

סטטיסטיקה



תוכן העניינים

1.	סטטיסטיקה תיאורית-הקדמה	1
4.	סטטיסטיקה תיאורית- הצגה של נתונים	4
12.	סטטיסטיקה תיאורית- סכימה	12
4.	סטטיסטיקה תיאורית -מדדי מרכז (ללא ספר)	4
16.	סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן	16
19.	סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן	19
21.	סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה	21
23.	סטטיסטיקה תיאורית - טרנספורמציה לינארית	23
26.	סטטיסטיקה תיאורית שאלות אמריקאיות	26
32.	יסודות ההסתברות	32
36.	פעולות בין מאורעות (חיתוך ואיחוד) - מאורעות זרים ומכילים	36
44.	הסתברות מותנית-במרחב מדגם אחיד	44
47.	הסתברות מותנית - מרחב לא אחיד	47
50.	הערכת כלים אבחנתיים	50
54.	דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה	54
59.	תלות ואי תלות בין מאורעות	59
63.	שאלות מסכמות בהסתברות	63
66.	המשתנה המקרי הבדיד - פונקציית ההסתברות	66
69.	המשתנה המקרי הבדיד - תוחלת - שונות וסטיית תקן	69
72.	המשתנה המקרי הבדיד -תוחלת של פונקציה של משתנה מקרי בדיד	72
75.	המשתנה המקרי הבדיד- טרנספורמציה לינארית	75
78.	תוחלת ושונות של סכום משתנים מקריים	78
81.	התפלגויות בדידות מיוחדות -התפלגות בינומית	81

תוכן העניינים

85	24. התפלגויות בדידות מיוחדות - התפלגות אחידה
88	25. המשתנה המקרי הבדיד - שאלות מסכמות
91	26. המשתנה המקרי הרציף - התפלגויות כלליות (שימוש באינטגרלים)
100	27. התפלגויות רציפות מיוחדות - התפלגות מעריכית
103	28. התפלגויות רציפות מיוחדות - התפלגות אחידה
106	29. התפלגויות רציפות מיוחדות - התפלגות נורמלית
114	30. הסקה סטטיסטית - הקדמה
117	31. התפלגות הדגימה ומשפט הגבול המרכזי
137	32. מושגי יסוד באמידה
142	33. רווח סמך לתוחלת (ממוצע)
152	34. רווח סמך לפרופורציה
158	35. רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים
160	36. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים
162	37. שאלות מסכמות על רווחי סמך
165	38. מבוא לבדיקת השערות על פרמטרים
171	39. בדיקת השערות על תוחלת (ממוצע)
189	40. בדיקת השערות על פרופורציה
196	41. בדיקת השערות על הפרש פרופורציות
200	42. בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים
204	43. בדיקת השערות לתוחלת ההפרש במדגמים מזווגים
208	44. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות
211	45. שאלות מסכמות בבדיקת השערות
224	46. ניתוח שונות חד כיוונית
233	47. מבחני חי בריבוע
247	48. מבחנים אפרמטרים למדגמים מזווגים
254	49. מבחנים אפרמטריים למדגמים בלתי תלויים

סטטיסטיקה

פרק 1 - סטטיסטיקה תיאורית-הקדמה

תוכן העניינים

1. כללי 1

סטטיסטיקה תיאורית – הקדמה:

רקע:

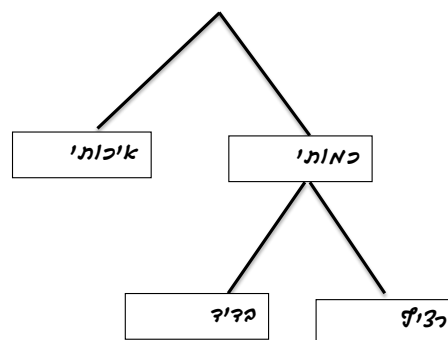
בסטטיסטיקה תיאורית אנו חוקרים קבוצה מסוימת, שיכולה להיות קבוצת ילדים בגן, קבוצת מניות בתיק, כלל התושבים בעיר מסוימת וכולי. בין ישות לישות בקבוצה ישנם גורמים היכולים לקבל מספר ערכים. גורמים אלה נקראים משתנים. למשל, בין מניה למניה בתיק משתנה התשואה היומית של המניה, הוותק של המניה, תחום המניה וכדומה. בסטטיסטיקה תיאורית אנחנו נתבונן בקבוצה מסוימת ובתוך הקבוצה הזו נאסוף נתונים לגבי משתנה מסוים ונלמד להציג את הנתונים ולנתח אותם מכל מיני אספקטים.

דוגמה:

בתיק מניות 10 מניות. מנהל התיק פרסם את התשואה של כל מניה בשנת 2011.

- (1) מי הקבוצה הנחקרת?
- (2) מה גודל הקבוצה?
- (3) מה המשתנה הנחקר?

סוגי משתנים:



משתנה איכותי

משתנה שלערכיו אין משמעות של יותר או פחות, אין עניין כמותי לערכים המתקבלים. כמו: מקום מגורים של אדם (רעננה, תל אביב, אשדוד..), מין האדם (זכר, נקבה) ומצב משפחתי (רווק, נשוי, גרוש, אלמן).

משתנה כמותי

משתנה שערכיו הם מספרים, להם יש משמעות כמותית כמו: גובה אדם בס"מ, ציון בבחינה וכדומה.

את המשתנה הכמותי נסווג לשני סוגים:

1. משתנה בדיד – משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים. כמו: מספר ילדים למשפחה (1,2,3...). וציון בבחינה (מ-0 ועד 100 בקפוצות של 1).
2. משתנה רציף – משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים. הערכים מתקבלים ברצף וללא קפיצות של ערכים. כמו: גובה בס"מ – אם למשל, הגובה הנמוך ביותר הוא 150 ועד 190 ס"מ בקבוצה הגבהים הם ברצף. גם בין 160 ל-161 ס"מ יש רצף אינסופי של ערכים אפשריים לגובה (160.33 ס"מ הוא גם גובה אפשרי), משקל בק"ג, מהירות בקמ"ש וכולי.

שאלות:

(1) סווגו את המשתנים הבאים לפי: איכותי / כמותי בדיד / כמותי רציף:

- א. מספר הדירות בבניין.
- ב. גיל אדם בשנים.
- ג. אחוז האבטלה בעיר.
- ד. מקצוע לימוד מועדף.

(2) להלן התפלגות מספר האיחורים לעבודה בחודש של העובדים בחברת "סטאר":

מספר האיחורים	מספר העובדים
0	17
1	23
2	85
3	50
4	25

בחברה 200 עובדים.

- א. מהו המשתנה הנחקר כאן?
- ב. האם מדובר במשתנה איכותי או כמותי?
- אם הוא כמותי האם הוא בדיד או רציף?

(3) להלן רשימה של משתנים כמותיים, ציינו ליד כל אחד אם הוא רציף או בדיד:

- א. שכר עובד ב-ש.
- ב. ציון בחינת בגרות.
- ג. תוצאה בהטלת קובייה.
- ד. מהירות ריצה בתחרות.
- ה. שיעור התמיכה בממשלה.

תשובות סופיות:

- (1) א. כמותי בדיד. ב. כמותי רציף. ג. כמותי רציף. ד. איכותי.
- (2) א. מספר איחורים. ב. כמותי בדיד.
- (3) א. רציף. ב. בדיד. ג. בדיד. ד. רציף. ה. רציף.

סטטיסטיקה

פרק 2 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי 4

סטטיסטיקה תיאורית – הצגה של נתונים:

רקע:

דרכים להצגת נתונים שנאספו:

רשימה של תצפיות:

התצפית היא הערך שנצפה עבור ישות מסוימת בקבוצה. רושמים את התצפיות שהתקבלו כרשומה, יעיל שיש מספר מועט של תצפיות. ההצגה הזו רלבנטית לכל סוגי המשתנים. למשל, להלן מספר החדרים בבניין בן 5 דירות: 3, 4, 3, 5, 4.

טבלת שכיחויות בדידה:

שם המשתנה- X	שכיחות – $f(x)$	שכיחות יחסית באחוזים
X_1	f_1	$\frac{f_1}{N} \cdot 100$
X_2	f_2	$\frac{f_2}{N} \cdot 100$
X_3	f_3	$\frac{f_3}{N} \cdot 100$
$\vdots$	$\vdots$	$\vdots$
X_k	f_k	$\frac{f_k}{N} \cdot 100$
סה"כ	$N = \sum_{i=1}^k f_i$	100%

רושמים את התצפיות בטבלה שבה עמודה אחת מבטאת את ערכי המשתנה והשנייה את השכיחות. יעיל עבור משתנה איכותי וכמותי בדיד וכשיש מספר רב של תצפיות. לא יעיל למשתנה כמותי רציף.

דוגמה:

להלן התפלגות הציונים בכיתה מסוימת:

הציון - X	מספר התלמידים – השכיחות - f	F_i	$\frac{f_i}{n}$
5	2	2	$0.08=2/25$
6	4	6	$0.16=4/25$
7	8	14	$0.32=8/25$
8	5	19	$0.2=5/25$
9	4	23	$0.16=4/25$
10	2	25	$0.08=2/25$

שכיחות מצטברת – צבירה של השכיחויות.

השכיחויות F_i – השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי:

$\frac{f_i}{n}$ – איזה חלק מהתצפיות בקבוצה שוות לערך.

טבלת שכיחויות במחלקות:

משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד וטבלת שכיחויות תהיה ארוכה מידי.

דוגמה:

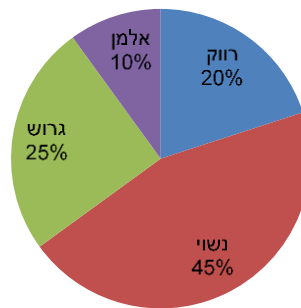
נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, בדקות. להלן ההתפלגות שהתקבלה:

זמן בדקות	מספר הילדים
0.5-3.5	20
3.5-9.5	18
9.5-19.5	14
19.5-29.5	8

דיאגרמת עוגה:

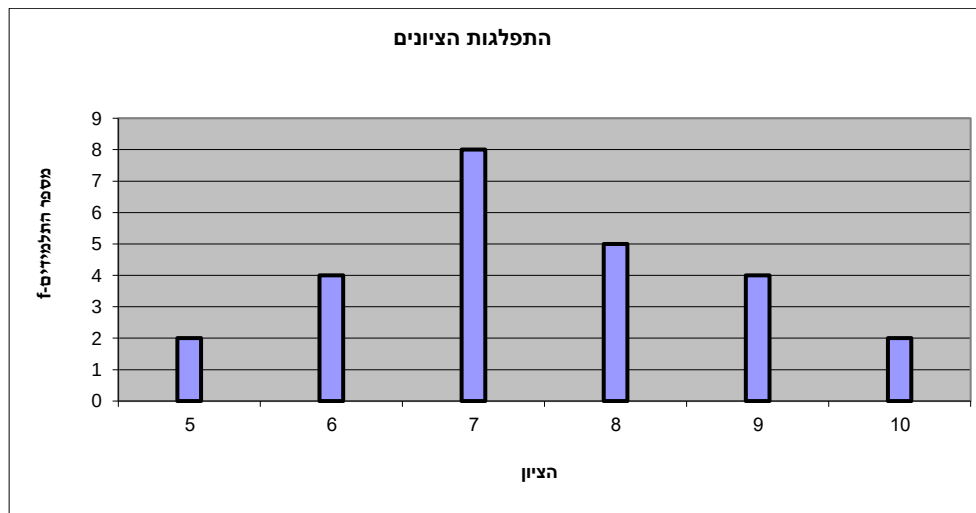
זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח", שהוא פרופורציונלי לשכיחות היחסית של ערך המשתנה בנתונים.

התפלגות המצב המשפחתי



דיאגרמת מקלות:

הציר האופקי הוא הציר של המשתנה והציר האנכי של השכיחות, כך שהגובה של המקל מעיד על השכיחות. רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף, וכן בסולמות מדידה עבור משתנה מסולם סדר.

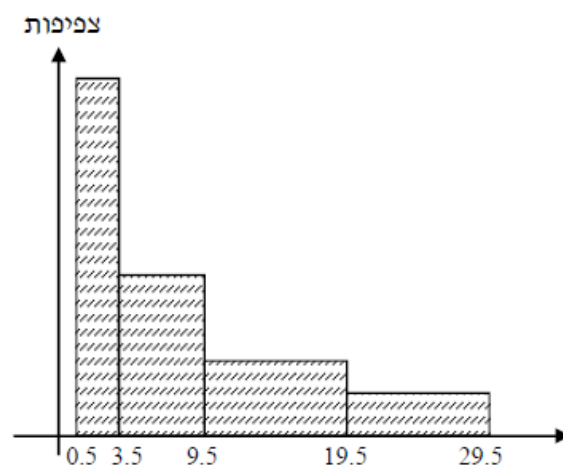


היסטוגרמה:

היסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות, והיא רלוונטית למשתנה כמותי רציף.

בהיסטוגרמה הציר האופקי הוא הציר של המשתנה והציר האנכי הוא הציר של הצפיפות. הצפיפות מחושבת בכל מחלקה על ידי חלוקת השכיחות ברוחב של כל המחלקה, והיא נותנת את מספר התצפיות הממוצע בכל מחלקה ליחידה. אם המחלקות הן שוות ברוחב, ניתן לשרטט את ההיסטוגרמה לפי השכיחות ואין צורך בצפיפות.

צפיפות	מצטברת	שכיחות	אמצע	רוחב	X
6.6667	20	20	2	3	0.5 - 3.5
3	38	18	6.5	6	3.5 - 9.5
1.4	52	14	14.5	10	9.5 - 19.5
0.8	60	8	24.5	10	19.5 - 29.5



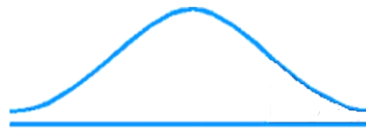
פוליגון – מצולעון:

אם נחבר את אמצע קצה כל מלבן בקווים ישרים. נותן מראה חזותי לצורה של התפלגות המשתנה.

צורות התפלגות נפוצות:

התפלגות סימטרית פעמונית

רוב התצפיות במרכז, וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. לדוגמה, ציוני IQ.

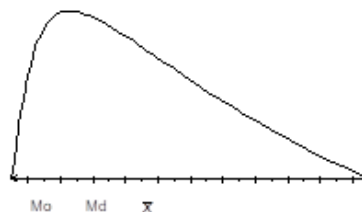


ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

התפלגות אסימטרית ימנית (חיובית)

רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. לדוגמה, שכר במשק.

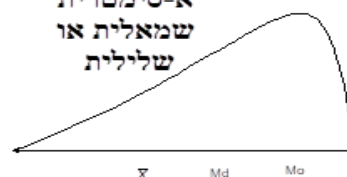
התפלגות א-סימטרית
ימנית או חיובית



התפלגות אסימטרית שמאלית (שלילית)

רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. לדוגמה, אורך חיים.

התפלגות
א-סימטרית
שמאלית או
שלילית



שאלות:

- (1) בסקר צפייה בטלוויזיה התקבלו התוצאות הבאות : 25 צפו בערוץ הראשון, 25 צפו בערוץ 10, 75 צפו בערוץ השני, 50 צפו באחד מערוצי הכבלים ו-25 לא צפו בטלוויזיה בזמן הסקר.
- א. רשמו את טבלת השכיחות ואת השכיחות היחסית.
- ב. תארו את הנתונים באופן גרפי.

- (2) להלן נתונים על התפלגות המקצוע המועדף של תלמידי שכבה ו' בבית הספר "מעוף":

מספר התלמידים	המקצוע
44	מתמטיקה
20	תנ"ך
12	אנגלית
26	היסטוריה

- א. מהו המשתנה הנחקר?
- ב. מהי פרופורציית התלמידים שמעדיפים תנ"ך?

- (3) להלן התפלגות ההשכלה במקום עבודה מסוים:

מספר העובדים	השכלה
60	נמוכה
120	תיכונית
20	אקדמאית

- א. מהו המשתנה הנחקר?
- מאיזה סולם הוא?
- ב. תארו את הנתונים באופן גרפי.

- (4) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
- 6, 5, 8, 7, 6, 7, 8, 6, 7, 5, 4, 6, 10, 9, 8, 6, 7.
- א. מהו המשתנה? האם הוא בדיד או רציף?
- ב. תארו את הרשימה בטבלת שכיחויות.
- ג. הוסיפו שכיחויות יחסיות לטבלה.
- ד. תארו את הנתונים באופן גרפי.

תשובות סופיות:

1) א. להלן טבלה: ב. עיין גרף מלא בסרטון הוידאו.

x	$f(x)$	$\frac{f(x)}{n}$	%
ערוץ 1	25	$\frac{25}{200}$	12.5%
ערוץ 10	25	$\frac{25}{200}$	12.5%
ערוץ 2	75	$\frac{75}{200}$	37.5%
כבלים	50	$\frac{50}{200}$	25%
לא צפו	25	$\frac{25}{200}$	12.5%
סה"כ	200	1	100%

- 2) א. מקצוע מועדף. ב. 19.6%.
- 3) א. משתנה נחקר: השכלה, סוג: סדר. ב. עיין גרף מלא בסרטון הוידאו.

- 4) א. המשתנה : ציון, משתנה בדיד.
ד. עיין גרף מלא בסרטון הוידאו.

ב+ג. להלן טבלה :

x	$f(x)$	$\frac{f(x)}{n}$	%
4	1	$\frac{1}{20}$	5%
5	2	$\frac{2}{20}$	10%
6	6	$\frac{6}{20}$	30%
7	4	$\frac{4}{20}$	20%
8	4	$\frac{4}{20}$	20%
9	2	$\frac{2}{20}$	10%
10	1	$\frac{1}{20}$	5%
סה"כ	20	20	100%

סטטיסטיקה

פרק 3 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי 12

סטטיסטיקה תיאורית – סכימה:

רקע:

בסטטיסטיקה ישנה צורת רישום מקובלת לסכום של תצפיות: $\sum_{i=1}^n X_i$.

נסביר את צורת הרישום על ידי הדוגמה הבאה:

i	X_i
1	5
2	0
3	1
4	3
5	2

(הסבר מלא מופיע בסרטונים באתר).

שאלות:

- 1) בבניין 5 דירות. לכל דירה רשמו את מספר החדרים שיש בדירה (X), ומספר הנפשות החיות בדירה (Y). חשבו:

מספר דירה	X	Y
1	2	1
2	3	1
3	2	2
4	4	3
5	3	2

א. $\sum_{i=1}^3 X_i$

ב. $\sum_{i=1}^5 Y_i$

ג. $\sum_{i=1}^4 X_i$

ד. $\left(\sum_{i=1}^4 X_i \right)^2$

ה. $\sum X_i$

ו. $\sum X_i Y_i$

ז. $\sum (X_i) \sum (Y_i)$

- (2) נתון לוח ערכי המשתנים X_i ו- Y_i , כאשר: $i = 1, 2, \dots, 6$, ונתונים הקבועים:
 $a = 2$, $b = 5$. חשבו את הנוסחאות הבאות:

i	1	2	3	4	5	6
X_i	3	2	4	-2	1	4
Y_i	2	0	0	1	-5	2

א. $\sum_{i=1}^4 y_i$

ב. $\sum_{i=1}^6 a$

ג. $\sum_{i=1}^6 x_i y_i$

ד. $\sum_{i=1}^6 (x_i + y_i)$

ה. $\sum_{i=1}^6 x_i + a$

- (3) קבעו לכל זהות האם היא נכונה:

א. $\sum_{i=1}^n bX_i = b \cdot \sum_{i=1}^n X_i$

ב. $\sum_{i=1}^n a = a \cdot n$

ג. $\left(\sum_{i=1}^n X_i \right)^2 = \sum_{i=1}^n X_i^2$

(4) נתון: $\sum_{i=1}^{10} X_i = 80$, $\sum_{i=1}^{10} X_i^2 = 1640$

חשבו: $\sum_{i=1}^{10} (X_i - 4)^2$

תשובות סופיות:

(1	א. 7.	ב. 9.	ג. 11.	ד. 121.
	ה. 14.	ו. 27.	ז. 126.	
(2	א. 3.	ב. 12.	ג. 7.	
	ד. 12.	ה. 14.		
(3	א. נכונה.	ב. נכונה.	ג. לא נכונה.	
(4	1160.			

סטטיסטיקה

פרק 4 - סטטיסטיקה תיאורית - מדדי מרכז

תוכן העניינים

1. כללי (ללא ספר)

סטטיסטיקה

פרק 5 - סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן

תוכן העניינים

1. כללי 16

סטטיסטיקה תיאורית – מדדי פיזור – הטווח, השונות וסטיית התקן:

רקע:

המטרה: למדוד את הפיזור של הנתונים, כלומר כמה הם רחוקים זה מזה ושונים זה מזה.

הטווח / תחום (RANGE):

ההפרש בין התצפית הגבוהה ביותר לנמוכה ביותר: $R = X_{\max} - X_{\min}$.

שונות וסטיית תקן:

שונות היא ממוצע ריבועי של הסטיות מהממוצע וסטיית התקן היא שורש של השונות.

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 \quad \text{עבור סדרת נתונים}$$

דוגמאות:

(1) נחשב את השונות של סדרת המספרים הבאה: 5, 4, 9.

$$S_x^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{\sum x^2 \cdot f}{n} - \bar{x}^2 \quad \text{עבור טבלת שכיחויות}$$

(2) להלן התפלגות הציונים בכיתה מסוימת בה ממוצע הציונים הוא 7.44.

הציון X	השכיחות F	$x^2 \cdot F$
5	2	50
6	4	144
7	8	392
8	5	320
9	4	324
10	2	200
סה"כ		1430

$$S_x^2 = \frac{\sum x^2 f(x)}{n} - \bar{x}^2 = \frac{1430}{25} - 7.44^2 = 1.8464$$

$$S = \sqrt{S_x^2} = \sqrt{1.8464} = 1.3588$$

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

שאלות:

- (1) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
 6, 5, 8, 7, 6, 6, 7, 8, 5, 4, 6, 10, 9, 8, 6, 7.
 חשבו את השונות, סטיית התקן והטווח של הציונים.

- (2) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

מספר מכוניות למשפחה	1	2	3	4	5
שכיחות	65	150	220	140	55

- א. חשבו סטיית התקן.
 ב. חשבו את הטווח של הנתונים.
 הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם.
- (3) בחברה העוסקת בטלמרקטינג בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4 שנים וסטיית התקן היא שנתיים.
 א. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים עם וותק של 4 שנים להתפלגות?
 ב. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים אשר אחד עם וותק של 0 שנים והשני עם וותק של 8 שנים להתפלגות?
- (4) נתונה רשימה של 5 תצפיות, אך רק עבור 4 מהן נרשמו הסטיות שלהן מהממוצע: 2, 3, 2, 1-. חשבו את השונות של חמש התצפיות.
- (5) בשכונה בדקו בכל דירה את מספר החדרים לדירה. בשכונה 200 דירות.

מספר חדרים	פרופורציה
1	0.1
2	0.2
3	0.4
4	0.15
5	

- א. מה הממוצע של מספר החדרים לשכונה בדירה?
 ב. חשבו את סטיית התקן של מספר החדרים לדירה.
 ג. חלק מבעלי הדירות בנות 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע(יקטין, יגדל, לא ישנה) על כל מדד שחישבתם בסעיפים הקודמים

תשובות סופיות:

- (1) שונות : 2.19, סטיית תקן : 1.48, טווח : 6.
- (2) א. סטיית תקן : 1.106. ב. טווח : 4.
- (3) א. ממוצע לא ישתנה, סטיית התקן תקטן.
ב. ממוצע לא ישתנה, סטיית התקן תגדל.
- (4) 10.8.
- (5) א. 3.05. ב. 1.16. ג. ממוצע: יקטן, סטיית התקן: תגדל.

סטטיסטיקה

פרק 6 - סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן

תוכן העניינים

1. כללי 19

סטטיסטיקה תיאורית – מדדי מיקום יחסי – ציון תקן:

רקע:

המטרה למדוד איך תצפית ממוקמת ביחס לשאר התצפיות בהתפלגות.

ציון תקן:

הנוסחה לציון תקן של תצפית היא: $Z = \frac{X - \bar{X}}{S}$.

- ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע. כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע:
- ציון תקן חיובי אומר שהתצפית מעל הממוצע.
 - ציון תקן שלילי אומר שהתצפית מתחת לממוצע.
 - ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

דוגמה (פתרון בהקלטה):

במקום עבודה מסוים, ממוצע המשכורות הוא 8 אלף ₪, עם סטית תקן של אלפיים ₪. באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים, עם סטית תקן של 1.5 שנים. ערן מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערן יותר, באופן יחסי, משכיל או משתכר?

שאלות:

- (1) תלמידי כיתה ח' ניגשו למבחן בלשון ולמבחן במתמטיקה. להלן התוצאות שהתקבלו:

המקצוע	ממוצע	סטיית תקן
לשון	74	12
מתמטיקה	80	16

עודד קיבל: 68 בלשון ו-70 במתמטיקה.

- א. באיזה מקצוע עודד טוב יותר באופן יחסי לשכבה שלו?
 ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שיהיה שקול לציונו בלשון?

- (2) במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום. להלן טבלה המסכמת את המידע שנאסף על שני המשתנים:

ממוצע	תפוקה	מספר פועלים
ממוצע	48	15
סטיית תקן	10	2

- באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים.
 מה יותר חריג באותו היום, יחסית לשאר הימים שנבדקו: נתוני התפוקה או כמות הפועלים?
 א. התפוקה.
 ב. כמות הפועלים.
 ג. חריגים באותה מידה.
 ד. חסרים נתונים כדי לדעת זאת.

- (3) הגובה הממוצע של המתגייסים לצבא הוא 175 סנטימטר עם סטיית תקן של 10 סנטימטר. המשקל הממוצע הוא 66 ק"ג עם סטיית תקן של 8 ק"ג. ערך התגייס כשגובהו 180 ס"מ ומשקלו 59 ק"ג.
 א. במה ערך חריג יותר ביחס לשאר המתגייסים, גובהו או משקלו?
 ב. כמה ערך אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

תשובות סופיות:

- (1) א. לשון. ב. 72.
 (2) ב'.
 (3) א. משקל. ב. 70.

סטטיסטיקה

פרק 7 - סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה

תוכן העניינים

1. כללי 21

סטטיסטיקה תיאורית – מדדי מיקום יחסי – אחוזונים בטבלה בדידה:

רקע:

האחוזון (המאון) ה- p הוא הערך בנתונים המחלק את הנתונים בצורה כזאת, שעד אליו (כולל) יש $p\%$ מהנתונים. מסמנים את האחוזון ה- p ב- X_p .

חישוב האחוזון מתוך נתונים בטבלת שכיחויות בדידה:

האחוזון הוא הערך שבו בפעם הראשונה השכיחות היחסית המצטברת (באחוזים) גדולה או שווה ל- $p\%$.

דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו:

שכיחות יחסית מצטברת	שכיחות מצטברת	$F(x)$	# תוכניות החיסכון
		100	0
		75	1
		25	2
		25	3
		25	4

א. מצאו את האחוזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

שאלות:

(1) להלן התפלגות של משתנה כלשהו:

$F(x)$	X
10	0
40	1
30	2
15	3
5	4

מצאו להתפלגות את:

- א. האחוזון ה-60.
- ב. המאון ה-40.
- ג. העשירון העליון.
- ד. הטווח בין הרבעונים.

(2) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

מספר מכוניות למשפחה	1	2	3	4	5
שכיחות	65	150	220	140	55

חשבו את:

- א. העשירון התחתון.
- ב. האחוזון ה-30.
- ג. הערך ש-20% מהתצפית גדולות ממנו.
- ד. רבעון עליון.

תשובות סופיות:

- (1) א. 2. ב. 1. ג. 3. ד. 1.
- (2) א. 1. ב. 2. ג. 4. ד. 4.

סטטיסטיקה

פרק 8 - סטטיסטיקה תיאורית - טרנספורמציה לינארית

תוכן העניינים

1. כללי 23

סטטיסטיקה תיאורית – טרנספורמציה לינארית:

רקע:

מצב שבו מבצעים שינוי מסוג הוספה (או החסרה) של קבוע, והכפלה (או חילוק) של קבוע, לכל התצפיות: $y = a \cdot x + b$. כך יושפעו המדדים השונים:

$$\begin{aligned} MR_y &= a \cdot MR + b \\ MO_y &= a \cdot MO + b \\ \bar{y} &= a \cdot \bar{x} + b \\ Md_y &= a \cdot Md_x + b \end{aligned} \quad \text{מדדי המרכז:}$$

$$\begin{aligned} R_y &= |a| R_x \\ S_y &= |a| S_x \\ S_y^2 &= a^2 S_x^2 \end{aligned} \quad \text{מדדי הפיזור:}$$

$$\begin{aligned} Y_p &= a \cdot X_p + b \\ Z_Y &= \frac{a}{|a|} Z_x \end{aligned} \quad \text{מדדי המיקום היחסי:}$$

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 (פתרון בהקלטה):

השכר הממוצע של עובדים הינו 9000 ₪ וטווח 6000 ₪. חשבו את המדדים הללו לאחר שהעלו את כל המשכורות ב-10% ואחר כך קנסו אותם ב-100 ₪.

שאלות:

- (1) עבור סדרת נתונים התקבל: $\bar{x} = 80, S = 15, MO = 70$. הוחלט להכפיל את כל התצפיות ב-4 ולהחסיר מהתוצאה 5. חשבו את המדדים הללו לאחר השינוי.
- (2) בחברה מסוימת השכר הממוצע הוא 40 ₪ לשעה עם סטיית תקן של 5 ₪ לשעה. הוחלט להעלות את כל המשכורות ב-10%, אך זה לא סיפק את העובדים ולכן הם קיבלו לאחר מכן תוספת של 2 ₪ לשעה. מה הממוצע ומהי השונות של השכר לשעה לאחר כל השינויים.
- (3) במבחן מסוים הציון החציוני היה 73, טווח הציונים היה 40 נקודות והעשירון העליון היה הציון 87. כיוון שהציונים בבחינה היו נמוכים, המורה החליט לתת פקטור של 4 נק' לכל התלמידים. חשבו את המדדים לאחר הפקטור.
- (4) דגמו מקו ייצור 50 קופסאות של גפרורים. בדקו בכל קופסא בה יש 40 גפרורים את כמות הגפרורים הפגומים. התקבל שבממוצע יש 3 גפרורים פגומים בקופסא, עם סטיית תקן של 1.5 גפרורים. מה יהיה הממוצע ומה תהיה סטיית התקן של מספר התקינים בקופסא?
- (5) חברת בזק הציעה את ההצעה הבאה: שלושים שקלים דמי מנוי חודשיים קבועים וכן 10 אגורות לכל דקה של שיחה יוצאת. אדם בדק במשך שנה את דקות השיחות היוצאות שלו, וקיבל שבממוצע חודשי יש לו 600 דקות שיחות יוצאות עם שונות של 2500 דקות רבועות, כמו כן בחודש ינואר ציון התקן היה 2. חשבו את המדדים הללו עבור חשבון הטלפון החודשי של אותו אדם בשקלים אם היה משתמש בחבילה המוצעת לו על ידי בזק.
- (6) הוכיחו שאם כל התצפיות בהתפלגות עברו טרנספורמציה לינארית: $Y_i = a \cdot X_i + b$, אזי הממוצע והשונות של כלל התצפיות לאחר הטרנספורמציה יהיו בהתאמה:
- $$\bar{y} = a \cdot \bar{x} + b, S_y^2 = a^2 S_x^2$$

תשובות סופיות:

- (1) ממוצע: 315, סטיית תקן: 60, שכיח: 275.
- (2) ממוצע: 46, שונות: 30.25.
- (3) טווח: 40, חציון: 77, עשירון עליון: 91.
- (4) ממוצע: 37, סטיית תקן: 1.5.
- (5) ממוצע: 90, שונות: 25, ציון תקן: 2.
- (6) $a^2 \cdot S_x^2$

סטטיסטיקה

פרק 9 - סטטיסטיקה תיאורית שאלות אמריקאיות

תוכן העניינים

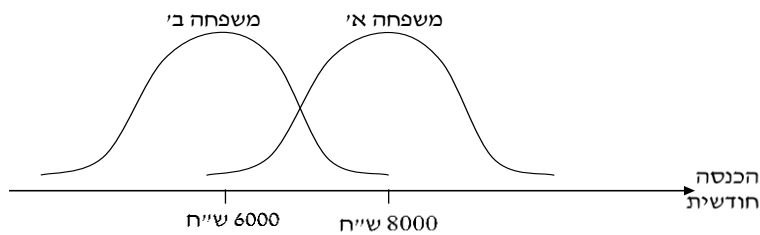
1. כללי 26

סטטיסטיקה תיאורית – שאלות אמריקאיות:

שאלות:

שאלות 1-3 מתייחסות לקטע הבא:

להלן שתי עקומות המתארות את התפלגות ההכנסות החודשיות של שתי משפחות שנבחרו באקראי:



- (1) לאיזו משפחה הכנסה שכיחה גבוהה יותר?
 - א. משפחה א'.
 - ב. משפחה ב'.
 - ג. לשתיהן אותה הכנסה שכיחה.
 - ד. לא ניתן לדעת – אין מספיק נתונים.
- (2) באיזו משפחה ההכנסה החציונית שווה להכנסה הממוצעת?
 - א. משפחה א'.
 - ב. משפחה ב'.
 - ג. בשתיהן ההכנסה החציונית שווה להכנסה הממוצעת.
 - ד. לא ניתן לדעת – אין מספיק נתונים.
- (3) באיזו משפחה סטית התקן של ההכנסה החודשית גבוהה יותר?
 - א. משפחה א'.
 - ב. משפחה ב'.
 - ג. לשתיהן אותה סטית תקן.
 - ד. לא ניתן לדעת – אין מספיק נתונים.

הנתונים הבאים מתייחסים לשאלות 4-6:

להלן נתונים חלקיים של טבלת שכיחויות:
כמו כן, נתון כי הממוצע הוא 1.66.

$F(x)$	x
?	0
10	1
6	2
15	3
?	4
50	סה"כ

(4) השכיח של הנתונים הוא:

- א. 0.
- ב. 15.
- ג. ישנם שני שכיחים: 0 ו-3.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של השכיח.

(5) חציון הנתונים הוא:

- א. 2.
- ב. 1.5.
- ג. 25.5.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(6) הטווח של הנתונים:

- א. 11.
- ב. 3.
- ג. 4.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(7) בהתפלגות אסימטרית ימנית של משתנה כמותי רציף, הערך המתאים למאון ה-30, ציון התקן שלו הוא בהכרח:

- א. שלילי.
- ב. חיובי.
- ג. אפס.
- ד. לא ניתן לדעת ללא הנתונים.

- 8) סדרת נתונים סטטיסטיים מונה 10 תצפיות. נתון כי סדרת הנתונים סימטרית סביב הממוצע. ממוצע הסדרה-40 ושונות הסדרה-100. בשלב מאוחר יותר נוספו שתי תצפיות נוספות לסדרה : 50 ו-30. השונות של 12 התצפיות :
- א. תקטן.
 - ב. תגדל.
 - ג. לא תשתנה.
 - ד. לא ניתן לחשב את השונות ללא ידיעת התצפיות.

- 9) הוספת גודל קבוע לכל תצפיות סדרת נתונים :
- א. תגדיל את סטיית התקן.
 - ב. תקטין את סטיית התקן.
 - ג. לא תשנה את סטיית התקן.
 - ד. לא ניתן לדעת.

הנתונים הבאים מתייחסים לשאלות 10-11:

להלן נתונים על ציוני תלמידים שנבחנו במועדים שונים בסטטיסטיקה :

שם התלמיד	ציון	ממוצע הציונים במועד בו נבחן	סטיית התקן של הציונים במועד בו נבחן
צבי	50	50	12
סטף	82	80	5
שרית	65	60	15
לובה	60	63	1.5
מיטב	70	70	10

- 10) התלמיד הטוב ביותר ביחס לנבחנים באותו מועד בו נבחן הוא :
- א. מיטב.
 - ב. צבי.
 - ג. לובה.
 - ד. שרית.
 - ה. סטף.

- 11) פנינה נבחנה עם סטף וציון התקן שלה שווה לציון התקן של שרית לכן ציונה הוא :
- א. 80.55.
 - ב. 65.
 - ג. 80.
 - ד. 81.66.

הנתונים הבאים מתייחסים לשאלות 12-15:

בבדיקת פתע של משרד הבריאות במפעל שוקולד, נמצא ש:

שוקולד פגום	0	1	2	3	4	5	6	7
מס' קופסאות	35	63	48	12	13	11	10	8

12) מהו החציון של מספר הפגומים בקופסא:

- א. 1.
- ב. 2.
- ג. 4.
- ד. לא ניתן לדעת.

13) מהו הרבעון התחתון של מספר הפגומים בקופסא?

- א. 1.
- ב. 2.
- ג. 3.
- ד. 4.
- ה. לא ניתן לדעת.

14) מספר הפגומים בקופסא הוא משתנה:

- א. סדר.
- ב. שמי.
- ג. כמותי בדיד.
- ד. כמותי רציף.

15) השכיח של מספר הפגומים בקופסא:

- א. 63.
- ב. 1.
- ג. 200.
- ד. לא ניתן לדעת.

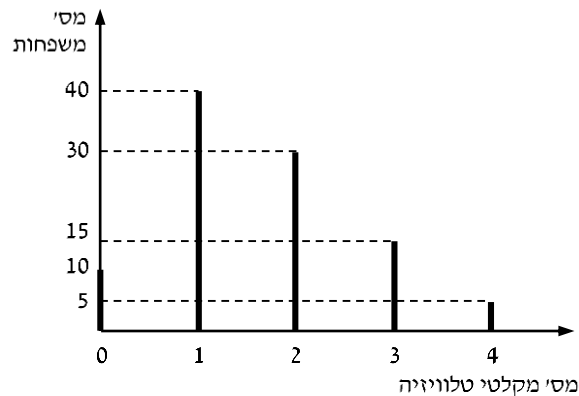
16) ביחס לציר המספרים, רוב הערכים בהתפלגות א-סימטרית ימנית נמצאים:

- א. בערכים הגבוהים.
- ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
- ג. בערכים הנמוכים.
- ד. לא ניתן לדעת.
- ה. אף לא תשובה מהנ"ל נכונה.

- 17) בוצע מחקר על מספר העובדים בחברות מזון לעומת חברות תקשורת. החציון והממוצע בשתיהן שווה 8. איזה מהטענות הבאות היא הנכונה והמלאה ביותר:
- השכיחות ב-2 החברות זהה אך שונה מ-8.
 - השכיח ב-2 החברות זהה אך לא ניתן לדעת מהו.
 - השכיח בשתי חברות הינו בהכרח 8.
 - שכיח בחברה אחת שונה מ-8 ובשנייה הוא 8.
 - אף תשובה אינה נכונה.

הנתונים הבאים מתייחסים לשאלות 18 עד 22:

נערך סקר על מספר מקלטי הטלוויזיה הנמצאים בבית. תוצאות הסקר נתונות בדיאגרמת מקלות הבאה:



18) המשתנה הנחקר כאן הוא:

- משתנה שמי.
- משתנה מסולם סדר.
- משתנה כמותי בדיד.
- משתנה כמותי רציף.

19) הטווח של ההתפלגות הוא:

- 35.
- 4.
- 3.
- 2.

20) ממוצע מספר מקלטי הטלוויזיה למשפחה הוא :

- א. 1.65
- ב. 1.5
- ג. 1
- ד. 2

21) השכיח של התפלגות זו היא :

- א. 40
- ב. 1.5
- ג. 1
- ד. 2

22) מסתבר שיש בין 2 ל-5 משפחות נוספות שאין להם מקלטי טלוויזיה ויש לצרף את המשפחות הללו להתפלגות. כיצד הנתון זה ישפיע על סטיית התקן?

- א. יקטין אותו.
- ב. יגדיל אותו.
- ג. לא ישנה אותו.
- ד. אין לדעת.

תשובות סופיות:

1) א'	2) ג'	3) ג'	4) ג'	5) ב'
6) ג'	7) א'	8) ג'	9) ב'	10) ה'
11) ד'	12) ג'	13) א'	14) ג'	15) ב'
16) ג'	17) ה'	18) ג'	19) ב'	20) א'
21) ג'	22) ב'			

סטטיסטיקה

פרק 10 - יסודות ההסתברות

תוכן העניינים

1. כללי 32

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 הטלויזיה 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

פרק 11 - פעולות בין מאורעות (חיתוך ואיחוד) - מאורעות זרים ומכילים

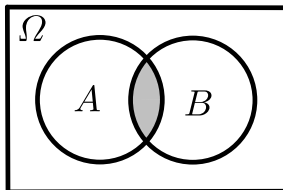
תוכן העניינים

1. כללי 36

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

רקע:

פעולת חיתוך:

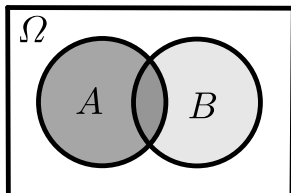


נותנת את המשותף בין המאורעות הנחתכים.
חיתוך בין המאורע A למאורע B יסומן כך: $A \cap B$.
מדובר בתוצאות שנמצאות ב- A וגם ב- B .

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:



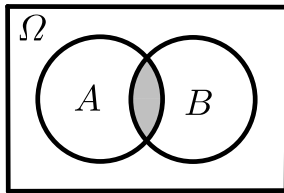
נותנת את כל האפשרויות שנמצאות לפחות באחת מהמאורעות, ומסומנת: $A \cup B$.
הפעולה נותנת את אשר נמצא ב- A או ב- B .
כלומר, לפחות אחד מהמאורעות קורה.

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:

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

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

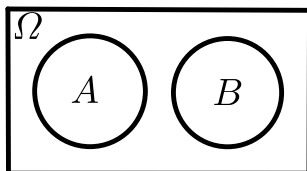
$$P(A \cap B) = 1 - P(\bar{A} \cup \bar{B})$$

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

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.

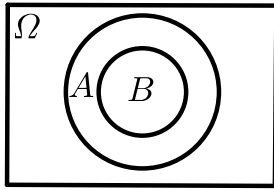
ההסתברות של חיתוך המאורעות היא אפס: $P(A \cap B) = 0$.

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:

$$(1) \text{ א. } A \cap B = \{EG, EF, FE, GE\}$$

$$\text{ב. } A \cup B = \{EG, EF, EE, FE, GE, EG, GF\}$$

$$(2) \text{ א. } B \cap \bar{A} \quad \text{ב. } A \cap \bar{B} \quad \text{ג. } A \cap B \quad \text{ד. } A \cup B \quad \text{ה. } \bar{A} \cap \bar{B} \quad \text{ו. } \bar{B}$$

$$(3) \text{ א. } \bar{B} = 5, 6, 7, 8, 9, B = 0, 1, 2, 3, 4, A = 0, 2, 4, 6, 8$$

$$A \cup B = 0, 2, 4, 6, 8, 1, 3, A \cap B = 0, 2, 4$$

$$\text{ב. } P(A \cup B) = 0.7, P(A \cap B) = 0.3, P(\bar{B}) = 0.5, P(B) = 0.5, P(A) = 0.5$$

$$(4) \text{ א. } \bar{\bar{A}} = A, A \cap \phi = \phi, A \cup \phi = A, A \cap \Omega = A, A \cup \Omega = \Omega$$

$$A \cup \bar{A} = \Omega, \bar{\phi} = \Omega, A \cap \bar{A} = \phi$$

$$(5) \text{ א. } A \cap B: \text{גובה בין } 1.7 \text{ ל-} 1.8. \text{ ב. } A \cup B: \text{כל גובה אפשרי.}$$

$$\text{ג. } \bar{A} = \bar{A} \cap B: \text{גובה לכל היותר } 1.7. \text{ ד. } \bar{A} \cup \bar{B}: \text{לכל היותר } 1.7 \text{ או לפחות } 1.8.$$

$$\text{ה. } A = \bar{\bar{A}}: \text{גובה מעל } 1.7.$$

$$(6) \text{ א. } A \cap B \cap C \quad \text{ב. } A \cap \bar{B} \cap \bar{C} \quad \text{ג. } A \cup B \cup C$$

$$\text{ד. } \bar{C} \quad \text{ה. } (A \cap B \cap \bar{C}) \cup (B \cap C \cap \bar{A}) \cup (A \cap C \cap \bar{B})$$

$$(7) \text{ א. } P(A \cup B) = 0.24 \quad \text{ב. } P(A \cap B) = 0.04 \quad \text{ג. } P(B \cap \bar{A}) = 0.16$$

$$(8) \text{ א. } P(A \cap B) = 10\% \quad \text{ב. } P(A \cup B) = 50\% \quad \text{ג. } P(\bar{A} \cap \bar{B}) = 50\%$$

$$(9) \text{ א. } P(A \cap B) = 0.2 \quad \text{ב. } P(\bar{A} \cap \bar{B}) = 0.3 \quad \text{ג. } P(A \cup \bar{B}) = 0.3$$

$$(10) \text{ א. לא. ב. כן. ג. כן. ד. לא.}$$

$$(11) \text{ א. כן. ב. } P(\bar{A} \cap B) = 0.3$$

$$(12) \text{ הטענה הנכונה היא ג'.$$

$$(13) \text{ א. } 0.05 \quad \text{ב. } 0.95$$

$$(14) \text{ א. } P(A \cap B) = 0.06 \quad \text{ב. לא. ג. } P((A \cap \bar{B}) \cup (B \cap \bar{A})) = 0.43$$

$$(15) \text{ } P(B) = \frac{1}{5}, P(A) = \frac{2}{5}$$

$$(16) \text{ א. נכון. ב. לא נכון. ג. לא נכון. ד. נכון.}$$

סטטיסטיקה

פרק 12 - הסתברות מותנית-במרחב מדגם אחד

תוכן העניינים

1. כללי 44

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?

תשובות סופיות:

(1) 0.2

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

(3) 0.5

(4) 0

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

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

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

סטטיסטיקה

פרק 13 - הסתברות מותנית - מרחב לא אחיד

תוכן העניינים

1. כללי 47

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?

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. $\frac{23}{30}$

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

ה. 0.3 ו. 0.72

סטטיסטיקה

פרק 14 - הערכת כלים אבחנתיים

תוכן העניינים

1. הערכת כלים אבחנתיים 50

הערכת כלים אבחנתיים:

רקע:

אנו מנסים לאבחן תכונה מסוימת באמצעות כלי מסוים (למשל, לאבחן האם לאדם יש קורונה באמצעות ערכת אבחון ביתית). נגדיר מדדים סטטיסטיים שונים שנותנים אינדיקציה לאיכות כלי האבחנה. נסמן ב-A: האדם קיבל תשובה חיובית, כלומר אובחן כבעל התכונה. נסמן ב-B: האדם הוא בעל התכונה.

רגישות – Sensitivity:

ההסתברות שאדם בעל התכונה יקבל תשובה חיובית, כלומר: $Sensitivity = P(A|B)$.

סגוליות – Specificity:

ההסתברות שאדם ללא התכונה יקבל תשובה שלילית, כלומר: $Specificity = P(\bar{A}|\bar{B})$.

דוגמה (פתרון בהקלטה):

5,000 נשים השתתפו במחקר שבו הן השתמשו בערכה לבדיקת היריון ביתית של חברה מסוימת ביום ה-14 של המחזור החודשי. בדעיבד מתוך 5,000 הנשים 80 היו בהיריון. 70 נשים קיבלו תשובה חיובית מערכת הבדיקה הביתית. 65 מהן אכן היו בהיריון. מה הסגוליות ומה הרגישות של בדיקת ההיריון הביתית?

הערך המנבא החיובי – Positive Predictive Value:

ההסתברות שאדם שקיבל תשובה חיובית הוא אכן בעל התכונה. הערך המנבא החיובי נקרא גם יכולת אבחון (יכולת דיאגנוסטית): $PPV = P(B|A)$.

הערך המנבא השלילי – Negative Predictive Value:

ההסתברות שאדם שקיבל תשובה שלילית אינו בעל התכונה: $NPV = P(\bar{B}|\bar{A})$.

המשך הדוגמה (פתרון בהקלטה):

מהו הערך המנבא החיובי ומהו הערך המנבא השלילי של ערכת הבדיקה הביתית?

ההסתברות לתוצאה חיובית מדומה – False Positive:

ההסתברות שאדם שאינו בעל התכונה יקבל תשובה חיובית: $FP = P(A|\bar{B})$.

ההסתברות לתוצאה שלילית מדומה – False Negative:

ההסתברות שאדם בעל התכונה יקבל תשובה שלילית: $FN = P(\bar{A}|B)$.

המשך הדוגמה (פתרון בהקלטה):

מה ההסתברות לתוצאה חיובית מדומה ומה ההסתברות לתוצאה שלילית מדומה בבדיקה באמצעות ערכת הבדיקה הביתית?

שאלות:

- (1) 2,000 גברים נבדקו בשיטה חדשה לגילוי סרטן המעי הגס. מתוך 150 גברים שביופסיה הוכיחה בוודאות שהם חולים 100 נמצאו חולים באמצעות הבדיקה החדשה. 80 גברים שהוכח שהם בריאים באמצעות ביופסיה קיבלו תשובה חיובית באמצעות השיטה החדשה.
מצאו את הסגוליות, הרגישות, הערך המנבא החיובי, הערך המנבא השלילי, ההסתברות לתוצאה חיובית מדומה.
- (2) הסיכוי של ערכת בדיקה ביתית של חברת קאנו לגלות מחלה מסוימת הוא 98%. לאדם בריא יש סיכוי של 5% לקבל בבדיקה תשובה חיובית. 5% ממשתמשי הערכה חולים במחלה.
א. מה הרגישות של הערכה?
ב. מה הסגוליות של הערכה?
ג. מה יכולת האבחון של הערכה?
- (3) הסיכוי של תהליך אבחון של הפרעת קשב לזהות סטודנטים בעלי הפרעת קשב הוא 0.95. הסיכוי שלו לזהות בטעות גם סטודנטים ללא הפרעת קשב כבעלי הפרעת קשב הוא 0.01. ידוע ש-7% מהסטודנטים סובלים מהפרעת קשב.
א. מה ההסתברות לתשובה חיובית מדומה?
ב. מה ההסתברות לתשובה שלילית מדומה?
- (4) נערכה בדיקה בשיטה חדשה לגילוי מלנומה בקרב 800 נבדקים – 400 נבדקים עם ביופסיה מוכחת של מלנומה ויתר הנבדקים ידועים ללא מלנומה. 30 מהנבדקים נמצאו בבדיקה החדשה חולים במלנומה. מתוך ה-30 רק 10 חולים באמת.
א. מה הרגישות של הבדיקה החדשה?
ב. מה הסגוליות של הבדיקה החדשה?
ג. מה היכולת הדיאגנוסטית של הבדיקה החדשה?
- (5) קבוצה של נבדקים שידוע ש-25% מהם נשאי HIV נבדקה בבדיקה חדשה המאפשרת קבלת תוצאה באופן מיידי. בבדיקה החדשה נמצאו 20% מהנבדקים כנשאי HIV. הסגוליות של הבדיקה החדשה היא 76%.
א. מה הרגישות של הבדיקה החדשה?
ב. מה ה-PPV של הבדיקה החדשה?
ג. מה ה-NPV של הבדיקה החדשה?

תשובות סופיות:

- (1) סגוליות: 0.9568, רגישות: $\frac{2}{3}$, הערך המנבא החיובי: $\frac{5}{9}$,
 הערך המנבא השלילי: 0.9725, ההסתברות לתוצאה חיובית מדומה: 0.0043.
- (2) א. 0.98 ב. 0.95 ג. 0.5078
- (3) א. 0.01 ב. 0.05
- (4) א. 0.25 ב. 0.95 ג. $\frac{1}{3}$
- (5) א. 0.08 ב. 0.1 ג. 0.7125

סטטיסטיקה

פרק 15 - דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה

תוכן העניינים

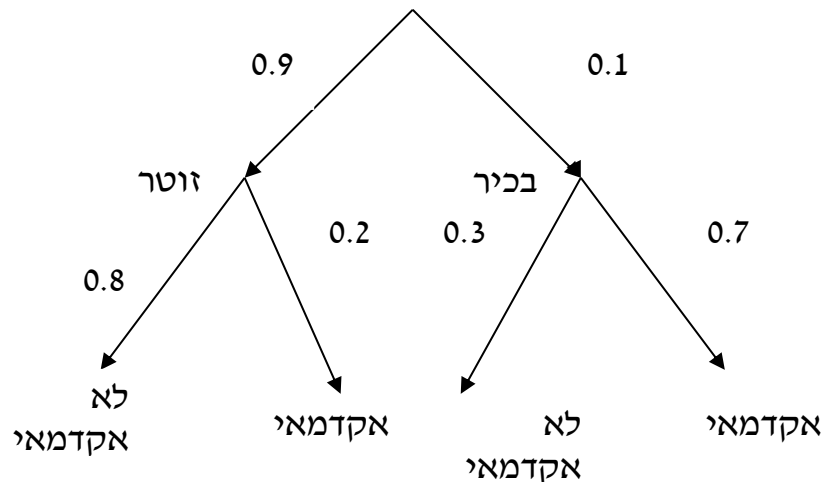
1. כללי 54

דיאגרמת עצים – נוסחת בייס ונוסחת ההסתברות השלמה:

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. 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 מותנה:

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

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 סוכרייה. אם היא בטעם תות אוכלים אותה ומוציאים סוכרייה נוספת, ואם היא בטעם לימון מחזירים אותה לשקית ומוציאים סוכרייה 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?

תשובות סופיות:

		ב. $\frac{23}{49}$	א. $\frac{2}{7}$	(1)
ד. 0.2	ג. 0.241	ב. 58%	א. 6%	(2)
		ב. 0.5	א. 0.544	(3)
	ג. 0.9722	ב. 0.09375	א. 9%	(4)
	ג. 0.2442	ב. 0.3488	א. 0.14	(5)
		ב. 0.8125	א. 70%	(6)
	ג. 0.7543	ב. 0.3158	א. 0.57	(7)
ד. 0.8778	ג. 0.3137	ב. 0.2889	א. 0.0886	(8)

סטטיסטיקה

פרק 16 - תלות ואי תלות בין מאורעות

תוכן העניינים

1. אי תלות בין מאורעות (מורחב) 59

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 יחדו?
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.

(9) זוג מעוניין להביא ילד לעולם, בבדיקה גנטית שהם עשו לאב התגלה שהאב אינו נשא של מחלה Q . מערכים את הסיכוי של האם להיות נשאית למחלה Q להיות 0.2. אם האם נשאית היא תלד בכל פעם ילד חולה בסיכוי 0.5 באופן בלתי תלוי בין הלידות. האם ילדה שני ילדים. האם המאורעות "הילד הראשון בריא" ו-"הילד השני בריא" הם מאורעות בלתי תלויים?

(10) מטילים פעמיים מטבע עם הסתברות p לעץ בכל הטלה, $0 < p < 1$.

A – יצא עץ בהטלה ראשונה.

B – יצאו תוצאות שונות.

עבור איזה ערכים של p המאורעות A ו- B בלתי תלויים?

(11) הוכח אם A ו- B בלתי תלויים, אזי $\bar{A}$, $\bar{B}$ בלתי תלויים.

(12) נתונה מערכת חשמלית שבשרטוט, כל מתג יכול להיות פתוח או סגור

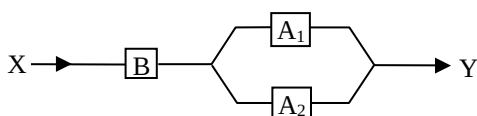
בהסתברויות שונות, אך באופן בלתי תלוי זה בזה.

להלן ההסתברות של כל מתג להיות סגור :

$$P(A_1) = 0.7, P(A_2) = 0.8, P(B) = 0.9$$

א. מה ההסתברות שיעבור זרם במערכת החשמלית?

ב. אם לא עובר זרם במערכת, מה הסיכוי שמתג B סגור?



(13) מטילים שתי קוביות הוגנות. נגדיר שלושה מאורעות :

A – תוצאה של קובייה ראשונה זוגית.

B – תוצאה של קובייה שניה אי זוגית.

C – סכום התוצאות של שתי הקוביות זוגי.

האם המאורעות בלתי-תלויים?

(14) ענה על הסעיפים הבאים :

א. המאורעות A ו- B הם מאורעות זרים של ניסוי כלשהו.
חוזרים על אותו ניסוי שוב באופן בלתי תלוי זה בזה.
הוכיחו שהסיכוי שמאורע A יתרחש בניסוי לפני שמאורע B יתרחש

$$\frac{P(A)}{P(A) + P(B)}$$

בניסוי הוא :

ב. מטילים קובייה הוגנת פעם אחרי פעם, מה הסיכוי שתתקבל תוצאה זוגית עוד לפני שתתקבל התוצאה 3?
ג. מטילים קובייה הוגנת פעם אחרי פעם, מה הסיכוי שתתקבל תוצאה זוגית עוד לפני שתתקבל תוצאה גדולה מ-4?

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.
- (9) תלויים.
- (10) 0.5
- (11) שאלת הוכחה.
- (12) א. 0.846 ב. 0.3506
- (13) תלויים.
- (14) א. שאלת הוכחה. ב. $\frac{3}{4}$ ג. $\frac{1}{2}$

סטטיסטיקה

פרק 17 - שאלות מסכמות בהסתברות

תוכן העניינים

1. כללי 63

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.
 - ד. אם ההסתברות שרק המכונית החדשה שלה היא מתוצרת אירופאית?
 - ה. 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 ייצאו 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 ייצאו רק לבאולינג וביום למחרת לא ייצאו 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:

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

סטטיסטיקה

פרק 18 - המשתנה המקרי הבדיד - פונקציית ההסתברות

תוכן העניינים

1. כללי 66

המשתנה המקרי הבדיד – פונקציית ההסתברות:

רקע:

משתנה מקרי בדיד:

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

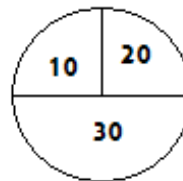
מתארים את המשתנה המקרי על ידי פונקציית ההסתברות.

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 ב-ש. 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 .

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	$P(X)$

סטטיסטיקה

פרק 19 - המשתנה המקרי הבדיד - תוחלת - שונות וסטיית תקן

תוכן העניינים

1. כללי 69

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.

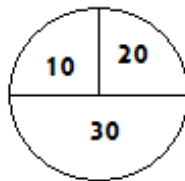
$$\text{מגדירים שונות באופן הבא: } V(X) = \sum_i (x_i - \mu)^2 P(x_i) = \sum_i x_i^2 P(x_i) - \mu^2 = \sigma^2$$

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: $V(X)$.

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	$P(X)$

סטטיסטיקה

פרק 20 - המשתנה המקרי הבדיד - תוחלת של פונקציה של משתנה מקרי
בדיד

תוכן העניינים

1. ראשי.....72

תוחלת של פונקציה של משתנה מקרי בדיד:

רקע:

יהי X משתנה מקרי, ותהי $g(X)$ פונקציה של X . אז:

$$\begin{aligned} E(g(X)) &= g(x_1)P(X=x_1) + g(x_2)P(X=x_2) + g(x_3)P(X=x_3) + \dots \\ &= \sum_i g(x_i) \cdot P(x_i) \end{aligned}$$

כאשר $x_1, x_2, x_3, \dots$ הם הערכים שהמשתנה X מקבל.

דוגמה (פתרון בהקלטה):

נתון:

X	0	1	2
$P(X)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

מצאו התפלגות ותוחלת של: $Y = X^2$.

שאלות:

- (1) מסובבים רולטה עליה המספרים 1 עד 4. יהיה X המספר שהתקבל לאחר סיבוב הרולטה. התפלגות X היא כדלהלן:

X	4	3	2	1
$P(X)$	0.3	0.4	0.2	0.1

א. חשבו את: $E(X)$, $E\left(\frac{1}{X}\right)$.

ב. האם: $E\left(\frac{1}{X}\right) = \frac{1}{E(X)}$?

- (2) יהי X משתנה מקרי בעל פונקציית ההסתברות הבאה:

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

חשבו את התוחלת של:

א. X^2 .

ב. 2^X .

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

א. מצאו את ערכו של A .

ב. חשבו את: $E([X - E(X)]^2)$.

- (4) בכל יום משחק ערן משחק יחיד בכל אחת מהאפליקציות הבאות: TWODOTS ו-PIANOTILES. בכל אחד מהמשחקים ישנם שלבים שיש לעבור. משחק בודד מסתיים בהצלחה אם ערן עבר את שלב, ובכישלון אם ערן לא עבר את השלב. ההסתברות שבאפליקציית TWODOTS ערן יעבור שלב היא 0.6 בכל יום. ההסתברות שבאפליקציית PIANOTILES ערן יעבור שלב היא 0.35 בכל יום. נניח שמעבר שלב בכל אחד מהמשחקים בלתי תלוי במשחק אחר. נסמן ב- W את מספר המשחקים שערן יעבור שלב בהם מחר.

א. חשבו את $E(W)$.

ב. חשבו את $E(W^3)$.

(5) יהי X משתנה מקרי בדיד עם תוחלת ושונויות סופיים: $Y = aX + b$, כאשר $a \neq 0$.
 a, b הינם פרמטרים. יש להוכיח ש: $E(Y) = aE(X) + b$, $V(Y) = a^2 \cdot V(X)$.

(6) אלעד צופה בסדרה בת 6 פרקים. 3 פרקים מתוך ה-6 הם פרקים שצולמו בישראל ו-3 פרקים אחרים צולמו בבולגריה. פרק אחד מבין הפרקים שצולמו בבולגריה מצולם כולו ביער. אלעד צופה בפרקי הסדרה בסדר אקראי, עד אשר הוא מגיע לפרק שצולם ביער בבולגריה. נגדיר את W כמספר הפרקים שצולמו בבולגריה שבהם יצפה אלעד.
 א. מהי התפלגות W ?
 ב. חשבו: $E(W^3)$.

(7) למיקה יש 20 חולצות ו-3 מגירות. כאשר מיקה מסדרת את 20 החולצות במגירות היא בוחרת עבור כל חולצה, באופן מקרי ובלתי תלוי בחולצות האחרות, את המגירה אליה תכניס את החולצה (כל אחת מהמגירות יכולה להכיל את כל החולצות).
 נסמן ב- X את מספר המגירות המכילות בדיוק 10 חולצות.
 מצאו את התפלגות X ואת: $E(\sqrt{X+2})$.

(8) מטבע מוטל 10 פעמים. X = מספר הפעמים שהתקבלה התוצאה ראש.
 א. בנו את פונקציית ההסתברות של X .
 ב. הרווח במשחק הוא 4^X . מצאו את התוחלת של הרווח במשחק.
 רמז: היעזרו בבינום של ניוטון: $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$.

תשובות סופיות:

- (1) א. $E\left(\frac{1}{X}\right) = 0.4083$, $E(X) = 2.9$. ב. לא.
- (2) א. 3.2. ב. 3.45.
- (3) א. 10. ב. 1.
- (4) א. 0.95. ב. 2.21.
- (5) הוכחה.
- (6) א. $X \sim U(1,3)$. ב. 12.
- (7) $E(\sqrt{X+2}) = 1.4659$.
- (8) א. $X \sim B(10, 0.5)$. ב. 2.5^{10} .

סטטיסטיקה

פרק 21 - המשתנה המקרי הבדיד - טרנספורמציה לינארית

תוכן העניינים

1. כללי 75

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):

חישובנו 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 הנ"ל?

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: $85,727.5^2$.

סטטיסטיקה

פרק 22 - תוחלת ושונות של סכום משתנים מקריים

תוכן העניינים

1. כללי 78

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 סכום 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:

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

$$0 \text{ אחר } K$$

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

פרק 23 - התפלגויות בדידות מיוחדות - התפלגות בינומית

תוכן העניינים

1. כללי 81

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?

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

סטטיסטיקה

פרק 24 - התפלגויות בדידות מיוחדות - התפלגות אחידה

תוכן העניינים

1. כללי 85

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?

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 ב. 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. $\frac{7}{30}$

סטטיסטיקה

פרק 25 - המשתנה המקרי הבדיד - שאלות מסכמות

תוכן העניינים

1. כללי 88

המשתנה המקרי הבדיד – שאלות מסכמות:

שאלות:

(1) נתון כי: $X \sim B\left(4, \frac{1}{2}\right)$, $Y \sim B\left(10, \frac{1}{4}\right)$.

- חשבו את התוחלת וסטיית התקן של X .
 - חשבו את התוחלת וסטיית התקן של $W = 2X - 4$.
 - חשבו את התוחלת של $T = X + Y$.
- האם ניתן לדעת מה סטיית התקן של T ?

(2) ערן משחק בקזינו בשתי מכונות הימורים, בכל מכונה משחק אחד (במכונה א' ובמכונה ב'). הסיכוי שלו לנצח במשחק במכונה א' הינו 0.08 והסיכוי שלו לנצח רק במכונה א' הינו 0.05. הסיכוי שלו להפסיד בשני המשחקים ביום מסוים הוא 0.88.

- מה הסיכוי שערן ינצח בשני המשחקים?
- מה התוחלת ומהי השונות של מספר הניצחונות של ערן?
- אם ערן נכנס לקזינו 5 פעמים ובכל פעם שיחק את שני המשחקים, מה ההסתברות שערן ינצח בשני המשחקים בדיוק פעם אחת מתוך חמשת הפעמים?

(3) 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 לא להשתמש בו שוב. נסמן ב- X את מספר הניסיונות עד שהדלת תפתח.

- בנו את פונקציית ההסתברות של X .
- חשבו את התוחלת והשונות של X .
- כל ניסיון לפתוח הדלת אורך חצי דקה. מה התוחלת ומה השונות של הזמן הכולל לפתיחת הדלת?

- (6) חברת הפקות של סרטים הפיקה 3 סרטים, אשר הופקו לטלוויזיה המקומית. חברת ההפקות מנסה למכור את הסרטים הללו לחו"ל. להלן ההסתברויות למכירת הסרטים לחו"ל:
- הסרט "הצב" יימכר לחו"ל בסיכוי של 0.6.
- הסרט "לעולם לא" יימכר לחו"ל בסיכוי של 0.7.
- הסרט "מוות פתאומי" יימכר לחו"ל בסיכוי של 0.2.
- ידוע כי כל סרט עלה להפקה חצי מיליון שקלים. כמו כן, כל סרט הביא להכנסה של 200,000 שקלים מהטלוויזיה המקומית. במידה וסרט יימכר לחו"ל, כל סרט יימכר ב-600,000 שקלים.
- א. בנו את פונקציית ההסתברות של מספר הסרטים שיימכרו לחו"ל.
- ב. מהי התוחלת והשונות של מספר הסרטים שיימכרו?
- ג. מהי התוחלת ומהי סטיית התקן של הרווח (במאות אלפי שקלים) של חברת ההפקה?
- (8) מבחן בנוי משני חלקים: בחלק א' 10 שאלות ובחלק ב' 10 שאלות. תלמיד התכוון רק לחלק א' של המבחן ובחלק זה בכל שאלה יש סיכוי של 0.8 שיענה נכון, בחלק השני לכל שאלה יש 4 תשובות כשרק אחת נכונה. בחלק זה הוא מנחש את התשובות.
- א. מהי ההסתברות שבחלק הראשון הוא יענה נכון על 7 שאלות בדיוק?
- ב. מהי ההסתברות שבחלק השני הוא יענה נכון על פחות מ-3 שאלות?
- ג. מה התוחלת ומהי השונות של מספר התשובות הנכונות בחלק הראשון?
- ד. מהי התוחלת ומהי השונות של מספר התשובות הנכונות בבחינה כולה?
- (9) יהי X משתנה מקרי המקיים: $E(X) = 2$ וכן: $V(X) = 1$. חשבו: $E(X - 5)^2$.

תשובות סופיות:

- (1) א. תוחלת: 2, סטיית תקן: 1. ב. תוחלת: 0, סטיית תקן: 2.
ג. תוחלת: 4.5, סטיית תקן: לא ניתן.
- (2) א. 0.03. ב. תוחלת: 0.15, שונות: 0.1875.
ג. 0.1328.
- (3) א. ראו טבלה: ב. תוחלת: 3, שונות: 2.
- | | | | | | |
|-----|-----|-----|-----|-----|--------|
| 5 | 4 | 3 | 2 | 1 | x |
| 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | $P(x)$ |
- ג. תוחלת: 1.5, שונות: 0.5.
- (6) א. ראו טבלה: ב. תוחלת: 1.5, שונות: 0.61.
- | | | | | |
|-------|-------|-------|-------|--------|
| 3 | 2 | 1 | 0 | x |
| 0.084 | 0.428 | 0.392 | 0.092 | $P(x)$ |
- ג. תוחלת: 0, סטיית תקן: 4.68.
- (8) א. 2.013. ב. 0.5256. ג. תוחלת: 8, שונות: 1.6.
ד. תוחלת: 10.5, שונות: 3.475.
- (9) 10.

סטטיסטיקה

פרק 26 - המשתנה המקרי הרציף- התפלגויות כלליות (שימוש באינטגרלים)

תוכן העניינים

1. כללי 91

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:

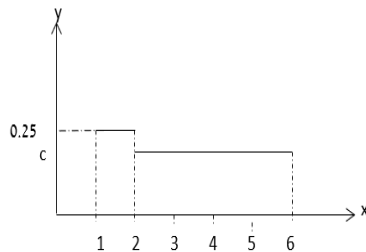
i. $P(x < 4)$

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:

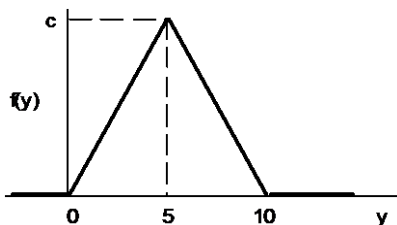
$$f(x) = \begin{cases} cx & 0 \leq x \leq b \\ 0 & \text{else} \end{cases}$$

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 .

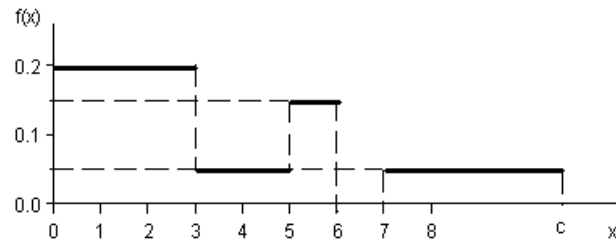
ג. חשבו את ההסתברויות הבאות:

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

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

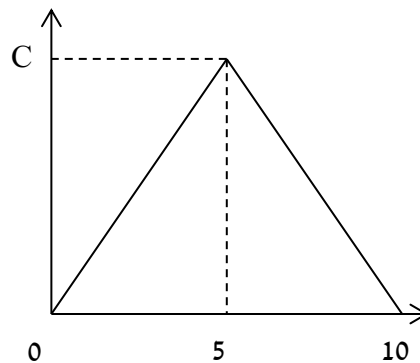
הסיקו מהו העשירון עליון: $y_{0.9}$.

(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:

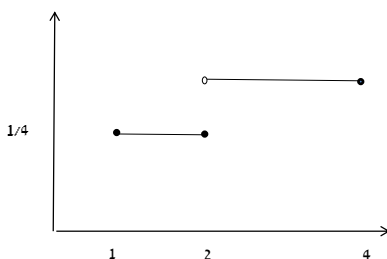
$$f(x) = 0.5 \cdot e^{2x}, \quad -\infty \leq x \leq \ln(c)$$

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

סטטיסטיקה

פרק 27 - התפלגויות רציפות מיוחדות- התפלגות מעריכית

תוכן העניינים

1. כללי 100

התפלגויות רציפות מיוחדות – התפלגות מעריכית:

רקע:

התפלגות זו היא התפלגות רציפה המאפיינת את הזמן עד להתרחשות מאורע מסוים. λ - הוא ממוצע מספר האירועים המתרחשים ביחידת זמן (אותו פרמטר מהתפלגות הפואסונית): $X \sim \exp(\lambda)$ כאשר $\lambda > 0$.

התפלגות זו צריכה להיות נתונה בתרגיל או שיאמר שמספר האירועים ביחידת זמן מתפלג פואסונית ואז הזמן עד להתרחשות המאורע הבא מתפלג מעריכית.

פונקציית הצפיפות של ההתפלגות:

$$f(x) = \lambda e^{-\lambda x} \quad \text{לכל } x \geq 0.$$

פונקציית ההתפלגות המצטברת: $F(t) = p(x \leq t) = 1 - e^{-\lambda t}$.

$$\text{התוחלת: } E(x) = \frac{1}{\lambda}$$

$$\text{השונות: } V(x) = \frac{1}{\lambda^2}$$

$$\text{להתפלגות זו יש תכונת חוסר הזיכרון: } P(X > a+b \mid X > a) = P(X > b).$$

דוגמה (פתרון בהקלטה):

אורך חיי סוללה מתפלג מעריכית עם תוחלת של 8 שעות.

א. מה ההסתברות שסוללה תחזיק מעמד פחות מ-9 שעות?

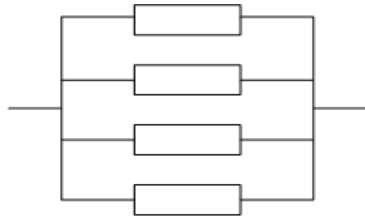
ב. מה סטיית התקן של אורך חיי הסוללה?

ג. אם סוללה כבר חייה מעל שעתיים, מה הסיכוי שהיא תחייה מעל 7 שעות בסך הכול?

שאלות:

- (1) הזמן שלוקח במערכת עד שתקלה מתרחשת מתפלג מעריכית עם תוחלת של 0.5 שעה.
- מה ההסתברות שהתקלה הבאה תתרחש תוך יותר מ-0.5 שעה?
 - מה ההסתברות שהתקלה הבאה תתרחש תוך פחות משעה?
 - מצא את הזמן החציוני להתרחשות תקלה במערכת.
- (2) הזמן שעובר בכביש מסוים עד להתרחשות תאונה מתפלג מעריכית עם תוחלת של 24 שעות.
- מהי סטיית התקן של הזמן עד להתרחשות תאונה?
 - מה ההסתברות שהתאונה הבאה תתרחש תוך פחות מיממה?
 - מהי ההסתברות שהתאונה הבאה תתרחש תוך לפחות יומיים?
- (3) משך הזמן X (בדקות) שסטודנטים עובדים רצוף על מחשב מתפלג מעריכית עם תוחלת של 30 דקות.
- מה הסיכוי שעבודת סטודנט על המחשב תארך פחות מרבע שעה?
 - מה הסיכוי שעבודת סטודנט על המחשב תארך בין רבע שעה לחצי שעה?
 - אם סטודנט עובד על המחשב כבר יותר מ-10 דקות, מה ההסתברות שמשך כל עבודתו יעלה על 30 דקות?
 - מהו הזמן שבסיכוי של 90% הסטודנט יעבוד פחות ממנו?
- (4) בממוצע מגיעים לחדר מיון 4 חולים בשעה בזרם פואסוני.
- שולה המזכירה הגיעה לחדר המיון. מה ההסתברות שזמן ההמתנה שלה לחולה הבא יהיה יותר מ-20 דקות?
 - אם שולה המתינה יותר מרבע שעה לחולה הבא. מה ההסתברות שתמתין בסך הכל יותר מחצי שעה?
 - מה ההסתברות שבין החולה הראשון לשני יש להמתין יותר מרבע שעה ובין החולה שני לשלישי יש להמתין פחות מרבע שעה?

(5) מערכת חשמלית כוללת 4 רכיבים אלקטרוניים זהים הפועלים במקביל כמתואר בשרטוט:



על מנת שהמערכת תפעל בצורה תקינה נדרש שלפחות אחד מהמרכיבים יהיה תקין. אורך החיים של כל רכיב מתפלג מעריכית עם ממוצע של 100 שעות.

א. מה ההסתברות שהמערכת תפעל בצורה תקינה במשך 100 שעות לפחות?

ב. מעוניינים להוסיף במקביל עוד רכיב למערכת. עלות הוספת רכיב היא K ₪.

כמו כן אם המערכת עבדה פחות מ-100 שעות נגרם הפסד של A ₪.

מה התנאי שבו יהיה כדאי להוסיף את הרכיב למערכת?

תשובות סופיות:

(1)	א. 0.368	ב. 0.865	ג. 0.347	
(2)	א. 24 שעות.	ב. 0.632	ג. 0.135	
(3)	א. 0.393	ב. 0.239	ג. 0.513	ד. 69.08
(4)	א. 0.264	ב. 0.368	ג. 0.233	
(5)	א. 0.8403	ב. $K < 0.0588A$		

סטטיסטיקה

פרק 28 - התפלגויות רציפות מיוחדות-התפלגות אחידה

תוכן העניינים

1. כללי 103

התפלגויות רציפות מיוחדות – התפלגות אחידה:

רקע:

זו התפלגות שפונקציית הצפיפות שלה קבועה בין a לבין b .

$$X \sim U(a, b)$$

פונקציית הצפיפות:

$$f(x) = \frac{1}{b-a}$$

$$a \leq x \leq b$$

פונקציית ההתפלגות המצטברת:

$$F(t) = \frac{t-a}{b-a}$$

התוחלת :

$$E(X) = \frac{a+b}{2}$$

השונות:

$$V(x) = \frac{(b-a)^2}{12}$$

דוגמה (פתרון בהקלטה):

X - משתנה מקרי רציף המתפלג באופן אחיד בין 20 ל-40.

מה הסיכוי ש- X קטן מ-25?

מה התוחלת והשונות של X ?

$$a = 20, b = 40$$

$$X \sim U(20, 40)$$

$$א. P(x < 25) = f(25) = \frac{25 - 20}{40 - 20} = 0.25$$

$$E(x) = \frac{20 + 40}{2} = 30$$

$$ב. V(x) = \frac{(40 - 20)^2}{12} = 33\frac{1}{3}$$

שאלות:

- (1) משך (בדקות) הפסקה בשיעור, X , מתפלג: $U(13,16)$.
- מהי התוחלת ומהי סטיית התקן של משך ההפסקה?
 - מהי ההסתברות שהפסקה תמשך יותר מ-15 דקות?
 - מהי ההסתברות שמשך ההפסקה יסטה מהתוחלת בפחות מדקה?
- (2) רכבת מגיעה לתחנה בשעות היום כל עשר דקות. אדם הגיע לתחנה בזמן אקראי.
- הסבר כיצד מתפלג זמן ההמתנה לרכבת?
 - אם זמן ההמתנה לרכבת ארך יותר מ-5 דקות, מהי ההסתברות שבסך הכל האדם ימתין לרכבת פחות מ-8 דקות?
 - מה תוחלת מספר הימים שיעברו עד הפעם הראשונה שהאדם ימתין לרכבת יותר מ-9 דקות?
- (3) מכונה אוטומטית ממלאת גביעי גלידה. משקל הגלידה לגביע מתפלג אחיד בין 100-110 גרם (המשקל הוא של גלידה ללא הגביע).
- מה ההסתברות שמשקל הגלידה בגביע יהיה מעל 108 גרם?
 - נתון שהגלידה בגביע עם משקל נמוך מ-107 גרם. מה ההסתברות שמשקל הגלידה יהיה מעל 105 גרם?
 - מה העשירון העליון של משקל הגלידה בגביע?
 - עלות גביע גלידה היא 0.5 שקל. כל גרם של גלידה עולה 0.22 אגורות. מהי התוחלת ומהי סטיית התקן של עלות הגביע ביחד עם הגלידה?

תשובות סופיות:

- (1) א. תוחלת: 14.5, שונות: 0.866. ב. $\frac{1}{3}$. ג. $\frac{2}{3}$.
- (2) א. $X \sim U(0,10)$. ב. 0.6. ג. 10.
- (3) א. 0.2. ב. $\frac{2}{7}$. ג. 109.
- ד. תוחלת: 73.1 אגורות, סטיית תקן: 0.635 אגורות.

סטטיסטיקה

פרק 29 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

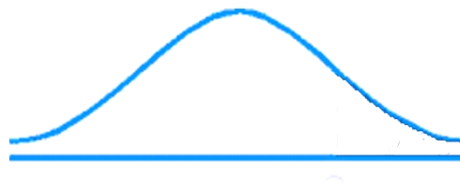
תוכן העניינים

1. כללי 106

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: $X \sim N(\mu, \sigma^2)$.

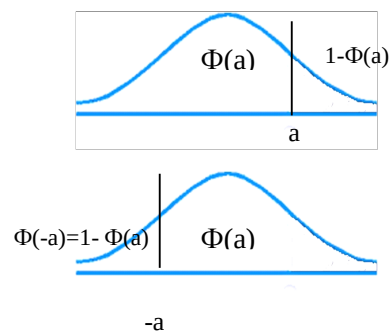
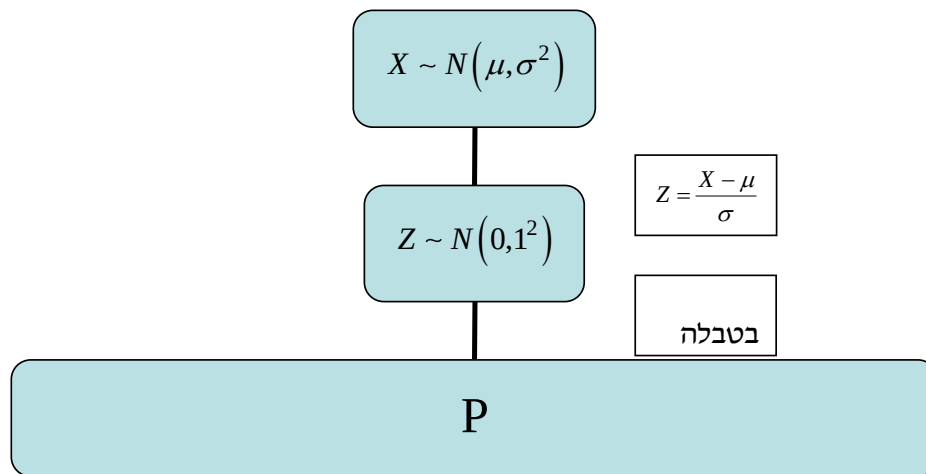
$$\text{נוסחת פונקציית הצפיפות: } f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

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 : $Z \sim N(0, 1^2)$.

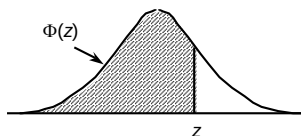
$$\text{תהליך התקנון מבוצע על ידי הנוסחה הבאה: } Z = \frac{X - \mu}{\sigma}$$

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.
- מה העשירון העליון של הציונים במבחן האינטליגנציה?
 - מה העשירון התחתון של ההתפלגות?
 - מהו הציון ש-20% מהנבחנים מקבלים מעליו?
 - מהו האחוזון ה-20?
 - 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.
- מה ממוצע נפח משקה בבקבוק?
 - 5% מהבקבוקים המיוצרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאיזה נפח שולחים בקבוק לבדיקה?
 - 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.
- מהו ממוצע אורך חיי מכשיר?
 - מהי סטית בתקן של אורך חיי מכשיר?
 - מה הסיכוי שמכשיר אקראי יחיה פחות מ-460 שעות?
 - מהו המאיון העליון של אורך חיי מכשיר?
 - 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.
- לאיזו התפלגות הממוצע הגבוה ביותר?
 - במה מבין המדדים הבאים התפלגות 1 ו-2 זהות?
 - בעשירון העליון.
 - בממוצע.
 - בשונות.
 - לאיזו התפלגות סטיית התקן הקטנה ביותר?
 - 1.
 - 2.
 - 3.
 - 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 סטית 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

פרק 30 - הסקה סטטיסטית - הקדמה

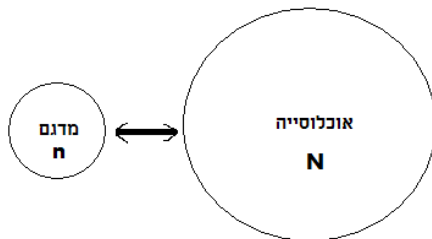
תוכן העניינים

1. כללי 114

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.

- א. מי האוכלוסייה?
- ב. מה המשתנה?
- ג. מה הפרמטרים?
- ד. מהו גודל המדגם?
- ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?
- ו. 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 הטלויזיה למשפחה ב"שוב "העוגן". נגדיר את X להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלויזיה 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

פרק 31 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

1. התפלגות ממוצע המדגם ומשפט הגבול המרכזי 117
2. התפלגות סכום תצפיות בלתי תלויות ומשפט הגבול המרכזי 125
3. התפלגות מספר ההצלחות במדגם - קירוב נורמלי להתפלגות הבינומית 128
4. התפלגות פרופורציית ההצלחות במדגם 133

התפלגות ממוצע המדגם ומשפט הגבול המרכזי:

רקע:

בפרק זה נדון בהתפלגות של ממוצע המדגם: $\bar{x} = \frac{\sum x_i}{n}$.

מכיוון שממדגם למדגם אנו יכולים לקבל ממוצע מדגם שונה, אזי ממוצע המדגם הוא משתנה מקרי ויש לו התפלגות.

גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים.

להלן רשימה של פרמטרים החשובים לפרק זה:

ממוצע האוכלוסייה נסמן ב- μ (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב- σ^2 .

סטיית תקן של אוכלוסייה: σ .

תכונות התפלגות:

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה: $E(\bar{x}) = \mu_{\bar{x}} = \mu$.

שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב- n .

תכונה זו נכונה רק במדגם מקרי: $V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$.

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

אם נוציא שורש לשונות נקבל סטיית תקן של ממוצע המדגם שנקראת גם

טעות תקן: $\sigma(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$.

דוגמה (פתרון בהקלטה):

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

א. מיהי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

דגימה מהתפלגות נורמאלית:

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

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

ממוצע המדגם גם יתפלג נורמאלית:

דוגמה (פתרון בהקלטה):

משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
מה ההסתברות שבמדגם של 4 תינוקות אקראיים בעת הולדתם המשקל הממוצע של התינוקות יהיה מתחת ל-3.5 ק"ג?

משפט הגבול המרכזי:

אם אוכלוסייה מתפלגת כלשהו עם ממוצע μ ושונות σ^2 אזי עבור מדגם מספיק

$$\bar{x} \rightsquigarrow N\left(\mu, \frac{\sigma^2}{n}\right) \quad (n \geq 30)$$

ממוצע המדגם מתפלג בקירוב נורמאלי:

דוגמה (פתרון בהקלטה):

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

שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מיהי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
 - מהו תוחלת ממוצע המדגם?
 - מהי טעות התקן?

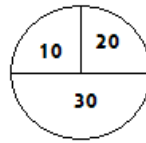
- (2) להלן התפלגות מספר מקלטי הטלויזיה למשפחה בישוב מסוים :

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

- בנו את פונקציית ההסתברות של X .
 - חשבו את התוחלת, השונות וסטיית התקן של X .
 - אם נדגום 4 משפחות מהישוב עם החזרה מה תהיה התוחלת, מהי השונות ומהי סטיית התקן של ממוצע המדגם?
- (3) אם נטיל קובייה פעמיים ונתבונן בממוצע התוצאות שיתקבלו, מה תהיה התוחלת ומה תהיה סטיית התקן של ממוצע זה?
- (4) משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
- מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?
- נתון כי ביום מסוים נולדו 4 תינוקות.
- מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
 - הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?

- (5) הגובה של המתגייסים לצה"ל מתפלג נורמאלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.
- א. מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?
 - ב. מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?
 - ג. מה ההסתברות שהגובה הממוצע שלהם יסטה מתוחלת הגבהים בפחות מ-5 ס"מ?
 - ד. מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?
- (6) הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים.
- לצורך הפתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.
- א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
 - ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?
 - ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
 - ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?
- (7) נפח היין בבקבוק מתפלג נורמאלית עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.
- א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ"ק?
 - ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ"ק?
 - ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ"ק?
 - ד. בקבוקי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?
- (8) משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4.
- א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
 - ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
 - ג. הסבר את ההבדל בתשובות של שני הסעיפים.

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



- אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.
- א. בנו את פונקציית ההסתברות של סכום הזכייה במשחק בודד.
- ב. מה התוחלת ומה השונות של סכום הזכייה?
- ג. אם האדם ישחק את המשחק 5 פעמים מה התוחלת ומה השונות של ממוצע סכום הזכייה בחמשת המשחקים?
- ד. אם האדם משחק את המשחק 50 פעם מה ההסתברות שבסה"כ יזכה ב-1050 ₪ ומעלה?

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

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

- 12) אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ.
- א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל-78 ס"מ?
- ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש-100 המוטות יספיקו למלאכה?
- ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. היעזרו במשפט הגבול המרכזי.

13) נתון משתנה מקרי בדיד בעל פונקציית ההסתברות הבאה:

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

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

14 נתון ש- $X \sim N(\mu, \sigma^2)$. דגמו 5 תצפיות מאותה התפלגות והתבוננו במוצע

המדגם $\bar{X}$. לכן: $P(\bar{X} > \mu)$ יהיה (בחרו בתשובה הנכונה):

- א. 0.
- ב. 0.5.
- ג. 1.
- ד. לא ניתן לדעת.

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

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

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

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

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

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

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

תשובות סופיות:

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

(2) א. להלן טבלה:

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

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

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

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

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

(5) א. 0 ב. 0 ג. 0.9544 ד. 178.205

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

(7) א. 0 ב. 0.1587 ג. 0.1587 ד. 0.5

(8) א. 0.5468 ב. 0.6826

(9) א. להלן טבלה:

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

א. התוחלת: 22.5, השונות: 68.75. ב. התוחלת: 22.5, השונות: 13.75.

ד. 0.8997

(10) 0.0475

(11) 0.1814

(12) א. 0.9772 ב. 0.0228 ג. 0.271

(13) 0.5

(14) ב'

(15) ד'

(16) ג'

(17) א. 2.429 ב. 0.25

(18) 0.6826

התפלגות סכום תצפיות בלתי תלויות ומשפט הגבול המרכזי:

רקע:

כעת נדון בסטטיסטי המבטא את סכום התצפיות במדגם: $T = \sum_{i=1}^n X_i$.
 כאשר כל התצפיות נדגמו באקראי מאותה אוכלוסייה, כלומר, היו: $X_1, \dots, X_n$ -
 משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה μ ושונותה σ^2 אזי:
 התוחלת והשונות של סכום התצפיות: $E(T) = n\mu$, $V(T) = n\sigma^2$.

דגימה מתוך התפלגות נורמלית:

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

משפט הגבול המרכזי:

אם X מתפלג כלשהו וידוע כי: $E(X) = \mu$, $V(X) = \sigma^2$, אזי עבור מדגם מספיק גדול (לפחות 30): $T \rightsquigarrow N(n\mu, n\sigma^2)$.

דוגמה (פתרון בהקלטה):

- בעיר מסוימת המשכורת הממוצעת של עובד הינה 8000 ₪. עם סטיית תקן של 2000 ₪. נדגמו 100 עובדים מהעיר שמפקידים את משכורותיהם לסניף בנק.
- מה התוחלת וסטיית התקן של סך המשכורות שיופקדו לסניף הבנק על ידי העובדים הללו?
 - מה ההסתברות שלסניף יופקד פחות מ-780 אלף ₪ ע"י אותם עובדים?

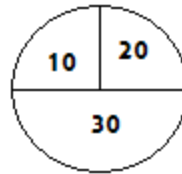
שאלות:

- (1) המשקל באוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת של 60 ק"ג וסטיית תקן של 10 ק"ג.
- א. מה הסיכוי שאדם אקראי מהאוכלוסייה ישקול מתחת ל-65 ק"ג?
- ב. מה הסיכוי שהמשקל הממוצע של 4 אנשים אקראיים יהיה מתחת ל-65 ק"ג?
- ג. מה הסיכוי שהמשקל הכולל של 4 אנשים אקראיים יהיה מתחת ל-240 ק"ג?
- (2) נפח יין בבקבוק מתפלג נורמאלית עם תוחלת של 750 מ"ל וסטיית תקן של 20 מ"ל. אדם קנה מארז של 4 בקבוקי יין.
- א. מהי התוחלת ומהי סטיית התקן של נפח היין במארז?
- ב. את היין שבמארז האדם מזג לכלי שקיבולתו 3.1 ליטר. מה ההסתברות שהיין יגלוש מהכלי?
- ג. אם לא היה נתון שנפח היין מתפלג נורמאלית. האם התשובה לסעיף א' הייתה משתנה? האם התשובה לסעיף ב' הייתה משתנה?
- (3) בספר כלשהו 500 עמודים. קצב הקריאה הממוצע הוא עמוד אחד ב-4 דקות עם סטיית תקן של 1 דקות.
- א. מה ההסתברות לסיים את הפרק הראשון (40 עמודים) תוך שעתיים וחצי?
- ב. מהו האחוזון ה-95 לזמן סיום קריאת הספר?
- (4) במגדל נבנו 40 יחידות דיור. כמו כן נבנו 135 מקומות חנייה לבניין. להלן פונקציית ההסתברות של מספר המכוניות ליחידת דיור:

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

- נניח שמספר המכוניות ליחידת דיור בלתי תליות זו בזו ועם אותה פונקציית הסתברות לכל יחידת דיור (אין צורך בתיקון רציפות).
- א. מהי ההסתברות שיהיה מקום בחניון המגדל לכל מכוניות הבניין?
- ב. בהינתן ויש מקום במגדל לכל המכוניות, מה הסיכוי שבפועל מספר המכוניות נמוך מ-130?

(5) בקזינו ישנה רולטה עליה מסומנים המספרים הבאים :



אדם מסובב את הרולטה וזוכה בסכום הרשום.

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

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

תשובות סופיות:

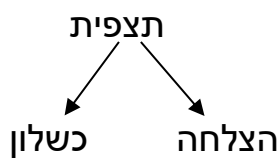
(1)	א. 0.6915	ב. 0.8413	ג. 0.5
(2)	א. תוחלת 3000 מ"ל, סטיית תקן 40 מ"ל.	ב. 0.0062	ג. סעיף א' - לא משתנה, סעיף ב' - לא פתיר, התבסס על התפלגות נורמלית.
(3)	א. 0.0571	ב. 2036.8	
(4)	א. 0.883	ב. 0.7949	
(5)	א. 0.8997	ב. תוחלת 1.111, שונות 0.1239	

התפלגות מספר ההצלחות במדגם – קירוב נורמלי להתפלגות הבינומית:

רקע:

תזכורת על התפלגות בינומית:

בפרק זה נדון בהתפלגות מספר ההצלחות במדגם אקראי (תצפיות בלתי תלויות זו בזו). את מספר ההצלחות במדגם נסמן ב- Y . מחלקים כל תצפית במדגם להצלחה או כישלון.



כעת מה שמשתנה מתצפית לתצפית הוא משתנה דיכוטומי (משתנה שיש לו שני ערכים). הסיכוי להצלחה יסומן עם הפרמטר p וכישלון יסומן ע"י הפרמטר $q = 1 - p$. מבצעים מדגם אקראי בגודל n : $Y \sim B(n, p)$.

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

תוחלת: $E(y) = np$.

שונות: $V(y) = npq$.

קירוב נורמלי עבור התפלגות בינומית:

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

$$1. \quad n \cdot p \geq 5$$

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

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

$$\text{אז: } Z_y = \frac{y - np}{\sqrt{npq}}$$

תיקון רציפות:

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

$$1. \quad p(Y = a) \cong p\left(a - \frac{1}{2} \leq Y \leq a + \frac{1}{2}\right)$$

$$2. \quad P(Y \leq a) \cong P(Y \leq a + 0.5)$$

$$3. \quad P(Y \geq a) \cong P(Y \geq a - 0.5)$$

הערות:

- התנאים למעבר מבינומי לנורמלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:

$$1. \quad n \cdot p \geq 5$$

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

- ישנם מרצים שנותנים את התנאי המחמיר הבא:

$$1. \quad n \cdot p \geq 10$$

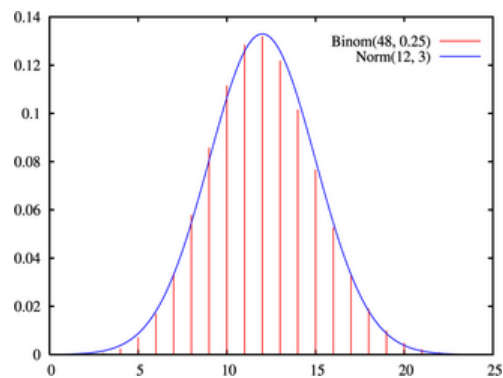
$$2. \quad n \cdot (1 - p) \geq 10$$

- וישנם מרצים שהתנאי שהם נותנים הוא: $(n \geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור מהתפלגות בינומית לנורמלית.
- הערה נוספת היא לגבי תיקון רציפות. ישנם מרצים שלא מחייבים לבצע תיקון רציפות שהמדגמים גדולים (בדרך כלל מעל 100 תצפיות) בפתרונות שאציג תמיד אבצע תיקון רציפות במעבר מבינומי לנורמלי כיוון שכך הפתרון יהיה יותר מדויק (בכל מקרה שהמדגמים גדולים העניין זניח).

דוגמה (הפתרון בהקלטה):

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

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



שאלות:

- (1) נתון ש-20% מאוכלוסייה מסוימת אקדמאית. נבחרו באקראי 10 אנשים באותה אוכלוסייה.
 - א. מה ההסתברות ששלושה מהם אקדמאים?
 - ב. מה ההסתברות שלכל היותר אחד מהם אקדמאי?
 - ג. מה התוחלת ומהי סטיית התקן של מספר האקדמאים במדגם?
- (2) במפעל 10% מהמוצרים פגומים. נלקחו 100 מוצרים באקראי מקו הייצור.
 - א. מה ההסתברות שנדגמו לפחות 6 מוצרים פגומים?
 - ב. מה ההסתברות שמספר המוצרים הפגומים יהיה לכל היותר 11 במדגם?
- (3) ציוני פסיכומטרי בקרב הנרשמים למוסד מסוים מתפלגים נורמאלית עם ממוצע 500 וסטיית תקן 100. למוסד מסוים הוחלט לקבל אך ורק סטודנטים שקיבלו מעל 600 בפסיכומטרי. 100 סטודנטים אקראיים נרשמו למוסד. מה ההסתברות שלפחות 20 יתקבלו?
- (4) מטילים מטבע 50 פעמים.
 - א. מה ההסתברות לקבל לכל היותר 30 עצים?
 - ב. מה ההסתברות לקבל 28 עצים לפי התפלגות הבינומית ולפי הקירוב הנורמאלי?
- (5) במטוס מקום ל-400 נוסעים. נרשמו לטיסה 430 אנשים (overbooking). מנתונים סטטיסטיים ידוע שהסיכוי שאדם שנרשם לטיסה אכן יגיע הוא 0.9.
 - א. מה ההסתברות שלא יהיו מקומות ישיבה לכל האנשים שהגיעו לטיסה?
 - ב. מה צריך להיות גודל המטוס כדי שבסיכוי שלפחות 95% המטוס יספיק לכמות הנרשמים?
- (6) מפעל לייצור ארטיקים טוען שהסיכוי שארטיק שהוא מייצר יהיה פגום הוא 0.01. מוכר הזמין 1000 ארטיקים מהמפעל. מה ההסתברות שהמוכר יקבל לפחות 980 ארטיקים תקינים אם טענת המפעל מוצדקת?
- (7) מהמר מטיל קובייה הוגנת 100 פעמים. בכל הטלה, אם מתקבל תוצאה זוגית בקובייה המהמר זוכה בשקל. אחרת, המהמר משלם שקל. המהמר הטיל את הקובייה 100 פעמים מה הסיכוי שהרווח של המהמר יהיה לכל היותר 10?

תשובות סופיות:

- (1) א. 0.201 ב. 0.3758 ג. התוחלת: 2, סטיית התקן: 1.2649.
- (2) א. 0.9332 ב. 0.6915
- (3) 0.1611
- (4) א. 0.9406 ב. בינומית - 0.0788, קירוב לנורמלית - 0.0778.
- (5) א. 0.015 ב. 0.398
- (6) 0.9996
- (7) 0.8643

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.

סטטיסטיקה

פרק 32 - מושגי יסוד באמידה

תוכן העניינים

1. כללי 137

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: 168, 184, 192, 171, 180, 177, 187, 168, 177 ו-175.
- 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: $\sum_{i=1}^{20} X_i = 162$, $\sum_{i=1}^{20} X_i^2 = 1502.2$.
- 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

פרק 33 - רווח סמך לתוחלת (ממוצע)

תוכן העניינים

- 1. רווח סמך כששונות האוכלוסיה ידועה 142
- 2. קביעת גודל מדגם 147
- 3. רווח סמך כששונות האוכלוסיה לא ידועה 149

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?

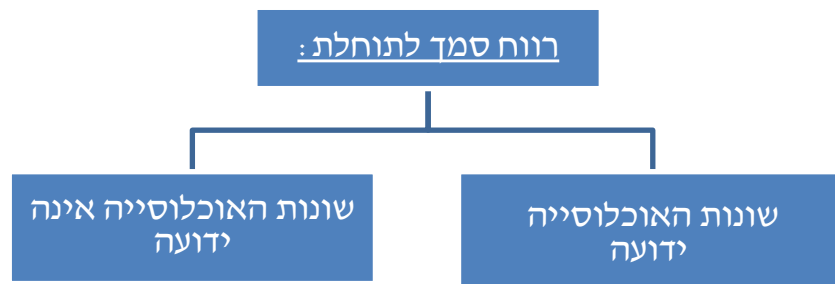
תשובות סופיות:

- (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:

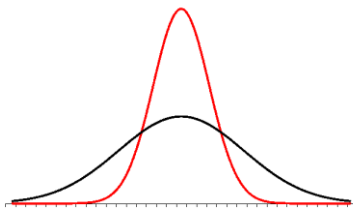


בפרק זה נעסוק במקרה ששונות האוכלוסייה (σ^2) אינה ידועה לנו.

מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול.

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



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

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.
 - א. 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: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X_i^2 = 643.19$.
 - א. 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?
- א. סטטיסטיקאי א.
ב. סטטיסטיקאי ב.
ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.
ד. 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

פרק 34 - רווח סמך לפרופורציה

תוכן העניינים

152	1. רווח הסמך לפרופורציה
155	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 ($\hat{p}$).

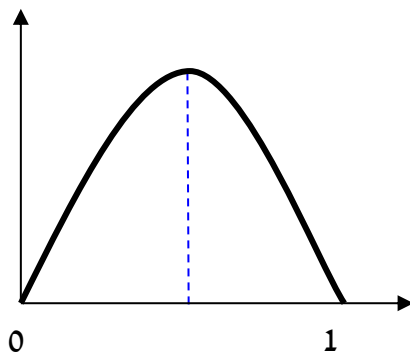
$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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 חישובתם 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 חישובת 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

פרק 35 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

1. כששונניות האוכלוסיה לא ידועות ובהנחת שוויון שונניות.....158

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

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

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

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

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
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

פרק 36 - רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים

תוכן העניינים

1. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 160

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.

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

$$\text{כאשר דרגות החופש: } df = n - 1$$

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

פרק 37 - שאלות מסכמות על רווחי סמך

תוכן העניינים

1. שאלות מסכמות על רווחי סמך 162

שאלות מסכמות על רווחי סמך:

שאלות:

- (2) 200 אנשים נשאלו כמה פעמים ביום הם שותים כוס קפה. להלן התפלגות התשובות:

מספר פעמים	0	1	2	3	4	5
מספר אנשים	86	34	28	22	20	10

- א. תנו רווח סמך לממוצע מספר כוסות הקפה שאנשים נוהגים לשתות ביום. $\alpha = 0.05$.
- ב. אדם השותה לפחות 4 כוסות קפה ביום נקרא "מכור לקפה". בנו רווח סמך לאחוז "המכורים לקפה". $\alpha = 0.1$.
- (3) חוקר בנה רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת בשנה. רווח הסמך שהתקבל הוא: $81 < p < 91$. רווח הסמך הנ"ל התבסס על מדגם של 500 איש.
- א. כמה אנשים במדגם טענו שכלל לא התקררו השנה?
- ב. באיזו רמת סמך נבנה רווח הסמך?
- ג. בנו רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת השנה ברמת סמך של 96% על סמך תוצאות המדגם.
- (4) ציוני IQ בארה"ב מתפלגים נורמאלית עם תוחלת 100. במדגם של 20 ישראלים שנבחנו במבחן ה-IQ התקבלו התוצאות הבאות:
- $$\sum_{i=1}^{20} x_i = 2040, \quad \sum_{i=1}^{20} x_i^2 = 210740$$
- א. אמדו ברמת ביטחון של 90% את ממוצע ציוני בחינת ה-IQ בישראל – מהי ההנחה הדרושה לפתרון?
- ב. על סמך רווח הסמך של סעיף א' האם תקבלו את הטענה שבישראל ממוצע הציונים שונה מארה"ב?
- ג. מה היה קורה לרווח הסמך אם היינו מגדילים את רמת הסמך שלו?

(5) להלן תוצאות מדגם שבדק עבור כל משפחה האם יש לה בבית מכשיר טאבלט:

אזור מגורים	גוש דן	שאר הארץ
גודל המדגם	200	240
מספר משפחות בעלי טאבלט	160	168

- א. בנו רווח סמך להבדל בין אחוז המשפחות עם טאבלט בגוש דן ואחוז המשפחות בעלי טאבלט בשאר חלקי הארץ. ברמת סמך של 98%.
- ב. בנו רווח סמך לפרופורציות משפחות בעלות טאבלט בכלל הארץ ברמת סמך של 95%.

(7) בנק מתלבט האם לפתוח סניף באזור A או באזור B. לצורך פתרון נניח שסטיית התקן של המשכורת באזור A היא 1200 ובאזור B 1500. הבנק דגם 50 אנשים מאזור A, המשכורת הממוצעת שהתקבלה במדגם היא 6,800 ₪. כמו כן, נדגמו 40 אנשים מאזור B, המשכורת הממוצעת שהתקבלה במדגם היא 6,600 ₪.

- א. בנו רווח סמך ברמת סמך של 95% להפרש הממוצעים של המשכורות בשני האזורים. האם על סמך רווח הסמך ניתן להמליץ לבנק היכן לפתוח את הסניף. אם כן, היכן?
- ב. בנו רווח סמך לתוחלת המשכורת באזור A ברמת סמך של 95%.

(8) להלן מדגם של שכר הדירה ב-₪ של 5 דירות שלושה חדרים בשכונת בבלי בתל אביב:

7500	6500	7000	7500	8000	שנת 2012
7700	6800	7800	8200	8000	שנת 2013

בנו רווח סמך ברמת סמך של 95% לתוחלת עליית שכר הדירה משנת 2012 לשנת 2013 בשכונת בבלי. ניתן להניח ששכר הדירה בשכונה מתפלג נורמלית.

תשובות סופיות:

- (2) א. $1.21 \leq \mu \leq 1.65$ ב. $10.85\% \leq p \leq 19.15\%$
- (3) א. 70 ב. 0.9988 ג. $83\% < p < 89\%$
- (4) א. $97.4 \leq \mu \leq 106.6$ ב. לא ג. יגדל.
- (5) א. $0.5\% \leq p_1 - p_2 \leq 19.5\%$ ב. $0.704 \leq p \leq 0.786$
- (7) א. $-372 \leq \mu_A - \mu_B \leq 772$, לא ב. $6467 \leq \mu_A \leq 7133$
- (8) $-21 \leq \mu_D \leq 821$

סטטיסטיקה

פרק 38 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

165	1. הקדמה
169	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. $H_0: \mu = 500$
 ב. לא דחינו את H_0 .
 ג. טעות מסוג שני.
- (2) א. לא ניתן לדעת.
 ב. טעות מסוג ראשון.
- (3) א. משפחות כיום.
 ב. מס' הילדים.
 ג. תוחלת מספר הילדים למשפחה כיום.
 ד. $H_0: \mu = 2.3$
 $H_1: \mu < 2.3$
 ה. לא לדחות את H_0 . ו. טעות מסוג שני.

סטטיסטיקה

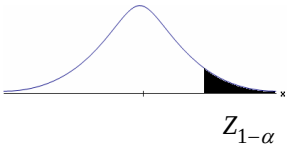
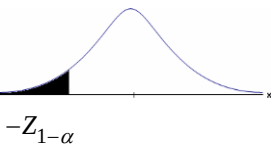
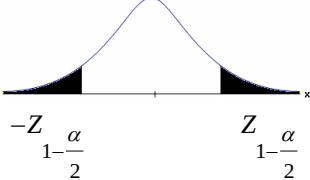
פרק 39 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

1. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה ידועה 171
2. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה ידועה) 175
3. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה לא ידועה 180
4. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה לא ידועה) 184
5. הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע) 187

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 שזוּבְלו 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.

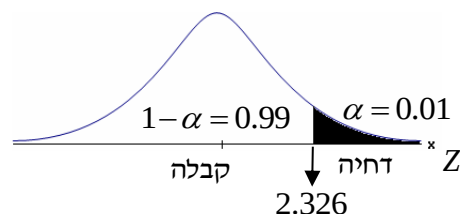
השערות:
 $H_0: \mu = 10$
 $H_1: \mu > 10$

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.

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 נורמאלית 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.

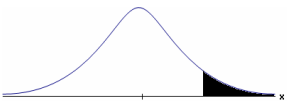
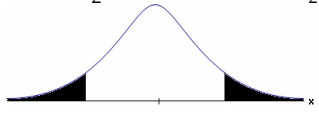
- (6) בבדיקת השערות התקבל שה- $p\text{-value} = 0.02$.
 מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%?
 בחרו בתשובה הנכונה.
- א. יקבל את השערת האפס בכל מקרה.
 - ב. ידחה את השערת האפס מקרה.
 - ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
 - ד. לא ניתן לדעת כי אין מספיק נתונים.
- (7) מובהקות התוצאה (PV) 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'.

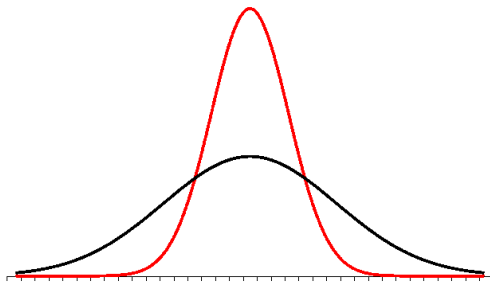
בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה לא ידועה:

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.

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

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

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

$$d.f = n-1$$

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 .

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

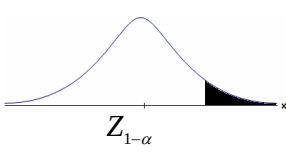
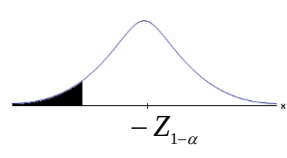
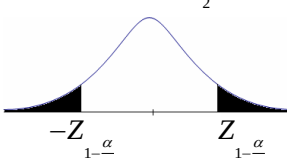
פרק 40 - בדיקת השערות על פרופורציה

תוכן העניינים

189	1. התהליך.....
192	2. מובהקות התוצאה - אלפא מינימלית.....

התהליך:

רקע:

השערת האפס: השערת 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

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

סטטיסטי המבחן:

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'.

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

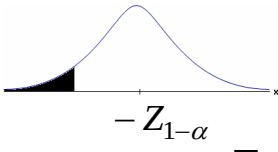
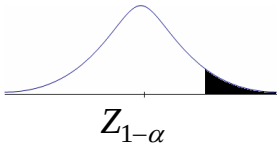
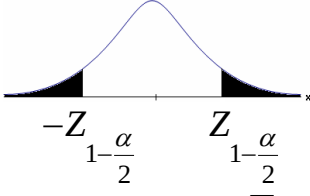
פרק 41 - בדיקת השערות על הפרש פרופורציות

תוכן העניינים

1. כללי 196

בדיקת השערות על הפרש פרופורציות

רקע

השערת האפס : השערה אלטרנטיבית:	1. מדגמים בלתי תלויים	2. מדגמים גדולים	
$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 > 0$	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 < 0$	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 \neq 0$	
תנאים:	1. מדגמים בלתי תלויים	2. מדגמים גדולים	כלל ההכרעה: אזור הדחייה של:
	$Z_{\hat{p}_1 - \hat{p}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ - דוחים את H_0	$Z_{\hat{p}_1 - \hat{p}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ - דוחים את H_0	
	או $Z_{\hat{p}_1 - \hat{p}_2} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\hat{p}_1 - \hat{p}_2} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ - דוחים את H_0		

$$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} \quad \text{סטטיסטי המבחן:}$$

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} \quad \text{כאשר הפרופורציה המשוקללת:}$$

חלופה אחרת לכלל הכרעה:

כלל ההכרעה: אזור הדחייה של H_0	
$\hat{p}_1 - \hat{p}_2 < 0 - Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$	$\hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$
	$\hat{p}_1 - \hat{p}_2 < 0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}} \text{ או } \hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$
$\hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$	

התפלגות של $\hat{p}_1 - \hat{p}_2$: $\hat{p}_1 - \hat{p}_2 \sim N(p_1 - p_2, \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2})$

$$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

תקנון:

$$Z_{\hat{p}_1 - \hat{p}_2} \Big|_{H_0} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

דוגמה (פתרון בהקלטה):

נדגמו 80 סטודנטים שנבחנו במיקרו-כלכלה. מתוכם 60 עברו את הבחינה.
 נדגמו 100 סטודנטים שנבחנו בסטטיסטיקה א'. מתוכם 82 עברו את הבחינה.
 האם שיעור העוברים את הבחינה בסטטיסטיקה גבוה מאשר מהבחינה במיקרו כלכלה? בדקו ברמת מבוהקות של 10%.

שאלות

- (1) במדגם של 200 גברים, 8% היו מובטלים. במדגם של 180 נשים, 10% מהן היו מובטלות.
האם קיים הבדל מובהק בין פרופורציית המובטלים לפרופורציית המובטלות?
בדקו ברמת מובהקות של 5%.
- (2) אחוז בעלי רישיון נהיגה בקרב האוכלוסייה הבוגרת הינו 60%. במדגם של 300 בוגרים מתל אביב 204 היו בעלי רישיון נהיגה.
במדגם של 220 בוגרים מירושלים 100 היו בעלי רישיון נהיגה.
א. ברמת מובהקות של 5% האם תקבלו את הטענה שאחוז בעלי הרישיון בתל אביב גבוה מהאחוז הארצי?
ב. ברמת מובהקות של 10% האם תקבלו את הטענה שאחוז בעלי הרישיון נהיגה בתל אביב גבוה מאחוז בעלי רישיון הנהיגה בירושלים?
- (3) נדגמו 500 בוגרים מתוכם 200 גברים והיתר נשים. במדגם התקבל: מתוך הגברים ל-48% תעודת בגרות. מתוך הנשים ל-58% תעודת בגרות. מטרת המחקר היא לבדוק האם שיעור הזכאיות לבגרות גבוה משיעור הזכאים.
א. מהי מובהקות התוצאה?
ב. מה תהיה המסקנה ברמת מובהקות של 8%?
- (4) במדגם שנערך על 100 פרות מחוות בדרום הארץ התקבל כי 20 פרות נושאות וירוס מסוים. במדגם שנערך על 200 פרות מחוות בצפון הארץ התקבל כי 10 מתוכן נושאות וירוס גם כן.
א. בנו מבחן ברמת מובהקות של 5% לבדיקת הטענה כי הווירוס תקף את פרות הדרום באופן משמעותי יותר מאשר את הפרות בצפון הארץ.
ב. מהי המסקנה לבדיקת הטענה של סעיף א ומהי הטעות האפשרית במסקנה?
ג. מהי עוצמת המבחן אם שיעור הפרות בדרום עם הווירוס גבוה ב-10% משיעור הפרות בצפון עם הווירוס?
ד. כיצד העוצמה תשתנה אם נגדיל את רמת המובהקות?

תשובות סופיות

(1)	לא נדחה את H_0 .
(2)	א. נדחה H_0 . ב. נדחה H_0 .
(3)	א. 0.0139 ב. נדחה H_0 .
(4)	א. ראה סרטון. ב. נדחה H_0 . ג. 0.8238 ד. תגדל.

סטטיסטיקה

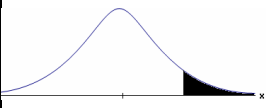
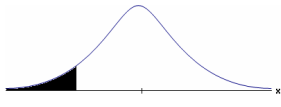
פרק 42 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסיה לא ידועות ומניחים שהן שוות..... 200

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.

הקבוצה	60W	100W
$\bar{X}$	1007	956
S	80	72
n	13	15

- 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 .

סטטיסטיקה

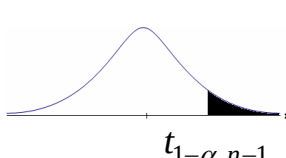
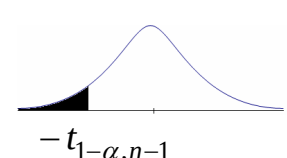
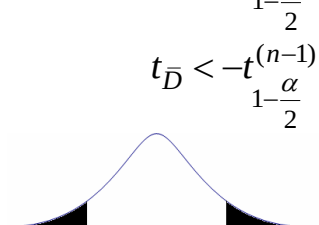
פרק 43 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

תוכן העניינים

1. בדיקת השערות למדגמים מזווגים..... 204

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

בדיקת השערות למדגמים מזווגים – רקע

השערת האפס: השערה אלטרנטיבית:	$H_0: \mu_D = C$ $H_1: \mu_D > C$	$H_0: \mu_D = C$ $H_1: \mu_D < C$	$H_0: \mu_D = C$ $H_1: \mu_D \neq C$
תנאים:	1. σ_D אינה ידועה 2. $D \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  - דוחים את H_0	$t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$  - דוחים את H_0	$t_{\bar{D}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{D}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$  - דוחים את H_0
חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:	$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ ו $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}, \quad t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

סטטיסטי המבחן:

דוגמה (פתרון בהקלטה):

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה 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'.

סטטיסטיקה

פרק 44 - הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות

תוכן העניינים

1. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות..... 208

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 :

מרצה X	82	90	68
מרצה Y	68	81	64

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 :
 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 .

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

פרק 45 - שאלות מסכמות בבדיקת השערות

תוכן העניינים

1. שאלות רב ברירה (אמריקאיות) 211

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 $\alpha = 0.05$?

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס לא הייתה נדחית.
- ג. ההשערה המחקרית הייתה נדחית.
- ד. 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:

- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. 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 :

- א. השערת האפס נכונה.
- ב. השערת האפס נדחתה.
- ג. השערת האפס לא נדחתה.
- ד. 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 :

- 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.

(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?

- א. טעות מסוג ראשון.
- ב. טעות מסוג שני.
- ג. טעות מסוג שלישי.
- ד. אין טעות במסקנתו.

(28) בבדיקת השערות התקבל שה- $p\text{-value} = 0.02$. מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר 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.

(29) מובהקות התוצאה (PV) היא גם :

- א. רמת המובהקות המינימאלית לדחות השערת האפס.
- ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
- ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.

(30) בבדיקת השערות מסוימת התקבל $p \text{ value} = 0.0254$, לכן :

- 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 .

(31) רמת המובהקות במחקר הייתה 2% לכן.

- א. בסיכוי של 2% נדחה את השערת האפס.
- ב. בסיכוי של 2% לא נדחה את השערת האפס.
- ג. בסיכוי של 2% השערת האפס לא נכונה.
- ד. 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:

$$H_0: \mu_1 - \mu_2 = 0; H_1: \mu_1 - \mu_2 \neq 0$$

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 **במציאות** 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

פרק 46 - ניתוח שונות חד כיוונית

תוכן העניינים

1. כללי 224

ניתוח שונות חד כיוונית

רקע תיאורטי

ניתוח שונות (חד כיוונית) הוא מבחן להשוואת תוחלות $(\mu_1, \dots, \mu_k)$ של k אוכלוסיות שונות. לכן, בניתוח שונות, השערות המחקר הן:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k \quad (\text{התוחלות של כל האוכלוסיות שוות})$$

$$H_1: \quad \text{אחרת} \quad (\text{לפחות שתיים מהתוחלות שונות})$$

ההנחות הדרושות לביצוע התהליך:

(1) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.

(2) כל האוכלוסיות הן עם אותה שונות σ^2 .

(3) המדגמים בלתי תלויים זה בזה.

ישנו משתנה המבדיל בין הקבוצות השונות, הוא המשתנה הבלתי תלוי הנקרא גורם (factor). משתנה זה הוא קטגוריאלי עם k רמות (levels). כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה: נסמן ב- n_i את גודל המדגם בקבוצה i .

$$n = \sum_{i=1}^k n_i \quad \text{- מספר התצפיות סך הכול (בכל המדגמים).}$$

$$\bar{X}_1, \dots, \bar{X}_k \quad \text{- ממוצע המדגם הראשון, ..., ממוצע המדגם ה-} k \text{-י.}$$

$$\bar{X} \quad \text{- ממוצע כללי (של כל המדגמים).}$$

$$SS_B = \sum_{i=1}^k n_i [\bar{X}_i - \bar{X}]^2 \quad \text{סכום ריבועים בין הקבוצות:}$$

$$SS_W = \sum_{i=1}^k n_i [n_i - 1] \cdot \hat{S}_i^2 \quad \text{סכום ריבועים בתוך הקבוצות:}$$

$$SS_T = \sum_{i=1}^k \sum_{j=1}^{n_j} [X_{ij} - \bar{X}]^2 \quad \text{סכום ריבועים כללי:}$$

$$SST = SSB + SSW$$

יש למלא את טבלת ניתוח השונות הבאה :

מקור השונות	סכום הריבועים SS	דרגות חופש df	ממוצע הריבועים MS	F
B - בין הקבוצות	SSB	$k - 1$	$\frac{SSB}{k - 1}$	$\frac{MSB}{MSW}$
W - בתוך הקבוצות	SSW	$n - k$	$\frac{SSW}{n - k}$	
T - סה"כ	SST	$n - 1$		

$$F = \frac{\frac{SSB}{k-1}}{\frac{SSW}{n-k}} \sim F(k-1, n-k)$$

אזור דחיית $H_0 : 1-\alpha : F_{(k-1, n-k)}$

שאלות

- (1) מחקר מעוניין להשוות בין שלוש תרופות לשיכוך כאבים במטרה לבדוק האם קיים הבדל בין התרופות מבחינת הזמן בדקות שלוקח עד שהתרופה משפיעה. לצורך הבדיקה נלקחו 15 אנשים שסובלים מכאבי ראש. אנשים אלה חולקו באקראי לשלוש קבוצה: קבוצה 1 קיבלה "אקמול" קבוצה 2 קיבלה "אופטלגין" קבוצה 3 קיבלה "נורופן". כל אדם במחקר מסר את מספר הדקות עד שהתרופה השפיעה עליו.
- א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר?
- ב. מהו המבחן הסטטיסטי המתאים כאן? רשמו את ההשערות.
- ג. מה הן ההנחות הדרושות כדי לבצע את המבחן הסטטיסטי שהצעת בסעיף הקודם?

- (2) בעיר מסוימת שלושה בתי ספר תיכון. ראש העיר התעניין לבדוק האם קיים הבדל בהצלחה של בתי הספר במקצוע מתמטיקה. לצורך כך הוא דגם מספר תלמידים שנבחנו במבחן הבגרות במתמטיקה ברמה של 3 יחידות בעירו ובדק עבור כל תלמיד מה ציון הבגרות שלו במתמטיקה.

"הס"	"רביץ"	"המתמיד"
85	98	78
83	62	65
74	55	70
85	80	90
75		56

- א. מהו המבחן הסטטיסטי המתאים?
- ב. מהו גודל המדגם? מהו המשתנה הבלתי תלוי (factor) כמה רמות יש לו?
- ג. חשבו את הממוצע ואת סטיית התקן של הציונים בכל אחד מהמדגמים.
- ד. מלאו את טבלת ANOVA.
- ה. רשמו את כלל ההכרעה למבחן שהוצע בסעיף א ברמת מובהקות של 5%.
- ו. האם קיים הבדל בין בתי הספר בעיר מבחינת רמת הצלחת התלמידים במקצוע המתמטיקה? ענה על סמך הסעיפים הקודמים.

- (3) מעוניינים לבדוק האם יש הבדל בהשפעה של שיטות טפול שונות על לחץ הדם הסיסטולי (SBP) באוכלוסייה של קשישים. נבדקו 4 שיטות שונות. בטבלה המצורפת מרוכזים ממצאי המחקר.

השיטה	A	B	C	D
גודל המדגם	12	14	8	12
הממוצע	178	172	180	182
סטיית התקן	4	8	5	3

- א. רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.
- ב. מה מסקנת המחקר ברמת מובהקות של 5%?
- ג. האם יש צורך לבצע השוואות מרובות?

(4) שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות.

האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות?

בדקו ברמת מובהקות של 5%.

האופה	ניר	מוזס	שלום
סכום הזמנים	206	212	182
סכום ריבועי הזמנים	10644	11250	8982

(5) להלן טבלת ניתוח שונות חד כיוונית. במחקר בחנו 4 סוגי סוללות. רצו לבדוק האם לסוג הסוללה השפעה על תוחלת אורך החיים שלה. הפעילו את כל הסוללות על אותו מכשיר ובדקו את אורך החיים של כל סוללה בשעות. מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.317	3	3.439	1.361	.279
Within Groups	60.648	24	2.527		
Total	70.964	27			

(6) להלן טבלת ANOVA בטבלה הושמטו חלקים. השלימו את החלקים בטבלה שהושמטו ומסומנים באותיות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450	ב	ג	ה	.000
Within Groups	א	17	ד		
Total	522.950	19			

7) חברת תרופות לקחה 15 אנשים ברמת בריאות דומה. החברה חילקה את האנשים ל שלוש קבוצות שוות בגודלן. לכל קבוצה ניתנה אותה תרופה במינון שונה (dosage). המינונים שניתנו הם: 10 מ"ג, 20 מ"ג ו-30 מ"ג. לאחר שעה מזמן לקיחת התרופה נבדק קצב פעימות הלב של כל אדם (pulse). הנתונים הוזנו לתוכנה סטטיסטית והתקבלו התוצאות הבאות:

ANOVA						pulse			
pulse						Tukey HSD ^a			
	Sum of Squares	df	Mean Square	F	Sig.	dosage	N	Subset for alpha = 0.05	
								1	2
Between Groups	414.400	2	207.200	19.733	.000	30.00	5	71.0000	
Within Groups	126.000	12	10.500			20.00	5		80.2000
Total	540.400	14				10.00	5		83.4000
						Sig.		1.000	.299
						Means for groups in homogeneous subsets are displayed.			
						a. Uses Harmonic Mean Sample Size = 5.000.			

Post Hoc Tests

Multiple Comparisons

		pulse Tukey HSD				
(I) dosage	(J) dosage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
10.00	20.00	3.20000	2.04939	.299	-2.2675	8.6675
	30.00	12.40000*	2.04939	.000	6.9325	17.8675
20.00	10.00	-3.20000	2.04939	.299	-8.6675	2.2675
	30.00	9.20000*	2.04939	.002	3.7325	14.6675
30.00	10.00	-12.40000*	2.04939	.000	-17.8675	-6.9325
	20.00	-9.20000*	2.04939	.002	-14.6675	-3.7325

*. The mean difference is significant at the 0.05 level.

- בדקו ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.
- הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.
- האם יש צורך במחקר בהשוואת מרובות. נמקו!
- לטבלת ANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "טוקי". ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

- (8) בעיר מסוימת רצו לבדוק האם קיים הבדל ברמה של התלמידים בין בתי הספר השונים בעיר. ביצעו מדגם מכל בית ספר ונתנו מבחן זהה לכל הנדגמים. לאחר מכן ריכזו את הנתונים בתוכנה סטטיסטית והפעילו ניתוח שונות. מצורפים הפלטים שהתקבלו. ענו על הסעיפים הבאים :
- כמה בתי ספר יש בעיר?
 - כמה תלמידים השתתפו בסך הכול במחקר?
 - האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%
 - בביטחון של 95% אילו בתי ספר שונים זה מזה ברמת התלמידים? נמקו והסבירו.

Oneway

ANOVA

grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7799.600	4	1949.900	13.586	.000
Within Groups	2870.400	20	143.520		
Total	10670.000	24			

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

(I) school	(J) school	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	5.40000	7.57681	.971	-20.2543	31.0543
	3.00	36.80000*	7.57681	.003	11.1457	62.4543
	4.00	36.40000*	7.57681	.003	10.7457	62.0543
	5.00	-2.60000	7.57681	.998	-28.2543	23.0543
2.00	1.00	-5.40000	7.57681	.971	-31.0543	20.2543
	3.00	31.40000*	7.57681	.011	5.7457	57.0543
	4.00	31.00000*	7.57681	.013	5.3457	56.6543
	5.00	-8.00000	7.57681	.888	-33.6543	17.6543
3.00	1.00	-36.80000*	7.57681	.003	-62.4543	-11.1457
	2.00	-31.40000*	7.57681	.011	-57.0543	-5.7457
	4.00	-.40000	7.57681	1.000	-26.0543	25.2543
	5.00	-39.40000*	7.57681	.001	-65.0543	-13.7457
4.00	1.00	-36.40000*	7.57681	.003	-62.0543	-10.7457
	2.00	-31.00000*	7.57681	.013	-56.6543	-5.3457
	3.00	.40000	7.57681	1.000	-25.2543	26.0543
	5.00	-39.00000*	7.57681	.001	-64.6543	-13.3457
5.00	1.00	2.60000	7.57681	.998	-23.0543	28.2543
	2.00	8.00000	7.57681	.888	-17.6543	33.6543
	3.00	39.40000*	7.57681	.001	13.7457	65.0543
	4.00	39.00000*	7.57681	.001	13.3457	64.6543

*. The mean difference is significant at the 0.05 level.

Homogeneous Subsets

grade

Scheffe^a

school	N	Subset for alpha = 0.05	
		1	2
3.00	5	45.0000	
4.00	5	45.4000	
2.00	5		76.4000
1.00	5		81.8000
5.00	5		84.4000
Sig.		1.000	.888

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

תשובות סופיות

(1) א. משתנה בלתי תלוי : סוג התרופה. ב. ניתוח שונות חד כיוונית

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : otherwise$$

משתנה תלוי : הזמן עד להשפעת התרופה בדקות.

ג. 1. מדגמים בלתי תלויים.

2. שוויון שונויות.

3. משתנים מתפלגים נורמלית.

(2) א. המבחן לניתוח שונות חד כיוונית.

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. משתנים מתפלגים נורמלית.

3. שוויון שונויות.

ב. גודל המדגם : 14. משתנה ב"ת : בית הספר, בעל 3 רמות.

$$g. \bar{X} = 71.8, \hat{S} = 12.93, \bar{X} = 73.75, \hat{S} = 19.29, \bar{X} = 80.4, \hat{S} = 5.46$$

ד. להלן טבלה :

מקור השונות	SS	df	MS	F
B	200.6	2	100.3	
W	1904.75	11	173.2	
סה"כ	2105.35	13		0.58

$$h. F > 3.98$$

ו. נקבל את H_0 .

(3) א. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
 $H_1 : otherwise$
 ב. נדחה את H_0 . ג. כן.

הנחות :

1. מדגמים בלתי תלויים.

2. שוויון שונויות.

3. משתנים מתפלגים נורמלית.

(4) נקבל את H_0 : נכריע שאין הבדל מובהק בין האופים מבחינת תוחלת זמן הכנה.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad (5)$$

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. שוויון שונות.

3. משתנים מתפלגים נורמלית.

נקבל את H_0 : לסוג סוללה אין השפעה של תוחלת החיים ברמת ביטחון של 10%.

(6) להלן טבלה :

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450	2 ב	178.725 ג	18.36 ה	.000
Within Groups	165.5 א	17	9.735 ד		
Total	522.950	19			

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad (7)$$

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. משתנים מתפלגים נורמלית.

3. שוויון שונות.

נדחה את H_0 : ברמת ביטחון של 5% קיים הבדל במינונים השונים מבחינת תוחלת הדופק.

$$\mu_{20} = \mu_{10} > \mu_{30} \quad \text{ד.} \quad \text{ג. כן.} \quad \text{ב. ראה וידאו.} \quad \text{א.} \quad (7)$$

$$25 \quad \text{ב.} \quad 5 \quad \text{א.} \quad (8)$$

ג. נדחה את H_0 : יש לפחות שני בתי ספר בעיר עם תוחלת רמת ציונים שונה.

$$(\mu_3 = \mu_4) < (\mu_1 = \mu_2 = \mu_3) \quad \text{ד.}$$

סטטיסטיקה

פרק 47 - מבחני חי בריבוע

תוכן העניינים

1. מבחן טיב התאמה 233
2. מבחן טיב התאמה והקשר שלו לבדיקת השערות על פרופורציה אחת 238
3. מבחן לאי תלות 240
4. קשר בין מבחן אי תלות לבדיקת השערות להפרש פרופורציות 245

מבחני חי בריבוע

מבחן טיב התאמה – רקע

מבחן זה בא לבדוק האם אוכלוסייה מסוימת מתפלגת לפי התפלגות נתונה. המשתנה הנחקר מחולק למספר קטגוריות ויש לבדוק האם תוצאות המדגם תואמות להתפלגות הנתונה.

מבנה המבחן:

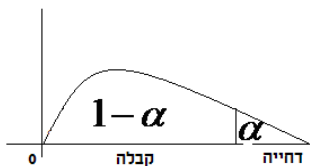
השערות:

המשתנה מתפלג לפי התפלגות מסוימת - H_0 .

אחרת - H_1 .

כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש. $d.f = K - 1$, כאשר K - מספר הקטגוריות.



הערך הקריטי הוא: $\chi^2_{1-\alpha, K-1}$, כלומר האחוזון ה- $1-\alpha$ בהתפלגות חי בריבוע שדרגות החופש הן $K-1$.

אם $\chi^2 > \chi^2_{1-\alpha, K-1}$, דוחים את השערת האפס.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad \text{סטטיסטי המבחן}$$

O_i - השכיחות שנצפתה במדגם בקטגוריה i .

p_i - הסתברות לקטגוריה i לפי השערת האפס.

$E_i = np_i$ - שכיחות צפויה במדגם לקטגוריה i בהנחת השערת האפס.

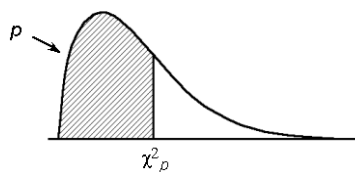
הערה:

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.

דוגמה (פתרון בהקלטה):

במדינה מסוימת שלוש מפלגות. בפרלמנט הנוכחי התפלגות מספר המושבים היא 30% למפלגה A, 60% למפלגה B ו-10% למפלגה C. לקראת הבחירות המתוכננות בשבוע הבא נעשה סקר שכלל 300 אזרחים. בסקר התקבל ש-40% יצביעו למפלגה A, 50% למפלגה B ו-10% למפלגה C. האם תוצאות הסקר תואמות להתפלגות המושבים בפרלמנט הנוכחי? בדקו ברמת מובהקות של 5%.

טבלת התפלגות חי-בריבוע – ערכי החלוקה



df	p												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.004393	0.004393	0.004393	0.004393	0.004393	0.004393	0.004393	0.004393	0.004393	0.004393	0.004393	0.004393
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) במטרה לבדוק האם קובייה הוגנת, מטילים אותה 120 פעמים. התקבל 17 פעמים 1, 23 פעמים 2, 20 פעמים 3, 25 פעמים 4, 18 פעמים 5 ו-17 פעמים 6. מה מסקנתכם ברמת מובהקות של 5%?
- (2) מפעל מייצר סוכריות בצבעים כחול, אדום, ירוק וכתום. מעוניינים לבדוק שפרופורציית הסוכריות הכחולות גדולה פי 2 מכל צבע אחר. לצורך כך נדגמו באקראי 200 סוכריות והתקבל: 70 כחולות, 50 אדומות, 40 ירוקות והיתר כתומות. מה מסקנתכם ברמת מובהקות של 5%?
- (3) משרד החינוך טוען שבקרב השכירים במשק היחס בין השכירים בעלי השכלה נמוכה, תיכונית ואקדמאית הוא 1:2:1 בהתאמה. במדגם של 200 שכירים התקבלו 56 אנשים בעלי השכלה נמוכה, 105 בעלי השכלה תיכונית והיתר בעלי השכלה גבוהה.
- א. על סמך תוצאות המדגם, האם התפלגות ההשכלה היא כמו שמשדרד החינוך מפרסם? בדוק ברמת מובהקות של 5%.
- ב. בנו רווח סמך ברמת סמך של 95% לפרופורציית השכירים במשק בעלי השכלה אקדמאית.
- (4) 200 איש נתבקשו לבחור ספרה באקראי והנה התוצאות שהתקבלו:
- 18 איש בחרו בספרה 0, 24 איש בחרו בספרה 1, 17 איש בחרו בספרה 2, 19 איש בחרו בספרה 3, 20 איש בחרו בספרה 4, 18 איש בחרו בספרה 5, 22 איש בחרו בספרה 6 והיתר בחרו בספרות 7-9.
- א. על סמך התוצאות הללו האם בחירת הספרות אקראית? בדקו ברמת מובהקות של 2.5%.
- ב. תנו הערכה למובהקות התוצאה.
- ג. אם נגדיל את גודל המדגם פי 2 ונשמור על אותם יחסים של כמות האנשים במדגם שבחרו בספרות, כיצד הדבר ישפיע על ערכו של הסטטיסטי χ^2 ? מה תהיה המסקנה במקרה זה?

- (5) מעוניינים לבדוק האם קובייה היא הוגנת. הטילו את הקובייה פעמיים והתבוננו בסכום הוצאות. חזרו על התהליך 72 פעמים.

להלן התוצאות שהתקבלו במדגם:

מה המסקנה ברמת מובהקות של 5%?

מספר הטלות	סכום התוצאות
20	2-5
17	6-8
20	9-10
15	11-12

- (6) בפנס יש 4 סוללות. בבדיקה שנערכה ב-400 פנסים נמצאו סוללות פגומות לפי השכיחויות הבאות:

מספר הסוללות הפגומות	0	1	2	3 ומעלה
שכיחות	276	104	12	8

מעוניינים לבדוק על סמך תוצאות מדגם אלה האם הסיכוי לסוללה פגומה הוא 20%. מהי רמת המובהקות המינימלית עבורה נכריע שהסיכוי לסוללה פגומה אינו 20%?

- (7) להלן השערות מחקר: $H_0: X \sim N(40, 2^2)$, $H_1: \text{else}$.

מספר הדגימות	X	מתחת 36	36-40	40-44	מעל 44
מספר הדגימות	3A	50A	45A	2A	

תוצאות המדגם הן:

מהו ערכו המקסימלי של A עבורו נקבל את H_0 ברמת מובהקות של 5%?

תשובות סופיות

- (1) לא נדחה H_0 .
- (2) לא נדחה H_0 .
- (3) א. לא נדחה H_0 . ב. (0.14, 0.25).
- (4) א. לא נדחה H_0 . ב. בין 0.95 ל-0.975.
- ג. יגדל פי 2; המסקנה לא תשתנה.
- (5) נכריע שהקובייה אינה הוגנת.
- (6) 0.005.
- (7) 14.

הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה – רקע

אם אנו רוצים לבצע מבחן טיב התאמה על משתנה שיש לו שתי קטגוריות בלבד (משתנה דיכוטומי), הדבר זהה לתהליך של בדיקת השערות דו צדדית על פרופורציה בודדת.

דוגמה (פתרון בהקלטה) :

הטילו מטבע 80 פעמים וקיבלו 48 פעמים את התוצאה "ראש".
בדקו האם המטבע הוא הוגן ברמת מובהקות של 5%.

א. באמצעות מבחן טיב התאמה.

ב. באמצעות מבחן Z לפרופורציה בודדת.

שאלות

(1) בסקר שנעשה על 320 נשאלים, 43.75% טענו שהחיה המועדפת עליהם היא כלב. עד היום היה נהוג לחשוב ש-40% מהאנשים מעדיפים כלבים. בדקו ברמת מובהקות של 5% האם הסקר יישנה את הסברה שהייתה נהוגה עד היום לגבי העדפת כלב.

א. באמצעות מבחן טיב התאמה.

ב. באמצעות מבחן על פרופורציה.

(2) לסוכנות מכוניות שלושה סניפים ברחבי הארץ. המכוניות נמכרות בסניפים השונים. מתוך 100 מכוניות נמצא ש-65 נמכרו בסניף תל-אביב, 23 בסניף ירושלים והיתר בסניף חיפה.

א. בדקו ברמת מובהקות של 5% האם שיעור המכוניות שנמכרות בסניף ת"א גדול פי 2 מכל סניף אחר.

ב. בדקו באמצעות מבחן טיב התאמה האם 60% מהמכוניות נהוגות להימכר בסניף תל אביב. האם יש דרך אחרת לבדוק את ההשערה?

(3) בתחרות ריצה בית ספרית שלושה מסלולי ריצה.

ב-50 תחרויות בדקו באיזה מסלול היה הניצחון.

התוצאות שהתקבלו מסוכמות בטבלה הבאה:

המסלול	1	2	3
מספר הניצחונות	20	15	15

א. בדקו ברמת מובהקות של 5% האם יש מסלול מועדף לניצחון.

ב. בדקו ברמת מובהקות של 5% האם הסיכוי לנצח במסלול מספר 1

גבוה מ- $\frac{1}{3}$.

תשובות סופיות

(1) לא נדחה H_0 .

(2) א. נדחה H_0 . ב. לא נדחה H_0 .

(3) א. לא נדחה H_0 . ב. לא נדחה H_0 .

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 .

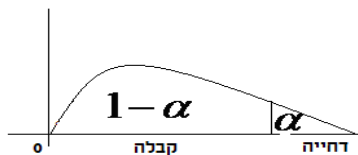
כלל הכרעה:

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.

$$E_i = \frac{f(x) \cdot f(y)}{n}$$

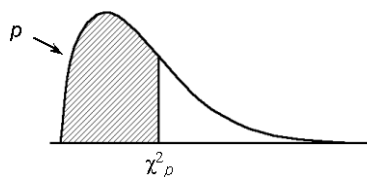
הערה:

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
פגומים	70	60	50
תקינים	800	700	600

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) בחנות בגדים A בדקו את התפלגות הצבעים של הבגדים הנמכרים ביום מסוים. כמו כן בדקו את התפלגות הצבעים בחנות שכנה B:

מספר פריטים / צבע	שחור	לבן	אדום	כחול
חנות A	15	20	15	50
חנות B	60	20	10	20

- א. בדקו ברמת מובהקות של 5% האם התפלגות הצבעים בחנות A היא ביחס של 1:1:1:3 לטובת הכחול.
 ב. בדקו ברמת מובהקות של 2.5% האם קיים הבדל בין החנויות מבחינת התפלגות הצבעים של הפריטים הנמכרים.

- (6) 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.

- (7) 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) א. נסיק שהתפלגות הצבעים בחנות היא כמו שמצוין.
 ב. נסיק שיש הבדל בין החנויות מבחינת התפלגות הצבעים.
 (6) נכון
 (7) 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	$f(y)$

הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות – רקע

מבחן לאי תלות שבו לכל משתנה יש שתי קטגוריות שקול לבדיקת השערות דו צדדית על הפרש פרופורציות כאשר השערת האפס היא שהפרופורציות שוות. כל זאת, כמובן, אם התנאים למבחנים מתקיימים.

דוגמה (פתרון בהקלטה) :

בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? האם אפשר לבדוק זאת גם על ידי מבחן לאי תלות וגם על ידי בדיקת השערות לשתי פרופורציות?

שאלות

- 1) בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? בדוק ברמת מובהקות של 5%.
א. על ידי מבחן לאי תלות.
ב. על ידי בדיקת השערות לשתי פרופורציות.
- 2) נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.
א. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.
ב. האם ניתן להגיד שהסיכוי להימצא במצב בריאותי תקין גבוה יותר כאשר עוסקים בפעילות גופנית סדירה לעומת המצב שלא עוסקים בפעילות גופנית סדירה? בדוק ברמת בטחון של 90%.

תשובות סופיות

- (1) נדחה את השערת האפס לפירוט הדרך ראה וידאו.
 (2) א. להלן טבלה: ב. נדחה את השערת האפס.

	לא תקין	תקין	פעילות/מצב בריאותי
60	10	50	סדירה
140	50	90	לא סדירה
200	60	140	

סטטיסטיקה

פרק 48 - מבחנים אפרמטרים למדגמים מזווגים

תוכן העניינים

1. מבחן ווילקוקסון..... 247
2. מבחן ווילקוקסון - על ידי שימוש בטבלה לערכים קריטיים..... 250

מבחן ווילקוקסון למדגמים מזווגים לפי שיטת המנייה – רקע

מתי נשתמש במבחן זה?

מבחן זה לא דורש הנחה של התפלגות נורמלית, אולם דורש ערכים מספריים המאפשרים חישוב הפרש בין ערכי X לערכי Y . מבחן זה הוא הגרסה הלא פרמטרית למבחן T למדגמים מזווג. נשתמש במבחן זה שיש משתנה כמותי שאינו מתפלג נורמלית או שיש משתנה מסולם סדר והמדגם הוא מזווג.

דוגמה (פתרון בהקלטה) :

שני קונדיטורים מתחרים על מקום עבודה. נתנו לשניהם להכין 8 מאפים שונים כאשר כל אחד מהמאפים נאפה על ידי שניהם. בסופו של דבר בעל הקונדיטוריה נתן ציון לכל אחד מהאופים בעבור כל אחד מהמאפים. להלן הציונים שהתקבלו, ורוצים לבדוק האם אופה א טוב יותר מאופה ב.

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

א. מהו המבחן הסטטיסטי המתאים?

ב. מהן השערות המחקר?

חישוב סטטיסטי המבחן:

- נחשב את ההפרשים D_i לכל תצפית.
- נוציא מהמדגם את כל התצפיות עם ההפרשים ששווים ל-0.
- נדרג את ההפרשים הנותרים מהקטן אל הגדול בלי להתייחס לסימן ההפרש, כלומר מדרגים את הערכים המוחלטים של ההפרשים. הפרשים זהים מקבלים דרגה זהה שהיא הדרגה הממוצעת של המקומות שהם תופסים.
- מסכמים את הדרגות של ההפרשים החיוביים ($W+$) ואת הדרגות של ההפרשים השליליים ($W-$).

דוגמה (פתרון בהקלטה) :

חשבו את W על סמך תוצאות המדגם.

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

חישוב מובהקות התוצאה :

כדי לחשב את מובהקות התוצאה נצטרך לדון בהתפלגות של W בהנחת השערת האפס. אם השערת האפס נכונה, לכל D_i יש את אותו סיכוי לקבל ערך חיובי או ערך שלילי כלומר 2 אפשרויות. ל n^* דרגות יש $2n^*$ אפשרויות שונות, שוות הסתברות. לפי רעיון זה ניתן יהיה לחשב את מובהקות התוצאה.

דוגמה (פתרון בהקלטה) :

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

שאלות

- (1) נדגמו 8 לקוחות שקיבלו שירות ממוקד טלפוני. לקוחות אלה נתבקשו לתת הערכה על יעילות השירות ועל האדיבות שבשירות. הציונים ניתנו בסקאלה מ-1 (הערכה הנמוכה) עד 10 (הערכה הגבוהה ביותר). להלן התוצאות שהתקבלו:

5	7	5	2	3	4	8	7	הערכה על יעילות השירות	X
4	7	10	8	6	7	7	8	הערכה על אדיבות השירות	Y

בדקו ברמת מובהקות של 5% האם קיים הבדל בין הערכה על יעילות השירות להערכה על אדיבות השירות?

- (2) רוצים לבדוק האם תרופה חדשה להקלת כאבי ראש יעילה יותר מתרופה מוכרת. לצורך כך נלקח מדגם בן 9 אנשים, שנתבקשו להשתמש בתרופה החדשה ובתרופה המוכרת, ולהשוות את יעילותה של התרופה החדשה ליעילות התרופה המוכרת. האנשים במחקר היו צריכים לתת הערכה של יעילות בסקלה של מ-1 עד 100. התוצאות שקיבל היו:

9	8	7	6	5	4	3	2	1	הנבדק
86	100	69	81	75	80	100	90	95	תרופה חדשה
60	60	50	70	75	49	65	76	80	תרופה מוכרת

האם התרופה החדשה משפרת את היעילות ביותר מ 10 נקודות? בדקו ברמת מובהקות של 1%.

תשובות סופיות

- (1) לא נדחה H_0 .
 (2) לא נדחה H_0 .

מבחן ווילקוקסון למדגמים מזווגים (על ידי שימוש בטבלה של ערכים

קריטיים) – רקע

מתי נשתמש במבחן זה?

מבחן זה לא דורש הנחה של התפלגות נורמלית, אולם דורש ערכים מספריים המאפשרים חישוב הפרש בין ערכי X לערכי Y . מבחן זה הוא הגרסה הלא פרמטרית למבחן T למדגמים מזווג. נשתמש במבחן זה שיש משתנה כמותי שאינו מתפלג נורמלית או שיש משתנה מסולם סדר על מדגם מזווג.

דוגמה (פתרון בהקלטה) :

שני קונדיטורים מתחרים על מקום עבודה. נתנו לשניהם להכין 8 מאפים שונים כאשר כל אחד מהמאפים נאפה על ידי שניהם. בסופו של דבר בעל הקונדיטוריה נתן ציון לכל אחד מהאופים בעבור כל אחד מהמאפים. להלן הציונים שהתקבלו, ורוצים לבדוק שאופה א טוב יותר מאופה ב.

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

א. מהו המבחן הסטטיסטי המתאים?

ב. מהן השערות המחקר?

חישוב סטטיסטי המבחן:

- (1) נחשב את ההפרשים D_i לכל תצפית.
- (2) נוציא מהמדגם את כל התצפיות עם ההפרשים ששווים ל-0.
- (3) נדרג את ההפרשים הנותרים מהקטן אל הגדול בלי להתייחס לסימן ההפרש, כלומר מדרגים את הערכים המוחלטים של ההפרשים. הפרשים זהים מקבלים דרגה זהה שהיא הדרגה הממוצעת של המקומות שהם תופסים.
- (4) מסכמים את הדרגות של ההפרשים החיוביים (W^+) ואת הדרגות של ההפרשים השליליים (W^-).
- (5) W יהיה W^+ או W^- , זה שאמור להיות יותר קטן לפי השערת המחקר או הקטן מבניהם אם ההשערה היא דו צדדית.

דוגמה (פתרון בהקלטה) :

חשבו את W על סמך תוצאות המדגם.

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

כלל הכרעה :

במבחן ווילקוקסון זה כלל ההכרעה הוא : נדחה את H_0 אם $W \leq W_c$.
 כאשר, W_c - הערך הקריטי ; W - הסטטיסטי.
 את הערכים הקריטיים נחלץ מתוך טבלה מתאימה :

n_1	חד-צדדי $\alpha = 0.01$ דו-צדדי $\alpha = 0.02$	חד-צדדי $\alpha = 0.025$ דו-צדדי $\alpha = 0.05$	חד-צדדי $\alpha = 0.05$ דו-צדדי $\alpha = 0.10$
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

דוגמה (פתרון בהקלטה) :

- א. רשמו את כלל ההכרעה המתאים ברמת מובהקות של 5%.
- ב. מה המסקנה ברמת מובהקות של 5%?

שאלות

- (1) נדגמו 8 לקוחות שקיבלו שירות ממוקד טלפוני. לקוחות אלה נתבקשו לתת הערכה על יעילות השירות ועל האדיבות שבשירות. הציונים ניתנו בסקאלה מ-1 (הערכה הנמוכה) עד 10 (הערכה הגבוהה ביותר). להלן התוצאות שהתקבלו:

5	7	5	2	3	4	8	7	הערכה על יעילות השירות	X
4	7	10	8	6	7	7	8	הערכה על אדיבות השירות	Y

בדקו ברמת מובהקות של 5% האם קיים הבדל בין הערכה על יעילות השירות להערכה על אדיבות השירות?

- (2) סטודנטים נתבקשו לתת חוות דעתם על רמת הקושי של הקורס (סקאלה של 1-5 כאשר 5=קשה ביותר) ועל רמת הקושי של הבחינות באותה סקאלה. הסטודנטים טוענים שהבחינה הייתה ברמה גבוהה יותר מהרמה של הקורס. להלן תוצאות המדגם:

4	5	1	2	3	4	2	3	4	1-קושי קורס
2	3	5	5	5	3	4	4	4	2-קושי בחינה

בדקו ברמת מובהקות של 5% את טענת הסטודנטים.

- (3) רוצים לבדוק את הטענה שהציונים במבחן בסטטיסטיקה ב גבוהים מאשר בסטטיסטיקה א. נלקחו 10 סטודנטים שסיימו את סטטיסטיקה ב. עבור כל סטודנט נבדק מה הציון בסטטיסטיקה א ומה הציון בסטטיסטיקה ב. להלן התוצאות שהתקבלו:

80	78	84	65	67	82	94	68	74	62	א
82	79	86	80	67	77	90	80	80	70	ב

- א. בדקו ברמת מובהקות של 5% באמצעות מבחן ווילקוקסון.
- ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם יוחלט לתת פקטור של 2 נקודות לכל הסטודנטים בשני המועדים?
- ג. כיצד הייתה משתנה התשובה אם מסתבר שנפלה טעות ועבור הסטודנט הראשון ברשימה יש להחליף בנתונים את הציון של סטטיסטיקה ב עם סטטיסטיקה א?

- (4) רוצים לבדוק האם תרופה חדשה להקלת כאבי ראש יעילה יותר מתרופה מוכרת. לצורך כך נלקח מדגם בן 9 אנשים, שנתבקשו להשתמש בתרופה החדשה ובתרופה המוכרת, ולהשוות את יעילותה של התרופה החדשה ליעילות התרופה המוכרת.
- האנשים במחקר היו צריכים לתת הערכה של יעילות בסקלה של מ-1 עד 100. התוצאות שקיבל היו:

הנבדק	1	2	3	4	5	6	7	8	9
תרופה חדשה	95	90	100	80	75	81	69	100	86
תרופה מוכרת	80	76	65	49	75	70	50	60	60

האם התרופה החדשה משפרת את היעילות ביותר מ 10 נקודות? בדקות ברמת מובהקות של 1%.

תשובות סופיות

- (1) לא נדחה H_0 .
- (2) לא נדחה H_0 .
- (3) א. לא נדחה H_0 . ב. לא משתנה. ג. לא משתנה.
- (4) לא נדחה H_0 .

סטטיסטיקה

פרק 49 - מבחנים אפרמטריים למדגמים בלתי תלויים

תוכן העניינים

1. מבחן ווילקוקסון למדגמים בלתי תלויים..... 254

מבחן ווילקוקסון למדגמים בלתי תלויים – רקע

מבחן ווילקוקסון למדגמים בלתי תלויים נכנס לקטגוריות המבחנים האפרמטריים. מבחן זה רלבנטי כאשר רוצים להשוות בין שתי אוכלוסיות על סמך שני מדגמים בלתי תלויים. המשתנה התלוי הוא משתנה כמותי שאינו מתפלג נורמלית או משתנה מסולם סדר. מבחן זה הוא החלופה האפרמטרית למבחן הפרמטרי להשוואת תוחלות על סמך שני מדגמים בלתי תלויים.

דוגמה:

מחקר חינוכי מעוניין להשוות בין 2 שיטות חינוך. המחקר רוצה לבדוק האם קיים הבדל ברמת ביטחון העצמי של הילדים בשיטות החינוך השונות. נבחרו באקראי 5 ילדים שחונכו בשיטת A. כמו כן נדגמו באקראי 5 ילדים שחונכו בשיטת B. פסיכולוגים בחנו את 10 הילדים ונתנו ציון לביטחון העצמי בסקאלה של 1-20. מהן ההשערות ומהו המבחן הסטטיסטי המתאים?

שיטה A	שיטה B
16	14
17	14
20	19
10	9
18	8

כדי לבצע את המבחן יש לחשב על סמך תוצאות המדגם את סטטיסטי המבחן שנסמן באות U.

השלבים לחישוב סטטיסטי המבחן:

1. נסדר את כלל התצפיות של המחקר בסדר עולה מהנמוך ביותר לגבוה ביותר אך יש לדעת כל תצפית מאיזה מדגם היא באה.
2. נדרג את כלל התצפיות של המחקר (אם יש תצפיות עם ערכים זהים הדירוג שלהן יהיה ממוצע המקומות שהם תופסים)
3. נחשב את W_1 - סכום הדירוגים של התצפיות השייכות למדגם 1,
- ו- W_2 - סכום הדירוגים של התצפיות השייכות למדגם 2.
4. נחשב את הגדלים הבאים:

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2} \quad U_2 = W_2 - \frac{n_2(n_2+1)}{2}$$

5. הסטטיסטי U : במבחן דו צדדי $U = \min(U_1, U_2)$ במבחן חד צדדי U יהיה ה- U_i , שאמור להיות יותר קטן לפי השערת המחקר.

כדי להגיע למסקנה יש שני סוגים של טבלאות סטטיסטיות.
 טבלה מהסוג הראשון: עוזרת לנו לחשב את מובהקות התוצאה לאחר שחישבנו את
 ה- U הסטטיסטי.

טבלה מהסוג השני שקובעת מראש את הערך הקריטי של U שנסמן ב U_c .

טבלה מהסוג הראשון	טבלה מהסוג השני
$n_1 - n_2 \leq 8$	n_1 או $n_2 \geq 9$
עוזרת לחשב על סמך תוצאות המדגם את P_v אם $P_v \leq \alpha$ דוחים את H_0 .	נותנת את הערך הקריטי U_c כל הכרעה: נדחה את H_0 אם $U \leq U_c$

דוגמה:

מה המסקנה ברמת מובהקות של 5%?

טבלאות למציאת מובהקות התוצאה במבחן ווילקוקסון למדגמים בלתי תלויים

 $n_2 = 3$

u	n_1		
	1	2	3
0	0.250	0.100	0.050
12	0.500	0.200	0.100
3	0.750	0.400	0.200
4		0.600	0.350
5			0.500
			0.650

 $n_2 = 4$

u	n_1			
	1	2	3	4
0	0.200	0.067	0.028	0.014
1	0.400	0.133	0.057	0.029
2	0.600	0.267	0.114	0.057
3		0.400	0.200	0.100
4		0.600	0.314	0.171
5			0.429	0.243
6			0.571	0.343
7				0.443
8				0.557

 $n_2 = 5$

u	n_1				
	1	2	3	4	5
0	0.167	0.047	0.018	0.008	0.004
1	0.333	0.095	0.036	0.016	0.008
2	0.500	0.190	0.071	0.032	0.016
3	0.667	0.286	0.125	0.056	0.028
4		0.429	0.196	0.095	0.048
5		0.571	0.286	0.143	0.075
6			0.393	0.206	0.111
7			0.500	0.278	0.155
8			0.607	0.365	0.210
9				0.452	0.274
10				0.548	0.345
11					0.421
12					0.500
13					0.579

$$n_2 = 6$$

u	n_1					
	1	2	3	4	5	6
0	0.143	0.036	0.012	0.005	0.002	0.001
1	0.286	0.071	0.024	0.010	0.004	0.002
2	0.428	0.143	0.048	0.019	0.009	0.004
3	0.571	0.214	0.083	0.033	0.015	0.008
4		0.321	0.131	0.057	0.026	0.013
5		0.429	0.190	0.086	0.041	0.021
6		0.571	0.274	0.129	0.063	0.032
7			0.357	0.176	0.089	0.047
8			0.452	0.238	0.123	0.066
9			0.548	0.305	0.165	0.090
10				0.381	0.214	0.120
11				0.457	0.268	0.155
12				0.545	0.331	0.197
13					0.396	0.242
14					0.465	0.294
15					0.535	0.350
16						0.409
17						0.469
18						0.531

$$n_2 = 7$$

u	n_1						
	1	2	3	4	5	6	7
0	0.125	0.028	0.008	0.003	0.001	0.001	0.000
1	0.250	0.056	0.017	0.006	0.003	0.001	0.001
2	0.375	0.111	0.033	0.012	0.005	0.002	0.001
3	0.500	0.167	0.058	0.021	0.009	0.004	0.002
4	0.625	0.250	0.092	0.036	0.015	0.007	0.003
5		0.333	0.133	0.055	0.024	0.011	0.006
6		0.444	0.192	0.082	0.037	0.017	0.009
7		0.556	0.258	0.115	0.053	0.026	0.013
8			0.333	0.158	0.074	0.037	0.019
9			0.417	0.206	0.101	0.051	0.027
10			0.500	0.264	0.134	0.069	0.036
11			0.583	0.324	0.172	0.090	0.049
12				0.394	0.216	0.117	0.064
13				0.464	0.265	0.147	0.082
14				0.538	0.319	0.183	0.104
15					0.378	0.223	0.130
16					0.438	0.267	0.159
17					0.500	0.314	0.191
18					0.562	0.365	0.228
19						0.418	0.267
20						0.473	0.310
21						0.527	0.355
22							0.402
23							0.451
24							0.500
25							0.549

$$n_2 = 8$$

u	n_1							
	1	2	3	4	5	6	7	8
0	0.111	0.022	0.006	0.002	0.001	0.000	0.000	0.000
1	0.222	0.044	0.012	0.004	0.002	0.001	0.000	0.000
2	0.333	0.089	0.024	0.008	0.003	0.001	0.001	0.000
3	0.444	0.133	0.042	0.014	0.005	0.002	0.001	0.001
4	0.556	0.200	0.067	0.024	0.009	0.004	0.002	0.001
5		0.267	0.097	0.036	0.015	0.006	0.003	0.001
6		0.356	0.139	0.055	0.023	0.010	0.005	0.002
7		0.444	0.188	0.077	0.033	0.015	0.007	0.003
8		0.556	0.248	0.107	0.047	0.021	0.010	0.005
9			0.315	0.141	0.064	0.030	0.014	0.007
10			0.387	0.184	0.085	0.041	0.020	0.010
11			0.461	0.230	0.111	0.054	0.027	0.014
12			0.539	0.285	0.142	0.071	0.036	0.019
13				0.341	0.177	0.091	0.047	0.025
14				0.404	0.217	0.114	0.060	0.032
15				0.467	0.262	0.141	0.076	0.041
16				0.533	0.311	0.172	0.095	0.052
17					0.362	0.207	0.116	0.065
18					0.416	0.245	0.140	0.080
19					0.472	0.286	0.168	0.097
20					0.528	0.331	0.198	0.117
21						0.377	0.232	0.139
22						0.426	0.268	0.164
23						0.475	0.306	0.101
24						0.525	0.347	0.221
25							0.389	0.253
26							0.433	0.287
27							0.478	0.323
28							0.522	0.360
29								0.399
30								0.439
31								0.480
32								0.520

טבלה למציאת U_c

ברמת מובהקות של 5% למבחן חד צדדי או ברמת מובהקות של 10% למבחן דו צדדי במבחן ווילקוקסון למדגמים בלתי תלויים.

n_1	n_2											
	9	10	11	12	13	14	15	16	17	18	19	20
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

שאלות

- (1) מעוניינים להשוות בין שתי קבוצות כדורסל. נלקחו 5 משחקים מקבוצה א' ושישה משחקים מקבוצה ב'. נבדק בכל משחק ועבור כל קבוצה מספר הנקודות שצברה במשחק. בדקו ברמת מובהקות של 5% האם קיים הבדל בין הקבוצות מבחינת הניקוד שצברה במשחק.

קבוצה א	קבוצה ב
68	82
82	74
78	82
94	64
87	67
	65

- (2) מעוניינים לבדוק האם קורס קיץ באנגלית משפר את יכולות האנגלית לתלמידי חטיבת ביניים. נלקחו 20 ילדים בגיל חטיבת הביניים ברמת אנגלית דומה. 12 מהם נשלחו לקורס קיץ והיתר לא. בסוף הקיץ כולם נבחנו במבחן באנגלית הציון הגבוה ביותר התקבל בקרב אחד שלא עשה את הקורס ושבעת הציונים הנמוכים ביותר היו גם בקרב תלמידים שלא עשו את הקורס. מה המסקנה ברמת מובהקות של 5%?
- (3) במחקר לבדיקת יעילות ויטמין C נבחרו 15 מתנדבים מבין עובדי המפעל. תשעה מהם נבחרו מקרית וקיבלו טיפול שוטף בוויטמין C, ואילו שאר המתנדבים (קבוצת הביקורת) קבלו גלולת סוכר. במשך שלוש שנות המחקר היו מספר ימי ההיעדרות בגלל ההצטננות:
 קבוצת הטיפול: 1, 3, 9, 3, 4, 0, 8, 12, 16.
 קבוצת הביקורת: 19, 7, 28, 13, 23, 12.
 בדקו ברמת מובהקות של 5% שמספר ימי המחלה במשך שלוש שנים מצטמצם ביותר מ-4 ימים עם לקיחת ויטמין C.

תשובות סופיות

- (1) לא נדחה את H_0 .
 (2) נדחה את H_0 .
 (3) לא נדחה את H_0 .

\\