מה ההסתברות שתלמיד אקראי שניבחר בכיתה קיבל את הציון 8? $\frac{f}{n} = \frac{5}{25} = 0.2$.

מה ההסתברות שתלמיד אקראי שניבחר בכיתה יכשל? $\frac{f}{n} = \frac{2}{25} = 0.08$.

הסתברות למאורע משלים: הסתברות לקבוצת המשלים של המאורע ביחס למרחב המדגם: $P(\bar{A}) = 1 - P(A)$. למשל, בדוגמה הקודמת הסיכוי לעבור את הבחינה יכול

להיות מחושב לפי הסיכוי להיכשל: $P(A) = 1 - \frac{2}{25} = \frac{23}{25}$.

שאלות:

(1) מהאותיות E, F ו- G יש ליצור מילה בת 2 אותיות, לא בהכרח בת משמעות.

א. הרכיבו את כל המילים האפשריות.

ב. רשמו את המקרים למאורע:

i. במילה נמצאת האות E .

ii. במילה האותיות שונות.

ג. רשמו את המקרים למאורע $\bar{A}$.

(2) מטילים זוג קוביות.

א. רשמו את מרחב המדגם של הניסוי. האם מרחב המדגם אחיד?

ב. רשמו את כל האפשרויות לאירועים הבאים:

i. סכום התוצאות 7.

ii. מכפלת התוצאות 12.

ג. חשבו את הסיכויים לאירועים שהוגדרו בסעיף ב'.

(3) נבחר באקראי ספרה מבין הספרות 0-9.

א. מה ההסתברות שהספרה שנבחרה גדולה מ-5?

ב. מה ההסתברות שהספרה שנבחרה היא לכל היותר 3?

ג. מה ההסתברות שהספרה שנבחרה היא אי זוגית?

(4) להלן התפלגות מספר מקלטי הטלוויזיה עבור כל משפחה בישוב מסוים:

מספר משפחות	22	28	18	22	10
מספר מקלטים	0	1	2	3	4

נבחרה משפחה באקראי מהישוב.

א. מה ההסתברות שאין מקלטים למשפחה?

ב. מה ההסתברות שיש מקלטים למשפחה?

ג. מה ההסתברות שיש לפחות 3 מקלטים למשפחה?

(5) להלן התפלגות מספר המכוניות למשפחה ביישוב "עדן":

מספר משפחות	20	40	100	30	10
מספר מכוניות	0	1	2	3	4

נבחרה משפחה אקראית מן הישוב.

א. מה ההסתברות שאין לה מכוניות?

ב. מה ההסתברות שבבעלות המשפחה לפחות 3 מכוניות?

ג. מה הסיכוי שבבעלותה פחות מ-3 מכוניות?

- (6) נטיל מטבע רגיל 3 פעמים. בצד אחד של המטבע מוטבע עץ ובצד השני פלי.
 א. רשמו את מרחב המדגם של הניסוי. האם מרחב המדגם הוא אחיד?
 ב. רשמו את כל האפשרויות לאירועים הבאים:
 i. התקבל פעם אחת עץ.
 ii. התקבל לפחות פלי אחד.
 ג. מהו המאורע המשלים ל-D?
 ד. חשבו את הסיכויים לאירועים שהוגדרו בסעיפים ב-ג.

תשובות סופיות:

- (1) א. $\Omega = \{EE, EF, EG, FE, FF, FG, GE, GF, GG\}$
 ב. $A = \{EE, EF, EG, FE, GE\}$, $B = \{EF, EG, FE, FG, GE, GF\}$
 ג. $\bar{A} = \{FF, FG, GF, GG\}$
- (2) א. $\Omega = \left\{ \begin{pmatrix} (1,1) & (2,1) & (3,1) & (5,1) & (4,1) & (6,1) \\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{pmatrix} \right\}$
 ב. $A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$, $C = \{(2,6), (3,4), (4,3), (6,2)\}$
 ג. הסיכוי ל-A: $\frac{1}{6}$. הסיכוי ל-B: $\frac{1}{9}$.
- (3) א. 0.4 ב. 0.4 ג. 0.5
 (4) א. 0.22 ב. 0.78 ג. 0.32
 (5) א. 0.1 ב. 0.2 ג. 0.8
 (6) א. $\Omega = \{PPP, PPE, PEP, EPP, PEE, EPE, EEP, EEE\}$
 ב. $A = \{PPE, PEP, EPP\}$, $D = \{PPP, PPE, PEP, EPP, PEE, EPE, EEP\}$
 ג. $\bar{D} = \{EEE\}$
 ד. $\frac{1}{8}$

סטטיסטיקה והסתברות

פרק 2 - פעולות בין מאורעות (חיתוך ואיחוד) - מאורעות זרים ומכילים

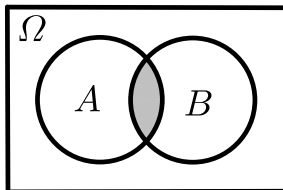
תוכן העניינים

1. כללי 5

פעולות בין מאורעות (חיתוך ואיחוד) – מאורעות זרים ומכילים:

רקע:

פעולת חיתוך:

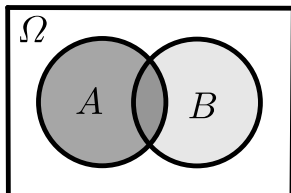


נותנת את המשותף בין המאורעות הנחתכים.
חיתוך בין המאורע A למאורע B יסומן כך: $A \cap B$.
מדובר בתוצאות שנמצאות ב- A וגם ב- B .

דוגמה:

בהטלת קובייה, למשל, האפשרויות לקבל לפחות 5 הן: $A = \{5, 6\}$.
האפשרויות לקבל תוצאה זוגית הן: $B = \{2, 4, 6\}$.
החיתוך שביניהם הוא: $A \cap B = \{6\}$.

פעולת איחוד:



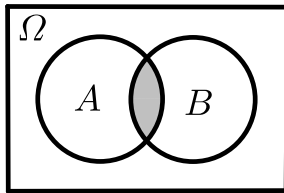
נותנת את כל האפשרויות שנמצאות לפחות באחת מהמאורעות, ומסומנת: $A \cup B$.
הפעולה נותנת את אשר נמצא ב- A או ב- B .
כלומר, לפחות אחד מהמאורעות קורה.

דוגמה:

בהטלת קובייה האפשרויות לקבל לפחות 5 הן: $A = \{5, 6\}$.
האפשרויות לקבל תוצאה זוגית: $B = \{2, 4, 6\}$.
האפשרויות לקבל לפחות 5 וגם תוצאה זוגית: $A \cup B = \{2, 4, 5, 6\}$.

דוגמה (הפתרון נמצא בהקלטה):

סטודנט ניגש בסמסטר לשני מבחנים. מבחן בסטטיסטיקה ומבחן בכלכלה. ההסתברות שלו לעבור את המבחן בסטטיסטיקה הוא 0.9, ההסתברות שלו לעבור את המבחן בכלכלה הוא 0.8 וההסתברות לעבור את המבחן בסטטיסטיקה ובכלכלה היא 0.75.
מה ההסתברות שלו לעבור את המבחן בסטטיסטיקה בלבד?
מה ההסתברות שלו להיכשל בשני המבחנים?
מה ההסתברות לעבור לפחות מבחן אחד?

נוסחת החיבור לשני מאורעות:

ההסתברות של איחוד מאורעות תחושב ע"י הקשר הבא:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

חוקי דה מורגן לשני מאורעות:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

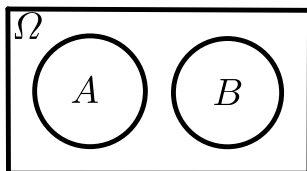
$$P(A \cap B) = 1 - P(\bar{A} \cup \bar{B})$$

$$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$$

שיטת ריבוע הקסם:

השיטה רלבנטית רק אם יש שני מאורעות במקביל בדומה לתרגיל הקודם:

	$\bar{A}$	A	
B	$P(\bar{A} \cap B)$	$P(A \cap B)$	$P(B)$
$\bar{B}$	$P(\bar{A} \cap \bar{B})$	$P(A \cap \bar{B})$	$P(\bar{B})$
	$P(\bar{A})$	$P(A)$	1

מאורעות זרים:

מאורעות זרים הם כאשר אין להם אף איבר משותף:
 $A \cap B = \{ \}$. כלומר, הם לא יכולים להתרחש בו זמנית.

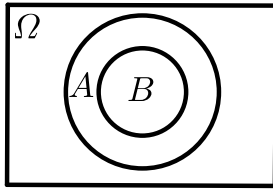
ההסתברות של חיתוך המאורעות היא אפס: $P(A \cap B) = 0$.

ההסתברות של איחוד המאורעות תחושב: $P(A \cup B) = P(A) + P(B)$.

דוגמה:

בהטלת קובייה, האפשרויות לקבל לפחות 5 הן: $A = \{5, 6\}$ והאפשרות לקבל 3

היא: $B = \{3\}$, ולכן החיתוך ביניהם הוא אפס, כלומר: $A \cap B = \{ \}$.

מאורעות מוכלים:

נתונים שני מאורעות A ו- B , השונים מאפס.
נאמר שהמאורע B מוכל במאורע A אם כל איברי
המאורע B כלולים במאורע A ונרשום: $B \subset A$.

מאורע A מכיל את מאורע B כל התוצאות שנמצאות ב- B
מוכלות בתוך מאורע A .

קשר זה מסומן באופן הבא: $B \subset A$.

$$A \cap B = B \quad P(A \cap B) = P(B)$$

$$A \cup B = A \quad P(A \cup B) = P(A)$$

$$\begin{aligned} A &= \{2, 4, 6\} \\ B &= \{2, 4\} \end{aligned} \quad \text{למשל:}$$

שאלות:

- (1) מהאותיות E, F ו- G יוצרים מילה בת 2 אותיות – לא בהכרח בת משמעות. נגדיר את המאורעות הבאים:
- A - במילה נמצאת האות E .
 - B - במילה אותיות שונות.
- א. רשמו את כל האפשרויות לחיתוך A עם B .
- ב. רשמו את כל האפשרויות לאיחוד של A עם B .
- (2) תלמיד ניגש בסמסטר לשני מבחנים מבחן בכלכלה ומבחן בסטטיסטיקה. נגדיר את המאורעות הבאים:
- A - לעבור את המבחן בסטטיסטיקה.
 - B - לעבור את המבחן בכלכלה.
- היעזרו בפעולות חיתוך, איחוד ומשלים בלבד כדי להגדיר את המאורעות הבאים וסמנו בדיאגרמת וון את השטח המתאים:
- א. התלמיד עבר רק את המבחן בכלכלה.
 - ב. התלמיד עבר רק את המבחן בסטטיסטיקה.
 - ג. התלמיד עבר את שני המבחנים.
 - ד. התלמיד עבר לפחות מבחן אחד.
 - ה. התלמיד נכשל בשני המבחנים.
 - ו. התלמיד נכשל בכלכלה.
- (3) נתבקשתם לבחור ספרה באקראי. נגדיר את A להיות הספרה שנבחרה היא זוגית. נגדיר את B להיות הספרה שנבחרה קטנה מ-5.
- א. רשמו את כל התוצאות למאורעות הבאים:
- $$A \cup B, A \cap B, \bar{B}, B, A$$
- ב. חשבו את ההסתברויות לכל המאורעות מהסעיף הקודם.
- (4) נסמן ב- Ω את מרחב המדגם וב- ϕ קבוצה ריקה.
- נתון כי A הינו מאורע בתוך מרחב המדגם.
- להלן מוגדרים מאורעות שפתרונם הוא Ω או ϕ או A .
- קבעו עבור כל מאורע מה הפתרון שלו:
- $$A \cup \bar{A}, \bar{\phi}, A \cap \bar{A}, A \cup \Omega, A \cap \Omega, A \cup \phi, A \cap \phi, \bar{A}$$

(5) הוגדרו המאורעות הבאים :

A - אדם שגובהו מעל 1.7 מטר

B - אדם שגובהו מתחת ל-1.8 מטר.

קבעו את גובהם של האנשים הבאים :

א. $A \cap B$

ב. $A \cup B$

ג. $\overline{A} \cap B$

ד. $\overline{A} \cup \overline{B}$

ה. $\overline{\overline{A}}$

(6) נגדיר את המאורעות הבאים :

A - אדם דובר עברית.

B - אדם דובר ערבית.

C - אדם דובר אנגלית.

השתמשו בפעולות איחוד, חיתוך והשלמה לתיאור המאורעות הבאים :

א. אדם דובר את כל שלוש השפות.

ב. אדם דובר רק עברית.

ג. אדם דובר לפחות שפה אחת מתוך השפות הללו.

ד. אדם אינו דובר אנגלית.

ה. קבוצת התלמידים שדוברים שתי שפות בדיוק (מהשפות הנ"ל).

(7) שתי מפלגות רצות לכנסת הבאה. מפלגת "גדר" תעבור את אחוז החסימה בהסתברות של 0.08 ומפלגת "עתיד" תעבור את אחוז החסימה בהסתברות של 0.20. בהסתברות של 76% שתי המפלגות לא תעבורנה את אחוז החסימה.

א. מה ההסתברות שלפחות אחת מהמפלגות תעבור את אחוז החסימה?

ב. מה ההסתברות ששתי המפלגות תעבורנה את אחוז החסימה?

ג. מה ההסתברות שרק מפלגות "עתיד" תעבור את אחוז החסימה?

(8) במקום עבודה מסוים 40% מהעובדים הם גברים. כמו כן, 20% מהעובדים הם אקדמאים. 10% מהעובדים הינן נשים אקדמאיות.

א. איזה אחוז מהעובדים הם גברים אקדמאיים?

ב. איזה אחוז מהעובדים הם גברים או אקדמאיים?

ג. איזה אחוז מהעובדים הם נשים לא אקדמאיות?

(9) הסיכוי של מניה A לעלות הנו 0.5 ביום מסוים והסיכוי של מניה B לעלות ביום מסוים הנו 0.4. בסיכוי של 0.7 לפחות אחת מהמניות תעלה ביום מסוים. חשבו את ההסתברויות הבאות לגבי שתי המניות הללו ביום מסוים:

א. ששתי המניות תעלנה.

ב. שאף אחת מהמניות לא תעלנה.

ג. שמניה A בלבד תעלה.

(10) מטילים זוג קוביות, אדומה ושחורה. נגדיר את המאורעות הבאים:

A - בקובייה האדומה התקבלה התוצאה 4 ובשחורה 2.

B - סכום התוצאות משתי הקוביות הוא 6.

C - מכפלת התוצאות בשתי הקוביות היא 10.

א. האם A ו-B מאורעות זרים?

ב. האם המאורע B מכיל את המאורע A?

ג. האם A ו-C מאורעות זרים?

ד. האם A ו-C מאורעות משלימים?

(11) עבור המאורעות A ו-B ידועות ההסתברויות הבאות: $P(A) = 0.6$,

$$P(\bar{A} \cap \bar{B}) = 0.1, P(B) = 0.3$$

א. האם A ו-B מאורעות זרים?

ב. חשבו את $P(\bar{A} \cap B)$.

(12) מטבע הוטל פעמיים. נגדיר את המאורעות הבאים:

A - קיבלנו עץ בהטלה הראשונה.

B - קיבלנו לפחות עץ אחד בשתי ההטלות.

איזו טענה נכונה?

א. A ו-B מאורעות זרים.

ב. A ו-B מאורעות משלימים.

ג. B מכיל את A.

ד. A מכיל את B.

(13) בהגרלה חולקו 100 כרטיסים. על 3 מהם רשום חופשה ועל 2 מהם רשום מחשב. שאר הכרטיסים ריקים. אדם קיבל כרטיס אקראי.

א. מה הסיכוי לזכות בחופשה או במחשב? האם המאורעות הללו זרים?

ב. מה ההסתברות לא לזכות בפרס?

14 נתון כי: $P(A) = 0.3$, $P(B) = 0.25$, $P(A \cup B) = 0.49$

א. חשבו את הסיכוי ל- $P(A \cap B)$.

ב. האם A ו- B מאורעות זרים?

ג. מה ההסתברות שרק A יקרה או שרק B יקרה?

15 A ו- B מאורעות זרים. נתון ש: $2 \cdot P(B \cap \bar{A}) = P(A \cap \bar{B}) = P(\bar{A} \cap \bar{B})$

מה הסיכוי למאורע A ומה ההסתברות למאורע B ?

16 קבעו אילו מהטענות הבאות נכונות:

א. $A \cap B = B \cap A$

ב. $\overline{A \cup B} = A \cap B$

ג. $A \cap B \cap C = A \cap B \cap (C \cup B)$

ד. $\overline{A \cap B \cap C} = \bar{A} \cup \bar{B} \cup \bar{C}$

תשובות סופיות:

- (1) א. $A \cap B = \{EG, EF, FE, GE\}$ ב. $A \cup B = \{EG, EF, EE, FE, GE, EG, GF\}$
- (2) א. $B \cap \bar{A}$ ב. $A \cap \bar{B}$ ג. $A \cap B$ ד. $A \cup B$ ה. $\bar{A} \cap \bar{B}$ ו. $\bar{B}$
- (3) א. $\bar{B} = 5, 6, 7, 8, 9, B = 0, 1, 2, 3, 4, A = 0, 2, 4, 6, 8$
 $A \cup B = 0, 2, 4, 6, 8, 1, 3, A \cap B = 0, 2, 4$
- ב. $P(A \cup B) = 0.7, P(A \cap B) = 0.3, P(\bar{B}) = 0.5, P(B) = 0.5, P(A) = 0.5$
- (4) $\bar{\bar{A}} = A, A \cap \phi = \phi, A \cup \phi = A, A \cap \Omega = A, A \cup \Omega = \Omega$
 $A \cup \bar{A} = \Omega, \bar{\phi} = \Omega, A \cap \bar{A} = \phi$
- (5) א. $A \cap B$: גובה בין 1.7 ל-1.8. ב. $A \cup B$: כל גובה אפשרי.
 ג. $\bar{A} = \bar{A} \cap B$: גובה לכל היותר 1.7. ד. $\bar{A} \cup \bar{B}$: לכל היותר 1.7 או לפחות 1.8.
 ה. $A = \bar{\bar{A}}$: גובה מעל 1.7.
- (6) א. $A \cap B \cap C$ ב. $A \cap \bar{B} \cap \bar{C}$ ג. $A \cup B \cup C$ ד. $\bar{C}$ ה. $(A \cap B \cap \bar{C}) \cup (B \cap C \cap \bar{A}) \cup (A \cap C \cap \bar{B})$
- (7) א. $P(A \cup B) = 0.24$ ב. $P(A \cap B) = 0.04$ ג. $P(B \cap \bar{A}) = 0.16$
- (8) א. $P(A \cap B) = 10\%$ ב. $P(A \cup B) = 50\%$ ג. $P(\bar{A} \cap \bar{B}) = 50\%$
- (9) א. $P(A \cap B) = 0.2$ ב. $P(\bar{A} \cap \bar{B}) = 0.3$ ג. $P(A \cup \bar{B}) = 0.3$
- (10) א. לא. ב. כן. ג. כן. ד. לא.
- (11) א. כן. ב. $P(\bar{A} \cap B) = 0.3$
- (12) הטענה הנכונה היא ג'.
- (13) א. 0.05. ב. 0.95.
- (14) א. $P(A \cap B) = 0.06$ ב. לא. ג. $P((A \cap \bar{B}) \cup (B \cap \bar{A})) = 0.43$
- (15) $P(B) = \frac{1}{5}, P(A) = \frac{2}{5}$
- (16) א. נכון. ב. לא נכון. ג. לא נכון. ד. נכון.

סטטיסטיקה והסתברות

פרק 3 - קומבינטוריקה - כלל המכפלה

תוכן העניינים

1. כללי 13

קומבינטוריקה – כלל המכפלה:

רקע:

כלל המכפלה:

כלל המכפלה הוא כלל שבאמצעותו אפשר לחשב את גודל המאורע או גודל מרחב המדגם.

אם לתהליך יש k שלבים: n_1 אפשרויות לשלב הראשון, n_2 אפשרויות לשלב השני... n_k אפשרויות לשלב k :

מספר האפשרויות לתהליך כולו יהיה: $n_1 \cdot n_2 \cdot n_3 \cdots n_k$

למשל, כמה אפשרויות יש למשחק בו מטילים קובייה וגם מטבע? (הסבר בהקלטה)

$$n_1 = 6, n_2 = 2$$

$$n_1 \cdot n_2 = 6 \cdot 2 = 12$$

למשל, כמה לוחיות רישוי בני 5 תווים ניתן ליצור כאשר התו הראשון הוא אות אנגלית והיתר ספרות? (הסבר בהקלטה)

$$n_1 = 26, n_2 = 10, n_3 = 10, n_4 = 10, n_5 = 10$$

$$n_1 \cdot n_2 \cdot n_3 \cdot n_4 \cdot n_5 = 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 260,000$$

שאלות:

- (1) חשבו את מספר האפשרויות לתהליכים הבאים :
- הטלת קובייה פעמים.
 - מספר תלת ספרתי.
 - בחירת בן ובת מכתה שיש בה שבעה בנים ועשר בנות.
 - חלוקת שני פרסים שונים לעשרה אנשים שונים כאשר אדם לא יכול לקבל יותר מפרס אחד.
- (2) במסעדה מציעים ארוחה עסקית.
- בארוחה עסקית יש לבחור מנה ראשונה, מנה עיקרית ושתייה. האופציות למנה ראשונה הן : סלט ירקות, סלט אנטיפסטי ומרק היום. האופציות למנה עיקרית הן : סטייק אנטריקוט, חזה עוף בגריל, לזניה בשרית ולזניה צמחונית. האופציות לשתייה הן : קפה, תה ולימונדה.
- כמה ארוחות שונות ניתן להרכיב בעזרת התפריט הזה?
 - אדם מזמין ארוחה אקראית. חשב את ההסתברויות הבאות :
 - בארוחה סלט ירקות, לזניה בשרית ולימונדה.
 - בארוחה סלט, לזניה ותה.
- (3) בוחרים באקראי מספר בין חמש ספרות. חשבו את ההסתברויות הבאות :
- המספר הוא זוגי.
 - במספר כל הספרות שונות.
 - במספר כל הספרות זהות.
 - במספר לפחות שתי ספרות שונות.
 - במספר לפחות שתי ספרות זהות.
 - המספר הוא פלינדרום (מספר הנקרא מימין ומשמאל באות הצורה).
- (4) חישה אנשים אקראיים נכנסו למעלית בבניין בן 8 קומות. חשבו את ההסתברויות הבאות :
- כולם ירו בקומה החמישית.
 - כולם ירדו באותה קומה.
 - כולם ירדו בקומה אחרת.
 - ערן ודני ירדו בקומה השישית והיתר בשאר הקומות.

- (5) במפלגה חמישה עשר חברי כנסת. יש לבחור שלושה חברי כנסת לשלושה תפקידים שונים. בכמה דרכים ניתן לחלק את התפקידים הבאים אם:
- א. חבר כנסת יכול למלא יותר מתפקיד אחד.
 - ב. חבר כנסת לא יכול למלא יותר מתפקיד אחד.
- (6) מטילים קובייה 4 פעמים.
- א. מה ההסתברות שכל התוצאות תהינה זהות?
 - ב. מה ההסתברות שכל התוצאות תהינה שונות?
 - ג. מה ההסתברות שלפחות שתי תוצאות תהינה זהות?
 - ד. מה ההסתברות שלפחות שתי תוצאות תהינה שונות?
- (7) יש ליצור מילה בת חמש אותיות, לא בהכרח עם משמעות מאותיות ה-ABC (26 אותיות).
- א. מה ההסתברות שבמילה שנוצרה אין האותיות A, D ו-L?
 - ב. מה ההסתברות שבמילה שנוצרה כל האותיות זהות?
 - ג. מה ההסתברות שבמילה שנוצרה לפחות שתי אותיות שונות זו מזו?
 - ד. מה ההסתברות שהמילה היא פלינדרום? (מילה אשר משמאל לימין, ומימין לשמאל נקראת אותו הדבר).

תשובות סופיות:

- (1) א. 36 ב. 900 ג. 70 ד. 90
- (2) א. 36 ב. i. $\frac{1}{36}$ ב. ii. $\frac{1}{9}$
- (3) א. 0.5 ב. 0.3024 ג. 0.0001 ד. 0.9999 ה. 0.6976 ו. 0.01
- (4) א. $\frac{1}{8^5}$ ב. $\frac{1}{8^4}$ ג. 0.205 ד. $\frac{1 \cdot 1 \cdot 7^3}{8^5}$
- (5) א. 3375 ב. 2730
- (6) א. $\frac{1}{216}$ ב. $\frac{5}{18}$ ג. $\frac{13}{18}$ ד. $\frac{215}{216}$
- (7) א. $\frac{23^5}{26^5}$ ב. $\frac{1}{26^4}$ ג. $1 - \frac{1}{26^4}$ ד. $\frac{1}{26^2}$

סטטיסטיקה והסתברות

פרק 4 - קומבינטוריקה- תמורה - סידור עצמים בשורה

תוכן העניינים

1. כללי..... 17

קומבינטוריקה – תמורה – סידור עצמים בשורה:

רקע:

תמורה:

מספר האפשרויות לסדר n עצמים שונים בשורה: $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$.

הערה: $0! = 1$.

דוגמאות (הפתרונות בהקלטה):

- בכמה דרכים שונות ניתן לסדר את האותיות: a, b, c, d ?
- בכמה דרכים שונות ניתן לסדר את האותיות: a, b, c, d , כך שהאותיות: a, b יהיו ברצף?
- בכמה דרכים שונות ניתן לסדר את האותיות: a, b, c, d , כך שהאותיות: a, b יופיעו בתור הרצף ba ?

שאלות:

- (1) חשבו : בכמה אופנים
א. אפשר לסדר 4 ספרים שונים על מדף?
ב. אפשר לסדר חמישה חיילים בטור?
- (2) סידרו באקראי 10 דיסקים שונים על מדף שמתוכם שניים בשפה העברית.
א. מה ההסתברות שהדיסקים בעברית יהיו צמודים זה לזה?
ב. מה ההסתברות שהדיסקים בעברית לא יהיו צמודים זה לזה?
ג. מה ההסתברות ששני הדיסקים בעברית יהיו כל אחד בקצה השני של המדף?
- (3) בוחנים 5 בנים ו-4 בנות בכיתה ומדרגים אותם לפי הציון שלהם בבחינה. נניח שאין תלמידים בעלי אותו ציון.
א. מהו מספר הדירוגים האפשריים?
ב. מהו מספר הדירוגים האפשריים אם מדרגים בנים ובנות בנפרד?
- (4) מסדרים 10 ספרים שונים על מדף.
א. בכמה אופנים ניתן לסדר את הספרים על המדף?
שני ספרים מתוך ה-10 הם ספרים בסטטיסטיקה.
ב. מה ההסתברות שאם נסדר את הספרים באקראי, הספרים בסטטיסטיקה יהיו צמודים זה לזה?
ג. מה ההסתברות שהספרים בסטטיסטיקה לא יהיו צמודים זה לזה?
ד. מה ההסתברות שהספרים בסטטיסטיקה יהיו בקצות המדף (כל ספר בקצה אחר)?
- (5) אדם יצר בנגן שלו פלייליסט (רשימת השמעה) של 12 שירים שונים. 4 בשפה העברית, 5 באנגלית ו-3 בצרפתית. האדם הריץ את הפלייליסט באקראי.
א. מה ההסתברות שכל השירים באנגלית יופיעו כשירים הראשונים כמקשה אחת?
ב. מה ההסתברות שכל השירים באנגלית יופיעו ברצף (לא חובה ראשונים)?
ג. מה ההסתברות ששירים באותה השפה יופיעו ברצף (כלומר כל השירים באנגלית ברצף, כל השירים בעברית ברצף וכך גם השירים בצרפתית)?

- 6) 4 בנים ו-4 בנות התיישבו באקראי בשורת כיסאות 1-8 בקולנוע.
- א. מה ההסתברות שיוסי ומיכל לא ישבו זה לצד זה?
- ב. מה ההסתברות שהבנים יתיישבו במקומות האי-זוגיים?
- ג. מה ההסתברות שכל הבנים ישבו זה לצד זה?
- ד. מה ההסתברות שהבנים ישבו זה לצד זה והבנות תשבנה זו לצד זו?

תשובות סופיות:

- 1) א. 0.24 ב. 0.120
- 2) א. 0.2 ב. 0.8 ג. 0.022
- 3) א. 0.362880 ב. 0.2880
- 4) א. 0.3628800 ב. 0.2 ג. 0.8 ד. $\frac{1}{45}$
- 5) א. $\frac{1}{792}$ ב. $\frac{1}{99}$ ג. $\frac{1}{4620}$
- 6) א. 0.75 ב. 0.014 ג. $\frac{1}{14}$ ד. $\frac{1}{35}$

סטטיסטיקה והסתברות

פרק 5 - קומבינטוריקה - דגימה סידורית ללא החזרה ועם החזרה

תוכן העניינים

1. כללי 20

קומבינטוריקה – דגימה סידורית ללא החזרה ועם החזרה:

רקע:

מדגם סידור בדגימה עם החזרה:

מספר האפשרויות בדגימת k עצמים מתוך n עצמים שונים כאשר הדגימה היא עם החזרה והמדגם סדור הוא: n^k .

דוגמה:

בחרים שלושה תלמידים מתוך עשרה לייצג ועד בו תפקידים שונים, תלמיד יכול למלא יותר מתפקיד אחד.

כמה ועדים שונים ניתן להרכיב? $10^3 = 1,000$, $k = 3$, $n = 10$.

מדגם סידור ללא החזרה:

מספר האפשרויות בדגימת k עצמים שונים מתוך n עצמים שונים ($n \geq k$) כאשר המדגם סדור ואין החזרה של עצמים נדגמים הינו:

$$(n)_k = n(n-1)(n-2)\dots(n-(k-1)) = \frac{n!}{(n-k)!}$$

דוגמה:

שלושה תלמידים נבחרים מתוך 10 לייצג וועד בו תפקידים שונים.

תלמיד לא יכול למלא יותר מתפקיד אחד: $\frac{10!}{7!} = 10 \cdot 9 \cdot 8 = 720$.

שאלות:

- (1) במפלגה 20 חברי כנסת, מעוניינים לבחור שלושה חברי כנסת לשלושה תפקידים שונים.
- א. חבר כנסת יכול למלא יותר מתפקיד אחד.
כמה קומבינציות ישנן לחלוקת התפקידים?
- ב. חבר כנסת לא יכול למלא יותר מתפקיד אחד.
כמה קומבינציות יש לחלוקת התפקידים?
- (2) במשחק מזל יש 4 משבצות ממוספרות מ-A (עד D). בכל משבצת יש למלא סיפרה (0-9). הזוכה הוא זה שניחש נכונה את כל הספרות בכל המשבצות בהתאמה.
- א. מה ההסתברות לזכות במשחק?
- ב. מה ההסתברות שבאף משבצת לא תהיה את הספרה 3 במספר הזוכה?
- ג. מה ההסתברות שהתוצאה 4 תופיע לפחות פעם אחת במספר הזוכה?
- (3) קבוצה מונה 22 אנשים, מה ההסתברות שלפחות לשניים מהם יהיה יום הולדת באותו התאריך?
- (4) שלושה אנשים קבעו להיפגש במלון הילטון בסינגפור. הבעיה היא שבסינגפור ישנם 5 מלונות הילטון.
- א. מה ההסתברות שכל השלושה ייפגשו?
- ב. מה ההסתברות שכל אחד יגיע לבית מלון אחר?
- (5) בכיתה 40 תלמידים. מעוניינים לבחור חמישה מהם לוועד כיתה. בכמה דרכים ניתן להרכיב את הוועד אם:
- א. בוועד 5 תפקידים שונים ותלמיד יכול למלא יותר מתפקיד אחד.
- ב. בוועד 5 תפקידים שונים ותלמיד לא יכול למלא יותר מתפקיד אחד.

תשובות סופיות:

- (1) א. 8000 ב. 6840
- (2) א. 0.0001 ב. 0.6561 ג. 0.3439
- (3) 0.476
- (4) א. 0.04 ב. 0.48
- (5) א. 40^5 ב. 78,960,960

סטטיסטיקה והסתברות

פרק 6 - קומבינטוריקה - דגימה ללא סדר וללא החזרה

תוכן העניינים

1. כללי 22

קומבינטוריקה – דגימה ללא סדר וללא החזרה:

רקע:

מדגם לא סדור בדגימה ללא החזרה:

מספר האפשרויות לדגום k עצמים שונים מתוך n עצמים שונים כאשר אין

משמעות לסדר העצמים הנדגמים ואין החזרה: $\frac{n!}{(n-k)!k!} = \binom{n}{k} = \frac{(n)_k}{k!}$

דוגמה:

מתוך 10 תלמידים יש לבחור שלושה נציגים לוועד ללא תפקידים מוגדרים:

$$\frac{10!}{7!3!} = \binom{10}{3} = 120$$

הערות:

$$\binom{n}{k} = \binom{n}{n-k} \quad (1)$$

$$\binom{n}{n-1} = \binom{n}{1} = n \quad (2)$$

$$\binom{n}{n} = \binom{n}{0} = 1 \quad (3)$$

שאלות:

- (1) בכיתה 15 בנות ו-10 בנים. יש לבחור 5 תלמידים שונים מהכיתה לנציגות הכיתה. בכמה דרכים אפשר להרכיב את הנציגות, אם:
- א. אין שום הגבלה לבחירה.
 - ב. מעוניינים ש-3 בנות ו-2 בנים ירכיבו את המשלחת.
 - ג. לא יהיו בנים במשלחת.
- (2) סטודנט מעוניין לבחור 5 קורסי בחירה בסמסטר זה. לפניו רשימה של 10 קורסים לבחירה: 5 במדעי הרוח, 3 במדעי החברה, 2 במתמטיקה.
- א. כמה בחירות שונות הוא יכול ליצור לעצמו?
 - ב. כמה בחירות יש לו בהן 3 קורסים הם ממדעי הרוח?
 - ג. כמה בחירות יש לו אם 2 מהן לא ממדעי הרוח?
 - ד. כמה בחירות יש לו אם 2 ממדעי הרוח, 2 ממדעי החברה ו-1 ממתמטיקה?
- (3) בכיתה 30 תלמידים מתוכם 12 תלמידים ו-18 תלמידות.
- יש לבחור למשלחת 4 תלמידים מהכיתה. התלמידים נבחרים באקראי.
- א. מה ההסתברות שהמשלחת תורכב רק מבנות?
 - ב. מה ההסתברות שבמשלחת תהיה רק בת אחת?
 - ג. מה ההסתברות שבמשלחת תהיה לפחות בת אחת?
- (4) במשחק הלוטו יש לבחור 5 מספרים מתוך 45. המספרים הם 1-45.
- א. מה ההסתברות שבמשחק הזוכה כל המספרים הם זוגיים?
 - ב. מה ההסתברות שבמספר הזוכה יש לכל היותר מספר זוגי אחד?
 - ג. מה ההסתברות שבמספר הזוכה לפחות פעם אחת יש מספר זוגי?
 - ד. מה ההסתברות שבמספר הזוכה כל המספרים גדולים מ-30?
- (5) בחפיסת קלפים ישנם 52 קלפים: 13 בצבע שחור בצורת עלה, 13 בצבע אדום בצורת לב, 13 בצבע אדום בצורת יהלום ו-13 בצבע שחור בצורת תלתן. מכל צורה (מתוך ה-4) יש 9 קלפים שמספרם 2-10, שאר הקלפים הם: נסיך, מלכה, מלך ואס (בעצם מדובר בקופסת קלפים רגילה ללא ג'וקר). שני אנשים משחקים פוקר. כל אחד מקבל באקראי 5 קלפים (ללא החזרה).
- א. מה ההסתברות שעודד יקבל את כל המלכים וערן את כל המלכות?
 - ב. מה ההסתברות שאחד השחקנים יקבל את הקלף אס-לב?
 - ג. מה ההסתברות שערן יקבל קלפים שחורים בלבד ועודד יקבל שני קלפים שחורים בדיוק?
 - ד. מה ההסתברות שערן יקבל לפחות 3 קלפים שהם מספר (אס אינו מספר)?

- (6) במכללה 4 מסלולי לימוד. בכל מסלול לימוד 5 מזכירות. יש ליצור וועד של 5 מזכירות מתוך כלל המזכירות במכללה. יוצרים וועד באופן אקראי. חשבו את ההסתברויות הבאות:
- כל המזכירות בוועד יהיו ממסלול "מדעי ההתנהגות".
 - כל המזכירות בוועד יהיו מאותו המסלול.
 - מכל מסלול תבחר לפחות מזכירה אחת.

תשובות סופיות:

(1)	א. 53130	ב. 20475	ג. 3003
(2)	א. 252	ב. 100	ג. 100
(3)	א. 0.1117	ב. 0.1445	ג. 0.9819
(4)	א. 0.02	ב. 0.187	ג. 0.972
(5)	א. 0	ב. 0.1923	ג. 0.009
(6)	א. $6.45 \cdot 10^{-5}$	ב. $2.58 \cdot 10^{-4}$	ג. 0.3225

סטטיסטיקה והסתברות

פרק 7 - קומבינטוריקה - שאלות מסכמות

תוכן העניינים

1. כללי 25

קומבינטוריקה – שאלות מסכמות:

שאלות:

- (1) בכיתה 40 תלמידים. מעוניינים לבחור חמישה מהם לוועד כיתה. בכמה דרכים ניתן להרכיב את הוועד אם:
 - א. בוועד 5 תפקידים שונים ותלמיד יכול למלא יותר מתפקיד אחד.
 - ב. בוועד 5 תפקידים שונים ותלמיד לא יכול למלא יותר מתפקיד אחד.
 - ג. אין תפקידים שונים בוועד.
- (2) במשרד 30 עובדים, יש לבחור ארבעה עובדים למשלחת לחו"ל. בכמה דרכים ניתן להרכיב את המשלחת?
 - א. במשלחת ארבע משימות שונות שיש למלא וכל עובד יכול למלא יותר ממשימה אחת.
 - ב. כמו בסעיף א' רק הפעם עובד לא יכול למלא יותר ממשימה אחת.
 - ג. מעוניינים לבחור ארבעה עובדים שונים למשלחת שבה לכולם אותו התפקיד.
- (3) מעוניינים להרכיב קוד סודי. הקוד מורכב מ-2 ספרות שונות ו-3 אותיות שונות באנגלית (26 אותיות אפשריות).
 - א. כמה קודים שונים ניתן להרכיב?
 - ב. כמה קודים שונים ניתן להרכיב אם הקוד מתחיל בספרה ונגמר בספרה?
 - ג. כמה קודים ניתן להרכיב אם הספרות חייבות להיות צמודות זו לזו?
 - ד. בכמה קודים הספרות לא מופיעות ברצף?
- (4) בארונות 4 מגירות. ילד התבקש על ידי אמו לסדר 6 משחקים בארונות. הילד מכניס את המשחקים באקראי למגירות השונות. כל מגירה יכולה להכיל את כל המשחקים יחד.
 - א. מה ההסתברות שהילד יכניס את כל המשחקים למגירה העליונה?
 - ב. מה ההסתברות שהילד יכניס את כל המשחקים לאותה מגירה?
 - ג. מה ההסתברות ש"דומינו" יוכנס למגירה העליונה ויתר המשחקים לשאר המגירות.
 - ד. מה ההסתברות ש"דומינו" לא יוכנס למגירה העליונה?

- (5) בעיר מסוימת מתמודדות למועצת העיר 4 מפלגות שונות: "הירוקים", "קדימה", "העבודה" ו"הליכוד". 6 אנשים אינם יודעים למי להצביע, ולכן בוחרים באקראי מפלגה כלשהי.
- מה ההסתברות שכל ה-6 יבחרו באותה מפלגה?
 - מה ההסתברות שמפלגת ה"ירוקים" לא תקבל קולות?
 - מה ההסתברות שמפלגת ה"ירוקים" תקבל בדיוק 3 קולות וכל מפלגה אחרת תקבל קול 1 בלבד?
 - מה ההסתברות שמפלגת "הירוקים" תקבל 2 קולות, מפלגת "העבודה" תקבל 2 קולות ומפלגת "הליכוד" תקבל 2 קולות?
- (6) 5 חברים נפגשו ורצו לראות סרט. לרשותם ספריה המונה 8 סרטים שונים. כל אחד התבקש לבחור סרט באקראי.
- מה ההסתברות שכולם יבחרו את אותו הסרט?
 - מה ההסתברות שכולם יבחרו את "הנוסע השמיני"?
 - מה ההסתברות שכל אחד יבחר סרט אחר?
 - מה הסיכוי שלפחות שניים יבחרו את אותו הסרט?
 - מה ההסתברות שיוסי וערן יבחרו את "הנוסע השמיני" וכל השאר סרטים אחרים?
 - מה ההסתברות שהנוסע השמיני לא יבחר על ידי אף אחד מהחברים?
 - לקחו את 8 הסרטים ויצרו מהם רשימה. נתון שברשימה 3 סרטי אימה, מה ההסתברות שברשימה שנוצרה יופיעו 3 סרטי האימה ברצף?
- (7) בקבוצה 10 אנשים. יש ליצור שתי וועדות שונות מתוך הקבוצה: אחת בת 4 אנשים והשנייה בת 3 אנשים. כל אדם יכול להיבחר רק לוועדה אחת. חשבו את מס' הדרכים השונות ליצירת הוועדות הללו כאשר:
- אין בוועדות תפקידים.
 - בכל וועדה יש תפקיד אחד של אחראי הוועדה.
 - בכל וועדה כל התפקידים שונים.
- (8) 4 גברים ו-3 נשים מתיישבים על כסאות בשורה של כסאות תיאטרון. בכל שורה 10 כסאות. בכמה דרכים שונות ניתן לבצע את ההושבה:
- ללא הגבלה.
 - כל הגברים ישבו זה ליד זה וגם כל הנשים תשבנה זו ליד זו.
 - שני גברים בקצה אחד ושני הגברים האחרים בקצה שני.
- (9) בהגרלה ישנם 10 מספרים מ-1 עד 10. נבחרו באקראי 5 מספרים. מה ההסתברות שהמספר 7 הוא השני בגודלו מבין המספרים שנבחרו?

- 10) 6 אנשים עלו לאוטובוס שעוצר ב-10 תחנות.**
כל אדם בוחר באופן עצמאי ואקראי באיזו תחנה לרדת.
א. מה ההסתברות שכל אחד יורד בתחנה אחרת?
ב. מה ההסתברות שבדיוק 3 ירדו בתחנה החמישית?
ג. מה ההסתברות שרונית תרד בתחנה השנייה והשאר לא?
ד. מה ההסתברות שכולם ירדו בתחנות 5,6 ולפחות אחד בכל אחת מהתחנות הללו?

- 11)** ברכבת 4 מקומות ישיבה עם כיוון הנסיעה 4 מקומות ישיבה נגד כיוון הנסיעה.

חלון	מעבר		

- 4 זוגות התיישבו במקומות אלו באקראי.

א. בכמה דרכים שונות ניתן להתיישב?

ב. מה ההסתברות שהזוג כהן ישבו זה לצד זה עם כיוון הנסיעה?

ג. מה ההסתברות שהזוג כהן ישבו זה לצד זה?

ד. מה ההסתברות שהזוג כהן ישבו כל אחד ליד החלון? (בכל שורה יש חלון).

ה. מה ההסתברות שהזוג כהן יישבו כך שכל אחד בכיוון נסיעה מנוגד?

ו. מה ההסתברות שהזוג כהן יישבו אחד מול השני פנים מול פנים.

ז. מה ההסתברות שכל הגברים ייסעו עם כיוון הנסיעה וכל הנשים תשבנה נגד כיוון הנסיעה?

ח. מה ההסתברות שכל זוג ישב אחד מול השני?

- 12) סיסמא מורכבת מ-5 תווים, תווים אלו יכולים להיות ספרה (0-9) ואותיות ה-ABC (26 אותיות). כל תו יכול לחזור על עצמו יותר מפעם אחת.**
- א. כמה סיסמאות שונות יש?
- ב. כמה סיסמאות שונות יש שבהן כל התווים שונים?
- ג. כמה סיסמאות שונות יש שבהן לפחות ספרה אחת ולפחות אות אחת?

- 13)** מתוך קבוצה בת n אנשים רוצים לבחור 3 אנשים לוועדה. בכמה דרכים שונות ניתן לבצע את הבחירה? בטא את תשובתך באמצעות n .
- א. בוועדה אין תפקידים ויש לבחור 3 אנשים שונים לוועדה.
ב. בוועדה תפקידים שונים. וכל אדם לא יכול למלא יותר מתפקיד אחת.
ג. בוועדה תפקידים שונים ואדם יכול למלא יותר מתפקיד אחד.

- 14)** שני אנשים מטילים כל אחד מטבע n פעמים. בטאו באמצעות n את הסיכוי שלכל אחד מהם אותו מספר פעמים של התוצאה "ראש".

- 15)** יוצרים קוד עם a ספרות (מותר לחזור על אותה ספרה בקוד).
חשבו את ההסתברויות הבאות (בטאו את תשובותיכם באמצעות a):
- א. בקוד אין את הספרה 5.
 - ב. בקוד מופיעה הספרה 3.
 - ג. בקוד לא מופיעות ספרות אי זוגיות.

תשובות סופיות:

- (1) א. 102,400,000 ב. 78,960,960 ג. 658008
- (2) א. 810,000 ב. 657,720 ג. 27,405
- (3) א. 14,040,000 ב. 1,404,000 ג. 5,616,000 ד. 8,424,000
- (4) א. 0.00024 ב. 0.00098 ג. 0.05933 ד. 0.75000
- (5) א. 0.00098 ב. 0.17798 ג. 0.02929 ד. 0.02197
- (6) א. $\frac{1}{4096}$ ב. $\frac{1}{32,768}$ ג. 0.205 ד. 0.795
- ה. 0.0105 ו. 0.5129 ז. 0.1071
- (7) א. 4,200 ב. 50,400 ג. 604,800
- (8) א. 604,800 ב. 2,880 ג. 2,880
- (9) 0.238
- (10) א. 0.1512 ב. 0.014 ג. 0.059 ד. $\frac{62}{10^6}$
- (11) א. 40,320 ב. 0.1071 ג. 0.2142 ד. 0.0357 ה. 0.5714 ו. 0.1429 ז. 0.0143 ח. 0.0095
- (12) א. 60,466,176 ב. 45,239,040 ג. 48,484,800
- (13) א. $\frac{n!}{3!(n-3)}$ ב. $n \cdot (n-1)(n-2)$ ג. n^3
- (14) $\frac{1}{4^n} \cdot \sum_{i=0}^n \binom{n}{i}^2$
- (15) א. 0.9^a ב. $1-0.9^a$ ג. 0.5^a

סטטיסטיקה והסתברות

פרק 8 - הסתברות מותנית-במרחב מדגם אחיד

תוכן העניינים

1. כללי 30

הסתברות מותנית – במרחב מדגם אחיד:

רקע:

לעיתים אנו נדרשים לחשב הסתברות למאורע כלשהו כאשר ברשותנו אינפורמציה לגבי מאורע אחר. הסתברות מותנית הינה סיכוי להתרחשות מאורע כלשהו כאשר ידוע שמאורע אחר התרחש / לא התרחש.

ההסתברות של A בהינתן ש- B כבר קרה: $P(A|B)$.

כשמרחב המדגם אחיד: $P(A|B) = \frac{|A \cap B|}{|B|}$

דוגמה (פתרון בהקלטה):

נטיל קובייה.

נגדיר:

A - התוצאה זוגית.

B - התוצאה גדולה מ-3.

נרצה לחשב את: $P(A|B)$.

שאלות:

- (1) נבחרה ספרה זוגית באקראי. מה הסיכוי שהספרה גדולה מ-6?
- (2) יוסי הטיל קובייה.
מה הסיכוי שקיבל את התוצאה 4, אם ידוע שהתוצאה שהתקבלה זוגית?
- (3) הוטלו צמד קוביות. נגדיר:
A - סכום התוצאות בשתי ההטלות הינו 7.
B - מכפלת התוצאות 12.
חשבו את $P(A|B)$.
- (4) מטבע הוטל פעמיים.
ידוע שהתקבל לכל היותר ראש אחד, מה הסיכוי שהתקבלו שני ראשים?
- (5) זוג קוביות הוטלו והתקבל שהתוצאות זהות.
מה הסיכוי שלפחות אחת התוצאות 5?
- (6) זוג קוביות הוטלו והתקבל לפחות פעם אחת 4.
מה הסיכוי שאחת התוצאות 5?
- (7) נבחרה משפחה בת שני ילדים, שמהם אחד הוא בן.
מה ההסתברות שבמשפחה שני בנים בקרב הילדים?
- (8) נבחרה משפחה בת שלושה ילדים, ונתון שהילד האמצעי בן.
מה הסיכוי שיש בנות בקרב הילדים?
- (9) בכיתה 6 בנים ו-7 בנות. נבחרו 4 ילדים מהכיתה.
אם ידוע שנבחרו 2 בנים ו-2 בנות, מה הסיכוי שאלעד לא נבחר?
- (10) חמישה חברים יצאו לבית קולנוע והתיישבו זה לצד זה באקראי,
בכיסאות מספר 5 עד 9. ידוע שערך ודין התיישבו זה ליד זה.
מה ההסתברות שהם יושבים בכיסאות מספר 6 ו-7?

תשובות סופיות:

(1) 0.2

(2) $\frac{1}{3}$

(3) 0.5

(4) 0

(5) $\frac{1}{6}$

(6) $\frac{2}{11}$

(7) $\frac{1}{3}$

(8) $\frac{3}{4}$

(9) $\frac{2}{3}$

(10) $\frac{1}{4}$

סטטיסטיקה והסתברות

פרק 9 - הסתברות מותנית - מרחב לא אחיד

תוכן העניינים

1. כללי 33

הסתברות מותנית – מרחב לא אחיד:

רקע:

הסיכוי שמאורע A יתרחש, בהינתן שמאורע B כבר קרה: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

במונה: הסיכוי לחיתוך של שני המאורעות, זה הנשאל וזה הנתון שהתרחש.
במכנה: הסיכוי למאורע נתון שהתרחש.

דוגמה (פתרון בהקלטה):

נבחרו משפחות שיש להם שתי מכוניות. ל-30% מהמשפחות הללו המכונית הישנה יותר היא מתוצרת אירופה ואצל 60% מהמשפחות הללו המכונית החדשה יותר מתוצרת אירופה. כמו כן, בקרב 15% מהמשפחות הללו שתי המכוניות הן מתוצרת אירופאית. אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית?

שאלות:

- (1) תלמיד ניגש בסמסטר לשני מבחנים: מבחן בכלכלה ומבחן בסטטיסטיקה. נגדיר את המאורעות הבאים:
- A - לעבור את המבחן בסטטיסטיקה.
 - B - לעבור את המבחן בכלכלה.
- 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. חשבו את הסיכויים למאורעות הבאים:
- א. התלמיד עבר בסטטיסטיקה, מה ההסתברות שהוא עבר בכלכלה?
 - ב. התלמיד עבר בכלכלה, מה ההסתברות שהוא עבר בסטטיסטיקה?
 - ג. התלמיד עבר בכלכלה, מה ההסתברות שהוא נכשל בסטטיסטיקה?
 - ד. התלמיד נכשל בסטטיסטיקה, מה ההסתברות שהוא נכשל בכלכלה?
 - ה. התלמיד עבר לפחות מבחן אחד, מה ההסתברות שהוא יעבור את שניהם?
- (2) במדינה שתי חברות טלפון סלולארי: "סופט" ו"בל". 30% מהתושבים הבוגרים רשומים אצל חברת "בל", 60% מהתושבים הבוגרים רשומים אצל חברת "סופט" ול-15% מהתושבים הבוגרים אין טלפון סלולארי כלל.
- א. איזה אחוז מהתושבים הבוגרים רשומים אצל שתי החברות?
 - ב. נבחר אדם שרשום אצל חברת "סופט", מה ההסתברות שהוא רשום גם אצל חברת "בל"?
 - ג. אם אדם לא רשום אצל חברת "בל", מה ההסתברות שהוא כן רשום בחברת "סופט"?
 - ד. אם אדם רשום אצל חברה אחת בלבד, מה ההסתברות שהוא רשום בחברת "סופט"?
- (3) במכללה שני חניונים: חניון קטן וחניון גדול. בשעה 08:00 יש סיכוי של 60% שבחניון הגדול יש מקום, סיכוי של 30% שבחניון הקטן יש מקום וסיכוי של 20% שבשני החניונים יש מקום.
- א. מה ההסתברות שיש מקום בשעה 08:00 רק בחניון הגדול של המכללה?
 - ב. ידוע שבחניון הקטן יש מקום בשעה 08:00, מה הסיכוי שבחניון הגדול יש מקום?
 - ג. אם בשעה 08:00 בחניון הגדול אין מקום, מה ההסתברות שבחניון הקטן יהיה מקום?
 - ד. נתון שלפחות באחד מהחניונים יש מקום בשעה 08:00, מה ההסתברות שבחניון הגדול יש מקום?

(4) נלקחו 200 שכירים ו-100 עצמאים. מתוך השכירים 20 הם אקדמאיים, ומתוך העצמאיים 30 הם אקדמאיים.

- א. בנו טבלת שכיחות משותפת לנתונים.
- ב. נבחר אדם אקראי מה ההסתברות שהוא שכיר?
- ג. מה ההסתברות שהוא שכיר ולא אקדמאי?
- ד. מה ההסתברות שהוא שכיר או אקדמאי?
- ה. אם האדם שנבחר הוא עצמאי מהי ההסתברות שהוא אקדמאי?
- ו. אם האדם שנבחר הוא לא אקדמאי, מה ההסתברות שהוא שכיר?

(5) חברה מסוימת פרסמה את הנתונים הבאים לגבי האזרחים מעל גיל 21 :
 40% מהאנשים מחזיקים כרטיס "ויזה", 52% מחזיקים כרטיס "ישראלכרט",
 20% מחזיקים כרטיס "אמריקן אקספרס", 15% מחזיקים ויזה וגם ישראלכרט,
 8% מחזיקים ישראלכרט וגם אמריקן אקספרס ו-7% מחזיקים כרטיס ויזה וגם
 אמריקן אקספרס. כמו כן, 5% מחזיקים בשלושת הכרטיסים הנ"ל.
 א. אם לאדם יש ויזה, מה הסיכוי שאין לו ישראלכרט?
 ב. אם לאדם שני כרטיסי אשראי, מה הסיכוי שאין לו ישראלכרט?
 ג. אם לאדם לפחות כרטיס אחד, מה הסיכוי שאין לו ישראלכרט?

תשובות סופיות:

(1) א. 0.833 ב. 0.9375 ג. 0.0625 ד. 0.5 ה. 0.789

(2) א. 5% ב. 0.0833 ג. 0.786 ד. 0.6875

(3) א. 0.4 ב. $\frac{2}{3}$ ג. 0.25 ד. $\frac{6}{7}$

(4) א. להלן טבלה: ב. $\frac{2}{3}$ ג. 0.6 ד. $\frac{23}{30}$

סה"כ	לא אקדמאי	אקדמאי	
200	180	20	שכיר
100	70	30	עצמאי
300	250	50	סה"כ

ה. 0.3 ו. 0.72

(5) א. 0.625 ב. 0.133 ג. 0.402

סטטיסטיקה והסתברות

פרק 10 - דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה

תוכן העניינים

1. כללי 37

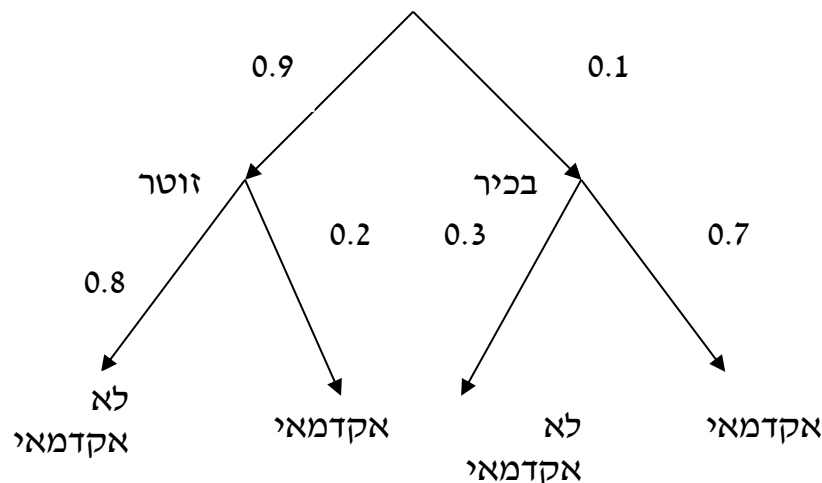
דיאגרמת עצים – נוסחת בייס ונוסחת ההסתברות השלמה:

רקע:

נשתמש בשיטה זו כאשר יש תרגיל שבו התרחשות המאורעות היא בשלבים, כך שכל תוצאה של כל שלב תלויה בשלב הקודם, פרט לשלב הראשון:

דוגמה:

בחברה מסוימת 10% מוגדרים בכירים והיתר מוגדרים זוטרים. מבין הבכירים 20% הם אקדמאים ומבין הזוטרים 80% הם אקדמאים ומבין הזוטרים 20% הם אקדמאים. נשרטט עץ שיתאר את הנתונים, השלב הראשון של העץ אינו מותנה בכלום ואילו השלב השני מותנה בשלב הראשון.



כדי לקבל את הסיכוי לענף מסוים נכפיל את כל ההסתברויות על אותו ענף. נבחר אדם באקראי מאותה חברה.

(1) מה הסיכוי שהוא בכיר אקדמאי ? $0.1 \cdot 0.7 = 0.07$.

(2) מה הסיכוי שהוא זוטר לא אקדמאי ? $0.9 \cdot 0.8 = 0.72$.

כדי לקבל את הסיכוי לכמה ענפים נחבר את הסיכויים של כל ענף (רק אחרי שבתוך הענף הכפלנו את ההסתברויות).

(3) מה הסיכוי שהוא אקדמאי ? $0.1 \cdot 0.7 + 0.9 \cdot 0.2 = 0.25$.

(4) נבחר אקדמאי מה ההסתברות שהוא עובד זוטר? מדובר כאן על שאלה בהסתברות מותנה ולכן נשתמש בעיקרון של הסתברות

$$P(zutar | academay) = \frac{0.9 \cdot 0.2}{0.25} = \frac{0.18}{0.25} = 0.72 \text{ מותנה :}$$

נוסחת ההסתברות השלמה:

בהינתן B , מאורע כלשהו, וחלוקה של מרחב המדגם Ω ל- $A_1, \dots, A_n$ כך ש- $\bigcup_i A_i = \Omega$,

$$. P(B) = \sum_{i=1}^n P(A_i) \cdot P\left(\frac{B}{A_i}\right) \quad \text{אזי:}$$

נוסחת בייס:

$$. P\left(\frac{A_j}{B}\right) = \frac{P(A_j) P\left(\frac{B}{A_j}\right)}{\sum_{i=1}^n P(A_i) \cdot P\left(\frac{B}{A_i}\right)}$$

שאלות:

- (1) בשקית סוכריות 4 סוכריות תות ו-3 לימון. מוציאים באקראי סוכריה. אם היא בטעם תות אוכלים אותה ומוציאים סוכריה נוספת, ואם היא בטעם לימון מחזירים אותה לשקית ומוציאים סוכריה נוספת.

א. מה ההסתברות שהסוכריה הראשונה שהוצאה בטעם תות והשנייה בטעם לימון?

ב. מה ההסתברות שהסוכריה השנייה בטעם לימון?
- (2) באוכלוסייה מסוימת 30% הם ילדים, 50% בוגרים והיתר קשישים. לפי נתוני משרד הבריאות הסיכוי שילד יחלה בשפעת במשך החורף הוא 80%, הסיכוי שמבוגר יחלה בשפעת במשך החורף הוא 40% והסיכוי שקשיש יחלה בשפעת במשך החורף הוא 70%.

א. איזה אחוז מהאוכלוסייה הינו קשישים שלא יחלו בשפעת במשך החורף?

ב. מה אחוז האנשים שיחלו בשפעת במשך החורף?

ג. נבחר אדם שחלה במשך החורף בשפעת, מה ההסתברות שהוא קשיש?

ד. נבחר ילד, מה ההסתברות שהוא לא יחלה בשפעת במשך החורף?
- (3) בכד א' 5 כדורים כחולים ו-5 כדורים אדומים. בכד ב' 6 כדורים כחולים ו-4 כדורים אדומים. בוחרים באקראי כד, מוציאים ממנו כדור ומבלי להחזירו מוציאים כדור נוסף.

א. מה ההסתברות ששני הכדורים שיוצאו יהיו בצבעים שונים?

ב. אם הכדורים שהוצאו הם בצבעים שונים, מה ההסתברות שהכדור השני שהוצא יהיה בצבע אדום?
- (4) חברת סלולר מסווגת את לקוחותיה לפי 3 קבוצות גיל: נוער, בוגרים ופנסיונרים. נתון כי: 10% מהלקוחות בני נוער, 70% מהלקוחות בוגרים והיתר פנסיונרים. מתוך בני הנוער 90% מחזיקים בסמארט-פון, מתוך האוכלוסייה הבוגרת ל-70% יש סמארט-פון ומתוך אוכלוסיית הפנסיונרים 30% מחזיקים בסמארט-פון.

א. איזה אחוז מלקוחות החברה הם בני נוער עם סמארט-פון?

ב. נבחר לקוח אקראי ונתון שיש לו סמארט-פון. מה ההסתברות שהוא פנסיונר?

ג. אם ללקוח אין סמארט-פון, מה ההסתברות שהוא לא בן נוער?

- (5) כדי להתקבל למקום עבודה יש לעבור שלושה מבחנים. המבחנים הם בשלבים, כלומר לאחר כישלון במבחן מסוים אין אפשרות לגשת למבחן הבא אחריו. 70% מהמועמדים עוברים את המבחן הראשון. מתוכם, 50% עוברים את המבחן השני. מבין אלה שעוברים את המבחן השני 40% עוברים את המבחן השלישי.
- א. מה ההסתברות להתקבל לעבודה?
 ב. מועמד לא התקבל לעבודה. מה ההסתברות שהוא נכשל במבחן הראשון?
 ג. מועמד לא התקבל לעבודה. מה ההסתברות שהוא עבר את המבחן השני?
- (6) משרד הבריאות פרסם את הנתונים הבאים:
- מתוך אוכלוסיית הילדים והנוער 80% חולים בשפעת בזמן החורף.
 מתוך אוכלוסיית המבוגרים (עד גיל 65) 60% חולים בשפעת בזמן החורף.
 30% מהתושבים הם ילדים ונוער. 50% הם מבוגרים. היתר קשישים.
 כמו כן נתון ש 68% מהאוכלוסייה תחלה בשפעת בחורף.
- א. מה אחוז החולים בשפעת בקרב האוכלוסייה הקשישה?
 ב. נבחר אדם שלא חלה בשפעת, מה ההסתברות שהוא לא קשיש?
- (7) רדאר שנמצא על החוף צריך לקלוט אנייה הנמצאת ב-1 מ-4 האזורים: A, B, C, D. אם האנייה נמצאת באזור A הרדאר מזהה אותה בסיכוי 0.8, סיכוי זה פוחת ב-0.1 ככל שהאנייה מתקדמת באזור. כמו כן נתון שבהסתברות חצי האנייה נמצאת באזור D, בהסתברות 0.3 באזור C, באזור B היא נמצאת בסיכוי 0.2, אחרת היא נמצאת באזור A.
- א. מה הסיכוי שהאנייה תתגלה ע"י הרדאר?
 ב. אם האנייה התגלתה ע"י הרדאר, מה ההסתברות שהיא נמצאת באזור C?
 ג. אם האנייה התגלתה ע"י הרדאר, מה הסיכוי שהיא לא נמצאת באזור B?
- (8) סימפטום X מופיע בהסתברות של 0.4 במחלה A, בהסתברות של 0.6 במחלה B ובהסתברות של 0.5 במחלה C. סימפטום X מופיע אך ורק במחלות הללו, אדם לא יכול לחלות ביותר ממחלה אחת מבין המחלות הללו. לקליניקה מגיעים אנשים כדלקמן: 8% חולים במחלה A, 10% במחלה B, 2% במחלה C והיתר בריאים. כמו כן נתון שבמחלה A, סימפטום X מתגלה בסיכוי של 80%, ובמחלות B, C הסימפטום מתגלה בסיכוי של 90% בכל מחלה.
- א. מה ההסתברות שאדם הגיע לקליניקה וגילו אצלו את סימפטום X?
 ב. אם התגלה אצל אדם סימפטום X, מה ההסתברות שהוא חולה במחלה A?
 ג. אם לאדם יש את סימפטום X, מה ההסתברות שהוא חולה במחלה A?
 ד. אם לא גילו אצל אדם את סימפטום X, מה ההסתברות שהוא בריא?

- (9) סטודנט ניגש למבחן אמריקאי. הסיכוי שהוא יודע תשובה לשאלה מסוימת הוא P , ואם הוא לא יודע את התשובה הוא מנחש. בכל מקרה הוא עונה על השאלה. נתון שלשאלה יש k תשובות אפשריות. אם הסטודנט ענה נכון על השאלה, מה הסיכוי שהוא ידע אותה?
- (10) אדם משחק נגד שני מתמודדים, רונית ודולב. האדם צריך לשחק שלושה משחקים ויש לו לבחור איזה סדר משחקים עדיף לו:
- דולב, רונית, דולב.
 - רונית, דולב, רונית.
- בכל משחק מישו חייב לנצח(אין תיקו). האדם ינצח בטורניר רק אם ינצח בשני משחקים ברציפות. נתון שדולב שחקן טוב יותר מאשר רונית. איזו אפשרות עדיפה יותר על האדם כדי לנצח בטורניר?

תשובות סופיות:

(1)	א. $\frac{2}{7}$	ב. $\frac{23}{49}$
(2)	א. 6%	ב. 58%
(3)	א. 0.544	ב. 0.5
(4)	א. 9%	ב. 0.09375
(5)	א. 0.14	ב. 0.3488
(6)	א. 70%	ב. 0.8125
(7)	א. 0.57	ב. 0.3158
(8)	א. 0.0886	ב. 0.2889
(9)	$\frac{kp}{1+p(k-1)}$	
(10)	א'.	

סטטיסטיקה והסתברות

פרק 11 - תלות ואי תלות בין מאורעות

תוכן העניינים

1. כללי 42

תלות ואי תלות בין מאורעות:

רקע:

אם מתקיים ש: $P(B|A) = P(B)$, נגיד שמאורע B בלתי תלוי ב- A .
 הדבר גורר גם ההפך: $P(A|B) = P(A)$, כלומר, גם A אינו תלוי ב- B .
 כשהמאורעות בלתי תלויים מתקיים ש: $P(A \cap B) = P(A) \cdot P(B)$.
 הוכחה לכך: $P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A) \cdot P(B)$.

נשתמש בנוסחאות של מאורעות בלתי תלויים רק אם נאמר במפורש שהמאורעות בלתי תלויים בתרגיל או שמההקשר אפשר להבין ללא צל של ספק שהמאורעות בלתי תלויים.

למשל,

חוקר מבצע שני ניסויים בלתי תלויים הסיכוי להצליח בניסוי הראשון הנו 0.7 והסיכוי להצליח בניסוי השני הוא 0.4.

א. מה הסיכוי להצליח בשני הניסויים יחדו?

כיוון שהמאורעות הללו בלתי תלויים:

$$P(A \cap B) = P(A) \cdot P(B) = 0.7 \cdot 0.4 = 0.28$$

ב. מה הסיכוי להיכשל בשני הניסויים?

באופן דומה:

$$P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B}) = (1 - 0.7)(1 - 0.4) = 0.18$$

הרחבה: אי תלות בין n מאורעות:

n מאורעות $A_1, \dots, A_n$ הם בלתי תלויים אם ורק אם: $P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$.

שאלות:

- (1) נתון: $P(A) = 0.2$, $P(B) = 0.5$, $P(A \cup B) = 0.6$.
האם המאורעות הללו בלתי תלויים?
- (2) תלמיד ניגש לשני מבחנים שהצלחתם לא תלויה זו בזו.
הסיכוי שלו להצליח במבחן הראשון הוא 0.7 והשני 0.4.
א. מה הסיכוי להצליח בשני המבחנים יחד?
ב. מה הסיכוי שניכשל בשני המבחנים?
- (3) במדינה מסוימת יש 8% אבטלה, נבחרו באקראי שני אנשים מהמדינה.
א. מה ההסתברות ששניהם מובטלים?
ב. מה ההסתברות שלפחות אחד מהם מובטל?
- (4) מוצר צריך לעבור בהצלחה ארבע בדיקות בלתי תלויות לפני שיווק, אחרת הוא נפסל ולא יוצא לשוק. הסיכוי לעבור בהצלחה כל אחת מהבדיקות הוא 0.8. בכל מקרה מבוצעות כל 4 הבדיקות.
א. מה הסיכוי שהמוצר יפסל?
ב. מה ההסתברות שהמוצר יעבור בהצלחה לפחות בדיקה אחת?
- (5) במדינה מסוימת יש 8% אבטלה, נבחרו באקראי חמישה אנשים מהמדינה.
א. מה ההסתברות שכולם מובטלים במדגם?
ב. מה ההסתברות שלפחות אחד מהם מובטל?
- (6) עבור שני מאורעות A ו- B המוגדרים על אותו מרחב מדגם נתון ש:
 $P(A|B) = 0.6$, $P(A \cap \bar{B}) = 0.3$, $P(A \cup B) = 0.9$.
האם A ו- B מאורעות בלתי תלויים?
- (7) הוכיח שאם: $P(A/B) = P(B/A)$, אז: $P(A) = P(B)$.

(8) קבעו אילו מהטענות הבאות נכונות. נמק!

- א. אם : $P(A \cup B) = P(A) \cdot P(B)$, אזי המאורעות בלתי תלויים.
 ב. מאורע A כלול במאורע B : $0 < P(B) < 1$, $P(A) > 0$, לכן : $P(A/B) < P(A)$.
 ג. A ו- B מאורעות זרים שסיכוייהם חיוביים לכן הם מאורעות תלויים.
 ד. A ו- B מאורעות תלויים שסיכוייהם חיוביים לכן A ו- B מאורעות זרים.
 ה. $\bar{P}(\bar{A} \cap \bar{B}) = 1 - P(A) - P(B)$ לכן A ו- B מאורעות זרים.

תשובות סופיות:

- (1) כן.
 (2) א. 0.28 ב. 0.18
 (3) א. 0.0064 ב. 0.1536
 (4) א. 0.5904 ב. 0.9984
 (5) א. 0.08^5 ב. 0.3409
 (6) לא, הם תלויים.
 (7) שאלת הוכחה.
 (8) א. לא נכון. ב. לא נכון. ג. נכון. ד. לא נכון. ה. נכון.

סטטיסטיקה והסתברות

פרק 12 - שאלות מסכמות בהסתברות

תוכן העניינים

1. כללי 45

שאלות מסכמות בהסתברות:

שאלות:

- (1) נלקחו משפחות שיש להם שתי מכוניות. ל-30% מהמשפחות הללו המכונית הישנה יותר היא מתוצרת אירופה ואצל 60% מהמשפחות הללו המכונית החדשה יותר מתוצרת אירופה. כמו כן 15% מהמשפחות הללו שתי המכוניות הן מתוצרת אירופאית.

 - א. מה ההסתברות שמשפחה אקראית בת שתי מכוניות תהיה ללא מכוניות מתוצרת אירופה?
 - ב. מה ההסתברות שלפחות מכונית אחת תהיה אירופאית?
 - ג. ידוע שלמשפחה יש מכונית אירופאית.
 - מה ההסתברות שרק המכונית החדשה שלה היא מתוצרת אירופאית?
 - ד. אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית?
- (2) במדינת "שומקום" 50% מהחלב במרכולים מיוצר במחלבה א', 40% במחלבה ב' והיתר במחלבה ג'. 3% מתוצרת מחלבה א' מגיעה חמוצה למרכולים ואילו במחלבה ב' 10%.

כמו כן ידוע שבמדינת "שומקום" בסך הכול 7.5% מהחלב חמוץ.

 - א. איזה אחוז מהחלב שמגיע למרכול ממחלבה ג' חמוץ?
 - ב. אם נרכש חלב חמוץ במרכול. מה הסיכוי שהוא יוצר במחלבה ג'?
 - ג. ברכישת חלב נמצא שהוא אינו חמוץ. מה הסיכוי שהוא יוצר במחלבה א'?
 - ד. האם המאורעות: "חלב חמוץ" ו-"יוצר במחלבה א'" בלתי תלויים?
- (3) רוני ורונה יצאו לבלות במרכז בילויים עם מספר אפשרויות בילוי: בהסתברות של 0.3 הם ייצאו לבאולינג, בהסתברות של 0.5 הם ייצאו לבית קפה ובהסתברות של 0.7 הם ייצאו לפחות לאחד מהם (באולינג/קפה).

 - א. מה ההסתברות שהם יצאו רק לבאולינג?
 - ב. האם המאורעות "לצאת לבאולינג" ו-"לצאת לבית קפה" זרים?
 - ג. האם המאורעות "לצאת לבאולינג" ו-"לצאת לבית קפה" תלויים?
 - ד. מה ההסתברות שיום אחד הם יצאו רק לבאולינג וביום למחרת לא יצאו לאף אחד מהמקומות?

(4) 70% מהנבחנים בסטטיסטיקה עוברים את מועד א'. כל מי שלא עובר את מועד א' ניגש לעשות מועד ב', מתוכם 80% עוברים אותו. מבין אלה שנכשלים בשני המועדים 50% נרשמים לקורס מחדש, והיתר פורשים מהתואר.

- מה הסיכוי שסטודנט אקראי עבר את הקורס?
- אם סטודנט אקראי עבר הקורס, מה הסיכוי שעבר במועד ב'?
- מה אחוז הסטודנטים שפורשים מהתואר?
- נבחרו 2 סטודנטים אקראיים רונית וינאי, מה ההסתברות שרונית עברה במועד א' ושינאי עבר במועד ב'?

(5) באוכלוסייה מסוימת 40% הם גברים והיתר הן נשים. מבין הגברים 10% מובטלים. בסך הכול 13% מהאוכלוסייה מובטלת.

- מה אחוז האבטלה בקרב הנשים?
- נבחר אדם מובטל, מה ההסתברות שזו אישה?
- נגדיר את המאורעות הבאים: A - נבחר אדם מובטל, B - נבחר גבר. האם המאורעות הללו זרים? והאם הם בלתי תלויים?

(6) בתיבה 10 מטבעות, מתוכם 7 מטבעות רגילים (ראש, זנב) ו-3 מטבעות שבשני צדדיהם טבוע ראש. אדם בוחר באקראי מטבע ומטיל אותו פעמיים. נסמן ב-A את ההטלה הראשונה ראש, וב-B את ההטלה השנייה ראש.

- חשבו את הסיכויים למאורעות A ו-B.
- האם המאורע A ו-B בלתי תלויים?
- ידוע שבהטלה הראשונה התקבל ראש, מה ההסתברות שהמטבע שהוטל הוא מטבע הוגן?

(7) ערן מעוניין למכור את רכבו והוא מפרסם מודעה באינטרנט ומודעה בעיתון. מבין אלה שמעוניינים לרכוש רכב משומש 30% יראו את המודעה באינטרנט, 50% יראו את המודעה בעיתון ו-72% יראו את המודעה בלפחות אחת מהמדיות.

- מה אחוז האנשים, מאלה שמעוניינים לרכוש רכב משומש, שיראו את 2 המודעות?
- אם אדם ראה את המודעה באינטרנט, מה ההסתברות שהוא לא ראה את המודעה בעיתון?
- האם המאורעות: "לראות את המודעה באינטרנט" ו-"לראות את המודעה בעיתון" בלתי תלויים?
- אדם שראה את המודעה באינטרנט בלבד יתקשר לערן בהסתברות של 0.7, אם הוא ראה את המודעה בעיתון בלבד הוא יתקשר לערן בהסתברות של 0.6 ואם הוא ראה את שתי המודעות הוא יתקשר לערן בהסתברות של 0.9.
 - מה ההסתברות שאדם המעוניין לרכוש רכב משומש יתקשר לערן?
 - אדם המעוניין לרכוש רכב משומש התקשר לערן. מה ההסתברות שהוא ראה את שתי המודעות?

תשובות סופיות:

א. 0.25	ב. 0.75	ג. 0.6	ד. 0.5
(1)			
א. 0.2	ב. 0.267	ג. 0.524	ד. תלויים.
(2)			
א. 0.2	ב. אינם זרים.	ג. תלויים.	ד. 0.06
(3)			
א. 0.94	ב. 0.255	ג. 0.03	ד. 0.168
(4)			
א. 15%	ב. 0.692	ג. לא זרים ותלויים.	
(5)			
א. 0.65	ב. תלויים.	ג. 0.5384	
(6)			
א. 8%	ב. 0.733	ג. תלויים.	ד. i. 0.478
(7)			
			ד. ii. 0.15

סטטיסטיקה והסתברות

פרק 13 - המשתנה המקרי הבדיד - פונקציית ההסתברות

תוכן העניינים

1. כללי 48

המשתנה המקרי הבדיד – פונקציית ההסתברות:

רקע:

משתנה מקרי בדיד:

משתנה מקרי בדיד הינו משתנה היכול לקבל כמה ערכים בודדים בהסתברויות שונות.

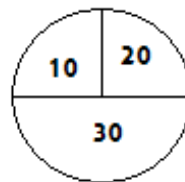
מתארים את המשתנה המקרי על ידי פונקציית ההסתברות.

פונקציית ההסתברות:

פונקציה המתאימה לכל ערך אפשרי של המשתנה את ההסתברות שלה. סכום ההסתברויות על פונקציית ההסתברות חייב להיות 1.

דוגמה (פתרון בהקלטה):

בקזינו יש רולטה כמתואר בשרטוט:



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה בש"ח. בנו את פונקציית ההסתברות של סכום הזכייה במשחק בודד.

שאלות:

- (1) ידוע שביישוב מסוים התפלגות מספר המכוניות למשפחה היא:
- 50 משפחות אינן מחזיקות במכונית.
 - 70 משפחות עם מכונית אחת.
 - 60 משפחות עם 2 מכוניות.
 - 20 משפחות עם 3 מכוניות.
- בוחרים באקראי משפחה מהיישוב, נגדיר את X להיות מספר המכוניות של המשפחה שנבחרה. בנו את פונקציית ההסתברות של X .
- (2) מהאותיות: A, B, C יוצרים קוד דו תווי.
- א. כמה קודים ניתן ליצור?
 - ב. רשמו את כל הקודים האפשריים.
 - ג. נגדיר את X להיות מספר הפעמים שהאות B מופיעה בקוד.
- בנו את פונקציית ההסתברות של X .
- (3) 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 המבחנים שהסטודנט עבר. בנו את פונקציית ההסתברות של X .
- (4) הסיכוי לזכות במשחק מסוים הינו 0.3. אדם משחק את המשחק עד אשר הוא מנצח אך בכל מקרה הוא לא משחק את המשחק יותר מ-4 פעמים. נגדיר את X להיות מספר הפעמים שהוא שיחק את המשחק. בנו את פונקציית ההסתברות של X .
- (5) חברה לניהול פרויקטים מנהלת 3 פרויקטים במקביל. הסיכוי שפרויקט א' יצליח הינו 0.7, הסיכוי שפרויקט ב' יצליח הינו 0.8, והסיכוי שפרויקט ג' יצליח הינו 0.9. נתון שהצלחת כל פרויקט בלתי תלויה זו בזו. נגדיר את X להיות מספר הפרויקטים שיצליחו. בנו את פונקציית ההסתברות של X .
- (6) להלן פונקציית הסתברות של משתנה מקרי כלשהו: $k = 1, 2, \dots, 4$: $P(X = k) = \frac{k}{A}$. מצאו את ערכו של A .

- (7) בגן ילדים 8 ילדים, מתוכם 5 בנים ו-3 בנות.
 בוחרים באקראי 3 ילדים להשתתף בהצגה.
 נגדיר את X כמספר הבנים שנבחרו להצגה.
 בנו את פונקציית ההסתברות של X .
- (8) בסקר שנערך בדקו בקרב אנשים האם הם צופים במהדורת החדשות של ערוצים 1,2,10. להלן הנתונים:
 20% צופים בערוץ 2.
 8% צופים בערוץ 1.
 10% צופים בערוץ 10.
 כמו כן נתון ש 1% צופים בשלושת המהדורות גם יחד.
 10% צופים בשתי המהדורות מתוך השלושה.
 נגדיר את X להיות מספר המהדורות מבין 3 המהדורות המדוברות שאדם אקראי צופה. בנו את פונקציות ההסתברות של X .

תשובות סופיות:

(1) להלן טבלה:

3	2	1	0	X
0.1	0.3	0.35	0.25	$P(X)$

(2) להלן טבלה:

2	1	0	X
$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$	$P(X)$

(3) להלן טבלה:

2	1	0	X
0.75	0.20	0.05	$P(X)$

(4) להלן טבלה:

4	3	2	1	X
0.343	0.147	0.21	0.3	$P(X)$

(5) להלן טבלה:

3	2	1	0	X
0.504	0.398	0.092	0.006	$P(X)$

(6) 10.

(7) להלן טבלה:

4	3	2	1	X
$\frac{10}{56}$	$\frac{30}{56}$	$\frac{15}{56}$	$\frac{1}{56}$	$P(X)$

(8) להלן טבלה:

4	3	2	1	X
0.01	0.1	0.15	0.74	$P(X)$

סטטיסטיקה והסתברות

פרק 14 - המשתנה המקרי הבדיד - תוחלת - שונות וסטיית תקן

תוכן העניינים

1. כללי 52

המשתנה המקרי הבדיד – תוחלת, שונות וסטיית תקן:

רקע:

תוחלת:

ממוצע של פונקציית ההסתברות, אם נבצע את התהליך אינסוף פעמים כמה בממוצע נקבל. התוחלת היא צפי של המשתנה המקרי.

$$\text{מגדירים תוחלת באופן הבא: } E(X) = \sum_i x_i P(x_i) = \mu$$

שונות:

תוחלת ריבועי הסטיות מהתוחלת – נותן אינדיקציה על הפיזור והסיכון של פונקציית ההסתברות.

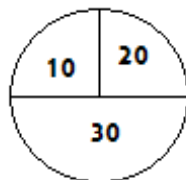
$$\text{מגדירים שונות באופן הבא: } V(X) = \sum_i (x_i - \mu)^2 P(x_i) = \sum_i x_i^2 P(x_i) - \mu^2 = \sigma^2$$

סטיית תקן:

שורש של השונות – הפיזור הממוצע הצפוי סביב התוחלת. מסמנים: $STD = \sigma$.

דוגמה:

בקזינו רולטה כמוראה בשרטוט. אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה ב-ש. הסתברות לקבלת הסכומים השונים:



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

$$E(X) = 10 \cdot 0.25 + 20 \cdot 0.25 + 30 \cdot 0.5 = 22.5 = \mu$$

$$V(X) = \sum_i (x_i - \mu)^2 P(x_i) =$$

$$= (10 - 22.5)^2 \cdot 0.25 + (20 - 22.5)^2 \cdot 0.25 + (30 - 22.5)^2 \cdot 0.5 = 68.75 = \sigma^2$$

$$\text{כדי לחשב את סטיית התקן נוציא שורש לשונות: } \sigma_x = \sqrt{V(X)} = \sqrt{68.75} = 8.29$$

שאלות:

- (1) אדם משחק במשחק מזל.
נגדיר את X להיות סכום הזכייה.
להלן פונקציית ההסתברות של X :

40	20	0	-30	X
0.2	0.3	0.1	0.4	$P(X)$

מהי התוחלת, השונות וסטית התקן של X ?

- (2) בישוב מסוים שני סניפי בנק: בנק פועלים ובנק לאומי. מתוך האוכלוסייה הבוגרת בישוב, ל-50% חשבון בנק בסניף הפועלים, ל-40% חשבון בנק בסניף לאומי ול-20% מהתושבים הבוגרים אין חשבון באף אחד מהסניפים. יהי X מס' סניפי הבנק שלבוגר בישוב יש בהם חשבון. חשבו את: $E(X)$.

- (3) ידוע של-20% מהמשפחות יש חיבור לווייני בביתם. בסקר אדם מחפש לראיין משפחה המחוברת ללוויין. הוא מטלפן באקראי למשפחה וממשיך עד אשר הוא מגיע למשפחה המחוברת ללוויין. בכל מקרה הסוקר לא יתקשר ליותר מ-5 משפחות. נגדיר את X להיות מספר המשפחות שאליהן האדם יתקשר.
א. בנו את פונקציית ההסתברות של X .
ב. חשבו את התוחלת וסטית התקן של X .

- (4) לאדם צרור מפתחות. בצרור 5 מפתחות אשר רק אחד מתאים לדלת של ביתו. האדם מנסה את המפתחות באופן מקרי. לאחר שניסה מפתח מסוים הוא מוציא אותו מהצרור כדי שלא ישתמש בו שוב.
נסמן ב- X את מספר הניסיונות עד שהדלת תפתח.
א. בנו את פונקציית ההסתברות של X .
ב. חשבו את התוחלת והשונות של X .

(5) נתונה פונקציית ההסתברות של המשתנה המקרי X :

8	6	4	2	X
0.2		0.3		$E(X)$

כמו כן נתון ש: $E(X) = 4.2$.

א. מצאו את ההסתברויות החסרות בטבלה.

ב. חשבו את: $V(X)$.

(6) משתנה מקרי בדיד מקבל את הערכים 5-0 ו-5. נתון שהתוחלת של המשתנה 0 ושהשונות היא 10. מצא את פונקציית ההסתברות.

(7) להלן ההתפלגות של משתנה מקרי:

X	P
1	$\frac{1}{4}$
3	$\frac{1}{2}$
K	$\frac{1}{4}$

מהו הערך שייתן ערך מינימלי לשונות של X ?

תשובות סופיות:

(1) תוחלת: 2, שונות: 796.

(2) 0.9.

(3) א. ראו סרטון. ב. תוחלת: 3.36, סטיית תקן: 1.603.

(4) א. ראו טבלה: ב. תוחלת: 3, שונות: 2.

5	4	3	2	1	X
0.2	0.2	0.2	0.2	0.2	$P(X)$

(5) א. ראו טבלה: ב. 5.16.

8	6	4	2	X
0.2	0.1	0.3	0.4	$P(X)$

(6) ראו טבלה:

5	0	-5	X
0.2	0.6	0.2	$P(X)$

(7) 2.33.

סטטיסטיקה והסתברות

פרק 15 - המשתנה המקרי הבדיד - טרנספורמציה לינארית

תוכן העניינים

1. כללי 56

המשתנה המקרי הבדיד – טרנספורמציה לינארית:

רקע:

טרנספורמציה לינארית היא מצב שבו מבצעים הכפלה של קבוע ו/או הוספה של קבוע על המשתנה המקורי (כולל גם חלוקה של קבוע והחסרה של קבוע).

בניסוח מתמטי נאמר כי אם משתנה אקראי Y מיוצג ע"י משתנה אקראי X כאשר a, b הם קבועים כלשהם: $Y = aX + b$, אזי מתקיימים:

$$E(Y) = aE(X) + b \quad (1)$$

$$V(Y) = a^2 \cdot V(X) \quad (2)$$

$$\sigma_Y = |a| \sigma_X \quad (3)$$

שלבי העבודה:

- (1) נזהה שמדובר בטרנספורמציה לינארית (שינוי קבוע לכל התצפיות).
- (2) נרשום את כלל הטרנספורמציה לפי נתוני השאלה.
- (3) נפשט את הכלל ונזהה את ערכי a ו- b .
- (4) נציב בנוסחאות שלעיל בהתאם למדדים שנשאלים.

דוגמה – הרולטה:

בהמשך לנתוני שאלת הרולטה נתון שעלות השתתפות במשחק 15 ₪. מהי התוחלת והשונות של הרווח במשחק?

פתרון (בהקלטה):

חישבנו קודם ש: $V(X) = 68.75 = \sigma^2$, $E(X) = 22.5 = \mu$.

שאלות:

(1) סטודנט ניגש ל-5 קורסים הסמסטר. נניח שכל קורס שסטודנט מסיים מזכה אותו ב-4 נקודות אקדמאיות. חשבו את התוחלת והשונות של סך הנקודות שיצבור הסטודנט כאשר נתון שתוחלת מספר הקורסים שיסיים היא 3.5 עם שונות 2.

(2) תוחלת סכום הזכייה במשחק מזל הינה 10 עם שונות 3. הוחלט להכפיל את סכום הזכייה במשחק. עלות השתתפות במשחק הינה 12. מה התוחלת ומהי השונות של הרווח במשחק?

(3) תוחלת של משתנה מקרי הינה 10 וסטית התקן 5. הוחלט להוסיף 2 למשתנה ולאחר מכן להעלות אותו ב-10%. מהי התוחלת ומהי סטיית התקן לאחר השינוי?

(4) X הינו משתנה מקרי. כמו כן נתון ש- $E(X) = 4$ ו- $V(X) = 3$.
 Y הינו משתנה מקרי חדש, עבורו: $Y = 7 - X$. חשבו את: $E(Y)$ ו- $V(Y)$.

(5) אדם החליט לבטח את רכבו; שווי הרכב 100,000 ₪. להלן התביעות האפשריות והסתברותן: בהסתברות של 0.001 תהיה תביעה טוטאלוסט (כל שווי הרכב). בהסתברות של 0.02 תהיה תביעה בשווי מחצית משווי הרכב. בהסתברות של 5% תהיה תביעה בשווי רבע משווי הרכב. אחרת אין תביעה בכלל. החברה מאפשרת תביעה אחת בשנה. נסמן ב- X את גובה התביעה השנתית, באלפי ₪.
 א. בנו את פונקציית ההסתברות של X .
 ב. חשבו את התוחלת והשונות של גובה התביעה.
 ג. פרמיית הביטוח היא 4,000 ₪.
 מהי התוחלת ומהי השונות של רווח חברת הביטוח לביטוח הרכב הני"ל?

(6) יהי X מספר התשובות הנכונות במבחן בו 10 שאלות. פונקציית ההסתברות של X נתונה בטבלה הבאה:

10	9	8	7	6	5	X
		0.3	0.2	0.2	0.1	$P(X)$

כמו כן, נתון שצפי מספר התשובות הנכונות בבחינה הוא 7.35.

- השלימו את פונקציית ההסתברות.
 - חשבו את השונות מספר התשובות הנכונות בבחינה.
 - הציון בבחינה מחושב באופן הבא:
- כל שאלה נכונה מזכה ב-10 נקודות. לכל שאלה שגויה, מופחתת נקודה. מהי התוחלת ומה השונות של הציון בבחינה?

(7) להלן פונקציית הסתברות של משתנה מקרי כלשהו: $k=1,2,3,4$: $P(X=k) = \frac{k}{A}$

- א. מצא את ערכו של A .
- ב. חשב את התוחלת והשונות של המשתנה הנחקר.
- ג. חשב את: $E(X^3)$.
- ד. חשב את התוחלת והשונות של המשתנה הבא: $\frac{X}{2} - 4$.

תשובות סופיות:

- (1) תוחלת: 14, שונות: 32.
- (2) תוחלת: 8, שונות: 12.
- (3) תוחלת: 13.2, סטיית תקן: 5.5.
- (4) תוחלת: 3, שונות: 3.
- (5) א. להלן טבלה: ב. תוחלת: 2350, שונות: $85,727.5^2$.

0	25	50	100	X
0.929	0.05	0.02	0.001	$P(X)$

- ג. תוחלת: 1650, שונות: $85,727.5^2$.
- (6) ב. $V(X) = 1.8275$.
- (7) א. $A=10$. ב. $E(X)=3$, $V(X)=1$. ג. $E(X^3)=35.4$, $V(X^3)=616.84$. ד. $E(Y)=-2.5$, $V(Y)=0.25$.

סטטיסטיקה והסתברות

פרק 16 - תוחלת ושונות של סכום משתנים מקריים

תוכן העניינים

1. כללי 59

תוחלת ושונות של סכום משתנים מקריים:

רקע:

אם: $X_1, X_2, \dots, X_n$ משתנים מקריים אזי:

$$E(T) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

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

$$V(T) = V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$$

דוגמה:

אדם משחק בשני משחקי מזל בלתי תלויים. תוחלת סכות הזכייה של המשחק הראשון היא 7 עם סטיית תקן 3. תוחלת סכום הזכייה של המשחק השני היא 2- עם סטיית תקן 4. מה התוחלת ומהי השונות של סכום הזכייה הכולל של שני המשחקים יחד?

שאלות:

(1) הרווח ממניה א' הוא עם תוחלת של 5 ושונות 10. הרווח ממניה ב' הוא עם תוחלת של 4 ושונות. ידוע שההשקעות של שתי המניות בלתי תלויות זו בזו. מה התוחלת והשונות של הרווח הכולל מהשקעה בשתי המניות יחד?

(2) X ו- Y הם משתנים בלתי תלויים, סטיית התקן של X היא 3. סטיית התקן של Y היא 4. מהי סטיית התקן של $X+Y$?

(3) אדם משחק בשני משחקי מזל בלתי תלויים זה בזה:
 X - סכום הזכייה במשחק הראשון.
 Y - סכום הזכייה במשחק השני.
 נתון:

$$\sigma(X) = 3, \quad E(x) = 10$$

$$\sigma(Y) = 4, \quad E(y) = 12$$

מהי התוחלת ומהי סטיית התקן של סכום הזכייה בשני המשחקים?

(4) ברולטה הסיכוי לזכות ב- 30 ₪ הוא חצי, ב-10 ₪ רבע וכך גם ב-20 ₪. מה היא התוחלת והשונות של סכום הזכייה הכולל לאדם המשחק ברולטה 4 פעמים?

(5) נתון משתנה מקרי בעל פונקציית ההסתברות הבאה:

$$K = 2, 3, 4, 5, \quad P(X = K) = \frac{A}{K-1}$$

$$0 \quad \text{אחר} \quad \text{ת}$$

מצאו את ערכו של A .

א. חשבו את התוחלת והשונות של X .

ב. נלקחו n משתנים מקריים בלתי תלויים מההתפלגות הנ"ל. בטאו באמצעות n את תוחלת והשונות של סכום המשתנים.

תשובות סופיות:

- (1) תוחלת: 9, שונות: 15.
- (2) 5.
- (3) תוחלת: 22, שונות: 5.
- (4) תוחלת: 90, שונות: 275.
- (5) א. $A = \frac{12}{25} = 0.48$. ב. תוחלת: 2.92, שונות: 1.1136.
- ג. תוחלת: 2.92, שונות: $1.1136n$.

סטטיסטיקה והסתברות

פרק 17 - התפלגויות בדידות מיוחדות - התפלגות בינומית

תוכן העניינים

1. כללי 62

התפלגויות בדידות מיוחדות – התפלגות בינומית:

רקע:

נגדיר את המושג ניסוי ברנולי :

ניסוי ברנולי הנו ניסוי שיש לו שתי תוצאות אפשריות: "הצלחה" ו"כישלון". למשל מוצר פגום או תקין, אדם עובד או מובטל, עץ או פלי בהטלת מטבע וכדומה. בהתפלגות בינומית חוזרים על אותו ניסוי ברנולי n פעמים באופן בלתי תלוי זה בזה. מגדירים את X להיות מספר ההצלחות שהתקבלו בסך הכול. נסמן ב- P את הסיכוי להצלחה בניסוי בודד, וב- Q את הסיכוי לכישלון בניסוי בודד. ואז נגיד ש: $X \sim B(n, p)$.

פונקציית ההסתברות של X :

$$P(X = K) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, 2, \dots, n \quad \text{לכל}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}; \quad n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1; \quad 0! = 1; \quad \text{כאשר}$$

לגודל: $\binom{n}{k}$ ניתן לחשב באמצעות המחשבון.

$$E(X) = np \quad \text{תוחלת:}$$

$$V(X) = npq \quad \text{שונות:}$$

שימו לב, כדי לזהות שמדובר בהתפלגות בינומית צריכים להתקיים כל התנאים הבאים:

- (1) חוזרים על אותו ניסוי ברנולי באופן בלתי תלוי זה בזה.
- (2) חוזרים על הניסוי n פעמים.
- (3) X – מוגדר כמספר ההצלחות המתקבלות בסך הכול.

דוגמה (פתרון בהקלטה):

במדינה מסוימת ל-80% מהתושבים יש רישיון נהיגה. נבחרו 10 תושבים אקראיים מהמדינה.

- א. מה ההסתברות שבדיוק ל-9 מהם יש רישיון נהיגה?
- ב. מה ההסתברות שלפחות ל-9 מהם יש רישיון נהיגה?
- ג. מהי התוחלת ומהי סטיית התקן של מספר התושבים שנדגמו ושיש להם רישיון נהיגה?

שאלות:

- (1) במדינה 10% מהאוכלוסייה מובטלת. נבחרו 5 אנשים באקראי מאותה אוכלוסייה. נגדיר את X להיות מספר המובטלים שהתקבלו במדגם.
- מהי ההתפלגות של X ?
 - מה ההסתברות שיהיה בדיוק מובטל אחד?
 - מה ההסתברות שכולם יעבדו במדגם?
 - מה ההסתברות ששלושה יעבדו במדגם?
 - מה ההסתברות שלפחות אחד יהיה מובטל?
 - מה תוחלת ומהי השונות של מספר המובטלים במדגם?
- (2) על פי נתוני משרד התקשורת ל-70% מהאוכלוסייה יש סמארטפון. נבחרו 10 אנשים באקראי. נגדיר את X כמספר האנשים שנדגמו עם סמארטפון.
- מהי ההתפלגות של X ? הסבירו.
 - מה ההסתברות שבמדגם ל-8 אנשים יש סמארט-פון?
 - מה ההסתברות שבמדגם לפחות ל-9 יהיו סמארט-פון?
 - מה התוחלת ומה סטיית התקן של מספר האנשים שנדגמו ולהם סמארט-פון?
- (3) בבית הימורים יש שורה של 6 מכונות מזל מאותו סוג. משחק במכונת מזל כזו עולה 5 ₪. ההסתברות לזכות ב-20 ₪ בכל אחת מהמכונות היא 0.1 וההסתברות להפסיד את ההשקעה היא 0.9 בכל מכונה. מהמר נכנס לבית ההימורים ומכניס 5 ₪ לכל אחת מ-6 המכונות.
- מה ההסתברות שיפסיד בכל המכונות?
 - מה ההסתברות שיזכה בדיוק בשתי מכונות?
 - מה ההסתברות שיזכה ביותר כסף מה-30 ₪ שהשקיע?
 - מהן התוחלת וסטיית התקן של הרווח נטו של המהמר (הזכיות בניכוי ההשקעה)?
- (4) במדינה מסוימת התפלגות ההשכלה בקרב האוכלוסייה מעל גיל 30 היא כזו:
- | השכלה | נמוכה | תיכונית | תואר I | תואר II ומעלה |
|-----------|-------|---------|--------|---------------|
| פרופורציה | 0.1 | 0.6 | 0.2 | 0.1 |
- נבחרו 20 אנשים אקראיים מעל גיל 30.
- מה ההסתברות ש-5 מהם אקדמאים?
 - מה התוחלת של מס' בעלי ההשכלה הנמוכה?

- (5) במכללה מסוימת 20% מהסטודנטים גרים בת"א. מבין הסטודנטים שגרים בת"א 30% מגיעים ברכבם, ומבין הסטודנטים שלא גרים בת"א 50% מגיעים ברכבם למכללה.
- א. השומר בשער המכללה בודק לכל סטודנט את תיקו בהיכנסו למכללה. מה ההסתברות שבקרב 5 סטודנטים שנבדקו ע"י השומר רק 1 מתוכם הגיע למכללה ברכבו?
- ב. בהמשך לסעיף הקודם מה ההסתברות שרוב הסטודנטים בקרב ה-5 הגיעו למכללה ברכבם?
- (6) במבחן אמריקאי 20 שאלות. סטודנט ניגש למבחן והסיכוי שהוא יודע שאלה כלשהי הוא 0.8. אם הוא לא יודע הוא מנחש את התשובה. לכל שאלה 4 תשובות אפשריות שרק אחת מהן נכונה.
- א. מה הסיכוי לענות על שאלה מסוימת נכון?
- ב. מה הסיכוי שיענה נכונה על בדיוק 16 שאלות?
- ג. על כל שאלה שענה נכון התלמיד מקבל 5 נקודות, על כל שאלה ששגה מופחתת נקודה, מה התוחלת ומהי השונות של ציון התלמיד?
- (7) 5% מקו היצור פגום. המוצרים נארזים בתוך קופסת קרטון. בכל קופסא 10 מוצרים שונים. הקופסאות נארזות בתוך מכולה. בכל מכולה 20 קופסאות.
- א. מה ההסתברות שבקופסא אקראית לפחות מוצר פגום אחד?
- ב. מה התוחלת ומהי סטיית התקן של מספר הקופסאות במכולה בהן לפחות מוצר פגום אחד?
- (8) מטבע הוגן מוטל 5 פעמים. נגדיר את X כמספר הפעמים שהתקבל עץ. חשבו את: $E(X^2)$.

תשובות סופיות:

- (1) א. $X \sim B(n=5, p=0.1)$ ב. 0.32805 ג. 0.59049
- ד. 0.0729 ה. 0.40954 ו. תוחלת: 0.5, שונות: 0.45
- (2) א. 0.5314 ב. 0.2335 ג. 0.1493 ד. תוחלת: 7, סטיית תקן: 1.449
- (3) א. 0.5314 ב. 0.0984 ג. 0.1143 ד. תוחלת: 18-, סטיית תקן: 14.697
- (4) א. 0.1789 ב. 2
- (5) א. 0.1956 ב. 0.4253
- (6) א. 0.85 ב. 0.182 ג. תוחלת: 82, שונות: 91.8
- (7) א. 0.401 ב. תוחלת: 8.025, סטיית תקן: 2.193
- (8) 7.5

סטטיסטיקה והסתברות

פרק 18 - התפלגויות בדידות מיוחדות - התפלגות אחידה

תוכן העניינים

1. כללי 66

התפלגויות בדידות מיוחדות – התפלגות אחידה:

רקע:

התפלגות אחידה הינה התפלגות שבה לכל תוצאה יש את אותה הסתברות. הערכים המתקבלים בהתפלגות הם החל מ- a ועד b בקפיצות של אחד.
 $X \sim U(a, b)$

פונקציית ההסתברות: $P(X = K) = \frac{1}{b - a + 1}$, $K = a, a + 1, \dots, b$

תוחלת: $E(X) = \frac{a + b}{2}$

שונות: $V(X) = \frac{(b - a + 1)^2 - 1}{12}$

דוגמה (פתרון בהקלטה) :

אדם בוחר מספר אקראי בין 1 ל-100 כולל.
 מהי פונקציית ההסתברות של המספר ומה הצפי שלו?

שאלות:

- (1) במשחק הלוטו 45 כדורים ממוספרים מ-1 ועד 45. נתבונן במשתנה X - המספר של הכדור הראשון שנשלף על ידי המכונה.
- חשבו את $P(X = 2)$.
 - חשבו את $P(X \leq 30)$.
 - חשבו את $P(X > 4 | X \leq 10)$.
 - חשבו את $P(X = k)$.
- (2) קוסם מבקש לבחור מספר שלם אקראי בין 1 ל-100.
- בהנחה שאין כאן מניפולציות של הקוסם, מהי התוחלת ומהי סטיית התקן של המספר שיבחר?
 - הקוסם ביקש משישה אנשים לבחור מספר:
 - מה ההסתברות ששלושה מהם יבחרו מספר הגדול מ-80?
 - מה התוחלת ומהי סטיית התקן של סכום המספרים שהאנשים בחרו?
- (3) יהי X התוצאה בהטלת קובייה.
- מהי ההתפלגות של X ?
 - מה התוחלת של X ?
 - קובייה הוטלה 4 פעמים. מה התוחלת ומה השונות של סכום התוצאות ב-4 ההטלות?
- (4) בכד 10 כדורים שרק אחד בצבע אדום. כדורים הוצאו ללא החזרה עד שהתקבל הכדור האדום. מה התוחלת ומהי השונות של מספר הכדורים שהוצאו?
- (5) יש לבחור מספר אקראי בין 1 ל-50, כולל.
- מה הסיכוי שהמספר 4 יבחר?
 - מה הסיכוי שהמספר שיבחר גדול מ-20?
 - אם נבחר מספר גדול מ-20, מה ההסתברות שהוא קטן מ-28?
- (6) הוכיחו שאם: $X \sim U(a, b)$, אז מתקיים ש: $E(X) = \frac{a+b}{2}$.

תשובות סופיות:

- (1) א. $\frac{1}{45}$ ב. $\frac{30}{45}$ ג. 0.6
- (2) א. תוחלת: 50.5, סטיית תקן: 28.87.
 ב. i. 0.08192 ii. תוחלת: 303, סטיית תקן: 70.71
- (3) א. $X \sim U(1,6)$ ב. 3.5 ג. תוחלת: 14, שונות: 11.66
- (4) תוחלת: 5.5, שונות: 8.25
- (5) א. $\frac{1}{50}$ ב. $\frac{30}{50}$ ג. $\frac{7}{30}$
- (6) שאלת הוכחה.

סטטיסטיקה והסתברות

פרק 19 - התפלגויות בדידות מיוחדות- התפלגות פואסונית

תוכן העניינים

1. כללי 69

התפלגויות בדידות מיוחדות – התפלגות פואסונית:

רקע:

התפלגות פואסונית היא התפלגות שמאפיינת את מספר האירועים שמתרחשים ביחידת זמן.

λ - פרמטר המאפיין את ההתפלגות הנ"ל. הפרמטר מייצג את קצב האירועים ביחידת זמן. כלומר, כמה אירועים בממוצע קורים ביחידת זמן: $X \sim \text{pois}(\lambda)$.
התפלגות פואסונית חייבת להופיע כנתון בשאלה ולכן לא יהיה צורך לזהותה.

פונקציית ההסתברות של ההתפלגות הפואסונית נתונה:

$$P(X = K) = \frac{e^{-\lambda} \cdot \lambda^K}{K!}, \quad K = 0, 1, 2, \dots, \infty$$

התוחלת והשונות של ההתפלגות:

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

תכונות מיוחדות של ההתפלגות:

- בהתפלגות הזו הפרמטר λ פרופורציונלי לאינטרוול הזמן שעליו דנים.
- אינטרוולי זמן לא חופפים בלתי תלויים זה בזה.

דוגמה (פתרון בהקלטה):

במוקד טלפוני מתקבלות פניות בקצב של 5 פניות לדקה. מספר הפניות בדקה מתפלג פואסונית.

- מה ההסתברות שבדקה כלשהי תתקבל פניה 1?
- מה ההסתברות שבשתי דקות יגיעו 12 פניות?
- מה ההסתברות שבדקה אחת תגיע פניה 1 ובשתי דקות שלאחר מכן 12 פניות?
- מה התוחלת וסטיית התקן של מספר הפניות בדקה?

שאלות:

- (1) במוקד טלפוני מתקבלות פניות בקצב של 5 פניות לדקה. מספר הפניות בדקה מתפלג פואסונית.
 - א. מה ההסתברות שבדקה תתקבל פניה 1?
 - ב. מה ההסתברות שבדקה תתקבל לפחות פניה 1?
 - ג. מה ההסתברות שבדקה יתקבלו לכל היותר 2 פניות?
 - ד. מה שונות מספר הפניות בדקה?
- (2) מספר הטעויות לעמוד בעיתון מתפלג פואסונית עם ממוצע של 4 טעויות לעמוד. בחלק מסוים של עיתון ישנם 5 עמודים.
 - א. מה ההסתברות שבחלק זה ישנן בדיוק 18 טעויות?
 - ב. אם בעמוד הראשון אין טעויות, מה ההסתברות שבסך הכול בחלק ישנן 15 טעויות?
 - ג. אם בחלק של העיתון נמצאו בסך הכול 18 טעויות, מה ההסתברות ש-5 מהן בעמוד הראשון?
- (3) מספר תאונות הדרכים הקטלניות במדינת ישראל מתפלג פואסונית עם סטיית תקן של 2 תאונות לשבוע.
 - א. מה תוחלת מספר התאונות בשבוע?
 - ב. מהי ההסתברות שבחודש (הניחו שבחודש יש 4 שבועות) יהיה בדיוק שבוע אחד בו יהיו 3 תאונות דרכים קטלניות?
- (4) לחנות AM:PM השכונתית מספר הלקוחות שנכנסים מתפלג פואסונית עם ממוצע של 2 לקוחות לדקה.
 - א. מה ההסתברות שבדקה כלשהי יהיו בדיוק 3 לקוחות?
 - ב. מה ההסתברות שבדקה כלשהי יגיח לפחות לקוח אחד?
 - ג. מה ההסתברות שבדקה כלשהי יהיו לכל היותר שני לקוחות?
 - ד. מהי התוחלת ומה סטיית התקן של מספר הלקוחות שנכנסים לחנות בדקה?
- (5) מספר הלידות בבית חולים מתפלג פואסונית עם תוחלת של 8 לידות ביום.
 - א. מה ההסתברות שביום א' נולדו 10 תינוקות וביום ב' נולדו 7 תינוקות?
 - ב. מיילדת עובדת במשמרות של 8 שעות. מה ההסתברות שבמשמרת שלה נולדו 3 תינוקות?
 - ג. מהי התוחלת של מספר הימים בשבוע בהם נולדים ביום עשרה תינוקות?

- (6) במערכת אינטרנט לתשלום חשבונות, מספר החשבונות המשולמים בשעה מתפלג פואסונית עם תוחלת של 30.
- א. כמה שעות צפויות לעבור עד אשר תתקבל שעה עם בדיוק 33 חשבונות?
- ב. בין השעה 08:00 ל-08:20 היו 18 חשבונות, מה ההסתברות שבין 08:00 ל-08:10 היו בדיוק 6 חשבונות?

תשובות סופיות:

(1)	א. 0.0337	ב. 0.9933	ג. 0.1246	ד. 5.
(2)	א. 0.084	ב. 0.099	ג. 0.151	
(3)	א. 4.	ב. 0.407		
(4)	א. 0.1804	ב. 0.8647	ג. 0.6767	ד. תוחלת: 2, סטיית תקן: 1.41.
(5)	א. 0.0139	ב. 0.2196	ג. 0.6948	
(6)	א. 16.7	ב. 0.0708		

סטטיסטיקה והסתברות

פרק 20 - המשתנה המקרי הרציף- התפלגויות כלליות (שימוש באינטגרלים)

תוכן העניינים

1. כללי 72

המשתנה המקרי הרציף – התפלגויות כלליות (שימוש באינטגרלים)

רקע:

בפרק זה נעסוק בהתפלגות של משתנים מקריים רציפים (גובה אדם אקראי, זמן תגובה וכו'). משתנים רציפים הם משתנים שבתחום מסוים מקבלים רצף אינסופי של ערכים אפשריים בניגוד למשתנים בדידים. נתאר את המשתנה המקרי הרציף על ידי פונקציה הנקראת פונקציית צפיפות. באופן כללי נסמן פונקציית צפיפות של משתנה רציף כלשהו ב- $f(x)$. השטח שמתחת לפונקציית הצפיפות נותן את ההסתברות. פונקציית צפיפות חייבת להיות לא-שלילית והשטח הכולל שמתחת לפונקציה יהיה תמיד 1.

הגדרות יסודיות:

יהא משתנה רציף X בעל פונקציית צפיפות $f(x)$.

פונקציית ההתפלגות מצטברת:

פונקציית ההתפלגות המצטברת מוגדרת באופן הבא: $F(t) = p(X \leq t) = \int_{-\infty}^t f(x) dx$.
כמו כן מתקיים: $p(X > t) = 1 - F(t)$ ו- $p(a < X < b) = F(b) - F(a)$.

תוחלת ושונות של משתנה רציף:

תוחלת של משתנה רציף תחושב באופן הבא: $E(X) = \int_{-\infty}^{\infty} X \cdot f(x) dx = \mu$.
שונות של משתנה רציף תחושב באופן הבא: $V(X) = \int_{-\infty}^{\infty} X^2 \cdot f(x) dx - \mu^2 = \sigma^2$.

תוחלת של פונקציה של X :

תוחלת של פונקציית משתנה רציף X , המסומנת : $g(x)$, תחושב באופן

$$\text{הבא : } E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

אחוזונים :

האחוזון ה- p הוא ערך (נסמן אותו : x_p), שהסיכוי ליפול מתחתיו הוא p .

$$\text{כלומר : } p(X \leq x_p) = p$$

ריענון מתמטי :

נוסחאות לחישוב שטחים

$$\text{שטח משולש : גובה } (h) \text{ כפול הבסיס } (a) \text{ חלקי } 2 : S_{\text{triangle}} = \frac{h \cdot a}{2}$$

$$\text{שטח מלבן : אורך } (a) \text{ כפול רוחב } (b) : S_{\text{rectangle}} = a \cdot b$$

משוואת קו ישר :

משוואת ישר מפורשת תסומן : $y = mx + n$, כאשר m הוא שיפוע הישר ו- n היא נקודת החיתוך של הישר עם ציר ה- y .

$$\text{שיפוע ישר העובר דרך שתי נקודות : } (x_1, y_1), (x_2, y_2) \text{ הוא : } m = \frac{y_2 - y_1}{x_2 - x_1}$$

משוואת ישר שעובר דרך נקודה ספציפית (x_1, y_1) ושיפועו הוא m , תחושב באופן

$$\text{הבא : } y - y_1 = m(x - x_1)$$

אינטגרלים מידיים:

$$\int a dx = ax + c$$

$$\int x^n dx = \frac{x^{n+1}}{n+1} + c \quad n \neq -1$$

$$\int \frac{1}{x} dx = \ln |x| + c$$

$$\int e^x dx = e^x + c$$

$$\int k^x dx = \frac{k^x}{\ln k} + c$$

$$\int \cos x dx = \sin x + c$$

$$\int \sin x dx = -\cos x + c$$

$$\int \tan x dx = -\ln |\cos x| + c$$

$$\int \cot x dx = \ln |\sin x| + c$$

$$\int \frac{1}{\cos^2 x} dx = \tan x + c$$

$$\int \frac{1}{\sin^2 x} dx = -\cot x + c$$

$$\int \frac{1}{\cos x} dx = \ln \left| \frac{1}{\cos x} + \tan x \right| + c$$

$$\int \frac{1}{x^2 + a^2} dx = \frac{1}{a} \arctan \left(\frac{x}{a} \right) + c$$

$$\int \frac{1}{\sqrt{a^2 - x^2}} dx = \arcsin \left(\frac{x}{a} \right) + c$$

$$\int \frac{f'}{f} dx = \ln |f| + c$$

$$\int e^f \cdot f' dx = e^f + c$$

$$\int \sin f \cdot f' dx = -\cos(f) + c$$

$$\int \sqrt{f} \cdot f' dx = \frac{2}{3} f^{\frac{3}{2}} + c$$

$$\int (ax+b)^n dx = \frac{1}{a} \frac{(ax+b)^{n+1}}{n+1} + c \quad n \neq -1$$

$$\int \frac{1}{ax+b} dx = \frac{1}{a} \ln |ax+b| + c$$

$$\int e^{ax+b} dx = \frac{1}{a} e^{ax+b} + c$$

$$\int k^{ax+b} dx = \frac{1}{a} \frac{k^{ax+b}}{\ln k} + c$$

$$\int \cos(ax+b) dx = \frac{1}{a} \sin(ax+b) + c$$

$$\int \sin(ax+b) dx = -\frac{1}{a} \cos(ax+b) + c$$

$$\int \tan(ax+b) dx = -\frac{1}{a} \ln |\cos(ax+b)| + c$$

$$\int \cot(ax+b) dx = \frac{1}{a} \ln |\sin(ax+b)| + c$$

$$\int \frac{1}{\cos^2(ax+b)} dx = \frac{1}{a} \tan(ax+b) + c$$

$$\int \frac{1}{\sin^2(ax+b)} dx = -\frac{1}{a} \cot(ax+b) + c$$

$$\int \frac{1}{\sin x} dx = \ln \left| \frac{1}{\sin x} - \cot x \right| + c$$

$$\int \frac{1}{x^2 - a^2} dx = \frac{1}{2a} \ln \left| \frac{x-a}{x+a} \right| + c$$

$$\int \frac{1}{\sqrt{x^2 \pm a^2}} dx = \ln |x + \sqrt{x^2 \pm a^2}| + c$$

$$\int f \cdot f' dx = \frac{1}{2} f^2 + c$$

$$\int \cos f \cdot f' dx = \sin(f) + c$$

$$\int \frac{f'}{\sqrt{f}} dx = 2\sqrt{f} + c$$

$$\int u \cdot v' dx = u \cdot v - \int u' \cdot v dx$$

שאלות:

(1) X הינו משתנה רציף עם פונקציית צפיפות כמוצג בשרטוט:

א. מצאו את ערכו של c .

ב. בנו את פונקציית ההתפלגות המצטברת.

ג. חשבו את ההסתברויות הבאות:

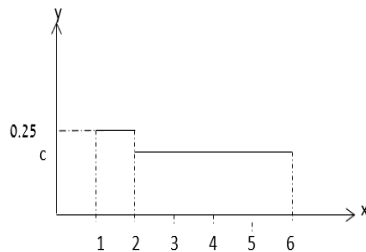
i. $P(x < 4)$

ii. $P(x > 1.5)$

iii. $P(1.5 < x < 5)$

iv. $P(5 < x < 10)$

v. מצאו את החציון של המשתנה.



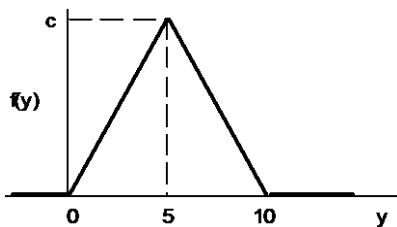
(2) נתון משתנה מקרי רציף A שפונקציית הצפיפות שלו היא: $f(x) = \begin{cases} cx & 0 \leq x \leq b \\ 0 & \text{else} \end{cases}$

וידוע ש- $P(0 < X < 1) = \frac{1}{4}$

א. מצאו במפורש את פונקציית הצפיפות של X .

ב. מצאו את החציון של X .

ג. מה הסיכוי ש- X קטן מ-0.5?



(3) נתונה פונקציית צפיפות של משתנה מקרי Y :

א. מצאו את c .

ב. מצאו את פונקציית ההתפלגות המצטברת של Y .

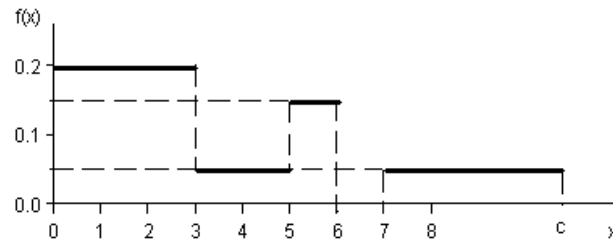
ג. חשבו את ההסתברויות הבאות:

$P(Y > 4)$, $P(7.5 \leq Y \leq 15.5)$, $P(Y \leq 3.0)$, $P(Y = 7.0)$

ד. מצאו את העשירון התחתון: $y_{0.1}$, הרבעון התחתון: $y_{0.25}$ והחציון של Y .

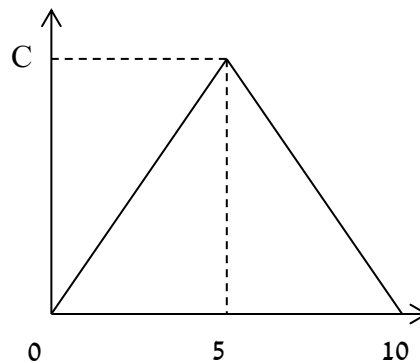
הסיקו מהו העשירון עליון: $y_{0.9}$.

(4) נתונה פונקציית צפיפות של משתנה מקרי X :



- א. מצאו ערך c שעבורו תתקבל פונקציית צפיפות.
- ב. מצאו את פונקציית ההתפלגות המצטברת.
- ג. חשבו את ההסתברויות הבאות:
 $P(1.0 < X \leq 5.0)$, $P(X \geq -2.0)$, $P(X \geq 4)$.

(5) נתונה פונקציית הצפיפות הבאה:



- א. מה ערכו של c ?
- ב. מצאו אינטרוול (תחום) סימטרי סביב הערך 5, שהסיכוי ליפול בו הינו 0.5.

(6) נתונה פונקציית צפיפות: $f(X) = \frac{2}{x}$, המוגדרת מ-1 עד K .

- א. מצאו את ערכו של K .
- ב. בנו את פונקציית ההתפלגות המצטברת.
- ג. חשבו את הסיכוי ש- X לפחות 1.5.
- ד. מצאו את העשירון התחתון של ההתפלגות.
- ה. מה התוחלת של X ?

(7) נתונה פונקציית צפיפות הבאה: $f(X) = AX^2(10 - X)$, $0 < X < 10$.

A הינו קבוע חיובי.

א. מצאו את A.

ב. חשב את: $P(x > 5 | x > 2)$.

ג. מה התוחלת ומהי השונות של X?

(8) פונקציית הצפיפות של משתנה מקרי רציף X:

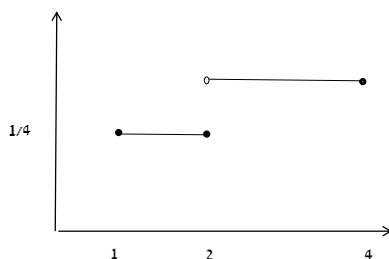
$$f(x) = 0.5 \cdot e^{2x}, \quad -\infty \leq x \leq \ln(c)$$

א. מצאו את ערכו של c.

ב. מצאו את פונקציית ההתפלגות המצטברת של ההתפלגות.

ג. חשב: $P(X > 0)$.

ד. מהו הרבעון העליון של ההתפלגות?



(9) נתונה פונקציית הצפיפות הבאה של משתנה מקרי X:

א. רשמו את נוסחת פונקציית הצפיפות.

ב. בנו את פונקציית ההתפלגות המצטברת.

ג. מצאו את החציון של ההתפלגות.

ד. חשבו את התוחלת והשונות של המשתנה.

ה. חשבו את: $E(X^3)$.

(10) במפעל מייצרים מוצר A. זמן תהליך הייצור של המוצר בשעות הוא בעל

$$f(x) = 6x(1-x), \quad 0 \leq x \leq 1$$

א. מה ההסתברות שזמן הייצור של מוצר A אקראי יהיה קטן מ-20 דקות?

ב. מה ההסתברות שזמן הייצור של מוצר A אקראי יהיה בדיוק חצי שעה?

ג. נבחרו חמישה מוצרים אקראיים מסוג A. מה תוחלת מספר המוצרים

שזמן הייצור שלהם יהיה גדול מ-20 דקות?

(11) זמן ההמתנה בדקות של לקוח בתור למכולת השכונתית מתפלג עם פונקציית

$$F(t) = 1 - e^{-0.2t}$$

א. שרטטו את פונקציית ההתפלגות המצטברת.

ב. מה הסיכוי שזמן ההמתנה יהיה לפחות רבע שעה?

ג. אם חיכיתי בתור כבר 10 דקות מה ההסתברות שאאלץ לחכות בסך הכול

פחות מרבע שעה?

ד. מהו הזמן ש-90% מהלקוחות מחכים מתחתיו?

(12) פונקציית הצפיפות של משתנה מקרי נתונה על ידי הנוסחה הבאה :

$$f(x) = \begin{cases} 0 & x < 4 \\ bx - 4b & 4 \leq x \leq 5 \\ b & 5 < x \leq 6 \\ 0 & x > 6 \end{cases}$$

- א. מצאו את b .
- ב. חשבו את התוחלת של X .
- ג. y הוא משתנה אינדיקטור המקבל את הערך 1 אם X קטן מ-5. מהי השונות של y ?

(13) נתונה פונקציית הצפיפות הבאה :

$$f(x) = \begin{cases} \frac{x^2}{4} & 1 \leq x \leq 2 \\ kx & 2 < x \leq 3 \end{cases}$$

- א. מצאו את ערכו של k .
- ב. מצאו את פונקציית ההתפלגות המצטברת.
- ג. חשבו $P(x > 2.5)$.

(14) להלן משתנה מקרי בעל פונקציית צפיפות הבאה : $f(x) = \frac{1}{b-a}$, $a \leq x \leq b$

- א. מצאו את פונקציית ההתפלגות המצטברת.
- ב. חשב את התוחלת והשונות של ההתפלגות.
- ג. מצאו את התוחלת של $\frac{1}{X}$.

תשובות סופיות:

$$F(t) = \begin{cases} 0 & t < 1 \\ (t-1)0.25 & 1 \leq t \leq 2 \\ 0.25 + (t-2) \cdot \frac{3}{16} & 2 < t \leq 6 \\ 1 & t > 6 \end{cases} \quad \text{ב.} \quad \text{א. } \frac{3}{16} \quad \text{ג. } \frac{5}{8} \quad \text{ד. } \frac{1}{3}$$

$$\text{א. } b=2, c=0.5 \quad \text{ב. } 1.41 \quad \text{ג. } 0.0625 \quad \text{ד. } \frac{1}{3}$$

$$F(t) = \begin{cases} 0 & t < 0 \\ 0.02t^2 & 0 \leq t \leq 5 \\ 1 - 0.02(t-10)^2 & 5 < t \leq 10 \\ 1 & t > 10 \end{cases} \quad \text{ב.} \quad \text{א. } 0.2 \quad \text{ג. } 0, 0.18, 0.125, 0.32$$

ד. עשירון תחתון: 2.24, רבעון תחתון: 3.54, החציון: 5, עשירון עליון: 7.76.

$$F(t) = \begin{cases} 0 & t < 0 \\ 0.2t & 0 < t \leq 3 \\ 0.6 + (t-3) \cdot 0.05 & 3 < t \leq 5 \\ 0.7 + (t-5) \cdot 0.15 & 5 < t \leq 6 \\ 0.85 & 6 < t \leq 7 \\ 0.85 + (t-7) \cdot 0.05 & 7 < t \leq 10 \\ 1 & t > 10 \end{cases} \quad \text{ב.} \quad \text{א. } 0.10 \quad \text{ג. } 0.5$$

$$\text{א. } c=0.2 \quad \text{ב. } 5 \pm 1.46$$

$$F(t) = \begin{cases} 0 & t < 1 \\ 2 \cdot \ln t & 1 \leq t \leq e^{\frac{1}{2}} \\ 1 & t > e^{\frac{1}{2}} \end{cases} \quad \text{ב.} \quad \text{א. } e^{\frac{1}{2}} \quad \text{ג. } 0.189$$

$$\text{ד. } 1.051 \quad \text{ה. } 1.297$$

$$\text{א. } 0.0012 \quad \text{ב. } 0.7067 \quad \text{ג. תוחלת: 6, שונות: 4}$$

$$(8) \quad \text{א. 2.} \quad \text{ב.} \quad F(t) = \begin{cases} \frac{1}{4}e^{2t} & t \leq \ln(2) \\ 1 & t > \ln(2) \end{cases} \quad \text{ג. 0.75} \quad \text{ד. 0.549}$$

$$(9) \quad \text{א.} \quad F(t) = \begin{cases} \frac{1}{4} & 1 \leq x \leq 2 \\ \frac{3}{8} & 2 < x \leq 4 \\ 0 & \text{אחרת} \end{cases} \quad \text{ב.} \quad F(t) = \begin{cases} 0 & t < 1 \\ (t-1)0.25 & 1 \leq t \leq 2 \\ 0.25 + (t-2) \cdot \frac{3}{8} & 2 < t \leq 4 \\ 1 & t > 4 \end{cases}$$

$$\text{ג. } 2\frac{2}{3} \quad \text{ד. תוחלת: } 2.625, \text{ שונות: } 0.6927 \quad \text{ה. } 23.4375$$

$$(10) \quad \text{א. } \frac{7}{27} \quad \text{ב. } 0 \quad \text{ג. } 3.704$$

$$(11) \quad \text{א. עיין סרטוט בוידאו} \quad \text{ב. } 0.0498 \quad \text{ג. } 0.6321 \quad \text{ד. } 11.51$$

$$(12) \quad \text{א. } \frac{2}{3} \quad \text{ב. } 5.22 \quad \text{ג. } \frac{2}{9}$$

$$(13) \quad \text{א. } \frac{1}{6} \quad \text{ב.} \quad F(t) = \begin{cases} 0 & t < 1 \\ \frac{t^3-1}{12} & 1 \leq t \leq 2 \\ \frac{7}{12} + \frac{t^2-4}{12} & 2 < t \leq 3 \\ 1 & t > 3 \end{cases} \quad \text{ג. } 0.229$$

$$(14) \quad \text{א.} \quad F(t) = \begin{cases} 0 & t < a \\ \frac{(t-b)}{b-a} & a \leq t \leq b \\ 1 & t > b \end{cases} \quad \text{ב. תוחלת: } E(X) = \frac{a+b}{2}, \text{ שונות: } V(X) = \frac{(b-a)^2}{12}$$

$$\text{ג. } \frac{\ln\left(\frac{b}{a}\right)}{b-a}$$

סטטיסטיקה והסתברות

פרק 21 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

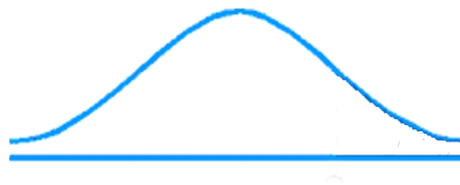
תוכן העניינים

1. כללי 81

התפלגויות רציפות מיוחדות – התפלגות נורמלית:

רקע:

התפלגות נורמלית הינה התפלגות של משתנה רציף. ישנם משתנים רציפים מסוימים שנהוג להתייחס אליהם כנורמליים כגון: זמן ייצור, משקל תינוק ביום היוולדו ועוד. פונקציית הצפיפות של ההתפלגות הנורמלית נראית כמו פעמון:



לעקומה זו קוראים גם עקומת גאוס ועקומה אחת נבדלת מהשנייה באמצעות הממוצע וסטיית התקן שלה.

אלה הם הפרמטרים שמאפיינים את ההתפלגות: $X \sim N(\mu, \sigma^2)$.

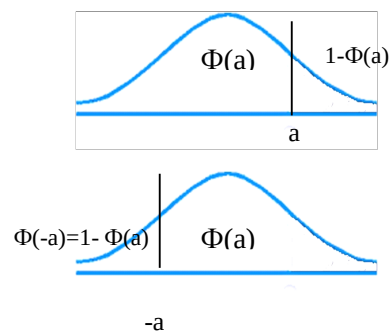
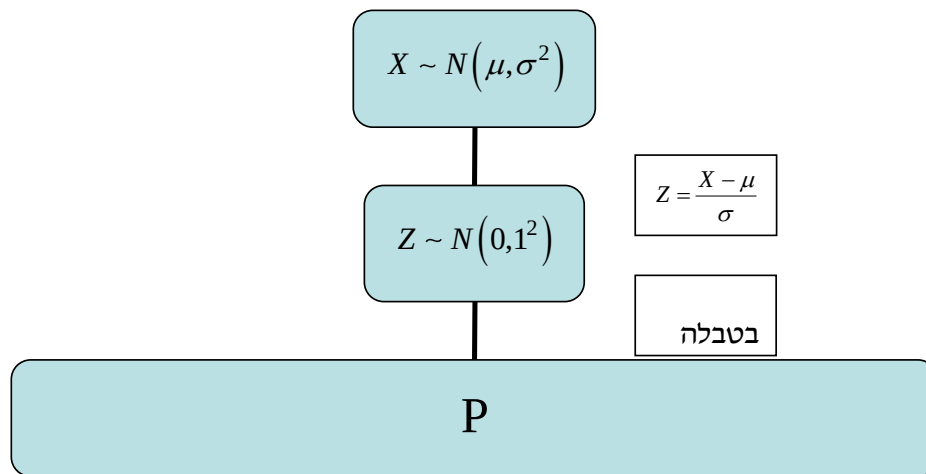
$$\text{נוסחת פונקציית הצפיפות: } f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלוונטיים שמתחת לעקומה. כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון. התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת, והיא תסומן באות Z : $Z \sim N(0, 1^2)$.

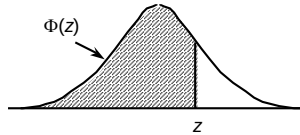
$$\text{תהליך התקנון מבוצע על ידי הנוסחה הבאה: } Z = \frac{X - \mu}{\sigma}$$

אחרי תקנון מקבלים ערך הנקרא ציון תקן. ציון התקן משמעו בכמה סטיות תקן הערך סוטה מהממוצע.

לאחר חישוב ציון התקן של ערך מסוים נעזרים בטבלה של ההתפלגות הנורמלית הסטנדרטית לחישוב השטח הרצוי, ובאופן כללי נתאר את הסכמה הבאה:



טבלת ההתפלגות המצטברת הנורמלית סטנדרטית – ערכי $\Phi(z)$:



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

z	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.891	4.417
$\Phi(z)$	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995

דוגמה (הפתרון בהקלטה):

משקל חפיסות שוקולד המיוצרות בחברה מתפלג נורמלית עם ממוצע 100 גרם בסטיית תקן של 8 גרם.

- (1) מה אחוז חפיסות השוקולד ששוקלות מתחת ל-110 גרם?
- (2) מה אחוז חפיסות השוקולד ששוקלות מעל 110 גרם?
- (3) מה אחוז חפיסות השוקולד ששוקלות מתחת ל-92 גרם?
- (4) מהו המשקל ש-90% מהחפיסות בקו הייצור שוקלים פחות מהם?

שאלות:

(1) הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטית תקן של 10 ס"מ.

- א. מה אחוז האנשים שגובהם מתחת ל-182.4 ס"מ?
- ב. מה אחוז האנשים שגובהם מעל 190 ס"מ?
- ג. מה אחוז האנשים שגובהם בדיוק 173.6 ס"מ?
- ד. מה אחוז האנשים שגובהם מתחת ל-170 ס"מ?
- ה. מה אחוז האנשים שגובהם לכל היותר 170 ס"מ?

(2) נתון שהזמן שלוקח לתרופה מסוימת להשפיע מתפלג נורמלית עם ממוצע של 30 דקות ושונות של 9 דקות רבועות.

- א. מהי פרופורציית המקרים בהן התרופה תעזור אחרי יותר משעה?
- ב. מה אחוז מהמקרים שבהן התרופה תעזור בין 35 ל-37 דקות?
- ג. מה הסיכוי שהתרופה תעזור בדיוק תוך 36 דקות?
- ד. מה שיעור המקרים שבהן ההשפעה של התרופה תסטה מ-30 דקות בפחות מ-3 דקות?

(3) המשקל של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 60 ק"ג וסטיית תקן של 8 ק"ג.

- א. מה אחוז האנשים שמשקלם נמוך מ-55 ק"ג?
- ב. מהי פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
- ג. מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל-70 ק"ג?
- ד. לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ-4 ק"ג?
- ה. מה הסיכוי שאדם אקראי ישקול מתחת ל-140 ק"ג?

(4) משקל תינוקות ביום היוולדם מתפלג נורמלית עם ממוצע של 3300 גרם וסטית תקן 400 גרם.

- א. מצאו את העשירון העליון.
- ב. מצאו את האחוזון ה-95.
- ג. מצאו את העשירון התחתון.

- (5) ציוני מבחן אינטליגנציה מתפלגים נורמלית עם ממוצע 100 ושונות 225.
- א. מה העשירון העליון של הציונים במבחן האינטליגנציה?
 - ב. מה העשירון התחתון של ההתפלגות?
 - ג. מהו הציון ש-20% מהנבחנים מקבלים מעליו?
 - ד. מהו האחוזון ה-20?
 - ה. מהו הציון ש-5% מהנבחנים מקבלים מתחתיו?
- (6) נפח משקה בבקבוק מתפלג נורמלית עם סטיית תקן של 20 מ"ל, ונתון ש-33% מהבקבוקים בעלי נפח שעולה על 508.8 מ"ל.
- א. מה ממוצע נפח משקה בבקבוק?
 - ב. 5% מהבקבוקים המיוצרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאיזה נפח שולחים בקבוק לבדיקה?
 - ג. 1% מהבקבוקים עם הנפח הקטן ביותר נתרמים לצדקה, מהו הנפח המקסימלי לצדקה?
- (7) אורך חיים של מכשיר מתפלג נורמלית. ידוע שמחצית מהמכשירים חיים פחות מ-500 שעות, כמו כן ידוע ש-67% מהמכשירים חיים פחות מ-544 שעות.
- א. מהו ממוצע אורך חיי מכשיר?
 - ב. מהי סטית בתקן של אורך חיי מכשיר?
 - ג. מה הסיכוי שמכשיר אקראי יחיה פחות מ-460 שעות?
 - ד. מהו המאיון העליון של אורך חיי מכשיר?
 - ה. 1% מהמכשירים בעלי אורך החיים הקצר ביותר נשלח למעבדה לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשליחת מכשיר למעבדה?
- (8) להלן שלוש התפלגויות נורמליות של שלוש קבוצות שונות ששורטטו באותה מערכת צירים. ההתפלגויות מוספרו כדי להבדיל ביניהן.
- א. לאיזו התפלגות הממוצע הגבוה ביותר?
 - ב. במה מבין המדדים הבאים התפלגות 1 ו-2 זהות?
 - i. בעשירון העליון.
 - ii. בממוצע.
 - iii. בשונות.
 - ג. לאיזו התפלגות סטיית התקן הקטנה ביותר?
 - i. 1.
 - ii. 2.
 - iii. 3.
 - iv. אין לדעת.



- (9)** הזמן שלוקח לאדם להגיע לעבודתו מתפלג נורמלית עם ממוצע של 40 דקות וסטית תקן של 5 דקות.
- א. מה ההסתברות שמשך הנסיעה של האדם לעבודתו יהיה לפחות שלושת רבעי השעה?
- ב. אדם יצא לעבודתו בשעה 08:10 מביתו. הוא צריך להגיע לעבודתו בשעה 09:00. מה הסיכוי שיאחר לעבודתו?
- ג. אם ידוע שזמן נסיעתו לעבודה היה יותר משלושת רבעי השעה. מה ההסתברות שזמן הנסיעה הכולל יהיה פחות מ-50 דקות?
- ד. מה הסיכוי שבשבוע (חמישה ימי עבודה) בדיוק פעם אחת יהיה זמן הנסיעה לפחות שלושת רבעי השעה?
- (10)** ההוצאה החודשית לבית אב בעיר "טרירה" מתפלגת נורמלית עם ממוצע של 2000 דולר וסטית תקן של 300 דולר. בחרו באקראי 5 בתי אב. ההסתברות שלפחות אחד מהם מוציא בחודש מעל ל- T דולר היא 0.98976.
- א. מה ערכו של T ?
- ב. מה הסיכוי שההוצאה החודשית של בית אב בעיר תהיה לפחות סטיית תקן אחת מעל T ?
- ג. מסתבר שנפלה טעות בנתונים, ויש להוסיף 100 דולר להוצאות החודשית של כל בתי האב בעיר. לאור זאת, מה ההסתברות שההוצאה החודשית של בית אב נמוכה מ-1800 דולר?
- (11)** אורך שיר אקראי המשודר ברדיו מתפלג נורמלית עם תוחלת של 3.5 דקות וסטית תקן של שלושים שניות.
- א. מה ההסתברות שאורך של שיר אקראי המנוגן ברדיו יהיה בין 3 ל-2.5 דקות?
- ב. מהו הטווח הבין רבעוני של אורך שיר המשודר ברדיו?
- ג. ביום מסוים מנוגנים 200 שירים ברדיו. כמה שירים מתוכם תצפה שיהיו באורך הנמוך מ-3.5 דקות?
- ד. בשעה מסוימת שודרו 8 שירים. מה ההסתברות שרבע מהם בדיוק היו ארוכים מ-4 דקות והיתר לא?

תשובות סופיות:

1	א. 89.25%	ב. 2.28%	ג. 0.	ד. 50%	ה. 50%
2	א. 0%	ב. 3.76%	ג. 0%	ד. 68.26%	
3	א. 26.43%	ב. 89.44%	ג. 39.44%	ד. 0.383	
	ה. 100%				
4	א. 3812.8	ב. 3958	ג. 2787.2		
5	א. 119.2	ב. 80.8	ג. 112.6	ד. 87.4	
6	א. 500	ב. 532.9	ג. 453.48		
7	א. 500	ב. 100	ג. 0.3446	ד. 733	
	ה. 267				
8	א. 3	ב. בממוצע.	ג. 1		
9	א. 0.1587	ב. 0.0228	ג. 0.8563	ד. 0.3975	
10	א. 1925	ב. 0.2266	ג. 0.1587		
11	א. 0.1359	ב. 0.675	ג. 100	ד. 0.25	

סטטיסטיקה והסתברות

פרק 22 - הסקה סטטיסטית - הקדמה

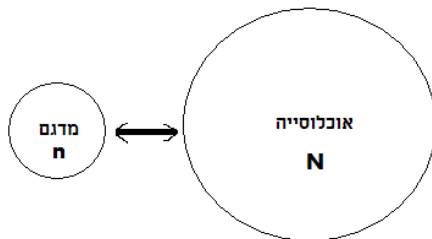
תוכן העניינים

1. כללי 89

הסקה סטטיסטית – הקדמה:

רקע:

אוכלוסייה:



קבוצה שאליה מפנים שאלה מחקרית. למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסוכרת מתעניינת באוכלוסיית חולי הסוכרת בעולם.

מדגם:

חלק מתוך האוכלוסייה. למשל, אם נדגום באקראי 10 אנשים מתוך חולי הסוכרת אז זהו מדגם מתוך אוכלוסיית חולי הסוכרת.

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, אנו מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה. הדגימה בקורס תהיה דגימה מקרית - הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכוי להיכלל במדגם.

סטטיסטי:

גודל המחושב על המדגם.

פרמטר:

גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים:

פרמטר (אוכלוסייה)	סטטיסטי (מדגם)	ממוצע
μ	$\bar{X}$	
P	$\hat{p}$	פרופורציה (שכיחות יחסית)

פרמטר הוא גודל קבוע גם אם אנו לא יודעים אותו סטטיסטי הוא משתנה ממדגם למדגם ולכן יש לו התפלגות הנקראת התפלגות הדגימה.

דוגמה (פתרון בהקלטה):

25% מאזרחי המדינה תומכים בהצעת החוק של חבר כנסת מסוים. הוחלט לדגום 200 אזרחים ומתוכם לבדוק מהו אחוז התומכים בהצעת החוק.

- מי האוכלוסייה?
- מה המשתנה?
- מה הפרמטרים?
- מהו גודל המדגם?
- מהו הסטטיסטי שמתכננים להוציא מהמדגם?
- האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

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.
- מי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
- (2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה ב"שוב "העוגן". נגדיר את X להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.
- מי היא האוכלוסייה ומהו המשתנה הנחקר?
 - מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

מספר המשפחות	מספר מקלטים
50	0
250	1
350	2
300	3
50	4
סך הכול $N = 1000$	

- (3) נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.
- מהי האוכלוסייה?
 - מה המשתנה באוכלוסייה?
 - מהם הפרמטרים?
 - מהו הסטטיסטי?

תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון. ג. ממוצע: 78, סטיית תקן: 15. ד. 2.
- (2) א. האוכלוסייה: 1000 משפחות ב"שוב העוגן, המשתנה הנחקר: מס' מקלטים. ב. $\bar{X}$ = ממוצע מדגם.
- (3) א. השכירים במדינה. ב. השכלה: אקדמאי, לא אקדמאי. ג. שיעור ההצלחות באוכלוסייה: 0.2. ג. מס' האקדמאים במדגם.

סטטיסטיקה והסתברות

פרק 23 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

- 1. התפלגות פרופורציית ההצלחות במדגם 92
- 2. חוק המספרים הגדולים 96

התפלגות פרופורציית ההצלחות במדגם:

רקע:

בפרק זה נדון בהתפלגות הדגימה של פרופורציית המדגם.
 Y - מספר ההצלחות במדגם (למשל, מספר המובטלים במדגם).

$$\hat{p} = \frac{Y}{n} \text{ - פרופורציית ההצלחות במדגם.}$$

למשל, שיעור המובטלים במדגם - $n = 200$:

מספר המובטלים : $Y = 20$.

$$\hat{p} = \frac{20}{200} = 0.1 \text{ : פרופורציית המובטלים במדגם}$$

נסמן ב- p את שיעור ההצלחה באוכלוסייה וב- q את שיעור הכישלונות באוכלוסייה.
 נבצע מדגם מקרי (הנחה שהתצפיות בלתי תלויות זו בזו) ונתבונן בהתפלגות של פרופורציית המדגם.

התוחלת, השונות וסטיית התקן של פרופורציית המדגם:

$$E(\hat{p}) = p, \quad V(\hat{p}) = \frac{pq}{n}$$

משפט הגבול המרכזי עבור הפרופורציה המדגמית:

$$\text{אם: } np \geq 5 \text{ \& } nq \geq 5, \text{ אזי: } \hat{p} \sim N\left(p, \frac{pq}{n}\right) \quad Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

הערות:

- התנאים לקרוב הנורמאלי הם נזילים, כלומר משתנים ממרצה אחד לשני.
 התנאי שהצגתי כאן הוא הפופולרי ביותר:
 1. $n \cdot p \geq 5$
 2. $n \cdot (1 - p) \geq 5$
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
 1. $n \cdot p \geq 10$
 2. $n \cdot (1 - p) \geq 10$
- וישנם מרצים המשתמשים בתנאי: $(n \geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור לנורמלית.

- כיוון שפרופורציה אינה חייבת להיות מספר שלם בהכרח לא נהוג לבצע כאן תיקון רציפות.

דוגמה (פתרון בהקלטה):

לפי נתוני משרד החינוך בעיר ירושלים ל-60% מתלמידי התיכון זכאים לתעודת בגרות. נדגמו 200 תלמידי תיכון.

- מה ההסתברות שהשכיחות היחסית ($\hat{p}$) של הזכאים לבגרות במדגם תעלה על 60%?
- מה ההסתברות שפרופורציית הזכאים לבגרות במדגם תעלה על 70%?

שאלות:

- (1) במדינה מסוימת 10% מכלל האוכלוסייה הינם מובטלים. נדגמו באקראי 140 אנשים מהמדינה.
 - א. מה התוחלת ומהי השונות של פרופורציות המובטלים שנדגמו?
 - ב. מה ההסתברות שבמדגם לפחות 10% יהיו מובטלים?
 - ג. מה ההסתברות שלכל היותר 9% מהמדגם יהיו מובטלים?
- (2) נניח כי 30% מהאוכלוסייה תומכים בהצעת חוק מסוימת. אם נדגום מהאוכלוסייה 200 איש. חשבו את ההסתברויות הבאות:
 - א. לפחות 35% יתמכו בהצעת החוק במדגם.
 - ב. לכל היותר 25% יתמכו בהצעת החוק במדגם.
 - ג. יותר מ-27% יתמכו בהצעת החוק במדגם.
- (3) לפי נתוני משרד התקשורת 40% מהאוכלוסייה מחזיקים בטלפון נייד מסוג "סמארטפון". נדגמו 400 אנשים מהאוכלוסייה.
 - א. מה ההסתברות שבמדגם לכל היותר ל-40% יש סמארטפון?
 - ב. מה ההסתברות שבמדגם לרוב יש סמארטפון?
 - ג. מה ההסתברות שפרופורציית בעלי הסמארטפון במדגם תסטה מהפרופורציה באוכלוסייה בלא יותר מ-4%?
 - ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם הינו מגדילים את גודל המדגם?
- (4) נתון כי 80% מבתי האב מחוברים לאינטרנט. נדגמו 400 בתי אב אקראיים.
 - א. מה ההסתברות שלפחות 340 מהם מחוברים לאינטרנט?
 - ב. מה ההסתברות שפרופורציית המחוברים לאינטרנט במדגם תסטה מהפרופורציה האמיתית ביותר מ-4%?
 - ג. כמה בתי אב יש לדגום כדי שהסטייה בין הפרופורציה המדגמית לפרופורציה האמיתית לא תעלה על 3% בהסתברות של 90%?
 - ד. מהו העשירון התחתון של התפלגות פרופורציית המדגם?
- (5) נתון שציוני פסיכומטרי מתפלגים נורמלית עם תוחלת 500 וסטיית תקן 100. ב"מועדון ה-700" נכללים נבחנים שמקבלים ציון מעל 700 בפסיכומטרי. מה הסיכוי שבמועד בו נבחנו 2000 נבחנים אקראיים יהיו לפחות 3% המשתייכים למועדון?

6 נתון ש- $X \sim B(n, p)$, ונגדיר את המשתנה הבא: $\hat{P} = \frac{X}{n}$.

א. הוכיחו ש: $E(\hat{P}) = p$, $V(\hat{P}) = \frac{p(1-p)}{n}$.

ב. מה p המביא את $V(\hat{P})$ להיות במקסימום?

תשובות סופיות:

- 1 א. התוחלת: 0.1, השונות: 0.00064 ב. 0.5 ג. 0.3446
- 2 א. 0.0618 ב. 0.0618 ג. 0.8238
- 3 א. 0.5 ב. 0 ג. 0.8968 ד. גדלה.
- 4 א. 0.0062 ב. 0.0456 ג. 0.481 ד. 0.77436
- 5 0.0154
- 6 א. שאלת הוכחה. ב. 0.5

חוק המספרים הגדולים:

רקע:

חוק המספרים הגדולים מתייחס להשפעת הגדלת גודל המדגם על הסיכוי של פרופורציית המדגם (או ממוצע המדגם) להיות קרובה מהפרופורציה האמיתית (או מהממוצע האימתי).

החוק לגבי פרופורציה:

נניח שמבצעים מדגם מקרי מתוך אוכלוסייה אינסופית בה פרופורציית ההצלחות היא p , ככל שהמדגם גדול יותר: כך הסיכוי שפרופורציית המדגם ($\hat{p}$) תהיה בקרבת הפרופורציה באוכלוסייה (P) גבוה יותר.

ולכן הסיכוי לקבל ערך חריג הרחוק מהפרופורציה של האוכלוסייה קטן יותר. בכתיבה מתמטית רושמים את חוק המספרים הגדולים לגבי הפרופורציה באופן

$$\text{הבא: } \lim_{n \rightarrow \infty} \hat{P}_n = P$$

בספרות המקצועית קוראים לחוק הזה החוק החזק של המספרים הגדולים. את החוק החלש של המספרים הגדולים רושמים באופן הבא: $P(|\hat{P} - P| < \varepsilon) \rightarrow 1$ כ- $n \rightarrow \infty$.

הערה: ככל שהמדגם גדל הסיכוי שפרופורציית המדגם תהיה בדיוק הפרופורציה האמיתית הולך וקטן.

החוק לגבי ממוצע: נניח שמבצעים מדגם מקרי מתוך התפלגות שבה התוחלת μ והשונות סופית. ככל שהמדגם גדול יותר, כך הסיכוי שממוצע המדגם ($\bar{X}$) יהיה בקרבת הממוצע באוכלוסייה (μ) גבוה יותר. לכן, הסיכוי לקבל ערך חריג הרחוק מהממוצע של האוכלוסייה קטן יותר. בכתיבה מתמטית רושמים את חוק המספרים הגדולים לגבי הממוצע באופן הבא: $\lim_{n \rightarrow \infty} \bar{X}_n = \mu$.

בספרות המקצועית קוראים לחוק הזה החוק החזק של המספרים הגדולים. את החוק החלש של המספרים הגדולים רושמים באופן הבא: $P(|\bar{X} - \mu| < \varepsilon) \rightarrow 1$ כ- $n \rightarrow \infty$.

דוגמה (פתרון בהקלטה):

באוכלוסייה מסוימת 20% מהאוכלוסייה מובטלת. איזה סיכוי יותר גבוה? במדגם בגודל 100 פרופורציית המובטלים תסטה מהפרופורציה האמיתית בלא יותר מ-4%. במדגם בגודל 200 פרופורציית המובטלים תסטה מהפרופורציה האמיתית בלא יותר מ-4%. הסבירו.

שאלות:

- (1) באוכלוסייה מסוימת 20% מהאוכלוסייה מובטלת. איזה סיכוי יותר גבוה?
א. אחד מתוך מדגם של חמישה יהיה מובטל.
ב. שניים מתוך עשרה יהיו מובטלים. הסבירו וחשבו.
- (2) באוכלוסייה מסוימת 20% מהאוכלוסייה מובטלת. איזה סיכוי יותר גבוה?
א. לפחות 3 מתוך 10 יהיו מובטלים
ב. לפחות 30 מתוך 100 יהיו מובטלים. הסבירו.
- (3) גובה של אוכלוסייה מסוימת מתפלג נורמלית עם ממוצע 170 ס"מ וסטיית תקן של 10 ס"מ. דוגמים 4 אנשים באקראי.
א. מה ההסתברות שהגובה הממוצע שלהם יהיה מעל 176 ס"מ.
ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם הינו מגדילים את גודל המדגם? נמקו.
- (4) ידוע שבהצעת חוק מסוימת תומכים 60% מהציבור. נדגמו באקראי 10 אנשים.
א. מה ההסתברות שבדיוק 60% מהמדגם תומכים בהצעת החוק?
ב. כיצד התשובה הייתה משתנה אם היו דוגמים 20 אנשים?
- (5) שני חוקרים ביצעו מדגם מאותה אוכלוסייה. חוקר א דגם 20 תצפיות והשני דגם 40 תצפיות וכל אחד מהם חישב את ממוצע המדגם: $\bar{X}_{20}$ ו- $\bar{X}_{40}$. ידוע שהתפלגות היא נורמלית ושהתוחלת באוכלוסייה הינה 500. בסעיפים הבאים נמקו אילו הסתברויות מהזוגות המוצגים גדולה יותר או שהן שוות ונמקו.
א. $P(\bar{X}_{20} > 500)$ או $P(\bar{X}_{40} > 500)$.
ב. $P(480 < \bar{X}_{20} < 520)$ או $P(480 < \bar{X}_{40} < 520)$.
ג. $P(\bar{X}_{20} < 490)$ או $P(\bar{X}_{40} < 490)$.
- (6) נתון ש: $X \sim G(P=0.1)$. מבצעים מדגם אקראי בגודל n מההתפלגות הזו ומחשבים את ממוצע המדגם: $\bar{X}_n$. הוכיחו: $\lim_{n \rightarrow \infty} \bar{X}_n = 10$.

תשובות סופיות:

- (1) אחד מתוך מדגם של חמישה יהיה מובטל.
- (2) לפחות 3 מתוך 10 יהיו מובטלים.
- (3) א. 0.1151. ב. קטנה.
- (4) א. 0.2508. ב. קטן.
- (5) א. $P(\bar{X}_{40} > 500) = P(\bar{X}_{20} > 500)$.
- ב. $P(480 < \bar{X}_{40} < 520) > P(480 < \bar{X}_{20} < 520)$.
- ג. $P(\bar{X}_{40} < 490) < P(\bar{X}_{20} < 490)$.
- (6) שאלת הוכחה.

סטטיסטיקה והסתברות

פרק 24 - מושגי יסוד באמידה

תוכן העניינים

1. כללי 99

מושגי יסוד באמידה:

רקע:

כזכור מהמפגש הקודם, פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסוימת. כמו ממוצע הגבהים בקרב מתגייסים לצה"ל - μ . כמו פרופורציית התומכים בממשלה בקרב אזרחי המדינה - p . בדרך כלל הפרמטרים הם גדלים שאינם ידועים באמת, ולכן מבצעים מדגמים במטרה לאמוד אותם. אין אפשרות לחשב אותם הניסיון הוא בלהעריך כמה הם שווים ככל שניתן.

- נסמן באופן כללי פרמטר באות θ ואומד ב- $\hat{\theta}$. הוא סטטיסטי המחושב על המדגם ובאמצעותו נאמוד את θ .
- שגיאת אמידה: $|\hat{\theta} - \theta|$ - ההפרש בין האומד לאמת (הפרמטר).

דוגמה (פתרון בהקלטה):

בכנסת ה-19 קיבלה מפלגת העבודה 15 מנדטים. בערוץ 10 ברגע סגירת הקלפיות העריכו את מספר המנדטים של המפלגה להיות 17 מנדטים וזאת על סמך תוצאות מדגם של הערוץ.

- א. מה הפרמטר בדוגמה זו?
- ב. מהי טעות האמידה של ערוץ 10?
- $\hat{\theta}$ יהיה אומד חסר הטיה ל- θ אם התוחלת של $\hat{\theta}$ תהיה שווה ל- θ : $E(\hat{\theta}) = \theta$.
- טעות התקן של אומד היא סטיית התקן שלו, כלומר: $\sigma(\hat{\theta}) = S.E$

פרמטרים מרכזיים והאומדים שלהם:

ממוצע האוכלוסייה μ :

האומד הנקודתי שלו יהיה: ממוצע המדגם: $\bar{x} = \frac{\sum x}{n}$.

$E(\bar{x}) = \mu$ לכן $\bar{x}$ הינו אומד חסר הטיה ל- μ . כמו כן, טעות תקן: $\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = SE$.

פרופורציה באוכלוסייה p :

האומד הנקודתי שלו יהיה: פרופורציה במדגם: $\hat{p} = \frac{y}{n}$.

$E(\hat{p}) = p$, לכן $\hat{p}$ הינו אומד חסר הטיה ל- p . כמו כן טעות התקן: $\sigma(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}}$.

שונות האוכלוסייה σ^2 :

האומד הנקודתי שלו יהיה: $S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$.

$E(S^2) = \sigma^2$ ולכן S^2 הינו אומד חסר הטיה ל- σ^2 : $S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$.

הערה: אומד הוא הנוסחה הכללית לאמידת הפרמטר ואומדן הוא הערך הספציפי שהתקבל במדגם מסוים.

דוגמה (פתרון בהקלטה):

נדגמו 10 משפחות בתל אביב ונבדק עבור כל משפחה מספר הילדים שלה.
להלן התוצאות שהתקבלו: 2, 1, 3, 2, 1, 4, 5, 2, 1, 3.
אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים:

1. ממוצע מספר הילדים למשפחה בתל אביב.
2. שונות מספר הילדים למשפחה בתל אביב.
3. פרופורציית המשפחות בנות שני ילדים.

שאלות:

- (1) מתוך 500 טירונים, נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - מהי טעות התקן של האומדן כשהמדגם בגודל 500?
 - מהו האומדן לפרמטר?
 - מהי טעות האמידה?
- (2) לפי נתוני היצרן, מקרר צורך בממוצע 2400 וואט לשעה עם סטיית תקן של 500 וואט לשעה.
- במדגם של 25 מקררים של היצרן התקבל ממוצע של 2342 וואט לשעה.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - מהי טעות התקן של האומדן?
 - מהו האומדן לפרמטר?
 - מהי טעות האמידה?
- (3) נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו: 175, 168, 184, 192, 171, 180, 177, 187, 168, 177.
- מצאו אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
 - מצאו אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
 - מצאו אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.
- (4) נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים.
- להלן התוצאות שהתקבלו: $\sum_{i=1}^{20} X_i^2 = 1502.2$, $\sum_{i=1}^{20} X_i = 162$.
- אמדו את השכר הממוצע של השכירים במשק.
 - אמדו את סטיית התקן של שכר השכירים במשק.
- (5) במטרה לאמוד את ממוצע האוכלוסייה, דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישבו את הממוצע שלהם. מהי טעות התקן?
- סטיית התקן של האוכלוסייה.
 - סטיית התקן של ממוצע האוכלוסייה.
 - סטיית התקן של המדגם.
 - סטיית התקן של ממוצע המדגם.

6) משקל הממוצע של אוכלוסייה מסוימת הוא 75 ק"ג עם שונות של 25. אם יבחרו כל המדגמים האפשריים בגודל 10 מאוכלוסייה זו סטיית התקן של ממוצעי המדגמים תהייה:

- א. 3.
- ב. 2.5.
- ג. 1.581.
- ד. אין מספיק נתונים לדעת.

7) במדגם מקרי, מתי סכום ריבועי הסטיות מהממוצע, $\sum_{i=1}^n (x_i - \bar{x})^2$, מחולק ב- $n-1$?

- א. כאשר n קטן.
- ב. כאשר תצפיות המדגם אינן בלתי תלויות.
- ג. כאשר האוכלוסייה אינה מתפלגת נורמאלית.
- ד. כאשר מעוניינים באומד חסר הטיה לשונות האוכלוסייה ממנה הוצא המדגם.
- ה. כאשר מעוניינים לחשב את שונות התפלגות הדגימה של ממוצע המדגם.

8) $X_1, X_2, \dots, X_{16}$ מדגם מקרי מתוך אוכלוסייה בעלת ממוצע μ לא ידוע ושונות: $\sigma^2 = 64$. טעות התקן של האומד ל- μ היא:

- א. 16.
- ב. 8.
- ג. 4.
- ד. 2.

9) מהו אומד חסר הטיה?

- א. אומד שערכו שווה לממוצע התפלגות הדגימה שלו.
- ב. אומד שערכו שווה לערך הפרמטר באוכלוסייה.
- ג. אומד שממוצע התפלגות הדגימה שלו שווה לערך הפרמטר באוכלוסייה.
- ד. אומד שהסיכוי שערכו יהיה גבוה מערך הפרמטר באוכלוסייה שווה לסיכוי שיהיה נמוך ממנו.

א. 0.25	ב. 0.019	ג. 0.24	ד. 0.01
א. אוכלוסייה : מקררים של יצרן, תוחלת : 2400, סטיית תקן : 500.			
ב. 100		ג. 2342	ד. 58
א. 177.9	ב. 64.1	ג. 0.4	
א. 8.1	ב. 3.16		
ד'			
ג'			
ד'			
ד'			
ג'			

סטטיסטיקה והסתברות

פרק 25 - רווח סמך לתוחלת (ממוצע)

תוכן העניינים

- 1. רווח סמך כששונות האוכלוסיה ידועה 104
- 2. קביעת גודל מדגם 109
- 3. רווח סמך כששונות האוכלוסיה לא ידועה 111

רווח סמך כששונות האוכלוסייה ידועה:

רקע:

ממוצע המדגם הוא אומדן לממוצע האוכלוסייה, אך לא באמת ניתן להבין ממנו על גודלו של ממוצע האוכלוסייה. ההסתברות שממוצע המדגם יהיה בדיוק כמו הממוצע האמיתי הוא אפסי.

מה שנהוג לעשות כדי לאמוד את ממוצע האוכלוסייה, זה לבנות רווח סמך.

נבנה מרווח בטחון שהסיכוי שהפרמטר μ ייכלל בתוכו הוא: $1-\alpha$.

$1-\alpha$: נקרא רמת בטחון או רמת סמך. כך ש: $P(A \leq \mu \leq B) = 1-\alpha$.

A - גבול התחתון של רווח הסמך.

B - הגבול העליון של רווח הסמך.

$L = B - A$ - אורך רווח הסמך.

דוגמה (פתרון בהקלטה):

חוקר דגם 25 חיילים שנבחנו במבחן הפסיכומטרי. הוא בנה רווח סמך לממוצע הציונים במבחן הפסיכומטרי בקרב אוכלוסיית החיילים וקיבל בין 510 ל-590. רווח הסמך נבנה ברמת סמך של 95%.

1. מהי אוכלוסיית המחקר?
2. מה המשתנה באוכלוסייה?
3. מה הפרמטר שהחוקר רצה לאמוד?
4. מהו רווח הסמך?
5. מה אורך רווח הסמך?
6. מהי רמת הביטחון של רווח הסמך?

בפרק זה נרצה לבנות רווח סמך לתוחלת (μ) במקרה ש- σ^2 (שונות האוכלוסייה) ידועה.

פרמטר אותו נרצה לאמוד: μ .

אומד נקודתי: $\bar{x}$.

תנאים לבניית רווח הסמך: $X \sim N$ או $n \geq 30$.

σ^2 (שונות האוכלוסייה) ידועה.

נוסחה לרווח הסמך: $\bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

דוגמה (פתרון בהקלטה):

על פי נתוני היצרן אורך חיי סוללה מתפלג נורמאלית עם סטיית תקן של 1 שעה. מעוניינים לאמוד את תוחלת חיי סוללה. נדגמו באקראי 4 סוללות, אורך החיים הממוצע שהתקבל הוא 13.5 שעות. בנו רווח סמך ברמת סמך של 95% לתוחלת אורך חיי סוללה.

שגיאת האמידה המקסימלית: $\varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

ε - נותן את שגיאת האמידה המקסימלית, דבר שנקרא גם טעות סטטיסטית, טעות דגימה.

דוגמה (פתרון בהקלטה):

בהמשך לשאלה עם הסוללות. מה ניתן להגיד בביטחון של 95% על שגיאת האמידה?

קשרים מתמטיים ברווח הסמך:

- אורך רווח הסמך הוא פעמיים שגיאת האמידה המקסימלית: $L = 2\varepsilon$.
- ממוצע המדגם נופל תמיד באמצע רווח הסמך: $\bar{X} = \frac{A+B}{2}$.
- ככל שמספר התצפיות (n) גבוה יותר, כך יש יותר אינפורמציה ולכן האומד יותר מדויק, ולכן נקבל רווח סמך יותר קצר.
- ככל שרמת הביטחון $(1-\alpha)$ גבוהה יותר, כך: $Z_{1-\frac{\alpha}{2}}$ יותר גבוה, ורווח הסמך יותר ארוך.

שאלות:

- (1) חוקר התעניין לאמוד את השכר הממוצע במשק. על סמך מדגם הוא קבע שבביטחון של 95% כי השכר הממוצע במשק נע בין 9200 ל-9800 ₪.
- מי האוכלוסייה במחקר?
 - מה המשתנה הנחקר?
 - מה הפרמטר שאותו רוצים לאמוד?
 - מה רווח הסמך לפרמטר?
 - מהי רמת הסמך לפרמטר?
 - מה אורך רווח הסמך?
 - מה הסיכוי שטעות הדגימה תעלה על 300 ₪?
- (2) מעוניינים לאמוד את התפוקה היומית הממוצעת של מפעל מסוים ברמת סמך של 95%. במדגם אקראי של 100 ימים התקבלה תפוקה ממוצעת 4950 מוצרים ביום. לצורך פתרון הנח שסטיית התקן האמתית ידועה ושווה 150 מוצרים ביום. בנו את רווח הסמך.
- (3) מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלי עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
- בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
 - בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
 - הסבירו כיצד ומדוע השתנה רווח הסמך.
- (4) דגמו 200 עובדים מהמשק הישראלי. השכר הממוצע שלהם היה 9700 ₪. נניח שסטיית התקן של השכר במשק היא 3000 ₪.
- בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.
 - מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?
 - מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?
 - אם היינו מגדילים את גודל המדגם ובונים רווח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרווח הסמך מכיל את הפרמטר?

- (5) בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (105,99). שחזרו את:
- א. ממוצע המדגם.
 - ב. שגיאת האמידה המקסימאלית.
 - ג. רמת הסמך.
- (6) זמן החלמה מאנגינה מתפלג עם סטיית תקן של יומיים. חברת תרופות מעוניינת לחקור אנטיביוטיקה חדשה שהיא פיתחה. במחקר השתתפו 60 אנשים שחלו באנגינה וקיבלו את האנטיביוטיקה החדשה. בממוצע הם החלימו לאחר 4 ימים.
- א. בנו רווח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.
 - ב. מה היה קורה לאורך רווח הסמך אם היה תקציב להגדלת גודל המדגם פי 4? הסבירו.
 - ג. מה היה קורה לאורך רווח הסמך אם היינו בונים את רווח הסמך ברמת סמך גדולה יותר? הסבירו.
- (7) חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא: $82 < \mu < 92$. נתון שסטיית התקן בהתפלגות שווה ל-10 ושהמדגם מתבסס על 16 תצפיות. התפלגות המשתנה היא נורמאלית.
- א. מהו ממוצע המדגם?
 - ב. מהי רמת הסמך של רווח הסמך שנבנה?
 - ג. מה הסיכוי ששגיאת האמידה באמידת ממוצע האוכלוסייה תעלה על 5?
- (8) חוקר בנה רווח סמך לתוחלת כאשר השונות בהתפלגות ידועה ברמת סמך של 95%. אם החוקר כעת יבנה על סמך אותם נתונים רווח סמך ברמת סמך קטנה מ-95%, איזה מהמשפטים הבאים לא יהיה נכון.
- א. אורך רווח הסמך החדש יהיה קטן יותר.
 - ב. גודל המדגם יהיה כעת קטן יותר.
 - ג. המרחק בין ממוצע המדגם לקצות רווח הסמך יהיו קטנים יותר ברווח הסמך החדש.
 - ד. רמת הביטחון לבנות רווח הסמך החדש תהיה קטנה יותר.

(9) חוקר בנה רווח סמך ל- μ וקיבל: $48 < \mu < 54$. מה נכון בהכרח:

א. $\mu = 51$.

ב. $\bar{X} = 6$.

ג. $\bar{X} = 51$.

ד. אורך רווח הסמך הינו 3.

(10) איזה מהגורמים הבאים אינו משפיע על גודלו של רווח בר סמך, כאשר שונות האוכלוסייה ידועה (בחרו בתשובה הנכונה):

א. רמת הביטחון.

ב. סטיית התקן באוכלוסייה.

ג. מספר המשתתפים.

ד. סטיית התקן במדגם.

תשובות סופיות:

(1) א. העובדים במשק. ב. שכר ב-ש. ג. μ . ד. $9200 < \mu < 9800$.

ה. 0.95. ו. 600. ז. 0.05.

(2) $4920.6 < \mu < 4979.4$

(3) א. $223.42 < \mu < 236.58$. ב. $222.16 < \mu < 237.84$.

ג. ראה סרטון.

(4) א. $9284 < \mu < 10,116$. ב. הסטיה המירבית בין $\bar{x}$ ל- μ היא 416 שם בבטחון של 95%.

ג. 800. ד. לא.

(5) א. 102. ב. 3. ג. 0.9544.

(6) א. $4.42 < \mu < 83.5$. ב. יקטן פי 2. ג. גדל.

(7) א. 87. ב. 5. ג. 0.9544.

(8) ב'.

(9) ג'.

(10) ד'.

קביעת גודל מדגם:

רקע:

אם מעוניינים לאמוד את ממוצע האוכלוסייה כאשר סטיית התקן של האוכלוסייה ידועה: σ ברמת סמך של $1-\alpha$ ושגיאת אמידה שלא תעלה על ε מסוים, נציב

$$.n \geq \left(\frac{z_{1-\frac{\alpha}{2}} \cdot \sigma}{\varepsilon} \right)^2 \quad \text{בנוסחה הבאה:}$$

כדי להציב בנוסחה צריך שהמשתנה הנחקר יתפלג נורמלית או שהמדגם ייצא בגודל של לפחות 30 תצפיות.

דוגמה (פתרון בהקלטה):

חברת תעופה מעוניינת לאמוד את תוחלת משקל המטען של נוסע. נניח שמשקל מטען של נוסע מתפלג נורמאלית עם סטיית תקן של 2 ק"ג. כמה נוסעים יש לדגום אם מעוניינים שבביטחון של 98% הסטייה המרבית בין ממוצע המדגם לממוצע האמיתי לא יעלה על 0.5 ק"ג? (תשובה: 87).

שאלות:

- (1) משתנה מקרי מתפלג נורמאלית עם סטיית תקן ידועה 12. מה צריך להיות גודל המדגם כדי לבנות רווח סמך ברמת סמך של 98% שאורכו לא יעלה על 2?
- (2) מעוניינים לאמוד את הדופק הממוצע של מתגייסים לצבא. מעוניינים שבביטחון של 95% שגיאת האמידה המרבית תהיה 0.5. נניח שהדופק מתפלג נורמאלית על סטיית תקן של 3 פעימות לדקה.
- א. כמה מתגייסים יש לדגום?
- ב. אם ניקח מדגם הגדול פי 4 מהמדגם של סעיף א ונאמוד את הממוצע באותה רמת סמך כיצד הדבר ישפיע על שגיאת האמידה?
- (3) יהי X משתנה מקרי עם ממוצע μ וסטיית תקן σ . חוקר רוצה לבנות רווח בר סמך ל- μ ברמת ביטחון של 0.95, כך שהאורך של הרווח יהיה 0.5σ . מהו גודל המדגם הנדרש?

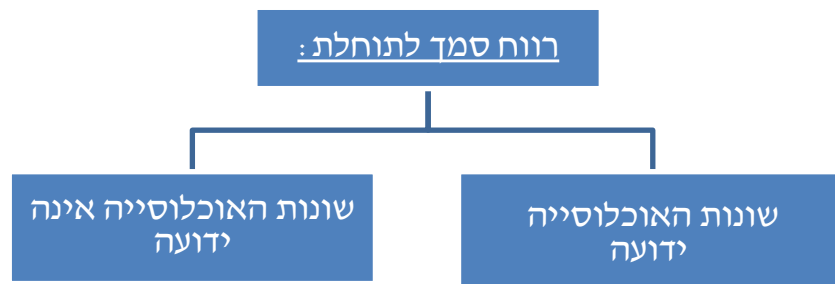
תשובות סופיות:

- (1) 780.
- (2) א. 139. ב. הדבר יקטין את ε פי 2.
- (3) $n = 62$.

רווח סמך כששונות האוכלוסייה לא ידועה:

רקע:

בבואנו לבנות רווח סמך לתוחלת אנו צריכים להתמקד בשני המצבים הבאים:

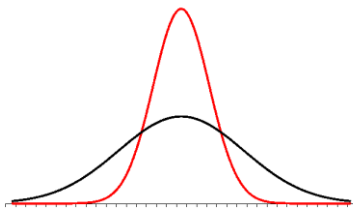


בפרק זה נעסוק במקרה ששונות האוכלוסייה (σ^2) אינה ידועה לנו.

מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול.

רווח סמך: $\bar{X} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \cdot \frac{S}{\sqrt{n}}$



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

התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה

להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים.

התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן: $df = n-1$.

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה.

כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

דוגמה (פתרון בהקלטה):

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי כיתות ח' נורמאלית.

במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח'. להלן התוצאות

שהתקבלו בדקות: 4.7, 5.2, 4.6, 5.3.

בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

שאלות:

- (1) מחקר מעוניין לדעת כיצד תרופה מסוימת משפיעה על קצב פעימות הלב. ל-5 אנשים שנטלו את התרופה מדדו את הדופק והתקבל מספר פעימות לדקה: 84, 88, 84, 79, 89. הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.
- א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הנ"ל.
- ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?
- ג. בהמשך לסעיף א', אם היינו בונים את רווח הסמך ברמת ביטחון של 99%, כיצד הדבר היה משפיע על רווח הסמך?
- (2) במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן: $S = 13$ ס"מ.
- בנו רווח סמך ברמת סמך של 90% לתוחלת גובה המתגייסים לצבא האמריקאי. מה יש להניח לצורך פתרון?
- (3) אדם מעוניין לאמוד את זמן הנסיעה הממוצע שלו לעבודה. לצורך כך הוא דוגם 5 ימים שזמן הנסיעה בהם בדקות הוא: 30, 40, 32, 34, 27.
- א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?
- ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים?
- (4) ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.
- א. בנו רווח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.
- ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמתית של כלל הנבחנים.
- ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.
- (5) נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X_i^2 = 643.19$.
- בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

- (6) נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו. להלן התוצאות שהתקבלו: $\bar{x} = 13.8$, $S = 2$. בנו רווח סמך ברמת סמך של 96% לממוצע ההשכלה של אזרחים מעל גיל 50.
- (7) שני סטטיסטיקאים בנו רווח בר-סמך לאותו פרמטר μ . לכל אחד מהסטטיסטיקאים מדגם אחר, אך באותו גודל 10. שניהם קבעו אותה רמת סמך. סטטיסטיקאי א': הניח $\sigma = 20$. סטטיסטיקאי ב': חישב לפי המדגם וקיבל $S = 20$. למי משני הסטטיסטיקאים יהיה רווח סמך ארוך יותר?
- סטטיסטיקאי א'.
 - סטטיסטיקאי ב'.
 - אותו אורך רווח סמך לשני הסטטיסטיקאים.
 - תלוי בתוצאות המדגם של כל סטטיסטיקאי.
- (8) נתון ש: $X \sim N(\mu, \sigma^2)$ ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגמית 10. אורך רווח הסמך שהתקבל הוא: 8.765. מהי רמת הביטחון של רווח הסמך?

תשובות סופיות:

- (1) א. $79.88 < \mu < 89.72$ ב. כן. ג. הוא היה גדל.
- (2) ראה בסרטון.
- (3) א. צריך להניח שהמשתנה מתפלג נורמלית. ב. לא ניתן לדעת.
- (4) א. $96.63 < \mu < 107.37$ ב. $96.90 < \mu < 107.10$ ג. ראה בסרטון.
- (5) $3.149 < \mu < 3.351$
- (6) $13.42 < \mu < 14.18$
- (7) ב'.
- (8) 90%.

סטטיסטיקה והסתברות

פרק 26 - רווח סמך לפרופורציה

תוכן העניינים

114	1. רווח הסמך לפרופורציה
117	2. קביעת גודל מדגם

רווח הסמך לפרופורציה:

רקע:

המטרה היא לאמוד את P – פרופורציה באוכלוסייה.

האומד הנקודתי:

$$\hat{p} = \frac{y}{n} \quad (Y - \text{מספר ההצלחות שבמדגם}).$$

$$\text{רווח הסמך ל- } p : \hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

תנאי לבניית רווח הסמך:

מדגם של לפחות 30 תצפיות (לעיתים נותנים תנאי של מספר הצלחות ומספר כשלונות לפחות 5 או לפחות 10).

$$\text{האומד לטעות התקן: } \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\text{מתקיים ש: } \hat{p} = \frac{A+B}{2}, \quad L = 2\varepsilon$$

דוגמה (פתרון בהקלטה):

במטרה לאמוד את אחוז המובטלים במשק נדגמו 200 אזרחים, מתוכם התקבל ש-24 היו מובטלים.

א. בנו רווח סמך לאחוז המובטלים באוכלוסייה ברמת סמך של 95%.

ב. מהו האומד לטעות התקן?

שאלות:

- (1) נדגמו 200 דירות בעיר חיפה. 48 מתוכן נמצאו כבעלות ממ"ד.
 א. בנו רווח סמך ברמת סמך של 95% לאחוז הדירות בחיפה עם ממ"ד.
 ב. על סמך סעיף א' מה ניתן לומר על שגיאת האמידה המקסימאלית?
 ג. בהנחה ובחיפה 80 אלף דירות, בנו רווח סמך ברמת סמך של 95% למספר הדירות בחיפה עם ממ"ד בפועל.
- (2) במדגם של 300 אנשי היי-טק התקבל ש-180 מהם אקדמאים.
 א. בנו רווח סמך לפרופורציית אקדמאים ברמת סמך של 95% (בקרב אנשי הייטק).
 ב. כיצד רווח הסמך של סעיף א' היה משתנה אם היינו מקטינים את רמת הסמך?
 ג. כיצד רווח הסמך היה משתנה אם היינו מגדילים את גודל המדגם?
- (3) במדגם של 400 נהגים התקבל רווח סמך לפרופורציית הנהגים החדשים:
 $0.08 < p < 0.18$
 א. כמה נהגים במדגם היו נהגים חדשים?
 ב. מהי רמת הסמך של רווח הסמך שנבנה?
- (4) במסגרת מערכת הבחירות בארה"ב נשאלו 840 אנשים עבור איזה מועמד יצביעו. 510 אנשים ענו כי יצביעו בעד ברק אובמה. בסקר פורסם שתתכן סטייה של $\pm 3\%$ מתוצאות האמת. באיזו רמת ביטחון הסקר השתמש?
- (5) במדגם של 300 נשים בגילאי 35-40 נמצא ש-140 היו נשואות, 80 היו גרושות, 60 רווקות והיתר אלמנות.
 א. מצאו רווח סמך ברמה של 90% לאחוז הגרושות באוכלוסייה הנחקרת.
 ב. מצאו רווח סמך ברמה של 99% לסיכוי שבאוכלוסייה הנחקרת תמצא אישה לא נשואה?
- (6) ביצעו מדגם באוכלוסייה. שיעור ההצלחות במדגם היה 10% ורווח הסמך ניבנה ברמת סמך של 95%. אורכו הינו 8.3156%. מהו גודל המדגם שנלקח?

תשובות סופיות:

- (1) א. $18.1\% < p < 29.9\%$
 ב. בביטחון של 95% שגיאת האמידה היא לכל היותר 0.059.
 ג. $14,480 < \mu < 23,920$
- (2) א. $0.545 \leq p \leq 0.655$ ב. האורך שלו היה קטן. ג. לא ניתן לדעת.
- (3) א. 52. ב. 0.997
- (4) 0.925
- (5) א. $30.9\% > p > 22.5\%$ ב. $60.72\% > p > 45.91\%$
- (6) 200

קביעת גודל מדגם:

רקע:

בפרק זה נדון איך קובעים גודל מדגם שבאים לאמוד פרופורציה באוכלוסייה מסוימת: החוקר קובע מראש את רמת הסמך הרצויה: $1-\alpha$. החוקר קובע מראש את הטעות הסטטיסטית המרבית שבה הוא מעוניין: ε (או את אורך רווח הסמך).

$L = 2\varepsilon$ - אורך רווח הסמך.

ε - טעות אמידה מרבית: המרחק המקסימאלי (הסטייה) בין הפרמטר (p) לאומד ($\hat{p}$).

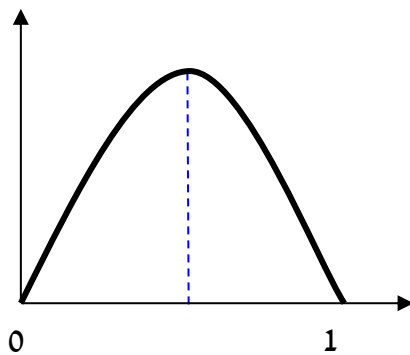
$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

$$n \geq \left(\frac{2 \cdot z_{1-\frac{\alpha}{2}} \cdot \sqrt{\hat{p}(1-\hat{p})}}{L} \right)^2 \quad \text{נקבל ש:}$$

הבעיה שאין אנו יודעים את $\hat{p}$.

נתבונן בביטוי: $\hat{p}(1-\hat{p})$.



כיוון שאין לנו ידע מוקדם על $\hat{p}$ נציב את המקרה השמרני ביותר שממקסם את הביטוי עבור: $\hat{p} = 0.5$.

$$n \geq \left(\frac{2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{0.5 \cdot 0.5}}{L} \right) \Rightarrow n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{L} \right)$$

אך אם תהיה לנו אינפורמציה מוקדמת על הפרופורציה נציב את הערך הקרוב ביותר ל-0.5 האפשרי.

דוגמה (פתרון בהקלטה):

מעוניינים לאמוד את שיעור האבטלה במשק. האמידה צריכה להתבצע ברמת סמך של 90% ועם שגיאת אמידה שלא תעלה על 4%.

א. מהו גודל המדגם המינימאלי שיש לקחת?

ב. חזור לסעיף א' אם ידוע שהאבטלה לא אמורה לעלות על 20%.

שאלות:

- (1) הממשלה אומדת מדי חודש את אחוז התמיכה בה. מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמיתי באוכלוסייה ביותר מ-3%, וזאת בביטחון של 95%?
- (2) משרד התקשורת מעוניין לדעת מה שיעור בתי האב עם אינטרנט.
 - א. כמה בתי אב יש לדגום אם מעוניינים שבביטחון של 90% אורך רווח הסמך לא יעלה על 8%?
 - ב. חזרו על סעיף א' אם ידעו שלפני חמש שנים ל-80% מבתי האב היה אינטרנט וכיום יש להניח שיש ליותר אינטרנט.
- (3) ערוץ טלוויזיה מעוניין לאמוד את הרייטינג של הערוץ בפריים טיים. המטרה שבביטחון של 95% הסטייה המרבית בין האומד לרייטינג האמיתי לא תעלה על 4%.
 - א. כמה מכשירי PEOPLE METER יש להתקין לצורך האמידה?
 - ב. לפי הערכה מוקדמת הרייטינג של הערוץ לא יכול לעלות על 20%. בהנחה ומכשיר כזה עולה 500 ₪ ליחידה מה החיסכון הכספי מאינפורמציה זאת?
- (4) ענו על הסעיפים הבאים:
 - א. כמה אזרחים יש לדגום כדי לאמוד את אחוז התמיכה בממשלה עם אורך רווח הסמך שלא עולה על 9% ברמת סמך של 90%?
 - ב. בהנחה ובוצע מדגם שאת גודלו חישבתם בסעיף א והתקבל שאחוז התמיכה בממשלה במדגם הנו 42%. בנו רווח סמך לאחוז התמיכה בממשלה ברמת סמך של 95%.
 - ג. על סמך סעיף ב', האם תקבלו את הטענה שמיעוט האוכלוסייה תומך הממשלה?
- (5) משרד הבריאות מתכנן לבצע מדגם שמטרתו לבדוק את הסיכוי לחלות בשפעת עם לקיחת חיסון נגד שפעת. הוא מעוניין שבסיכוי של 98% טעות האמידה לא תעלה על 3%.
 - א. כמה מחוסנים יש לדגום?
 - ב. משרד הבריאות ביצע את המדגם שאת גודלו חישבתם בסעיף הקודם וקיבל ש-15% מבין אלה שקיבלו חיסון נגד שפעת בכל זאת חלו במשך החורף בשפעת. בנו ברמת סמך של 98% את הסיכוי לחלות בחורף בשפעת עם לקיחת חיסון נגד שפעת.
 - ג. בהמשך לסעיף הקודם. מהי טעות האמידה המרבית בביטחון של 98% מדוע הוא קטן מ-3%?

תשובות סופיות:

- (1) 1.1068
- (2) א. 0.423 ב. 0.271
- (3) א. 0.601 ב. 108,000 ₪.
- (4) א. 0.335 ב. $0.367 < p < 0.473$.
- ג. בביטחון של 0.95 ניתן להגיד שמיעוט באוכלוסייה תומך בממשלה.
- (5) א. 0.1509 ב. 0.15 ± 0.02 ג. ראה סרטון.

סטטיסטיקה והסתברות

פרק 27 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

1. כששונניות האוכלוסיה ידועות.....120
2. כששונניות האוכלוסיה לא ידועות ובהנחת שוויון שונניות.....122
3. כששונניות האוכלוסיה לא ידועות ובהנחת שוויון שונניות.....124

כששונויות האוכלוסייה ידועות:

רקע:

המטרה היא לאמוד את פער התוחלות: $\mu_1 - \mu_2$, כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

האומד נקודתי: $\bar{x}_1 - \bar{x}_2$.

התנאים לבניית רווח הסמך:

1. σ_1^2, σ_2^2 ידועות.

2. $X_1, X_2 \sim N$ או $n_1, n_2 > 30$.

3. שני מדגמים בלתי תלויים.

רווח סמך: $(\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של $1-\alpha$, לא קיים הבדל בין התוחלות.

דוגמה (פתרון בהקלטה):

נדגמו 100 תושבים מאזור A והמשכורת הממוצעת הייתה שם 9200 ₪. כמו כן נדגמו 120 תושבים מאזור B וממוצע המשכורות שהתקבל שם 8700 ₪. לצורך פתרון נניח שסטיית התקן של המשכורות באוכלוסיית שני האזורים היא 1800 ₪. אמדו ברמת סמך של 90% את הפרש השכר הממוצע בין אזור A לאזור B.

שאלות:

- (1) מעוניינים לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. ידוע שציוני הפסיכומטרי מתפלגים נורמאלית עם סטיית תקן 100. במדגם של 16 נבחנים חיילים התקבל ממוצע 543. במדגם של 20 תלמידי תיכון התקבל ממוצע 508. בנו רווח סמך לפער תוחלות הציונים בין חיילים לתלמידי תיכון ברמת סמך של 90%. מה ניתן להסיק מרווח סמך זה?
- (2) ציוני IQ מתוכננים כך שיתפלגו נורמאלית עם סטיית תקן של 15. במדגם של 20 נבחנים ישראלים התקבל ממוצע ציונים 104. במדגם של 23 נבחנים אמריקאיים התקבל ממוצע ציונים 99. א. בנו רווח סמך ברמת סמך של 95% לפער בין ישראל לארה"ב בממוצע הציונים במבחן ה-IQ. ב. האם קיים הבדל בין ישראלים לאמריקאים מבחינת ממוצע הציונים?
- (3) חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. ידוע שרמת הקשיות של ברגים מתפלגת נורמלית עם סטיית תקן של 4 יחידות. במדגם של 15 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות ובמדגם של 12 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25. עבור אילו רמות בטחון יקבע שאין הבדל בין שני סוגי הברגים מבחינת ממוצע רמת הקשיות שלהם?

תשובות סופיות:

- (1) $(-20, 90)$.
- (2) א. $-3.99 < \mu_1 - \mu_2 < 13.99$.
ב. לא נוכל לטעון בביטחון של 95% שקיים הבדל בין ישראל לארה"ב.
- (3) רמות בטחון הגבוהות מ-0.9476.

כששונויות האוכלוסייה לא ידועות ובהנחת שוויון שונויות:

רקע:

המטרה היא לאמוד את פער התוחלות: $\mu_1 - \mu_2$, כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

האומד נקודתי: $\bar{X}_1 - \bar{X}_2$.

התנאים לבניית רווח הסמך:

$$1. \sigma_1^2 = \sigma_2^2$$

$$2. X_1, X_2 \sim N$$

3. מדגמים בלתי תלויים.

השונויות המשוקללת: כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השונויות הזו על ידי שקלול שתי השונויות של שני המדגמים על ידי

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$d.f = n_1 + n_2 - 2$$

$$\text{רווח סמך: } (\bar{X}_1 - \bar{X}_2) \pm t_{\frac{n_1+n_2-2}{2}, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של $1 - \alpha$, לא קיים הבדל בין התוחלות.

דוגמה (פתרון בהקלטה):

מחקר מעוניין לבדוק האם קיים הבדל בין תל אביב לבאר שבע מבחינת ההכנסה הממוצעת של אקדמאים. להלן תוצאות המדגם שנעשה:

מספר האקדמאים	תל אביב	באר שבע
10	20	
ממוצע הכנסות של אקדמאים	11,000	9500
סטיית התקן של הכנסות אקדמאים	200	250

בנו רווח סמך ברמת ביטחון של 90% להפרש תוחלות ההכנסה בשני האזורים. הניחו שהשכר מתפלג נורמלית עם אותה שונות בכל אחד מהאזורים.

שאלות:

- (1) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

המדינה	ישראל	ארה"ב
גודל המדגם	15	15
סכום הציונים	1560	1470
סכום ריבועי הציונים	165,390	147,560

מצאו רווח סמך ברמת סמך של 95% לסטייה בין ממוצע הציונים בישראל לממוצע הציונים בארה"ב. רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

- (2) להלן 4 תצפיות על משתנה X שמתפלג: $N(\mu_x, \sigma^2)$, ומשתנה Y שמתפלג: $N(\mu_y, \sigma^2)$.

X	22	20	21	25
Y	18	25	17	12

חשבו רווח סמך ל- $\mu_y - \mu_x$ ברמת הסמך 90%, בהנחה ששני המדגמים בלתי תלויים.

תשובות סופיות:

- (1) הנחות:
- השונות שווה.
 - שהציונים מתפלגים נורמלית.
 - המדגמים אינם תלויים זה בזה.
- $$-5.52 < \mu_1 - \mu_2 < 17.52$$
- (2)
- $$-9.6 < \mu_y - \mu_x < 1.6$$

כששונויות האוכלוסייה לא ידועות ובהנחת שוויון שונויות:

רקע:

המטרה היא לאמוד את פער התוחלות: $\mu_1 - \mu_2$, כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

האומד נקודתי: $\bar{X}_1 - \bar{X}_2$.

התנאים לבניית רווח הסמך:

$$1. \sigma_1^2 = \sigma_2^2.$$

$$2. X_1, X_2 \sim N.$$

3. מדגמים בלתי תלויים.

השונויות המשוקללת: כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השונויות הזו על ידי שקלול שתי השונויות של שני המדגמים על ידי

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{הנוסחה הבאה:}$$

$$d.f = n_1 + n_2 - 2 \quad \text{דרגות החופש:}$$

$$\text{רווח סמך: } (\bar{X}_1 - \bar{X}_2) \pm t_{1-\frac{\alpha}{2}}^{n_1+n_2-2} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של $1 - \alpha$, לא קיים הבדל בין התוחלות.

דוגמה (פתרון בהקלטה):

מחקר מעוניין לבדוק האם קיים הבדל בין תל אביב לבאר שבע מבחינת ההכנסה הממוצעת של אקדמאים. להלן תוצאות המדגם שנעשה:

מספר האקדמאים	תל אביב	באר שבע
ממוצע הכנסות של אקדמאים	11,000	9500
סטיית התקן של הכנסות אקדמאים	200	250

בנו רווח סמך ברמת ביטחון של 90% להפרש תוחלות ההכנסה בשני האזורים. הניחו שהשכר מתפלג נורמלית עם אותה שונות בכל אחד מהאזורים.

שאלות:

- (1) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

המדינה	ישראל	ארה"ב
גודל המדגם	15	15
סכום הציונים	1560	1470
סכום ריבועי הציונים	165,390	147,560

מצאו רווח סמך ברמת סמך של 95% לסטייה בין ממוצע הציונים בישראל לממוצע הציונים בארה"ב. רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

- (2) להלן 4 תצפיות על משתנה X שמתפלג: $N(\mu_x, \sigma^2)$, ומשתנה Y שמתפלג: $N(\mu_y, \sigma^2)$.

X	22	20	21	25
Y	18	25	17	12

חשבו רווח סמך ל- $\mu_y - \mu_x$ ברמת הסמך 90%, בהנחה ששני המדגמים בלתי תלויים.

תשובות סופיות:

- (1) הנחות:
- השונות שווה.
 - שהציונים מתפלגים נורמלית.
 - המדגמים אינם תלויים זה בזה.
- $$-5.52 < \mu_1 - \mu_2 < 17.52$$
- (2)
- $$-9.6 < \mu_y - \mu_x < 1.6$$

סטטיסטיקה והסתברות

פרק 28 - רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים

תוכן העניינים

1. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 126

רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים:

רקע:

מדגם מזווג: מדגם אחד שבו יש n צמדים. כל תצפית במדגם תנפק זוג ערכים: X ו- Y .

ניצור משתנה חדש: $D = x - y$.

הפרמטר שנרצה לאמוד: μ_D .

התנאים לבניית רווח הסמך:

$$1. \quad x, y \sim N$$

2. המדגם מזווג.

$$\text{נוסחת רווח הסמך: } \bar{D} \pm t_{1-\frac{\alpha}{2}}^{n-1} \frac{S_D}{\sqrt{n}}$$

כאשר דרגות החופש: $df = n - 1$.

דוגמה (פתרון בהקלטה):

מעוניינים לבדוק האם יש הבדל בין מהירות הריצות של שתי תוכנות מחשב. לקחו 5 קבצים אקראיים והריצו אותם בשתי התוכנות:

הקובץ	1	2	3	4	5
הזמן בתוכנה הראשונה	25	48	49	46	38
הזמן בתוכנה השנייה	27	46	42	40	48

הניחו כי זמני הריצות מתפלגים נורמלית. מצאו רווח סמך של 95% להפרש תוחלת הזמן בין שתי התוכנות.

שאלות:

- (1) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן הציונים בסמסטר א' ו-ב' :

82	75	90	68	74	סמסטר א'
100	76	87	84	80	סמסטר ב'

נניח שהציונים מתפלגים נורמאלית.

- א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א' לבין סמסטר ב'.
- ב. האם על סמך רווח הסמך קיים הבדל בין הסמסטרים מבחינת תוחלת הציונים?
- ג. מה צריך לשנות בנתונים כדי שהמדגמים יהיו בלתי תלויים?
- (2) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נדגמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות :

חברה/ מדינה	ארה"ב	קנדה	הולנד	פולין	מצרים	סין	יפן
בזק - X	1.5	2.1	2.2	3	3.5	3.2	4.2
קווי זהב - Y	1.4	2	1.9	3.1	3.3	3.2	4.2

בהנחה והמחירים מתפלגים נורמלית עבור כל חברה, בנו רווח סמך ברמת סמך של 90% לתוחלת הפרש המחירים של שתי החברות.

תשובות סופיות:

- (1) א. $-19 < \mu_0 < 38$. ב. בביטחון של 95% לא קיים הבדל. ג. ראה הסבר בסרטון.
- (2) $-0.013 < \mu < 0.185$.

סטטיסטיקה והסתברות

פרק 29 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

128	1. הקדמה
132	2. סוגי טעויות

הקדמה:

רקע:

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

שלב א: נוהה 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 אזורי 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 הדחייה 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:

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

- 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:

מציאות	הכרעה		
		H_0	H_1
	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. $H_0: \mu = 500$
 $H_1: \mu < 500$ 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.
 ה. לא לדחות את H_0 . ו. טעות מסוג שני.
- ד. $H_0: \mu = 2.3$
 $H_1: \mu < 2.3$

סטטיסטיקה והסתברות

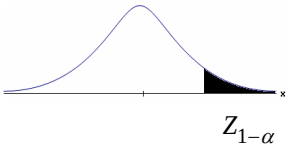
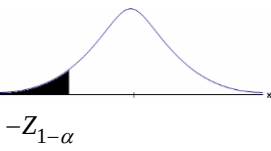
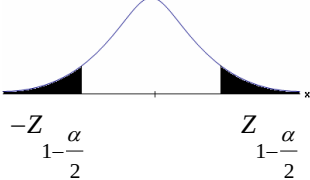
פרק 30 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

1. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה ידועה 134
2. סיכוי לטעויות ועוצמה (ששונות האוכלוסיה ידועה) 138
3. קביעת גודל מדגם (ששונות האוכלוסיה ידועה) 144
4. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה ידועה) 147
5. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה לא ידועה 152
6. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה לא ידועה) 156
7. ניתוח פלטים 159
8. הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע) 164

בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה ידועה:

רקע:

השערת האפס: השערה אלטרנטיבה:	$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
תנאים:	1. σ ידועה 2. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$Z_{\bar{x}} > Z_{1-\alpha}$  $Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}} < -Z_{\frac{\alpha}{2}}$ או $Z_{\bar{x}} > Z_{\frac{\alpha}{2}}$  $-Z_{\frac{\alpha}{2}}$ $Z_{\frac{\alpha}{2}}$ דוחים את H_0

$$Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

סטטיסטי המבחן:

חלופה אחרת לכלל הכרעה:

$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$	נדחה H_0 אם מתקיים:

דוגמה:

יבול העגבניות מתפלג נורמלית עם תוחלת של 10 טון לדונם וסטיית תקן של 2.5 טון לדונם בעונה. משערים ששיטת זיבול חדשה תעלה את תוחלת היבול לעונה מבלי לשנות את סטיית התקן. נדגמו 4 חלקות שזוכלו בשיטה החדשה. היבול הממוצע שהתקבל היה 12.5 טון לדונם. בדקו את ההשערה ברמת מובהקות של 1%.

פיתרון:

אוכלוסייה: עגבניות.

המשתנה: $X =$ יבול העגבניות בטון לעונה.

הפרמטר: $\mu =$ תוחלת היבול בשיטת הזיבול החדשה.

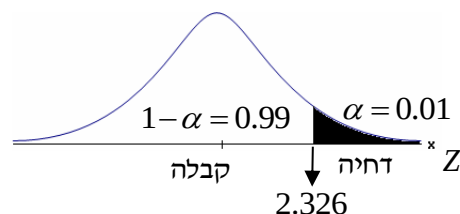
השערות:
 $H_0: \mu = 10$
 $H_1: \mu > 10$

תנאים:

1. $X \sim N$.

2. $\sigma = 2.5$.

כלל הכרעה:



נדחה את H_0 אם $Z_{\bar{X}} > 2.326$

תוצאות: $n = 4$, $\bar{x} = 12.5$

סטטיסטי המבחן: $Z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

נציב: $Z_{\bar{X}} = \frac{12.5 - 10}{\frac{2.5}{\sqrt{4}}} = 2 < 2.326$

מסקנה:

לא נדחה H_0 (נקבל H_0).

ברמת מובהקות של 1% לא נוכל לקבל את הטענה ששיטת הזיבול החדשה מעלה את תוחלת היבול של העגבניות.

שאלות:

- (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. בהנחה שגם בשיטתו סטיית התקן תהיה 15 מה מסקנתכם ברמת מובהקות של 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.
 - א. מה מסקנתכם ברמת מובהקות של 2.5%?
 - ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?
- (3) מהנדס האיכות מעוניין לבדוק אם מכונה מכיילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?
- (4) המשקל הממוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ובשימוש בדיאטה מסוימת שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם זה היה 84 ק"ג. יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.
- (5) לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדגם של 25 ברגים העובי הממוצע היה 4.07 מ"מ. קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט. הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.
- (6) במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5% מה תמיד נכון? בחרו בתשובה הנכונה.
 - א. הגדלת רמת המובהקות לא תשנה את מסקנת המחקר.
 - ב. הגדלת רמת המובהקות תשנה את מסקנת המחקר.
 - ג. הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
 - ד. הקטנת רמת המובהקות תשנה את מסקנת המחקר.

(7) חוקר ערך מבחן דו צדדי ברמת מובהקות של α והחליט לדחות את השערת האפס.

אם החוקר היה עורך מבחן צדדי ברמת מובהקות של $\frac{\alpha}{2}$ אזי בהכרח:

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס הייתה לא נדחית.
- ג. לא ניתן לדעת מה תהיה מסקנתו במקרה זה.

(8) שני סטטיסטיקאים בדקו השערות: $H_0: \mu = \mu_0$ כנגד $H_1: \mu > \mu_0$,

עבור שונות ידועה ובאותה רמת מובהקות.

שני החוקרים קבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.

- א. אם חוקר א' החליט לדחות את H_0 , מה יחליט חוקר ב'? נמקו.
- ב. אם חוקר א' יחליט לא לדחות את H_0 , מה יחליט חוקר ב'? נמקו.

תשובות סופיות:

(1) נקבל H_0 , בר"מ של 5% לא נקבל את הטענה של המורה ששיטת הלימוד שלו מעלה את ממוצע הציונים.

(2) א. נדחה H_0 , בר"מ של 2.5% נקבל את תלונת אגודת הצרכנים בדבר הפחתת נפח המשקה בבקבוק.

ב. הגדלנו את רמת המובהקות לכן אנחנו נשארים בדחייה של H_0 והמסקנה לא תשתנה.

(3) נדחה H_0 , בר"מ של 5% נקבע שהמכונה לא מאופסת.

(4) נדחה H_0 , בר"מ של 0.1 נקבל את הטענה שהדיאטה יעילה ומפחיתה את המשקל הממוצע.

(5) נקבל H_0 , בר"מ של 0.05 נכריע שתוחלת עובי הבורג מתים למפרט.

(6) א'.

(7) ג'.

(8) א. לדחות. ב. לא ניתן לדעת.

סיכוי לטעויות ועוצמה (ששונות האוכלוסייה ידועה):

רקע:

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

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

$$\alpha = P(H_0 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_0}(H_0 \text{ לדחות})$$

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ לקבל את } H_0 \mid \text{לדחות את } H_0) = P_{H_1}(H_0 \text{ לקבל})$$

רמת בטחון:

$$(\alpha - 1) = P(H_0 \text{ לקבל את } H_0 \mid \text{לדחות את } H_0) = P_{H_0}(H_0 \text{ לקבל})$$

עוצמה:

$$(\beta - 1) = \pi = P(H_0 \text{ לדחות את } H_0 \mid \text{לדחות את } H_0) = P_{H_1}(H_0 \text{ לדחות})$$

התהליך לחישוב סיכוי לטעות מסוג שני:

השערת האפס: השערה אלטרנטיבה:	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
תנאים:	1. σ ידועה 2. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
חישוב β :	$P_{\mu_1} \left(\bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$	$P_{\mu_1} \left(\bar{X} > \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$	$P_{\mu_1} \left(\mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$

התפלגות ממוצע המדגם: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

התקנון:

דוגמה:

בתחילת השנה חשבון הטלפון הסלולארי הממוצע לאדם היה 200 ₪ עם סטיית תקן של 80 ₪ לחודש. בעקבות כניסתן של חברות טלפון סלולארית חדשות מעוניינים לבדוק האם כיום ממוצע חשבון הטלפון הסלולארי פחת. לצורך בדיקה דגמו באקראי 36 אנשים וחשבון הטלפון הסלולארי שלהם היה 150 ₪ בממוצע לחודש.

א. רשמו את השערות המחקר ובנו כלל הכרעה במונחי חשבון ממוצע מדגמי ברמת מובהקות של 5%.

ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?

ג. נניח שבמציאות כיום החשבון הממוצע הוא 160 ₪. מה הסיכוי לבצע טעות מסוג שני?

ד. אם נקטין את רמת המובהקות מסעיף א', כיצד הדבר ישפיע על התשובה מסעיף ג'?

פתרון:

א. אוכלוסייה: משלמי חשבון טלפון סלולאר כיום.

המשתנה: X = חשבון הטלפון החודשי בשקלים.

הפרמטר: μ .

השערות:
 $H_0: \mu = 200$
 $H_1: \mu < 200$

תנאים:

1. $\mu = 200$.

2. $n = 36$.

$$\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, K = \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\alpha = 0.05$$

$$Z_{1-\alpha} = Z_{0.95} = 1.645$$

$$\text{נציב: שקלים } K = 200 - 1.645 \cdot \frac{80}{\sqrt{36}} = 178.07$$

כלל ההכרעה: דחה את H_0 אם שקלים $\bar{X} < 178.07$.

ב. ברמת מובהקות של 5% נכריע שאכן ממוצע חשבון הטלפון הסלולרי פחת מתחילת השנה.

ג. השערות:
 $H_0: \mu_0 = 200$
 $H_1: \mu < 200$

כלל ההכרעה: נדחה את H_0 אם $\bar{X} < 178.07$.

$$H_1: \bar{X} \sim N\left(160, \frac{80^2}{36}\right)$$

$$Z = \frac{178.07 - 160}{\frac{80}{\sqrt{36}}} = 1.36$$

$$\beta = P_{H_1} \left(\overset{\text{לקבל את}}{H_0} \right) = P_{H_1} (\bar{X} > 178.07) = 1 - \phi(136) = 1 - 0.9131 = 0.0869$$

ד. הקטנת α מגדילה את β .

שאלות:

(1) נתון ש: $X \sim N(\mu, \sigma^2 = 1)$.

- להלן השערות של חוקר לגבי הפרמטר μ : $H_0: \mu = 5$, $H_1: \mu = 7$. מעוניינים ליצור כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.
- א. עבור אילו ערכים של X שידגם נדחית השערת H_0 ?
- ב. מה הסיכוי לבצע טעות מסוג שני?
- ג. אם במדגם התקבל ש- $X = 6.9$ מה תהיה המסקנה ומה הטעות האפשרית?

- (2) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. מעוניינים לבדוק אם כיום ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה.
- א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי ברמת מובהקות של 5%.
- ב. בהמשך לסעיף א' מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?
- ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העצמה של הכלל מסעיף א'?

- (3) להלן נתונים על תהליך של בדיקת השערות על תוחלת:

$$H_0: \mu = 200, H_1: \mu \neq 200, \sigma = 30, n = 225$$

- א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.
- ב. בהמשך לסעיף א', מהי העצמה אם התוחלת שווה ל-195?
- ג. הסבירו, ללא חישוב, איך העצמה תשתנה אם רמת המובהקות תהיה 5%?

- (4) מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50

מ"מ וסטית תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומוודדים את קוטרו, בכדי לבדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוילת כנדרש או שקוטר הצינורות קטן מהדרוש.

- א. רשמו את ההשערות ואת כלל ההכרעה ברמת מובהקות של 5%.
- ב. אם ביום כלשהו מכונת הייצור התקלקלה והיא מייצרת את הצינורות בקוטר שתוחלתו 48 מ"מ בלבד (סטית התקן לא השתנתה), מה ההסתברות שהתקלה לא תתגלה בביקורת האיכות? כיצד נקראת הסתברות זו?
- ג. הסבירו ללא חישוב כיצד התשובה לסעיף ב' תשתנה אם רמת המובהקות תגדל.
- ד. הסבירו ללא חישוב כיצד התשובה לסעיף ב' תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

- (5) להלן השערות של מחקר: $H_0: \mu = 50$, $H_1: \mu = 58$. מעוניינים לדגום 100 תצפיות. ידוע שסטיית התקן של ההתפלגות הינה 20.
- א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10%. מהי רמת המובהקות?
- ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו)?
- i. סטיית התקן הייתה יותר גדולה.
- ii. הסיכוי לטעות מסוג שני גדול יותר.

השאלות שלהלן הן שאלות רב-ברירה, בחרו בתשובה הנכונה ביותר:

- (6) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:
- א. הסיכוי לטעות מסוג ראשון גדל.
- ב. העוצמה של המבחן גדלה.
- ג. הסיכוי לטעות מסוג שני גדל.
- ד. תשובות א' ו-ב' נכונות.
- (7) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן:
- א. השערת האפס נכונה.
- ב. השערת האפס נדחתה.
- ג. השערת האפס לא נדחתה.
- ד. אף אחת מהתשובות לא נכונה בהכרח.

- (8) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

$1 - \beta$	α
גדולה	א. גדולה
קטנה	ב. גדולה
גדולה	ג. קטנה
קטנה	ד. קטנה

- (9) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית H_0 קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך:
- א. הן α , והן $1 - \beta$, יקטנו.
- ב. α יישאר ללא שינוי ואילו $1 - \beta$ יגדל.
- ג. α יגדל ואילו $1 - \beta$ יקטן.
- ד. הן α והן $1 - \beta$ יגדלו.

- (10) ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח שלחץ הדם בקרב עיתונאים גבוה יותר מהממוצע באוכלוסייה. הוא לקח מדגם של 60 עיתונאים וקיבל ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?
- טעות מסוג ראשון.
 - טעות מסוג שני.
 - טעות מסוג שלישי.
 - אין טעות במסקנתו.

תשובות סופיות:

- (1) א. מעל 6.645. ב. 0.3594.
ג. דחינו את H_0 , תתכן טעות מסוג ראשון.
- (2) א. נדחה H_0 אם $\bar{X} < 2.24$. ב. נדחה H_0 ג. 1.
- (3) א. נדחה H_0 אם $\bar{X} > 203.29$ או $\bar{X} < 196.71$. ב. 0.8051. ג. תקטן.
- (4) א. נדחה H_0 אם $\bar{X} < 48.9$. ב. 0.0885. ג. תקטן. ד. תקטן.
- (5) א. 0.0033. ב. i. רמת המובהקות הייתה קטנה.
ב. ii. רמת המובהקות הייתה גדלה.
- (6) ד'.
- (7) ג'.
- (8) ג'.
- (9) א'.
- (10) ב'.

קביעת גודל מדגם (ששונות האוכלוסייה ידועה):

רקע:

השערות המחקר הן: $H_0: \mu = \mu_0$, $H_1: \mu = \mu_1$.

סטיית התקן של האוכלוסייה ידועה σ ומעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β .

$$\text{הנוסחה הבאה נותנת את גודל המדגם הרצוי: } n \geq \left(\frac{(Z_{1-\alpha} + Z_{1-\beta}) \times \sigma}{\mu_0 - \mu_1} \right)^2$$

דוגמה:

משרד החינוך מפעיל בגן חובה שיטת חינוך שפותחה בשנת 1995. לפי שיטת חינוך זו תוחלת הציון במבחן אוצר מילים לגיל הרך הוא 70. אנשי חינוך החליטו לבדוק שיטת חינוך שפותחה בהולנד הנותנת שם תוחלת ציון אוצר מילים של 80. נניח שציוני מבחן זה מתפלגים נורמאלית עם $\sigma = 17$. כדי לבדוק האם גם בישראל הפעלת שיטת החינוך ההולנדית תעבוד בגנים, רוצים לבנות מחקר ברמת מובהקות של 5%. כמו כן, מעוניינים שאם בהפעלת השיטה ההולנדית תוחלת הציונים תעלה לכדי 80, המחקר יגלה זאת בסיכוי של 90%. כמה ילדי גן חובה דרושים למחקר?

פתרון:

האוכלוסייה: ילדי גן חובה.

המשתנה: $X =$ ציון במבחן אוצר מילים.

הפרמטר: μ .

השערות: $H_0: \mu = 70$
 $H_1: \mu = 80$

$$X \sim N(\mu, \sigma^2 = 17^2)$$

אם בהפעלת השיטה ההולנדית התוחלת תעלה ל-80, נגלה זאת בסיכוי 90%.

$$n \geq \left(\frac{(Z_{1-\alpha} + Z_{1-\beta}) \times \sigma}{\mu_0 - \mu_1} \right)^2$$

$$\alpha = 0.05$$

$$1 - \beta = 0.9$$

$$\mu_0 = 70$$

$$\mu_1 = 80$$

$$\sigma = 17$$

$$Z_{1-\alpha} = Z_{0.95} = 1.645$$

$$Z_{1-\beta} = Z_{0.9} = 1.282$$

$$n \geq \left(\frac{(1.645 + 1.282) \times 17}{70 - 80} \right)^2 = 24.76 \quad \text{נציב:}$$

$$\text{לכן, } n_{\min} = 25.$$

שאלות:

- (1) במבחן אינטליגנציה הציונים מתפלגים נורמאלית עם סטיית תקן 8 וממוצע 100. פסיכולוג מעוניין לבדוק את הטענה שבאוכלוסיות במצב סוציו אקונומי נמוך תוחלת הציונים היא 95. אם מעוניינים לגלות את הטענה בהסתברות של לפחות 99% כשרמת המובהקות היא 5% מהו גודל המדגם הדרוש?
- (2) משרד התקשורת טוענים שאדם מדבר בממוצע 180 דקות בחודש בטלפון הסלולרי. חברות הטלפון הסלולרי טוענות שאינפורמציה זו אינה נכונה ואדם מדבר בממוצע פחות: כ-160 דקות. לצורך פתרון נניח שסטיית התקן של זמן השיחה החודשי ידוע ושווה ל-60 דקות. כמה אנשים יש לדגום כך שאם טענת משרד התקשורת נכונה נדחה אותה בסיכוי של 5% (איך קוראים להסתברות זאת?) כמו כן אם טענת חברות הטלפון הסלולרית נכונה המחקר יגלה זאת בסיכוי של 90% (איך קוראים להסתברות זאת?).
- (3) השערות המחקר הן: $H_0: \mu = \mu_0$, $H_1: \mu = \mu_1$. כמו כן נתון שהמשתנה מתפלג נורמלית עם סטיית התקן ידועה σ מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β . הוכיחו שגודל המדגם הרצוי לכך יהיה:
- $$n \geq \left(\frac{(Z_{1-\alpha} + Z_{1-\beta}) \times \sigma}{\mu_0 - \mu_1} \right)^2$$

תשובות סופיות:

- (1) 41.
(2) 78.
(3) שאלת הוכחה.

מובהקות תוצאה – אלפא מינימלית (ששונות האוכלוסייה ידועה):

רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה:

באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v . את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות. המסקנה של המחקר תקבע לפי העיקרון הבא: אם $p_v \leq \alpha$, דוחים את H_0 . מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

(לקבל את תוצאות המדגם וקיצוני) $p_v = P_{H_0}$.

אם ההשערה היא דו צדדית:

(לקבל את תוצאות המדגם וקיצוני) $p_v = 2P_{H_0}$.

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס: השערה אלטרנטיבה:		$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	
1. σ ידועה		תנאים:	
2. $X \sim N$ או מדגם מספיק גדול		p-value	
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \Leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \Leftarrow \bar{x} < \mu_0$	

כאשר בהנחת השערת האפס: $Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$, $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$

דוגמה:

המשקל הממוצע של מתגייסים לצבא לפני 20 שנה היה 65 ק"ג. מחקר מעוניין לבדוק האם כיום המשקל הממוצע של מתגייסים גבוה יותר. נניח שמשקל המתגייסים מתפלג נורמאלי עם סטיית תקן של 12 ק"ג. במדגם של 16 מתגייסים התקבל משקל ממוצע של 71 ק"ג.

א. מהי מובהקות התוצאה?

ב. מה המסקנה אם רמת המובהקות היא 5% ואם רמת המובהקות היא 1%?

פתרון:

א. אוכלוסייה: המתגייסים לצבא כיום.

משתנה: X = משקל בק"ג.

פרמטר: μ .

השערות:
 $H_0: \mu = 65$
 $H_1: \mu > 65$

תנאים:

1. $X \sim N$.

2. $\sigma = 12$.

תוצאות מדגם:

$$n = 16$$

$$\bar{X} = 71$$

$$P_V = P_{H_0} \left(\begin{array}{c} \text{לתוצאות} \\ \text{המדגם} \\ \text{וקיצוני} \end{array} \right) = P_{H_0} (\bar{X} \geq 71) = 1 - \phi(2) = 1 - 0.9772 = 0.0228$$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{71 - 65}{\frac{12}{\sqrt{16}}} = 2$$

$$\alpha_{\min} = 0.0228$$

שאלות:

- (1) להלן השערות של מחקר: $H_0: \mu = 70$, $H_1: \mu > 70$. המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות: $n = 100$, $\bar{x} = 74$. מהי מובהקות התוצאה?
- (2) השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪. מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?
- (3) אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ-100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטיית התקן של משקל החטיף היא 12 גרם. האדם מתכוון לשקול 100 חפיסות חטיפים ולאחר מכן להגיע להחלטה. לאחר הבדיקה הוא קיבל משקל הממוצע של 98.5 גרם.
 - א. רשמו את השערות המחקר.
 - ב. מהי רמת המובהקות המינימלית עבורה דוחים את השערת האפס?
 - ג. מהי רמת המובהקות המקסימלית עבורה נקבל את השערת האפס?
 - ד. מה המסקנה ברמת מובהקות של 5?
- (4) מכונה לחיתוך מוטות במפעל חותכת מוטות באורך שמתפלג נורמלית עם תוחלת אליה כוונה המכונה וסטיית תקן 2 ס"מ. ביום מסוים כוונה המכונה לחתוך מוטות באורך 80 ס"מ. אחראי האיכות מעוניין לבדוק האם המכונה מכוילת. לצורך כך נדגמו מקו הייצור 16 מוטות שנחתכו אורכן הממוצע היה 81.7 ס"מ.
 - א. מהי רמת המובהקות המינימלית עבורה נכריע שהמכונה לא מכוילת?
 - ב. אם נוסיף עוד תצפית שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?
 - ג. הכרע ברמת מובהקות של 5% האם המכונה מכוילת.
- (5) אם מקבלים בחישובים אלפא מינימלית (P value) קטנה מאד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון/לא נכון? נמק.

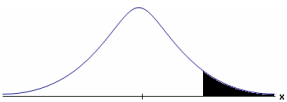
- (6) בבדיקת השערות התקבל שה- $p\text{-value} = 0.02$.
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%?
בחרו בתשובה הנכונה.
- א. יקבל את השערת האפס בכל מקרה.
 - ב. ידחה את השערת האפס מקרה.
 - ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
 - ד. לא ניתן לדעת כי אין מספיק נתונים.
- (7) מובהקות התוצאה (PV) היא גם (בחרו בתשובה הנכונה):
- א. רמת המובהקות המינימאלית לדחות השערת האפס.
 - ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
 - ג. רמת המובהקות שנקבעת מראש על ידי החוקר שטרם קיבל את תוצאות המחקר.
 - ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.
- (8) בבדיקת השערות מסוימת התקבל: $p\text{ value} = 0.0254$ לכן (בחרו בתשובה הנכונה):
- א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .
 - ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .
 - ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .
 - ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

תשובות סופיות:

- (1) 0.0228.
- (2) עבור כל רמת מובהקות סבירה.
- (3) א. $H_0: \mu = 100$
 $H_1: \mu < 100$
 ב. 0.1056 ג. 0.1056.
- ד. נכריע שיש עמידה בהתחייבות של החברה.
- (4) א. 0.0006 ב. יקטן. ג. נכריע שאין כיול.
- (5) נכון.
- (6) א'.
- (7) א'.
- (8) ג'.

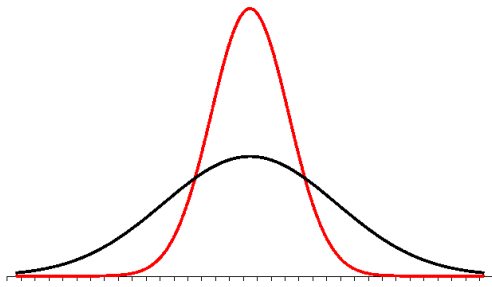
בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה לא ידועה:

רקע:

השערת האפס: השערה אלטרנטיבה:	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
תנאים:	1. σ אינה ידועה 2. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$t_{\bar{x}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{x}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{x}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{x}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ - דוחים את H_0
חלופה לכלל ההכרעה: נדחה H_0 אם מתקיים:	$\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ או $\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$

$$. t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} : \text{סטטיסטי המבחן}$$

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



התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויה במושג שנקרא דרגות חופש.

דרגות החופש הן: $df = n - 1$.

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה. כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

דוגמה (פתרון בהקלטה):

מפעל קיבל הזמנה לייצור משטחים בעובי של 0.1 ס"מ. כדי לבדוק האם המפעל עומד בדרישה נדגמו 10 משטחים ונמצא שהעובי הממוצע הוא 0.104 עם אומדן לסטיית תקן 0.002 ס"מ.

א. מהן השערות המחקר?

ב. מה ההנחה הדרושה לצורך פתרון?

ג. בדוק ברמת מובהקות של 5%.

שאלות:

(1) משך זמן ההחלמה בלקיחת אנטיביוטיקה מסוימת הוא 120 שעות בממוצע עם סטיית תקן לא ידועה. מעוניינים לבדוק האם אנטיביוטיקה אחרת מקטינה את משך זמן ההחלמה. במדגם של 5 חולים שלקחו את האנטיביוטיקה האחרת התקבלו זמני ההחלמה הבאים: 90, 95, 100, 80, 125 שעות. מה מסקנתכם ברמת מובהקות של 5%. מהי ההנחה הדרושה לצורך הפתרון?

(2) משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום היוולדם בישראל 3300 גר'. 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:

$$n = 20$$

$$\bar{x} = 3120$$

$$S = 280$$

מה מסקנתכם ברמת מובהקות של 5% מה יש להניח לצורך פתרון?

(3) ציוני מבחן אינטליגנציה מתפלגים נורמלית. בארה"ב ממוצע הציונים הוא 100. במדגם שנעשה על 23 נבחנים ישראלים, התקבל ממוצע ציונים 104.5 וסטיית התקן המדגמית 16. האם בישראל ממוצע הציונים שונה מארה"ב? הסיקו ברמת מובהקות של 5%.

(4) באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{10} X_i = 750$$

$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$

נתון שההתפלגות היא נורמלית.

בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80.

- (5) ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם. ליאור השתמש בטבלה של התפלגות Z . רוני השתמשה בטבלה של התפלגות t . מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.
- אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
 - אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
 - שני החוקרים בהכרח יגיעו לאותה מסקנה.
 - לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

- (6) נתון ש: $X \sim N(\mu, \sigma^2)$ כמו כן נתונות ההשערות הבאות: $H_0: \mu = \mu_0$. $H_1: \mu < \mu_0$. חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות. בחר בתשובה הנכונה:
- א. כעת בברור הוא ידחה את השערת האפס.
 - ב. כעת הוא דווקא יקבל את השערת האפס.
 - ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

תשובות סופיות:

- (1) נדחה H_0 .
- (2) נדחה H_0 .
- (3) נקבל H_0 .
- (4) נקבל H_0 .
- (5) ב'.
- (6) ג'.

מובהקות תוצאה – אלפא מינימלית (ששונות האוכלוסייה לא ידועה):

רקע:

נזכיר שהמסקנה של המחקר תיקבע לפי העיקרון הבא: אם $p_v \leq \alpha$ דוחים את H_0 . מובהקות התוצאה היא הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

(לקבל את תוצאות המדגם וקיצוני) $p_v = P_{H_0}$.

אם ההשערה היא דו צדדית:

(לקבל את תוצאות המדגם וקיצוני) $p_v = 2P_{H_0}$.

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס: השערה אלטרנטיבה:		$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
תנאים:		1. אינה ידועה או 2. מדגם מספיק גדול $X \sim N$		
p-value		אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \Leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \Leftarrow \bar{x} < \mu_0$	$P_{H_0}(\bar{X} \leq \bar{x})$	$P_{H_0}(\bar{X} \geq \bar{x})$

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$$

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

$$d.f = n-1$$

דוגמה:

ממוצע זמן הנסיעה של אדם לעבודה הינו 40 דקות. הוא מעוניין לבדוק דרך חלופית שאמורה להיות יותר מהירה. לצורך כך הוא דוגם 5 ימים שבהם הוא נוסע בדרך החלופית. זמני הנסיעה שקיבל בדקות הם: 34, 40, 30, 32, 27. הניחו שזמן הנסיעה מתפלג נורמלי.

- א. רשמו את השערות המחקר.
- ב. מצאו חסמים למובהקות התוצאה.
- ג. מה המסקנה ברמת מובהקות של 5%?

פתרון:

אוכלוסייה: כלל הנסיעות לעבודה בדרך החלופית.

משתנה: X = זמן נסיעה בדקות.

תנאים: $X \sim N$.

פרמטר: μ .

א. השערות:
 $H_0: \mu = 40$
 $H_1: \mu < 40$

ב. תוצאות המדגם:

$$n = 5, \bar{X} = \frac{\sum X_i}{n} = \frac{34 + 40 + \dots}{5} = 32.6$$

$$S^2 = \frac{\sum X_i^2 - n \cdot \bar{X}^2}{n-1} = \frac{34^2 + 40^2 - \dots - 5 \cdot 32.6^2}{5-1} = 23.4$$

$$S = \sqrt{23.4}$$

$$t_{\bar{X}} = \frac{\frac{\bar{X} - \mu_0}{S}}{\frac{1}{\sqrt{n}}} = \frac{32.6 - 40}{\frac{4.88}{\sqrt{5}}} = -3.39$$

$$P_V = P_{H_0} = (\bar{X} \leq 32.6) = P(t \leq -3.39)$$

$$d.f = 5 - 1 = 4$$

$$1\% < P_V < 2.5\%$$

$P_V < \alpha = 0.05$, לכן דוחים את H_0 .

מסקנה: בר"מ של 5% נכריע שהדרך החלופית מהירה יותר.

שאלות:

- (1) קו ייצור אריזות סוכר נארזות כך שהמשקל הממוצע של אריזות הסוכר צריך להיות אחד קילוגרם. בכל יום דוגמים מקו הייצור 5 אריזות במטרה לבדוק האם קו הייצור תקין. בבדיקה דגמו 5 אריזות סוכר ולהלן משקלן בגרמים: 1024, 1008, 1005, 996, 997.
- א. רשמו את השערות המחקר.
- ב. מהי מובהקות התוצאה? הצג חסמים.
- ג. מה המסקנה ברמת מובהקות של 5%?
- (2) חוקר בדק את הטענה כי פועלים העובדים במשמרת לילה איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. במדגם מיקרי של 25 פועלים שעבדו במשמרת לילה נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטית תקן של 3 שעות.
- מהי ה- α המינימלית שלפיה ניתן להחליט שאכן העובדים במשמרת לילה איטיים יותר?
- (3) הגובה של מתגייסים לצה"ל מתפלג נורמלית. במדגם של 25 מתגייסים מדדו את הגבהים שלהם בס"מ והתקבלו התוצאות הבאות:
- $$\bar{x} = 176.2, \sum (x_i - \bar{x})^2 = 2832$$
- מטרת המחקר היא לבדוק האם תוחלת הגבהים של המתגייסים גבוה מ-174 ס"מ באופן מובהק.
- מהי בקרוב מובהקות התוצאה ועל פיה מה תהיה המסקנה ברמת מובהקות של 6%?

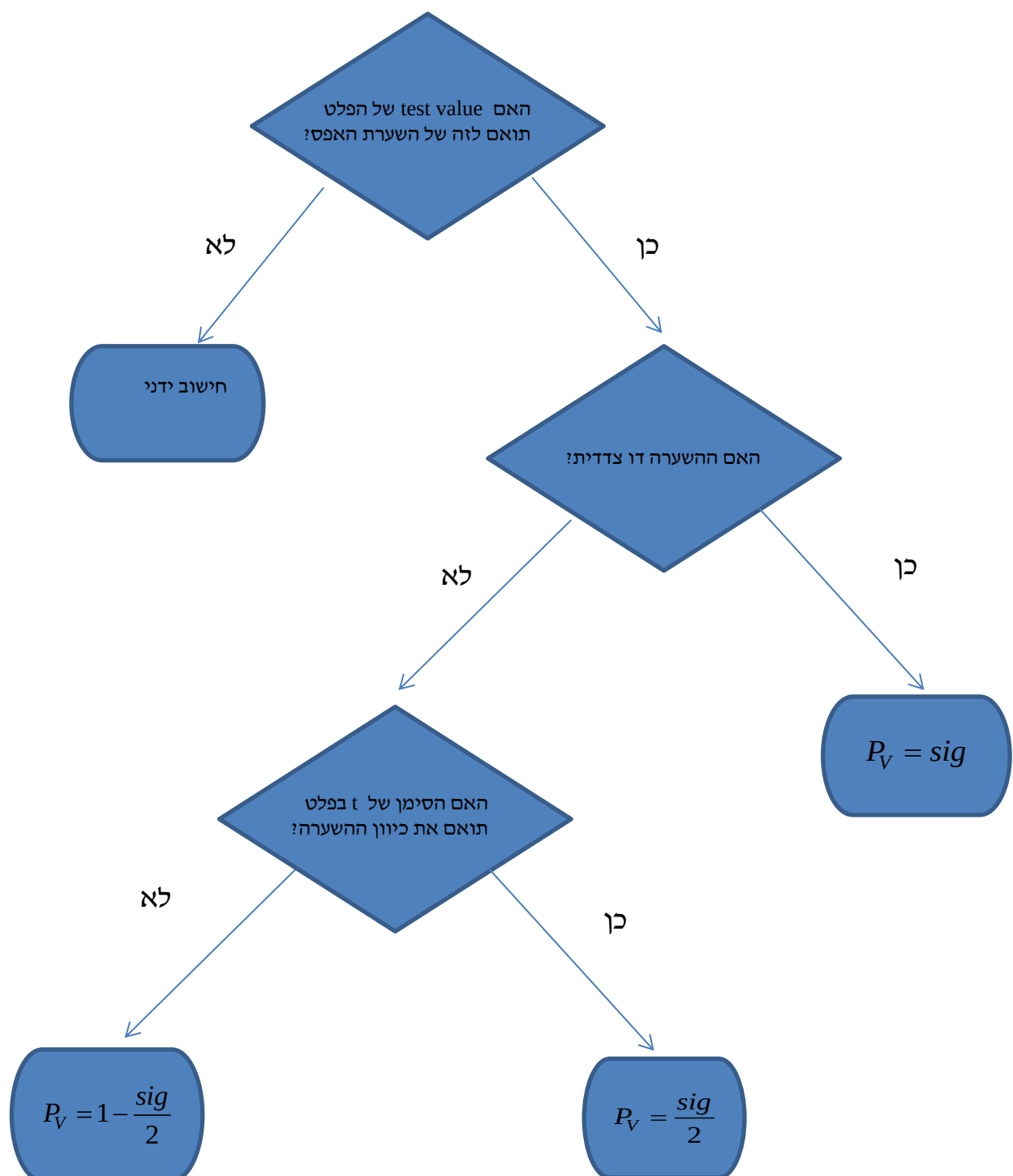
תשובות סופיות:

- (1) א. $H_0: \mu = 1000$
 ב. $20\% \leq P_v \leq 50\%$
 ג. ברמת מובהקות של 5% לא נוכל לקבוע שקו הייצור אינו תקין.
- (2) 10%.
- (3) 1.01, נקבל את H_0 .

ניתוח פלטים:

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS:



דוגמה (פתרון בהקלטה):

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
$\bar{X}$	25	87.6400	64.90434	12.98087

One-Sample Test

Test Value = 60						
				Mean Difference	95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)		Lower	Upper
$\bar{X}$	2.129	24	.044	27.64000	.8488	54.4312

ממוצע הציונים במבחן המיצב בחשבון הוא 60. הוחלט לדגום כיתה אקראית של 25 תלמידים וללמד אותם בשיטת לימוד חדשה.

- א. מהו רווח הסמך לממוצע הציונים בחשבון אם יוחלט ליישם את שיטת הלימוד החדשה?
- ב. מהו P_V לבדיקת יעילותה של שיטת הלימוד החדשה?
- ג. מה יוכרע ברמת מובהקות של 5% לגבי יעילותה של שיטת הלימוד החדשה?

שאלות:

- 1) באוניברסיטה גדולה גיל הסטודנטים לתואר ראשון מתפלג נורמאלי. בעבר פורסם שהגיל הממוצע של הסטודנטים הינו 23. להלן פלט תוכנת SPSS על מדגם של 16 סטודנטים אקראיים מתואר ראשון:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
age	16	23.4375	2.50250	.62562

One-Sample Test

	Test Value = 23					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
age	.699	15	.495	.43750	-.8960	1.7710

- א. מהו המבחן הסטטיסטי שנעשה כאן?
 ב. מה ערכו של הפרמטר לפי השערת האפס?
 ג. רשום רווח סמך ברמת סמך של 95% לתוחלת גיל הסטודנטים באוניברסיטה לתואר ראשון.
 ד. בדוק ברמת מובהקות של 5% האם הגיל הממוצע כיום שונה מבעבר?

- 2) קבוצת ילדים בגיל 6 קיבלה משימה לביצוע. עבור כל ילד בדקו כמה זמן לקח לו לסיים את המשימה בדקות. להלן תוצאות הניתוח הסטטיסטי:

One-Sample Test

	Test Value = 4.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
time	-1.853	24	.076	-.09200	-.1944	.0104

- א. כמה ילדים השתתפו במחקר?
 ב. מצא רווח סמך ברמת סמך של 95% לתוחלת זמן ביצוע המשימה עבור ילדים בני 6.
 ג. מה יש להניח כדי שרווח הסמך מסעיף א' יהיה תקף?
 ד. בדוק ברמת מובהקות של 5% שזמן ביצוע המשימה הממוצע נמוך מ-4.5 דקות.

3) להלן פלט מחשב עבור ניתוח סטטיסטי שנעשה בתוכנת SPSS. הניתוח הוא עבור מדגם אקראי של קבוצת נבחנים בבגרות באנגלית.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
grade	???	???	19.62787	2.95901

One-Sample Test

	Test Value = 75					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
grade	???	43	.017	-7.34091	-13.3083	???

- א. השלימו את הגדלים החסרים המסומנים בסמני שאלה בפלט.
 ב. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים שונה מ-75?
 ג. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים קטנה מ-75?
 ד. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים גדולה מ-75?

4) יצרן סיגריות מפרסם כי תוחלת הניקוטין בסיגריות שהוא מיצר קטנה מ-27 מ"ג. בבדיקה מקרית של 5 סיגריות מתוצרתו נמצאו כמויות הניקוטין הבאות: 21, 21, 20, 24, 22 מ"ג. הניחו כי כמות הניקוטין בסיגריות מפולג נורמאלי.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
nicotine	5	21.6000	1.51658	.67823

One-Sample Test

	Test Value = 27					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
nicotine	-7.962	4	.001	-5.40000	-7.2831	-3.5169

- א. האם ברמת מובהקות של 5% ניתן להסיק שיש אמת בפרסום?
 ב. אם היינו מוסיפים עוד תצפית שערכה 20. כיצד הדבר היה משפיע על הערך Sig ועל המסקנה?
 ג. בדקו האם ניתן להגיד שתוחלת רמת הניקוטין שונה מ-26 ברמת מובהקות של 5%.

תשובות סופיות:

- (1) א. הסקה של תוחלת אחת. ב. 23. ג. $(22.104, 24.771)$. ד. נקבל H_0 .
- (2) א. 25. ב. $(4.3056, 4.5104)$. ג. המשתנה מתפלג נורמלית. ד. נדחה H_0 .
- (3) א. $n = 44, \bar{X} = 67.66, t = -2.48, upper = -1.3735$. ב. 0.017. ג. 0.0085. ד. 0.9915.
- (4) א. נכריע שיש אמת בפרסום. ב. המסקנה לא תשתנה. ג. נכריע שהתוחלת שונה מ-26.

הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע):

רקע:

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α על μ :

$$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$$

על ידי בניית רווח סמך ברמת סמך של $1-\alpha$ ל- μ :

אם μ_0 נופל ברווח $\leftarrow$ נקבל את H_0 .

אם μ_0 לא נופל ברווח $\leftarrow$ נדחה את H_0 .

דוגמה:

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו :

$$H_0: \mu = 80, H_1: \mu \neq 80, \alpha = 5\%$$

החוקר בנה רווח סמך ברמה של 90% וקיבל: $79 < \mu < 84$.
האם אפשר לדעת מה מסקנתו, ואם כן מהי?

פתרון (פתרון מלא בהקלטה):

רווח הסמך ברמת סמך של 90% מכיל "80".

ברמת סמך של 95% רווח הסמך יגדל ויכיל "80".

לכן, ברמת מובהקות של 5% נקבל H_0 .

שאלות:

- (1) חוקר רצה לבדוק את ההשערות הבאות: $H_0: \mu = 90$, $H_1: \mu \neq 90$. החוקר בנה רווח סמך לתוחלת ברמת סמך של 95% וקיבל את רווח הסמך הבא: (87, 97).
אם החוקר מעוניין לבצע בדיקת השערות ברמת מובהקות של 1% האם ניתן להגיע למסקנה ע"ס רווח הסמך? נמקו.
- (2) חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בדם. ידוע כי מספר מיליגרם הסוכר בסמ"ק דם הוא משתנה מקרי שמתפלג נורמלית עם סטיית תקן 10.4 מ"ג. נלקח מדגם של 60 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרם סוכר היה 115.5 מ"ג לסמ"ק.
א. בנה רווח סמך ברמת סמך 95% לתוחלת רמת הסוכר בדם אצל הניזונים מדיאטה זו.
ב. ידוע שתוחלת רמת הסוכר בדם באוכלוסיה היא 90 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א שהדיאטה משפיעה על רמת הסוכר בדם? הסבירו.
- (3) יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצלין היא 200 מ"ג לקפסולה. משרד הבריאות ביצע מדגם של 8 קפסולות אקראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצלין לקפסולה עם סטיית תקן מדגמית של 5 מ"ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.
א. בנו רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.
ב. בדקו ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

תשובות סופיות:

- (1) נקבל השערת.
(2) א. $112.87 \leq \mu \leq 118.13$.
ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.
(3) א. $191.8 \leq \mu \leq 200.2$. ב. נכריע שיש אמת בפרסום.

סטטיסטיקה והסתברות

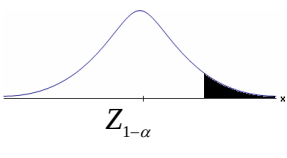
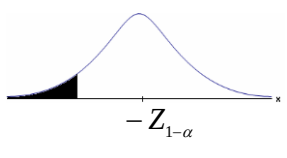
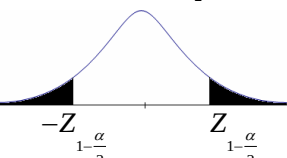
פרק 31 - בדיקת השערות על פרופורציה

תוכן העניינים

166	1. התהליך
169	2. סיכוי לטעויות ועוצמה
173	3. קביעת גודל מדגם
175	4. מובהקות התוצאה - אלפא מינימלית

התהליך:

רקע:

השערת האפס: השערה אלטרנטיבית:	$H_0: p = p_0$ $H_1: p > p_0$	$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p \neq p_0$
תנאים:	$np_0 \geq 5 \text{ \& } n(1-p_0) \geq 5$		
כלל ההכרעה: אזור הדחייה של H_0 :	$Z_{\hat{p}} > Z_{1-\alpha}$  -דוחים את H_0	$Z_{\hat{p}} < -Z_{1-\alpha}$  -דוחים את H_0	או $Z_{\hat{p}} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\hat{p}} > Z_{1-\frac{\alpha}{2}}$  -דוחים את H_0

סטטיסטי המבחן:
$$Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

חלופה אחרת לכלל הכרעה:

כלל ההכרעה – אזור הדחייה של H_0 :		
$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ $\hat{p} < p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$

דוגמה (פתרון בהקלטה):

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

שאלות:

- (1) במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. ברמת מובהקות של 5% האם השנה הקשו על תנאי הקבלה?
- (2) במדגם של 300 אזרחים 57% מתנגדים להצעת חוק מסוימת. לאור נתונים אלה האם רוב האזרחים מתנגדים להצעת החוק? בדקו ברמת מובהקות של 10%.
- (3) הטילו מטבע 50 פעמים וקיבלו 28 פעמים עץ. האם המטבע הוגן ברמת מובהקות של 5%?
- (4) קפיטריה במכללה מסוימת מעריכה כי אחוז הסטודנטים שקונים קפה בקפיטריה הינו 20%. נערך סקר אשר כלל 200 סטודנטים. התברר כי 33 מהם רוכשים קפה בקפיטריה. מטרת הסקר הייתה לבדוק את אמיתות הערכה של הקפיטריה.
 - א. רשמו את ההשערות.
 - ב. בדקו את ההשערות ברמת מובהקות של 10%.
 - ג. מה תהיה המסקנה אם נקטין את רמת המובהקות?
- (5) חבר כנסת רוצה להעביר חוק. לצורך כך הוא דוגם 400 אזרחים במטרה לבדוק האם רוב האזרחים תומכים בחוק. במדגם התקבל ש-276 אזרחים תומכים בחוק.
 - א. מה מסקנתכם ברמת מובהקות של 5%?
 - ב. האם ניתן לדעת מה תהיה המסקנה אם רמת המובהקות תהיה גדולה יותר? הסבירו.
- (6) שני חוקרים בדקו את ההשערות הבאות: $H_0: p = p_0$, $H_1: p > p_0$. חוקר א' השתמש ברמת מובהקות α_1 וחוקר ב' ברמת מובהקות α_2 החוקר הראשון דחה את H_0 ואילו החוקר השני קיבל את H_0 . שניהם התבססו על אותם תוצאות של מדגם. בחר בתשובה הנכונה:
 - א. $\alpha_1 = \alpha_2$.
 - ב. $\alpha_1 > \alpha_2$.
 - ג. $\alpha_1 < \alpha_2$.
 - ד. המצב המתואר לא אפשרי.

תשובות סופיות:

- (1) נדחה H_0 .
- (2) נדחה H_0 .
- (3) נקבל H_0 .
- (4) א. $H_0: p = 0.2$
 ב. $H_1: p \neq 0.2$
 ג. המסקנה לא תשתנה.
- (5) א. נדחה H_0 .
 ב. המסקנה לא תשתנה.
- (6) ג'.

סיכוי לטעויות ועוצמה:

רקע:

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

$$\alpha = P(H_0 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_0}(H_0)$$

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ לקבל את } H_1 \mid \text{לדחות את } H_0) = P_{H_1}(H_0)$$

רמת בטחון:

$$(1-\alpha) = P(H_0 \text{ לקבל את } H_0 \mid \text{לדחות את } H_0) = P_{H_0}(H_0)$$

עוצמה:

$$\pi = (1-\beta) = P(H_0 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_1}(H_0)$$

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

התהליך לחישוב סיכוי לטעות מסוג שני:

השערת האפס: השערה אלטרנטיבית:	$H_0: p = p_0$ $H_1: p \neq p_0$	$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p > p_0$
תנאים:	$np_0 \geq 5 \& n(1-p_0) \geq 5$		
כלל ההכרעה: אזור הדחייה של H_0 :	$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ או $\hat{p} < p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$

חישוב β :
$P_{H_1} \left(\hat{p} < p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \right)$
$P_{H_1} \left(p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p} < p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \right)$
$P_{H_1} \left(\hat{p} > p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \right)$

כאשר: $\hat{p} \sim N \left(p, \frac{p(1-p)}{n} \right)$

והתקנון: $Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

דוגמה (פתרון בהקלטה):

רופאי שיניים טוענים שיותר ממחצית האוכלוסייה הבוגרת בארץ אינם מבקרים אצל רופא שיניים באופן קבוע, כנדרש. כדי לבדוק טענה זו, נערך סקר בקרב 150 אנשים בוגרים.

א. רשמו את ההשערות וכלל הכרעה ברמת מובהקות של 10%.

ב. מהי עוצמת המבחן אם מסתבר ש 60% מהאוכלוסייה אינם מבקרים אצל רופא שיניים באופן קבוע.

שאלות:

- (1) משרד הבריאות פרסם ש-10% מתושבי המדינה סובלים ממחלת האסטמה. מחקר דורש לבדוק האם בחיפה, בגלל זיהום האוויר, שיעור הסובלים מאסטמה גבוה יותר. לצורך המחקר נבדקו 260 מתושבי חיפה.
 - א. רשמו את השערות המחקר, וצרו מבחן ברמת מובהקות של 5% לבדיקתן.
 - ב. מהי עצמת המבחן של סעיף א' בהנחה ובחיפה 16% מהתושבים סובלים מאסטמה?
 - ג. כיצד תשנה התשובה לסעיף ב' אם מסתבר שבחיפה 18% סובלים מאסטמה?
 - ד. בהמשך לסעיף א' האם נכון לומר שבהסתברות של 5% ההשערה שבחיפה 10% מהתושבים סובלים מאסטמה אינה נכונה?
- (2) אחוז הסובלים מתופעות הלוואי מתרופה מסוימת הוא 15%. חברת תרופות טוענת שפיתחה תרופה שאמורה לצמצם את אחוז הסובלים מתופעות לוואי. לצורך בדיקת הטענה הוחלט לבצע מחקר שיכלול 120 חולים שיקבלו את התרופה הנבדקת.
 - נניח שהתרופה נבדקת אכן מורידה את פרופורציות הסובלים מתופעות הלוואי ל-10%, מהי עצמת המבחן עבור רמת מובהקות של 5%?
- (3) בעיר מסוימת היו 20% אקדמאים. בעקבות פתיחת מכללה בעיר לפני כמה שנים מעוניינים לבדוק האם אחוז האקדמאים גדל. מעוניינים שהמחקר יכלול 200 אנשים והוא יהיה ברמת מובהקות של 5%.
 - א. חשבו את הסיכוי לבצע טעות מסוג שני בהנחה והיום יש 28% אקדמאים.
 - ב. כיצד התשובה לסעיף הקודם תשתנה אם נגדיל את רמת המובהקות?
- (4) 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?
- (5) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן (בחרו בתשובה הנכונה):
 - א. השערת האפס נכונה.
 - ב. השערת האפס נדחתה.
 - ג. השערת האפס לא נדחתה.
 - ד. אף אחת מהתשובות לא נכונה בהכרח.
- (6) קבעו אם הטענה הבאה נכונה: בבדיקת השערות לא ניתן לבצע בו זמנית טעות מסוג ראשון וטעות מסוג שני.

תשובות סופיות:

- (1) א. $H_0 : p = 0.1$
 ב. 0.9015 ג. תגדל. ד. טענה לא נכונה.
 $H_1 : p > 0.1$
- (2) 0.4404
- (3) א. 0.1446 ב. תקטן.
- (4) חוקר א'.
- (5) ג'.
- (6) נכונה.

קביעת גודל מדגם:

רקע:

השערות המחקר הן: $H_0: p = p_0$, $H_1: p = p_1$: מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β .

הנוסחה הבאה נותנת את גודל המדגם הרצוי:

$$n \geq \left(\frac{Z_{1-\alpha} \sqrt{p_0 q_0} + Z_{1-\beta} \sqrt{p_1 q_1}}{p_0 - p_1} \right)^2$$

דוגמה (פתרון בהקלטה):

רוצים לבדוק האם אחוז האנשים השוהים בשמש ללא הגנה ירד בעקבות הפרסומת על נזקי השמש.
 בעבר 60% מהאוכלוסייה שהתה בשמש ללא הגנה. מה גודל המדגם המינימלי שיש לקחת כדי לבדוק שהאחוז הנ"ל ירד ל-48% אם מעוניינים שהסיכוי לטעות מסוג ראשון יהיה 5% והסיכוי לטעות מסוג שני יהיה 1%?

שאלות:

- (1) משרד התמ"ת פרסם שאחוז האבטלה במשק היום עומד על 8%. לעומתו, משרד הפנים טוען שחלה עלייה בשיעור האבטלה עד לכדי 11%. כדי לבדוק מי מבניהם צודק, מה צריך להיות גודל המדגם שיענה על שני התנאים הבאים:
- א. אם משרד התמ"ת צודק, נדחה את טענתו בסיכוי של 10%.
- ב. אם משרד הפנים צודק, נדחה את טענתו בסיכוי של 4%.
- (2) מפעיל קזינו מפרסם שהסיכוי לזכות במכונת מזל הינו 0.42. אדם טוען שהסיכויים לזכות במשחק נמוכים יותר. כמה פעמים יש לשחק את המשחק כדי שאם טענת מפעיל הקזינו נכונה נקבל את טענת האדם בסיכוי של 1% ואם במציאות הסיכוי לזכות במכונה הוא 0.3 נקבל את מפעיל הקזינו בסיכוי של 8%?

תשובות סופיות:

(1) .891

(2) .224

מובהקות התוצאה – אלפא מינימלית:

רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה: באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v . את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות. המסקנה של המחקר תקבע לפי העיקרון הבא:

אם $p_v \leq \alpha$ דוחים את H_0 .

מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

לקבל את תוצאות המדגם וקיצוני $p_v = P_{H_0}$.

אם ההשערה היא דו צדדית:

לקבל את תוצאות המדגם וקיצוני $p_v = 2P_{H_0}$.

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס: השערה אלטרנטיבית:	$H_0 : p = p_0$ $H_1 : p > p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p \neq p_0$
תנאים:	$np_0 \geq 5 \& n(1-p_0) \geq 5$		
p-value	$P_{H_0}(\hat{p} \geq \hat{p})$	$P_{H_0}(\hat{p} \leq \hat{p})$	אם $2 \cdot P_{H_0}(\hat{p} \geq \hat{p}) \Leftarrow \hat{p} > p_0$ אם $2 \cdot P_{H_0}(\hat{p} \leq \hat{p}) \Leftarrow \hat{p} < p_0$

כאשר בהנחת השערת האפס: $\hat{P} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$

התקנון: $Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

דוגמה (פתרון בהקלטה):

ישנה טענה שיש הבדל בין אחוז הבנים ואחוז הבנות הפונים ללמוד להנדסאי מחשבים. לשם כך נלקח מדגם מקרי של 200 תלמידים הלומדים מחשבים והתברר כי 112 מהם בנים.

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

שאלות:

- (1) במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. רוצים לבדוק האם השנה הקשו על תנאי הקבלה.
א. מהי מובהקות התוצאה?
ב. מה תהיה המסקנה ברמת מובהקות של 1% וברמת מובהקות של 5%?
- (2) נהוג לחשוב ש-60% מהילדים בגיל שלוש קמים מהמיטה במהלך הלילה לפחות פעם אחת. ישנה טענה שללא שנת צהריים פחות מ-60% מהילדים בגיל זה יקומו לפחות פעם אחת במהלך הלילה. נדגמו 80 ילדים בגיל 3 אשר אינם ישנים בצהריים מתוכם התקבל ש-41 קמו במהלך הלילה.
א. מהי רמת המובהקות המינימלית עבורה תתקבל הטענה במחקר?
ב. מהי רמת המובהקות המקסימלית עבורה לא תתקבל טענת המחקר?
ג. עבור אילו רמות מובהקות נקבל את טענת המחקר?
ד. מה תהיה מסקנת המחקר ברמת מובהקות של 6%?
- (3) במטרה לבדוק האם מטבע הוא הוגן מטילים אותו 80 פעמים. התקבל ש-60 מההטלות הראו עץ. רשמו את השערות המחקר, חשבו את מובהקות התוצאה והסיקו מסקנה ברמת מובהקות של 5%.
- (4) בבדיקת השערות על פרופורציה התקבל שה- $p\text{-value} = 0.02$.
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 5%:
(בחרו בתשובה הנכונה)
א. יקבל את השערת האפס
ב. ידחה את השערת האפס.
ג. לא ניתן לדעת כי אין מספיק נתונים.
- (5) קבעו אם הטענה הבאה נכונה:
"במבחן לבדיקת השערות חד-צדדי התקבל ערך $p\text{-value}$ של 3%,
לכן אם היינו מבצעים מבחן דו-צדדי (כאשר יתר הנתונים ללא שינוי),
היינו מקבלים ערך $p\text{-value}$ של 6%".
- (6) במפעל 10% מהעובדים נפגעים לפחות פעם אחת בשנה מתאונות עבודה. לאור זאת, המפעל החליט לצאת בתוכנית לצמצום שיעור הנפגעים. תכנית זו נוסתה על 100 עובדים. מתוכם 12 נפגעו בתאונות עבודה במשך השנה. מהי רמת המובהקות הקטנה ביותר עבורה יוחלט שהתכנית יעילה?

תשובות סופיות:

- (1) א. 0.0455.
 ב. ברמת מובהקות של 1% : לא דוחים את H_0 .
 ברמת מובהקות של 5% : נדחה את H_0 .
- (2) א. 0.0548. ב. 0.0548. ג. מעל 0.0548.
 ד. נכריע לטובת טענת המחקר.
- (3) $p_v = 0$, נדחה את H_0 .
- (4) ב'.
- (5) הטענה נכונה.
- (6) 0.7486.

סטטיסטיקה והסתברות

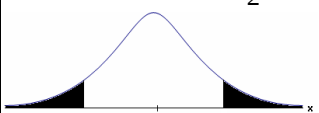
פרק 32 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

- 1. כששונויות האוכלוסייה ידועות.....179
- 2. כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות.....183
- 3. ניתוח פלטים.....187

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונות של האוכלוסייה ידועות – רקע

$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 > c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 < c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 \neq c$	השערת האפס: השערה אלטרנטיבית:
מדגמים בלתי תלויים σ_1, σ_2 ידועות $X_1, X_2 \sim N$ או מדגמים מספיק גדולים			תנאים:
$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ דוחים את H_0	או $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$ דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0 : $-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{סטטיסטי המבחן:}$$

חלופה אחרת לכלל הכרעה:

נדחה H_0 אם מתקיים:	
$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ או $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	

התפלגות הפרש הממוצעים: $\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{התקנון:}$$

דוגמה (פתרון בהקלטה):

בשנת 2004 הפער בין השכר הממוצע של הגברים לנשים היה 3000 ₪ לטובת הגברים. מעוניינים לבדוק האם כיום הצטמצם הפער בין הגברים לנשים מבחינת השכר הממוצע. נדגמו 100 עובדים גברים. שכרם הממוצע היה 9,072 ₪. נדגמו 80 עובדות, שכרן הממוצע היה 7809 ₪. לצורך פתרון נניח שסטיות התקן של השכר ידועות ושוות ל-2000 ₪ באוכלוסיית הנשים ו-3000 ₪ באוכלוסיית הגברים. מה המסקנה ברמת מבוהקות של 5%?

שאלות

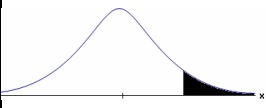
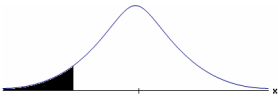
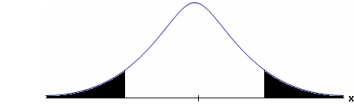
- (1) מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלה נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה. במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות. במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות. לצורך פתרון הניחו שבכל אזור, סטיית התקן היא שעה 1 ביום. בדקו את טענת המחקר ברמת מובהקות של 1%.
- (2) ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 100. מכון ללימוד פסיכומטרי טוען שהוא יכול לשפר את ממוצע הציונים ביותר מ-30 נקודות. במדגם של 20 נבחנים שניגשו למבחן ללא הכנה במכון התקבל ממוצע 508. במדגם של 25 נבחנים שעברו הכנה במכון התקבל ממוצע ציונים 561. מה מסקנתכם ברמת מובהקות של 5%.
- (3) במדגם אקראי של 20 ימים נבדקה התפוקה של מפעל ביום. התפוקה הממוצעת הייתה של 340 מוצרים ליום. במדגם אקראי של 20 ימים אחרים נבדקה התפוקה של המפעל בלילה והתפוקה הממוצעת הייתה 295. לצורך פתרון נניח שסטיית התקן של התפוקה ביום היא 40 מוצרים ובלילה 30 מוצרים.
 - א. מהי מובהקות התוצאה לבדיקה האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית.
 - ב. מה תהיה המסקנה ברמת מובהקות של 8%?
- (4) במחקר מקיף שנעשה באירופה נקבע שגברים גבוהים מנשים ב-8 ס"מ בממוצע. מחקר ישראלי מתעניין לבדוק האם בישראל הפער גדול יותר. לצורך המחקר נדגמו 40 גברים ו-40 נשים באקראי. כמו כן, נניח שסטיות התקן של הגברים והנשים ידועות ושוות ל-6 ס"מ אצל הנשים ו-12 ס"מ אצל הגברים.
 - א. מהן השערות המחקר ומהו כלל ההכרעה ברמת מובהקות של 10%?
 - ב. אם בישראל הפער בין גברים לנשים מבחינת הגובה הממוצע הוא 11 ס"מ, מה ההסתברות שהמחקר לא יגלה זאת? איך קוראים להסתברות הזאת?

תשובות סופיות

- (1) נדחה H_0 .
- (2) לא נדחה את H_0 .
- (3) א. 0 ב. נדחה את H_0 .
- (4) א. נדחה את H_0 , אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותר מ-10.72 ס"מ.
ב. 0.6331

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

השערת האפס: השערה אלטרנטיבית: תנאים:	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 > c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 < c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 \neq c$
1. מדגמים בלתי תלויים 2. σ_1, σ_2 לא ידועות אך שוות 3. המשתנים בכל אוכלוסייה מתפלגים נורמלית	$t_{\bar{x}_1 - \bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  $t_{1-\alpha, n_1+n_2-2}$ דוחים את H_0	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$  $-t_{1-\alpha, n_1+n_2-2}$ דוחים את H_0	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ או $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  $-t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ $t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ דוחים את H_0

$$t_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

סטטיסטי המבחן:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

השונות המשוקללת:

חלופה אחרת לכלל הכרעה:

נדחה H_0 אם מתקיים:	
$\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ או $\bar{x}_1 - \bar{x}_2 < c - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$
$\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	

דוגמה (פתרון בהקלטה):

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשתמשים בה כיום לבניית בניינים. לצורך בדיקת טענת המחקר נדגמו 10 יחידות של מתכות מהסוג הישן ו-12 יחידות של מתכות מהסוג החדש. להלן תוצאות המדגם:

טמפרטורת ההתכה הממוצעת במתכת הישנה 1170 מעלות עם אומד חסר הטיה לשונות $S^2 = 200$.

טמפרטורת ההתכה הממוצעת במתכת החדשה 1317 מעלות עם אומד חסר הטיה לשונות $S^2 = 260$.

נניח לצורך פתרון שטמפרטורת ההתכה מתפלגת נורמאלית עם אותה שונות במתכות השונות. בדקו ברמת מובהקות של 5%.

שאלות

- (1) להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (במ"ר):

120	94	90	130	95	112	120	2012
	69	74	105	91	82	100	2013

בדקו שבשנת 2013 הייתה ירידה משמעותית בשטחי הדירות לעומת שנת 2012 עבור רמת מובהקות של 5%.
הניחו ששטחי הדירות בכל שנה מתפלגים נורמלית עם אותה שונות.

- (2) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

המדינה	ישראל	ארה"ב
גודל המדגם	15	15
סכום הציונים	1560	1470
סכום ריבועי הציונים	165,390	147,560

בדקו ברמת מובהקות של 5% האם קיים הבדל של נקודה בין ישראלים לאמריקאים מבחינת ממוצע הציונים במבחן ה-IQ לטובת ישראל.
רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

- (3) להלן תוצאות מדגם הבדק אורך חיים של נורות מסוג W60 ומסוג W100. אורך החיים נמדד בשעות.

100W	60W	הקבוצה
956	1007	$\bar{X}$
72	80	S
15	13	n

- א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפתרון.
ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג 100?
ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.

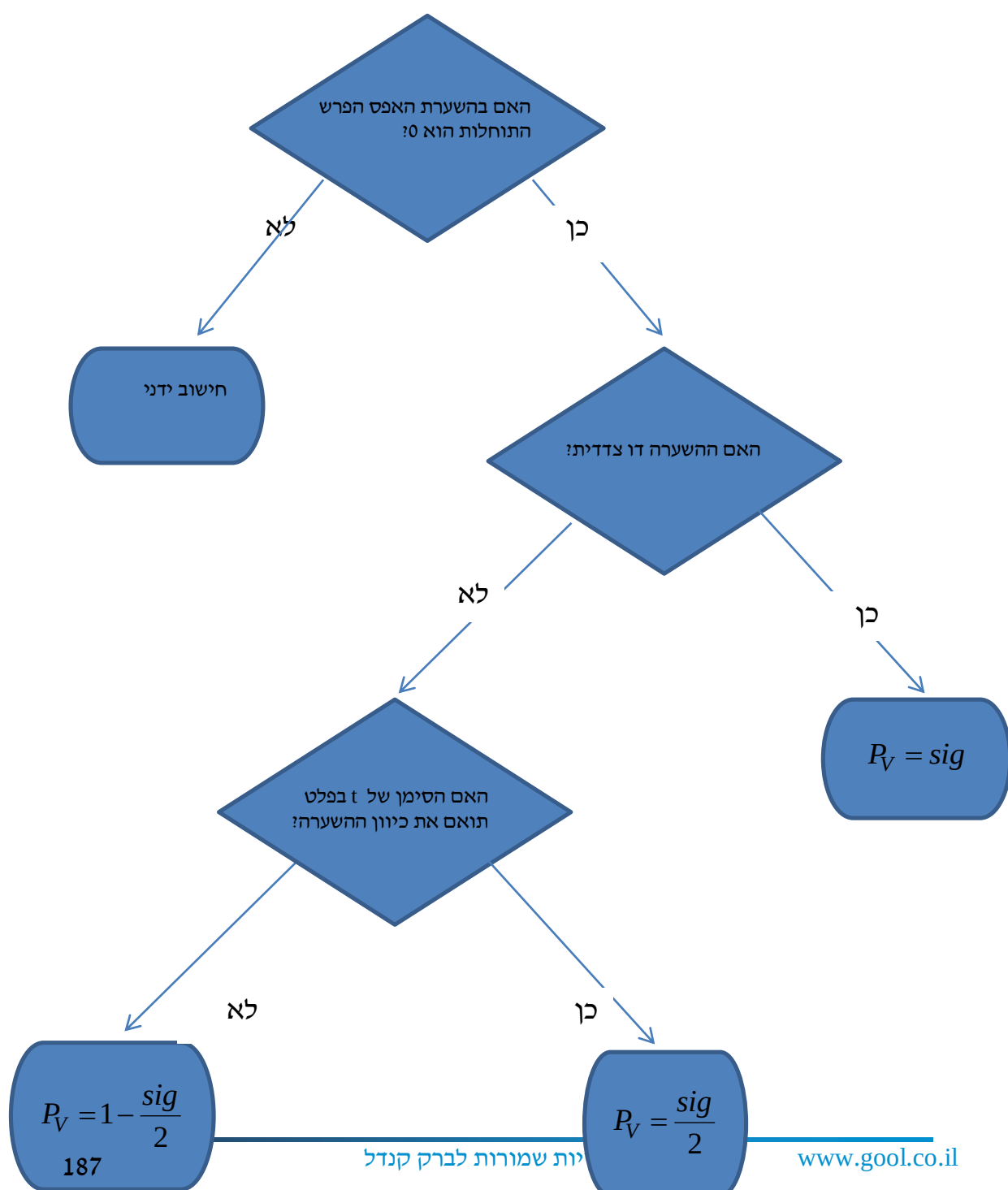
תשובות סופיות

- (1) נדחה את H_0 .
- (2) הנחות:
1. סטיות התקן שוות.
 2. המשתנים מתפלגים נורמלית.
- נקבל את H_0 .
- (3) א. נדחה את H_0 . ב. רמת מובהקות של לפחות 5%.
- ג. לא נדחה את H_0 .

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

ניתוח פלטים – רקע

מובהקות התוצאה על סמך הפלט:



דוגמה (פתרון בהקלטה) :

בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים משני אזורים שונים במדינה על מס' האחים והאחיות שלהם. להלן הפלט שהתקבל :

Group Statistics

	Region of the United States	N	Mean	Std. Deviation	Std. Error Mean
Number of Brothers and Sisters	North East	676	3.76	2.939	.113
	South East	410	4.05	2.993	.148

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Number of Brothers and Sisters	Equal variances assumed	.173	.677	-1.583	1084	.114	-.293	.185	-.657	.070
	Equal variances not assumed			-1.576	850.945	.115	-.293	.186	-.658	.072

- א. מהו המבחן הסטטיסטי שנעשה כאן?
- ב. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?
- ג. בדוק האם קיים הבדל בין "South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.
- ד. מהי מובהקות התוצאה לבדיקת הטענה שההפרש הממוצע בין "South East" לבין "North East" חיובי?

שאלות

- 1) להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים. רמת האופטימיות נמדדה בסולם ציונים של 1 עד 5.

Group Statistics

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
optimizm	MALE	633	2.6053	.49781	.01979
	FEMALE	568	2.5503	.48483	.02034

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	1190.977	.053	.05500	.02838	-.00068	.11067

- א. האם ניתן להניח ששוונות האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?
- ב. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?
- ג. מצא את הגבול העליון של רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.
- ד. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

(2) פסיכולוגים טוענים שאנשים שניגשים למבחן אינטליגנציה יותר מפעם אחת נוטים לקבל ציונים גבוהים יותר. להלן הפלט שהתקבל:

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
grade	A	9	96.8889	9.40006	3.13335
	B	11	108.4545	11.46616	3.45718

Independent Samples Test										
Levene's Test for Equality of Variances				t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
grade	Equal variances assumed	.206	.656	-2.428	18	.026	-11.56566	4.76333	-21.57304	-1.55828
	Equal variances not assumed			-2.479	17.997	.023	-11.56566	4.66583	-21.36832	-1.76299

T-Test

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.

ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?

ג. האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק

ברמת מובהקות של $\alpha = 0.05$.

ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha = 0.01$.

3) כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל בממוצע גיל הנישואין (הראשונים), בין נשים הגרות בערים מרכזיות לבין נשים הגרות בערים מרוחקות מהמרכז. לשם כך נדגמו 50 כלות מכל אחת משתי ערים עיר א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שלהלן:

T-Test

Group Statistics

מקום המגורים	N	Mean	Std. Deviation	Std. Error Mean
גיל הנישואין עיר א	50	24.8072	1.38978	.19654
עיר ב	50	23.0131	1.62070	.22920

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
גיל הנישואין	Equal variances assumed	.330	.567	5.942	98	.000	1.79415	.30193	1.19497	2.39332
	Equal variances not assumed			5.942	95.772	.000	1.79415	.30193	1.19480	2.39350

- א. מהו המבחן הסטטיסטי שנעשה כאן?
- ב. מצא רווח סמך ברמת סמך של 95% להפרש בין עיר א לעיר ב מבחינת גיל הנשים הממוצע בנישואין הראשונים.
- ג. האם ניתן לומר ברמת מובהקות של 1% שנשים בערים מרכזיות מתחתנות בגיל מאוחר יותר מאשר נשים הגרות בערים מרוחקות?

4) להלן פלט של תוכנת SPSS :

T-Test

	N	Mean	Std. Deviation	Std. Error Mean
X	26	36.3077	13.23259	2.59513
Y	24	46.4583	20.96369	4.27920

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2- tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	4.446	.040	-2.164	???	.044	-10.15064	???	-20.03781	-.26347
Equal variances not assumed			-2.038	38.267	.048	???	5.00462	-20.27964	-.02164

א. השלימו את סימני השאלה בטבלה.

ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של X לזה של Y ?

ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X גדולה מהתוחלת של Y ?

ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X קטנה מהתוחלת של Y ?

תשובות סופיות

- (1) א. נקבל את H_0 ונכריע שיש שוויון שוניות.
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.
 ג. 0.11076
 ד. $2.5665 \leq \mu \leq 2.6441$
- (2) א. מבחן T להפרש ממוצעים במדגמים בלתי תלויים.
 ב. מבחן T למדגמים מזווגים.
 ג. נקבל את H_0 , נקבע לקיום שוויון שוניות.
 ד. נקבל את H_0 , לא נקבל את טענת הפסיכולוגים.
- (3) א. מבחן T להשוואת תוחלת במדגמים בלתי תלויים.
 ב. $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$ ג. כן.
- (4) א. 10.15, 4.69, -48
 ב. 0.04
 ג. 0.978
 ד. 0.022

סטטיסטיקה והסתברות

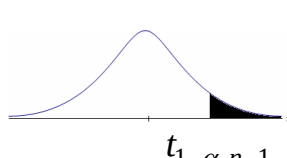
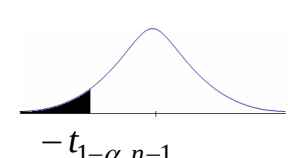
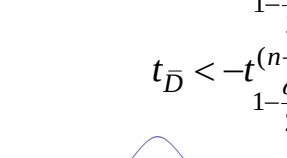
פרק 33 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

תוכן העניינים

194	1. בדיקת השערות למדגמים מזווגים
198	2. ניתוח פלטים

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

בדיקת השערות למדגמים מזווגים – רקע

$H_0: \mu_D = C$ $H_1: \mu_D > C$	$H_0: \mu_D = C$ $H_1: \mu_D < C$	$H_0: \mu_D = C$ $H_1: \mu_D \neq C$	השערת האפס: השערה אלטרנטיבית:
1. σ_D אינה ידועה 2. $D \sim N$ או מדגם מספיק גדול			תנאים:
$t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{D}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{D}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ - דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$, $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}, \quad t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

סטטיסטי המבחן:

דוגמה (פתרון בהקלטה):

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחירה נמוכים מהמחירים מרשת השיווק "שופרסל". לצורך הבדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות. להלן המחירים:

המוצר / רשת	מגה בעיר	שופרסל
שמפו	17	18
ג'ל כביסה	48	57
עוגת גבינה	35	35
לחם	12	10
קפה נמס	49	47
בקבוק יין	113	142
גבינה בולגרית	20	26

בהנחה והמחירים מתפלגים נורמאלית, בדקו ברמת מובהקות של 5% את טענת רשת "מגה בעיר".

שאלות

- (1) במטרה לבדוק האם קיים הבדל בין חברת X לחברת Y מבחינת המחירים לשיחות בינ"ל. נדגמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

חברה/מדינה	ארה"ב	קנדה	הולנד	פולין	מצרים	סין	יפן
X	1.5	2.1	2.2	3	3.5	3.2	4.2
Y	1.4	2	1.9	3.1	3.2	3.2	4.2

בהנחה והמחירים מתפלגים נורמלית בכל חברה, בדקו ברמת מובהקות של 5% האם קיים הבדל בין החברות מבחינת המחירים במוצא?

- (2) מכון המכין לפסיכומטרי טוען שהוא מעלה את ממוצע הציונים ביותר מ-30 נקודות. 8 נבחנים נבדקו לפני ואחרי שהם למדו במכון. להלן התוצאות שהתקבלו:

לפני	506	470	420	640	670	390	500	590
אחרי	570	540	430	610	680	510	520	580

מה מסקנתכם ברמת מובהקות 5%? הניחו שציוני פסיכומטרי מתפלגים נורמלית.

- (3) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן הציונים שלהם בסמסטר א' ו- ב':

סטטיסטיקה א'	74	68	90	75	82
סטטיסטיקה ב'	80	84	87	76	100

פורסם שתלמידים שמסיימים את סמסטר ב' משפרים במוצא את הציונים ב-5 נקודות לעומת סמסטר א'. הניחו שהציונים מתפלגים נורמלית.

- א. מהי מובהקות התוצאה לבדיקת הטענה שהשיפור הוא יותר מ-5 נקודות?
 ב. על סמך הסעיף הקודם, מהי רמת המובהקות המינימלית להכרעה שהשיפור הוא יותר מ-5 נקודות?
 ג. לאור זאת, מה המסקנה ברמת מובהקות של 10%?

- (4) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמלית ללא ידיעת השונות האמתית. המבחן שיש לבצע כאן הוא:

- א. מבחן Z למדגם יחיד.
 ב. מבחן T למדגם יחיד.
 ג. מבחן T למדגמים בלתי תלויים.
 ד. מבחן T למדגמים מזווגים.

- (5) בתחנת טיפת חלב מסוימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשיים נשקל בכל אחד מהמשקלים. להלן תוצאות השקילה (בק"ג):

משקל במכשיר 1	2.5	0.7	9.6	4.5
משקל במכשיר 2	0.5	1.7	6.9	3.5

נניח שהמשקלים מתפלגים נורמלית, המבחן שיש לבצע כאן הוא:

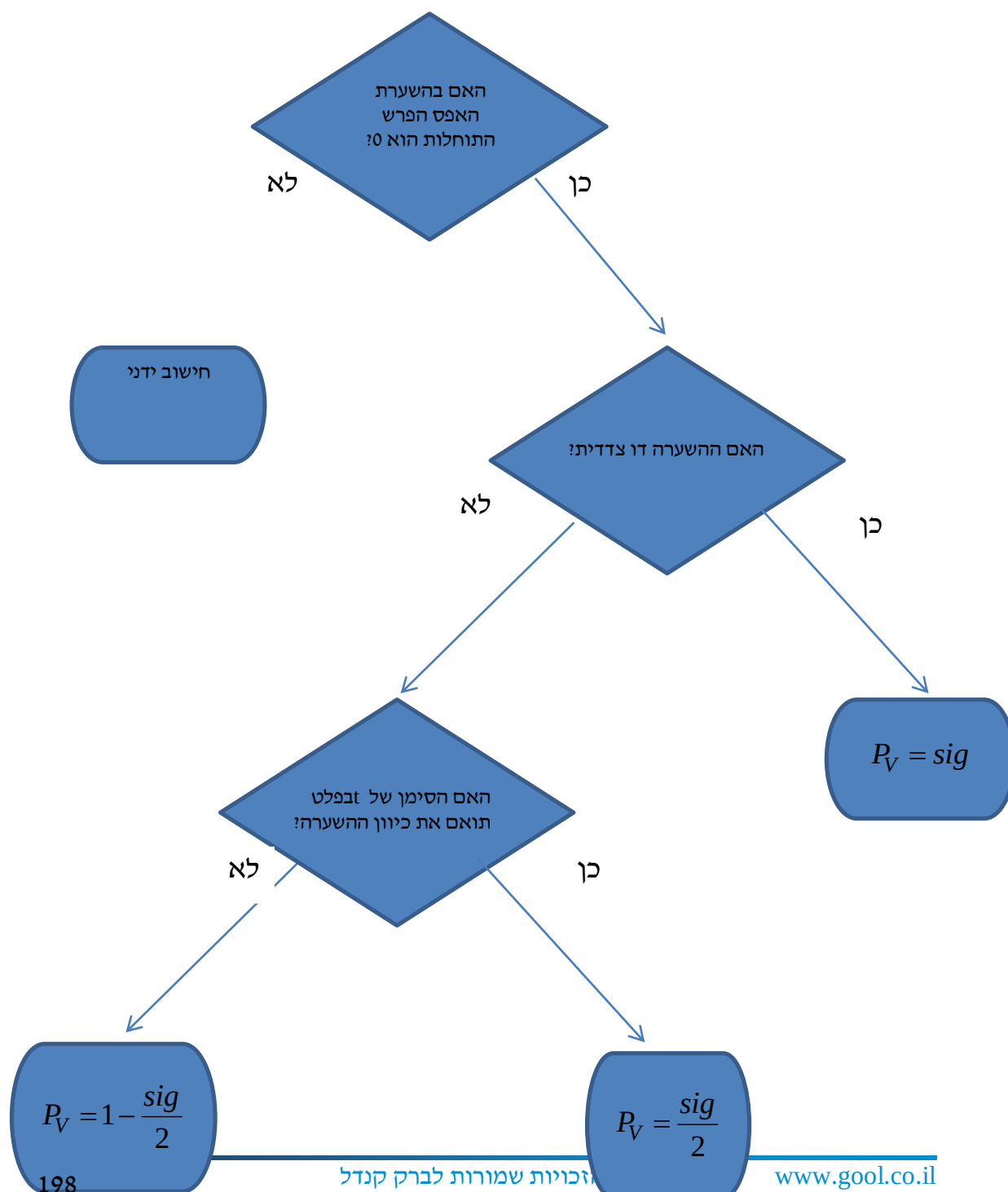
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.
- (6) כדי להשוות בין שני אצנים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים. המבחן שיש לבצע כאן הוא:
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.

תשובות סופיות

- (1) לא נדחה H_0 .
- (2) לא נדחה H_0 .
- (3) א. $0.25 \leq p \leq 0.5$ ב. 0.5 ג. לא נדחה H_0 .
- (4) ד'.
- (5) ד'.
- (6) ג'.

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

מדגמים מזווגים – ניתוח פלטים – רקע



דוגמה (פתרון בהקלטה) :

כדי לבדוק את ההשפעה של קורס לגמילה מעישון נלקח מדגם מקרי של 5 נבדקים. עבור כל אחד מהם נמדדה צריכת הסיגריות היומית לפני הקורס וחודשיים אחריו. הניחו שצריכת הסיגריות מתפלגת נורמלית. להלן התוצאות :

נבדק	1	2	3	4	5
לפני	40	22	25	28	30
אחרי	30	24	13	10	12

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	BEFORE	29.0000	5	6.85565	3.06594
	AFTER	17.8000	5	8.72926	3.90384

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	90% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	BEFORE - AFTER	11.20000	8.19756	3.66606	3.38452	19.01548	3.055	4	.038

בדקו ברמת מובהקות של 5% האם הקורס יעיל.

שאלות

- 1) בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים על השכלת הוריהם, להלן הפלט שהתקבל:

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Highest Year School Completed, Father - Highest Year School Completed, Mother	-.007	3.115	.100	-.203	.189	-.072	973	.943

- א. תנו אומדן להפרש הממוצעים.
 ב. תנו אומדן לטעות התקן של הפרש הממוצעים.
 ג. האם קיים הבדל מובהק בין השכלת האבות להשכלת האימהות ברמת מובהקות של 5%?

- 2) בתחרות קפיצה למים שופטים באופן קבוע שופט איטלקי ושופט דרום קוריאני. להלן פלט המנתח את הציונים ששופטים אלה נתנו בתחרויות השונות:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Italy	???	300	.86742	.05008
	South Korea	8.9183	???	.81992	.04734

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Italy - South Korea	-.42233	.36153	.02087	-.46341	-.38126	-20.234	???	???

- א. השלימו את החלקים החסרים בפלט (מסומנים בסימני שאלה).
 ב. בדקו את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.
 ג. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים?
 ד. בנו את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.

3) בדקו את ציוניהם של 44 נבדקים אקראיים במבחן הפסיכומטרי. פעם אחת לפני הכנה (Before) ופעם אחת אחרי הכנה (After).

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Before - After	-7.45455	19.28303	2.90703	-13.31712	-1.59197	-2.564	43	.014

- א. רשמו מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.
- ב. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.
- ג. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.
- ד. מצאו רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

(4) להלן פלט של תכנת SPSS :

T-Test

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 x	54.0000	6	5.86515	2.39444
y	46.5000	6	10.72847	4.37988

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 x - y	7.50000	??	4.72405	-4.64356	19.64356	??	5	.173

- א. מלא את החלקים החסרים בטבלה.
- ב. מהי רמת המובהקות המינימלית לקבלת הטענה שיש הבדל בין X ל- Y בממוצע?
- ג. האם התשובה לסעיף הקודם הייתה משתנה, ואם כן גדלה או קטנה, אם הינו מוסיפים עוד תצפית שההפרש בין X ל- Y הוא 0.
- ד. מהי מובהקות התוצאה לבדיקת הטענה ש X גדול מ- Y בממוצע?
- ה. מהי מובהקות התוצאה לבדיקת הטענה ש X קטן מ- Y בממוצע?
- ו. בנו רווח סמך לתוחלת של X ברמת סמך של 90%.

תשובות סופיות

- (1) א. -0.007 ב. 0.1 ג. אין הבדל מובהק.
- (2) א. $d.f = 299$, $n = 300$, $\bar{X} = 8.496$, $\text{Sig} = 0$.
- ב. נדחה את H_0 . ג. (0.46341, 0.38126) ד. (0.4708, 0.3738) .
- (3) א. ראה וידאו. ב. נדחה את H_0 . ג. לא נדחה את H_0 .
- ד. (1.592, 13.317) .
- (4) א. 1.5876 , 11.5715 ב. 0.173 ג. יגדל.
- ד. 0.0865 ה. 0.9135 ו. $49.18 < \mu < 58.82$.

סטטיסטיקה והסתברות

פרק 34 - הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות

תוכן העניינים

1. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות 204

הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α על $\mu_1 - \mu_2$:

$$H_0: \mu_1 - \mu_2 = C, \quad H_1: \mu_1 - \mu_2 \neq C$$

על ידי בניית רווח סמך ברמת סמך של $1 - \alpha$ ל- $\mu_1 - \mu_2$:

אם C נופל ברווח $\leftarrow$ נקבל את H_0 .

אם C לא נופל ברווח $\leftarrow$ נדחה את H_0 .

דוגמה (פתרון בהקלטה) :

חוקר ביצע בדיקת השערות לתוחלת ההפרש במדגם מזווג.

להלן השערותיו: $\alpha = 5\%$, $H_0: \mu_D = 80$, $H_1: \mu_D \neq 80$.

החוקר בנה רווח סמך ברמה של 90%, $78 < \mu_D < 83$.

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

שאלות

- (1) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'.
להלן ציוניהם בסמסטר א' ו- ב' :

סמסטר א	סמסטר ב
74	80
68	84
90	87
75	76
82	100

- א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א' לבין סמסטר ב'.
- ב. פורסם שתלמידים שמסיימים את סמסטר ב משפרים בממוצע את הציונים ב-5 נק' לעומת סמסטר א'. האם יש אמת בפרסום?

- (2) הוחלט להשוות הציונים אצל מרצה X ואצל מרצה Y. נבחרו באקראי 6 סטודנטים, 3 סטודנטים של מרצה X ו-3 סטודנטים של מרצה Y, עבורם התקבלו הציונים הבאים :

68	90	82	מרצה X
64	81	68	מרצה Y

- א. חשבו רווח סמך ברמת סמך 90% להפרש בין התוחלות של הציונים אצל שני המרצים.
- ב. האם ברמת מובהקות של 10% נכריע שיש הבדל בין תוחלות הציונים אצל שני המרצים?

שאלות רב-ברירה :

- (3) סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים.
- הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח $-2 < \mu_1 - \mu_2 < 4.5$. אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות :
- $H_0 : \mu_1 - \mu_2 = 0$; $H_1 : \mu_1 - \mu_2 \neq 0$ ברמת מובהקות 0.05, מסקנתו תהיה :
- א. לדחות את השערת האפס.
- ב. לא לדחות את השערת האפס.
- ג. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.
- ד. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

- (4) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו : $-0.0293 < \mu_D < 0.2145$, רווח הסמך הוא ברמת סמך של 95%.
לכן מסקנת המחקר היא :
- ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
 - ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
 - לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D .

תשובות סופיות

- (1) א. $-3.8 \leq \mu_D \leq 19$ ב. נכריע שיש אמת בפרסום.
- (2) א. $-8.5 \leq \mu_X - \mu_Y \leq 26.5$ ב. נכריע שאין הבדל.
- (3) ג'.
- (4) א'.

סטטיסטיקה והסתברות

פרק 35 - שאלות מסכמות בבדיקת השערות

תוכן העניינים

1. שאלות פתוחות מסכמות.....207
2. שאלות רב ברירה (אמריקאיות).....210

שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות

- (1) שני חוקרים נתבקשו לבדוק את ההשערות הבאות: $H_0: \mu = 520$, $H_1: \mu > 520$. כל חוקר בדק מדגם של 225 נחקרים. ידוע ש- $\sigma = 20$. חוקר א' קבע את כלל ההכרעה לפי $\alpha = 0.05$. חוקר ב' מחליט לדחות H_0 אם $\bar{X} > 522$.

- למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?
- מהי ההסתברות לטעות מסוג שני של חוקר ב' עבור $\mu_1 = 525$?
- הסבר ללא חישוב נוסף, האם ההסתברות לטעות מסוג שני עבור $\mu_1 = 525$, של חוקר א' שווה/קטנה/גדולה לזו של חוקר ב'?
- חוקר א' קיבל במדגם שלו $\bar{X} = 523$. מהי מסקנתו?

- (2) ידוע כי תוחלת מספר הלייקים היומי של דנה היא 12 עם סטיית תקן 5. דני טוען שהוא יותר פופולארי מדנה בכך שהוא מקבל יותר לייקים מדנה ביום. על-מנת לבדוק זאת ספר דני כמה לייקים הוא קיבל בכל יום במהלך 7 שבועות (כלומר, ב-49 ימים) וקיבל סך-הכול 637 לייקים. נניח כי סטיית התקן של מספר הלייקים שדני מקבל ביום זהה לסטיית התקן של דנה.
- מהי רמת המובהקות שכדאי לדני לדרוש, כדי שדנה תשתכנע בצדקת טענתו (שדני פופולרי יותר בכך שהוא מקבל יותר לייקים מדנה ביום)?
 - אם דני משער שתוחלת מספר ה"לייקים" שהוא מקבל ביום היא 14 וקובע רמת מובהקות 2.5%, מהי עוצמת המבחן של דני?

מוצר / רשת	A	B
1	5	5
2	4	5
3	5	3
4	7	4

- (3) ברצוננו להשוות בין רשתות A לבין B. לשם כך בחרנו 4 מוצרים, ובדקנו את מחיריהם בשתי הרשתות. להלן התוצאות:
- הניחו כי המחירים מתפלגים נורמלית.
- אם יש הנחות נוספות כדי לבצע את המבחן הפרמטרי רשמו אותן.
- בדקו האם קיים הבדל בין הרשתות מבחינת תוחלת המחירים. רמת מובהקות של 5%.
 - חזרו על הסעיף הקודם בהנחה ונבחרו בכל רשת מוצרים באקראי ולא בהכרח אותם מוצרים.

- 4) במטרה לבדוק האם סטודנטים הלומדים במכללות משקיעים יותר זמן ללימודים מאשר סטודנטים באוניברסיטאות נדגמו 12 סטודנטים ובדקו לכל סטודנט את הזמן שהוא משקיע ביום ללימודים. הזמנים נמדדו בדקות:

סטודנטים באוניברסיטאות	176	156	189	171	140	180
סטודנטים במכללות	180	190	191	186	204	150

- א. נסחו את ההשערות ובדוק אותן ברמת מובהקות של 5%. רשום את כלל ההכרעה ואת ההנחות הדרושות לביצוע המבחן הפרמטרי.
- ב. חשבו את p-value.
- ג. ישנה טענה שממוצע זמן ההשקעה בלימודים במכללות הוא 3.5 שעות ביום. בדוק את הטענה כאשר רמת המובהקות הינה 5%.

- 5) בשנת 2000 ל-60% היה מדיח כלים בבית. מחקר רוצה לבדוק האם כיום פרופורציית המשפחות עם מדיח כלים עלה. הוחלט לבצע מדגם אקראי של 150 משפחות.

- א. רשמו את השערות המחקר.
- ב. מה היא מסקנת המחקר מובהקות של 5% אם
- ל-102 משפחות היה מדיח כלים.
- ג. מהי הטעות האפשרית במסקנה מהסעיף הקודם. האם ניתן לדעת את הסתברותה?

גברים	2	0.9	0.7	1.2	1.1
נשים	1.4	1.1	1.9	1.8	1.2

ברמת
במדגם

- 6) נערך מחקר על הקשר בין עישון ויתר לחץ דם. נבדק מדגם מקרי של 200 מעשנים ונמצא כי 30 סבלו מיתר לחץ דם. ידוע שבאוכלוסייה 18% סובלים מיתר לחץ דם.
- א. בדקו ברמת מובהקות 0.1 את ההשערה כי אחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.
- ב. מהי רמת המובהקות המינימלית לקבלת הטענה שאחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.
- ג. מהי עצמת המבחן, אם אחוז הסובלים מיתר לחץ דם בקרב אוכלוסיית המעשנים היא בפועל 25%.

- 7) להלן התפלגות מספר הנסיעות לחופשה השנתית במדגם של משפחות ישראליות. בדקו ברמת מובהקות של 5%:

מספר הנסיעות	0	1	2	3	4
מספר המשפחות	84	102	26	20	12

- א. באיטליה משפחות נוסעות בממוצע פעמיים בשנה לחופשה. האם בישראל משפחות נוסעות פחות מאשר באיטליה?
- ב. בהולנד 80% מהמשפחות נוסעות לפחות פעם אחת בשנה לחופשה, האם בישראל אחוז המשפחות שנוסעות לפחות פעם אחת בשנה לחופשה נמוך מאשר בהולנד?

(8) נתון כי: $X \sim N(\mu, \sigma^2 = 10^2)$.

מעוניינים לבדוק את ההשערות: $H_0: \mu = 40$, $H_1: \mu > 40$.

דגמו 25 תצפיות מהאוכלוסייה והתקבל $\bar{X} = 45$.

א. חשבו את p-value (מובהקות התוצאה).

ב. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה: $H_1: \mu < 40$.

ג. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה: $H_1: \mu \neq 40$.

תשובות סופיות

(1) א. חוקר א'	ב. 0.0122	ג. גדלה.	ד. נדחה H_0 .
(2) א. לפחות 0.0808	ב. 0.7995		
(3) א. לא נדחה H_0 .	ב. לא נדחה H_0 .		
(4) א. לא נדחה H_0 .	ב. בין 5% ל-10%.	ג. נדחה H_0 .	
(5) א. $H_0: p = 0.6$ $H_1: p > 0.6$	ב. נדחה H_0 .	ג. טעות מסוג ראשון בסיכוי של 0.05.	
(6) א. לא נדחה H_0 .	ב. 0.8643	ג. 0.8749.	
(7) א. נדחה H_0 .	ב. נדחה H_0 .		
(8) א. 0.0062	ב. 0.9938	ג. 0.0124.	

שאלות סיכום – שאלות רב ברירה על בדיקת השערות

(1) בבדיקת השערה חד-צדדית ימנית ברמת מובהקות $\alpha = 0.01$, נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות אותם נתונים ברמת מובהקות $\alpha = 0.05$?

- השערת האפס הייתה נדחית.
- השערת האפס לא הייתה נדחית.
- ההשערה המחקרית הייתה נדחית.
- בהעדר נתונים נוספים, לא ניתן לדעת.

(2) על מנת לבדוק האם ההסתברות ללידת בן הינה חצי, נבחר מדגם מקרי של 200 ילדים, ונמצא שישנם 120 בנים. מהן ההשערות האלטרנטיביות להשערת האפס?

א. $H_1: p = 0.5$

ב. $H_1: p = 0.6$

ג. $H_1: p > 0.5$

ד. $H_1: p \neq 0.5$

(3) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמיתית. המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
- מבחן T למדגם יחיד.
- מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים מזווגים.

(4) כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות:

מהן ההשערות הנבדקות? (ההפרש חושב $X - Y$)

68	82	93	69	לפני הנישואין - X
71	84	88	80	לאחר הנישואין - Y

א. $H_1: \mu_d < 0, H_0: \mu_d = 0$

ב. $H_1: \mu_x - \mu_y < 0, H_0: \mu_x - \mu_y = 0$

ג. $H_1: \mu_x - \mu_y < 0, H_0: \mu_x - \mu_y = 0$

ד. $H_1: \mu_d > 0, H_0: \mu_d = 0$

(5) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :

- השערת האפס נכונה.
- השערת האפס נדחתה.
- השערת האפס לא נדחתה.
- אף אחת מהתשובות לא נכונה בהכרח.

(6) ידוע כי ילד בגיל שנתיים ישן בממוצע 9 שעות בלילה. במדגם של 20 תינוקות בני שנתיים המתגוררים בצפון נמצא, כי ממוצע שעות השינה בלילה הינו 10 עם סטיית תקן של 1.1 במדגם של 10 תינוקות בדרום נמצא, כי ממוצע שעות השינה בלילה הינו 7.9 עם סטיית תקן של 1.1. על מנת להשוות בין ממוצע שעות השינה של ילדים מהצפון לבין זה של כלל הילדים יש לערוך _____, ועל מנת להשוות בין ממוצע שעות השינה של ילדים מהדרום לזה של ילדים המתגוררים בצפון יש לערוך _____. יש להניח שההנחות הדרושות מתקיימות.

- מבחן Z למדגם יחיד ; מבחן T למדגם יחיד.
- מבחן T למדגם יחיד ; מבחן T למדגמים תלויים.
- מבחן T למדגם יחיד ; מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים בלתי תלויים ; מבחן T לממוצע יחיד.

(7) מובהקות התוצאה (PV) היא גם :

- רמת המובהקות המינימאלית לדחות השערת האפס.
- רמת המובהקות המקסימאלית לדחיית השערת האפס.
- רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- רמת המובהקות המינימאלית לאי דחיית השערת האפס.

(8) כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות :

68	82	93	69	לפני הנישואין
71	84	88	80	לאחר הנישואין

לבדיקת

באיזה התפלגות משתמשים

ההשערות, ובכמה דרגות חופש :

- ההתפלגות Z ללא דרגות חופש.
- ההתפלגות T ו-3 דרגות חופש.
- ההתפלגות T ו-6 דרגות חופש.
- ההתפלגות χ^2 ו-3 דרגות חופש.

- 9) שני סטטיסטיקאים בודקים השערות ברמת מובהקות $\alpha = 0.05$ על סמך אותו מדגם. סטטיסטיקאי א' בודק את ההשערה: $H_0: \mu = 20$ כנגד האלטרנטיבה $H_1: \mu \neq 20$ ומחליט לא לדחות את השערת האפס. סטטיסטיקאי ב' בודק את ההשערה $H_0: \mu \leq 20$ כנגד האלטרנטיבה $H_1: \mu > 20$ מה יחליט סטטיסטיקאי ב'?
- לדחות את השערת האפס.
 - לא לדחות את השערת האפס.
 - ללא נתונים נוספים אי אפשר לדעת מה יחליט.
- 10) חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?
- הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
 - הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
 - הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.
 - הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
- 11) רמת הכולסטרול בדמם של אנשים מתפלג נורמאלית עם תוחלת של 180 מ"ג (ל 100 סמ"ק דם). וסטיית תקן של 10 מ"ג. מעוניינים לבדוק את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר. נניח שסטיית התקן אצל צמחונים זהה לסטיית התקן של כלל האנשים. במדגם של 20 צמחונים התקבל ממוצע רמת כולסטרול 174.5 מ"ג. אם הוחלט לקבל את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר איזה סוג טעות אפשרית במסקנה?
- טעות מסוג ראשון.
 - טעות מסוג שני.
 - טעות מסוג שלישי.
 - לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמתית אצל הצמחונים.

12) בסקר שנערך התקבל ש 60% מתוך 220 נשאלים מבקרים אצל השיננית לפחות פעם אחת בשנה. עבור אילו רמות מובהקות ניתן יהיה לקבוע שרוב האוכלוסייה מבקרת אצל השיננית לפחות פעם בשנה?

- א. רמת מובהקות הגדולה מ-5%.
- ב. רמת מובהקות הקטנה מ-5%.
- ג. רמת מובהקות הגדלה מ-0.0015.
- ד. רמת מובהקות הקטנה מ-0.0015.

13) שני חוקרים העוסקים בתחום מחקרי משותף החליטו להסתמך על נתונים של מדגם שפורסם על ידי הלשכה המרכזית לסטטיסטיקה.

חוקר א' ניסח השערה דו צדדית ואילו חוקר ב' ניסח השערה חד צדדית. מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?

- א. אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב' באותה רמת מובהקות.
- ב. אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס באותה רמת מובהקות.
- ג. אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת האפס באותה רמת מובהקות.
- ד. אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת האפס בתנאי שרמת המובהקות כפולה בגודלה.

14) ידוע מנתוני העבר כי תוחלת הציונים בבחינה בפסיכולוגיה היא 79. הועלתה השערה כי תוחלת הציונים בקרב העולים החדשים נמוכה יותר. לצורך בדיקת הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75. מה משמעות הפרמטר בניסוח ההשערות?

- א. תוחלת ציוני העולים באוכלוסייה.
- ב. ממוצע ציוני העולים במדגם.
- ג. תוחלת ציוני האוכלוסייה מנתוני העבר.
- ד. ממוצע ציוני שאר האוכלוסייה במדגם.

15) חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?

- א. החוקר דחה את השערת H_0 כאשר היא הייתה נכונה.
- ב. החוקר דחה את השערת H_1 כאשר היא הייתה נכונה.
- ג. החוקר לא דחה את השערת H_0 כאשר היא הייתה לא נכונה.
- ד. המדגם של החוקר שייך בפועל להתפלגות הדגימה של H_1 .

- 16** חוקר ביקש לבחון האם תאומים זהים אשר הופרדו בילדותם שונים מתאומים זהים אשר גדלו יחדיו מבחינת מידת הפער בין התאומים בלחץ הדם. הוא דגם 20 זוגות תאומים מכל אוכלוסייה ומדד את הפרש בין לחץ הדם בכל זוג תאומים. מהו המבחן הסטטיסטי המתאים?
- מבחן T למדגמים בלתי תלויים עם 38 דרגות חופש.
 - מבחן T למדגמים מזווגים, עם 39 דרגות חופש.
 - מבחן T למדגמים בלתי תלויים עם 39 דרגות חופש.
 - מבחן T למדגמים מזווגים עם 38 דרגות חופש.

- 17** בינואר השנה פורסם שהשכר הממוצע במשק הוא 8,900 ₪. במדגם שנעשה בחודש יוני על 60 עובדים נרשם עבור כל עובד במדגם האם השכר שלו נמוך או לא נמוך מהשכר הממוצע שפורסם בחודש ינואר. מהו המבחן המתאים כדי לבדוק שרוב העובדים בחודש יוני קיבלו שכר הנמוך מהשכר הממוצע שפורסם בחודש ינואר?
- מבחן Z על פרופורציה.
 - מבחן T על תוחלת אחת.
 - מבחן T על שתי תוחלות במדגמים בלתי תלויים.
 - מבחן T על שתי תוחלות במדגמים תלויים.

- 18** שלושה חוקרים רצו לבדוק את השפעתו של שידור פרסומות נגד תאונות דרכים על מהירות הנהיגה של נהגים בישראל (השונות של מהירות הנהיגה בישראל אינה ידועה). עידו השווה את מהירות הנהיגה של קבוצת נהגים אחת, חודש לפני שידור הפרסומות וחודש לאחר שידור הפרסומות. רון השווה את מהירות הנהיגה של קבוצת נהגים, שראו את הפרסומות, למהירות הנהיגה של קבוצת נהגים, שלא ראו את הפרסומות. יואב השווה את מהירות הנהיגה של קבוצת נהגים בחודש בו שודרו הפרסומות, למהירות הנהיגה הממוצעת בישראל על פי נתוני משרד התחבורה. המבחנים בהם צריכים החוקרים להשתמש הם:
- שלושתם במבחן T למדגמים בלתי תלויים.
 - עידו במבחן T למדגמים מזווגים, ורון ויואב במבחן T למדגמים בלתי תלויים.
 - עידו במבחן T למדגמים מזווגים, רון במבחן T למדגמים בלתי תלויים ויואב במבחן T למדגם יחיד.
 - עידו במבחן T למדגמים מזווגים, רון ויואב במבחן T למדגם יחיד.

- 19** במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5%. מה תמיד נכון?
- הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
 - הגדלת רמת המובהקות תשנה את מסקנת המחקר.
 - הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
 - הקטנת רמת המובהקות תשנה את מסקנת המחקר.
- 20** חוקר ערך מבחן דו צדדי ברמת מובהקות של α והחליט לדחות את השערת האפס. אם החוקר היה עורך מבחן חד צדדי ברמת מובהקות של $\frac{\alpha}{2}$ אזי בהכרח:
- השערת האפס הייתה נדחית.
 - השערת האפס הייתה לא נדחית.
 - לא ניתן לדעת מה תהיה מסקנתו במקרה זה.
- 21** ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם.
- ליאור השתמש בטבלה של התפלגות Z .
 - רוני השתמשה בטבלה של התפלגות T .
 - מה נוכל לומר בנוגע להחלטת המחקר שלהם?
 - אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
 - אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
 - שני החוקרים בהכרח יגיעו לאותה מסקנה.
 - לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.
- 22** נתון ש $X \sim N(\mu, \sigma^2)$ כמו כן נתונות ההשערות הבאות: $H_0: \mu = \mu_0$, $H_1: \mu < \mu_0$.
- חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות.
- כעת בברור הוא ידחה את השערת האפס.
 - כעת הוא דווקא יקבל את השערת האפס.
 - כעת לא ניתן לדעת מה תהיה מסקנתו.
- 23** אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:
- הסיכוי לטעות מסוג ראשון גדל.
 - העוצמה של המבחן גדלה.
 - הסיכוי לטעות מסוג שני גדל.
 - תשובות א ו-ב נכונות.

(24) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :

- השערת האפס נכונה.
- השערת האפס נדחתה.
- השערת האפס לא נדחתה.
- אף אחת מהתשובות לא נכונה בהכרח.

(25) מה המצב הרצוי לחוקר המבצע בדיקת השערה :

$1 - \beta$	α
א. גדולה	א. גדולה
ב. קטנה	ב. גדולה
ג. גדולה	ג. קטנה
ד. קטנה	ד. קטנה

(26) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית H_0 קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך :

- הן α , והן $(1 - \beta)$, יקטנו.
- α יישאר ללא שינוי ואילו $(1 - \beta)$ יגדל.
- α יגדל ואילו $(1 - \beta)$ יקטן.
- הן α והן $(1 - \beta)$ יגדלו.

(27) ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח שלחץ הדם בקרב עיתונאים גבוה יותר מהממוצע באוכלוסייה. הוא לקח מדגם של 60 עיתונאים וקיבל ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?

- טעות מסוג ראשון.
- טעות מסוג שני.
- טעות מסוג שלישי.
- אין טעות במסקנתו.

(28) בבדיקת השערות התקבל שה- $p\text{-value} = 0.02$. מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה :

- יקבל את השערת האפס בכל מקרה.
- ידחה את השערת האפס מקרה.
- ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
- לא ניתן לדעת כי אין מספיק נתונים.

(29) מובהקות התוצאה (PV) היא גם :

- רמת המובהקות המינימאלית לדחות השערת האפס.
- רמת המובהקות המקסימאלית לדחיית השערת האפס.
- רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- רמת המובהקות המינימאלית לאי דחיית השערת האפס.

(30) בבדיקת השערות מסוימת התקבל $p \text{ value} = 0.0254$, לכן :

- ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .
- ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .
- ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .
- ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

(31) רמת המובהקות במחקר הייתה 2% לכן.

- בסיכוי של 2% נדחה את השערת האפס.
- בסיכוי של 2% לא נדחה את השערת האפס.
- בסיכוי של 2% השערת האפס לא נכונה.
- אף תשובה לא נכונה.

(32) נתון ש: $X \sim N(\mu, \sigma^2)$. כמו כן נתונות ההשערות הבאות: $H_0: \mu = \mu_0$, $H_1: \mu < \mu_0$.

- חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות.
- σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5%. אם הוא היה מגדיל את רמת המובהקות ל-10% אזי:
- א. כעת בברור הוא ידחה את השערת האפס.
- ב. כעת הוא דווקא יקבל את השערת האפס.
- ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

(33) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים

- זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה.
- לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית. מספר דרגות החופש במבחן הוא :

א. 9

ב. 19

ג. 18

ד. 8

34) בתחנת טיפת חלב מסוימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשיים נשקל בכל אחד מהמשקלים. להלן תוצאות השקילה (בק"ג):

משקל במכשיר 1	2.5	0.7	9.6	4.5
משקל במכשיר 2	0.5	1.7	6.9	3.5

נניח שהמשקלים מתפלגים נורמלית.

המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
- מבחן T למדגם יחיד.
- מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים מזווגים.

35) כדי להשוות בין שני אצים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים.

המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
- מבחן T למדגם יחיד.
- מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים מזווגים.

36) סטטיסטיקאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0. המשמעות של תוצאה זו היא:

- לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
- תמיד לדחות את השערת האפס כאשר היא נכונה.
- לעולם לא לדחות את השערת האפס כאשר היא נכונה.
- תמיד לדחות את השערת האפס כאשר היא לא נכונה.

37) סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים. הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח $-2 < \mu_1 - \mu_2 < 4.5$. אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות:

$$H_1: \mu_1 - \mu_2 \neq 0; H_0: \mu_1 - \mu_2 = 0$$

ברמת מובהקות 0.05 מסקנתו תהיה:

- לדחות את השערת האפס.
- לא לדחות את השערת האפס.
- שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.
- שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

(38) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפללים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו: $-0.0293 < \mu_D < 0.2145$ רווח הסמך הוא ברמת סמך של 95%. לכן מסקנת המחקר היא:

- א. ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
- ב. ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
- ג. לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D .

(39) אם רמת מובהקות של מבחן סטטיסטי הינה 0, הכוונה היא:

- א. תמיד נדחה H_0 כאשר היא נכונה, אך לא תמיד נדחה אותה כאשר היא לא נכונה.
- ב. לא נדחה את H_0 אף פעם.
- ג. לא נדחה את H_0 כאשר היא נכונה אך יתכן ונדחה אותה כאשר היא לא נכונה.
- ד. כל התשובות לא נכונות.

(40) חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות: $H_0: \mu = 10$, $H_1: \mu \neq 10$. לצורך בדיקה הוא לקח מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך תוצאות המדגם הוא חישב וקיבל: $t_x = -2.63$. לכן המסקנה היא:

- א. הוא ידחה H_0 ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
- ב. הוא ידחה H_0 ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025.
- ג. הוא ידחה H_0 ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01.
- ד. הוא לא ידחה H_0 ברמת מובהקות 0.1.

(41) האיגוד האמריקני לרפואת ילדים מפרסם הנחיות חדשות הקובעות כי יש ליטול תוספת יוד במהלך תקופת ההיריון וההנקה. מחסור במינרל זה עלול לגרום לפגיעה מוחית אצל העובר והתינוק. החלטה זו נקבעה על סמך מחקר בו השתתפו 1050 נשים שנטלו יוד במהלך תקופת ההיריון וההנקה. מתוך הנשים שהשתתפו במחקר, רק ל-21 נמצאו ילדים בעלי פגיעה מוחית לעומת 3% באוכלוסייה הכללית. בנוסף, פורסם שהאיגוד האמריקאי מגיע למסקנותיו על סמך רמת מובהקות של 0.5%. מה הסיכוי לבצע טעות מסוג ראשון במחקר?

- א. 0.005
- ב. 0.03
- ג. 0.0287
- ד. 0.05

- 42) חוקרת שיערה, כי משקלן של נשים כשנה לאחר החתונה גבוה ממשקלן בעת החתונה. החוקרת דגמה 15 נשים, ובדקה את משקלן בשתי נקודות הזמן (בעת החתונה, ושנה לאחריה), אך לא מצאה הבדל מובהק ברמת מובהקות 0.01. בהנחה, כי **במצאות** השערתה של החוקרת נכונה, **סביר** כי אם היא תגדיל את גודל המדגם, אזי:
- א. יקטן הסיכוי לטעות מסוג שני (β).
 - ב. תגדל רמת הביטחון ($1 - \alpha$).
 - ג. אף תשובה לא נכונה.
 - ד. כל התשובות נכונות.

43) איזה מהמשפטים הבאים נכון תמיד?

- א. $POWER + \alpha + \beta = 1$
- ב. $POWER = 0.5 - \beta$
- ג. $POWER + \alpha = 1$
- ד. $\beta + \alpha = 1$
- ה. הכול לא נכון.

44) מה נכון לומר לגבי הנחת שיוויון השונויות במבחן T למדגמים בלתי תלויים?

- א. היא אומרת שהשונויות המדגמיות שוות.
- ב. בלעדיה אין שום דרך לבדוק השערה על הפרש בין תוחלות.
- ג. היא חשובה הן עבור מדגמים מזווגים והן עבור מדגמים בלתי תלויים.
- ד. אף תשובה אינה נכונה.

45) חוקר החליט לא לדחות השערה ברמת מובהקות של α . במידה וחוקר זה היה בודק השערה זו ברמת מובהקות של 2α על סמך אותם נתונים, האם ההשערה תדחה?

- א. ההשערה תדחה.
- ב. ההשערה לא תדחה.
- ג. התשובה תלויה בעוצמת המבחן.
- ד. לא ניתן לדעת בוודאות אם ההשערה תדחה או לא.

46) חוקרת שיערה, כי בגילאי הגן בנות יותר תקשורתיות מבנים. אם החוקרת תדגום אקראית 30 בנים ו-30 בנות, ובמדגם יתקבל אותו ממוצע של ציון תקשורת. סטטיסטי המבחן יהיה:

- א. אפס
- ב. חיובי
- ג. שלילי
- ד. לא ניתן לדעת

(47) עוצמה שווה ל-1 פרושה :

- א. לעולם לא לדחות את השערת האפס כאשר היא נכונה.
- ב. תמיד לדחות את השערת האפס כאשר היא נכונה.
- ג. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

(48) מה מהבאים נכון לגבי מבחן T מדגמים מזווגים?

- א. כל התצפיות במחקר אינן תלויות זו בזו.
- ב. כל התצפיות במחקר תלויות זו בזו.
- ג. כל הצמדים של תצפיות במחקר אינם תלויים זה בזה.
- ד. התצפיות בתוך כל צמד אינן תלויות זו בזו.

(49) לבדיקת ההשערה החד צדדית על התוחלת של התפלגות נורמלית $H_0: \mu \geq 10$, נלקח מדגם והתקבלה רמת מובהקות מינימאלית לדחיית השערת האפס 0.058. לו רצינו לבדוק את ההשערה הדו צדדית $H_0: \mu = 10$, $H_1: \mu \neq 10$, אז על סמך תוצאת אותו המדגם ברמת מובהקות 0.05 :

- א. ניתן להכריע בין ההשערות רק אם שונות האוכלוסייה נתונה.
- ב. מקבלים את השערת האפס.
- ג. דוחים את השערת האפס.
- ד. לא ניתן להכריע בין ההשערות שכן חסרים נתונים.

(50) לבדיקת ההשערה החד צדדית ימנית $H_0: \mu = 55$, $H_1: \mu = 65$,

נלקח מדגם מקרי בגודל n מאוכלוסייה בעלת התפלגות נורמלית ושונות σ^2 . רמת המובהקות היא 5%. נמצא שהעוצמה היא 0.9. להלן 3 טענות :

- עבור מדגם בגודל n וברמת מובהקות 5% לבדיקת ההשערות : $H_0: \mu = 55$, $H_1: \mu = 60$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות : $H_0: \mu = 55$, $H_1: \mu = 65$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל n ורמת מובהקות 10% לבדיקת ההשערות : $H_0: \mu = 55$, $H_1: \mu = 65$ העוצמה תהיה קטנה מ-0.9.

- א. שלושת הטענות אינן נכונות.
- ב. טענות 2 ו-3 אינן נכונות וטענה 1 נכונה.
- ג. טענות 1 ו-2 נכונות וטענה 3 אינה נכונה.
- ד. טענות 1 ו-3 אינן נכונות וטענה 2 נכונה.

תשובות סופיות:

שאלה	תשובה	שאלה	תשובה
1	א	26	א
2	ד	27	ב
3	ד	28	א
4	א	29	א
5	ג	30	ג
6	ג	31	ד
7	א	32	א
8	ב	33	א
9	ג	34	ד
10	א	35	ג
11	א	36	א
12	ג	37	ג
13	א	38	א
14	א	39	ג
15	א	40	א
16	א	41	א
17	א	42	א
18	ג	43	ה
19	א	44	ד
20	ג	45	ד
21	ב	46	א
22	ג	47	ד
23	ד	48	ג
24	ג	49	ב
25	ג	50	ד

סטטיסטיקה והסתברות

פרק 36 - מבחני חי בריבוע

תוכן העניינים

1. מבחן לאי תלות..... 223

מבחן חי בריבוע לאי תלות בין משתנים – רקע

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים. שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות.

מבנה המבחן:

השערות:

אין תלות בין המשתנים H_0 .

יש תלות בין המשתנים H_1 .

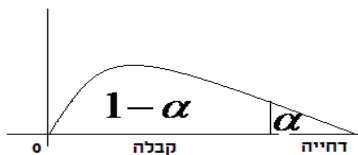
כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש $d.f = (r-1)(c-1)$. כאשר: r - מספר הקטגוריות של המשתנה שבשורות. c - מספר הקטגוריות של המשתנה שבעמודות.

הערך הקריטי הוא: $\chi^2_{1-\alpha, (r-1)(c-1)}$, כלומר האחוזון ה- $1-\alpha$ בהתפלגות חי בריבוע שדרגות החופש הן $(r-1)(c-1)$. אם $\chi^2 > \chi^2_{1-\alpha, (r-1)(c-1)}$ אז דוחים את השערת האפס.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

כאשר:



O_i - השכיחות נצפית במדגם בתא i .

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

$$E_i = \frac{f(x) \cdot f(y)}{n}$$

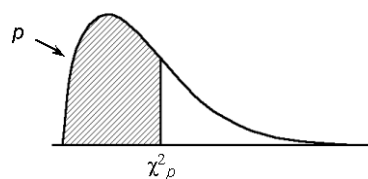
הערה:

תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.
 תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ 20% מהתאים E קטן מ-5.

דוגמה (הפתרון בהקלטה):

האם יש תלות בין המגדר לבין דעה מסוימת?
 יש לבדוק ברמת מובהקות של 5% על סמך תוצאות הסקר:

המגדר / דעה	בעד	נגד	נמנע	סה"כ
גברים	50	40	10	
נשים	20	60	20	
סה"כ				

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p 

df	p												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.00157	0.00982	0.02393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

שאלות

- (1) נבדקה התלות בין גודל הארגון לבין שביעות הרצון של העובדים. להלן התוצאות:

שביעות רצון גודל המפעל	נמוכה	בינונית	גבוהה	סה"כ
גדול	182	203	215	600
קטן	154	110	136	400
סה"כ	336	313	351	1000

מה המסקנה ברמת מובהקות של 2.5%?

- (2) מפעל עובד בשלוש משמרות. להלן מספר המוצרים הפגומים והתקינים בכל אחת מן המשמרות לפי מדגם שנעשה:

לילה	ערב	יום	
70	60	50	פגומים
800	700	600	תקינים

האם יש הבדל בין שיעורי הפגומים במשמרות השונות? הסיקו עבור רמת מובהקות $\alpha = 0.05$.

- (3) נדגמו 50 מוצרים ממפעל מסוים מתוך 30 מוצרים שיוצרו ביום 17 נבחרו לייצוא מתוך המוצרים שיוצרו בלילה 10 נבחרו לייצוא. האם יש קשר בין היות מוצר לייצוא למועד שבו הוא יוצר? בדקו ברמת בטחון של 95%.

- (4) במטרה לבדוק האם השתנו דפוסי ההצבעה למפלגות השונות בין שבוע שעבר לשבוע נלקחו שני סקרים אחד מהשבוע שעבר והאחר מהשבוע. להלן דפוסי ההצבעה שהתקבלו בסקרים אלה.

א. מהי רמת המובהקות המינמלית עבורה ניתן להחליט שהשתנו דפוסי ההצבעה משבוע שעבר לשבוע באופן מובהק?

ב. כיצד הייתה התשובה לסעיף א משתנה אם כל השכיחויות בטבלה של תוצאות המדגם היו מוכפלות פי 2?

ג. בנו רווח סמך לשיעור המצביעים למפלגה א השבוע ברמת סמך של 95%.

	מפלגה א	מפלגה ב	מפלגות אחרות	סה"כ
שבוע שעבר				
השבוע	143		253	550
סה"כ	243	314		1050

- (5) סטודנט קיבל בבדיקת השערות ערך χ^2 (chi-square) השוו לאפס. הסטודנט הסיק כי לא קיימת תלות בין שני המשתנים שבדק, בכל רמת מובהקות. נכון / לא נכון? נמקו.

- (6) להלן טבלת O של שני משתנים שהתקבל במדגם כלשהו :

$f(x)$	Y_4	Y_3	Y_2	Y_1	
200					X_1
200					X_2
	160	120	60	60	$f(y)$

מה צריכות להיות השכיחויות בתוך הטבלה כדי שמובהקות התוצאה (PV) תהיה 100%?

תשובות סופיות

- (1) נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.
- (2) נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.
- (3) נסיק שאין קשר בין היות מוצא לייצוא למועד שבו הוא יוצר.
- (4) א. 10% ב. קטן ג. (0.223, 0.297)
- (5) נכון
- (6) להלן טבלה :

$f(x)$	Y_4	Y_3	Y_2	Y_1	
200	80	60	30	30	X_1
200	-8	60	30	30	X_2
400	160	120	60	60	$f(y)$

סטטיסטיקה והסתברות

פרק 37 - משתנה דו-מימדי בדיד - פונקציית הסתברות משותפת

תוכן העניינים

1. כללי 228

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:

$$p(x=0, y=0) = 0.05$$

$$p(x=0, y=1) = 0$$

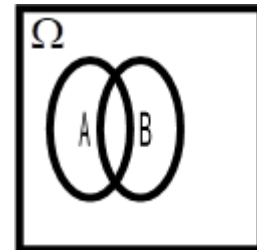
$$p(x=1, y=0) = 0.15$$

$$p(x=1, y=1) = 0.05$$

$$p(x=2, y=0) = 0$$

$$p(x=2, y=1) = 0.75$$

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



שימו לב שסכום כל ההסתברויות בפונקציית ההסתברות המשותפת הוא 1.

כעת נסכם את השורות ואת העמודות ונקבל את פונקציית 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 ביישוב 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.

$x \setminus y$	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.

$x \setminus y$	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:

$x \setminus y$	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 ג. תוחלת: 2.

$x \setminus y$	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.

$x \setminus y$	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 והסתברות

פרק 38 - משתנה דו מימדי בדיד - מתאם בין משתנים

תוכן העניינים

1. כללי 234

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.



מקדם מתאם -1 או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים, שניתן לבטאו על ידי הנוסחה: $y = ax + b$.

מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר לינארי מלא בו השיפוע a יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לינארי 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:

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

$$\text{cov}(x, y) = E[(x - \mu_x)(y - \mu_y)] = E(x \cdot y) - E(x) \cdot E(y) \quad \text{השונויות המשותפת:}$$

תכונות של השונויות המשותפת:

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

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

משתנים בלתי מתואמים:

משתנים בלתי מתואמים הם משתנים שמקדם המתאם שלהם אפס, וכדי שדבר כזה יקרה השונויות 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 טרנספורמציה לינארית על מקדם המתאם:

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

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

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

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

y	P_Y
0	0.2
1	0.8

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

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

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

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

$$E(xy) = 0 \cdot 0 \cdot 0.05 + 0 \cdot 1 \cdot 0 + 1 \cdot 0 \cdot 0.15 + 1 \cdot 1 \cdot 0.05 + 2 \cdot 0 \cdot 0 + 2 \cdot 1 \cdot 0.75 = 1.55$$

$$\text{cov}(x, y) = E(x \cdot y) - E(x) \cdot E(y) = 1.55 - 1.7 \cdot 0.8 = 0.19$$

$$\rho = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} = \frac{0.19}{0.557 \cdot 0.4} = 0.853$$

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 פונקציית 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 מתאומים?
- 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 מתאומים?

- (4) קובייה הוגנת הוטלה פעמיים.
יהי X ההטלה הגדולה מבין שתי התוצאות, ויהי Y מס' ההטלות בהן יצאה תוצאה זוגית.
- מצאו את פונקציית ההסתברות המשותפת של X ו- Y .
 - חשבו את מקדם המתאם של X ו- Y .
 - 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.
- בנו את פונקציית ההסתברות המשותפת ואת פונקציות ההסתברות השוליות.
 - האם המשתנים מתואמים?
 - מה מקדם המתאם בין X לבין Y ?
 - מה יהיה מקדם המתאם:
- בין מספר הדירות המשופצות למספר הדירות הזוגיות שנדגמו.
 - 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.

$x \setminus y$	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.

$x \setminus y$	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.

$x \setminus y$	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.

$x \setminus y$	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. $\frac{2}{3}$.

$x \setminus y$	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 והסתברות

פרק 39 - המשתנה המקרי הדו ממדי - קומבינציות ליניאריות

תוכן העניינים

1. כללי 241

המשתנה המקרי הדו מימדי – קומבינציות לינאריות:

רקע:

יהיו שני משתנים מקריים X ו- Y .
התוחלת והשונות של סכומם היא:

$$\begin{aligned}E(X+Y) &= E(X) + E(Y) \\V(X+Y) &= V(X) + V(Y) + 2 \cdot \text{cov}(X, Y)\end{aligned}$$

התוחלת והשונות של הפרשם היא:

$$\begin{aligned}E(X-Y) &= E(X) - E(Y) \\V(X-Y) &= V(X) + V(Y) - 2 \cdot \text{cov}(X, Y)\end{aligned}$$

קומבינציה ליניארית:

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

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

דוגמה (פתרון בהקלטה):

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

- השלימו את ההסתברויות החסרות.
- האם המשתנים תלויים?
- האם המשתנים בלתי מתואמים?
- חשבו את השונות המשותפת.
- חשבו את התוחלת והשונות של סכום המשתנים.
- 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.
- חשבו את השונות המשותפת בין הציון הכמותי לציון המילולי.
 - חשבו את התוחלת והשונות של סכום הציונים בחלק הכמותי ובחלק המילולי.
 - חשבו את התוחלת והשונות של הפרש הציונים בין החלק הכמותי לחלק המילולי.
 - 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: $\text{var}(X + 2Y) = 3$, $\text{var}(X - 2Y) = 2$.חשבו: $\text{cov}(X, Y)$.(4) מטילים קובייה n פעמים. נגדיר את המשתנים הבאים: X = מספר הפעמים שהתקבלה התוצאה 6. Y = מספר הפעמים שהתקבלה התוצאה 5בטאו את השונות המשותפת באמצעות n .

תשובות סופיות:

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

$x \setminus y$	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 והסתברות

פרק 40 - המשתנה המקרי הדו ממדי הבדיד - שאלות מסכמות

תוכן העניינים

1. שאלות מסכמות 244

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: $\rho = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$.

שונויות משותפת:

מגדירים את השונויות המשותפת:

$$\text{cov}(x, y) = E[(x - \mu_x)(y - \mu_y)] = E(x \cdot y) - E(x) \cdot E(y)$$

תכונות של השונויות המשותפת:

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

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

$$3. \text{cov}[(aX + b), (cY + d)] = a \cdot c \cdot \text{cov}(X, Y)$$

משתנים בלתי מתואמים:

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

השפעת טרנספורמציה לינארית על מקדם המתאם:

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

תוחלת ושונות של סכום משתנים:

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

קומבינציות לינאריות:

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

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

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

שאלות:

- (1) יש ליצור סיסמה בת 3 תווים. כל תו יכול להיבחר רק מתוך כלל התווים הבאים: 1, 2, A, B, C. 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 "טבורג". 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 "טבורג" 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 "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 : Y, X ו- Z :

$$\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$$

(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 : $\text{cov}(4Y, Q+6)$

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:

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

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

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

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

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

(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: ד. 1.

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

(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 והסתברות

פרק 41 - המשתנה המקרי הדו ממדי הרציף

תוכן העניינים

1. משתנה דו ממדי רציף..... 252

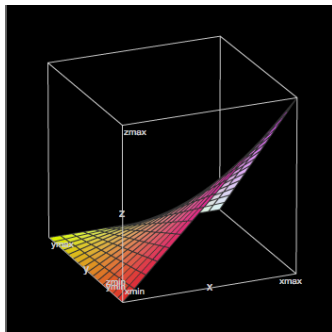
משתנה מקרי דו ממדי רציף:

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:

$$(1) \quad f(x, y) \geq 0 \quad \text{לכל } (x, y) \in R$$

$$(2) \quad \iint_R f(x, y) dx dy = 1$$



דוגמה (פתרון בהקלטה):

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

הראו שפונקציה זו יכולה להיות פונקציית צפיפות משותפת.

פונקציית צפיפות שולית:

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

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

דוגמה (פתרון בהקלטה):

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

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:

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

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

$$\text{ו-} Y \text{ יהיו בתחום הזה: } P[(x, y) \in A] = \iint_A f(x, y) dx dy$$

דוגמה (פתרון בהקלטה):

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

$$P(X < 0.5 \cap Y < 0.5) \quad \text{חשבו את הסיכוי:}$$

פונקציית התפלגות מצטברת משותפת:

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

$$F(s, t) = P(X \leq s \cap Y \leq t) = \int_{-\infty}^t \int_{-\infty}^s f(x, y) dx dy$$

דוגמה (פתרון בהקלטה):

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

מצאו את פונקציית ההתפלגות המצטברת המשותפת

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

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

$$\text{המקיימים: } f(y) > 0 \text{ על ידי: } f(x|y) = \frac{f(x, y)}{f(y)}$$

באופן דומה, פונקציית הצפיפות המותנית של Y בהינתן ש- $X = x$ לכל ערכי x

$$\text{המקיימים: } f(x) > 0 \text{ על ידי: } f(y|x) = \frac{f(x, y)}{f(x)}$$

דוגמה (פתרון בהקלטה):

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

$$\text{מצאו את: } f(x|y)$$

תוחלת מותנית:

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

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

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):

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

$$\text{מצאו את: } E(X|Y)$$

שאלות:

- (1) נתונה פונקציית הצפיפות הבאה: $f(x, y) = x + y$, המוגדרת בתחום שבו: $0 \leq x \leq 1$ וגם: $0 \leq y \leq 1$. הוכיחו שמדובר בפונקציית צפיפות.
- (2) נתונה פונקציית הצפיפות הבאה: $f(x, y) = Ax(x - y)$, המוגדרת בתחום שבו: $0 \leq x \leq 2$ וגם: $-x \leq y \leq x$. 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: $0 \leq x \leq 1$ וגם: $0 \leq y \leq 1$.
 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: $60 \leq x \leq y$ וגם: $60 \leq y \leq 100$.
 א. הראו שפונקציה זו מקיימת את התנאים של פונקציית צפיפות.
 ב. מצאו את פונקציית הצפיפות השולית של Y .
 ג. חשבו את $E(X)$, $V(X)$.
 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: $0 \leq x \leq y$, $f(x|y) = \frac{2x}{y^2}$.
מצאו את השונות המשותפת של X ו- Y .

(7) נתונים שני משתנים מקרים רציפים X ו- Y . פונקציות הצפיפות

$$f(x, y) = \begin{cases} x & 0 < y < 1 \\ 0 & 1 - y \leq x \leq 1 + y \\ 0 & \text{else} \end{cases}$$

המשותפות שלהם היא:

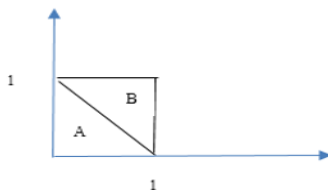
- מצאו את $f(x)$.
- מצאו את $f(y|x)$.
- מצאו את $E(Y|X)$.

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

- רשמו את פונקציית הצפיפות המשותפת.
- מצאו את פונקציית הצפיפות השולית של X ו- Y .
- חשבו את התוחלת של X ו- Y .
- האם X ו- Y משתנים בלתי מתואמים?
- 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?



- מצאו את $f(x)$.
- מצאו את $f(x|y)$.

(10) נתונה פונקציית הצפיפות המשותפת: $f(x, y) = cx$. פונקציה זו מוגדרת

בתחום שבו: $0 \leq x \leq 1$ וכן: $0 \leq y \leq x^2$.

- מצאו את הקבוע C .
- חשבו את ההסתברות ש- $6Y < 1 - X$.

11 נתונים X ו- Y שני משתנים מקריים רציפים כך ש: $Y \sim U(0,1)$

ו- $X|Y=y \sim U(0, \sqrt{y})$. 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: $f(x,y) = \frac{e^{-y} \cdot e^{-\frac{x}{y}}}{y}$, 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 $E(X^2|Y)$.

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: $E(|X-Y|^n) = \frac{2}{(n+1)(n+2)}$.

17 $X \sim \exp(1)$ וכן: $Y \sim \exp(1)$, הינם משתנים מקרים בלתי תלויים.

נגדיר את: $Z = \frac{X}{X+Y}$.

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. $E(X) = 73\frac{1}{3}$, $V(X) = 88\frac{8}{9}$

ד. לא. ה. 0.5 ו. 0.5625

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

ג. 0. ד. $\frac{\lambda}{\lambda + \mu}$

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

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

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

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

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

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

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

ה. לא.

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

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

$$(10) \quad \text{א. 4.} \quad \text{ב. 0.0947.}$$

$$(11) \quad \frac{7}{12}$$

$$(12) \quad \frac{1}{3}$$

$$(13) \quad e^{\frac{1}{2}}$$

$$(14) \quad \frac{25}{72}$$

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

$$\text{ב.} \quad y^2 + 1$$

$$\text{ג. 1.}$$

$$(16) \quad \text{שאלת הוכחה.}$$

$$(17) \quad \text{שאלת הוכחה.}$$