סטטיסטיקה ואמינות

תוכן העניינים

1. יסודות ההסתברות 1
2. פעולות בין מאורעות (חיתוך ואיחוד) - מאורעות זרים ומכילים 5
3. קומבינטוריקה -כלל המכפלה 14
4. קומבינטוריקה- תמורה - סידור עצמים בשורה 18
5. קומבינטוריקה -דגימה סידורית ללא החזרה ועם החזרה 21
6. קומבינטוריקה - דגימה ללא סדר וללא החזרה 23
7. קומבינטוריקה - דגימה ללא סדר ועם החזרה 26
8. קומבינטוריקה - שאלות מסכמות 30
9. הסתברות מותנית-במרחב מדגם אחיד 37
10. הסתברות מותנית - מרחב לא אחיד 40
11. דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה 44
12. תלות ואי תלות בין מאורעות 49
13. שאלות מסכמות בהסתברות 53
14. המשתנה המקרי הבדיד - פונקציית ההסתברות 58
15. המשתנה המקרי הבדיד - תוחלת - שונות וסטיית תקן 62
16. המשתנה המקרי הבדיד -תוחלת של פונקציה של משתנה מקרי בדיד 66
17. המשתנה המקרי הבדיד- טרנספורמציה לינארית 69
18. תוחלת ושונות של סכום משתנים מקריים 72
19. התפלגויות בדידות מיוחדות -התפלגות בינומית 75
20. התפלגויות בדידות מיוחדות -התפלגות גיאומטרית 79
21. התפלגויות בדידות מיוחדות- התפלגות פואסונית 82
22. התפלגויות בדידות מיוחדות-התפלגות היפרגאומטרית 85
23. התפלגויות בדידות מיוחדות -התפלגות בינומית שלילית 88

תוכן העניינים

24. המשתנה המקרי הבדיד - שאלות מסכמות 91
25. המשתנה המקרי הרציף- התפלגויות כלליות (שימוש באינטגרלים) 98
26. התפלגויות רציפות מיוחדות- התפלגות מעריכית 107
27. התפלגויות רציפות מיוחדות - התפלגות נורמלית 110
28. התפלגות מינימום ומקסימום 118
29. שאלות בהסתברות על מערכות 122
30. סטטיסטיקה תיאורית-הקדמה 125
31. סטטיסטיקה תיאורית- הצגה של נתונים 128
32. סטטיסטיקה תיאורית- סכימה 139
33. סטטיסטיקה תיאורית- מדדי מרכז ..... (ללא ספר)
34. סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן 143
35. סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני 146
36. סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה 150
37. סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן 152
38. סטטיסטיקה תיאורית- תרשים קופסא 154
39. הסקה סטטיסטית - הקדמה 156
40. התפלגות הדגימה ומשפט הגבול המרכזי 159
41. מושגי יסוד באמידה 179
42. רווח סמך לתוחלת (ממוצע) 184
43. רווח סמך לפרופורציה 194
44. רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים 200
45. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 202
46. שאלות מסכמות על רווחי סמך 204
47. מבוא לבדיקת השערות על פרמטרים 206
48. בדיקת השערות על תוחלת (ממוצע) 212
49. בדיקת השערות על פרופורציה 244
50. בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים 257
51. בדיקת השערות לתוחלת ההפרש במדגמים מזווגים 272

52. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות 282
53. שאלות מסכמות בבדיקת השערות 285
54. מבחני חי בריבוע 301
55. ניתוח שונות חד כיוונית 305
56. מקדם המתאם (מדד קשר) הלינארי ומובהקותו 314
57. רגרסיה 334
58. מדדי קשר-רגרסיה -שונות מוסברת ושונות לא מוסברת 337
59. מבחן הבינום 340

סטטיסטיקה ואמינות

פרק 1 - יסודות ההסתברות

תוכן העניינים

1. כללי 1

הגדרות יסודיות:

רקע:

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

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

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.

© כל הזכויות שמורות לברק קנדל

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: $P(A)=1-\frac{2}{25}=\frac{23}{25}$.

© כל הזכויות שמורות לברק קנדל

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 ההסברות שהספרה שנבחרה גדולה מ-5?

ב. מה ההסתברות שהספרה שנבחרה היא לכל היותר 3?

ג. מה ההסתברות שהספרה שנבחרה היא אי זוגית?

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

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

נבחרה משפחה באקראי מהישוב.

א. מה ההסתברות שאין מקלטים למשפחה?

ב. מה ההסתברות שיש מקלטים למשפחה?

ג. מה ההסתברות שיש לפחות 3 מקלטים למשפחה?

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

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

נבחרה משפחה אקראית מן הישוב.

א. מה ההסתברות שאין לה מכוניות?

ב. מה ההסתברות שבבעלות המשפחה לפחות 3 מכוניות?

ג. מה הסיכוי שבבעלותה פחות מ-3 מכוניות?

© כל הזכויות שמורות לברק קנדל

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. $\Omega=\begin{Bmatrix}(1,1) & (2,1) & (3,1) & (5,1) & (4,1) & (6,1)\\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2)\\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3)\\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4)\\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5)\\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6)\end{Bmatrix}$.

ב. $C=\{(2,6),(3,4),(4,3),(6,2)\}$, $A=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$.

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. $D=\{PPP,PPE,PEP,EPP,PEE,EPE,EEP\}$, $A=\{PPE,PEP,EPP\}$.

ג. $\bar{D}=\{EEE\}$.

ד. $\frac{1}{8}$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

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

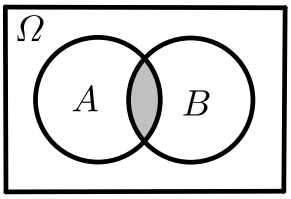
תוכן העניינים

1. כללי 5

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

רקע:

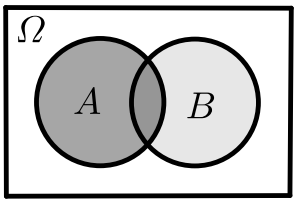
פעולת חיתוך:



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

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

פעולת איחוד:



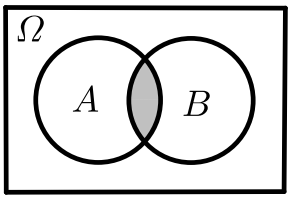
נותנת את כל האפשריות שנמצאות לפחות באחת מהמאורעות, ומסומנת: $A\cup B$.
הפעולה נותנת את אשר נמצא ב-A או B.
כלומר, לפחות אחד מהמאורעות קורה.

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?

© כל הזכויות שמורות לברק קנדל

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:

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

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

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

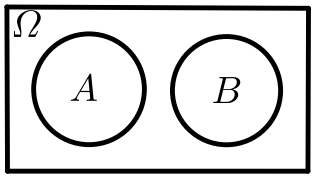
$$P(A\cup B)=1-P(\bar{A}\cap\bar{B})$$

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: $P(A\cap B)=0$.

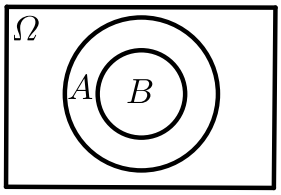
ההסתברות של איחוד המאורעות תחושב: $P(A\cup B)=P(A)+P(B)$.

דוגמה:

בהטלת קובייה, האפשרויות לקבל לפחות 5 הן: $A=\{5,6\}$ והאפשרות לקבל 3 היא: $B=\{3\}$, ולכן החיתוך ביניהם הוא אפס, כלומר: $A\cap B=\{\ \}$.

© כל הזכויות שמורות לברק קנדל

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: $B \subset A$.

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

למשל: $A = \{2,4,6\}$, $B = \{2,4\}$.

© כל הזכויות שמורות לברק קנדל

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 בדיאגראמת ון 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 , B , $\overline{B}$, $A\cap B$, $A\cup B$.
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:
$\overline{\overline{A}}$, $A\cap\phi$, $A\cup\phi$, $A\cap\Omega$, $A\cup\Omega$, $A\cap\overline{A}$, $\overline{\phi}$, $A\cup\overline{A}$.

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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: $P(A)=0.6$, $P(B)=0.3$, $P(\bar{A}\cap\bar{B})=0.1$.

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?

© כל הזכויות שמורות לברק קנדל

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: $2\cdot P(B\cap\overline{A})=P(A\cap\overline{B})=P(\overline{A}\cap\overline{B})$.

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

17) נתון ש-A ו-B מאורעות במרחב מדגם. נתון ש-$P(A)=0.3$, $P(B)=0.2$.

א. האם יתכן ש-$P(A\cup B)=0.4$?

ב. האם יתכן ש-$P(A\cup B)=0.6$?

ג. אם A ו-B זרים מה הסיכוי $P(A\cup B)$?

ד. אם A מכיל את B מה הסיכוי $P(A\cup B)$?

18) מתוך אזרחי המדינה הבוגרים ל-30% חשבון בבנק הפועלים. ל-28% חשבון בבנק לאומי ול-15% חשבון בבנק מזרחי. כמו כן נתון כי 6% מחזיקים חשבון בבנק לאומי ובבנק הפועלים. ל-5% חשבון בבנק פועלים ומזרחי. ול-4% חשבון בבנק לאומי ומזרחי. כמו כן ל-1% מהאוכלוסייה הבוגרת חשבון בנק בשלושת הבנקים יחד.

א. מה אחוז האזרחים להם חשבון בבנק לאומי בלבד?

ב. מה ההסתברות שאזרח כלשהו יחזיק חשבון בבנק פועלים ולאומי אבל לא בבנק מזרחי?

ג. מה ההסתברות שלאזרח יהיה חשבון בפועלים או במזרחי אבל לא בבנק לאומי?

ד. מה אחוז האזרחים שיש להם חשבון בנק אחד בלבד?

ה. מה אחוז האזרחים שיש להם בדיוק חשבון בשני בנקים בלבד?

ו. מה ההסתברות שלאזרח בוגר אין חשבון בנק באף אחד מהבנקים הללו?

ז. לאיזה אחוז מהאזרחים יש חשבון בנק בלפחות אחד מהבנקים הללו?

© כל הזכויות שמורות לברק קנדל

19) חברה מסוימת פרסמה את הנתונים הבאים לגבי האזרחים מעל גיל 21.
הנתונים שהתקבלו היו: 40% מהאנשים מחזיקים כרטיס "ויזה", 52% מחזיקים כרטיס "ישראכרט", 20% מחזיקים כרטיס "אמריקן אקספרס", 15% מחזיקים כרטיס ויזה וגם ישראכרט, 8% מחזיקים כרטיס ישראכרט וגם אמריקן אקספרס ו-7% מחזיקים כרטיס ויזה וגם אמריקן אקספרס.
כמו כן, 13% לא מחזיקים באף אחד משלושת הכרטיסים הנ"ל.

א. מה אחוז מחזיקי שלושת כרטיס האשראי גם יחד?

ב. מה אחוז מחזיקי ישראכרט וויזה אך לא את אמריקן אקספרס?

ג. מה אחוז מחזיקי כרטיס אחד בלבד?

20) הוכיחו: $P\left(\bar{A}\cap\bar{B}\right)=1-P\left(A\right)-P\left(B\right)+P\left(A\cap B\right)$.

21) A ו-B מאורעות במרחב המדגם. האם נכון לומר שהסיכוי שיתרחש בדיוק מאורע אחד הוא: $P\left(A\right)+P\left(B\right)-2P\left(A\cap B\right)$?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. $A\cap B=\{EG,EF,FE,GE\}$.
ב. $A\cup B=\{EG,EF,EE,FE,GE,EG,GF\}$.

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

3) א. $A=0,2,4,6,8$, $B=0,1,2,3,4$, $\bar{B}=5,6,7,8,9$,
$A\cap B=0,2,4$, $A\cup B=0,2,4,6,8,1,3$.
ב. $P(A)=0.5$, $P(B)=0.5$, $P(\bar{B})=0.5$, $P(A\cap B)=0.3$, $P(A\cup B)=0.7$.

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

5) א. $A\cap B$: גובה בין 1.7 ל-1.8. ב. $A\cup B$: כל גובה אפשרי .
ג. $\bar{A}=\bar{A}\cap B$: גובה לכל היותר 1.7. ד. $\bar{A}\cup\bar{B}$: לכל היותר 1.7 או לפחות 1.8.
ה. $A=\bar{\bar{A}}$: גובה מעל 1.7.

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

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

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

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

10) א. לא. ב. כן. ג. כן. ד. לא.

11) א. כן. ב. $P(\bar{A}\cap B)=0.3$.

12) הטענה הנכונה היא ג'.

13) א. 0.05. ב. 0.95.

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

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

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

17) א. כן. ב. לא. ג. $P(A\cup B)=0.5$. ד. $P(A\cup B)=0.3$.

18) א. 19%. ב. 0.05. ג. 0.31. ד. 46%. ה. 12%. ו. 0.41.
ז. 59%.

19) א. 5%. ב. 10%. ג. 67%.

20) שאלת הוכחה.

21) נכון.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 3 - קומבינטוריקה -כלל המכפלה

תוכן העניינים

1. כללי 14

קומבינטוריקה – כלל המכפלה:

רקע:

כלל המכפלה:

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

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

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

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

$n_1 = 6$, $n_2 = 2$

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

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

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

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

© כל הזכויות שמורות לברק קנדל

שאלות:

1) חשבו את מספר האפשרויות לתהליכים הבאים:

א. הטלת קובייה פעמים.

ב. מספר תלת ספרתי.

ג. בחירת בן ובת מכתה שיש בה שבעה בנים ועשר בנות.

ד. חלוקת שני פרסים שונים לעשרה אנשים שונים כאשר אדם לא יכול לקבל יותר מפרס אחד.

2) במסעדה מציעים ארוחה עסקית.
בארוחה עסקית יש לבחור מנה ראשונה, מנה עיקרית ושתייה.
האופציות למנה ראשונה הן: סלט ירקות, סלט אנטיפסטי ומרק היום.
האופציות למנה עיקרית הן: סטייק אנטריקוט, חזה עוף בגריל, לזניה בשרית ולזניה צמחונית. האופציות לשתייה הן: קפה, תה ולימונדה.

א. כמה ארוחות שונות ניתן להרכיב בעזרת התפריט הזה?

ב. אדם מזמין ארוחה אקראית. חשב את ההסתברויות הבאות:

i. בארוחה סלט ירקות, לזניה בשרית ולימונדה.

ii. בארוחה סלט, לזניה ותה.

3) בוחרים באקראי מספר בין חמש ספרות. חשבו את ההסתברויות הבאות:

א. המספר הוא זוגי.

ב. במספר כל הספרות שונות.

ג. במספר כל הספרות זהות.

ד. במספר לפחות שתי ספרות שונות.

ה. במספר לפחות שתי ספות זהות.

ו. המספר הוא פלינדרום (מספר הנקרא מימין ומשמאל באות הצורה).

4) חישה אנשים אקראיים נכנסו למעלית בבניין בן 8 קומות.
חשבו את ההסתברויות הבאות:

א. כולם ירו בקומה החמישית.

ב. כולם ירדו באותה קומה.

ג. כולם ירדו בקומה אחרת.

ד. ערן ודני ירדו בקומה השישית והיתר בשאר הקומות.

© כל הזכויות שמורות לברק קנדל

5) במפלגה חמישה עשר חברי כנסת. יש לבחור שלושה חברי כנסת לשלושה תפקידים שונים. בכמה דרכים ניתן לחלק את התפקידים הבאים אם:

א. חבר כנסת יכול למלא יותר מתפקיד אחד.

ב. חבר כנסת לא יכול למלא יותר מתפקיד אחד.

6) מטילים קובייה 4 פעמים.

א. מה ההסתברות שכל התוצאות תהינה זהות?

ב. מה ההסתברות שכל התוצאות תהינה שונות?

ג. מה ההסתברות שלפחות שתי תוצאות תהינה זהות?

ד. מה ההסתברות שלפחות שתי תוצאות תהיה שונות?

7) יש ליצור מילה בת חמש אותיות, לא בהכרח עם משמעות מאותיות ה-ABC (26 אותיות).

א. מה ההסתברות שבמילה שנוצרה אין האותיות D, A ו-L?

ב. מה ההסתברות שבמילה שנוצרה כל האותיות זהות?

ג. מה ההסתברות שבמילה שנוצרה לפחות שתי אותיות שונות זו מזו?

ד. מה ההסתברות שהמילה היא פלינדרום? (מילה אשר משמאל לימין, ומימין לשמאל נקראת אותו הדבר).

8) יוצרים קוד עם a ספרות (מותר לחזור על אותה ספרה בקוד).
חשבו את ההסתברויות הבאות: (בטאו את תשובותיכם באמצעות a).

א. בקוד אין את הספרה 5.

ב. בקוד מופיעה הספרה 3.

ג. בקוד לא מופיעות ספרות אי זוגיות.

9) במשחק מזל יש למלא טופס בו n משבצות. כל משבצת מסומנת בסימן V או X. בכמה דרכים שונות ניתן למלא את טופס משחק המזל?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. 36. ב. 900. ג. 70. ד. 90.

2) א. 36. ב. i. $\frac{1}{36}$. ב. ii. $\frac{1}{9}$.

3) א. 0.5. ב. 0.3024. ג. 0.0001. ד. 0.9999. ה. 0.6976. ו. 0.01.

4) א. $\frac{1}{8^5}$. ב. $\frac{1}{8^4}$. ג. 0.205. ד. $\frac{1\cdot 1\cdot 7^3}{8^5}$.

5) א. 3375. ב. 2730.

6) א. $\frac{1}{216}$. ב. $\frac{5}{18}$. ג. $\frac{13}{18}$. ד. $\frac{215}{216}$.

7) א. $\frac{23^5}{26^5}$. ב. $\frac{1}{26^4}$. ג. $1-\frac{1}{26^4}$. ד. $\frac{1}{26^2}$.

8) א. 0.9^a. ב. $1-0.9^a$. ג. 0.5^a.

9) 2^n.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 4 - קומבינטוריקה- תמורה - סידור עצמים בשורה

תוכן העניינים

1. כללי 18

קומבינטוריקה – תמורה – סידור עצמים בשורה:

רקע:

תמורה:

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

<u>הערה</u>: $0!=1$.

דוגמאות (הפתרונות בהקלטה):

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

© כל הזכויות שמורות לברק קנדל

שאלות:

1) חשבו : בכמה אופנים

א. אפשר לסדר 4 ספרים שונים על מדף?

ב. אפשר לסדר חמישה חיילים בטור?

2) סידרו באקראי 10 דיסקים שונים על מדף שמתוכם שניים בשפה העברית.

א. מה ההסתברות שהדיסקים בעברית יהיו צמודים זה לזה?

ב. מה ההסתברות שהדיסקים בעברית לא יהיו צמודים זה לזה?

ג. מה ההסתברות ששני הדיסקים בעברית יהיו כל אחד בקצה השני של המדף?

3) בוחנים 5 בנים ו-4 בנות בכיתה ומדרגים אותם לפי הציון שלהם בבחינה. נניח שאין תלמידים בעלי אותו ציון.

א. מהו מספר הדירוגים האפשריים?

ב. מהו מספר הדירוגים האפשריים אם מדרגים בנים ובנות בנפרד?

4) מסדרים 10 ספרים שונים על מדף.

א. בכמה אופנים ניתן לסדר את הספרים על המדף?

שני ספרים מתוך ה-10 הם ספרים בסטטיסטיקה.

ב. מה ההסתברות שאם נסדר את הספרים באקראי, הספרים בסטטיסטיקה יהיו צמודים זה לזה?

ג. מה ההסתברות שהספרים בסטטיסטיקה לא יהיו צמודים זה לזה?

ד. מה ההסתברות שהספרים בסטטיסטיקה יהיו בקצות המדף (כל ספר בקצה אחר)?

5) אדם יצר בנגן שלו פליליסט (רשימת השמעה) של 12 שירים שונים. 4 בשפה העברית, 5 באנגלית ו-3 בצרפתית. האדם הריץ את הפליליסט באקראי.

א. מה ההסתברות שכל השירים באנגלית יופיעו כשירים הראשונים כמקשה אחת?

ב. מה ההסתברות שכל השירים באנגלית יופיעו ברצף (לא חובה ראשונים)?

ג. מה ההסתברות ששירים באותה השפה יופיעו ברצף (כלומר כל השירים באנגלית ברצף, כל השירים בעברית ברצף וכך גם השירים בצרפתית)?

© כל הזכויות שמורות לברק קנדל

6) 4 בנים ו-4 בנות התיישבו באקראי בשורת כיסאות 1-8 בקולנוע.

א. מה ההסתברות שיוסי ומיכל לא ישבו זה לצד זה?

ב. מה ההסתברות שהבנים יתיישבו במקומות האי-זוגיים?

ג. מה ההסתברות שכל הבנים ישבו זה לצד זה?

ד. מה ההסתברות שהבנים ישבו זה לצד זה והבנות תשבנה זו לצד זו?

תשובות סופיות:

1) א. 24. ב. 120.

2) א. 0.2. ב. 0.8. ג. 0.022.

3) א. 362880. ב. 2880.

4) א. 3628800. ב. 0.2. ג. 0.8. ד. $\frac{1}{45}$.

5) א. $\frac{1}{792}$. ב. $\frac{1}{99}$. ג. $\frac{1}{4620}$.

6) א. 0.75. ב. 0.014. ג. $\frac{1}{14}$. ד. $\frac{1}{35}$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 5 - קומבינטוריקה -דגימה סידורית ללא החזרה ועם החזרה

תוכן העניינים

1. כללי 21

קומבינטוריקה – דגימה סידורית ללא החזרה ועם החזרה:

רקע:

מדגם סידור בדגימה עם החזרה:

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

דוגמה:

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

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

מדגם סידור ללא החזרה:

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

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

דוגמה:

שלושה תלמידים נבחרים מתוך 10 לייצג וועד בו תפקידים שונים.

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

© כל הזכויות שמורות לברק קנדל

שאלות:

1) במפלגה 20 חברי כנסת, מעוניינים לבחור שלושה חברי כנסת לשלושה תפקידים שונים.

א. חבר כנסת יכול למלא יותר מתפקיד אחד.
כמה קומבינציות ישנן לחלוקת התפקידים?

ב. חבר כנסת לא יכול למלא יותר מתפקיד אחד.
כמה קומבינציות יש לחלוקת התפקידים?

2) במשחק מזל יש 4 משבצות ממוספרות מ-A-D (A עד D). בכל משבצת יש למלא סיפרה (0-9). הזוכה הוא זה שניחש נכונה את כל הספרות בכל המשבצות בהתאמה.

א. מה ההסתברות לזכות במשחק?

ב. מה ההסתברות שבאף משבצת לא תהיה את הספרה 3 במספר הזוכה?

ג. מה ההסתברות שהתוצאה 4 תופיע לפחות פעם אחת במספר הזוכה?

3) קבוצה מונה 22 אנשים, מה ההסתברות שלפחות לשניים מהם יהיה יום הולדת באותו התאריך?

4) שלושה אנשים קבעו להיפגש במלון הילטון בסינגפור.
הבעיה היא שבסינגפור ישנם 5 מלונות הילטון.

א. מה ההסתברות שכל השלושה ייפגשו?

ב. מה ההסתברות שכל אחד יגיע לבית מלון אחר?

5) בכיתה 40 תלמידים. מעוניינים לבחור חמישה מהם לוועד כיתה.
בכמה דרכים ניתן להרכיב את הוועד אם:

א. בוועד 5 תפקידים שונים ותלמיד יכול למלא יותר מתפקיד אחד.

ב. בוועד 5 תפקידים שונים ותלמיד לא יכול למלא יותר מתפקיד אחד.

תשובות סופיות:

1) א. 8000. ב. 6840.

2) א. 0.0001. ב. 0.6561. ג. 0.3439.

3) 0.476.

4) א. 0.04. ב. 0.48.

5) א. 40^5. ב. 78,960,960.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 6 - קומבינטוריקה - דגימה ללא סדר וללא החזרה

תוכן העניינים

1. כללי 23

קומבינטוריקה – דגימה ללא סדר וללא החזרה:

רקע:

מדגם לא סדור בדגימה ללא החזרה:

מספר האפשרויות לדגום k עצמים שונים מתוך n עצמים שונים כאשר אין משמעות לסדר העצמים הנדגמים ואין החזרה: $\frac{n!}{(n-k)!k!}=\binom{n}{k}=\frac{(n)_k}{k!}$.

דוגמה:

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

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

הערות:

1) $\binom{n}{k}=\binom{n}{n-k}$.

2) $\binom{n}{n-1}=\binom{n}{1}=n$.

3) $\binom{n}{n}=\binom{n}{0}=1$.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) בכיתה 15 בנות ו-10 בנים. יש לבחור 5 תלמידים שונים מהכיתה לנציגות הכיתה. בכמה דרכים אפשר להרכיב את הנציגות, אם:

א. אין שום הגבלה לבחירה.
ב. מעוניינים ש-3 בנות ו-2 בנים ירכיבו את המשלחת.
ג. לא יהיו בנים במשלחת.

2) סטודנט מעוניין לבחור 5 קורסי בחירה בסמסטר זה. לפניו רשימה של 10 קורסים לבחירה: 5 במדעי הרוח, 3 במדעי החברה, 2 במתמטיקה.

א. כמה בחירות שונות הוא יכול ליצור לעצמו?
ב. כמה בחירות יש לו בהן 3 קורסים הם ממדעי הרוח?
ג. כמה בחירות יש לו אם 2 מהן לא ממדעי הרוח?
ד. כמה בחירות יש לו אם 2 ממדעי הרוח, 2 ממדעי החברה ו-1 ממתמטיקה?

3) בכיתה 30 תלמידים מתוכם 12 תלמידים ו-18 תלמידות.
יש לבחור למשלחת 4 תלמידים מהכיתה. התלמידים נבחרים באקראי.

א. מה ההסתברות שהמשלחת תורכב רק מבנות?
ב. מה ההסתברות שבמשלחת תהיה רק בת אחת?
ג. מה ההסתברות שבמשלחת תהיה לפחות בת אחת?

4) במשחק הלוטו יש לבחור 5 מספרים מתוך 45. המספרים הם 1-45.

א. מה ההסתברות שבמשחק הזוכה כל המספרים הם זוגיים?
ב. מה ההסתברות שבמספר הזוכה יש לכל היותר מספר זוגי אחד?
ג. מה ההסתברות שבמספר הזוכה לפחות פעם אחת יש מספר זוגי?
ד. מה ההסתברות שבמספר הזוכה כל המספרים גדולים מ-30?

5) בחפיסת קלפים ישנם 52 קלפים: 13 בצבע שחור בצורת עלה, 13 בצבע אדום בצורת לב, 13 בצבע אדום בצורת יהלום ו-13 בצבע שחור בצורת תלתן.
מכל צורה (מתוך ה-4) יש 9 קלפים שמספרם 2-10, שאר הקלפים הם; נסיך, מלכה, מלך ואס (בעצם מדובר בקופסת קלפים רגילה ללא ג'וקר).
שני אנשים משחקים פוקר. כל אחד מקבל באקראי 5 קלפים (ללא החזרה).

א. מה ההסתברות שעודד יקבל את כל המלכים וערן את כל המלכות?
ב. מה ההסתברות שאחד השחקנים יקבל את הקלף אס-לב?
ג. מה ההסתברות שערן יקבל קלפים שחורים בלבד ועודד יקבל שני קלפים שחורים בדיוק?
ד. מה ההסתברות שערן יקבל לפחות 3 קלפים שהם מספר (אס אינו מספר)?

© כל הזכויות שמורות לברק קנדל

6) במכללה 4 מסלולי לימוד. בכל מסלול לימוד 5 מזכירות. יש ליצור וועד של 5 מזכירות מתוך כלל המזכירות במכללה. יוצרים וועד באופן אקראי.
חשבו את ההסתברויות הבאות:

א. כל המזכירות בוועד יהיו ממסלול "מדעי ההתנהגות".
ב. כל המזכירות בוועד יהיו מאותו המסלול.
ג. מכל מסלול תבחר לפחות מזכירה אחת.

7) הוכיחו כי: $\binom{n}{k}+\binom{n}{k+1}=\binom{n+1}{k+1}$.

8) $2n$ בנים ו-$2n$ בנות מתחלקים ל-2 קבוצות.

א. בכמה דרכים שונות ניתן לבצע את החלוקה אם שתי הקבוצות צריכות להיות שוות בגודלן ויש בכל קבוצה מספר שווה של בנים ובנות?
ב. בכמה דרכים ניתן לבצע את החלוקה אם יש מספר שווה של בנים ובנות בכל קבוצה אבל הקבוצות לא בהכרח בגודל שווה.

תשובות סופיות:

1) א. 53130. ב. 20475. ג. 3003.

2) א. 252. ב. 100. ג. 100. ד. 60.

3) א. 0.1117. ב. 0.1445. ג. 0.9819.

4) א. 0.02. ב. 0.187. ג. 0.972. ד. 0.00246.

5) א. 0. ב. 0.1923. ג. 0.009. ד. 0.837.

6) א. $6.45\cdot10^{-5}$. ב. $2.58\cdot10^{-4}$. ג. 0.3225.

7) שאלת הוכחה.

8) א. $\binom{2n}{n}^2$. ב. $\sum_{i=1}^{n}\binom{2n}{i}^2$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 7 - קומבינטוריקה - דגימה ללא סדר ועם החזרה

תוכן העניינים

1. כללי 26

קומבינטוריקה – דגימה ללא סדר ועם החזרה:

רקע:

מספר האפשרויות לבחור k עצמים (לא בהכרח שונים) מתוך n עצמים שונים, ללא חשיבות לסדר העצמים הנדגמים, ועצם יכול להיבחר יותר מפעם אחת:

$$\binom{n+k-1}{k}=\binom{n+k-1}{n-1}.$$

דוגמה:

בכמה דרכים שונות ניתן לחלק 4 כדורים זהים לשלושה תאים שבכל תא יש מקום ליותר מכדור אחד? (פתרון והסבר הרעיון בהקלטה)

סיכום כללי של המצבים האפשריים לדגימה:

מספר האפשרויות לבחירת k עצמים מתוך אוכלוסייה של n עצמים שונים		
ביצוע הדגימה	**עם התחשבות בסדר הבחירה**	**ללא התחשבות בסדר הבחירה**
עם החזרה	n^k	$\binom{n+k-1}{k}=\binom{n+k-1}{n-1}$
ללא החזרה	$(n)_k=\frac{n!}{(n-k)!}$	$\binom{n}{k}=\frac{n!}{k!(n-k)!}$

© כל הזכויות שמורות לברק קנדל

שאלות:

1) בכמה דרכים יש להכניס 8 כדורים זהים לחמישה תאים כאשר תא יכול להכיל יותר מכדור אחד?

2) בכמה אופנים ניתן להכניס 5 מחברות זהות ל-3 תיקים שונים?

3) בכמה אופנים ניתן להכניס 8 כדורים לתוך 3 תאים שונים כאשר:

א. הכדורים זהים.

ב. הכדורים שונים זה מזה.

4) בכמה דרכים יש לסדר 10 משחקים ב-4 מגירות כאשר:

א. המשחקים שונים זה מזה.

ב. במשחקים זהים זה לזה.

5) מהו מספר הפתרונות השלמים האי שליליים למשוואה הבאה: $X_1 + X_2 = 3$.

6) מהו מספר הפתרונות השלמים האי-שליליים למשוואה הבאה:
$X_1 + X_2 + X_3 + X_4 = 20$.

7) במכירה פומבית הוצגו 4 פמוטי זהב זהים לחלוטין. על קניית היצירות התחרו 3 אספנים. אספן יכול היה לרכוש יותר מפמוט אחד. בהנחה וכל הפמוטים נמכרו, כמה אפשרויות מכירה לאספנים השונים ישנן?

8) נתונות האותיות: A, B, C ו-D. נרצה לבחור שתי אותיות מתוך קבוצת האותיות הללו כאשר מותר לבחור אותה אות יותר מפעם אחת אבל אין חשיבות לסדר האותיות שנבחרו. כמה דרכים ישנן לבחירה?

9) במשחק הלוטו החדש יש לבחור ארבעה מספרים מתוך המספרים 1-20.
אין חשיבות לסדר הפנימי של המספרים, אלא רק לגלות אילו מספרים עלו בגורל. מה הסיכוי לגלות את המספרים שעלו בגורל אם:

א. אסור לבחור את אותו מספר יותר מפעם אחת.

ב. מותר לחזור על אותו מספר יותר מפעם אחת.

© כל הזכויות שמורות לברק קנדל

10) ישנם 5 כדורים להכניס ל-6 תאים.
חשבו את מספר האפשרויות להכנסת הכדורים כאשר:

א. הכדורים שונים ותא יכול להכיל יותר מכדור אחד.
ב. הכדורים זהים ותא יכול להכיל יותר מכדור אחד.
ג. הכדורים שונים ותא לא יכול להכיל יותר מכדור אחד.
ד. הכדורים זהים ותא לא יכול להכיל יותר מכדור אחד.

11) ישנם k כדורים להכניס ל-n תאים $(n>k)$.
חשבו את מספר האפשרויות להכנסת הכדורים כאשר:

א. הכדורים שונים ותא יכול להכיל יותר מכדור אחד.
ב. הכדורים זהים ותא יכול להכיל יותר מכדור אחד.
ג. הכדורים שונים ותא לא יכול להכיל יותר מכדור אחד.
ד. הכדורים זהים ותא לא יכול להכיל יותר מכדור אחד.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) 495.

2) 21.

3) א. 45. ב. 6561.

4) א. 4^{10}. ב. 286.

5) 4.

6) 1771.

7) 15.

8) 10.

9) א. $\frac{1}{4845}$. ב. $\frac{1}{8855}$.

10) א. 7776 ב. 252. ג. 720. ד. 6.

11) א. n^k. ב. $\binom{n+k-1}{n-1}=\binom{n+k-1}{k}$. ג. $(n)_k=\frac{n!}{(n-k)!}$.

ד. $\binom{n}{k}=\frac{n!}{k!(n-k)!}$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 8 - קומבינטוריקה - שאלות מסכמות

תוכן העניינים

1. כללי 30

קומבינטוריקה – שאלות מסכמות:

שאלות:

1) בכיתה 40 תלמידים. מעוניינים לבחור חמישה מהם לוועד כיתה.
בכמה דרכים ניתן להרכיב את הוועד אם:
א. בוועד 5 תפקידים שונים ותלמיד יכול למלא יותר מתפקיד אחד.
ב. בוועד 5 תפקידים שונים ותלמיד לא יכול למלא יותר מתפקיד אחד.
ג. אין תפקידים שונים בוועד.

2) במשרד 30 עובדים, יש לבחור ארבעה עובדים למשלחת לחו"ל.
בכמה דרכים ניתן להרכיב את המשלחת?
א. במשלחת ארבע משימות שונות שיש למלא וכל עובד יכול למלא יותר ממשימה אחת.
ב. כמו בסעיף א' רק הפעם עובד לא יכול למלא יותר ממשימה אחת.
ג. מעוניינים לבחור ארבעה עובדים שונים למשלחת שבה לכולם אותו התפקיד.

3) מעוניינים להרכיב קוד סודי. הקוד מורכב מ-2 ספרות שונות ו-3 אותיות שונות באנגלית (26 אותיות אפשריות).
א. כמה קודים שונים ניתן להרכיב?
ב. כמה קודים שונים ניתן להרכיב אם הקוד מתחיל בספרה ונגמר בספרה?
ג. כמה קודים ניתן להרכיב אם הספרות חייבות להיות צמודות זו לזו?
ד. בכמה קודים הספרות לא מופיעות ברצף?

4) בארונית 4 מגירות. ילד התבקש על ידי אמו לסדר 6 משחקים בארונית.
הילד מכניס את המשחקים באקראי למגירות השונות.
כל מגירה יכולה להכיל את כל המשחקים יחד.
א. מה ההסתברות שהילד יכניס את כל המשחקים למגירה העליונה?
ב. מה ההסתברות שהילד יכניס את כל המשחקים לאותה מגירה?
ג. מה ההסתברות ש"דומינו" יוכנס למגירה העליונה ויתר המשחקים לשאר המגירות.
ד. מה ההסתברות ש"דומינו" לא יוכנס למגירה העליונה?

© כל הזכויות שמורות לברק קנדל

5) בעיר מסוימת מתמודדות למועצת העיר 4 מפלגות שונות: "הירוקים", "קדימה", "העבודה" ו"הליכוד". 6 אנשים אינם יודעים למי להצביע, ולכן בוחרים באקראי מפלגה כלשהי.

א. מה ההסתברות שכל ה-6 יבחרו באותה מפלגה?
ב. מה ההסתברות שמפלגת ה"ירוקים" לא תקבל קולות?
ג. מה ההסתברות שמפלגת ה"ירוקים" תקבל בדיוק 3 קולות וכל מפלגה אחרת תקבל קול 1 בלבד?
ד. מה ההסתברות שמפלגת "הירוקים תקבל 2 קולות, מפלגת "העבודה" תקבל 2 קולות ומפלגת "הליכוד" תקבל 2 קולות?

6) 5 חברים נפגשו ורצו לראות סרט. לרשותם ספריה המונה 8 סרטים שונים. כל אחד התבקש לבחור סרט באקראי.

א. מה ההסתברות שכולם יבחרו את אותו הסרט?
ב. מה ההסתברות שכולם יבחרו את "הנוסע השמיני"?
ג. מה ההסתברות שכל אחד יבחר סרט אחר?
ד. מה הסיכוי שלפחות שניים יבחרו את אותו הסרט?
ה. מה ההסתברות שיוסי וערן ייבחרו את "הנוסע השמיני" וכל השאר סרטים אחרים?
ו. מה ההסתברות שהנוסע השמיני לא ייבחר על ידי אף אחד מהחברים?
ז. לקחו את 8 הסרטים ויצרו מהם רשימה. נתון שברשימה 3 סרטי אימה, מה ההסתברות שברשימה שנוצרה יופיעו 3 סרטי האימה ברצף?

7) בקבוצה 10 אנשים. יש ליצור שתי וועדות שונות מתוך הקבוצה: אחת בת 4 אנשים והשנייה בת 3 אנשים. כל אדם יכול להיבחר רק לוועדה אחת.
חשבו את מס' הדרכים השונות ליצירת הוועדות הללו כאשר:

א. אין בוועדות תפקידים.
ב. בכל וועדה יש תפקיד אחד של אחראי הוועדה.
ג. בכל וועדה כל התפקידים שונים.

8) 4 גברים ו-3 נשים מתיישבים על כסאות בשורה של כסאות תיאטרון. בכל שורה 10 כסאות. בכמה דרכים שונות ניתן לבצע את ההושבה:

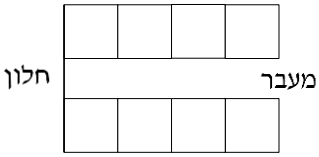
א. ללא הגבלה.
ב. כל הגברים ישבו זה ליד זה וגם כל הנשים תשבנה זו ליד זו.
ג. שני גברים בקצה אחד ושני הגברים האחרים בקצה שני.

9) בהגרלה ישנם 10 מספרים מ-1 עד 10. נבחרו באקראי 5 מספרים.
מה ההסתברות שהמספר 7 הוא השני בגודלו מבין המספרים שנבחרו?

© כל הזכויות שמורות לברק קנדל

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

11) ברכבת 4 מקומות ישיבה עם כיוון הנסיעה ו4 מקומות ישיבה נגד כיוון הנסיעה.
4 זוגות התיישבו במקומות אלו באקראי.



א. בכמה דרכים שונות ניתן להתיישב?
ב. מה ההסתברות שהזוג כהן ישבו זה לצד זה עם כיוון הנסיעה?
ג. מה ההסתברות שהזוג כהן ישבו זה לצד זה?
ד. מה ההסתברות שהזוג כהן ישבו כל אחד ליד החלון? (בכל שורה יש חלון).
ה. מה ההסתברות שהזוג כהן יישבו כך שכל אחד בכיוון נסיעה מנוגד?
ו. מה ההסתברות שהזוג כהן יישבו אחד מול השני פנים מול פנים.
ז. מה ההסתברות שכל הגברים ייסעו עם כיוון הנסיעה וכל הנשים תשבנה נגד כיוון הנסיעה?
ח. מה ההסתברות שכל זוג ישב אחד מול השני?

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

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

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

© כל הזכויות שמורות לברק קנדל

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

16) זוג קוביות הוטלו מספר פעמים. כמה פעמים יש להטיל את זוג הקוביות בכדי שבהסתברות של לפחות 0.5 תתקבל לפחות הטלה אחת (של הזוג) עם סכום תוצאות 12?

17) בוחרים באופן מקרי מספר בין 6 ספרות.
א. מה הסיכוי שהספרה 5 תופיע בדיוק פעם אחת במספר?
ב. מה הסיכוי שהספרה 4 תופיע לפחות פעם אחת וגם הספרה 0 תופיע לפחות פעם אחת במספר?

18) במשרד של דנה 5 תיקיות אותן היא מסדרת באקראי בטור. 3 תיקיות הן אדומות ו-2 תיקיות הן כחולות. דנה רשמה שני פתקים ושמה כל פתק במקום אקראי בין התיקיות (לכל פתק יש 4 אפשרויות למיקום).
א. מה הסיכוי ששני הפתקים יהיו במקומות שונים?
ב. מה הסיכוי שבין שני הפתקים יש שתי תיקיות אדומות ואין תיקיות כחולות?
ג. מה הסיכוי שבין שני הפתקים יש בדיוק תיקיה אחת?
ד. מה הסיכוי שבין שני הפתקים יש שתי תיקיות ואחת מהן כחולה?

19) לירון 6 עטים אותם הוא מכניס באקראי ל-3 קלמרים שונים.
לכל עט הוא בוחר באופן מקרי קלמר.
א. מה הסיכוי שיש בדיוק 2 קלמרים שבכל קלמר בדיוק 2 עטים?
ב. מה הסיכוי שיש בדיוק קלמר אחד שבו בדיוק 2 עטים?
ג. מה הסיכוי שיש בדיוק 3 קלמרים שבכל אחד בדיוק 2 עטים?

20) מסדרים n כדורים שונים ב n תאים שונים (תא יכול להכיל יותר מכדור אחד). מה הסיכוי שבתא i $(1 \le i \le n)$ יהיו בדיוק k כדורים?

21) בתחרות ריצה עלו לגמר 6 מתמודדים. רק בשלושת המקומות הראשונים זוכים במדליות. נניח שכל המתמודדים מסיימים את התחרות.
א. כמה אפשרויות יש לסיים את התחרות?
ב. כמה אפשרויות יש לכך שמתמודד מספר 6 יקבל מדליה?
ג. כמה אפשרויות יש לכך שמתמודד מספר 6 יקבל מדליה או שמתמודד מספר 2 יקבל מדליית זהב?

© כל הזכויות שמורות לברק קנדל

22) מטילים קובייה הוגנת k פעמים.

א. מה הסיכוי שהתוצאה הכי גדולה שהתקבלה היא j ?

ב. מה הסיכוי שהתוצאה הכי קטנה שהתקבלה היא i ?

ג. עבור $i \le j$, מה הסיכוי שהתוצאה הכי גדולה היא j וגם התוצאה הכי קטנה היא i ?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. 102,400,000. ב. 78,960,960. ג. 658008.

2) א. 810,000. ב. 657,720. ג. 27,405.

3) א. 14,040,000. ב. 1,404,000. ג. 5,616,000. ד. 8,424,000.

4) א. 0.00024. ב. 0.00098. ג. 0.05933. ד. 0.75000.

5) א. 0.00098. ב. 0.17798. ג. 0.02929. ד. 0.02197.

6) א. $\frac{1}{4096}$. ב. $\frac{1}{32,768}$. ג. 0.205. ד. 0.795
ה. 0.0105. ו. 0.5129. ז. 0.1071.

7) א. 4,200. ב. 50,400. ג. 604,800.

8) א. 604,800. ב. 2,880. ג. 2,880.

9) 0.238.

10) א. 0.1512. ב. 0.014. ג. 0.059. ד. $\frac{62}{10^6}$.

11) א. 40,320. ב. 0.1071. ג. 0.2142. ד. 0.0357.
ה. 0.5714. ו. 0.1429. ז. 0.0143. ח. 0.0095.

12) א. 60,466,176. ב. 45,239,040. ג. 48,484,800.

13) א. $\frac{n!}{3!(n-3)}$. ב. $n\cdot(n-1)(n-2)$. ג. n^3.

14) $\frac{1}{4^n}\cdot\sum_{i=0}^{n}\binom{n}{i}^2$.

15) א. 0.9^a. ב. $1-0.9^a$. ג. 0.5^a.

16) לפחות 25 פעמים.

17) א. 0.35721. ב. 0.1759.

18) א. 0.75. ב. 0.075. ג. 0.375. ד. 0.15.

19) א. 0. ב. $\frac{450}{729}$. ג. $\frac{90}{729}$.

20) $\frac{\binom{n}{k}(n-1)^{n-k}}{n^n}$.

21) א. 720. ב. 360. ג. 432.

© כל הזכויות שמורות לברק קנדל

22) א. $\frac{j^k-(j-1)^k}{6^k}$. ב. $\frac{(7-i)^k-(6-i)^k}{6^k}$.

ג. $\frac{(j-i+1)^k-2\cdot(j-i)^k+(j-i-1)^k}{6^k}$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 9 - הסתברות מותנית-במרחב מדגם אחיד

תוכן העניינים

1. כללי 37

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: $P(A|B)$.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) נבחרה ספרה זוגית באקראי. מה הסיכוי שהספרה גדולה מ-6?

2) יוסי הטיל קובייה.
מה הסיכוי שקיבל את התוצאה 4, אם ידוע שהתוצאה שהתקבלה זוגית?

3) הוטלו צמד קוביות. נגדיר:
A - סכום התוצאות בשתי ההטלות הינו 7.
B - מכפלת התוצאות 12.
חשבו את $P(A|B)$.

4) מטבע הוטל פעמיים.
ידוע שהתקבל לכל היותר ראש אחד, מה הסיכוי שהתקבלו שני ראשים?

5) זוג קוביות הוטלו והתקבל שהתוצאות זהות.
מה הסיכוי שלפחות אחת התוצאות 5?

6) זוג קוביות הוטלו והתקבל לפחות פעם אחת 4.
מה הסיכוי שאחת התוצאות 5?

7) נבחרה משפחה בת שני ילדים, שמהם אחד הוא בן.
מה ההסתברות שבמשפחה שני בנים בקרב הילדים?

8) נבחרה משפחה בת שלושה ילדים, ונתון שהילד האמצעי בן.
מה הסיכוי שיש בנות בקרב הילדים?

9) בכיתה 6 בנים ו-7 בנות. נבחרו 4 ילדים מהכיתה.
אם ידוע שנבחרו 2 בנים ו-2 בנות, מה הסיכוי שאלעד לא נבחר?

10) חמישה חברים יצאו לבית קולנוע והתיישבו זה לצד זה באקראי,
בכיסאות מספר 5 עד 9. ידוע שערן ודין התיישבו זה ליד זה.
מה ההסתברות שהם יושבים בכיסאות מספר 6 ו-7?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) 0.2.

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

3) 0.5.

4) 0.

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

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 10 - הסתברות מותנית - מרחב לא אחיד

תוכן העניינים

1. כללי 40

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?

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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

א. בנו טבלת שכיחות משותפת לנתונים.

ב. נבחר אדם אקראי מה ההסתברות שהוא שכיר?

ג. מה ההסתברות שהוא שכיר ולא אקדמאי?

ד. מה ההסתברות שהוא שכיר או אקדמאי?

ה. אם האדם שנבחר הוא עצמאי מהי ההסתברות שהוא אקדמאי?

ו. אם האדם שנבחר הוא לא אקדמאי, מה ההסתברות שהוא שכיר?

5) חברה מסוימת פרסמה את הנתונים הבאים לגבי האזרחים מעל גיל 21:
40% מהאנשים מחזיקים כרטיס "ויזה", 52% מחזיקים כרטיס "ישראכרט", 20% מחזיקים כרטיס "אמריקן אקספרס", 15% מחזיקים ויזה וגם ישראכרט, 8% מחזיקים ישראכרט וגם אמריקן אקספרס ו-7% מחזיקים כרטיס ויזה וגם אמריקן אקספרס. כמו כן, 5% מחזיקים בשלושת הכרטיסים הנ"ל.

א. אם לאדם יש ויזה, מה הסיכוי שאין לו ישראכרט?

ב. אם לאדם שני כרטיסי אשראי, מה הסיכוי שאין לו ישראכרט?

ג. אם לאדם לפחות כרטיס אחד, מה הסיכוי שאין לו ישראכרט?

© כל הזכויות שמורות לברק קנדל

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.

5) א. 0.625. ב. 0.133. ג. 0.402.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 11 - דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה

תוכן העניינים

1. כללי 44

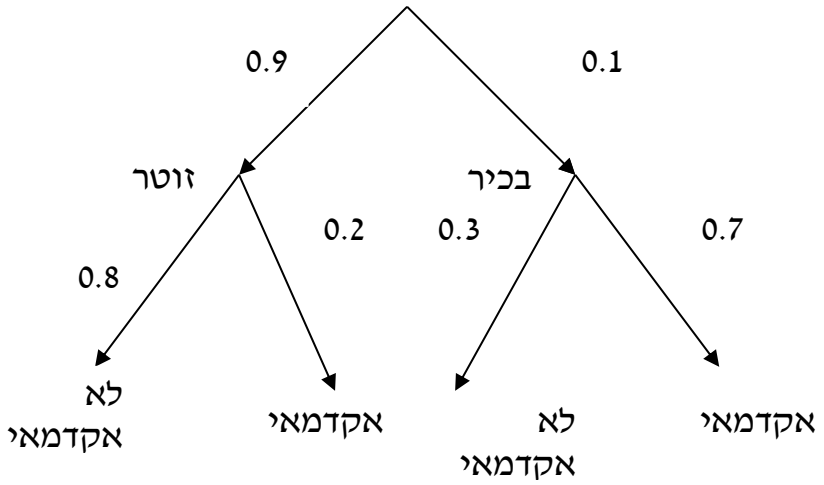
דיאגרמת עצים – נוסחת בייס ונוסחת ההסתברות השלמה:

רקע:

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

דוגמה:

בחברה מסוימת 10% מוגדרים בכירים והיתר מוגדרים זוטרים. מבין הבכירים 70% הם אקדמאים ומבין הזוטרים 20% הם אקדמאים. נשרטט עץ שיתאר את הנתונים, השלב הראשון של העץ אינו מותנה בכלום ואילו השלב השני מותנה בשלב הראשון.



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

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

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

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

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

4) נבחר אקדמאי מה ההסתברות שהוא עובד זוטר?
מדובר כאן על שאלה בהסתברות מותנה ולכן נשתמש בעיקרון של הסתברות מותנה: $P(zutar\,|\,academay)=\frac{0.9\cdot 0.2}{0.25}=\frac{0.18}{0.25}=0.72$.

© כל הזכויות שמורות לברק קנדל

נוסחת ההסתברות השלמה:

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

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

נוסחת בייס:

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

ב. מה ההסתברות שהסוכריה השנייה בטעם לימון?

2) באוכלוסייה מסוימת 30% הם ילדים, 50% בוגרים והיתר קשישים. לפי נתוני משרד הבריאות הסיכוי שילד יחלה בשפעת במשך החורף הוא 80%, הסיכוי שמבוגר יחלה בשפעת במשך החורף הוא 40% והסיכוי שקשיש יחלה בשפעת במשך החורף הוא 70%.

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

ב. מה אחוז האנשים שיחלו בשפעת במשך החורף?

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

ד. נבחר ילד, מה ההסתברות שהוא לא יחלה בשפעת במשך החורף?

3) בכד א' 5 כדורים כחולים ו-5 כדורים אדומים. בכד ב' 6 כדורים כחולים ו-4 כדורים אדומים. בוחרים באקראי כד, מוציאים ממנו כדור ומבלי להחזירו מוציאים כדור נוסף.

א. מה ההסתברות ששני הכדורים שיוצאו יהיו בצבעים שונים?

ב. אם הכדורים שהוצאו הם בצבעים שונים, מה ההסתברות שהכדור השני שהוצא יהיה בצבע אדום?

4) חברת סלולר מסווגת את לקוחותיה לפי 3 קבוצות גיל: נוער, בוגרים ופנסיונרים. נתון כי: 10% מהלקוחות בני נוער, 70% מהלקוחות בוגרים והיתר פנסיונרים. מתוך בני הנוער 90% מחזיקים בסמארט-פון, מתוך האוכלוסייה הבוגרת ל-70% יש סמארט-פון ומתוך אוכלוסיית הפנסיונרים 30% מחזיקים בסמארט-פון.

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

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

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

© כל הזכויות שמורות לברק קנדל

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 ש68% 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?

© כל הזכויות שמורות לברק קנדל

9) סטודנט ניגש למבחן אמריקאי. הסיכוי שהוא יודע תשובה לשאלה מסוימת הוא P, ואם הוא לא יודע את התשובה הוא מנחש. בכל מקרה הוא עונה על השאלה. נתון שלשאלה יש k תשובות אפשריות.
אם הסטודנט ענה נכון על השאלה, מה הסיכוי שהוא ידע אותה?

10) אדם משחק נגד שני מתמודדים, רונית ודולב. האדם צריך לשחק שלושה משחקים ויש לו לבחור איזה סדר משחקים עדיף לו:
- דולב, רונית, דולב.
- רונית, דולב, רונית.

בכל משחק מישהו חייב לנצח(אין תיקו). האדם ינצח בטורניר רק אם ינצח בשני משחקים ברציפות. נתון שדולב שחקן טוב יותר מאשר רונית.
איזו אפשרות עדיפה יותר על האדם כדי לנצח בטורניר?

תשובות סופיות:

1) א. $\frac{2}{7}$. ב. $\frac{23}{49}$.

2) א. 6%. ב. 58%. ג. 0.241. ד. 0.2.

3) א. 0.544. ב. 0.5.

4) א. 9%. ב. 0.09375. ג. 0.9722.

5) א. 0.14. ב. 0.3488. ג. 0.2442.

6) א. 70%. ב. 0.8125.

7) א. 0.57. ב. 0.3158. ג. 0.7543.

8) א. 0.0886. ב. 0.2889. ג. 0.3137. ד. 0.8778.

9) $\frac{kp}{1+p(k-1)}$.

10) א'.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 12 - תלות ואי תלות בין מאורעות

תוכן העניינים

1. אי תלות בין מאורעות (מורחב) 49

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: $P\left(\bigcap_{i=1}^{n}A_i\right)=\prod_{i=1}^{n}P(A_i)$.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) נתון: $P(A\cup B)=0.6$, $P(B)=0.5$, $P(A)=0.2$.
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 שיווקו, 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:
$P(A\cup B)=0.9$, $P(A\cap \overline{B})=0.3$, $P(A|B)=0.6$.
האם A ו- B מאורעות בלתי תלויים?

7) הוכיח שאם: $P(A/B)=P(B/A)$, אז: $P(A)=P(B)$.

© כל הזכויות שמורות לברק קנדל

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 חיובים לכן הם מאורעות תלויים.

ד. A ו- B מאורעות תלויים שסיכוייהם חיובים לכן A ו- B מאורעות זרים.

ה. $\overline{P}(\overline{A}\cap\overline{B})=1-P(A)-P(B)$ 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 למחלץ 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, $0<p<1$.

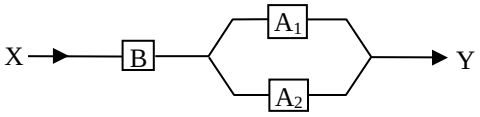
A – יצא עץ בהטלה ראשונה.

B – יצאו תוצאות שונות.

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

11) הוכח אם A ו-B בלתי תלויים, אזי $\overline{A}$, $\overline{B}$ 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?

© כל הזכויות שמורות לברק קנדל

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 בניסוי הוא: $\frac{P(A)}{P(A)+P(B)}$.

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 שתקבל 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. $\frac{1}{2}$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 13 - שאלות מסכמות בהסתברות

תוכן העניינים

1. כללי 53

שאלות מסכמות בהסתברות:

שאלות:

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

א. מה ההסתברות שמשפחה אקראית בת שתי מכוניות תהיה ללא מכוניות מתוצרת אירופה?

ב. מה ההסתברות שלפחות מכונית אחת תהיה אירופאית?

ג. ידוע שלמשפחה יש מכונית אירופאית.
מה ההסתברות שרק המכונית החדשה שלה היא מתוצרת אירופאית?

ד. אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית?

2) במדינת "שומקום" 50% מהחלב במרכולים מיוצר במחלבה א', 40% במחלבה ב' והיתר במחלבה ג'. 3% מתוצרת מחלבה א' מגיעה חמוצה למרכולים ואילו במחלבה ב' 10%.
כמו כן ידוע שבמדינת "שומקום" בסך הכול 7.5% מהחלב חמוץ.

א. איזה אחוז מהחלב שמגיע למרכול ממחלבה ג' חמוץ?

ב. אם נרכש חלב חמוץ במרכול. מה הסיכוי שהוא יוצר במחלבה ג'?

ג. ברכישת חלב נמצא שהוא אינו חמוץ. מה הסיכוי שהוא יוצר במחלבה א'?

ד. האם המאורעות: "חלב חמוץ" ו-"יוצר במחלבה א'" בלתי תלויים?

3) רוני ורונה יצאו לבלות במרכז בילוים עם מספר אפשרויות בילוי:
בהסתברות של 0.3 הם ייצאו לבאולינג, בהסתברות של 0.5 הם ייצאו לבית קפה ובהסתברות של 0.7 הם יצאו לפחות לאחד מהם (באולינג/קפה).

א. מה ההסתברות שהם יצאו רק לבאולינג?

ב. האם המאורעות "לצאת לבאולינג" ו-"לצאת לבית קפה" זרים?

ג. האם המאורעות "לצאת לבאולינג" ו-"לצאת לבית קפה" תלויים?

ד. מה ההסתברות שיום אחד הם יצאו רק לבאולינג וביום למחרת לא יצאו לאף אחד מהמקומות?

© כל הזכויות שמורות לברק קנדל

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.

א. מה הסיכוי שסטודנט אקראי עבר את הקורס?

ב. אם סטודנט אקראי עבר הקורס, מה הסיכוי שעבר במועד ב'?

ג. מה אחוז הסטודנטים שפורשים מהתואר?

ד. 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.

א. מה אחוז האבטלה בקרב הנשים?

ב. נבחר אדם מובטל, מה ההסתברות שזו אישה?

ג. 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.

א. חשבו את הסיכויים למאורעות A ו-B.

ב. האם המאורע A ו-B בלתי תלויים?

ג. 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.

א. מה אחוז האנשים, מאלה שמעוניינים לרכוש רכב משומש, שיראו את 2 המודעות?

ב. אם אדם ראה את המודעה באינטרנט, מה ההסתברות שהוא לא ראה את המודעה בעיתון?

ג. האם המאורעות: "לראות את המודעה באינטרנט" ו-"לראות את המודעה בעיתון" בלתי תלויים?

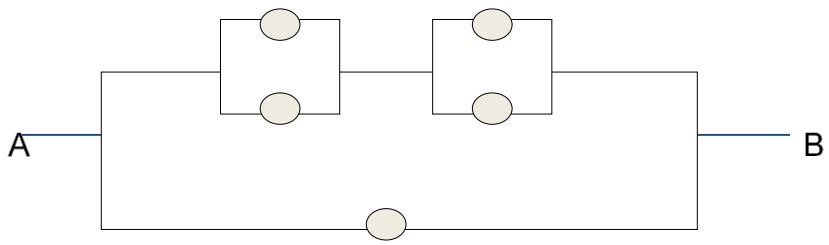
ד. אדם שראה את המודעה באינטרנט בלבד יתקשר לערן בהסתברות של 0.7, אם הוא ראה את המודעה בעיתון בלבד הוא יתקשר לערן בהסתברות של 0.6 ואם הוא ראה את שתי המודעות הוא יתקשר לערן בהסתברות של 0.9.

i. מה ההסתברות שאדם המעוניין לרכוש רכב משומש יתקשר לערן?

ii. אדם המעוניין לרכוש רכב משומש התקשר לערן. מה ההסתברות שהוא ראה את שתי המודעות?

© כל הזכויות שמורות לברק קנדל

8) נתונה המערכת החשמלית הבאה:



כל יחידה עובדת באופן בלתי תלוי ובהסתברות p.

כדי שהמערכת תפעל צריך לעבור זרם מהנקודה A לנקודה B.

הוכיחו שהסיכוי שהמערכת תפעל הוא: $P+(1-P)\left(2P-P^2\right)^2$.

9) ליאת מעוניינת לתרגל לבחינה בהסתברות. היא מצאה באינטרנט מאגר הכולל 25 שאלות מבחינות. השאלות ממוספרות ו-6 מתוכן עוסקות במשתנה מקרי רציף. ליאת החליטה לבחור באקראי 7 שאלות מהמאגר במטרה לפתור אותן. כל שאלה שלא עוסקת במשתנה הרציף תיפתר על ידי מיכל בסיכוי של 90%, אך אם השאלה עוסקת במשתנה הרציף היא תיפתר בסיכוי של 60%.

א. מה הסיכוי שהשאלות שנבחרו הן כולן ממוספרות בסדר עוקב?

ב. מה הסיכוי ששאלה 20 היא השאלה עם המספר המקסימלי מבין השאלות שנבחרו?

ג. ידוע שליאת בחרה 2 שאלות שעוסקות במשתנה הרציף והיתר לא. מה הסיכוי שתצליח לפתור 6 מתוך השאלות שבחרה?

10) נתונים שלושה מאורעות: A, B ו-C. ידוע ש: $P(A|B)=1$, $P(A|C)=1$.

תנו דוגמא ספציפית למאורעות: A, B ו-C שבה המאורעות B ו-C תלויים.

11) הוכיחו או הפריכו (על ידי דוגמה נגדית) את הטענה הבאה:

אם A ו-B בלתי תלויים, אז A ו- $\overline{B}$ בלתי תלויים.

12) משחקים משחק מזל פעמיים, כך שבכל משחק בודד יש אפשרות לנצח או להפסיד. הסיכוי לנצח בכל משחק הוא P כאשר: $0<P<1$.

נגדיר את המאורעות הבאים:

A - תוצאות המשחקים שונות זו מזו.

B - המשחק הראשון היה ניצחון.

מה ערכו של P, עבורו A ו-B יהיו מאורעות בלתי תלויים?

© כל הזכויות שמורות לברק קנדל

13) טל מניח בשורה N קוביות בצבעים שונים. בין שתי קוביות אקראיות כלשהן ערן מניח מכחול. הוכיחו שהסיכוי שהקובייה הכחולה והאדומה יהיו בצדדים שונים של המכחול הוא: $\frac{N+1}{3(N-1)}$.

14) הוכיחו באמצעות אינדוקציה את אי שוויון בול: $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$.

© כל הזכויות שמורות לברק קנדל

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.

8) שאלת הוכחה.

9) א. $\frac{19}{480,700}$. ב. $\frac{27,132}{480,700}$. ג. 0.4015.

10) ראו סרטון.

11) שאלת הוכחה.

12) $\frac{1}{2}$.

13) שאלת הוכחה.

14) שאלת הוכחה.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 14 - המשתנה המקרי הבדיד - פונקציית ההסתברות

תוכן העניינים

1. כללי 58

המשתנה המקרי הבדיד – פונקציית ההסתברות:

רקע:

משתנה מקרי בדיד:

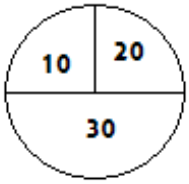
משתנה מקרי בדיד הינו משתנה היכול לקבל כמה ערכים בודדים בהסתברויות שונות.
מתארים את המשתנה המקרי על ידי פונקציית הסתברות.

פונקציית הסתברות:

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

דוגמה (פתרון בהקלטה):

בקזינו יש רולטה כמתואר בשרטוט:



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

© כל הזכויות שמורות לברק קנדל

שאלות:

1) ידוע שביישוב מסוים התפלגות מספר המכוניות למשפחה היא:
50 משפחות אינן מחזיקות במכונית.
70 משפחות עם מכונית אחת.
60 משפחות עם 2 מכוניות.
20 משפחות עם 3 מכוניות .
בוחרים באקראי משפחה מהיישוב, נגדיר את X להיות מספר המכוניות של המשפחה שנבחרה. בנו את פונקציית ההסתברות של X.

2) מהאותיות: A , B , C יוצרים קוד דו תווי.
א. כמה קודים ניתן ליצור?
ב. רשמו את כל הקודים האפשריים.
ג. נגדיר את X להיות מספר הפעמים שהאות B מופיעה בקוד.
בנו את פונקציית ההסתברות של X.

3) תלמיד ניגש בסמסטר לשני מבחנים: מבחן בכלכלה ומבחן בסטטיסטיקה.
כמו כן, נתון שהסיכוי לעבור את המבחן בכלכלה הנו 0.8, הסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9 והסיכוי לעבור את שני המבחנים הנו 0.75.
יהי X מספר המבחנים שהסטודנט עבר. בנו את פונקציית ההסתברות של X.

4) הסיכוי לזכות במשחק מסוים הינו 0.3. אדם משחק את המשחק עד אשר הוא מנצח אך בכל מקרה הוא לא משחק את המשחק יותר מ-4 פעמים.
נגדיר את X להיות מספר הפעמים שהוא שיחק את המשחק.
בנו את פונקציית ההסתברות של X.

5) חברה לניהול פרויקטים מנהלת 3 פרויקטים במקביל. הסיכוי שפרויקט א' יצליח הינו 0.7, הסיכוי שפרויקט ב' יצליח הינו 0.8, והסיכוי שפרויקט ג' יצליח הינו 0.9. נתון שהצלחת כל פרויקט בלתי תלויה זו בזו. נגדיר את X להיות מספר הפרויקטים שיצליחו. בנו את פונקציית ההסתברות של X.

6) להלן פונקציית הסתברות של משתנה מקרי כלשהו: $k=1,2...4$, $P(X=k)=\frac{k}{A}$.
מצאו את ערכו של A .

© כל הזכויות שמורות לברק קנדל

7) בגן ילדים 8 ילדים, מתוכם 5 בנים ו-3 בנות.
בוחרים באקראי 3 ילדים להשתתף בהצגה.
נגדיר את X כמספר הבנים שנבחרו להצגה.
בנו את פונקציית ההסתברות של X.

8) בסקר שנערך בדקו בקרב אנשים האם הם צופים במהדורת החדשות של ערוצים 1,2,10. להלן הנתונים:
20% צופים בערוץ 2.
8% צופים בערוץ 1.
10% צופים בערוץ 10.
כמו כן נתון ש 1% צופים בשלושת המהדורות גם יחד.
10% צופים בשתי המהדורות מתוך השלושה.
נגדיר את X להיות מספר המהדורות מבין 3 המהדורות המדוברות שאדם אקראי צופה. בנו את פונקציות ההסתברות של X.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) להלן טבלה:

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

2) להלן טבלה:

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

3) להלן טבלה:

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

4) להלן טבלה:

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

5) להלן טבלה:

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

6) 10.

7) להלן טבלה:

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

8) להלן טבלה:

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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 15 - המשתנה המקרי הבדיד - תוחלת - שונות וסטיית תקן

תוכן העניינים

1. כללי 62

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.

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.

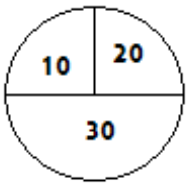
מגדירים שונות באופן הבא: $V(X)=\sum_i (x_i-\mu)^2 P(x_i)=\sum_i x_i^2 P(x_i)-\mu^2=\sigma^2$.

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-₪. הסתברות לקבלת הסכומים השונים:



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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

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 וסטיית 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.

© כל הזכויות שמורות לברק קנדל

5) נתונה פונקציית ההסתברות של המשתנה המקרי X:

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

כמו כן נתון ש: $E(X)=4.2$.

א. מצאו את ההסתברויות החסרות בטבלה.

ב. חשבו את: $V(X)$.

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

7) להלן ההתפלגות של משתנה מקרי:

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

מהו הערך שייתן ערך מינימלי לשונות של X?

© כל הזכויות שמורות לברק קנדל

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.

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

5) א. ראו טבלה: ב. 5.16.

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

6) ראו טבלה:

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

7) 2.33.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 16 - המשתנה המקרי הבדיד -תוחלת של פונקציה של משתנה מקרי בדיד

תוכן העניינים

1. ראשי 66

תוחלת של פונקציה של משתנה מקרי בדיד:

רקע:

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

$$E(g(X)) = g(x_1)P(X=x_1) + g(x_2)P(X=x_2) + g(x_3)P(X=x_3) + \cdots$$
$$= \sum_i g(x_i)\cdot P(x_i)$$

כאשר $x_1, x_2, x_3, \ldots$ 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: $Y = X^2$.

© כל הזכויות שמורות לברק קנדל

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

חשבו את התוחלת של:

א. X^2.

ב. 2^X.

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

א. מצאו את ערכו של A.

ב. חשבו את: $E\left(\left[X-E(X)\right]^2\right)$.

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 $E(W^3)$.

© כל הזכויות שמורות לברק קנדל

5) יהי X משתנה מקרי בדיד עם תוחלת ושונות 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: $E\left(\sqrt{X+2}\right)$.

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) $E\left(\sqrt{X+2}\right)=1.4659$.
8) א. $X\sim B(10,0.5)$. ב. 2.5^{10}.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 17 - המשתנה המקרי הבדיד- טרנספורמציה לינארית

תוכן העניינים

1. כללי 69

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:

1) $E(Y)=aE(X)+b$.
2) $V(Y)=a^2\cdot V(X)$.
3) $\sigma_Y=|a|\sigma_x$.

שלבי העבודה:

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

דוגמה – הרולטה:

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

פתרון (בהקלטה):

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

© כל הזכויות שמורות לברק קנדל

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?

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

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

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

א. השלימו את פונקציית ההסתברות.

ב. חשבו את השונות מספר התשובות הנכונות בבחינה.

ג. הציון בבחינה מחושב באופן הבא:
כל שאלה נכונה מזכה ב-10 נקודות. לכל שאלה שגויה, מופחתת נקודה.
מהי התוחלת ומה השונות של הציון בבחינה?

© כל הזכויות שמורות לברק קנדל

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

א. מצא את ערכו של A.

ב. חשב את התוחלת והשונות של המשתנה הנחקר.

ג. חשב את: $E(X^3)$.

ד. חשב את התוחלת והשונות של המשתנה הבא: $\frac{X}{2}-4$.

תשובות סופיות:

1) תוחלת: 14, שונות: 32.

2) תוחלת: 8, שונות: 12.

3) תוחלת: 13.2, סטיית תקן: 5.5.

4) תוחלת: 3, שונות: 3.

5) א. להלן טבלה: ב. תוחלת: 2350, שונות: $85,727.5^2$.

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

ג. תוחלת: 1650, שונות: $85,727.5^2$.

6) ב. $V(X)=1.8275$.

7) א. $A=10$. ב. $E(X)=3$, $V(X)=1$. ג. $E(X^3)=35.4$, $V(X^3)=616.84$.

ד. $E(Y)=-2.5$, $V(Y)=0.25$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 18 - תוחלת ושונות של סכום משתנים מקריים

תוכן העניינים

1. כללי 72

תוחלת ושונות של סכום משתנים מקריים:

רקע:

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

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

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

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

דוגמה:

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

© כל הזכויות שמורות לברק קנדל

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:

$$\sigma(X)=3 \quad , \quad E(x)=10$$
$$\sigma(Y)=4 \quad , \quad E(y)=12$$

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:

$$K=2,3,4,5 \ , \ P(X=K)=\frac{A}{K-1}$$
$$\text{אחר} \quad 0$$

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.

© כל הזכויות שמורות לברק קנדל

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: $1.1136n$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 19 - התפלגויות בדידות מיוחדות -התפלגות בינומית

תוכן העניינים

1. כללי 75

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:

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

כאשר: $0!=1$; $n!=n\cdot(n-1)\cdot(n-2)\cdot....\cdot 1$; $\binom{n}{k}=\frac{n!}{k!(n-k)!}$

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

תוחלת: $E(X)=np$

שונות: $V(X)=npq$

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

1) חוזרים על <u>אותו ניסוי ברנולי</u> באופן <u>בלתי תלוי</u> זה בזה.
2) חוזרים על הניסוי n פעמים.
3) X – מוגדר <u>כמספר ההצלחות המתקבלות</u> 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?

© כל הזכויות שמורות לברק קנדל

שאלות:

1) במדינה 10% מהאוכלוסייה מובטלת. נבחרו 5 אנשים באקראי מאותה אוכלוסייה. נגדיר את X להיות מספר המובטלים שהתקבלו במדגם.

א. מהי ההתפלגות של X ?

ב. מה ההסתברות שיהיה בדיוק מובטל אחד?

ג. מה ההסתברות שכולם יעבדו במדגם?

ד. מה ההסתברות ששלושה יעבדו במדגם?

ה. מה ההסתברות שלפחות אחד יהיה מובטל?

ו. 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.

א. מהי ההתפלגות של X ? הסבירו.

ב. מה ההסתברות שבמדגם ל-8 אנשים יש סמארט-פון?

ג. מה ההסתברות שבמדגם לפחות ל-9 יהיו סמארט-פון?

ד. 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.

א. מה ההסתברות שיפסיד בכל המכונות?

ב. מה ההסתברות שיזכה בדיוק בשתי מכונות?

ג. מה ההסתברות שיזכה ביותר כסף מה-30 ₪ שהשקיע?

ד. 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.

א. מה ההסתברות ש-5 מהם אקדמאים?

ב. מה התוחלת של מס' בעלי ההשכלה הנמוכה?

© כל הזכויות שמורות לברק קנדל

5) במכללה מסוימת 20% מהסטודנטים גרים בת"א. מבין הסטודנטים שגרים בת"א 30% מגיעים ברכבם, ומבין הסטודנטים שלא גרים בת"א 50% מגיעים ברכבם למכללה.

א. השומר בשער המכללה בודק לכל סטודנט את תיקו בהיכנסו למכללה. מה ההסתברות שבקרב 5 סטודנטים שנבדקו ע"י השומר רק 1 מתוכם הגיע למכללה ברכבו?

ב. בהמשך לסעיף הקודם מה ההסתברות שרוב הסטודנטים בקרב ה-5 הגיעו למכללה ברכבם?

6) במבחן אמריקאי 20 שאלות. סטודנט ניגש למבחן והסיכוי שהוא יודע שאלה כלשהי הוא 0.8. אם הוא לא יודע הוא מנחש את התשובה.
לכל שאלה 4 תשובות אפשריות שרק אחת מהן נכונה.

א. מה הסיכוי לענות על שאלה מסוימת נכון?

ב. מה הסיכוי שיענה נכונה על בדיוק 16 שאלות?

ג. על כל שאלה שענה נכון התלמיד מקבל 5 נקודות, על כל שאלה ששגה מופחתת נקודה, מה התוחלת ומהי השונות של ציון התלמיד?

7) 5% מקו היצור פגום. המוצרים נארזים בתוך קופסת קרטון. בכל קופסא 10 מוצרים שונים. הקופסאות נארזות בתוך מכולה. בכל מכולה 20 קופסאות.

א. מה ההסתברות שבקופסא אקראית לפחות מוצר פגום אחד?

ב. מה התוחלת ומהי סטיית התקן של מספר הקופסאות במכולה בהן לפחות מוצר פגום אחד?

8) מטבע הוגן מוטל 5 פעמים. נגדיר את X כמספר הפעמים שהתקבל עץ.
חשבו את: $E\left(X^2\right)$.

© כל הזכויות שמורות לברק קנדל

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) ב. 0.2335 ג. 0.1493 ד. תוחלת: 7, סטיית תקן: 1.449.

3) א. 0.5314 ב. 0.0984 ג. 0.1143.

ד. תוחלת: 18-, סטיית תקן: 14.697.

4) א. 0.1789 ב. 2.

5) א. 0.1956 ב. 0.4253.

6) א. 0.85 ב. 0.182 ג. תוחלת: 82, שונות: 91.8.

7) א. 0.401 ב. תוחלת: 8.025, סטיית תקן: 2.193.

8) 7.5.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 20 - התפלגויות בדידות מיוחדות -התפלגות גיאומטרית

תוכן העניינים

1. כללי 79

התפלגויות בדידות מיוחדות – התפלגות גיאומטרית:

רקע:

חוזרים באופן בלתי תלוי על אותו ניסוי ברנולי.
X – מוגדר להיות מספר הניסויים שבוצעו עד ההצלחה הראשונה, כולל.
נסמן ב- p את הסיכוי להצלחה בניסויי בודד וב- q את הסיכוי לכישלון בניסוי בודד.
$X \sim G(p)$.

פונקציית ההסתברות: $P(X=k)=pq^{k-1}$, $k=1,2,...\infty$.

תוחלת: $E(X)=\frac{1}{p}$.

שונות: $V(X)=\frac{q}{p^2}$.

תכונות חשובות:

אם X מתפלג על פי התפלגות גיאומטרית, אזי X הוא בעל תכונת חוסר זיכרון,
דהיינו, $P(X>k)=q^k \cdot P(X=(n+k)/X>k)=P(X=n)$.

דוגמה (פתרון בהקלטה):

בכד 10 כדורים ש-3 מהם ירוקים. אדם מוציא באקראי כדור אחר כדור עד שבידו כדור ירוק. ההוצאה היא עם החזרת הכדור לכד בכל פעם מחדש.

א. מהי ההתפלגות של מספר הכדורים שהוצאו?

ב. מה ההסתברות שהוצאו בדיוק 5 כדורים?

ג. מה ההסתברות שהוצאו יותר מ 5 כדורים?

ד. אם הוצאו יותר מ-3 כדורים . מה הסיכוי שהוצאו בדיוק 5 כדורים?

ה. מה התוחלת וסטיית התקן של מספר הכדורים שהוצאו?

© כל הזכויות שמורות לברק קנדל

שאלות:

1) קו ייצור המוני מייצר מוצרים כך ש-5% מהם פגומים. איש בקרת איכות דוגם באופן מקרי מוצרים מקו הייצור עד אשר בידו מוצר פגום.
חשבו את ההסתברויות הבאות:

א. שידגום 3 מוצרים.

ב. שידגום 4 מוצרים.

ג. שידגום 5 מוצרים.

ד. שידגום יותר מ-7 מוצרים.

ה. שידגום לא פחות מ-8 מוצרים.

2) צילום שמבוצע במכון הרנטגן "X-RAY" יתקבל תקין בהסתברות של 0.9.
אדם נכנס למכון כדי להצטלם, והוא ייצא מהמכון רק כאשר יש בידו תצלום תקין.

א. מה ההסתברות שיצטלם בסך הכול 3 פעמים?

ב. מה ההסתברות שהצטלם יותר מ-4 פעמים?

ג. מה התוחלת ומה השונות של מספר הצילומים שייבצע?

ד. כל צילום עולה למכון 50 ₪. אדם משלם על צילום תקין 100 ₪.
מה התוחלת ומה השונות של רווח המכון מאדם שהגיע להצטלם?

3) מטילים מטבע עד אשר מתקבלת התוצאה "עץ".

א. מה ההסתברות להטיל את המטבע לכל היותר 10 פעמים?

ב. מה ההסתברות להטיל את המטבע לכל היותר 5 פעמים,
אם ידוע שהמטבע הוטל לפחות 3 פעמים?

ג. אם ידוע שבשתי ההטלות הראשונות התקבלה התוצאה "פלי", מה ההסתברות שהאדם הטיל את המטבע 7 פעמים?

ד. מה תוחלת מספר הפעמים שהתקבלה התוצאה "פלי"?

4) 30% מהמכוניות בארץ הן בצבע לבן. בכל יום נכנסות לחניון כשלהו 10 מכוניות אקראיות.

א. מה ההסתברות שביום מסוים בדיוק מחצית מהמכוניות בחניון יהיו לבנות?

ב. מה תוחלת מספר הימים שיעברו מהיום עד שלראשונה מחצית מהמכוניות בחניון יהיו לבנות?

© כל הזכויות שמורות לברק קנדל

5) אדם משחק במשחק מזל עד אשר הוא מפסיד. הצפי הוא שישחק את המשחק 10 פעמים. מה הסיכוי להפסיד במשחק בודד?

א. מה ההסתברות שישחק את המשחק בדיוק 6 פעמים?

ב. מה ההסתברות שישחק את המשחק לכל היותר 12 פעמים?

ג. ידוע שהאדם שיחק את המשחק יותר מ-6 פעמים.
מה ההסתברות ששיחק את המשחק בדיוק 10 פעמים?

ד. מהי סטיית התקן של מספר הפעמים שישחק את המשחק?

6) במאפייה מייצרים עוגות גבינה ועוגות שוקולד שנארזות באריזות אטומות. 40% מהעוגות הן עוגות גבינה והיתר שוקולד. התווית על האריזה מודבקת בשלב מאוחר יותר של הייצור. אדם נכנס למפעל ובוחר באקראי עוגה.

א. מה ההסתברות שייאלץ לבחור 5 עוגות עד שקיבל עוגות שוקולד?

ב. אם הוא דגם פחות מ-7 עוגות עד שיקבל עוגת שוקולד, מה ההסתברות שבפועל הוא דגם יותר מ-4 עוגות?

ג. האדם דוגם עוגות עד אשר הוא מוצא עוגת שוקולד. ידוע שעוגת גבינה עולה ליצרן 50 שקלים ועוגת שוקולד 30 שקלים. מהי התוחלת ומהי השונות של עלות הייצור הכוללת של העוגות שדגם?

ד. בהמשך לסעיף הקודם, מהי התוחלת ומהי סטיית התקן של מספר עוגת הגבינה שדגם האדם?

תשובות סופיות:

1) א. 0.04512 ב. 0.0428 ג. 0.0407 ד. 0.6983 ה. 0.6983.

2) א. 0.009 ב. 0.0001 ג. תוחלת: 1.111, שונות: 0.1234.
ד. תוחלת: 44.4, שונות: 308.5.

3) א. 0.999 ב. 0.875 ג. 0.03125 ד. 1.

4) א. 0.1029 ב. 9.72.

5) א. 0.06 ב. 0.7176 ג. 0.0729 ד. 9.487.

6) א. 0.015 ב. 0.0215 ג. תוחלת: $63\frac{1}{3}$, שונות: $2777\frac{7}{9}$.
ד. תוחלת $\frac{2}{3}$, שונות 1.054.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 21 - התפלגויות בדידות מיוחדות- התפלגות פואסונית

תוכן העניינים

1. כללי 82

התפלגויות בדידות מיוחדות – התפלגות פואסונית:

רקע:

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

λ - פרמטר המאפיין את ההתפלגות הנ"ל. הפרמטר מייצג את קצב האירועים ביחידת זמן. כלומר, כמה אירועים בממוצע קורים ביחידת זמן: $X \sim pois(\lambda)$.

התפלגות פואסונית חייבת להופיע כנתון בשאלה ולכן לא יהיה צורך לזהותה.

פונקציית ההסתברות של ההתפלגות הפואסונית נתונה:

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

התוחלת והשונות של ההתפלגות:

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

תכונות מיוחדות של ההתפלגות:

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

דוגמה (פתרון בהקלטה):

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

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

א. מה ההסתברות שבדקה תתקבל פניה 1?

ב. מה ההסתברות שבדקה תתקבל לפחות פניה 1?

ג. מה ההסתברות שבדקה יתקבלו לכל היותר 2 פניות?

ד. מה שונות מספר הפניות בדקה?

2) מספר הטעויות לעמוד בעיתון מתפלג פואסונית עם ממוצע של 4 טעויות לעמוד.
בחלק מסוים של עיתון ישנם 5 עמודים.

א. מה ההסתברות שבחלק זה ישנן בדיוק 18 טעויות?

ב. אם בעמוד הראשון אין טעויות, מה ההסתברות שבסך הכול בכול החלק ישנן 15 טעויות?

ג. אם בחלק של העיתון נמצאו בסך הכול 18 טעויות, מה ההסתברות ש-5 מהן בעמוד הראשון?

3) מספר תאונות הדרכים הקטלניות במדינת ישראל מתפלג פואסונית עם סטיית תקן של 2 תאונות לשבוע.

א. מה תוחלת מספר התאונות בשבוע?

ב. מהי ההסתברות שבחודש (הניחו שבחודש יש 4 שבועות) יהיה בדיוק שבוע אחד בו יהיו 3 תאונות דרכים קטלניות?

4) לחנות AM:PM השכונתית מספר הלקוחות שנכנסים מתפלג פואסונית עם ממוצע של 2 לקוחות לדקה.

א. מה ההסתברות שבדקה כלשהי יהיו בדיוק 3 לקוחות?

ב. מה ההסתברות שבדקה כלשהי יגיח לפחות לקוח אחד?

ג. מה ההסתברות שבדקה כלשהי יהיו לכל היותר שני לקוחות?

ד. מהי התוחלת ומה סטיית התקן של מספר הלקוחות שנכנסים לחנות בדקה?

5) מספר הלידות בבית חולים מתפלג פואסונית עם תוחלת של 8 לידות ביום.

א. מה ההסתברות שביום א' נולדו 10 תינוקות וביום ב' נולדו 7 תינוקות?

ב. מיילדת עובדת במשמרות של 8 שעות. מה ההסתברות שבמשמרת שלה נולדו 3 תינוקות?

ג. מהי התוחלת של מספר הימים בשבוע בהם נולדים ביום עשרה תינוקות?

© כל הזכויות שמורות לברק קנדל

6) במערכת אינטרנט לתשלום חשבונות, מספר החשבונות המשולמים בשעה מתפלג פואסונית עם תוחלת של 30.

א. כמה שעות צפויות לעבור עד אשר תתקבל שעה עם בדיוק 33 חשבונות?

ב. בין השעה 08:00 ל-08:20 היו 18 חשבונות, מה ההסתברות שבין 08:00 ל-08:10 היו בדיוק 6 חשבונות?

תשובות סופיות:

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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 22 - התפלגויות בדידות מיוחדות-התפלגות היפרגאומטרית

תוכן העניינים

1. כללי 85

התפלגויות בדידות מיוחדות – התפלגות היפרגאומטרית:

רקע:

נתונה אוכלוסייה המכילה N פריטים , מתוכה D פריטים בעלי תכונה מסוימת – פריטים אלה נקראים "מיוחדים". בוחרים מאותה אוכלוסייה n פריטים ללא החזרה.
X מוגדר להיות מספר הפריטים ה"המיוחדים" שנדגמו.
משתנה מקרי היפרגאומטרי עם הפרמטרים (N,D,n) יסומן על ידי: $X \sim H(N,D,n)$.

פונקציית ההסתברות של ההתפלגות: $P(X=k)=\dfrac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$.

התוחלת של ההתפלגות: $E(X)=n\cdot\dfrac{D}{N}$.

השונות של ההתפלגות: $V(X)=n\cdot\dfrac{D}{N}\cdot\left(1-\dfrac{D}{N}\right)\cdot\dfrac{N-n}{N-1}$.

דוגמה (הפתרון בהקלטה):

בכתה 40 תלמידים, שמתוכם 10 בנות והשאר בנים. בוחרים קבוצה של ארבעה תלמידים שיסעו למשלחת.

א. כיצד מספר הבנים במשלחת מתפלג?
ב. מה התוחלת ומהי השונות של מספר הבנים במשלחת?
ג. מה הסיכוי שבמשלחת יהיו 3 בנים?

© כל הזכויות שמורות לברק קנדל

שאלות:

1) בכד 5 כדורים אדומים ו-4 כדורים ירוקים. מוציאים באקראי שלושה כדורים מהכד.

א. בנו את פונקציית ההסתברות של מספר הכדורים האדומים שהוצא בטבלה.

ב. חשבו את התוחלת והשונות של מספר הכדורים האדומים שהוצאו, פעם מתוך פונקציית ההסתברות ופעם מתוך הנוסחאות להתפלגות היפרגאומטרית.

ג. מה הייתה התוחלת והשונות של מספר הכדורים האדומים אם ההוצאה הייתה עם החזרה?

2) בחידון 10 שאלות משלושה תחומים שונים: 3 בתחום הספורט, 4 בתחום הבידור והיתר בתחום המדעים. משתתף בחידון שולף באקראי 4 שאלות.
נגדיר את X להיות מספר השאלות מתחום הספורט שנשלפו.

א. בנו את פונקציית ההסתברות של X בנוסחה (לא בטבלה).

ב. מה התוחלת וסטיית התקן של X?

ג. חשבו את ההסתברות הבאה: $P\left(X=2 \mid X>1\right)$.

3) נדגמו 6 אנשים מתוך אוכלוסייה שבה 60% בעלי רישיון נהיגה.
אנו מתעניינים במספר האנשים שנדגמו עם רישיון נהיגה.
זהו בסעיפים הבאים את ההתפלגות, וחשבו לכל התפלגות את התוחלת והשונות:

א. האוכלוסייה גדולה מאד.

ב. האוכלוסייה בת 10 אנשים.

4) בארגון עובדים 7 מהנדסים, 3 טכנאים ו-5 הנדסאים.
בוחרים באופן מקרי משלחת של 4 עובדים לכנס במדריד.

א. מהי ההסתברות שייבחרו רק מהנדסים?

ב. מה תוחלת מספר הטכנאים שייבחרו?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א.

x	0	1	2	3
$P(x)$	$\frac{4}{84}$	$\frac{30}{84}$	$\frac{40}{84}$	$\frac{10}{84}$

ב. תוחלת: $1\frac{2}{3}$, שונות: $\frac{5}{9}$.

ג. תוחלת: $1\frac{2}{3}$, שונות: $\frac{20}{27}$.

2) א. $\frac{\binom{3}{k}\cdot\binom{7}{4-k}}{\binom{10}{4}}$. ב. תוחלת: 1.5, סטיית תקן: 0.748. ג. 0.9.

3) א. תוחלת: 3.6, שונות: 1.44. ב. תוחלת: 3.6, שונות: 0.64.

4) א. 0.0256. ב. 0.8.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 23 - התפלגויות בדידות מיוחדות -התפלגות בינומית שלילית

תוכן העניינים

1. כללי 88

התפלגויות בדידות מיוחדות – התפלגות בינומית שלילית:

רקע:

בהתפלגות זו חוזרים על אותו ניסוי ברנולי בזה אחר זה באופן בלתי תלוי עד אשר מצליחים בפעם ה- r. X = מספר החזרות עד שהתקבלו r הצלחות: $X \sim NB(r,p)$.

פונקציית ההסתברות: $k = r, r+1, \ldots \infty$, $P(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$.

תוחלת: $E(X) = \frac{r}{p}$.

שונות: $V(X) = \frac{r(1-p)}{p^2}$.

דוגמה (פתרון בהקלטה):

קובייה מוטלת עד שמקבלים 3 פעמים תוצאה שגדולה מ-4.

א. מה הסיכוי להטיל את הקובייה 6 פעמים?

ב. מה תוחלת ושונות מספר הפעמים שנטיל את הקובייה?

© כל הזכויות שמורות לברק קנדל

שאלות:

1) בכד 4 כדורים שחורים ו-6 כדורים לבנים. כדור מוצא באקראי פעם אחר פעם ומוחזר בין הוצאה להוצאה. נסמן ב- X את מספר הכדורים שהוצאו עד שהתקבלו 2 כדורים לבנים בסך הכול (לא בהכרח ברצף).

א. חשבו את $P(X=2)$.

ב. חשבו את $P(X=3)$.

ג. חשבו את $P(X=4)$.

ד. חשבו את $P(X=k)$.

2) הסיכוי לזכות במשחק מזל הוא 0.4. אדם משחק במשחק ומפסיק ברגע שהוא ניצח פעמיים (לא בהכרח ברצף).

א. מה הסיכוי שישחק פעמיים?

ב. מה הסיכוי שישחק 3 פעמים?

ג. מה הסיכוי שישחק 4 פעמים?

ד. מה הסיכוי שישחק 5 פעמים?

ה. מה הסיכוי שישחק k פעמים?

3) הראו שההתפלגות הגאומטרית היא מקרה פרטי של ההתפלגות הבינומית השלילית.

4) מטבע מוטל שוב ושוב עד שמתקבל שלוש פעמים עץ בסך הכול.

א. בנו את פונקציית ההסתברות של מספר ההטלות הכולל.

ב. מהי התוחלת ומהי השונות של מספר ההטלות הכולל?

ג. חוזרים על התהליך שלעיל 5 פעמים. מה ההסתברות שפעמיים מתוך ה-5 חזרות נאלץ להטיל את המטבע בדיוק 4 פעמים?

5) יהיה X_i מספר החזרות עד ההצלחה הראשונה בניסיונות ברנוליים בלתי תלויים זה בזה, כאשר $i=1,2,\dots,n$.

הוכיחו שהתוחלת והשונות של $\sum_{i=1}^{n} X_i$ זהות לתוחלת והשונות של ההתפלגות הבינומית השלילית $NB(n,p)$.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. 0.36. ב. 0.288. ג. 0.0576. ד. $0.6^2 \cdot 0.4^{k-2}$.

2) א. 0.16. ב. 0.192. ג. 0.1728. ד. 0.13824. ה. $0.4^2 \cdot 0.6^{k-2}$.

3) שאלת הוכחה.

4) ב. תוחלת: 6, שונות: 6. ג. 0.1886.

5) שאלת הוכחה.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 24 - המשתנה המקרי הבדיד - שאלות מסכמות

תוכן העניינים

1. כללי 91

המשתנה המקרי הבדיד – שאלות מסכמות:

שאלות:

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

א. חשבו את התוחלת וסטיית התקן של X.

ב. $W = 2X - 4$, חשבו את התוחלת וסטיית התקן של W.

ג. $T = X + Y$, חשבו את התוחלת של T.
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.

א. מה הסיכוי שערן ניצח בשני המשחקים?

ב. מה התוחלת ומהי השונות של מספר הניצחונות של ערן?

ג. 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.

א. בנו את פונקציית ההסתברות של X.

ב. חשבו את התוחלת והשונות של X.

ג. כל ניסיון לפתוח הדלת אורך חצי דקה. מה התוחלת ומה השונות של הזמן הכולל לפתיחת הדלת?

4) מספר התקלות בשידור "ערוץ 1" מתפלג פואסונית בקצב של 6 תקלות ביום.

א. מה ההסתברות שביום מסוים הייתה לפחות תקלה אחת?

ב. מה ההסתברות שבשבוע (7 ימי שידור) יהיו בדיוק 6 ימים בהם לפחות תקלה אחת?

ג. מה תוחלת מספר הימים שיעברו מהיום ועד היום הראשון בו לפחות תהיה תקלה אחת?

© כל הזכויות שמורות לברק קנדל

5) בעל חנות גדולה בקניון שם לב ש-40% מהמוצרים בחנותו נרכשים עבור ילדים, 35% נרכשים עבור נשים ו-25% נרכשים עבור גברים. 10% מהמוצרים הנרכשים עבור ילדים הם מתוצרת חוץ, וכך גם 60% מהמוצרים הנרכשים עבור נשים ו-50% מאלה הנרכשים עבור גברים.

א. מה ההסתברות למכור בחנות זו מוצר מתוצרת חוץ?

ב. יהי X מספר המוצרים שימכרו בחנות זו מפתיחתה ביום א' בבוקר, עד שלראשונה יימכר מוצר מתוצרת הארץ (כולל).
מהי פונקציית ההסתברות של X?

ג. מהי תוחלת מס' המוצרים מתוצרת חוץ שימכרו, עד שלראשונה יימכר מוצר מתוצרת הארץ?

ד. ביום ב' נמכרו בחנות 7 מוצרים. מה ההסתברות שבדיוק 3 מהם הם מתוצרת חוץ?

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 "הצבי" יימכר לחו"ל בסיכוי של 0.6.
הסרט "לעולם לא" יימכר לחו"ל בסיכויי 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?

7) במפעל מייצרים סוכריות כך ש-20% מהסוכריות בטעם תות. הייצור הוא ייצור המוני. שאר הסוכריות בטעמים שונים, השקיות נארזות ובכל שקית בדיוק 5 סוכריות.

א. נבחרה שקית ונתון שבשקית פחות מ-3 סוכריות אדומות.
מה ההסתברות שבשקית סוכריה אדומה אחת?

ב. בוחרים באקראי שקית אחר שקית, במטרה למצוא שקית ללא סוכריות אדומות. מה ההסתברות שייאלצו לדגום יותר מ-6 שקיות?

© כל הזכויות שמורות לברק קנדל

8) מבחן בנוי משני חלקים: בחלק א' 10 שאלות ובחלק ב' 10 שאלות. תלמיד התכונן 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: $E(X-5)^2$.

10) הסיכוי לעבור מבחן נהיגה הינו P. בוחרים באקראי ארבעה נבחנים.

ההסתברות ששניים מהם יעברו את מבחן הנהיגה גבוה פי $\frac{8}{3}$ מהסיכוי שכל הארבעה יעברו את המבחן.

א. חשבו את ערכו של P.

ב. תלמיד ניגש לבחינה עד אשר הוא עובר אותה.
מה ההסתברות שיעבור את מבחן הנהיגה רק במבחן הרביעי?

ג. מה ההסתברות שיאלץ לגשת לפחות לחמישה מבחנים בסך הכול?

ד. מה התוחלת ומהי השונות של מספר המבחנים שבהם יכשל?

ה. ידוע שהתלמיד ניגש לשלושה מבחנים ועדיין לא עבר. מה ההסתברות שבסופו של דבר יעבור במבחן הנהיגה החמישי?

11) רובוט נמצא בנקודה 0 על ציר המספרים. הרובוט מבצע n צעדים ובכל צעד הוא נע בסיכוי P. ימינה ביחידה אחת ובסיכוי $1-P$ שמאלה ביחידה אחת.
נסמן ב- X את המספר עליו עומד הרובוט לאחר n צעדים.
רשמו את פונקציית ההסתברות של X באמצעות P ו- n.

12) למטבע יש סיכוי P לקבל את התוצאה ראש. מטילים את המטבע. אם יוצא ראש בפעם הראשונה מפסידים שקל ומפסיקים את המשחק. אחרת, ממשיכים לזרוק וזוכים במספר שקלים לפי מספר הפעמים שהטלנו את המטבע מההתחלה ועד שהתקבל ראש.

א. בנו את פונקציית ההסתברות של רווח המשחק (באמצעות P).

ב. בטאו את תוחלת הרווח באמצעות P.

ג. לאלו ערכי P המשחק כדאי?

© כל הזכויות שמורות לברק קנדל

13) מטבע הוגן מוטל עד שמתקבל $m+1$ פעמים עץ.
רשמו את פונקציית ההסתברות של מספר הפעמים שהתקבל פלי.

14) נתונות N מגירות ממוספרות מ-1 ועד N. מתוך n חולצות, יש לבחור באופן אקראי לכל חולצה מגירה. כל מגירה יכולה להכיל את כל החולצות.
נגדיר את X_1 - כמספר החולצות שהונחו במגירה מספר 1.
נגדיר את X_N - כמספר החולצות שהונחו במגירה מספר N.
חשבו את: $V\left(X_1+X_N\right)$.

15) n אנשים יושבים במסעדה. בזמן שמגיע העת לשלם, האנשים פועלים לפי העיקרון הבא: כל אחד מהם מטיל מטבע הוגן עד אשר אחד מהם מקבל תוצאה שונה מכל השאר והוא זה שמשלם.
מהי תוחלת מספר הסבבים שיבוצעו עד שימצא משלם?

16) הסיכוי לעבור בקורס מסוים את מועד א' הוא 0.7. סטודנט שנכשל במועד א' בהכרח ניגש למועד ב' ואז הסיכוי שלו לעבור אותו הוא 0.8. אם סטודנט נכשל במועד ב' הוא ניגש למועד מיוחד ואחרון.
נתון שלמועד א' נגשו כל 20 הסטודנטים הרשומים לקורס.
מהי התפלגות מספר הבחינות שיאלץ המרצה לחבר?

17) לקניון 3 כניסות שונות. בכל כניסה מספר האנשים שנכנסים לקניון מתפלג פואסונית באופן בלתי תלוי בכניסה האחרת. מספר האנשים שנכנסים בכניסה ה-i מתפלג פואסונית עם קצב של i אנשים בשנייה.
יהי Y מספר האנשים שנכנסים לקניון בשנייה מכל הכניסות יחדיו.
מצאו את: $E\left[\frac{1}{Y+1}\right]$.

18) לרני 20 טושים אותם הוא מכניס באקראי ל-3 קלמרים. לכל טוש נבחר קלמר באקראי ובאופן בלתי תלוי בטוש אחר. כל קלמר יכול להכיל עד 20 טושים.
נסמן ב-X את מספר הקלמרים שיש בהם בדיוק 10 טושים.
חשבו את $E\left(\sqrt{x+7}\right)$.

© כל הזכויות שמורות לברק קנדל

19) בשדרות רוטשילד החליטו לשתול n ברושים ו-2 אורנים אחד אחרי השני בשורה. סידור העצים בשורה נעשה באקראי. נגדיר את X להיות מספר הברושים, בין הברוש הגבוה ביותר לברוש הנמוך ביותר שנשתלו.

א. מצאו את ההתפלגות של X.

ב. הוכיחו שהתוחלת של X היא $\frac{n-2}{3}$.

© כל הזכויות שמורות לברק קנדל

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.

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

ג. תוחלת: 1.5, שונות: 0.5.

4) א. 0.9975.
ב. 0.0172.
ג. 1.0025.

5) א. 0.375.
ב. 0.6.
ג. 0.282.
ד. 0.282.

6) א. ראו טבלה:
ב. תוחלת: 1.5, שונות: 0.61.

x	0	1	2	3
$P(x)$	0.092	0.392	0.428	0.084

ג. תוחלת: 0, סטיית תקן: 4.68.

7) א. 0.4348.
ב. 0.0923.

8) א. 2.013.
ב. 0.5256.
ג. תוחלת: 8, שונות: 1.6.
ד. תוחלת: 10.5, שונות: 3.475.

9) 10.

10) א. 0.6.
ב. 0.0384.
ג. 0.0256.
ד. תוחלת: 0.67, שונות: 1.11.
ה. 0.24.

11) $P(X=k)=\binom{n}{\frac{k+n}{2}}\cdot p^{\frac{k+n}{2}}\cdot(1-p)^{\frac{n-k}{2}}$.

12) א. $P(X=k)=\begin{cases} P & k=-1 \\ (1-P)^{K-1}\cdot P & k=2,3,\dots,\infty \end{cases}$.
ב. $\frac{1-2p^2}{p}$.
ג. $0<p<\sqrt{\frac{1}{2}}$.

© כל הזכויות שמורות לברק קנדל

13) $k=0,1,\dots,\infty$, $P(X=k)=\binom{m+k}{m}\cdot\left(\frac{1}{2}\right)^{k+m+1}$.

14) $n\cdot\left(\frac{2}{N}\right)\cdot\left(1-\frac{2}{N}\right)$.

15) $\frac{2^n}{2n}$.

16) ראו טבלה:

X	1	2	3
$P(X)$	0.0008	0.2893	0.7099

17) $\frac{e^{-6}}{6}\left[e^6-1\right]$.

18) 2.675.

19) א. $k=0,1,\dots,n-2$, $P(X=k)=\frac{n-k-1}{\binom{n}{2}}$. ב. הוכחה.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 25 - המשתנה המקרי הרציף- התפלגויות כלליות (שימוש באינטגרלים)

תוכן העניינים

1. כללי 98

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 התפלגות מצטברת:

פונקציית ההתפלגות המצטברת מוגדרת באופן הבא: $F(t)=p(X\le t)=\int_{-\infty}^{t} f(x)dx$

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: $V(X)=\int_{-\infty}^{\infty} X^2\cdot f(x)dx-\mu^2=\sigma^2$.

© כל הזכויות שמורות לברק קנדל

תוחלת של פונקציה של X:

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

כלומר: $p(X \le x_p)=p$.

ריענון מתמטי:

נוסחאות לחישוב שטחים

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

שטח מלבן: אורך (a) 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.

שיפוע ישר העובר דרך שתי נקודות: $(x_1,y_1),(x_2,y_2)$ 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 הבא: $y-y_1=m(x-x_1)$.

© כל הזכויות שמורות לברק קנדל

אינטגרלים מידיים:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

שאלות:

1) X הינו משתנה רציף עם פונקציית צפיפות כמוצג בשרטוטו:

א. מצאו את ערכו של c.

ב. בנו את פונקציית ההתפלגות המצטברת.

ג. חשבו את ההסתברויות הבאות:

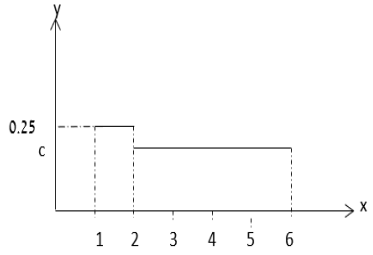
i. $P(x<4)$.

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: $f(x)=\begin{cases} cx & 0\le x\le b \\ 0 & else \end{cases}$,

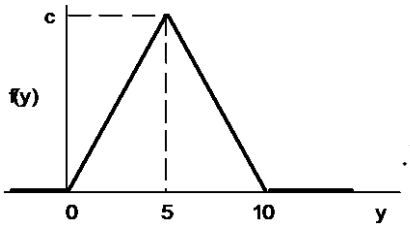
וידוע ש- $P(0<X<1)=\frac{1}{4}$.

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:

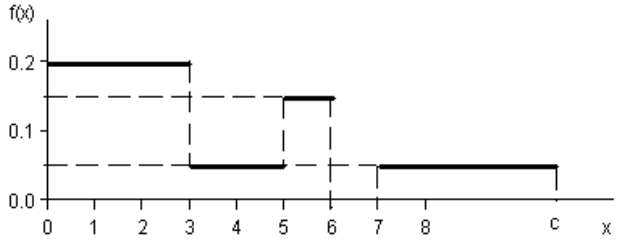
$P(Y=7.0)$, $P(Y\le 3.0)$, $P(7.5\le Y\le 15.5)$, $P(Y>4)$.

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

הסיקו מהו העשירון עליון: $y_{0.9}$.

© כל הזכויות שמורות לברק קנדל

4) נתונה פונקציית צפיפות של משתנה מקרי X:



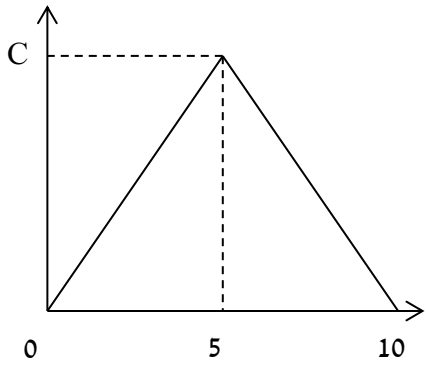
א. מצאו ערך c שעבורו תתקבל פונקציית צפיפות.

ב. מצאו את פונקציית ההתפלגות המצטברת.

ג. חשבו את ההסתברויות הבאות:

$P(1.0 < X \le 5.0)$, $P(X \ge -2.0)$, $P(X \ge 4)$.

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?

© כל הזכויות שמורות לברק קנדל

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:
$f(x)=0.5\cdot e^{2x}$, $-\infty\le X\le \ln(c)$.

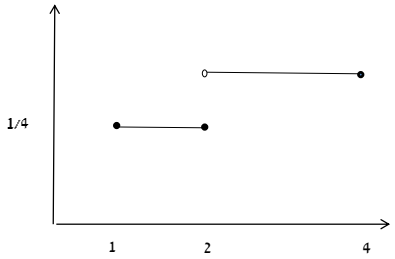
א. מצאו את ערכו של c.

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 פונקציית הצפיפות הבאה: $f(x)=6x(1-x)$, $0\le x\le 1$.

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 ההתפלגות המצטברת הבאה: $F(t)=1-e^{-0.2t}$.

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?

© כל הזכויות שמורות לברק קנדל

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

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

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

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: $f(x)=\frac{1}{b-a}\ ,\ a\le x\le b.$

א. מצאו את פונקציית ההתפלגות המצטברת.

ב. חשב את התוחלת והשונות של ההתפלגות.

ג. מצאו את התוחלת של $\frac{1}{X}$.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. $\frac{3}{16}$.

ב. $F(t)=\begin{cases}0 & t<1\\ (t-1)0.25 & 1\le t\le 2\\ 0.25+(t-2)\cdot\frac{3}{16} & 2<t\le 6\\ 1 & t>6\end{cases}$.

ג. i. $\frac{5}{8}$.

ii. $\frac{7}{8}$. iii. $\frac{11}{16}$. iv. $\frac{3}{16}$. ד. $3\frac{1}{3}$.

2) א. $c=0.5$, $b=2$. ב. 1.41. ג. 0.0625.

3) א. 0.2.

ב. $F(t)=\begin{cases}0 & t<0\\ 0.02t^2 & 0\le t\le 5\\ 1-0.02(t-10)^2 & 5<t\le 10\\ 1 & t>10\end{cases}$.

ג. 0, 0.18, 0.125, 0.32.

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

4) א. 10.

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

ג. 0.5.

5) א. $c=0.2$. ב. 5 ± 1.46.

6) א. $e^{\frac{1}{2}}$.

ב. $F(t)=\begin{cases}0 & t<1\\ 2\cdot\ln t & 1\le t\le e^{\frac{1}{2}}\\ 1 & t>e^{\frac{1}{2}}\end{cases}$.

ג. 0.189.

ד. 1.051. ה. 1.297.

7) א. 0.0012. ב. 0.7067. ג. תוחלת: 6, שונות: 4.

© כל הזכויות שמורות לברק קנדל

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

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

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

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

11) א. עיין סרטוט בוידאו ב. 0.0498 ג. 0.6321 ד. 11.51.

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

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

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

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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 26 - התפלגויות רציפות מיוחדות- התפלגות מעריכית

תוכן העניינים

1. כללי 107

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 מההתפלגות 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 התרחשות המאורע הבא מתפלג מעריכית.

פונקציית הצפיפות של ההתפלגות:

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

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

התוחלת: $E(x) = \frac{1}{\lambda}$.

השונות: $V(x) = \frac{1}{\lambda^2}$.

להתפלגות זו יש תכונת חוסר הזיכרון: $P(X > a+b \setminus X > a) = P(X > b)$.

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?

© כל הזכויות שמורות לברק קנדל

שאלות:

1) הזמן שלוקח במערכת עד שתקלה מתרחשת מתפלג מעריכית עם תוחלת של 0.5 שעה.

א. מה הסתברות שהתקלה הבאה תתרחש תוך יותר מ-0.5 שעה?

ב. מה ההסתברות שהתקלה הבאה תתרחש תוך פחות משעה?

ג. מצא את הזמן החציוני להתרחשות תקלה במערכת.

2) הזמן שעובר בכביש מסוים עד להתרחשות תאונה מתפלג מעריכית עם תוחלת של 24 שעות.

א. מהי סטית התקן של הזמן עד להתרחשות תאונה?

ב. מה ההסתברות שהתאונה הבאה תתרחש תוך פחות מיממה?

ג. 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.

א. מה הסיכוי שעבודת סטודנט על המחשב תארך פחות מרבע שעה?

ב. מה הסיכוי שעבודת סטודנט על המחשב תארך בין רבע שעה לחצי שעה?

ג. אם סטודנט עובד על המחשב כבר יותר מ-10 דקות, מה ההסתברות שמשך כל עבודתו יעלה על 30 דקות?

ד. מהו הזמן שבסיכוי של 90% הסטודנט יעבוד פחות ממנו?

4) בממוצע מגיעים לחדר מיון 4 חולים בשעה בזרם פואסוני.

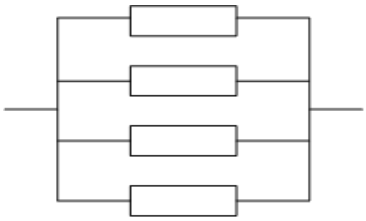
א. שולה המזכירה הגיעה לחדר המיון. מה ההסתברות שזמן ההמתנה שלה לחולה הבא יהיה יותר מ-20 דקות?

ב. אם שולה המתינה יותר מרבע שעה לחולה הבא. מה ההסתברות שתמתין בסך הכל יותר מחצי שעה?

ג. מה ההסתברות שבין החולה הראשון לשני יש להמתין יותר מרבע שעה ובין החולה שני לשלישי יש להמתין פחות מרבע שעה?

© כל הזכויות שמורות לברק קנדל

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. $K < 0.0588A$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 27 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

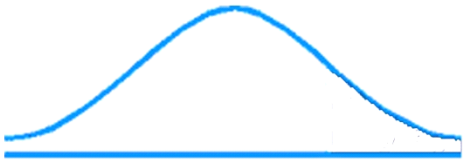
תוכן העניינים

1. כללי 110

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

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

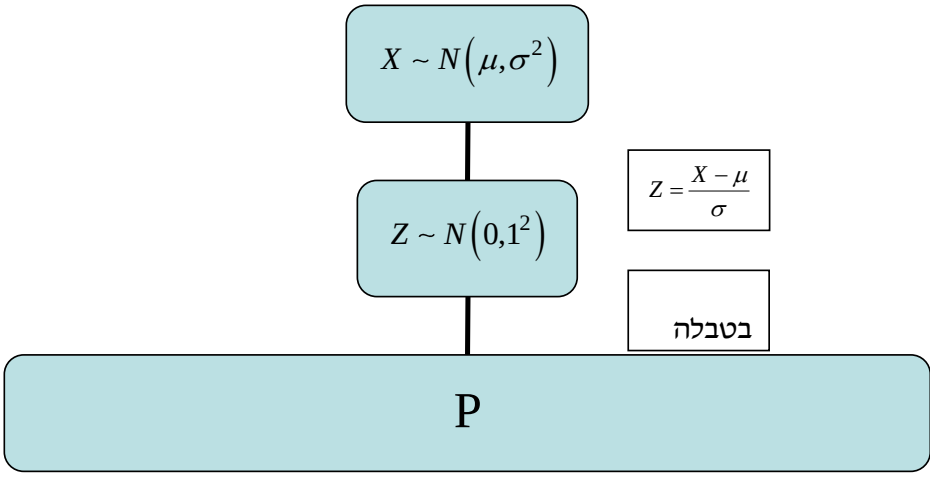
כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלוונטיים שמתחת לעקומה. כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון. התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת, והיא תסומן באות Z: $Z \sim N\left(0,1^2\right)$.

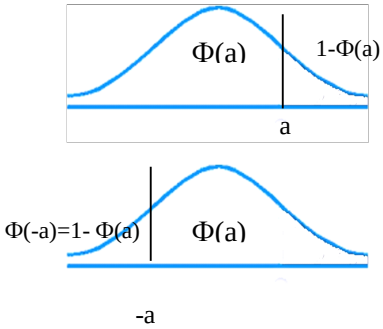
תהליך התקנון מבוצע על ידי הנוסחה הבאה: $Z=\frac{X-\mu}{\sigma}$.

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

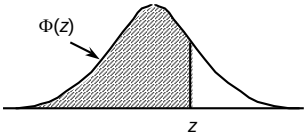
לאחר חישוב ציון התקן של ערך מסוים נעזרים בטבלה של ההתפלגות הנורמלית הסטנדרטית לחישוב השטח הרצוי, ובאופן כללי נתאר את הסכמה הבאה:

© כל הזכויות שמורות לברק קנדל


$X \sim N(\mu,\sigma^2)$
$Z = \frac{X-\mu}{\sigma}$
$Z \sim N(0,1^2)$
בטבלה
P


Φ(a)
1-Φ(a)
a
Φ(-a)=1- Φ(a)
Φ(a)
-a

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
Φ(z)	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995

© כל הזכויות שמורות לברק קנדל

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

© כל הזכויות שמורות לברק קנדל

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 וסטיית תקן 400 גרם.

א. מצאו את העשירון העליון.

ב. מצאו את האחוזון ה-95.

ג. מצאו את העשירון התחתון.

© כל הזכויות שמורות לברק קנדל

5) ציוני מבחן אינטליגנציה מתפלגים נורמלית עם ממוצע 100 ושונות 225.

א. מה העשירון העליון של הציונים במבחן האינטיליגנציה?

ב. מה העשירון התחתון של ההתפלגות?

ג. מהו הציון ש-20% מהנבחנים מקבלים מעליו?

ד. מהו האחוזון ה-20?

ה. מהו הציון ש-5% מהנבחנים מקבלים מתחתיו?

6) נפח משקה בבקבוק מתפלג נורמלית עם סטיית תקן של 20 מ"ל, ונתון ש-33% מהבקבוקים בעלי נפח שעולה על 508.8 מ"ל.

א. מה ממוצע נפח משקה בבקבוק?

ב. 5% מהבקבוקים המיוצרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאיזה נפח שולחים בקבוק לבדיקה?

ג. 1% מהבקבוקים עם הנפח הקטן ביותר נתרמים לצדקה, מהו הנפח המקסימלי לצדקה?

7) אורך חיים של מכשיר מתפלג נורמלית. ידוע שמחצית מהמכשירים חיים פחות מ-500 שעות, כמו כן ידוע ש-67% מהמכשירים חיים פחות מ-544 שעות.

א. מהו ממוצע אורך חיי מכשיר?

ב. מהי סטית בתקן של אורך חיי מכשיר?

ג. מה הסיכוי שמכשיר אקראי יחיה פחות מ-460 שעות?

ד. מהו המאיון העליון של אורח חיי מכשיר?

ה. 1% מהמכשירים בעלי אורך החיים הקצר ביותר נשלח למעבדה לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשליחת מכשיר למעבדה?

8) להלן שלוש התפלגויות נורמליות של שלוש קבוצות שונות ששורטטו באותה מערכת צירים. ההתפלגויות מוספרו כדי להבדיל ביניהן.

א. לאיזו התפלגות הממוצע הגבוה ביותר?

ב. במה מבין המדדים הבאים התפלגות 1 ו-2 זהות?

i. בעשירון העליון.

ii. בממוצע.

iii. בשונות.

ג. לאיזו התפלגות סטיית התקן הקטנה ביותר?

i. 1.

ii. 2.

iii. 3.

iv. אין לדעת.

© כל הזכויות שמורות לברק קנדל

9) הזמן שלוקח לאדם להגיע לעבודתו מתפלג נורמלית עם ממוצע של 40 דקות וסטית תקן של 5 דקות.

א. מה ההסתברות שמשך הנסיעה של האדם לעבודתו יהיה לפחות שלושת רבעי השעה?

ב. אדם יצא לעבודתו בשעה 08:10 מביתו. הוא צריך להגיע לעבודתו בשעה 09:00. מה הסיכוי שיאחר לעבודתו?

ג. אם ידוע שזמן נסיעתו לעבודה היה יותר משלושת רבעי השעה. מה ההסתברות שזמן הנסיעה הכולל יהיה פחות מ-50 דקות?

ד. מה הסיכוי שבשבוע (חמישה ימי עבודה) בדיוק פעם אחת יהיה זמן הנסיעה לפחות שלושת רבעי השעה?

10) ההוצאה החודשית לבית אב בעיר "טרירה" מתפלגת נורמלית עם ממוצע של 2000 דולר וסטית תקן של 300 דולר. בחרו באקראי 5 בתי אב. ההסתברות שלפחות אחד מהם מוציא בחודש מעל ל- T דולר היא 0.98976.

א. מה ערכו של T?

ב. מה הסיכוי שההוצאה החודשית של בית אב בעיר תהיה לפחות סטיית תקן אחת מעל T?

ג. מסתבר שנפלה טעות בנתונים, ויש להוסיף 100 דולר להוצאות החודשית של כל בתי האב בעיר. לאור זאת, מה ההסתברות שההוצאה החודשית של בית אב נמוכה מ-1800 דולר?

11) אורך שיר אקראי המשודר ברדיו מתפלג נורמלית עם תוחלת של 3.5 דקות וסטיית 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?

© כל הזכויות שמורות לברק קנדל

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	

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 28 - התפלגות מינימום ומקסימום

תוכן העניינים

1. כללי 118

התפלגות מינימום ומקסימום:

רקע:

התפלגות מקסימום:

נניח ש- X_i הינם משתנים מקריים בלתי תלויים בעלי אותה התפלגות רציפה.

נגדיר את: $U=\max\left(x_1,x_2,...x_n\right)$. מתקיים ש: $F_U\left(t\right)=\left(F_X\left(t\right)\right)^n$,

ולכן: $f\left(u\right)=n\cdot\left(F_x\left(u\right)\right)^{n-1}\cdot f_x\left(u\right)$.

התפלגות מינימום:

נניח ש- X_i הינם משתנים מקריים בלתי תלויים בעלי אותה התפלגות רציפה.

נגדיר את: $Z=\min\left(X_1,X_2...,X_n\right)$. מתקיים ש: $F_Z\left(t\right)=1-\left(1-F_X\left(t\right)\right)^n$,

ולכן: $f(z)=n\cdot\left[1-F_x\left(z\right)\right]^{n-1}\cdot f_x\left(z\right)$.

דוגמה (הפתרון בהקלטה):

$X_i\sim\exp\left(\lambda\right)$.

הוכיחו כי: $\min\left(x_i\right)\sim\exp\left(n\lambda\right)$, $i=1,2,...n$.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) ענו על הסעיפים הבאים:

א. הוכיחו שאם X_i מתפלג רציף עבור כל: $i=1,2,...,n$ באופן בלתי תלוי עם פונקציית צפיפות $f(x)$ ו- $Z=\min(X_1,X_2...X_n)$, מתקיים ש: $f(z)=n\left[1-F_X(z)\right]^{n-1}\cdot f_x(z)$.

ב. הוכיחו שאם X_i מתפלג רציף עבור כל: $i=1,2,...,n$ באופן בלתי תלוי עם פונקציית צפיפות $f(x)$ ו- $U=\max(x_1,x_2,...x_n)$, מתקיים ש: $f(u)=n\cdot\left(F_x(u)\right)^{n-1}\cdot f_x(u)$.

2) אורך חיי רכיב מתפלג מעריכית עם תוחלת של 30 יום.

א. מכשיר בנוי מ-3 רכיבים בלתי תלויים המחוברים במקביל. בנו את פונקצית ההתפלגות המצטברת של אורך חיי מכשיר.

ב. חזרו על סעיף א' אם הרכיבים מחוברים בטור.

ג. מה התוחלת והשונות של אורך חיי המכשיר המתואר בסעיף ב'?

3) בכיתה 30 תלמידים, כל תלמיד נרדם תוך זמן המתפלג אקספוננציאלית עם קצב של 8 הירדמויות בשעה. המורה צועק אחרי שנרדם התלמיד הראשון ועוזב את הכיתה שנרדם התלמיד האחרון.

א. מה הסיכוי שיצעק אחרי פחות מדקה?

ב. מה הסיכוי שיצא מהכיתה אחרי פחות מדקה?

4) 3 אנשים משתתפים בתחרות ריצה ל-100 מטרים. כל אחד מהם רץ את המרחק בזמן שהוא משתנה מקרי בעל התפלגות אחידה בתחום בין 10 ל-12 שניות.

א. מה הסיכוי שהמנצח סיים את הריצה בזמן הגבוה מ-10.5 שניות?

ב. מה הסיכוי שהמפסיד סיים את הריצה בזמן הנמוך מ-11.2 שניות?

ג. מהי התפלגות זמן הריצה של המפסיד בתחרות? מצאו את התוחלת והשונות שלו?

5) X_1,X_2 מתפלגים נורמאלית סטנדרטית.

נגדיר את: $Z=\min(X_1,X_2)$ ואת: $Y=\max(X_1,X_2)$.

א. חשבו $P(z>1)$.

ב. חשבו $P(Y>1)$.

ג. חשבו $P(Y>1/Y>0)$.

© כל הזכויות שמורות לברק קנדל

6) רונית נכנסת למכון יופי. היא מבצעת טיפול פדיקור ומניקור בו זמנית. משך זמן הפדיקור מתפלג מעריכית עם תוחלת של 20 דקות ומשך זמן המניקור מתפלג מעריכית עם תוחלת של 15 דקות. נניח שאין תלות במשך זמן הטיפול של המניקור והפדיקור.

א. מצאו את ההסתברות שמשך זמן הטיפול לא יעלה על שעה.

ב. ידוע שמשך זמן טיפול הפדיקור עלה על 10 דקות. מה ההסתברות שמשך זמן בטיפול במכון היופי לא יעלה על 20 דקות?

7) נתון ש: $X \sim \exp(\lambda)$ ו- $Y \sim \exp(\mu)$. $U = \min(x, y)$ כמו x, y בלתי תלויים.

הוכיחו כי: $U \sim \exp(\mu + \lambda)$.

8) X_1 ו- X_2 שני משתנים מקריים רציפים בלתי תלויים המתפלגים אחיד בין 0 ל-1.

נגדיר: $Y = \max(x_1, x_2)$.

חשבו את: $P(Y > 0.5)$.

9) נתון ש- $X_i \sim U(0,2)$ בלתי תלויים זה בזה כאשר: $i = 1, 2, \ldots, 5$.

מצאו את פונקציית הצפיפות של: $T = \max(X_i)$.

10) נתון משתנה מקרי X בעל פונקציית הצפיפות הבאה:

$$f(x) = \begin{cases} 3x^2 & o \le x \le 1 \\ 0 & \text{אחר} \end{cases}.$$

נגדיר את: $W = \max(X_i)$ כאשר: $i = 1, 2, \ldots, 10$.

חשבו את $E(W)$.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) הוכחה.

2) א. $F_u(t)=\left(1-e^{-\frac{1}{30}t}\right)^3$. ב. $Z\sim \exp\left(\lambda=\frac{1}{10}\right)$.

ג. תוחלת: 10, שונות: 100.

3) א. 0.9817. ב. 0.

4) א. 0.421875. ב. 0.216. ג. תוחלת: 11.5, שונות: 0.15.

5) א. 0.02518. ב. 0.2922. ג. 0.3896.

6) א. 0.9328. ב. 0.2898.

7) הוכחה.

8) 0.75.

9) $\frac{5}{2}\cdot\left(\frac{t}{2}\right)^4$.

10) $\frac{30}{31}$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 29 - שאלות בהסתברות על מערכות

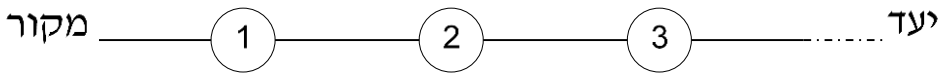
תוכן העניינים

1. כללי 122

מערכות חשמליות:

רקע:

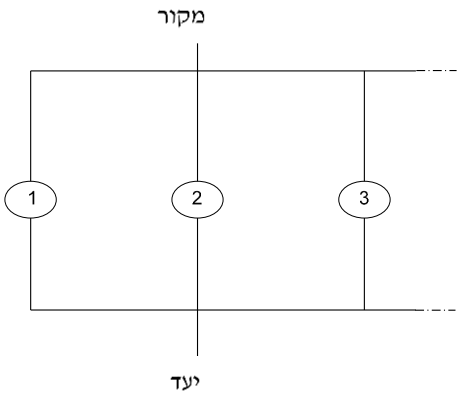
מערכת חשמלית בטור הינה מערכת חשמלית בה הרכיבים מסודרים באופן הבא:



נסמן ב- A_i את המאורע: רכיב i פועל.

כדי שהמערכת כולה תפעל צריך להתקיים ש: $\bigcap_{i=1}^{n} A_i$.

מערכת חשמלית במקביל הינה מערכת חשמלית בה הרכיבים מסודרים באופן הבא:



כדי שהמערכת החשמלית כולה תפעל צריך להתקיים ש: $\bigcup_{i=1}^{n} A_i$.

דוגמה (הפתרון בהקלטה):

במערכת חשמלית 4 רכיבים בלתי תלויים שלכל אחד מהם סיכוי P לפעול.
בטאו באמצעות P את הסיכוי שהמערכת תפעל.

א. כל הרכיבים מחוברים בטור זה לזה.

ב. כל הרכיבים מחוברים במקביל זה לזה.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) נתונים שלושה רכיבים חשמליים המחוברים בטור.
אורך החיים של כל מכשיר מתפלג באופן הבא:

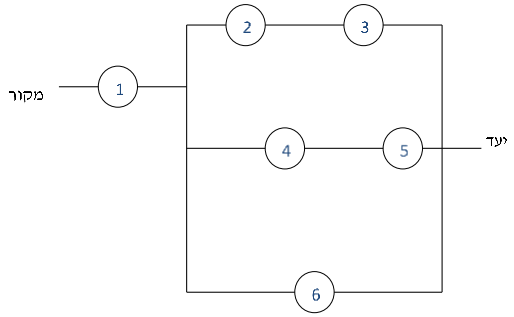
$X_1 \sim U(2,4)$.

$X_2 \sim N(3,1)$.

$X_3 \sim \exp(1)$.

כל רכיב פועל באופן בלתי תלוי זה לזה. כל הרכיבים הופעלו כעת. מה הסיכוי שבעוד 3 שעות המערכת תפעל?

2) המערכת החשמלית הבאה מכילה 6 רכיבים כמוראה בשרטוט:

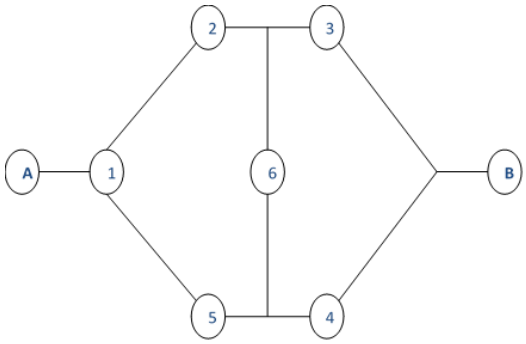


כל רכיב פועל באופן בלתי תלוי זה לזה. רכיבים מספר 1, 2, 6 פועלים בסיכוי 0.9. רכיב מספר 3 פועל בסיכוי 0.8. רכיבים מספר 4, 5 פועלים בסיכוי P.
מצאו את P, אם הסיכוי שהמערכת תפעל הוא 0.887148.

3) בין שני המחשבים A ו-B נמצאים 6 שרתים כמוראה בשרטוט. כל אחד מהשרתים תקין בסיכוי 0.9. על מנת שהודעה תצליח לעבור ממחשב A ל-B צריך להיות לפחות מסלול אחד שבו כל השרתים תקינים.

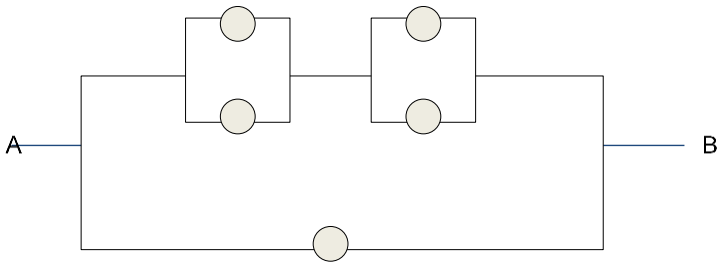
א. מה ההסתברות לכך שההודעה תעבור בהצלחה ממחשב A ל-B?

ב. ההודעה לא הצליחה לעבור ממחשב A למחשב B.
מה הסיכוי ששרת מספר 1 לא תקין?



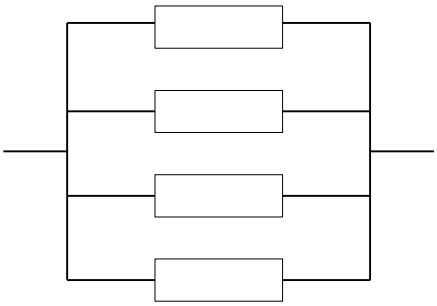
© כל הזכויות שמורות לברק קנדל

4) נתונה המערכת החשמלית הבאה:



כל יחידה עובדת באופן בלתי תלוי ובהסתברות P.
כדי שהמערכת תפעל צריך לעבור זרם מהנקודה A לנקודה B.
הוכיחו שהסיכוי שהמערכת תפעל הוא: $P+(1-P)\left(2P-P^2\right)^2$.

5) מערכת חשמלית כוללת 4 רכיבים אלקטרוניים זהים הפועלים במקביל כמוראה 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. נרצה 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) 0.1245.

2) 0.7.

3) א. 0.880632 ב. 0.837745.

4) שאלת הוכחה.

5) א. 0.8403 ב. $0.0588A > K$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 30 - סטטיסטיקה תיאורית-הקדמה

תוכן העניינים

1. כללי 125

סטטיסטיקה תיאורית – הקדמה:

רקע:

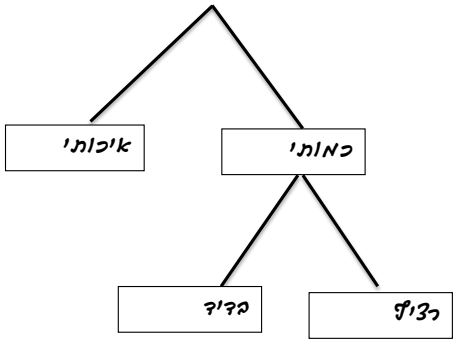
בסטטיסטיקה תיאורית אנו חוקרים קבוצה מסוימת, שיכולה להיות קבוצת ילדים בגן, קבוצת מניות בתיק, כלל התושבים בעיר מסוימת וכולי.
בין ישות לישות בקבוצה ישנם גורמים היכולים לקבל מספר ערכים. גורמים אלה נקראים משתנים. למשל, בין מניה למניה בתיק משתנה התשואה היומית של המניה, הוותק של המניה, תחום המניה וכדומה.
בסטטיסטיקה 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).

© כל הזכויות שמורות לברק קנדל

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 בקפיצות 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.

© כל הזכויות שמורות לברק קנדל

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-₪.

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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 31 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי 128

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: 4, 5, 3, 4, 3.

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	$\frac{f_x}{N}\cdot 100$
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.

© כל הזכויות שמורות לברק קנדל

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:

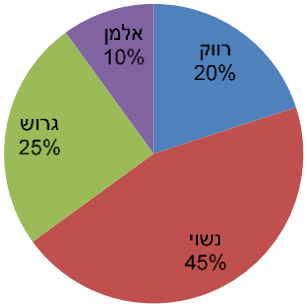
זמן בדקות	מספר הילדים
0.5-3.5	20
3.5-9.5	18
9.5-19.5	14
19.5-29.5	8

© כל הזכויות שמורות לברק קנדל

דיאגרמת עוגה:

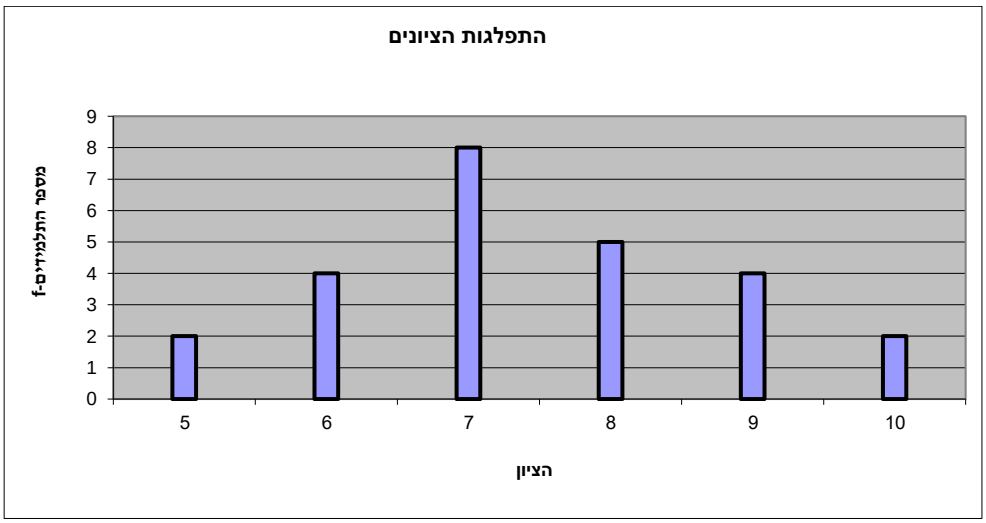
זהו התיאור הגרפי של משתנה איכותי. בדיאגראמת עוגה כל ערך במשתנה מקבל "נתח", שהוא פרופורציונלי לשכיחות היחסית של ערך המשתנה בנתונים.

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.

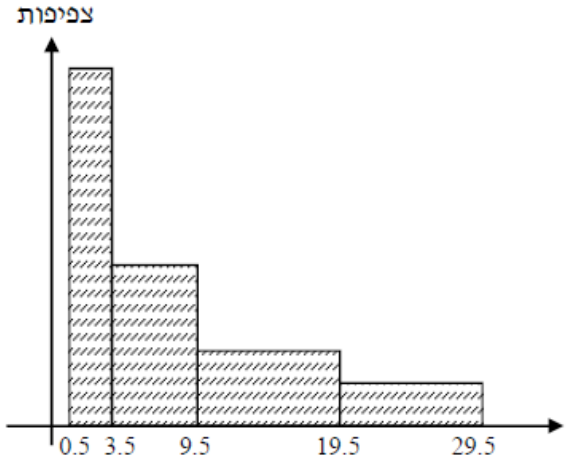


© כל הזכויות שמורות לברק קנדל

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 ליחדה. אם המחלקות הן שוות ברוחב, ניתן לשרטט את ההיסטוגרמה לפי השכיחות ואין צורך בצפיפות.

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



פוליגון – מצולעון:

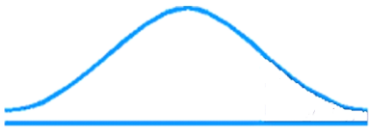
אם נחבר את אמצע קצה כל מלבן בקווים ישרים.
נותן מראה חזותי לצורה של התפלגות המשתנה.

© כל הזכויות שמורות לברק קנדל

צורות התפלגות נפוצות:

התפלגות סימטרית פעמונית

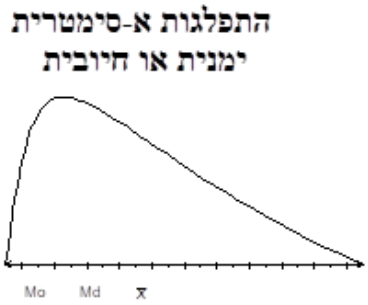
רוב התצפיות במרכז, וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. לדוגמה, ציוני IQ.



ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

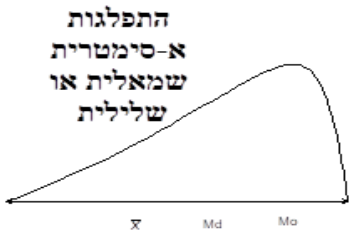
התפלגות אסימטרית ימנית (חיובית)

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



התפלגות אסימטרית שמאלית (שלילית)

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



© כל הזכויות שמורות לברק קנדל

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, 9, 8, 6, 7, 6, 7, 8, 5, 4, 6, 10, 9, 8, 6, 7.

א. מהו המשתנה? האם הוא בדיד או רציף?

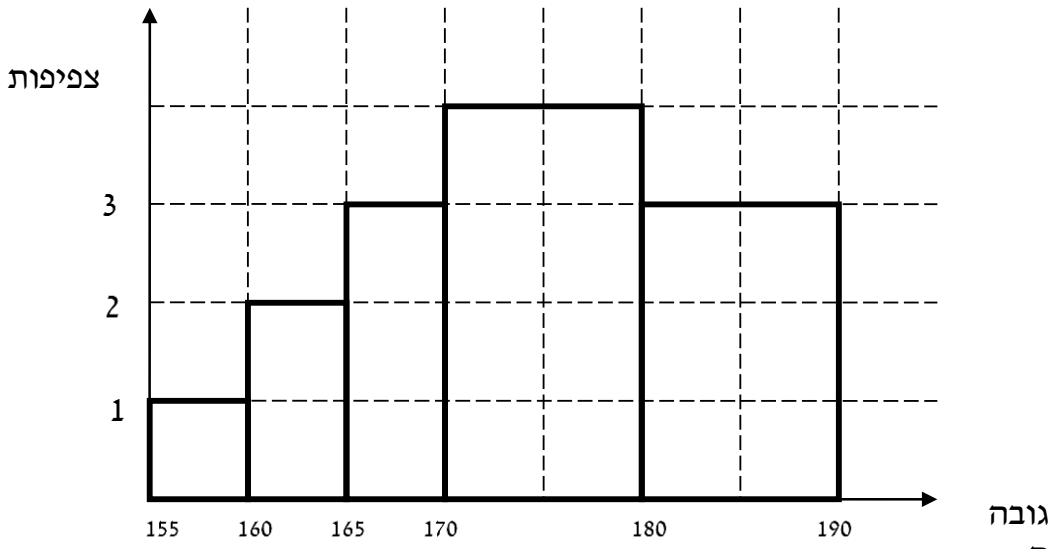
ב. תארו את הרשימה בטבלת שכיחויות.

ג. הוסיפו שכיחויות יחסיות לטבלה.

ד. תארו את הנתונים באופן גרפי.

© כל הזכויות שמורות לברק קנדל

5) להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת:



א. מהו המשתנה הנחקר? האם הוא בדיד או רציף?
ב. תארו את הנתונים בטבלת שכיחויות במחלקות.
ג. הוסיפו שכיחות יחסית לטבלה.
ד. הוסיפו את הצפיפות של כל מחלקה לטבלה.
ה. מהי צורת ההתפלגות של הגבהים?

6) להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

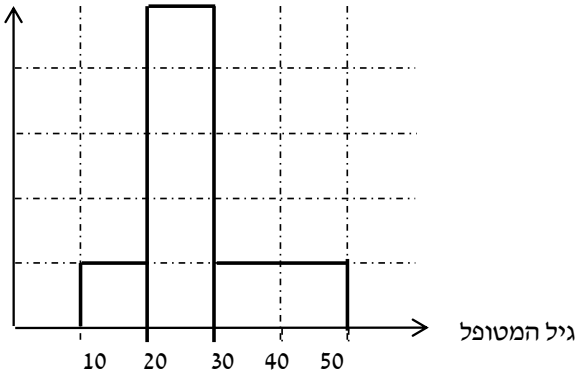
משקל	מספר מקרים
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

א. תארו את ההתפלגות באופן גרפי.
ב. מה ניתן להגיד על צורת ההתפלגות?

© כל הזכויות שמורות לברק קנדל

7) להלן גיל המטופלים של ד"ר שוורץ בשנים:
קנה מידה:

8 מטופלים=



א. מה המשתנה הנחקר? האם הוא בדיד או רציף?
ב. מהי הקבוצה הנחקרת?
ג. תרגמו את ההסיטוגרמה לטבלת שכיחות.
ד. מהי הפרופורציה של המטופלים של ד"ר שוורץ בגילאים 20-30?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. להלן טבלה:

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%

ב. עיין גרף מלא בסרטון הוידאו.

2) א. מקצוע מועדף.

ב. 19.6%.

3) א. משתנה נחקר: השכלה, סוג: סדר.

ב. עיין גרף מלא בסרטון הוידאו.

© כל הזכויות שמורות לברק קנדל

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%

5) א. גובה בס"מ, רציף.

ב+ג+ד. להלן טבלה:

ה. אסימטרית.

x	$f(x)$	$\frac{f(x)}{n}$	%	d
155-160	5	$\frac{5}{100}$	5%	1
160-165	10	$\frac{10}{100}$	10%	2
165-170	15	$\frac{15}{100}$	15%	3
170-180	40	$\frac{40}{100}$	40%	4
180-190	30	$\frac{30}{100}$	30%	3

© כל הזכויות שמורות לברק קנדל

6) א. עיין גרף מלא בסרטון הוידאו.
ב. סימטרית.

7) א. המשתנה: גיל בשנים, משתנה רציף.
ב. המטופלים של ד"ר שוורץ.
ד. להלן טבלה:
ה. 62.5%.

x	$f(x)$
10-20	8
20-30	40
30-50	16

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 32 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי 139

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).

© כל הזכויות שמורות לברק קנדל

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. $\sum(X_i)\sum(Y_i)$.

© כל הזכויות שמורות לברק קנדל

2) נתון לוח ערכי המשתנים X_i ו- Y_i, כאשר: $i=1,2,...,6$, 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: $\sum_{i=1}^{10} X_i^2 = 1640$, $\sum_{i=1}^{10} X_i = 80$.

חשבו: $\sum_{i=1}^{10} (X_i - 4)^2$.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 33 - סטטיסטיקה תיאורית- מדדי מרכז

תוכן העניינים

1. כללי ..... (ללא ספר)

סטטיסטיקה ואמינות

פרק 34 - סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן

תוכן העניינים

1. כללי 143

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.

עבור סדרת נתונים: $S_x^{\,2} = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \dfrac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2$.

דוגמאות:

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

עבור טבלת שכיחויות: $S_x^{\,2} = \dfrac{\sum(x-\bar{x})^2 f}{n} = \dfrac{\sum x^2 \cdot f}{n} - \bar{x}^2$.

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**

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

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

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
6, 5, 8, 7, 6, 9, 8, 6, 7, 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.

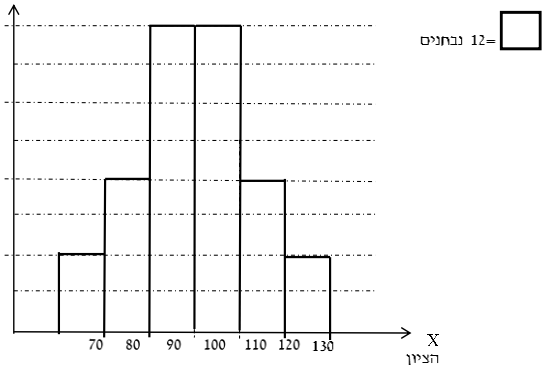
ג. חלק מבעלי הדירות בנות 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע (יקטין, יגדל, לא ישנה) על כל מדד שחישבתם בסעיפים הקודמים.

© כל הזכויות שמורות לברק קנדל

6) להלן התפלגות המשקל של קבוצה מסוימת בק"ג:
מהי סטיית התקן של התפלגות המשקל?

משקל	מספר מקרים
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

7) להלן התפלגות הציונים במבחן אינטליגנציה:



א. מה הממוצע ומה החציון של ההתפלגות?

ב. חשבו את סטיית התקן של הציונים.

ג. מסתבר שיש להוסיף 20 תצפיות לכל אחת משתי המחלקות 90-100 ו- 100-110. כיצד הדבר ישתנה את כל אחד מהמדדים של הסעיפים 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.

6) 7.73.

7) א. 100. ב. 12.96. ג. ממוצע: לא ישתנה, סטיית תקן: תקטן.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 35 - סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני

תוכן העניינים

1. טווח בינרבעוני 146

סטטיסטיקה תיאורית – מדדי פיזור – טווח בין רבעוני:

רקע:

הטווח הבין-רבעוני (יש הקוראים לו התחום הבין-רבעוני) נותן את הטווח בין הרבעונים בו נמצאים 50% מהתצפיות המרכזיות. הרעיון ליצור מדד פיזורי שלא רגיש לתצפיות החריגות ביותר. כדי לחשב את הטווח הבין-רבעוני יש למצוא את הרבעון התחתון והעליון של התפלגות התצפיות.

רבעון תחתון – ערך שמחלק את ההתפלגות לשניים.
25% מהמקרים נמוכים ממנו או שווים לו ו75% מהמקרים גבוהים או שווים לו.
סימון: Q_1.

רבעון עליון – ערך שמחלק את ההתפלגות לשניים.
75% מהמקרים נמוכים ממנו או שווים לו ו25% מהמקרים גבוהים או שווים לו.
סימון: Q_3.

הטווח הבין רבעוני הוא הפער בין שני הרבעונים: $IQR = Q_3 - Q_1$.

שלבים במציאת טווח בין-רבעוני בטבלת שכיחויות:

שלב א: נמצא את הרבעון התחתון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 25%.

שלב ב: נמצא את הרבעון העליון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 75%.

שלב ג: נמצא את הטווח הבין-רבעוני: נחסר את הרבעונים: $IQR = Q_3 - Q_1$.

דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו.
מהו הטווח הבין-רבעוני של מספר תוכניות החיסכון בסניף?

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

© כל הזכויות שמורות לברק קנדל

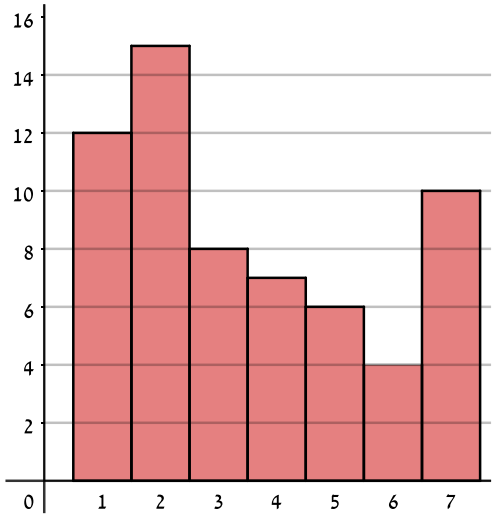
שאלות:

1) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

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

מהו הטווח הבין-רבעוני של מספר המכוניות למשפחה בישוב "הגורן"?

2) בסקר שנעשה בדקו את מספר ימי המחלה השנתיים של מורים בארץ.



א. מה מייצגים הערכים בציר האופקי?

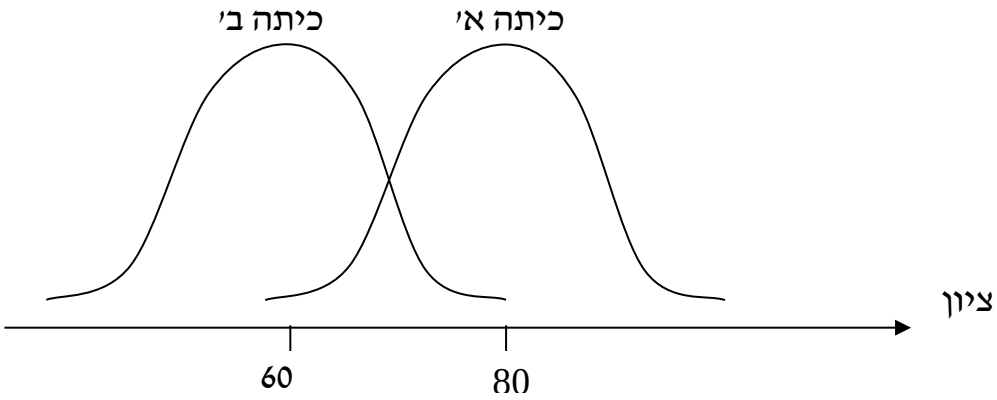
ב. מהו הטווח הבין-רבעוני של מספר ימי המחלה של המורים

ג. אם נוסיף 25 מורים אשר הצהירו שמספר ימי המחלה השנתיים שלהם הוא 4 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.

ד. אם מסתבר שחלק מהמורים בסקר הצהירו שהם חלו 7 ימים בשנה אבל בפועל הם חלו 8 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.

© כל הזכויות שמורות לברק קנדל

3) לפניך שתי עקומות המתארות את התפלגות הציונים בכל כיתה.
באיזו כיתה הטווח הבין-רבעוני גדול יותר?



א. כיתה א.
ב. כיתה ב'.
ג. לשתיהן אותו טווח בין-רבעוני.
ד. לא ניתן לדעת, אין מספיק נתונים.

4) הוספת גודל קבוע לכל תצפיות סדרת נתונים:
א. תגדיל את הטווח הבין-רבעוני.
ב. תקטין את הטווח הבין-רבעוני.
ג. לא תשנה הטווח הבין-רבעוני.
ד. לא ניתן לדעת מה יקרה לטווח הבין-רבעוני.

5) חושב הטווח הבין-רבעוני עבור התפלגות מסוימת והתקבלה התוצאה אפס.
לכן:
א. לפחות 50% מהתצפיות זהות.
ב. סטיית התקן היא אפס.
ג. ההתפלגות היא סימטרית.
ד. מצב זה כלל לא יתכן.

© כל הזכויות שמורות לברק קנדל

6) סניף מספר 543 של בנק "רווה" בדק ל-80 לקוחות את מספר הפעמים שכל לקוח נכנס לסניף הבנק במשך שבוע. התוצאות שהתקבלו הן:
50 אנשים נכנסו 0 פעמים לסניף.
20 אנשים נכנסו פעם אחת לסניף.
5 אנשים נכנסו פעמיים לסניף.
5 אנשים נכנסו יותר מפעמיים.
מהו הטווח הבין-רבעוני?
א. 60.
ב. 2.
ג. 50.
ד. 1.

7) התפלגות הציונים במבחן ווקסלר היא סימטרית לכן:
א. טווח הציונים הוא אפס.
ב. הטווח הבין רבעוני של הציונים אפס.
ג. סעיפים א ו-ב הם נכונים.
ד. אף סעיף אינו נכון.

תשובות סופיות:

1) 2.
2) א. מספר ימי המחלה השנתיים. ב. 3. ג. יקטן. ד. לא ישתנה.
3) ג'.
4) ג'.
5) א'.
6) ד'.
7) ד'.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 36 - סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה

תוכן העניינים

1. כללי 150

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:

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

א. מצאו את האחוזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) להלן התפלגות של משתנה כלשהו:

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

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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 37 - סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן

תוכן העניינים

1. כללי 152

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: $Z = \frac{X - \overline{X}}{S}$.

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. ערך מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים.
מה ערך יותר, באופן יחסי, משכיל או משתכר?

© כל הזכויות שמורות לברק קנדל

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:

	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.
ערן התגייס כשגובהו 180 ס"מ ומשקלו 59 ק"ג.

א. במה ערן חריג יותר ביחס לשאר המתגייסים, גובהו או משקלו?

ב. כמה ערן אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

תשובות סופיות:

1) א. לשון. ב. 72.

2) ב'.

3) א. משקל. ב. 70.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 38 - סטטיסטיקה תיאורית- תרשים קופסא

תוכן העניינים

1. כללי 154

סטטיסטיקה תיאורית – תרשים קופסא (Boxplot):

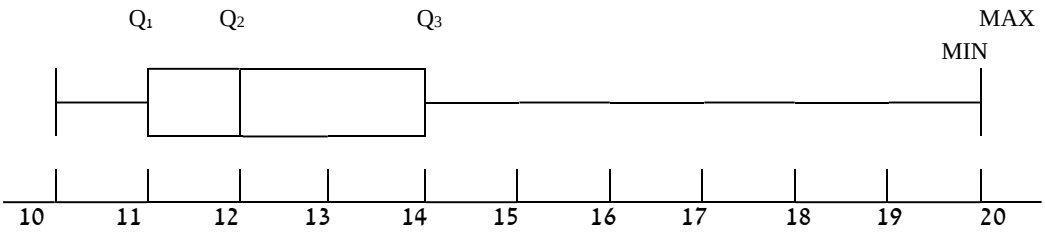
רקע:

תרשים קופסא הינו תרשים שבעזרתו ניתן לבחון :

1) את המרכז של ההתפלגות על ידי החציון (Q_2).

2) את הפיזור של הנתונים (הטווח והטווח הבין רבעוני).

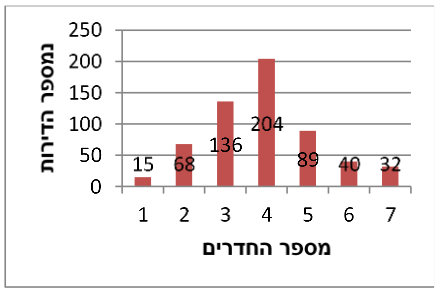
3) את צורת ההתפלגות (סימטרית \אסימטרית ימנית או אסימטרית שמאלית).



© כל הזכויות שמורות לברק קנדל

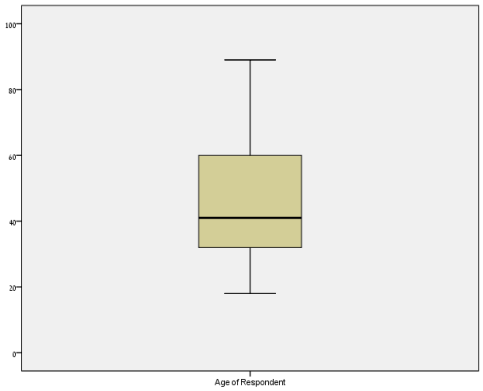
שאלות:

1) להלן התפלגות מספר החדרים לדירות שנבנו בשנת 2009 בעיר אשדוד:



א. מצאו את החציון, הרבעון התחתון והרבעון העליון של ההתפלגות.
ב. שרטטו דיאגרמת קופסא להתפלגות.
ג. מה ניתן לומר על צורת ההתפלגות?

2) להלן דיאגרמת קופסא המתארת את התפלגות הגיל (בשנים) באוכלוסייה מסוימת:



א. מה הגיל החציוני?
ב. מה בערך טווח הגילאים?
ג. מה ניתן להגיד על צורת ההתפלגות?

תשובות סופיות:

1) א. חציון: 4 , רבעון תחתון: 3, רבעון עליון: 5.
ב. ראה גרף מלא בסרטון וידאו. ג. כמעט סימטרית.

2) א. חציון: 40. ב. טווח: 70. ג. התפלגות אסימטרית ימנית.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

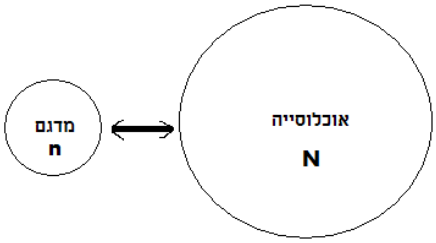
פרק 39 - הסקה סטטיסטית - הקדמה

תוכן העניינים

1. כללי 156

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.

© כל הזכויות שמורות לברק קנדל

הסימונים לפרמטר וסטטיסטי הם שונים:

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

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?

© כל הזכויות שמורות לברק קנדל

שאלות:

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב "העוגן".
נגדיר את X להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.

א. מיהי האוכלוסייה ומהו המשתנה הנחקר?

ב. 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.

א. מהי האוכלוסייה?

ב. מה המשתנה באוכלוסייה?

ג. מהם הפרמטרים?

ד. 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 בישוב 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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 40 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

1. התפלגות ממוצע המדגם ומשפט הגבול המרכזי 159
2. התפלגות סכום תצפיות בלתי תלויות ומשפט הגבול המרכזי 167
3. התפלגות מספר ההצלחות במדגם - קירוב נורמלי להתפלגות הבינומית 170
4. התפלגות פרופורציית ההצלחות במדגם 175

התפלגות ממוצע המדגם ומשפט הגבול המרכזי:

רקע:

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

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

גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים.

להלן רשימה של פרמטרים החשובים לפרק זה:

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

שונות אוכלוסייה נסמן ב- σ^2.

סטיית תקן של אוכלוסייה: σ.

תכונות התפלגות:

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

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

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

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

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

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

דוגמה (פתרון בהקלטה):

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

א. מיהי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטית התקן של ממוצע המדגם?

© כל הזכויות שמורות לברק קנדל

דגימה מהתפלגות נורמאלית:

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

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

דוגמה (פתרון בהקלטה):

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

משפט הגבול המרכזי:

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

דוגמה (פתרון בהקלטה):

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

א. מיהי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו תוחלת ממוצע המדגם?

ו. מהי טעות התקן?

2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב מסוים:

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

א. בנו את פונקציית ההסתברות של X.

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

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

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

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

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

נתון כי ביום מסוים נולדו 4 תינוקות.

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

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

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

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

© כל הזכויות שמורות לברק קנדל

5) הגובה של המתגייסים לצה"ל מתפלג נורמאלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.

א. מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?

ב. מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?

ג. מה ההסתברות שהגובה הממוצע שלהם יסטה מתחולת הגבהים בפחות מ-5 ס"מ?

ד. מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

6) הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים.
לצורך הפתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.

א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?

ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?

ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?

ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?

7) נפח היין בבקבוק מתפלג נורמאלית עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.

א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ"ק?

ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ"ק?

ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ"ק?

ד. בקבוקיי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?

8) משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4.

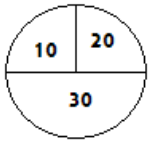
א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?

ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?

ג. הסבר את ההבדל בתשובות של שני הסעיפים.

© כל הזכויות שמורות לברק קנדל

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:

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

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

© כל הזכויות שמורות לברק קנדל

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

א. 0.
ב. 0.5.
ג. 1.
ד. לא ניתן לדעת.

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

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

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

2) א. להלן טבלה:

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

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

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

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

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

5) א. 0. ב. 0. ג. 0.9544. ד. 178.205.

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

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

8) א. 0.5468. ב. 0.6826.

9) א. להלן טבלה:

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

ב. התוחלת: 22.5, השונות: 68.75. ג. התוחלת: 22.5, השונות: 13.75.
ד. 0.8997.

10) 0.0475.

11) 0.1814.

12) א. 0.9772. ב. 0.0228. ג. 271.

13) 0.5.

14) ב'.

15) ד'.

16) ג'.

17) א. 2.429. ב. 0.25.

18) 0.6826.

© כל הזכויות שמורות לברק קנדל

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: $X_1,\ldots,X_n$ -

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

משפט הגבול המרכזי:

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

אזי עבור מדגם מספיק גדול (לפחות 30): $T\sim\succ N\left(n\mu,n\sigma^2\right)$.

דוגמה (פתרון בהקלטה):

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

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

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

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

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

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

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

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

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

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

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

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

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

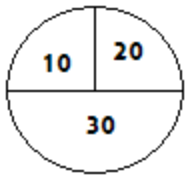
נניח שמספר המכוניות ליחידת דיור בלתי תלויות זו בזו ועם אותה פונקציית הסתברות לכל יחידת דיור (אין צורך בתיקון רציפות).

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

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

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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.

תצפית
הצלחה כשלון

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

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

תוחלת: $E(y)=np$.

שונות: $V(y)=npq$.

קירוב נורמלי עבור התפלגות בינומית:

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

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

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

אז: $Z_y = \frac{y-np}{\sqrt{npq}}$.

© כל הזכויות שמורות לברק קנדל

תיקון רציפות:

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

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

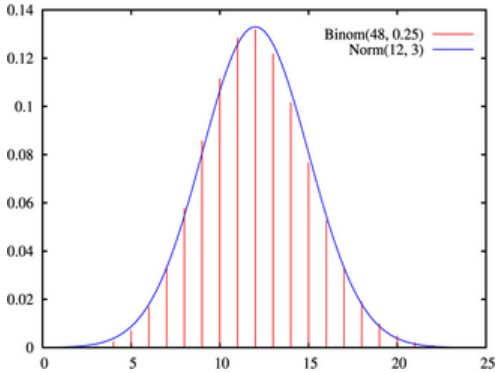
הערות:

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

© כל הזכויות שמורות לברק קנדל

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?



© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

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

2) א. 0.9332. ב. 0.6915.

3) 0.1611.

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

5) א. 0.015. ב. 398.

6) 0.9996.

7) 0.8643.

© כל הזכויות שמורות לברק קנדל

התפלגות פרופורציית ההצלחות במדגם:

רקע:

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

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

למשל, שיעור המובטלים במדגם - $n=200$:
מספר המובטלים: $Y=20$.
פרופורציית המובטלים במדגם: $\hat{p}=\frac{20}{200}=0.1$.

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:

$E(\hat{p})=p \;,\; V(\hat{P})=\frac{pq}{n}$.

משפט הגבול המרכזי עבור הפרופורציה המדגמית:

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

הערות:

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

© כל הזכויות שמורות לברק קנדל

- 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.

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

ב. מה ההסתברות שפרופורציית הזכאים לבגרות במדגם תעלה על 70%?

© כל הזכויות שמורות לברק קנדל

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 היות 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 סמאטרפון?

ג. מה ההסתברות שפרופורציית בעלי הסמרטפון 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?

© כל הזכויות שמורות לברק קנדל

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

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

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. 481. ד. 0.77436.

5) 0.0154.

6) א. שאלת הוכחה. ב. 0.5.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 41 - מושגי יסוד באמידה

תוכן העניינים

1. כללי 179

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-$\hat{\theta}$. $\hat{\theta}$ הוא סטטיסטי המחושב על המדגם ובאמצעותו נאמוד את θ.
- שגיאת אמידה: $\left|\hat{\theta}-\theta\right|$ - 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: $E\left(\hat{\theta}\right)=\theta$.
- טעות התקן של אומד היא סטיית התקן שלו, כלומר: $\sigma\left(\hat{\theta}\right)=S.E$

© כל הזכויות שמורות לברק קנדל

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 אומר 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 אומר 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: 3, 1, 2, 5, 4, 1, 2, 3, 1, 2.
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.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) מתוך 500 טירונים, נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.

א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?

ב. מהי טעות התקן של האומד כשהמדגם בגודל 500?

ג. מהו האומדן לפרמטר?

ד. 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.

א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?

ב. מהי טעות התקן של האומד?

ג. מהו האומדן לפרמטר?

ד. 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.

א. מצאו אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.

ב. מצאו אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.

ג. 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 : $\sum_{i=1}^{20} X_i = 162$, $\sum_{i=1}^{20} X^2_i = 1502.2$.

א. אמדו את השכר הממוצע של השכירים במשק.

ב. 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?

א. סטיית התקן של האוכלוסייה.

ב. סטיית התקן של ממוצע האוכלוסייה.

ג. סטיית התקן של המדגם.

ד. סטיית התקן של ממוצע המדגם.

© כל הזכויות שמורות לברק קנדל

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) $X_1, X_2, ..., X_{16}$ 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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. 0.25 ב. 0.019 ג. 0.24 ד. 0.01.
2) א. אוכלוסייה: מקררים של יצרן, תוחלת: 2400, סטיית תקן: 500.
ב. 100. ג. 2342 ד. 58.
3) א. 177.9 ב.64.1. ג.0.4.
4) א. 8.1 ב.3.16.
5) ד'.
6) ג'.
7) ד'.
8) ד'.
9) ג'.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 42 - רווח סמך לתוחלת (ממוצע)

תוכן העניינים

1. רווח סמך כששונות האוכלוסיה ידועה 184
2. קביעת גודל מדגם 189
3. רווח סמך כששונות האוכלוסיה לא ידועה 191

רווח סמך כששונות האוכלוסייה ידועה:

רקע:

ממוצע המדגם הוא אומד לממוצע האוכלוסייה, אך לא באמת ניתן להבין ממנו על גודלו של ממוצע האוכלוסייה. ההסתברות שממוצע המדגם יהיה בדיוק כמו הממוצע האמתי 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: $P(A \le \mu \le B)=1-\alpha$.
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?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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 ₪.

א. מי האוכלוסייה במחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר שאותו רוצים לאמוד?

ד. מה רווח הסמך לפרמטר?

ה. מהי רמת הסמך לפרמטר?

ו. מה אורך רווח הסמך?

ז. 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 נורמאלית עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.

א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.

ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.

ג. 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 ₪.

א. בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.

ב. מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?

ג. מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?

ד. אם היינו מגדילים את גודל המדגם ובונים רווח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרווח הסמך מכיל את הפרמטר?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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

א. $\mu = 51$.

ב. $\overline{X} = 6$.

ג. $\overline{X} = 51$.

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-₪. 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 ₪ בבטחון של 95%.
ג. 800. ד. לא.

5) א. 102. ב. 3. ג. 0.9544.

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

7) א. 87. ב. 5. ג. 0.9544.

8) ב'.

9) ג'.

10) ד'.

© כל הזכויות שמורות לברק קנדל

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

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 האמתי לא יעלה על 0.5 ק"ג? (תשובה: 87).

© כל הזכויות שמורות לברק קנדל

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.

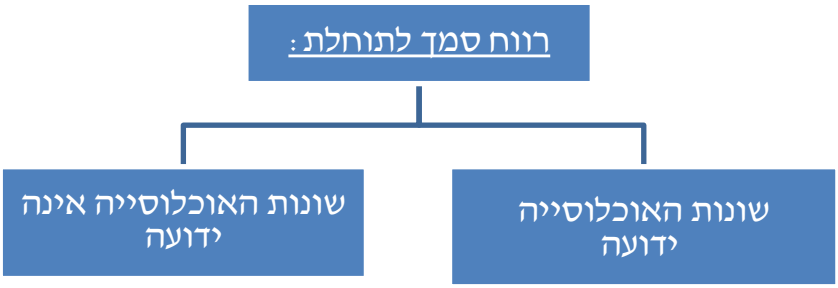
3) $n=62$.

© כל הזכויות שמורות לברק קנדל

רווח סמך כששונות האוכלוסייה לא ידועה:

רקע:

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

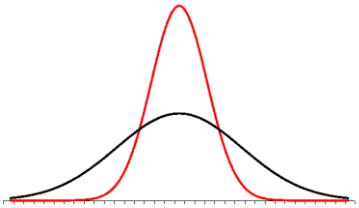


בפרק זה נעסוק במקרה ששונות האוכלוסייה $\left(\sigma^2\right)$ אינה ידועה לנו.

מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול.

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



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

התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים.

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

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה.

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

דוגמה (פתרון בהקלטה):

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי כיתות ח' נורמאלית.
במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח'. להלן התוצאות שהתקבלו בדקות: 5.3, 4.6, 5.2, 4.7.
בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) מחקר מעוניין לדעת כיצד תרופה מסוימת משפיעה על קצב פעימות הלב.
ל-5 אנשים שנטלו את התרופה מדדו את הדופק והתקבל מספר פעימות לדקה: 84, 88, 84, 79, 89.
הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.

א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הנ"ל.

ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?

ג. בהמשך לסעיף א', אם היינו בונים את רווח הסמך ברמת ביטחון של 99%, כיצד הדבר היה משפיע על רווח הסמך?

2) במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן: $S=13$ ס"מ.
בנו רווח סמך ברמת סמך של 90% לתוחלת גובה המתגייסים לצבא האמריקאי. מה יש להניח לצורך פתרון?

3) אדם מעוניין לאמוד את זמן הנסיעה הממוצע שלו לעבודה. לצורך כך הוא דוגם 5 ימים שזמן הנסיעה בהם בדקות הוא: 30, 40, 32, 34, 27.

א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?

ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים?

4) ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.

א. בנו רווח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.

ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמתית של כלל הנבחנים.

ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.

5) נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X^2_i = 643.19$.
בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

© כל הזכויות שמורות לברק קנדל

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 $\sigma=20$ 2.
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: $X \sim N\left(\mu,\sigma^2\right)$ 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%.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 43 - רווח סמך לפרופורציה

תוכן העניינים

1. רווח הסמך לפרופורציה 194
2. קביעת גודל מדגם 197

רווח הסמך לפרופורציה:

רקע:

המטרה היא לאמוד את P – פרופורציה באוכלוסייה.

האומד הנקודתי:

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

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).

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

מתקיים ש: $\hat{P}=\frac{A+B}{2}$, $L=2\varepsilon$.

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?

© כל הזכויות שמורות לברק קנדל

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 הינו 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?

© כל הזכויות שמורות לברק קנדל

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. $0.545 \le p \le 0.655$. 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.

© כל הזכויות שמורות לברק קנדל

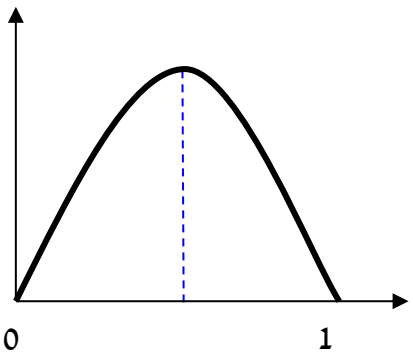
קביעת גודל מדגם:

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.

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

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%.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) הממשלה אומדת מדי חודש את אחוז התמיכה בה.
מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמתי 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 האמתי לא תעלה על 4%.

א. כמה מכשירי PEOPLE METER יש להתקין לצורך האמידה?

ב. לפי הערכה מוקדמת הרייטינג של הערוץ לא יכול לעלות על 20%.
בהנחה ומכשיר כזה עולה 500 ₪ ליחידה מה החיסכון הכספי מאינפורמציה זאת?

4) ענו על הסעיפים הבאים:

א. כמה אזרחים יש לדגום כדי לאמוד את אחוז התמיכה בממשלה עם אורך רווח הסמך שלא עולה על 9% ברמת סמך של 90%?

ב. בהנחה ובוצע מדגם שאת גודלו חישבתם בסעיף א והתקבל שאחוז התמיכה בממשלה במדגם הנו 42%. בנו רווח סמך לאחוז התמיכה בממשלה ברמת סמך של 95%.

ג. על סמך סעיף ב', האם תקבלו את הטענה שמיעוט האוכלוסייה תומך הממשלה?

5) משרד הבריאות מתכנן לבצע מדגם שמטרתו לבדוק את הסיכוי לחלות בשפעת עם לקיחת חיסון נגד שפעת. הוא מעוניין שבסיכוי של 98% טעות האמידה לא תעלה על 3%.

א. כמה מחוסנים יש לדגום?

ב. משרד הבריאות ביצע את המדגם שאת גודלו חישבת 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%?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) 1068.

2) א. 423. ב. 271.

3) א. 601. ב. 108,000₪.

4) א. 335. ב. $0.367 < p < 0.473$.

ג. בביטחון של 0.95 ניתן להגיד שמיעוט באוכלוסייה תומך בממשלה.

5) א. 1509. ב. 0.15 ± 0.02 ג. ראה סרטון.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 44 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסיה לא ידועות ובהנחת שוויון שונויות..... 200

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. $\sigma^2{}_1 = \sigma^2{}_2$.
2. $X_1, X_2 \sim N$.
3. מדגמים בלתי תלויים.

השונות המשוקללת: כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השונות הזו על ידי שקלול שתי השונויות של שני המדגמים על ידי הנוסחה הבאה: $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$.

דרגות החופש: $d.f = n_1 + n_2 - 2$.

רווח סמך: $(\bar{x}_1 - \bar{x}_2) \pm t_{1-\frac{\alpha}{2}}^{n_1+n_2-2} \cdot \sqrt{\frac{S^2{}_p}{n_1} + \frac{S^2{}_p}{n_2}}$.

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:

	תל אביב	באר שבע
מספר האקדמאים	20	10
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.

© כל הזכויות שמורות לברק קנדל

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: $N\left(\mu_x,\sigma^2\right)$, ומשתנה Y שמתפלג: $N\left(\mu_y,\sigma^2\right)$.

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:
1. השונות שווה.
2. שהציונים מתפלגים נורמלית.
3. המדגמים אינם תלויים זה בזה.

$-5.52<\mu_1-\mu_2<17.52$.

2) $-9.6<\mu_y-\mu_x<1.6$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 45 - רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים

תוכן העניינים

1. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 202

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. $x, y \sim N$.
2. המדגם מזווג.

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

כאשר דרגות החופש: $df = n-1$.

דוגמה (פתרון בהקלטה):

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

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

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

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) $-0.013 < \mu < 0.185$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 46 - שאלות מסכמות על רווחי סמך

תוכן העניינים

1. שאלות מסכמות על רווחי סמך 204

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:

$$\sum_{i=1}^{20} x_i = 2040 \ , \ \sum_{i=1}^{20} x_i^2 = 210740 .$$

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?

© כל הזכויות שמורות לברק קנדל

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%.

תשובות סופיות:

2) א. $1.21 \le \mu \le 1.65$. ב. $10.85\% \le p \le 19.15\%$.

3) א. 70. ב. 0.9988. ג. $83\% < p < 89\%$.

4) א. $97.4 \le \mu \le 106.6$. ב. לא. ג. יגדל.

7) א. $-372 \le \mu_A - \mu_B \le 772$, לא. ב. $6467 \le \mu_A \le 7133$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 47 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

1. הקדמה 206
2. סוגי טעויות 210

הקדמה:

רקע:

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

שלב א: נזהה את הפרמטר הנחקר.

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

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

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

1. **אזור דחייה:**
דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.

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

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

שלב ו: להסיק מסקנה בהתאם לתוצאות המדגם.

© כל הזכויות שמורות לברק קנדל

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

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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. $H_0: p = 0.08$
$H_1: p \neq 0.08$.

© כל הזכויות שמורות לברק קנדל

סוגי טעויות:

רקע:

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

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

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

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

הגדרת הטעויות:

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

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

דוגמה (פתרון בהקלטה):

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

א. רשמו את השערות המחקר.

ב. מה מסקנת המחקר?

ג. איזו סוג טעות יתכן וביצעו במחקר?

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

א. האם ניתן לדעת אם בוצע טעות במחקר?

ב. מה סוג הטעות האפשרית?

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

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מה השערות המחקר?

ה. מה מסקנת המחקר?

ו. מהי סוג הטעות האפשרית במחקר?

תשובות סופיות:

1) א. $H_0: \mu = 500$, $H_1: \mu < 500$. ב. לא דחינו את H_0 . ג. טעות מסוג שני.

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

3) א. משפחות כיום. ב. מס' הילדים.

ג. תוחלת מספר הילדים למשפחה כיום. ד. $H_0: \mu = 2.3$, $H_1: \mu < 2.3$.

ה. לא לדחות את H_0 . ו. טעות מסוג שני.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

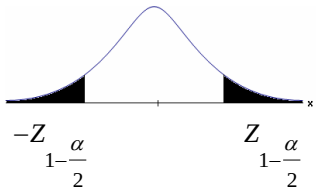
פרק 48 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

1. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה ידועה 212
2. סיכוי לטעויות ועוצמה (ששונות האוכלוסיה ידועה) 216
3. קביעת גודל מדגם (ששונות האוכלוסיה ידועה) 222
4. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה ידועה) 225
5. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה לא ידועה 230
6. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה לא ידועה) 234
7. הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע) 237
8. ניתוח פלטים 239

בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה ידועה:

רקע:

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

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

חלופה אחרת לכלל הכרעה:

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

© כל הזכויות שמורות לברק קנדל

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 שזובלו 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.

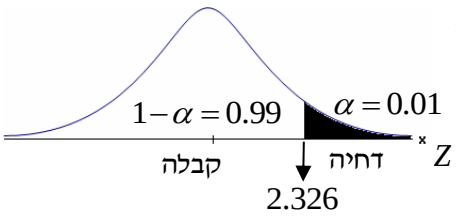
הפרמטר: μ = תוחלת היבול בשיטת הזיבול החדשה.

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

נציב: $Z_{\bar{X}} = \dfrac{1.25 - 10}{\frac{2.5}{\sqrt{4}}} = 2 < 2.326$

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.

© כל הזכויות שמורות לברק קנדל

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 מכוילת (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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

סיכוי לטעויות ועוצמה (ששונות האוכלוסייה ידועה):

רקע:

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

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

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

הסיכוי לבצע טעות מסוג 2:

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

רמת בטחון:

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

עוצמה:

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

© כל הזכויות שמורות לברק קנדל

התהליך לחישוב סיכוי לטעות מסוג שני:

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

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

התקנון: $Z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$.

דוגמה:

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

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

ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?

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

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

פתרון:

א. אוכלוסייה: משלמי חשבון טלפון סלולאר כיום.

המשתנה: X = חשבון הטלפון החודשי בשקלים.

הפרמטר: μ.

השערות: $\begin{matrix} H_0: \mu = 200 \\ H_1: \mu < 200 \end{matrix}$.

תנאים:

1. $\mu = 200$.
2. $n = 36$.

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

$\alpha = 0.05$

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

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

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

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

ג. השערות: $\begin{matrix} H_0: \mu_0 = 200 \\ H_1: \mu < 200 \end{matrix}$.

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

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

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

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

ד. הקטנת α מגדילה את β.

שאלות:

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

להלן השערות של חוקר לגבי הפרמטר μ: $H_0: \mu = 5$, $H_1: \mu = 7$.

מעוניינים ליצור כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.

א. עבור אילו ערכים של X שידגם נדחית השערת H_0?

ב. מה הסיכוי לבצע טעות מסוג שני?

ג. אם במדגם התקבל ש- $X = 6.9$ מה תהיה המסקנה ומה הטעות האפשרית?

2) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. מעוניינים לבדוק אם כיום ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות.
במדגם התקבל ממוצע 2.17 ילדים למשפחה.

א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי ברמת מובהקות של 5%.

ב. בהמשך לסעיף א' מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?

ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העצמה של הכלל מסעיף א'?

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

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

א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.

ב. בהמשך לסעיף א', מהי העצמה אם התוחלת שווה ל-195?

ג. הסבירו, ללא חישוב, איך העצמה תשתנה אם רמת המובהקות תהיה 5%?

4) מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50 מ"מ וסטית תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומודדים את קוטרם, בכדי לבדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוילת כנדרש או שקוטר הצינורות קטן מהדרוש.

א. רשמו את ההשערות ואת כלל ההכרעה ברמת מובהקות של 5%.

ב. אם ביום כלשהו מכונת הייצור התקלקלה והיא מייצרת את הצינורות בקוטר שתוחלתו 48 מ"מ בלבד (סטית התקן לא השתנתה), מה ההסתברות שהתקלה לא תתגלה בביקורת האיכות? כיצד נקראת הסתברות זו?

ג. הסבירו ללא חישוב כיצד התשובה לסעיף ב תשתנה אם רמת המובהקות תגדל.

ד. הסבירו ללא חישוב כיצד התשובה לסעיף ב תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

© כל הזכויות שמורות לברק קנדל

5) להלן השערות של מחקר: $H_0: \mu = 50$, $H_1: \mu = 58$.
מעוניינים לדגום 100 תצפיות. ידוע שסטיית התקן של ההתפלגות הינה 20.

א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10%.
מהי רמת המובהקות?

ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו)?

i. סטיית התקן הייתה יותר גדולה.

ii. הסיכוי לטעות מסוג שני גדול יותר.

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

6) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:

א. הסיכוי לטעות מסוג ראשון גדל.

ב. העוצמה של המבחן גדלה.

ג. הסיכוי לטעות מסוג שני גדל.

ד. תשובות א' ו-ב' נכונות.

7) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן:

א. השערת האפס נכונה.

ב. השערת האפס נדחתה.

ג. השערת האפס לא נדחתה.

ד. אף אחת מהתושבות לא נכונה בהכרח.

8) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

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

9) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית H_0 קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך:

א. הן α, והן $1-\beta$, יקטנו.

ב. α יישאר ללא שינוי ואילו $1-\beta$ יגדל.

ג. α יגדל ואילו $1-\beta$ יקטן.

ד. הן α והן $1-\beta$ יגדלו.

© כל הזכויות שמורות לברק קנדל

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

א. טעות מסוג ראשון.

ב. טעות מסוג שני.

ג. טעות מסוג שלישי.

ד. אין טעות במסקנתו.

תשובות סופיות:

1) א. מעל 6.645. ב. 0.3594.

ג. דחינו את H_0, תתכן טעות מסוג ראשון.

2) א. נדחה H_0 אם $\bar{X} < 2.24$. ב. נדחה H_0. ג. 1.

3) א. נדחה H_0 אם $\bar{X} > 203.29$ או $\bar{X} < 196.71$. ב. 0.8051. ג. תקטן.

4) א. נדחה H_0 אם $\bar{X} < 48.9$. ב. 0.0885. ג. תקטן. ד. תקטן.

5) א. 0.0033. ב. i. רמת המובהקות הייתה קטנה.

ב. ii. רמת המובהקות הייתה גדלה.

6) ד'.

7) ג'.

8) ג'.

9) א'.

10) ב'.

© כל הזכויות שמורות לברק קנדל

קביעת גודל מדגם (ששונות האוכלוסייה ידועה):

רקע:

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

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

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

דוגמה:

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

פתרון:

<u>האוכלוסייה</u>: ילדי גן חובה.

<u>המשתנה</u>: X = ציון במבחן אוצר מילים.

<u>הפרמטר</u>: μ.

<u>השערות</u>: $\begin{matrix}H_0:\mu=70\\H_1:\mu=80\end{matrix}$.

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

אם בהפעלת השיטה ההולנדית התוחלת תעלה ל80, נגלה זאת בסיכוי 90%.

© כל הזכויות שמורות לברק קנדל

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

$$\alpha = 0.05$$

$$1-\beta = 0.9$$

$$\mu_0 = 70$$

$$\mu_1 = 80$$

$$\sigma = 17$$

$$Z_{1-\alpha} = Z_{0.95} = 1.645$$
$$Z_{1-\beta} = Z_{0.9} = 1.282$$

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

לכן, $n_{\min} = 25$.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) במבחן אינטליגנציה הציונים מתפלגים נורמאלית עם סטיית תקן 8 וממוצע 100. פסיכולוג מעוניין לבדוק את הטענה שבאוכלוסיות במצב סוציו אקונומי נמוך תוחלת הציונים היא 95.
אם מעוניינים לגלות את הטענה בהסתברות של לפחות 99% כשרמת המובהקות היא 5% מהו גודל המדגם הדרוש?

2) משרד התקשורת טוענים שאדם מדבר בממוצע 180 דקות בחודש בטלפון הסלולרי. חברות הטלפון הסלולרי טוענות שאינפורמציה זו אינה נכונה ואדם מדבר בממוצע פחות: כ-160 דקות. לצורך פתרון נניח שסטיית התקן של זמן השיחה החודשי ידוע ושווה ל-60 דקות. כמה אנשים יש לדגום כך שאם טענת משרד התקשורת נכונה נדחה אותה בסיכוי של 5% (איך קוראים להסתברות זאת?) כמו כן אם טענת חברות הטלפון הסלולרית נכונה המחקר יגלה זאת בסיכוי של 90% (איך קוראים להסתברות זאת?).

3) השערות המחקר הן: $H_0:\mu=\mu_0$, $H_1:\mu=\mu_1$.
כמו כן נתון שהמשתנה מתפלג נורמלית עם סטיית התקן ידועה σ מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β. הוכיחו שגודל המדגם הרצוי לכך יהיה: $n\geq\left(\frac{\left(Z_{1-\alpha}+Z_{1-\beta}\right)\times\sigma}{\mu_0-\mu_1}\right)^2$.

תשובות סופיות:

1) 41.
2) 78.
3) שאלת הוכחה.

© כל הזכויות שמורות לברק קנדל

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 $p_v \le \alpha$, 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:**	$H_0:\mu=\mu_0$ $H_1:\mu\ne\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu<\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu>\mu_0$
תנאים:	1. σ ידועה 2. $X \sim N$ או מדגם מספיק גדול		
p-value	אם $\bar{x} > \mu_0 \Rightarrow 2\cdot P_{H_0}\left(\bar{X} \ge \bar{x}\right)$ אם $\bar{x} < \mu_0 \Rightarrow 2\cdot P_{H_0}\left(\bar{X} \le \bar{x}\right)$	$P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \ge \bar{x})$

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

© כל הזכויות שמורות לברק קנדל

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. <u>אוכלוסייה</u>: המתגייסים לצבא כיום.

<u>משתנה</u>: X = משקל בק"ג.

<u>פרמטר</u>: μ.

<u>השערות</u>: $\begin{matrix} H_0: \mu = 65 \\ H_1: \mu > 65 \end{matrix}$.

<u>תנאים</u>:

1. $X \sim N$.
2. $\sigma = 12$.

<u>תוצאות מדגם</u>:

$$n = 16$$

$$\bar{X} = 71$$

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

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

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

© כל הזכויות שמורות לברק קנדל

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: $\bar{x} = 74$, $n = 100$.
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 נורמאלית 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.

© כל הזכויות שמורות לברק קנדל

6) בבדיקת השערות התקבל שה- $p-value=0.02$.
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: $p\ value=0.0254$ 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 .

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) 0.0228.

2) עבור כל רמת מובהקות סבירה.

3) א. $\begin{matrix} H_0: \mu = 100 \\ H_1: \mu < 100 \end{matrix}$. 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'.

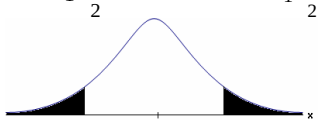
7) א'.

8) ג'.

© כל הזכויות שמורות לברק קנדל

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

רקע:

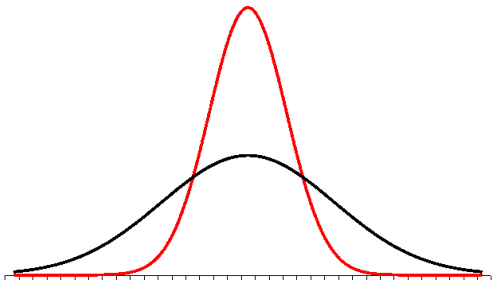
השערת האפס: **השערה אלטרנטיבה:**	$H_0:\mu=\mu_0$ $H_1:\mu\neq\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu<\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu>\mu_0$
תנאים:	1. σ אינה ידועה 2. $X\sim N$ או מדגם מספיק גדול		
כלל ההכרעה: **אזור הדחייה של H_0:**	$t_{\bar{x}}>t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{x}}<-t_{1-\frac{\alpha}{2}}^{(n-1)}$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

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

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

© כל הזכויות שמורות לברק קנדל

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%.

© כל הזכויות שמורות לברק קנדל

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: 125, 80, 100, 95, 90 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:

$$\sum_{i=1}^{10} X_i = 750$$
$$\sum_{i=1}^{10} \left(X_i - \bar{X}\right)^2 = 900$$

נתון שההתפלגות היא נורמלית.
בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80 .

© כל הזכויות שמורות לברק קנדל

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.

א. אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.

ב. אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.

ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.

ד. 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: $\begin{matrix} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{matrix}$.
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'.

© כל הזכויות שמורות לברק קנדל

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

רקע:

נזכיר שהמסקנה של המחקר תיקבע לפי העיקרון הבא: אם $p_v \le \alpha$ דוחים את H_0.
מובהקות התוצאה היא הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.
(לקבל את תוצאות המדגם וקיצוני) $p_v = P_{H_0}$.
אם ההשערה היא דו צדדית:
(לקבל את תוצאות המדגם וקיצוני) $p_v = 2P_{H_0}$.

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס: **השערה אלטרנטיבה:**	$H_0: \mu = \mu_0$ $H_1: \mu \ne \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
תנאים:	1. σ אינה ידועה או 2. מדגם מספיק גדול $X \sim N$		
p-value	אם $\bar{x} > \mu_0 \Rightarrow 2 \cdot P_{H_0}\left(\bar{X} \ge \bar{x}\right)$ אם $\bar{x} < \mu_0 \Rightarrow 2 \cdot P_{H_0}\left(\bar{X} \le \bar{x}\right)$	$P_{H_0}\left(\bar{X} \le \bar{x}\right)$	$P_{H_0}\left(\bar{X} \ge \bar{x}\right)$

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

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

$$d.f = n-1$$

© כל הזכויות שמורות לברק קנדל

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: 27, 32, 30, 40, 34.
הניחו שזמן הנסיעה מתפלג נורמלית.

א. רשמו את השערות המחקר.

ב. מצאו חסמים למובהקות התוצאה.

ג. מה המסקנה ברמת מובהקות של 5%?

פתרון:

<u>אוכלוסייה</u>: כלל הנסיעות לעבודה בדרך החלופית.

<u>משתנה</u>: X = זמן נסיעה בדקות.

<u>תנאים</u>: $X \sim N$.

<u>פרמטר</u>: μ.

א. <u>השערות</u>: $\begin{matrix} H_0: \mu = 40 \\ H_1: \mu < 40 \end{matrix}$.

ב. <u>תוצאות המדגם</u>:

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

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

$$S = \sqrt{23.4}$$

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

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

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

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

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

<u>מסקנה</u>: בר״מ של 5% נכריע שהדרך החלופית מהירה יותר.

© כל הזכויות שמורות לברק קנדל

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: 997, 996, 1005, 1008, 1024.

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:
$\sum(x_i-\bar{x})^2=2832$, $\bar{x}=176.2$.
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. $H_0:\mu=1000$, $H_1:\mu\neq1000$. ב. $20\%\le P_v\le 50\%$.

ג. ברמת מובהקות של 5% לא נוכל לקבוע שקו הייצור אינו תקין.

2) 10%.

3) 1.01, נקבל את H_0 .

© כל הזכויות שמורות לברק קנדל

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 ← נקבל את H_0 .

אם μ_0 לא נופל ברווח ← 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 .

© כל הזכויות שמורות לברק קנדל

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 פנצילין 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. $112.87 \le \mu \le 118.13$.

ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.

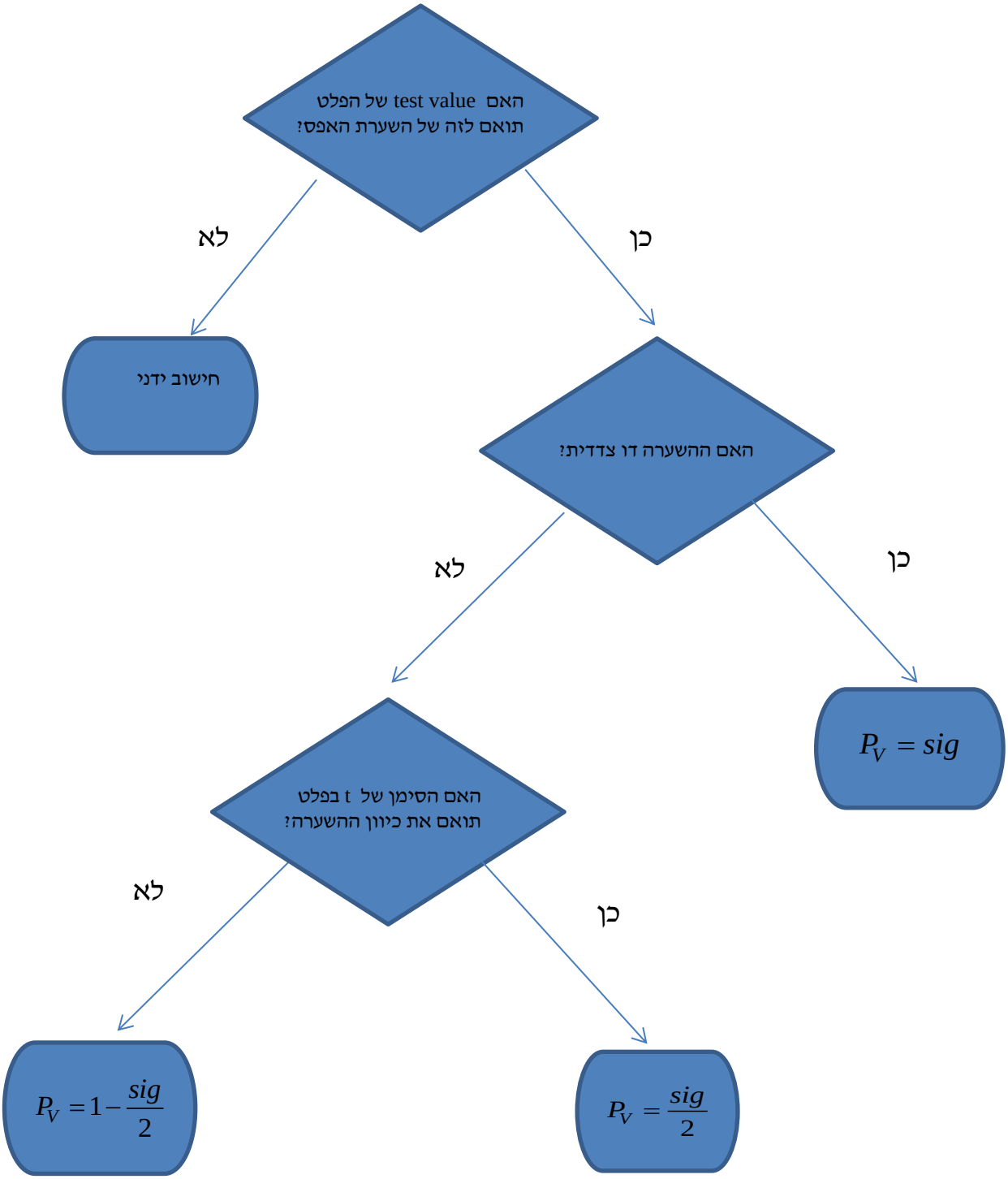
3) א. $191.8 \le \mu \le 200.2$. ב. נכריע שיש אמת בפרסום.

© כל הזכויות שמורות לברק קנדל

ניתוח פלטים:

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS :



דוגמה (פתרון בהקלטה):

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
X	25	87.6400	64.90434	12.98087

One-Sample Test

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

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

א. מהו רווח הסמך לממוצע הציונים בחשבון אם יוחלט ליישם את שיטת הלימוד החדשה?

ב. מהו P_V לבדיקת יעילותה של שיטת הלימוד החדשה?

ג. מה יוכרע ברמת מובהקות של 5% לגבי יעילותה של שיטת הלימוד החדשה?

© כל הזכויות שמורות לברק קנדל

שאלות:

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

One-Sample Statistics

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

One-Sample Test

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

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

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

One-Sample Test

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

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

© כל הזכויות שמורות לברק קנדל

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

One-Sample Statistics

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

One-Sample Test

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

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

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

One-Sample Statistics

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

One-Sample Test

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

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

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. הסקה של תוחלת אחת. ב. 23. ג. (22.104, 24.771).

ד. נקבל H_0.

2) א. 25. ב. (4.3056, 4.5104). ג. המשתנה מתפלג נורמלית.

ד. נדחה H_0.

3) א. $n=44$, $\bar{X}=67.66$, $t=-2.48$, $upper=-1.3735$.

ב. 0.017. ג. 0.0085. ד. 0.9915.

4) א. נכריע שיש אמת בפרסום. ב. המסקנה לא תשתנה.

ג. נכריע שהתוחלת שונה מ-26.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

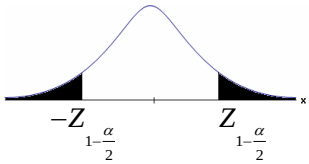
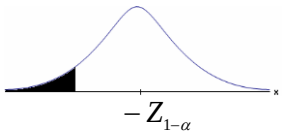
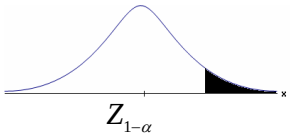
פרק 49 - בדיקת השערות על פרופורציה

תוכן העניינים

1. התהליך 244
2. סיכוי לטעויות ועוצמה 247
3. קביעת גודל מדגם 251
4. מובהקות התוצאה - אלפא מינימלית 253

התהליך:

רקע:

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

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

חלופה אחרת לכלל הכרעה:

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

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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) נדחה H_0.

2) נדחה H_0.

3) נקבל H_0.

4) א. $\begin{matrix} H_0: p=0.2 \\ H_1: p\neq 0.2 \end{matrix}$. ב. נקבל H_0. ג. המסקנה לא תשתנה.

5) א. נדחה H_0. ב. המסקנה לא תשתנה.

6) ג'.

© כל הזכויות שמורות לברק קנדל

סיכוי לטעויות ועוצמה:

רקע:

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

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

הסיכוי לבצע טעות מסוג 2:

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

רמת בטחון:

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

עוצמה:

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

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

התהליך לחישוב סיכוי לטעות מסוג שני:

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

© כל הזכויות שמורות לברק קנדל

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

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

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

דוגמה (פתרון בהקלטה):

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

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

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

© כל הזכויות שמורות לברק קנדל

שאלות:

1) משרד הבריאות פרסם ש-10% מתושבי המדינה סובלים ממחלת האסטמה. מחקר דורש לבדוק האם בחיפה, בגלל זיהום האוויר, שיעור הסובלים מאסטמה גבוה יותר. לצורך המחקר נבדקו 260 מתושבי חיפה.

א. רשמו את השערות המחקר, וצרו מבחן ברמת מובהקות של 5% לבדיקתן.

ב. מהי עצמת המבחן של סעיף א' בהנחה ובחיפה 16% מהתושבים סובלים מאסטמה?

ג. כיצד תשנה התשובה לסעיף ב' אם מסתבר שבחיפה 18% סובלים מאסטמה?

ד. בהמשך לסעיף א' האם נכון לומר שבהסתברות של 5% ההשערה שבחיפה 10% מהתושבים סובלים מאסטמה אינה נכונה?

2) אחוז הסובלים מתופעות הלוואי מתרופה מסוימת הוא 15%. חברת תרופות טוענת שפיתחה תרופה שאמורה לצמצם את אחוז הסובלים מתופעות לוואי. לצורך בדיקת הטענה הוחלט לבצע מחקר שיכלול 120 חולים שיקבלו את התרופה הנבדקת.
נניח שהתרופה נבדקת אכן מורידה את פרופורציות הסובלים מתופעות הלוואי ל-10%, מהי עצמת המבחן עבור רמת מובהקות של 5%?

3) בעיר מסוימת היו 20% אקדמאים. בעקבות פתיחת מכללה בעיר לפני כמה שנים מעוניינים לבדוק האם אחוז האקדמאים גדל. מעוניינים שהמחקר יכלול 200 אנשים והוא יהיה ברמת מובהקות של 5%.

א. חשבו את הסיכוי לבצע טעות מסוג שני בהנחה והיום יש 28% אקדמאים.

ב. כיצד התשובה לסעיף הקודם תשתנה אם נגדיל את רמת המובהקות?

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

5) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן (בחרו בתשובה הנכונה):

א. השערת האפס נכונה.

ב. השערת האפס נדחתה.

ג. השערת האפס לא נדחתה.

ד. אף אחת מהתושבות לא נכונה בהכרח.

6) קבעו אם הטענה הבאה נכונה: בבדיקת השערות לא ניתן לבצע בו זמנית טעות מסוג ראשון וטעות מסוג שני.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. $\begin{matrix} H_0: p=0.1 \\ H_1: p>0.1 \end{matrix}$. ב. 0.9015. ג. תגדל. ד. טענה לא נכונה.

2) 0.4404.

3) א. 0.1446 ב. תקטן.

4) חוקר א'.

5) ג'.

6) נכונה.

© כל הזכויות שמורות לברק קנדל

קביעת גודל מדגם:

רקע:

השערות המחקר הן: $H_0: p = p_0 \quad , \quad H_1: p = p_1$

מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β.

הנוסחה הבאה נותנת את גודל המדגם הרצוי: $n \geq \left(\frac{Z_{1-\alpha}\sqrt{p_0 q_0} + Z_{1-\beta}\sqrt{p_1 q_1}}{p_0 - p_1} \right)^2$.

דוגמה (פתרון בהקלטה):

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

© כל הזכויות שמורות לברק קנדל

שאלות:

1) משרד התמ"ת פרסם שאחוז האבטלה במשק היום עומד על 8%.
לעומתו, משרד הפנים טוען שחלה עלייה בשיעור האבטלה עד לכדי 11%.
כדי לבדוק מי מבניהם צודק, מה צריך להיות גודל המדגם שיענה על שני התנאים הבאים:

א. אם משרד התמ"ת צודק, נדחה את טענתו בסיכוי של 10%.

ב. אם משרד הפנים צודק, נדחה את טענתו בסיכוי של 4%.

2) מפעיל קזינו מפרסם שהסיכוי לזכות במכונת מזל הינו 0.42.
אדם טוען שהסיכויים לזכות במשחק נמוכים יותר.
כמה פעמים יש לשחק את המשחק כדי שאם טענת מפעיל הקזינו נכונה נקבל את טענת האדם בסיכוי של 1% ואם במציאות הסיכוי לזכות במכונה הוא 0.3 נקבל את מפעיל הקזינו בסיכוי של 8%?

תשובות סופיות:

1) 891.

2) 224.

© כל הזכויות שמורות לברק קנדל

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 $p_v \le \alpha$ 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:	$H_0: p = p_0$ $H_1: p \ne p_0$	$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p > p_0$
תנאים:	$np_0 \ge 5 \,\&\, n(1-p_0) \ge 5$		
p-value	אם $\hat{p} > p_0 \Rightarrow 2 \cdot P_{H_0}(\hat{P} \ge \hat{p})$ אם $\hat{p} < p_0 \Rightarrow 2 \cdot P_{H_0}(\hat{P} \le \hat{p})$	$P_{H_0}(\hat{P} \le \hat{p})$	$P_{H_0}(\hat{P} \ge \hat{p})$

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

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

© כל הזכויות שמורות לברק קנדל

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%?

© כל הזכויות שמורות לברק קנדל

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 $p-value$ של 3%,
לכן אם היינו מבצעים מבחן דו-צדדי (כאשר יתר הנתונים ללא שינוי),
היינו מקבלים ערך $p-value$ 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?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

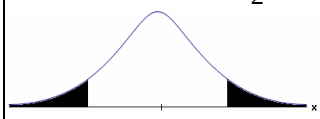
פרק 50 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסייה ידועות..... 257
2. כששוניויות האוכלוסיה לא ידועות ומניחים שהן שוות..... 261
3. ניתוח פלטים..... 265

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונויות של האוכלוסייה ידועות – רקע

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

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

© כל הזכויות שמורות לברק קנדל

חלופה אחרת לכלל הכרעה:

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

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

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

דוגמה (פתרון בהקלטה) :

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

© כל הזכויות שמורות לברק קנדל

שאלות

1) מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלה נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה.
במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות.
במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות.
לצורך פתרון הניחו שבכל אזור, סטיית התקן היא שעה 1 ביום.
בדקו את טענת המחקר ברמת מובהקות של 1%.

2) ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 100. מכון ללימוד פסיכומטרי טוען שהוא יכול לשפר את ממוצע הציונים ביותר מ-30 נקודות.
במדגם של 20 נבחנים שניגשו למבחן ללא הכנה במכון התקבל ממוצע 508.
במדגם של 25 נבחנים שעברו הכנה במכון התקבל ממוצע ציונים 561. מה מסקנתכם ברמת מובהקות של 5%.

3) במדגם אקראי של 20 ימים נבדקה התפוקה של מפעל ביום. התפוקה הממוצעת הייתה של 340 מוצרים ליום. במדגם אקראי של 20 ימים אחרים נבדקה התפוקה של המפעל בלילה והתפוקה הממוצעת הייתה 295. לצורך פתרון נניח שסטיית התקן של התפוקה ביום היא 40 מוצרים ובלילה 30 מוצרים.

א. מהי מובהקות התוצאה לבדיקה האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית.

ב. מה תהיה המסקנה ברמת מובהקות של 8%?

4) במחקר מקיף שנעשה באירופה נקבע שגברים גבוהים מנשים ב-8 ס"מ בממוצע. מחקר ישראלי מתעניין לבדוק האם בישראל הפער גדול יותר. לצורך המחקר נדגמו 40 גברים ו 40 נשים באקראי. כמו כן, נניח שסטיות התקן של הגברים והנשים ידועות ושוות ל-6 ס"מ אצל הנשים ו-12 ס"מ אצל הגברים.

א. מהן השערות המחקר ומהו כלל ההכרעה ברמת מובהקות של 10%?

ב. אם בישראל הפער בין גברים לנשים מבחינת הגובה הממוצע הוא 11 ס"מ, מה ההסתברות שהמחקר לא יגלה זאת? איך קוראים להסתברות הזאת?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) נדחה H_0.

2) לא נדחה את H_0.

3) א. 0 ב. נדחה את H_0.

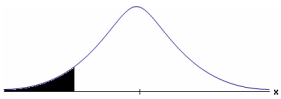
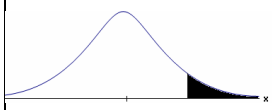
4) א. נדחה את H_0, אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותרמ-10.72 ס"מ.

ב. 0.6331

© כל הזכויות שמורות לברק קנדל

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

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

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

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

© כל הזכויות שמורות לברק קנדל

חלופה אחרת לכלל הכרעה:

נדחה H_0 אם מתקיים:	
$\bar{x}_1-\bar{x}_2 > c+t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$ **או** $\bar{x}_1-\bar{x}_2 < c-t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$	$\bar{x}_1-\bar{x}_2 < c-t_{1-\alpha}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$
$\bar{x}_1-\bar{x}_2 > c+t_{1-\alpha}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$	

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%.

© כל הזכויות שמורות לברק קנדל

שאלות

1) להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (במ"ר):

2012	120	112	95	130	90	94	120
2013	100	82	91	105	74	69	

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 W 60 דולקות בממוצע יותר מאשר נורות מסוג 100?

ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W 60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.

© כל הזכויות שמורות לברק קנדל

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%.

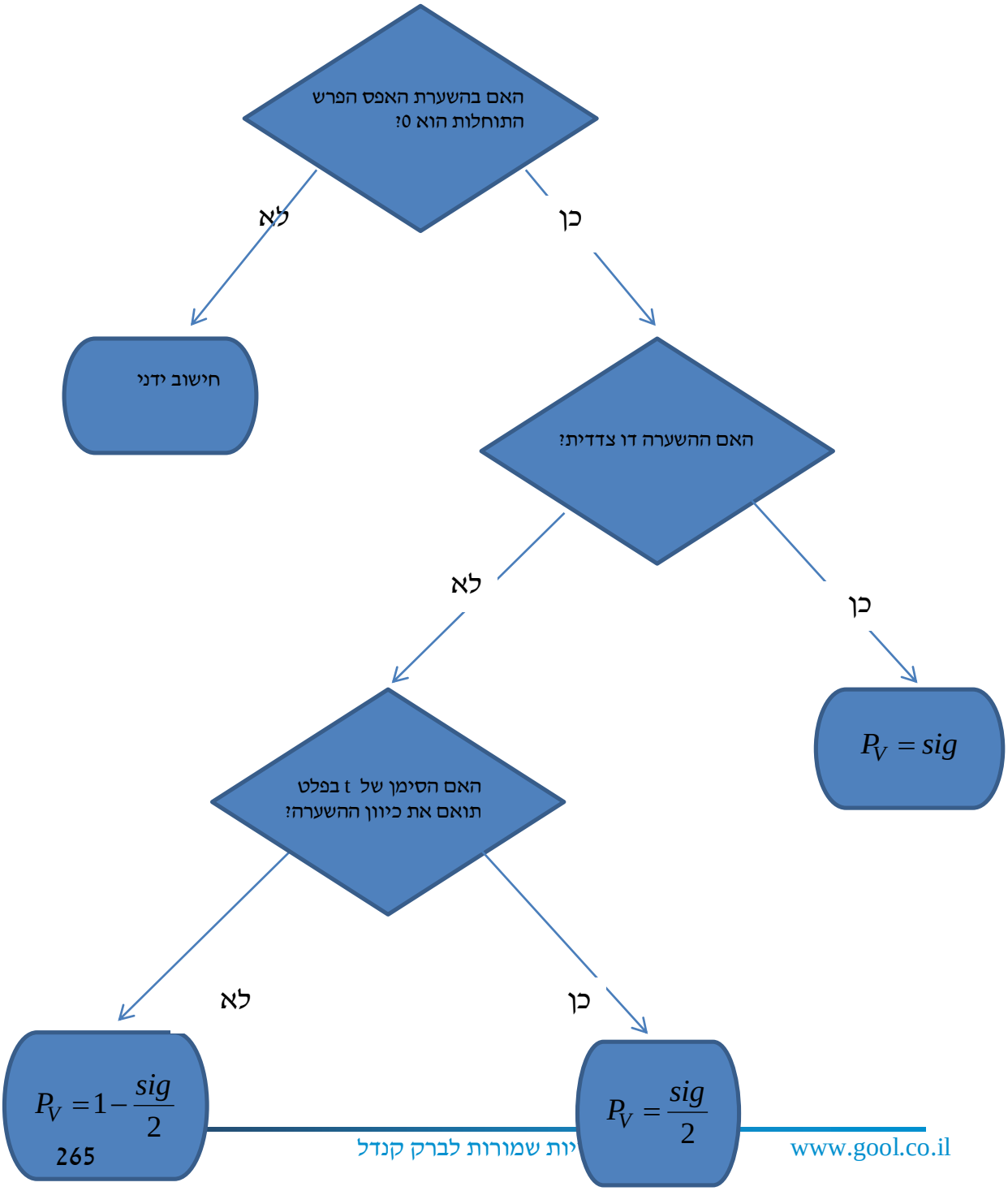
ג. לא נדחה את H_0.

© כל הזכויות שמורות לברק קנדל

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

ניתוח פלטים – רקע

מובהקות התוצאה על סמך הפלט:



דוגמה (פתרון בהקלטה) :

בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים משני אזורים שונים במדינה על מס' האחים והאחיות שלהם. להלן הפלט שהתקבל:

Group Statistics

	Region of the United States	N	Mean	Std. Deviation	Std. Error Mean
Number of Brothers and Sisters	North East	676	3.76	2.939	.113
	South East	410	4.05	2.993	.148

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Number of Brothers and Sisters	Equal variances assumed	.173	.677	-1.583	1084	.114	-.293	.185	-.657	.070
	Equal variances not assumed			-1.576	850.945	.115	-.293	.186	-.658	.072

א. מהו המבחן הסטטיסטי שנעשה כאן?

ב. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?

ג. בדוק האם קיים הבדל בין-"South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.

ד. מהי מובהקות התוצאה לבדיקת הטענה שההפרש הממוצע בין-" South East" לבין-"North East" חיובי?

© כל הזכויות שמורות לברק קנדל

שאלות

1) להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים. רמת האופטימיות נמדדה בסולם ציונים של 1 עד 5.

Group Statistics

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
optimizm	MALE	633	2.6053	.49781	.01979
	FEMALE	568	2.5503	.48483	.02034

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	1190.977	.053	.05500	.02838	-.00068	.11067

א. האם ניתן להניח ששונות האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?

ב. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?

ג. מצא את הגבול העליון של רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.

ד. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

© כל הזכויות שמורות לברק קנדל

2) פסיכולוגים טוענים שאנשים שניגשים למבחן אינטליגנציה יותר מפעם אחת נוטים לקבל ציונים גבוהים יותר. להלן הפלט שהתקבל:

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
grade	A	9	96.8889	9.40006	3.13335
	B	11	108.4545	11.46616	3.45718

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
grade	Equal variances assumed	.206	.656	-2.428	18	.026	-11.56566	4.76333	-21.57304	-1.55828
	Equal variances not assumed			-2.479	17.997	.023	-11.56566	4.66583	-21.36832	-1.76299

T-Test

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.

ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?

ג. האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של $\alpha = 0.05$.

ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha = 0.01$.

© כל הזכויות שמורות לברק קנדל

3) כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל בממוצע גיל הנישואין (הראשונים), בין נשים הגרות בערים מרכזיות לבין נשים הגרות בערים מרוחקות מהמרכז.
לשם כך נדגמו 50 כלות מכל אחת משתי ערים עיר א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שלהלן:

T-Test

Group Statistics

מקום המגורים		N	Mean	Std. Deviation	Std. Error Mean
גיל הנישואין	עיר א	50	24.8072	1.38978	.19654
	עיר ב	50	23.0131	1.62070	.22920

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
גיל הנישואין	Equal variances assumed	.330	.567	5.942	98	.000	1.79415	.30193	1.19497	2.39332
	Equal variances not assumed			5.942	95.772	.000	1.79415	.30193	1.19480	2.39350

א. מהו המבחן הסטטיסטי שנעשה כאן?
ב. מצא רווח סמך ברמת סמך של 95% להפרש בין עיר א לעיר ב מבחינת גיל הנשים הממוצע בנישואין הראשונים.
ג. האם ניתן לומר ברמת מובהקות של 1% שנשים בערים מרכזיות מתחתנות בגיל מאוחר יותר מאשר נשים הגרות בערים מרוחקות?

© כל הזכויות שמורות לברק קנדל

4) להלן פלט של תוכנת SPSS:

T-Test

	N	Mean	Std. Deviation	Std. Error Mean
X	26	36.3077	13.23259	2.59513
Y	24	46.4583	20.96369	4.27920

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-ailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	4.446	.040	-2.164	???	.044	-10.15064	???	-20.03781	-.26347
Equal variances not assumed			-2.038	38.267	.048	???	5.00462	-20.27964	-.02164

א. השלימו את סימני השאלה בטבלה.

ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של X לזה של Y?

ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X גדולה מהתוחלת של Y?

ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X קטנה מהתוחלת של Y?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. נקבל את H_0 ונכריע שיש שוויון שוניות.

ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.

ג. 0.11076 ד. $2.5665 \le \mu \le 2.6441$.

2) א. מבחן T להפרש ממוצעים במדגמים בלתי תלויים.

ב. מבחן T למדגמים מזווגים.

ג. נקבל את H_0, נקבע לקיום שוויון שוניות.

ד. נקבל את H_0, לא נקבל את טענת הפסיכולוגים.

3) א. מבחן T להשוואת תוחלת במדגמים בלתי תלויים.

ב. $1.19497 \le \mu_1 - \mu_2 \le 2.39332$ ג. כן.

4) א. 10.15, 4.69,-48 ב. 0.04

ג. 0.978 ד. 0.022.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

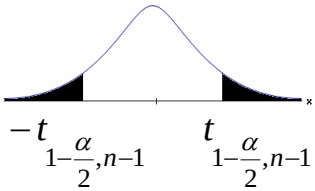
פרק 51 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

תוכן העניינים

1. בדיקת השערות למדגמים מזווגים 272
2. ניתוח פלטים 276

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

בדיקת השערות למדגמים מזווגים – רקע

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

סטטיסטי המבחן: $t_{\bar{D}} = \dfrac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$, $S_D{}^2 = \dfrac{\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}{n-1} = \dfrac{\sum_{i=1}^{n} D^2{}_i - n\bar{D}^2}{n-1}$

דוגמה (פתרון בהקלטה):

© כל הזכויות שמורות לברק קנדל

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".

© כל הזכויות שמורות לברק קנדל

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 בממוצע?

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 בממוצע 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 נורמאלית 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.

© כל הזכויות שמורות לברק קנדל

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	4.5	9.6	0.7	2.5
משקל במכשיר 2	3.5	6.9	1.7	0.5

נניח שהמשקלים מתפלגים נורמלית, 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.

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:

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

תשובות סופיות

1) לא נדחה H_0.

2) לא נדחה H_0.

3) א. $0.25 \le pv \le 0.5$ ב. 0.5 ג. לא נדחה H_0.

4) ד'.

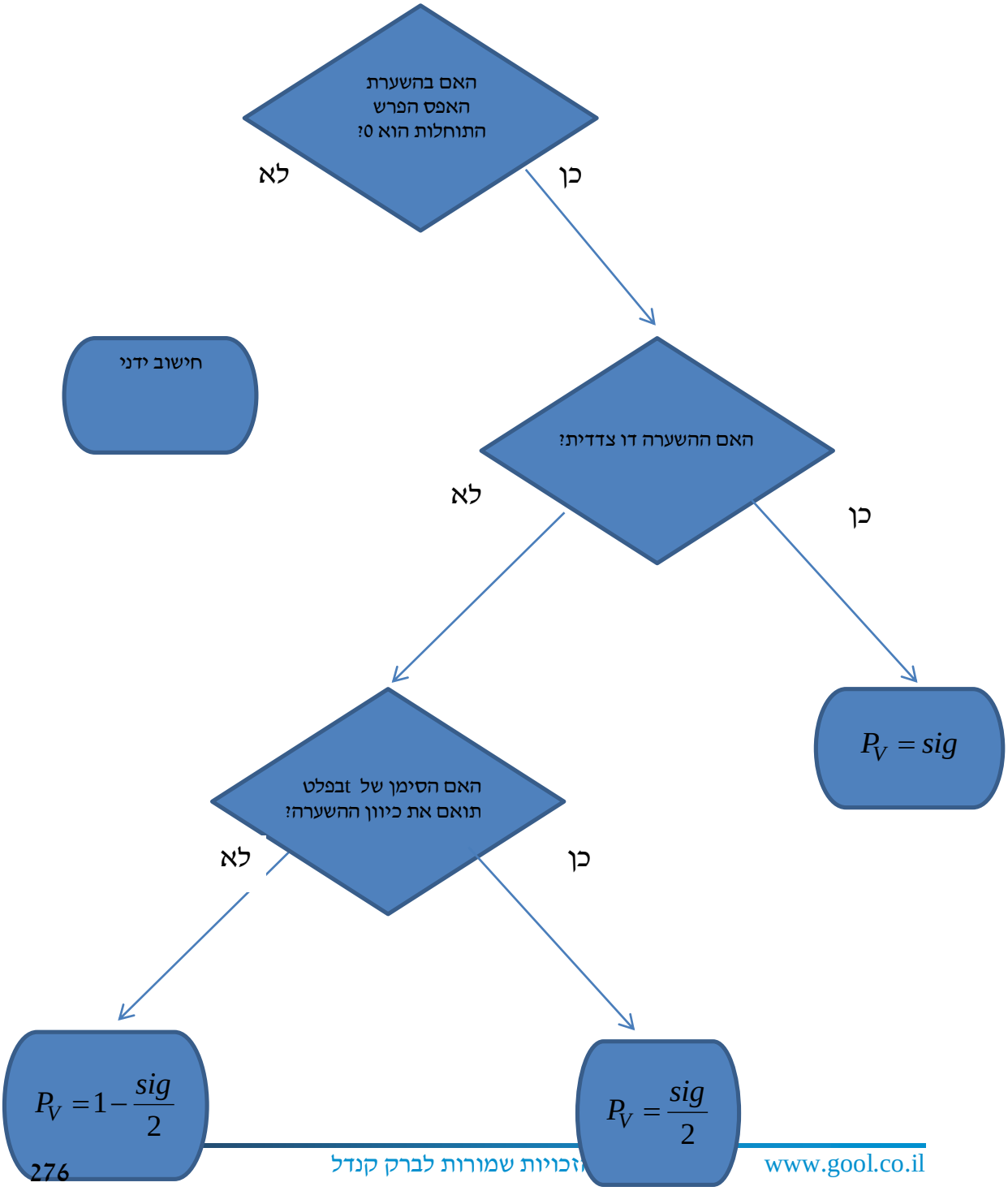
5) ד'.

6) ג'.

© כל הזכויות שמורות לברק קנדל

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

מדגמים מזווגים – ניתוח פלטים – רקע



זכויות שמורות לברק קנדל

דוגמה (פתרון בהקלטה) :

כדי לבדוק את ההשפעה של קורס לגמילה מעישון נלקח מדגם מקרי של 5 נבדקים. עבור כל אחד מהם נמדדה צריכת הסיגריות היומית לפני הקורס וחודשיים אחריו. הניחו שצריכת הסיגריות מתפלגת נורמלית. להלן התוצאות:

נבדק	1	2	3	4	5
לפני	40	22	25	28	30
אחרי	30	24	13	10	12

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	BEFORE	29.0000	5	6.85565	3.06594
	AFTER	17.8000	5	8.72926	3.90384

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				90% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 BEFORE - AFTER	11.20000	8.19756	3.66606	3.38452	19.01548	3.055	4	.038

בדקו ברמת מובהקות של 5% האם הקורס יעיל.

© כל הזכויות שמורות לברק קנדל

שאלות

1) בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים על השכלת הוריהם, להלן הפלט שהתקבל:

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Highest Year School Completed, Father - Highest Year School Completed, Mother	-.007	3.115	.100	-.203	.189	-.072	973	.943

א. תנו אומדן להפרש הממוצעים.

ב. תנו אומדן לטעות התקן של הפרש הממוצעים.

ג. האם קיים הבדל מובהק בין השכלת האבות להשכלת האימהות ברמת מובהקות של 5%?

2) בתחרות קפיצה למים שופטים באופן קבוע שופט איטלקי ושופט דרום קוריאני. להלן פלט המנתח את הציונים ששופטים אלה נתנו בתחרויות השונות:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Italy	???	300	.86742	.05008
	South Korea	8.9183	???	.81992	.04734

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Italy - South Korea	-.42233	.36153	.02087	-.46341	-.38126	-20.234	???	???

א. השלימו את החלקים החסרים בפלט (מסומנים בסימני שאלה).

ב. בדקו את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.

ג. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים?

ד. בנו את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.

© כל הזכויות שמורות לברק קנדל

3) בדקו את ציוניהם של 44 נבדקים אקראיים במבחן הפסיכומטרי. פעם אחת לפני הכנה (Before) ופעם אחת אחרי הכנה (After).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Before - After	-7.45455	19.28303	2.90703	-13.31712	-1.59197	-2.564	43	.014

א. רשמו מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.

ב. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.

ג. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.

ד. מצאו רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

© כל הזכויות שמורות לברק קנדל

4) להלן פלט של תכנת SPSS:

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	x	54.0000	6	5.86515	2.39444
	y	46.5000	6	10.72847	4.37988

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 x - y	7.50000	??	4.72405	-4.64356	19.64356	??	5	.173

א. מלא את החלקים החסרים בטבלה.

ב. מהי רמת המובהקות המינימלית לקבלת הטענה שיש הבדל בין X ל- Y בממוצע?

ג. האם התשובה לסעיף הקודם הייתה משתנה, ואם כן גדלה או קטנה,
אם הינו מוסיפים עוד תצפית שההפרש בין X ל- Y הוא 0.

ד. מהי מובהקות התוצאה לבדיקת הטענה ש X גדול מ- Y בממוצע?

ה. מהי מובהקות התוצאה לבדיקת הטענה ש X קטן מ- Y בממוצע?

ו. בנו רווח סמך לתוחלת של X ברמת סמך של 90%.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. -0.007 ב. 0.1 ג. אין הבדל מובהק.

2) א. $d.f = 299$, $n = 300$, $\bar{X} = 8.496$, $\text{Sig} = 0$.
ב. נדחה את H_0. ג. $(0.46341, 0.38126)$ ד. $(0.4708, 0.3738)$.

3) א. ראה וידאו. ב. נדחה את H_0. ג. לא נדחה את H_0.
ד. $(1.592, 13.317)$.

4) א. 1.5876, 11.5715 ב. 0.173 ג. יגדל.
ד. 0.0865 ה. 0.9135 ו. $49.18 < \mu < 58.82$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 52 - הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות

תוכן העניינים

1. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות 282

הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α על $\mu_1-\mu_2$:

$H_0:\mu_1-\mu_2=C, \qquad H_1:\mu_1-\mu_2\neq C$.

על ידי בניית רווח סמך ברמת סמך של $1-\alpha$ ל- $\mu_1-\mu_2$:

אם C נופל ברווח ← נקבל את H_0 .

אם C לא נופל ברווח ← נדחה את H_0 .

דוגמה (פתרון בהקלטה) :

חוקר ביצע בדיקת השערות לתוחלת ההפרש במדגם מזווג.

להלן השערותיו : $\alpha=5\%, \qquad H_1:\mu_D\neq 80, \qquad H_0:\mu_D=80$.

החוקר בנה רווח סמך ברמה של 90%, $78<\mu_D<83$.

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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 למוצע 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. $-3.8 \le \mu_D \le 19$ ב. נכריע שיש אמת בפרסום.

2) א. $-8.5 \le \mu_X - \mu_Y \le 26.5$ ב. נכריע שאין הבדל.

3) ג'.

4) א'.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 53 - שאלות מסכמות בבדיקת השערות

תוכן העניינים

1. שאלות פתוחות מסכמות 285
2. שאלות רב ברירה (אמריקאיות) 288

שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות

1) שני חוקרים נתבקשו לבדוק את ההשערות הבאות: $H_0: \mu = 520, \quad H_1: \mu > 520$.
כל חוקר בדק מדגם של 225 נחקרים. ידוע ש- $\sigma = 20$.
חוקר א' קבע את כלל ההכרעה לפי $\alpha = 0.05$.
חוקר ב' מחליט לדחות H_0 אם $\bar{X} > 522$.

א. למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?

ב. מהי ההסתברות לטעות מסוג שני של חוקר ב' עבור $\mu_1 = 525$.

ג. הסבר ללא חישוב נוסף, האם ההסתברות לטעות מסוג שני עבור $\mu_1 = 525$, של חוקר א' שווה/קטנה/גדולה לזו של חוקר ב'.

ד. חוקר א' קיבל במדגם שלו $\bar{X} = 523$. מהי מסקנתו?

2) ידוע כי תוחלת מספר הלייקים היומי של דנה היא 12 עם סטיית תקן 5.
דני טוען שהוא יותר פופולארי מדנה בכך שהוא מקבל יותר לייקים מדנה ביום. על-מנת לבדוק זאת ספר דני כמה לייקים הוא קיבל בכל יום במהלך 7 שבועות (כלומר, ב – 49 ימים) וקיבל סך-הכול 637 לייקים. נניח כי סטיית התקן של מספר הלייקים שדני מקבל ביום זהה לסטיית התקן של דנה.

א. מהי רמת המובהקות שכדאי לדני לדרוש, כדי שדנה תשתכנע בצדקת טענתו (שדני פופולרי יותר בכך שהוא מקבל יותר לייקים מדנה ביום).

ב. אם דני משער שתוחלת מספר ה"לייק"ים שהוא מקבל ביום היא 14 וקובע רמת מובהקות 2.5%, מהי עוצמת המבחן של דני?

3) ברצוננו להשוות בין רשתות A לבין B. לשם כך בחרנו 4 מוצרים, ובדקנו את מחיריהם בשתי הרשתות. להלן התוצאות:

מוצר / רשת	A	B
1	5	5
2	4	5
3	5	3
4	7	4

הניחו כי המחירים מתפלגים נורמלית.
אם יש הנחות נוספות כדי לבצע את המבחן הפרמטרי רשמו אותן.

א. בדקו האם קיים הבדל בין הרשתות מבחינת תוחלת המחירים. רמת מובהקות של 5%.

ב. חזרו על הסעיף הקודם בהנחה ונבחרו בכל רשת מוצרים באקראי ולא בהכרח אותם מוצרים.

© כל הזכויות שמורות לברק קנדל

4) במטרה לבדוק האם סטודנטים הלומדים במכללות משקיעים יותר זמן ללימודים מאשר סטודנטים באוניברסיטאות נדגמו 12 סטודנטים ובדקו לכל סטודנט את הזמן שהוא משקיע ביום ללימודים. הזמנים נמדדו בדקות:

סטודנטים באוניברסיטאות	176	156	189	171	140	180
סטודנטים במכללות	180	190	191	186	204	150

א. נסחו את ההשערות ובדוק אותן ברמת מובהקות של 5%. רשום את כלל ההכרעה ואת ההנחות הדרושות לביצוע המבחן הפרמטרי.

ב. חשבו את p-value.

ג. ישנה טענה שממוצע זמן ההשקעה בלימודים במכללות הוא 3.5 שעות ביום. בדוק את הטענה כאשר רמת המובהקות הינה 5%.

5) בשנת 2000 ל-60% היה מדיח כלים בבית. מחקר רוצה לבדוק האם כיום פרופורציית המשפחות עם מדיח כלים עלה. הוחלט לבצע מדגם אקראי של 150 משפחות.

א. רשמו את השערות המחקר.

ב. מה היא מסקנת המחקר ברמת מובהקות של 5% אם במדגם ל-102 משפחות היה מדיח כלים.

גברים	2	0.9	0.7	1.2	1.1
נשים	1.4	1.1	1.9	1.8	1.2

ג. מהי הטעות האפשרית במסקנה מהסעיף הקודם. האם ניתן לדעת את הסתברותה?

6) נערך מחקר על הקשר בין עישון ויתר לחץ דם. נבדק מדגם מקרי של 200 מעשנים ונמצא כי 30 סבלו מיתר לחץ דם.
ידוע שבאוכלוסייה 18% סובלים מיתר לחץ דם.

א. בדקו ברמת מובהקות 0.1 את ההשערה כי אחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.

ב. מהי רמת המובהקות המינימלית לקבלת הטענה שאחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.

ג. מהי עצמת המבחן, אם אחוז הסובלים מיתר לחץ דם בקרב אוכלוסיית המעשנים היא בפועל 25%.

7) להלן התפלגות מספר הנסיעות לחופשה השנתית במדגם של משפחות ישראליות. בדקו ברמת מובהקות של 5%:

מספר הנסיעות	0	1	2	3	4
מספר המשפחות	84	102	26	20	12

א. באיטליה משפחות נוסעות בממוצע פעמיים בשנה לחופשה. האם בישראל משפחות נוסעות פחות מאשר באיטליה?

ב. בהולנד 80% מהמשפחות נוסעות לפחות פעם אחת בשנה לחופשה, האם בישראל אחוז המשפחות שנוסעות לפחות פעם אחת בשנה לחופשה נמוך מאשר בהולנד?

© כל הזכויות שמורות לברק קנדל

8) נתון כי: $X \sim N(\mu, \sigma^2 = 10^2)$.

מעוניינים לבדוק את ההשערות: $H_0: \mu = 40, \quad H_1: \mu > 40$.

דגמו 25 תצפיות מהאוכלוסייה והתקבל $\bar{X} = 45$.

א. חשבו את p-value (מובהקות התוצאה).

ב. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה: $H_1: \mu < 40$

ג. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה: $H_1: \mu \neq 40$

תשובות סופיות

1) א. חוקר א' ב. 0.0122 ג. גדלה. ד. נדחה H_0.

2) א. לפחות 0.0808 ב. 0.7995.

3) א. לא נדחה H_0. ב. לא נדחה H_0.

4) א. לא נדחה H_0. ב. בין 5% ל-10%. ג. נדחה H_0.

5) א. $H_0: p = 0.6$, $H_1: p > 0.6$ ב. נדחה H_0. ג. טעות מסוג ראשון בסיכוי של 0.05.

6) א. לא נדחה H_0. ב. 0.8643 ג. 0.8749.

7) א. נדחה H_0. ב. נדחה H_0.

8) א. 0.0062 ב. 0.9938 ג. 0.0124.

© כל הזכויות שמורות לברק קנדל

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 האלטרנטיבית 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 $X - Y$)

לפני הנישואין- X	69	93	82	68
לאחר הנישואין- Y	80	88	84	71

א. $H_1 : \mu_d < 0, \quad H_0 : \mu_d = 0$

ב. $H_1 : \mu_X - \mu_Y < 0, \quad H_0 : \mu_X - \mu_Y = 0$

ג. $H_1 : \mu_X - \mu_Y < 0, \quad H_0 : \mu_X - \mu_Y = 0$

ד. $H_1 : \mu_d > 0, \quad H_0 : \mu_d = 0$

© כל הזכויות שמורות לברק קנדל

5) 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 מהתושבות 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.

א. מבחן Z למדגם יחיד; מבחן T למדגם יחיד.

ב. מבחן T למדגם יחיד; מבחן T למדגמים תלויים.

ג. מבחן T למדגם יחיד; מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים בלתי תלויים; מבחן T לממוצע יחיד.

7) 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.

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:

לפני הנישואין	69	93	82	68
לאחר הנישואין	80	88	84	71

באיזה התפלגות משתמשים לבדיקת ההשערות, ובכמה דרגות חופש:

א. ההתפלגות Z ללא דרגות חופש.

ב. ההתפלגות T ו-3 דרגות חופש.

ג. ההתפלגות T ו-6 דרגות חופש.

ד. ההתפלגות χ^2 ו-3 דרגות חופש.

© כל הזכויות שמורות לברק קנדל

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'?

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

ג. ללא נתונים נוספים אי אפשר לדעת מה יחליט.

10) חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?

א. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

ב. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

ג. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.

ד. 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?

א. טעות מסוג ראשון.

ב. טעות מסוג שני.

ג. טעות מסוג שלישי.

ד. לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמתית אצל הצמחוניים.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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?

א. מבחן T למדגמים בלתי תלויים עם 38 דרגות חופש.

ב. מבחן T למדגמים מזווגים, עם 39 דרגות חופש.

ג. מבחן T למדגמים בלתי תלויים עם 39 דרגות חופש.

ד. 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?

א. מבחן Z על פרופורציה.

ב. מבחן T על תוחלת אחת.

ג. מבחן T על שתי תוחלות במדגמים בלתי תלויים.

ד. 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:

א. שלושתם במבחן T למדגמים בלתי תלויים.

ב. עידו במבחן T למדגמים מזווגים, ורון ויואב במבחן T למדגמים בלתי תלויים.

ג. עידו במבחן T למדגמים מזווגים, רון במבחן T למדגמים בלתי תלויים ויואב במבחן T למדגם יחיד.

ד. עידו במבחן T למדגמים מזווגים, רון ויואב במבחן T למדגם יחיד.

© כל הזכויות שמורות לברק קנדל

19) במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5%.
מה תמיד נכון?

א. הגדלת רמת המובהקות לא תשתנה 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.

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:

א. השערת האפס הייתה נדחית.
ב. השערת האפס הייתה לא נדחית.
ג. 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?

א. אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
ב. אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.
ד. 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: $H_0: \mu=\mu_0, \quad H_1: \mu<\mu_0$.
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.

א. כעת בברור הוא ידחה את השערת האפס.
ב. כעת הוא דווקא יקבל את השערת האפס.
ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

23) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:

א. הסיכוי לטעות מסוג ראשון גדל.
ב. העוצמה של המבחן גדלה.
ג. הסיכוי לטעות מסוג שני גדל.
ד. תשובות א ו-ב נכונות.

© כל הזכויות שמורות לברק קנדל

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 מהתושבות לא נכונה בהכרח.

25) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

	α	$1-\beta$
א.	גדולה	גדולה
ב.	גדולה	קטנה
ג.	קטנה	גדולה
ד.	קטנה	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- $p-value=0.02$.

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.

© כל הזכויות שמורות לברק קנדל

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 $p\ value = 0.0254$, 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: $H_0: \mu = \mu_0, \quad H_1: \mu < \mu_0$.
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

© כל הזכויות שמורות לברק קנדל

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	4.5	9.6	0.7	2.5
משקל במכשיר 2	3.5	6.9	1.7	0.5

נניח שהמשקלים מתפלגים נורמלית.
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.

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:

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

36) סטטיסטיקאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0.
המשמעות של תוצאה זו היא:

א. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
ב. תמיד לדחות את השערת האפס כאשר היא נכונה.
ג. לעולם לא לדחות את השערת האפס כאשר היא נכונה.
ד. 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:

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.

© כל הזכויות שמורות לברק קנדל

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 מתפלים נורמלית בנו רווח סמך למוצע 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: $H_0: \mu = 10, \quad H_1: \mu \neq 10$.
לצורך בדיקה הוא לקח מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך תוצאות המדגם הוא חישב וקיבל: $t_{\bar{x}} = -2.63$. 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

© כל הזכויות שמורות לברק קנדל

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. $\text{POWER}+\alpha+\beta=1$

ב. $\text{POWER}=0.5-\beta$

ג. $\text{POWER}+\alpha=1$

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

© כל הזכויות שמורות לברק קנדל

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 **<u>נכון</u>** 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 $H_0:\mu\geq 10$, $H_1:\mu<10$. נלקח מדגם והתקבלה רמת מובהקות מינימאלית לדחיית השערת האפס 0.058. לו רצינו לבדוק את ההשערה הדו צדדית $H_0:\mu=10,\quad H_1:\mu\neq 10$, 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 $H_0:\mu=55,\quad H_1:\mu=65$, 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: $H_0:\mu=55,\quad H_1:\mu=60$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות: $H_0:\mu=55,\quad H_1:\mu=65$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל n ורמת מובהקות 10% לבדיקת ההשערות: $H_0:\mu=55,\quad H_1:\mu=65$ 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.

© כל הזכויות שמורות לברק קנדל

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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 54 - מבחני חי בריבוע

תוכן העניינים

1. מבחן לאי תלות 301

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:

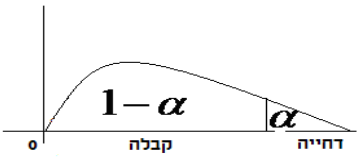
הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש $d.f=(r-1)(c-1)$.

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.

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



כאשר:

O_i - השכיחות נצפית במדגם בתא i.

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

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

© כל הזכויות שמורות לברק קנדל

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.

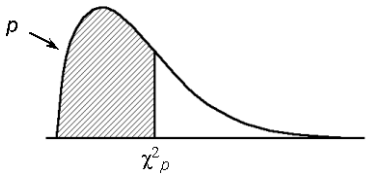
דוגמה (הפתרון בהקלטה):

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				

© כל הזכויות שמורות לברק קנדל

טבלת התפלגות חי-בריבוע – ערכי החלוקה $\chi^2 p$



df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
						p							
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	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

© כל הזכויות שמורות לברק קנדל

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:

	יום	ערב	לילה
פגומים	50	60	70
תקינים	600	700	800

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%.

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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 55 - ניתוח שונות חד כיוונית

תוכן העניינים

1. כללי 305

ניתוח שונות חד כיוונית

רקע תיאורטי

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות $\left(\mu_{1...},\mu_k\right)$ של k אוכלוסיות שונות.
לכן, בניתוח שונות, השערות המחקר הן:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ (התוחלות של כל האוכלוסיות שוות)

$H_1:$ אחרת (לפחות שתיים מהתוחלות שונות)

ההנחות הדרושות לביצוע התהליך:

2) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.
3) כל האוכלוסיות הן עם אותה שונות σ^2.
4) 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.

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

$\overline{X}_1$ - ממוצע המדגם הראשון,.....,$\overline{X}_k$ -ממוצע המדגם ה- k -י.

$\overline{X}$ - ממוצע כללי (של כל המדגמים).

סכום ריבועים בין הקבוצות: $SS_B = \sum_{i=1}^{k} n_i \left[\overline{X}_i - \overline{X}\right]^2$

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

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

$$SST = SSB + SSW$$

יש למלא את טבלת ניתוח השונות הבאה:

© כל הזכויות שמורות לברק קנדל

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	$n-1$		

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

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

© כל הזכויות שמורות לברק קנדל

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 1קיבלה "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?
מהו ה"גורם" וכמה רמות יש לו?

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.
להלן הציונים שהתקבלו:

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

א. מהו המבחן הסטטיסטי המתאים?
רשמו את ההשערות ואת ההנחות של המבחן.

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?

© כל הזכויות שמורות לברק קנדל

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			

© כל הזכויות שמורות לברק קנדל

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 ניבדק קצב פעימות הלב של כל אדם (pulse). הנתונים הוזנו לתוכנה סטטיסטית והתקבלו התוצאות הבאות:

ANOVA

pulse

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	414.400	2	207.200	19.733	.000
Within Groups	126.000	12	10.500		
Total	540.400	14			

pulse

Tukey HSD[a]

dosage	N	Subset for alpha = 0.05	
		1	2
30.00	5	71.0000	
20.00	5		80.2000
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.

א. בדקו ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.

ב. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.

ג. האם יש צורך במחקר בהשוואת מרובות. נמקו!

ד. לטבלת הANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "טוקי". ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

© כל הזכויות שמורות לברק קנדל

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:

א. כמה בתי ספר יש בעיר?

ב. כמה תלמידים השתתפו בסך הכול במחקר?

ג. האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%

ד. 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			

© כל הזכויות שמורות לברק קנדל

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

(I) school	(J) school	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval 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	Subset for alpha = 0.05 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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. משתנה בלתי תלוי: סוג התרופה.
משתנה תלוי: הזמן עד להשפעת התרופה בדקות.

ב. ניתוח שונות חד כיווני
$H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: otherwise$.

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 שונות.

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

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

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

ה. $F > 3.98$.

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.

© כל הזכויות שמורות לברק קנדל

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

5) $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_1: otherwise$

הנחות:
1. מדגמים בלתי תלויים.
2. שוויון שונויות.
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			

7) א. $H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: otherwise$

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.

ב. ראה וידאו. ג. כן. ד. $\mu_{20} = \mu_{10} > \mu_{30}$.

8) א. 5 ב. 25

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

ד. $(\mu_3 = \mu_4) < (\mu_1 = \mu_2 = \mu_3)$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 56 - מקדם המתאם (מדד קשר) הלינארי ומובהקותו

תוכן העניינים

1. מקדם המתאם הלנארי (פירסון) 314
2. חישוב מקדם המתאם הלנארי (פירסון) 325
3. בדיקת השערות על מקדם המתאם הלנארי 330

מקדם המתאם (מדד קשר) הלינארי ומובהקותו

מדד הקשר הלינארי (פירסון) – מבוא

מעוניינים לבדוק עד כמה קיים קשר מסוג קשר לינארי (קו ישר) בין שני משתנים.
שני המשתנים שאנו בודקים לגביהם קשר צריכים להיות משתנים כמותיים.
מבחינת סולמות מדידה כל משתנה נחקר צריך להיות מסולם רווחים או מנה.

בדרך כלל המשתנה המוצג כ- Y הוא המשתנה התלוי והמשתנה המוצג ב- X הוא המשתנה הבלתי תלוי.
תיאור גרפי לנתונים נעשה על ידי דיאגרמת פיזור. בדיאגרמת פיזור אנחנו מסמנים כל תצפית בנקודה לפי שיעור ה- X ושיעור ה- Y שלה.
דיאגרמת הפיזור נותנת אינדיקציה גרפית על הקשר בין שני המשתנים.

דוגמה (פתרון בהקלטה) :

בבניין 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגרות בדירה. להלן התוצאות שהתקבלו :

מספר חדרים בדירה	2	2	3	2	3	3	4	4
מספר הנפשות בדירה	0	1	2	2	3	4	4	5

1) כמה תצפיות ישנן בדוגמה?

2) כמה משתנים ישנם בדוגמה, מי הם?

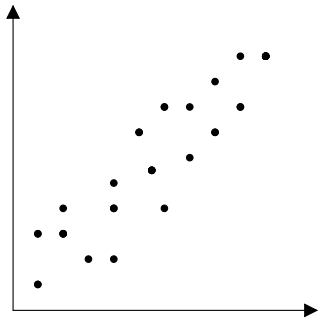
3) שרטטו לנתונים דיאגרמת פיזור.

4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

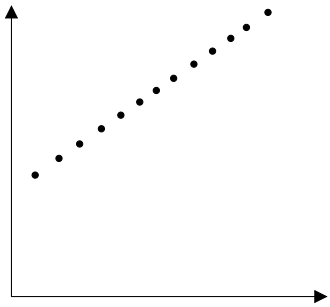
© כל הזכויות שמורות לברק קנדל

דיאגרמות פיזור לקשר בין משתנים וניתוחם

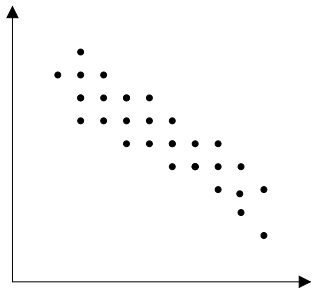
קשר לנארי חיובי חלקי



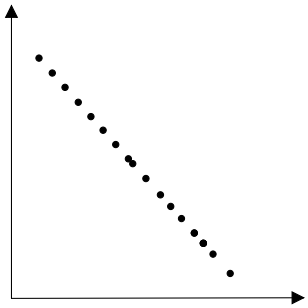
קשר לנארי חיובי מלא



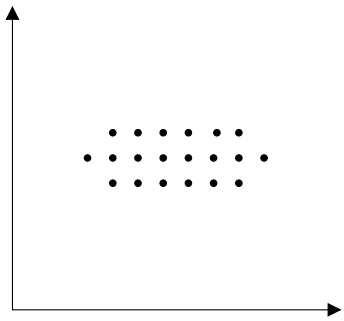
קשר לינארי שלילי חלקי



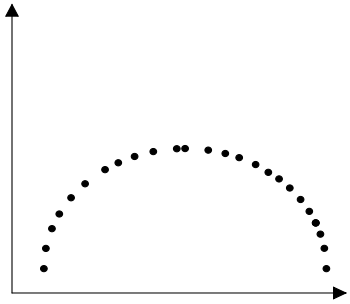
קשר לינארי שלילי מלא



אין קשר



אין קשר לינארי



משמעות מקדם המתאם:

כדי לבדוק עד כמה קיים קשר לינארי בין שני המשתנים ישנו מדד קשר שנקרא גם מקדם המתאם הלינארי הידוע גם בשם מקדם המתאם של פירסון.

מקדם מתאם זה מקבל ערכים בין 1 ל-1.

-1 0 1

© כל הזכויות שמורות לברק קנדל

מקדם מתאם 1-או 1 אומר שקיים קשר לינארי מלא בין המשתנים שניתן לבטאו על ידי נוסחה של קו ישר: $y = ax + b$.

מתאם חיובי מלא (מקדם מתאם 1):

קיים קשר לנארי מלא בו השיפוע a יהיה חיובי ואילו מתאם שלילי (מקדם מתאם-1) מלא אומר שקיים קשר לנארי מלא בו השיפוע a שלילי.

מתאם חיובי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט. ככל שמקדם המתאם קרוב לאפס עוצמת הקשר יותר חלשה וככל שהמדד רחוק יותר מהאפס העוצמה יותר חזקה. לסיכום, מקדם המתאם בודק את עוצמת הקשר הלנארי, ואת כיוון הקשר.

מקדם המתאם הלינארי אינו מושפע מיחידות המדידה. כל שינוי ביחידות המדידה של המשתנים, לא ישנה את מקדם המתאם.

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:

r - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר ρ.

קיומו של מתאם בין שני משתנים אינו מצביע על סיבתיות בהכרח. למשל, אם נמצא מתאם חיובי בין כמות הסוכרזית שאדם אוכל לבין במשקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי, כלומר אם נחליף את X ב- Y התוצאה תהיה זהה.

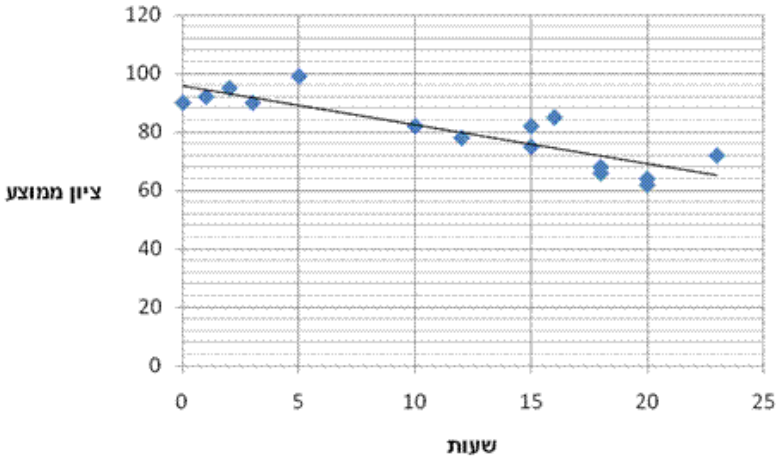
דוגמה (פתרון בהקלטה) :

- מה ניתן להגיד על מקדם המתאם של שני המשתנים על סמך דיאגרמת הפיזור ששרטטנו?
- אם היינו משנים את השרטוט כך שבציר האנכי היה המשתנה "מספר החדרים" ובציר האופקי היה "מספר הנפשות", האם הדבר היה משפיע על מדד הקשר של פירסון?

© כל הזכויות שמורות לברק קנדל

שאלות

1) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור:



א. מיהו המשתנה הבלתי תלוי?

ב. מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר? מה ניתן להגיד על עוצמת הקשר?

2) להלן טבלה המסכמת את מקדמי המתאם הלינארי בין ציוני מבחנים שונים שהתקבלו עבור תלמידים בכיתה מסוימת:

	ספורט	לשון	מתמטיקה
ספורט	?	-0.7	?
לשון	?	?	0.6
מתמטיקה	-0.1	?	?

א. השלימו את מקדמי המתאם שמסומנים בסימן שאלה בטבלה.

ב. בין אילו שני ציוני מקצועות שונים קיים מתאם בעל העוצמה החזקה ביותר?

3) במחקר נתבקשו לבדוק את הקשר בין מספר שעות התרגול של קורס לבין הציון הסופי שלו. להלן תוצאות מדגם שהתקבל:

שעות תרגול	ציון סופי
20	90
25	90
30	95
15	60
30	90
20	85
10	50

א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי בדוגמה זו?

ב. שרטטו דיאגרמת פיזור לנתונים.

ג. מה ניתן לומר על הקשר בין המשתנים במדגם?

ד. מסתבר שבסופו של דבר נתנו פקטור של 5 נקודות לציון הסופי. כיצד הדבר היה משנה את מקדם המתאם של המדגם?

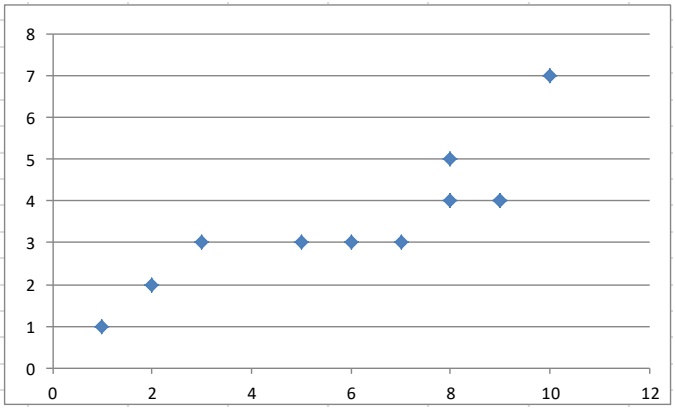
© כל הזכויות שמורות לברק קנדל

4) בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה 0.8.

א. השלימו את המשפט:
בחודש ינואר ככל שהטמפרטורה היומית נוטה לרדת, כך כמות המשקעים נוטה _____.

ב. הוחלט להעביר את הטמפרטורה למעלות פרנהייט על מנת שיוכלו להשוות אותה לנתונים מארה"ב. נוסחת המעבר היא $F^0 = 32 + \frac{9}{5}C^0$.
כיצד הדבר ישפיע על מקדם המתאם בין הטמפרטורה במעלות פרנייט לכמות המשקעים במ"מ?

5) להלן דיאגרמת פיזור המראה קשר בין שני משנים:



א. השלימו: ניתן לראות שהקשר הוא לינארי _____ (מלא\ חלקי) כיוון הקשר הוא (חיובי \שלילי).

ב. השלימו: אם היינו מוסיפים תצפית שערך ה- X שלה הוא 4 וערך ה- Y שלה הוא 7, מקדם המתאם של פירסון היה _____ (גדל\ קטן\ לא משתנה).

שאלות רב ברירה (יש לבחור את התשובה הנכונה):

6) חוקר אקלים דגם כמה ימים בשנה ומדד את הטמפרטורה בטורונטו שבקנדה ואת הטמפרטורה בסידני שבאוסטרליה באותו היום. הוא חישב ומצא מקדם מתאם שלילי בין הטמפרטורה היומית בטורונטו לבין הטמפרטורה היומית בסידני. משמעות מקדם המתאם השלילי במדגם:

א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני בימים שנדגמו.

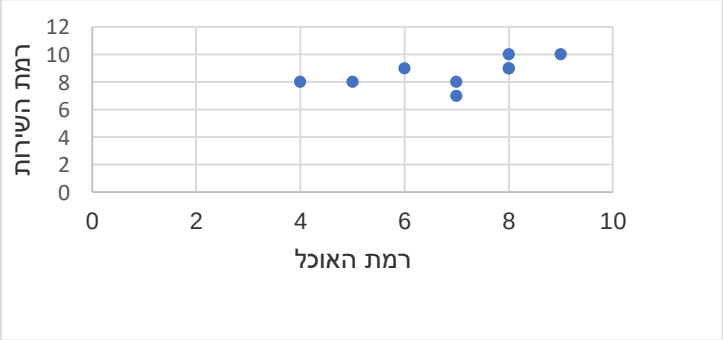
ב. במדגם, רוב הטמפרטורות בטורונטו היו שליליות.

ג. ההפרש בין הטמפרטורה בטורונטו לבין הטמפרטורה באוסטרליה, במדגם זה, הוא שלילי.

ד. במדגם יש נטייה שהטמפרטורה יורדת בטורונטו לטמפרטורה לעלות בסידני.

© כל הזכויות שמורות לברק קנדל

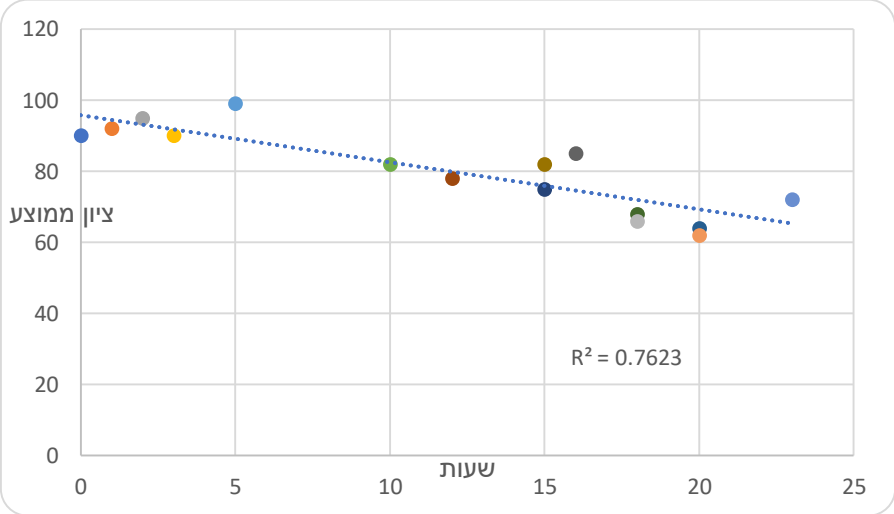
7) בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



מה יהיה ערכו של מקדם המתאם (r)?

א. $r=-0.3$
ב. $r=0$
ג. $r=1.125$
ד. $r=0.593$

8) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור.

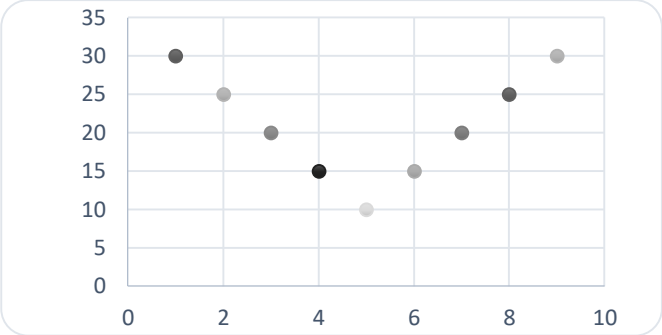


מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?

א. ככל שמבלים יותר הציון נוטה לרדת.
ב. אין קשר בין שעות הבילוי לציון.
ג. ככל שמבלים פחות הציון נוטה לרדת.
ד. ככל שהציון נוטה לרדת הסטודנט מבלה פחות.

© כל הזכויות שמורות לברק קנדל

9) התרשים הבא מתאר קשר בין שני משתנים, איזה מהמתאמים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?



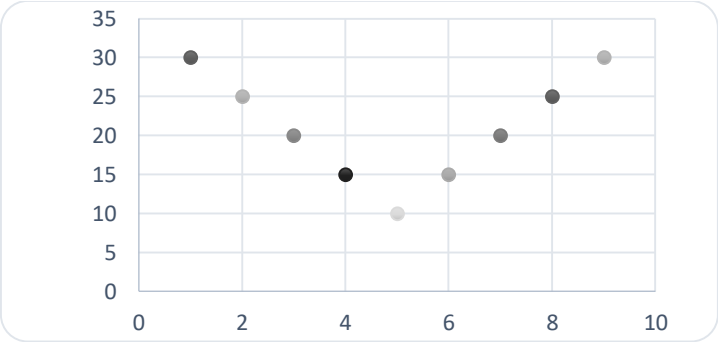
א. $r = 1$ היות ושני המשתנים יוצרים קוים ישרים.

ב. $r = 2$ היות ויש שני קוים בעלי קשר מושלם.

ג. $r = 0$ היות והקו יורד ואחר כך עולה באותו האופן.

ד. $r = \pm 1$ היות ויש קו עולה וגם קו יורד.

10) התרשים הבא מתאר דיאגרמת פיזור.



איזו טענה נכונה?

א. בתרשים מוצג הקשר בין שני משתנים.

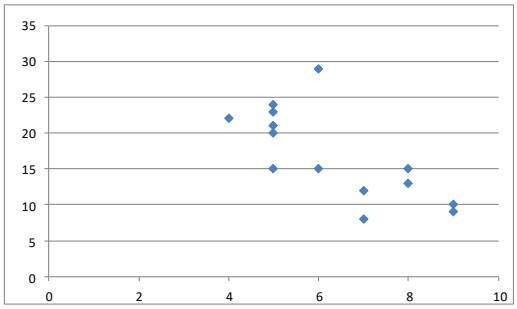
ב. בתרשים מוצג הקשר בין 9 משתנים.

ג. בתרשים מוצג הקשר בין 10 משתנים.

ד. אין לדעת כמה משתנים מוצגים בתרשים.

© כל הזכויות שמורות לברק קנדל

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים:



X - (משתנה בלתי תלוי בציר האופקי)

ו- Y (משתנה תלוי).

במדגם התקבל $r^2 = 0.52$.

11) לאור הנתונים המופיעים בדיאגרמה, איזה מבין הערכים הבאים מתאים להיות התוצאה של r?

א. -0.52

ב. 0.72

ג. -0.72

ד. 0.52

12) אם מקדם המתאם בין שני משתנים הוא 1, אזי:

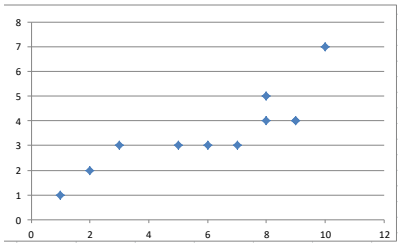
א. הערכים של המשתנים הם חיוביים.

ב. עבור כל תצפית ערך של משתנה אחד שווה לערך של המשתנה השני.

ג. הקשר הלינארי הוא בעוצמה חזקה.

ד. אף אחת מהתשובות לא בהכרח נכונה.

13) להלן דיאגרמת פיזור:



מה יהיה מקדם המתאם בין שני המשתנים?

א. 1

ב. 0.85

ג. 0.15

ד. 0

14) בבדיקת קשר בין שני משתנים התקבל: $r = -1$.

א. קיימת נוסחה לינארית הקושרת בין כל התצפיות.

ב. לא קיים קשר בין שני המשתנים.

ג. ככל שמשתנה אחד נוטה לרדת גם לשני יש נטייה לרדת.

ד. קיים קשר בין שני המשתנים, אך לא ניתן לדעת מאיזה סוג.

© כל הזכויות שמורות לברק קנדל

15) לפי הפתגם "רחוק מהעין, רחוק מהלב", יש קשר ___ בין קרבה פיזית לקרבה נפשית.

א. חיובי

ב. שלילי

ג. אפסי

ד. לא ניתן לדעת.

16) מבחן אמי"ר הינו מבחן מיון באנגלית של המרכז הארצי לבחינות והערכה. הציון המינימלי בבחינה הינו 150 והמקסימלי הינו 250. בקורס הכנה למבחן השתתפו 19 תלמידים. להלן הציונים שלהם על פי פלט שהתקבל:

		159
		170
		180
		185
		204
		224
		236
		212
		168
		189
		195
		163
		187
		206
		201
		223
		242
		203
		205
197.47	AVERAGE	
536.25	VARPA	

יש להוסיף עמודה נוספת לצד עמודת הציונים שתראה לכל תלמיד כמה נקודות חסרות לו כדי להשלים לציון המקסימלי בבחינה.

מה יהיה מקדם המתאם בין שתי העמודות (כלומר, מקדם המתאם בין הציון לבין הנקודות החסרות)?

א. -1

ב. 1

ג. -0.5

ד. 0.5

17) מקדם המתאם בין שטחי דירה למחיר שלהם חושב ונמצא 1.2. מה נובע מכך?

א. ככל שהדירה גדולה יותר בשטחה כך היא יקרה יותר.

ב. ככל שהדירה קטנה יותר בשטחה כך היא זולה יותר.

ג. לא קיים קשר בין שטח הדירה למחיר הדירה.

ד. מצב כזה שמתואר הנתונים לא אפשרי.

18) אם ניקח 10 אנשים ונרשום לכל אדם את הגובה במטר וכמו כן את הגובה בס"מ. מה יהיה מקדם המתאם בין גובה האדם במטר לגובה האדם בס"מ?

א. 1

ב. 0

ג. -1

ד. לא ניתן לדעת.

© כל הזכויות שמורות לברק קנדל

19) נמצא מתאם חיובי בעוצמה גבוהה בין X – ציון בבגרות בלשון ל Y – ציון בבגרות במתמטיקה. אילו מהמשפטים הבאים נכון?

א. ניתן לומר שאחת מהסיבות להבדלים שיש לסטודנטים במתמטיקה נובעים מההבדלים שיש להם בלשון.

ב. קיימת נוסחה של קו ישר שקושרת בין ציון בבגרות במתמטיקה לציון בבגרות בלשון.

ג. ללא יוצא מן הכלל, ניתן להגיד שכל תלמיד שמצליח יותר מתלמיד אחר בלשון גם יצליח יותר מאותו תלמיד במתמטיקה.

ד. אף אחד מהטענות שהוצגו אינה בהכרח נכונה.

20) עבור סדרה של תצפיות מדדו את X ואת Y. נמצא שעבור כל התצפיות שהערך של Y ירד הערך של X בהכרח ירד ללא יוצא מן הכלל. מקדם המתאם של פירסון יהיה בהכרח:

א. 1

ב. -1

ג. 0

ד. אף אחת מהתשובות.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. שעות בילוי. ב. הקשר חלקי, כיוון הקשר שלילי .

2) א. להלן טבלה : ב. ספורט ולשון.

	ספורט	לשון	מתמטיקה
ספורט	1	-0.7	0.1
לשון	-0.7	1	0.6
מתמטיקה	-0.1	0.6	1

3) א. ב״ת- מס׳ שעות התרגול, תלוי- ציון. ב. ראה גרף בפתרון וידאו.
ג. קשר לינארי חיובי חלקי. ד. מקדם המתאם לא היה משתנה.

4) א. לעלות. ב. לא ישפיע על מקדם המתאם.

5) א. חלקי, חיובי. ב. קטן.

6) ד׳. 7) ד׳. 8) א׳. 9) ג׳. 10) א׳.

11) ג׳. 12) ד׳. 13) ב׳. 14) א׳. 15) א׳.

16) א׳. 17) ד׳. 18) א׳. 19) ד׳. 20) ד׳.

© כל הזכויות שמורות לברק קנדל

מדדי קשר – מדד הקשר הלינארי (פירסון) – רקע

המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה.
בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו- Y הוא המשתנה המוסבר (התלוי).

דוגמה:

נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד – X מסבירה את ההכנסה שלו Y. במקרה זה שנות ההשכלה זהו המשתנה המסביר (או הבלתי תלוי) ואנחנו מעוניינים לבדוק כיצד שינויים בשנות ההשכלה של אדם יכולים להסביר את השינויים שלו בהכנסה, ולכן רמת ההכנסה זהו המשתנה המוסבר התלוי במשתנה המסביר אותו.

שלב ראשון: נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים.

דוגמה:

בבניין של 5 דירות בדקו את הנתונים הבאים:
X -מס' חדרים בדירה. Y-מס' נפשות הגרות בדירה.
להלן התוצאות שהתקבלו:

מס' דירה	X	Y
1	3	2
2	2	2
3	4	3
4	3	3
5	5	4

נשרטט מנתונים אלה דיאגרמת פיזור (הדיאגרמה המלאה בסרטון). נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן (הדיאגרמות המלאות בסרטון).

שלב שני: מחשבים את מקדם המתאם (מדד הקשר) שבודק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שניראה בשלב הראשון רק בעין.
המדד בודק את כיוון הקשר (חיובי או שלילי) ואת עוצמת הקשר (חלש עד חזק).
מקדם מתאם זה מקבל ערכים בין 1- ל-1.
מקדם מתאם 1- או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי הנוסחה: $y = bx + a$.

מתאם חיובי מלא (מקדם מתאם 1):

קיים קשר לנארי מלא בו השיפוע b יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לנארי מלא בו השיפוע b שלילי (מקדם מתאם 1-).

© כל הזכויות שמורות לברק קנדל

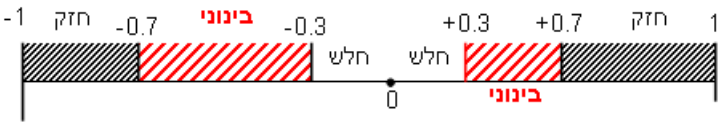
מתאם חיובי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

מתאם שלילי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

ככל שערך מקדם המתאם קרוב לאפס נאמר שעוצמת הקשר חלשה יותר וככל שמקדם המתאם רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר:



מקדם המתאם יסומן באות r.

כדי לחשב את מקדם המתאם, יש לחשב את סטיות התקן של כל משתנה ואת השונות המשותפת.

שונות משותפת: $COV(x,y)=\frac{\sum(x-\bar{x})(y-\bar{y})}{n}=\frac{\sum xy}{n}-\bar{x}\cdot\bar{y}$

שונות של המשתנה X: $s_x^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}=\frac{\sum_{i=1}^{n}x_i^2}{n}-\bar{x}^2$

שונות המשתנה Y: $S_Y^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}=\frac{\sum_{i=1}^{n}y_i^2}{n}-\bar{y}^2$

מקדם המתאם הלינארי: $r_{xy}=\frac{COV(x,y)}{S_x\cdot S_y}$

© כל הזכויות שמורות לברק קנדל

שאלות

1) להלן נתונים לגבי שישה תלמידים שנגשו למבחן. בדקו לגבי כל תלמיד את הציון שלו בסוף הקורס וכמו כן את מספר החיסורים שלו מהקורס.

מספר חיסורים	2	1	0	2	3	4
ציון	80	90	90	70	70	50

א. שרטטו דיאגראמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר ביו מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?

ב. חשבו את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?

ג. הסבירו, ללא חישוב, כיצד מקדם המתאם היה משתנה אם היה מתווסף תלמיד שהחסיר 4 פעמים וקיבל ציון 80?

2) במחקר רפואי רצו לבדוק האם קיים קשר בין רמת ההורמון X בדם החולה לרמת ההורמון Y שלו. לצורך כך מדדו את רמת ההורמונים ההלו עבור חמישה חולים. להלן התוצאות שהתקבלו:

X	Y
10	12
14	15
15	15
18	17
20	21

א. מה הממוצע של כל רמת הורמון?

ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

3) נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} X_i = 240 \qquad \sum_{i=1}^{20} Y_i = 200$$

$$\sum_{i-1}^{20} (X_i - \bar{X})^2 = 76 \qquad \sum_{i-1}^{20} (Y_i - \bar{Y})^2 = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y - \bar{Y}) = 60.8$$

א. חשב את מדד הקשר הלינארי בין X ל- Y. מיהו המשתנה התלוי?

ב. מה המשמעות של התוצאה שקיבלת בסעיף א?

4) נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} X_i = 240 \qquad \sum_{i=1}^{20} Y_i = 200$$

$$\sum_{i-1}^{20} X_i^2 = 2960 \qquad \sum_{i=1}^{20} Y_i^2 = 2080$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

חשבו את מדד הקשר הלינארי בין X ל- Y.

5) במוסד אקדמי ציון ההתאמה מחושב כך: מכפילים את הציון הממוצע בבגרות ב-3 ומפחיתים 2 נקודות. ידוע שעבור 40 מועמדים סטיית התקן של ממוצע הציון בבגרות הייתה 2.
מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם?

6) להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו.

א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.

ב. לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$, $S_x = S_y = 1$.
לכן, מדד הקשר של פירסון יהיה 1.

ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות רב-ברירה:

7) נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן:

א. הדבר מעיד שהציונים בכתה היו שליליים.

ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.

ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.

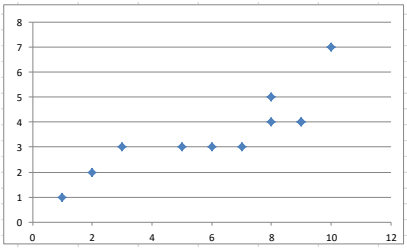
ד. אף אחת מהתשובות לא נכונה.

© כל הזכויות שמורות לברק קנדל

8) נלקחו 20 מוצרים ונבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש״ח (באותו היום ערך הדולר היה-4.2₪).
מהו מקדם המתאם בין המחיר בדולר למחיר בש״ח?

א. 1
ב. 0
ג. 4.2
ד. לא ניתן לדעת.

9) להלן דיאגראמת פיזור:
מה יהיה מקדם המתאם בין שני המשתנים?



א. 1
ב. 0.85
ג. 0.15
ד. 0

תשובות סופיות

1) א. משתנה תלוי: ציון, משתנה ב״ת: מס׳ חיסורים. ראה דיאגרמה בוידאו. ניתן להסיק שקיים קשר לינארי שלילי וחלקי בין מספר החיסורים לציון התלמיד.
ב. -0.9325. ג. הקשר יישאר לינארי שלילי חלקי אך עוצמתו תחלש.

2) א. $\bar{y}=16$, $\bar{x}=15.4$ ב. $r_{xy}=0.96$.

3) א. 0.8.

4) 0.8.

5) 1.

6) א. נכון. ב. לא נכון. ג. נכון.

7) ג׳.

8) א׳.

9) ב׳.

© כל הזכויות שמורות לברק קנדל

בדיקת השערות על מקדם המתאם הלינארי – רקע

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הלינארי וכיוונו בין שני המשתנים הנחקרים באוכלוסייה. כאשר:
r - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר ρ.

השערת האפס: תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים $H_0:\rho=0$.

ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית.

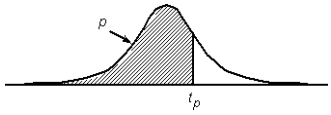
סטטיסטי המבחן: $t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\sim t(n-2)$

סטטיסטי זה מתפלג t עם $n-2$ דרגות חופש.

השערת האפס:	$H_0:\rho=0$	$H_0:\rho=0$	$H_0:\rho=0$
השערת המחקר:	$H_1:\rho\neq 0$	$H_1:\rho<0$	$H_1:\rho>0$
כלל ההכרעה: אזור דחייה של השערת האפס	$t\geq t_{1-\alpha}$ א ו $t\leq -t_{1-\alpha}$	$t\leq -t_{1-\alpha}$	$t\geq t_{1-\alpha}$

© כל הזכויות שמורות לברק קנדל

טבלת ערכים קריטיים של t - נספח: טבלת התפלגות T



P

דרגות חופש	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.000	3.078	6.314	12.709	31.821	63.657	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	6.859
6	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.405
8	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.291

© כל הזכויות שמורות לברק קנדל

שאלות

1) להלן נתונים על הוותק בעבודה (בשנים) ועל השכלה (בשנים) במדגם של 10 עובדים:

נחקר	1	2	3	4	5	6	7	8	9	10
X -וותק	13	18	2	8	16	9	5	28	17	24
Y -השכלה	12	14	17	8	11	12	13	8	12	15

מקדם המתאם חושב והתקבל: 0.31-.

א. האם קיים מתאם בין וותק העובד להשכלתו? בדקו ברמת מובהקות של 5%?

ב. אם הוותק של העובד היה נמדד בחודשים האם התשובה לסעיף א' הייתה משתנה?

2) מחקר התעניין לבדוק את הקשר בין גיל נשים בהריון לרמת ההמוגלובין שלהן בדם בזמן הריון. נדגמו 7 נשים והתקבלו התוצאות הבאות:

נחקרת	1	2	3	4	5	6	7
המוגלובין	14.7	13.5	9.7	12	10.8	13	10.3
גיל	39	34	30	29	28	26	23

במדגם חושב מדד הקשר של פירסון להיות 0.7.

א. האם ניתן לומר שבמדגם אם אישה היא יותר מבוגרת אזי בהכרח יש לה יותר המוגלובין בדם?

ב. האם ניתן לומר, ברמת מובהקות של 5%, שקיים מתאם בין גיל האישה שבהריון לבין רמת ההמוגלובין שלה בדם?

3) בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה 0.8-.

א. בדקו ברמת מובהקות של 2.5% האם קיים קשר לינארי שלילי בחודש ינואר בין הטמפרטורה במעלות צלזיוס לבין המשקעים במעלות צלזיוס.

ב. כיצד הייתה משתנה התשובה לסעיף א אם הינו מוסיפים עוד תצפיות למדגם?

ג. על סמך טבלת T המצורפת עבור אילו רמות מובהקות ניתן להחליט שקיים קשר לינארי שלילי מובהק?

4) מתווך דירות חישב את מקדם המתאם בין שטח דירה במרכז תל אביב לבין המחיר של הדירה עבור 17 דירות. מקדם המתאם שקיבל היה 0.6.

א. בדוק ברמת מובהקות של 5% האם ניתן להגיד שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה במרכז תל אביב?

ב. מהי מובהקות התוצאה לבדיקת ההשערה שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה בתל אביב.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. לא נדחה את H_0 .
ב. לא תשתנה.

2) א. לא
ב. לא נדחה את H_0 .

3) א. נדחה את H_0 .
ב. לא ניתן לדעת.
ג. לפחות 0.005.

4) א. נדחה את H_0 .
ב. $0.005 < P_V < 0.01$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 57 - רגרסיה

תוכן העניינים

1. כללי 334

מדדי קשר – רגרסיה ליניארית:

רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבוי. לבנות קו ניבויים הנקרא גם קו רגרסיה המנבא משתנה אחד על סמך האחר. מדובר בקו שמנבא את Y על סמך X.
השיטה למציאת הקו הנ"ל נקראת שיטת הריבועים הפחותים והקו המתקבל נקרא קו הרגרסיה או קו הניבויים או קו הריבועים הפחותים.

a - נותן את ערך Y כאשר X הנו אפס על גבי קו הניבויים. הוא נקרא החותך של הקו.

b - הוא שיפוע הקו נותן בכמה בעצם Y משתנה כאשר X גדל ביחידה אחת על גבי קו הניבויים.

להלן המשוואות למציאת הפרמטרים של קו הרגרסיה: $b = r\frac{S_r}{S_x}$, $Y = bX + a$.

לצורך בניית קו ניבויים לניבוי X על סמך Y נצטרך לעדכן את הנוסחאות בהתאם.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} Y_i = 200 \ , \ \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} \left(Y_i - \overline{Y}\right) = 76 \ , \ \sum_{i=1}^{20} \left(X_i - \overline{X}\right)^2 = 76 \ .$$

$$\sum_{i=1}^{20} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = 60.8$$

א. חשבו את מדד הקשר הלינארי בין X ל- Y. מיהו המשתנה התלוי?

ב. מצאו את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבירו את משמעות הפרמטרים של קו הרגרסיה.

ג. משפחת כהן הכניסה 15,000₪. מה ההוצאה הצפויה שלה?

2) נסמן ב- X את ההשכלה של אדם בשנות לימוד.
נסמן ב- Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:
$\overline{X} = 14 \ , \ \overline{Y} = 8 \ , \ \mathrm{COV}\left(X,Y\right) = 7.5 \ , \ S_Y = 5 \ , \ S_X = 2$.

א. חשבו את מדד הקשר של פירסון בין ההשכלה להכנסה.

ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?

ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000₪?

3) חוקר רצה לחקור את הקשר הקווי שבין הציון המבחן בסטטיסטיקה לבין מספר שעות ההכנה של הסטודנטים למבחן. במדגם של 100 סטודנטים שנבחנו בקורס נרשמו התוצאות הבאות: הציון הממוצע של הסטודנטים היה 65 עם סטיית תקן של 27. מספר שעות ההכנה הממוצע היה 30 עם סטיית תקן של 18. מקדם המתאם בין הציון לשעות ההכנה היה 0.8.

א. על פי משוואת הרגרסיה, שעת הכנה נוספת משפרת את ציון המבחן ב-?

ב. על פי משוואת הרגרסיה, תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון?

ג. מהו קו הרגרסיה לניבוי הציון לפי שעות ההכנה?

4) נתונים 2 משתנים X ו- Y. כמו כן נתון: $\overline{X} = 1.5$, $S_x = S_y = 4$,
וכן שקו הרגרסיה של Y על בסיס X הינו: $Y = -0.2X + 0.5$.
חשבו מהו מקדם המתאם בין X ל- Y.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. 0.8. ב. $Y = 0.8X + 0.4$. ג. 12.4 אלפי ₪.

2) א. 0.75. ב. 4.25 אלפי ₪. ג. 14.6 שנים.

3) א. 1.2. ב. 29. ג. $Y = 1.2X + 29$.

4) -0.2.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 58 - מדדי קשר-רגרסיה -שונות מוסברת ושונות לא מוסברת

תוכן העניינים

1. כללי 337

מדדי קשר – רגרסיה – שונות מוסברת ושונות לא מוסברת:

רקע:

המטרה ברגרסיה היא להסביר את השונות של המשתנה התלוי.
למשל, להסביר את השונות של המשכורת באמצעות הוותק או להסביר את השוני בציונים באמצעות כמות החיסורים.

r^2 - החלק מהשונות של המשתנה התלוי מוסבר. השונות המוסברת נקראת גם שונות ניבויים. השונות הלא מוסברת נקראת גם שונות טעויות.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) נמצא קשר חיובי בעוצמה של 0.7 בין שטח דירה למחירה. כמו כן, נתון שסטיית התקן של מחירי הדירות הינה 200.

א. איזה אחוז מהשונות של מחירי הדירות מוסבר על ידי שטח הדירה?

ב. איזה אחוז מהשונות של מחירי הדירות לא מוסבר על ידי שטח הדירה?

ג. מהי השונות המוסברות ומהי השונות הלא מוסברת של מחירי הדירות?

2) להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו!

א. אם שונות הטעויות שווה ל-0 (השונות הלא מוסברת) אז מקדם המתאם של פירסון יהיה 1.

ב. אם מקדם המתאם של פירסון בין שני משתנים הוא 1 אזי שונות הטעויות (השונות הלא מוסברת) תהיה 0.

ג. אם השונות המשותפת של X ושל Y היא 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות רב-ברירה:

3) בקשר בין שני משתנים התקבל: $r^2 = 0.64$, לכן:

א. ללא יוצא מן הכלל ככל שערכי משתנה אחד עולה השני יעלה.

ב. 64% מהשונות של משתנה אחד מוסבר על ידי המשתנה השני.

ג. הקשר בין שני המשתנים הוא בעוצמה של 0.64.

ד. כל התשובות נכונות.

4) אם מגדילים את r^2, ניתן לומר כי:

א. אחוז השונות המוסברת יקטן.

ב. אחוז השונות המוסברת יגדל.

ג. אחוז השונות המוסברת יישאר ללא שינוי.

ד. סטיית התקן משתנה.

ה. לא ניתן לדעת.

© כל הזכויות שמורות לברק קנדל

5) בקורס מבוא לכלכלה ניתנו במשך השנה שני מבחנים: מבחן בסוף סמסטר א' X ומבחן בסוף סמסטר ב' Y. כאשר בנו את קו הרגרסיה של הציון במבחן סוף סמסטר ב' לפי הציון במבחן סוף סמסטר א התקבלה שונות טעויות של 80, ושונות ניבויים של 20.
לפי נתונים אלו, מקדם המתאם בין הציון במבחן סוף סמסטר א' לבין הציון במבחן סוף סמסטר ב' הוא:

א. 0.44.

ב. 0.44 -.

ג. עוצמת ההקשר הלינארי היא 0.44, אך אין אפשרות לדעת את סימנה.

ד. אין אפשרות לחשב את מקדם המתאם.

ה. 0.35.

תשובות סופיות:

1) א. 49%. ב. 51%.
ג. שונות מוסברת: 19,600, שונות לא מוסברת: 20,400.

2) א. לא נכון. ב. נכון. ג. נכון.

3) ב'.

4) ב'.

5) ג'.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה ואמינות

פרק 59 - מבחן הבינום

תוכן העניינים

1. מבחן הבינום 340
2. חישוב רמת מובהקות ועוצמת מבחן במבחן הבינום 345

מבחנים אפרמטריים למדגם יחיד

מבחן הבינום – רקע

מבחן הבינום הינו מבחן סטטיסטי על הפרמטר p . הפרמטר p מייצג את פרופורציית ההצלחות באוכלוסייה כלומר, הסיכוי בניסוי בודד להצליח.
ההשערות האפשריות על הפרמטר הן :

השערת האפס:	$H_0 : p = p_0$	$H_0 : p = p_0$	$H_0 : p = p_0$
השערה אלטרנטיבית:	$H_1 : p \neq p_0$	$H_1 : p < p_0$	$H_1 : p > p_0$

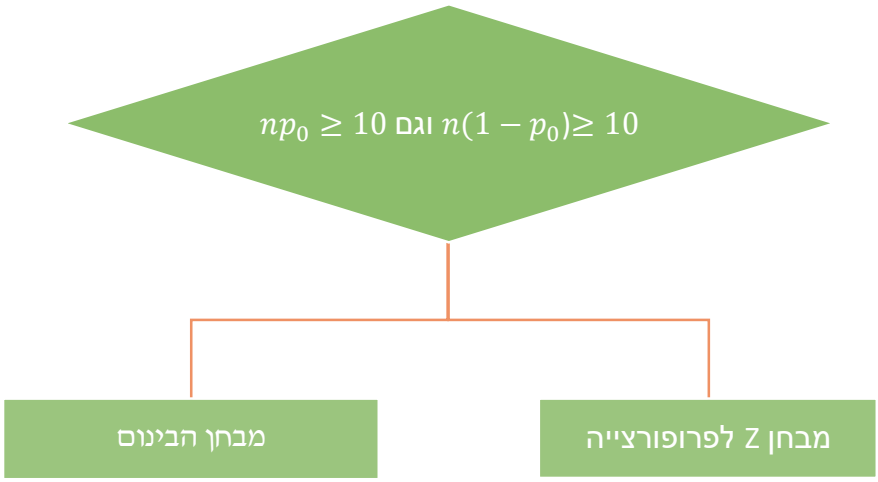
דוגמה:

במדינה אירופאית התנהל לפני 4 שנים משאל עם בדבר לגליזציה של הקנאביס.
במשאל העם 40% מהאזרחים היו בעד לגליזציה של הקנאביס ובשל כך חוק הלגליזציה לא עבר באותה המדינה. במדגם שנעשה כיום בו השתתפו 15 אזרחים מהמדינה 10 ענו שהם תומכים בלגליזציה של הקנאביס במדינה.
פרלמנטר מהמדינה חוקר האם כיום אחוז התומכים בלגליזציה של הקנאביס במדינה עלה יחסית לאחוז שהתקבל במשאל העם במדינה לפני 4 שנים.
רשמו את השערות המחקר.

© כל הזכויות שמורות לברק קנדל

במבחן הבינום מבצעים מדגם אקראי בגודל n ומתבוננים במספר ההצלחות שהתקבלו במדגם, אותן נסמן ב- Y. בהנחה והתצפיות במדגם בלתי תלויות זו בזו אנו אומרים ש: $Y \sim B(n, p)$.

מבחן הבינום נכנס לקטגוריה של מבחנים אפרמטרים והוא בא כחלופה למבחן הפרמטרי על פרופורציה אחת כאשר התנאים לקירוב הנורמלי אינם מתקיימים.



דוגמה:

מהו המבחן הסטטיסטי המתאים? נמקו.

הטכניקה הנוחה ביותר למבחן הבינום היא לחשב את מובהקות התוצאה ולדחות את השערת האפס אם $\alpha \geq PV$. מובהקות התוצאה היא הסיכוי לתוצאות של המדגם וקיצוני יותר בהנחת השערת האפס.

כזכור, פונקציית ההסתברות של ההתפלגות הבינומית היא:

$$P(Y=k)=\binom{n}{k}p^k(1-p)^{n-k} \qquad k=0,1,\ldots,n.$$

כאשר התוחלת של ההתפלגות הבינומית היא: $E(Y)=np$.

בעמוד הבא מצורפת טבלה של התפלגות בינומית מצטברת.

זו טבלה שיכולה לעזור בתהליך החישוב.

דוגמה:

חשבו את מובהקות התוצאה. מה תהיה מסקנת המחקר ברמת מובהקות של 5%?

© כל הזכויות שמורות לברק קנדל

טבלת התפלגות בינומית מצטברת

n	x	p = 0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0474	0.0064	0.0002
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

© כל הזכויות שמורות לברק קנדל

שאלות

1) אחוז האנשים שאוכלים גלידות בחודשי החורף הינו 30%, קיים חשש בקרב מיצרי הגלידות כי השנה פחת אחוז אוכלי הגלידות בחודשי החורף. לשם כך נדגמו 12 אנשים אשר מתוכם 2 טענו שהם אוכלים גלידה בחודשי החורף.

א. רשמו את השערות המוצגות בשאלה זו וציינו מהו המבחן הסטטיסטי המתאים.

ב. מה תהיה המסקנה ברמת מובהקות של 10%?

ג. מה תהיה המסקנה ברמת מובהקות גדולה יותר מ- 10%?

2) מטבע הוטל 15 פעמים במטרה לבדוק האם המטבע סימטרי.
בסך הכול התקבלו בהטלות 3 פעמים עץ.

א. בדקו ברמת מובהקות של 5% האם המטבע הוא סימטרי.

ב. עבור אילו רמות מובהקות נוכל להסיק שהמטבע אינו סימטרי?

3) אחוז המובטלים במשק לפני 5 שנים היה 10%. מעוניינים לבדוק האם כיום אחוז המובטלים קטן לעומת זה שהיה לפני 5 שנים. במדגם של 20 אנשים התקבל מובטל אחד ויחיד.

א. מהי רמת המובהקות הקטנה ביותר עבורה יוסק שכיום אחוז האבטלה נמוך מאשר לפני 5 שנים?

ב. מה תהיה המסקנה אם רמת המובהקות תהיה 5%?

4) נניח ש 40% מהאוכלוסייה מכירה את המוצרים של חברת "רמקס". החברה מתכננת לצאת בקמפיין פרסום שמטרתו להעלות את המודעות לקיומם של המוצרים של החברה באוכלוסייה. בזמן הקמפיין נדגמו 20 אנשים אקראיים מתוכם 70% טענו שהם מכירים את המוצרים של חברת "רמקס".

א. מהי מובהקות התוצאה?

ב. כיצד מובהקות התוצאה הייתה משתנה אם במדגם 80% היו טוענים שהם מכירים את המוצרים של חברת "רמקס"?

© כל הזכויות שמורות לברק קנדל

5) חברה לתוספי מזון דיווחה שנטילת מולטי ויטמין מקטינה את הסיכוי לחלות במחלות חורף במהלך החורף. לפי משרד הבריאות 70% מהאוכלוסייה חולים במחלות חורף במהלך החורף. במחקר השתתפו 20 אנשים אשר נטלו במהלך שנה מולטי ויטמין. במהלך החורף נמצא ש 12 מתוכם חלו במחלות חורף במהלך החורף. את המחקר יש לבצע ברמת מובהקות של 5%.
נסמן ב: PV את מובהקות התוצאה של המחקר.

א. האם המשפט הבא נכון?
"אם כלל האוכלוסייה הייתה נוטלת מולטי ויטמין אז בסיכוי של 5% ניתן להגיד ש-70% מהאוכלוסייה הייתה חולה במחלת חורף".

ב. האם המשפט הבא נכון?
"אם כלל האוכלוסייה הייתה נוטלת מולטי ויטמין אז בסיכוי של PV 70% מהאוכלוסייה הייתה חולה במחלת חורף".

ג. חשבו את PV.

תשובות סופיות

1) א. מבחן הבינום.
$H_0: p = 0.3$
$H_1: p < 0.3$
ב. לא נדחה את H_0 .
ג. לא ניתן לדעת.

2) א. נדחה את H_0 .
ב. לפחות 0.0352

3) א. 0.3917
ב. לא נדחה את H_0 .

4) א. 0.0065
ב. תקטן.

5) א. לא נכון.
ב. לא נכון.
ג. 0.2277.

© כל הזכויות שמורות לברק קנדל

חישוב הסתברויות לטעויות ועוצמת המבחן במבחן הבינום – רקע

מבחן הבינום הינו מבחן סטטיסטי על הפרמטר p .
הפרמטר p מייצג את פרופורציית ההצלחות באוכלוסייה כלומר, הסיכוי בניסוי בודד להצליח. ההשערות האפשריות על הפרמטר הן :

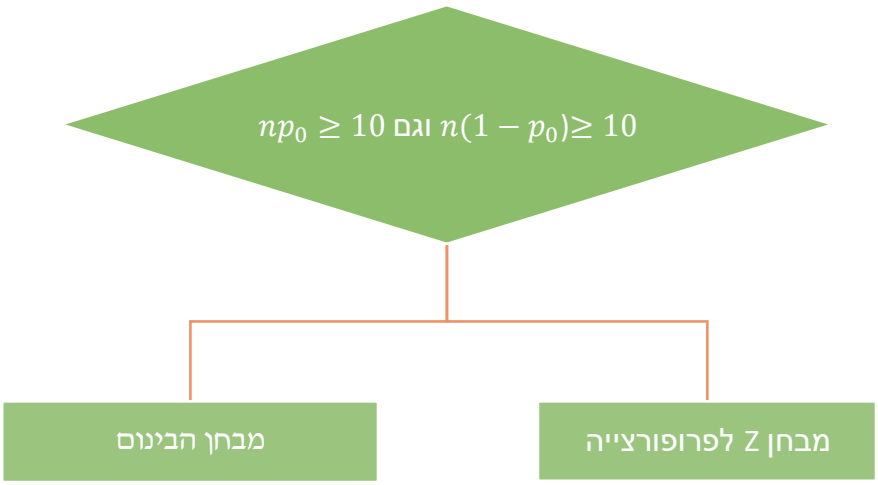
השערת האפס :	$H_0 : p = p_0$	$H_0 : p = p_0$	$H_0 : p = p_0$
השערה אלטרנטיבית :	$H_1 : p \neq p_0$	$H_1 : p < p_0$	$H_1 : p > p_0$

דוגמה :

בחברת פז 40% מעובדי החברה מאחרים לעבודה. מנכ"ל החברה מעוניין לצמצם את אחוז העובדים המאחרים בחברה. לצורך כך, בוחר המנכ"ל מחלקה אקראית במטרה לבדוק עליה שיטת תמריצים שאמורה לצמצם את אחוז המאחרים.
במחלקה שנבחרה 10 עובדים והופעלה עליה שיטת התמריצים הנבדקת.
הוסכם שאם מספר המאחרים במחלקה יהיה לכל היותר עובד אחד לאחר הפעלת שיטת התמריצים, תופעל שיטת התמריצים על כלל החברה.

- מהן השערות המחקר?

במבחן הבינום מבצעים מדגם אקראי בגודל n ומתבוננים במספר ההצלחות שהתקבלו במדגם, אותן נסמן ב- Y . בהנחה והתצפיות במדגם בלתי תלויות זו בזו אנו אומרים ש : $Y \sim B(n, p)$.
מבחן הבינום נכנס לקטגוריה של מבחנים אפרמטריים והוא בא כחלופה למבחן הפרמטרי על פרופורציה אחת כאשר התנאים לקירוב הנורמלי אינם מתקיימים.



דוגמה :

© כל הזכויות שמורות לברק קנדל

- מהו המבחן הסטטיסטי המתאים? נמקו.

כזכור, פונקציית ההסתברות של ההתפלגות הבינומית היא:

$$P(Y=k)=\binom{n}{k}p^k(1-p)^{n-k} \qquad k=0,1,\ldots,n$$.

בעמוד הבא מצורפת טבלה של התפלגות בינומית מצטברת. זו טבלה שיכולה לעזור בתהליך החישוב.
נזכיר את המצבים האפשריים בבדיקת השערות:

	הכרעה		
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

α=P(לדחות את H_0 | נכונה H_0)= P_{H_0}(לדחות H_0)

הסיכוי לבצע טעות מסוג 2:

β =P(לקבל את H_0 | נכונה H_1)=P_{H_1}(לקבל H_0)

רמת בטחון:

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

עוצמה:

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

דוגמה:

- מהי רמת המובהקות של המחקר המוצע?
- אם בשיטת התמריצים, הסיכוי לאחר הוא 10%, מה ההסתברות שהמחקר יגלה זאת?

© כל הזכויות שמורות לברק קנדל

טבלת הסתברות בינומית (מצטברת) עבור p -ים שונים

		p									
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0474	0.0064	0.0002
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

© כל הזכויות שמורות לברק קנדל

שאלות

1) במטרה לבדוק האם במוסד לימוד מסוים פרופורציית הבנים נמוכה מפרופורציית הבנות נדגמו באקראי 10 תלמידים. הוחלט שאם מספר הבנים במדגם יהיה לכל היותר 2 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות.

א. מהי רמת המובהקות במבחן המוצע?

ב. מהי העוצמה במבחן המוצע בהנחה ובמוסד 30% בנים?

ג. כיצד התשובות לסעיפים הקודמים ישתנו אם ישנו את המבחן כך שיוחלט שאם מספר הבנים במדגם יהיה לכל היותר 3 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות?
הסבירו ללא חישוב מחדש.

2) הסיכוי לזכות במשחק מסוים הינו 70%. מעוניינים לבדוק האם שיטת משחק מסוימת מעלה את סיכויי הזכייה לצורך כך מחליטים לשחק את המשחק 20 פעמים תוך שימוש בשיטה הנבדקת. מחליטים שאם מספר הזכיות במדגם יהיה לפחות 18 נקבל את הטענה שאכן השיטה עובדת והיא מעלה את סיכויי ההצלחה להיות 90%.
מה הסיכוי לבצע טעות מסוג ראשון ומה הסיכוי לבצע טעות מסוג שני במבחן שהוצע?

3) במטרה לבדוק האם מטבע הוא סימטרי הטילו אותו 8 פעמים. הוחלט שאם מספר העצים יהיה בין 1 ל 7 כולל, יוחלט שהמטבע הוגן, אחרת נחליט שהמטבע לא הוגן.

א. רשמו את השערות המחקר.

ב. מה ההסתברות להחליט שהמטבע אינו סימטרי למרות שבפועל המטבע הינו סימטרי?

ג. מהי ההסתברות שאם במטבע הסיכוי לעץ הוא 20%, אכן נגלה זאת?

ד. כיצד התשובה לסעיף הקודם תשתנה אם הסיכוי לעץ במטבע אף נמוך מ-20%?

4) נתון ש- $X \sim B(15, p)$ ההשערות של המחקר הן: $H_0: p = 0.3 \quad H_1: p > 0.3$.
כלל ההכרעה הוא: נדחה את השערת האפס אם $X \geq 8$.

א. מהי רמת המובהקות במחקר זה?

ב. מה יקרה לרמת המובהקות אם כלל ההכרעה יהיה: נדחה את השערת האפס אם $X \geq 7$? הסבירו ללא חישוב.

ג. מה יקרה לרמת המובהקות אם השערות המחקר יהיו $H_0: p = 0.3 \quad H_1: p \neq 0.3$ אבל כלל ההכרעה לא ישתנה?
הסבירו ללא חישוב.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. 0.0547 ב. 0.3828 ג. תגדלנה.

2) $\beta = 0.3231$ $\alpha = 0.0355$.

3) א. $H_0 : p = 0.5$, $H_1 : p \neq 0.5$ ב. 0.0078 ג. 0.1678

ד. העוצמה תגדל.

4) א. 0.05 ב. תגדל. ג. לא תשתנה.

© כל הזכויות שמורות לברק קנדל