<u>סטודנטים יקרים</u>

לפניכם ספר תרגילים בקורס סטטיסטיקה ב'. הספר הוא חלק מקורס חדשני וראשון מסוגו בארץ בנושא זה, המועבר ברשת האינטרנט On-line.

הקורס באתר כולל <u>פתרונות מלאים</u> לספר התרגילים, וכן את התיאוריה הרלוונטית לכל נושא ונושא.

הקורס כולו מוגש בסרטוני וידאו המלווים בהסבר קולי, כך שאתם רואים את התהליכים בצורה מובנית, שיטתית ופשוטה, ממש כפי שנעשה בשיעור פרטי, לדוגמה <u>לחצו כאן</u>.

את הקורס בנה מר ברק קנדל, מרצה מבוקש במוסדות אקדמיים שונים ובעל ניסיון עתיר בהוראת המקצוע.

אז אם אתם עסוקים מידי בעבודה, סובלים מלקויות למידה, רוצים להצטיין או פשוט אוהבים ללמוד בשקט בבית, אנחנו מזמינים אתכם לחוויית לימודים יוצאת דופן וחדשה לחלוטין, היכנסו עכשיו לאתר www.gool.co.il.

אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

גול זה בול. בִּשְׁבִילְךָ!

תוכן

פרק 1 - בעיות בסיסיות בהסתברות

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 :

בהטלת קובייה : $\{1,2,3,4,5,6\}$.
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, לקבל לפחות 5 : $A=\{5,6\}$
לקבל תוצאה זוגית : $B=\{2,4,6\}$

גודל מרחב המדגם : מספר התוצאות האפשריות במרחב המדגם:

בהטלת הקובייה : $|\Omega|=6$

גודל המאורע : מספר התוצאות האפשריות במאורע עצמו.

בהטלת הקובייה : $|A|=2$ $|B|=3$

מאורע משלים : מאורע המכיל את כל התוצאות האפשריות במרחב המדגם פרט לתוצאות במאורע אותו הוא משלים:
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.

<u>הסתברות במרחב מדגם אחיד</u> :

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

למשל, מה הסיכוי בהטלת קובייה לקבל לפחות 5? $p(A)=\frac{|A|}{|\Omega|}=\frac{2}{6}$

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

<u>הסתברות במרחב לא אחיד :</u>

יחושב לפי השכיחות היחסית : $\frac{f}{n}$

להלן התפלגות הציונים בכיתה מסוימת :

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

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

ב. מה ההסתברות שתלמיד אקראי שניבחר בכיתה יכשל?

$$\frac{f}{n}=\frac{2}{25}=0.08$$

<u>הסתברות למאורע משלים</u> :

$$p(\bar{A})=1-P(A)$$

למשל, בדוגמה הקודמת הסיכוי לעבור את הבחינה יכול להיות מחושב לפי הסיכוי להיכשל :

$$p(A)=1-\frac{2}{25}=\frac{23}{25}$$

תרגילים:

1. מהאותיות E, F ו-G יוצרים מילה בת 2 אותיות לא בהכרח בת משמעות.
   א. הרכב את כל המילים האפשריות.
   ב. רשום את המקרים למאורע:
   A- במילה נמצאת האות E.
   B- במילה האותיות שונות.
   ג. רשום את המקרים למאורע $\bar{A}$.

2. מטילים זוג קוביות.
   א. רשום את מרחב המדגם של הניסוי. האם המרחב מדגם הוא אחיד?
   ב. רשום את כל האפשרויות למאורעות הבאים:
   A- סכום התוצאות 7.
   C- מכפלת התוצאות 12.
   ג. חשב את הסיכויים למאורעות שהוגדרו בסעיף ב.

3. בוחרים 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 שנספרו עבור כל משפחה בישוב מסוים:

מספר משפחות	**מספר מקלטים**
22	0
28	1
18	2
22	3
10	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	**מספר מכוניות**
20	0
40	1
100	2
30	3
10	4

נבחרה משפחה אקראית מן הישוב.

א. מה ההסתברות שאין לה מכוניות?

ב. מה ההסתברות שבבעלות המשפחה לפחות 3 מכוניות?

ג. מה הסיכוי שבבעלותה פחות מ-3 מכוניות?

6. מטילים מטבע רגיל 3 פעמים. בצד אחד של המטבע מוטבע עץ ובצד השני פלי.

א. רשום את מרחב המדגם של הניסוי. האם המרחב מדגם הוא אחיד?

ב. רשום את כל האפשרויות למאורעות הבאים:

A- התקבל פעם אחת עץ.

D- התקבל לפחות פלי אחד.

ג. מהו המאורע המשלים ל –D.

ד. חשבו את הסיכויים למאורעות שהוגדרו בסעיפים ב- ג.

<u>פתרונות:</u>

<u>שאלה 2</u>

ג. הסיכוי ל-A: $\frac{1}{6}$

הסיכוי ל-B: $\frac{1}{9}$

<u>שאלה 3</u>

א. 0.4

ב. 0.4

ג. 0.5

<u>שאלה 4</u>

א. 0.22

ב. 0.78

ג. 0.32

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

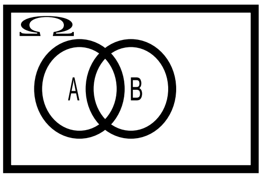
רקע:

פעולת חיתוך :

נותנת את המשותף בין המאורעות הנחתכים , חיתוך בין המאורע A למאורע B יסומן כך :

$A \bigcap B$

מדובר בתוצאות שנמצאות ב- A וגם ב-B.



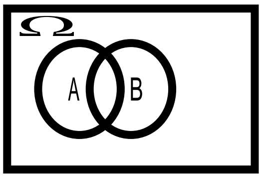
בהטלת קובייה , למשל, לקבל לפחות 5 : $A=\{5,6\}$

לקבל תוצאה זוגית : $B=\{2,4,6\}$

$$A \bigcap B=\{6\}$$

פעולת איחוד :

נותנת את כל האפשריות שנמצאות לפחות באחת מהמאורעות. הסימון הוא : $A \bigcup B$ נותנת את אשר נימצא ב-A או B. כלומר, לפחות אחד מהמאורעות קורה.



בהטלת קובייה , למשל, לקבל לפחות 5 : $A=\{5,6\}$

לקבל תוצאה זוגית : $B=\{2,4,6\}$

$$A \bigcup B=\{2,4,5,6\}$$

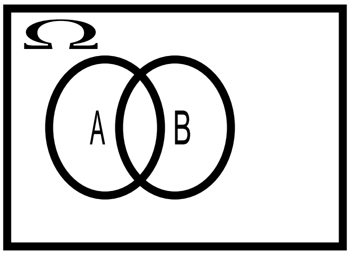
דוגמה (הפתרון נמצא בהקלטה)

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. ההסתברות לעבור את המבחן בסטטיסטיקה ובכלכלה היא 0.75.

א. מה ההסתברות שלו לעבור את המבחן בסטטיסטיקה בלבד?
ב. מה ההסתברות שלו להיכשל בשני המבחנים?
ג. מה ההסתברות לעבור לפחות מבחן אחד?

נוסחת החיבור לשני מאורעות :

$$p(A\bigcup B)=P(A)+P(B)-P(A\bigcap B)$$



חוקי דה מורגן לשני מאורעות:

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

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

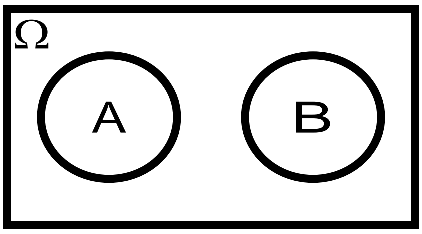
$$P(A\cup B)=1-P(\bar{A}\cap\bar{B})$$

<u>שיטת ריבוע הקסם</u>:

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

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

<u>מאורעות זרים :</u> מאורעות שאין להם מהמשותף: לא יכולים להתרחש בו זמנית.



$A\bigcap B=\{\}$

$P(A\bigcap B)=0$
$P(A\bigcup B)=P(A)+P(B)$

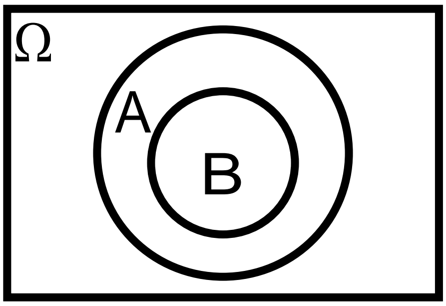
למשל , בהטלת קובייה

לקבל לפחות 5 : $A=\{5,6\}$

לקבל 3 : $B=\{3\}$

$A\bigcap B=\{\}$

<u>מאורעות מכילים :</u>



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

קשר זה מסומן באופן הבא : $\boxed{B \subset A}$

$A \cap B = B$
$A \cup B = A$

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

למשל:

$A = \{2, 4, 6\}$
$B = \{2, 4\}$

<u>תרגילים:</u>

1. מהאותיות E, F ו-G יוצרים מילה בת 2 אותיות לא בהכרח בת משמעות.
נגדיר את המאורעות הבאים :
E- במילה נמצאת האות E.
F- במילה אותיות שונות.
א. רשום את כל האפשרויות לחיתוך A עם B.
ב. רשום 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.
העזר בפעולות חיתוך , איחוד ומשלים בלבד כדי להגדיר את המאורעות הבאים וסמן בדיאגראמת וון את השטח המתאים :
א. התלמיד עבר רק את המבחן בכלכלה.
ב. התלמיד עבר רק את המבחן בסטטיסטיקה.
ג. התלמיד עבר את שני המבחנים.
ד. התלמיד עבר לפחות מבחן אחד.
ה. התלמיד נכשל בשני המבחנים.
ו. התלמיד נכשל בכלכלה.

3. נתבקשתם לבחור ספרה באקראי. נגדיר את A להיות הספרה שנבחרה היא זוגית. נגדיר את B להיות הספרה שנבחרה קטנה מ-5.
א. רשמו את כל התוצאות למאורעות הבאים :

$A =$

$B =$

$\bar{B} =$

$A \cap B =$

$A \cup B =$

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.

קבע עבור כל מאורע מה הפתרון שלו.

$\overline{\overline{A}}$

$A \cap \phi$

$A \cup \phi$

$A \cap \Omega$

$A \cup \Omega$

$A \cap \overline{A}$

$\overline{\phi}$

$A \cup \overline{A}$

5. הוגדרו המאורעות הבאים:

A= אדם שגובהו מעל 1.7 מטר

B=אדם גובהו מתחת ל-1.8 מטר

קבע 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.

השתמש 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 דוברי 2 שפות בדיוק (מהשפות הנ"ל).

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

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

ב. מה ההסתברות ששתי המפלגות תעבורנה את אחוז החסימה?

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

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

א. איזה אחוז מהעובדים הם גברים אקדמאיים?

ב. איזה אחוז מהעובדים הם גברים או אקדמאיים?

ג. איזה אחוז מהעובדים הם נשים לא אקדמאיות?

9. הסיכוי של מניה A לעלות הנו 0.5 ביום מסוים והסיכוי של מניה B לעלות ביום מסוים הנו 0.4. בסיכוי של 0.7 לפחות אחת מהמניות תעלה ביום מסוים. חשב 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 6.

C- מכפלת התוצאות בשתי הקוביות 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:

$$p(\bar{A}\cap\bar{B})=0.1 \quad p(B)=0.3 \quad p(A)=0.6$$

א. האם A ו-B מאורעות זרים?

ב. חשב את $p(\bar{A}\cap B)$

12. מטבע הוטל פעמיים. נגדיר את המאורעות הבאים:

A- קיבלנו עץ בהטלה הראשונה.

B- קיבלנו לפחות עץ אחד בשתי ההטלות.

איזו טענה נכונה?

א. A ו- B מאורעות זרים.

ב. A ו- B מאורעות משלימים.

ג. B מכיל את A.

ד. A מכיל את B.

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

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

ב. מה ההסתברות לא לזכות בפרס?

14.

$P(A)=0.3$
$P(B)=0.25$
$P(A\cup B)=0.49$

א. חשב את הסיכוי ל - $P(A\cap B)$

ב. האם A ו- B מאורעות זרים?

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

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

מה הסיכוי למאורע A ומה ההסתברות למאורע B?

16. קבע אילו מהטענות הבאות נכונות :

א. $A\bigcap B=B\bigcap A$

ב. $\overline{A\bigcup B}=A\bigcap B$

ג. $A\bigcap B\bigcap C=A\bigcap B\bigcap (C\bigcup B)$

ד. $\overline{A\bigcap B\bigcap C}=\bar{A}\bigcup \bar{B}\bigcup \bar{C}$

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

א. האם יתכן ש- 0.4= $p(A\bigcup B)$?

ב. האם יתכן ש -0.6= $p(A\bigcup B)$?

ג. אם A ו- B זרים מה הסיכוי $p(A\bigcup B)$?

ד. אם A מכיל את B מה הסיכוי $p(A\bigcup B)$?

18. מתוך אזרחי המדינה הבוגרים ל- 30% חשבון בבנק הפועלים. ל28% 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?

.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 הוכח: $p(\bar{A}\cap\bar{B})=1-P(A)-P(B)+P(A\cap B)$

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

פתרונות:

שאלה 7

א. 0.24

ב. 0.04

ג. 0.16

שאלה 8

א. 10%

ב. 50%

ג. 50%

שאלה 9

א. 0.2

ב. 0.3

ג. 0.3

שאלה 10

א. לא.

ב. כן.

ג. כן.

ד. לא.

שאלה 11

א. כן

ב. 0.3

שאלה 12

התשובה הנכונה ג

שאלה 13

א. 0.05

ב. 0.95

שאלה 14

א. 0.06

ב. לא זרים

ג. 0.43

שאלה 18

א.0.19

ב.0.05

ג. 0.31

ד. 0.46

ה.0.12

ו.0.41

ז.0.59

פרק 3 - קומבינטוריקה - כלל המכפלה

רקע:

כלל המכפלה:

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

אם לתהליך יש k שלבים : n_1 אפשריות 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)

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

תרגילים:

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 אנטרקוט, 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:
      1. בארוחה סלט ירקות, לזניה בשרית ולימונדה.
      2. 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 ספרות זהות.
   ו. המספר הוא פלינדרום (מספר הנקרא מימין ומשמאל באותה צורה).

4. חמישה אנשים אקראיים נכנסו למעלית בבנין בן 8 קומות. חשבו את ההסתברויות הבאות:
   א. כולם ירדו בקומה החמישית?
   ב. כולם ירדו באותה קומה?
   ג. כולם ירדו בקומה אחרת?
   ד. ערן ודני ירדו בקומה השישית והיתר בשאר הקומות?

5. במפלגה חמישה עשר חברי כנסת. יש לבחור שלושה חברי כנסת לשלושה תפקידים שונים. בכמה דרכים ניתן לחלק את התפקידים 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 של התוצאות תהינה שונות?
   ג. מה ההסתברות שלפחות שתי תוצאות תהינה זהות?
   ד. מה ההסתברות שלפחות שתי תוצאות תהינה שונות?

7. יש ליצור מילה בת חמש אותיות לא בהכרח עם משמעות מאותיות ה- ABC (26 אותיות) בת 5 אותיות.
   א. מה ההסתברות שבמילה שנוצרה אין האותיות A, D ו L?
   ב. מה ההסתברות שבמילה שנוצרה כל האותיות זהות?
   ג. מה ההסתברות שבמילה שנוצרה לפחות שתי אותיות שונות זו מזו?
   ד. מה ההסתברות שהמילה היא פלינדרום (מילה אשר משמאל לימין, ומימין לשמאל נקראת אותו הדבר).

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 בסימון V או בסימון X. בכמה דרכים שונות ניתן למלא את טופס משחק המזל?

פתרונות :

שאלה 1

א. 36

ב. 900

ג. 70

ד. 90

שאלה 2

א. 36

ב. 1. 1/36

2. 1/9

שאלה 3

א. 0.5

ב. 0.3024

ג. 0.0001

ד. 0.9999

ה. 0.6976

ו. 0.01

שאלה 4

א. 0.00003

ב. 0.00024

ג. 0.20508

ד. 0.01047

שאלה 5

א. 3,375

ב. 2,730

שאלה 6

א. 1/216

ב. 5/18

ג. 13/18

ד. 215/216

שאלה 7

א. 0.5417

ב. $\frac{1}{26^4}$

ד. 0.0015

שאלה 9

2^n

פרק 4 - קומבינטוריקה- תמורה - סידור עצמים בשורה

רקע:

תמורה:

מספר האפשריות לסדר n עצמים שונים בשורה :

$$n!=1\cdot 2\cdot 3\cdots(n-1)\cdot n$$

הערה: $0!=1$

למשל , בכמה דרכים שונות ניתן לסדר את האותיות a,b,c,d ? (הפתרון בהקלטה)

למשל , בכמה דרכים שונות ניתן לסדר את האותיות a,b,c,d , כך שהאותיות a,b יהיו ברצף? (הפתרון בהקלטה)

למשל , בכמה דרכים שונות ניתן לסדר את האותיות a,b,c,d , כך שהאותיות a,b יופיעו בתור הרצף ba ? (הפתרון בהקלטה)

תרגילים:

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 להם אותו ציון.

א. מהו מספר הדירוגים האפשריים?

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

4. מסדרים 10 ספרים שונים על מדף.

א. בכמה אופנים ניתן לסדר את הספרים על המדף?

שני ספרים מתוך ה-10 הם ספרים בסטטיסטיקה.

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

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

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

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

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

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

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

6. 4 בנים ו-4 בנות התיישבו באקראי בשורת קולנוע בכיסאות 1-8.

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?

פתרונות :

שאלה 1

א. 24

ב. 120

שאלה 2

א. 0.2

ב. 0.8

ג. 0.022

שאלה 3

א. 362,880

ב. 2,880

שאלה 4

א. 3,628,800

ב. 0.2

ג. 0.8

ד. $\frac{1}{45}$

שאלה 5

א. $\frac{1}{792}$

ב. $\frac{1}{99}$

ג. $\frac{1}{4620}$

שאלה 6

א. 0.75

ב. 0.014

ג. $\frac{1}{14}$

ד. $\frac{1}{35}$

פרק 5 - קומבינטוריקה - תמורה עם עצמים זהים

רקע:

תמורה עם חזרות :

אם יש בין העצמים שיש לסדר עצמים זהים יש לבטל את הסידור הפנימי שלהם על ידי חלוקה בסידורים הפנימיים שלהם.

מספר האופנים לסדר n עצמים בשורה , ש- n_1 מהם זהים מסוג 1 , n_2 זהים מסוג 2, ..., n_r זהים מסוג r :

$$\frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_r!}$$

למשל ,

כמה מילים ניתן ליצור מכל האותיות הבאות : W W T T K K ? (תשובה בהקלטה)

תרגילים:

1. במשחק יש לצבוע שתי משבצות מתוך המשבצות הבאות :

בכמה דרכים שונות ניתן לבצע את הצביעה?

2. בכמה אופנים שונים אפשר לסדר בשורה את האותיות ב ע ע ב ע ג?

3. בבית נורות מקום ל-6 נורות. בחרו שתי נורות אדומות, שתי נורות צהובות ושתי נורות כחולות.
כמה דרכים שונות יש לסדר את הנורות?

4. רוצים ליצור מספר מכל הספרות הבאות : 1,2,2,2,6
כמה מספרים כאלה אפשר ליצור?

5. במשחק בול פגיעה יש 10 משבצות, אדם צובע 4 משבצות מתוך ה-10. המשתתף השני צריך לנחש אילו 4 משבצות נצבעו. מה ההסתברות שבניחוש אחד יהיה בול פגיעה?

6. כמה אותות שונים , שכל אחד מורכב מ 10 דגלים שונים ניתן ליצור אם 4 דגלים הם לבנים , 3 כחולים , 2 אדומים ואחד שחור. דגלים שווי צבע זהים זה לזה לחלוטין.

<u>פתרונות:</u>

3. 90
4. 20
5. $\frac{1}{210}$
6. 12,600

פרק 6 - קומבינטוריקה - דגימה סידורית ללא החזרה ועם החזרה

רקע:

מדגם סדור בדגימה עם החזרה

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

למשל,
בוחרים שלושה תלמידים מתוך עשרה לייצג ועד בו תפקידים שונים, תלמיד יכול למלא יותר מתפקיד אחד.
כמה ועדים שונים ניתן להרכיב?

$n=10$

$k=3$

$10^3=1,000$

מדגם סדור ללא החזרה

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

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

למשל,
שלושה תלמידים נבחרים מתוך 10 לייצג וועד בו תפקידים שונים . תלמיד לא יכול למלא יותר מתפקיד אחד.

$$\frac{10!}{7!} = 10 \cdot 9 \cdot 8 = 720$$

תרגילים:

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.

פתרונות :

שאלה 1 :

א. 8000

ב. 6840

שאלה 2 :

א. 0.0001

ב. 0.6561

ג. 0.3439

שאלה 3 :

0.476

שאלה 4 :

א. 0.04

ב. 0.48

פרק 7 - 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. $$\binom{n}{n}=\binom{n}{0}=1$$

תרגילים :

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 במקצועות מדעי הרוח.
   3 במקצועות מדעי החברה.
   2 מתחום המתמטיקה.

   א. כמה בחירות שונות הוא יכול ליצור לעצמו?

   ב. כמה בחירות יש לו בהן 3 קורסים הם ממדעי הרוח?

   ג. כמה בחירות יש לו אם 2 מהן **לא** ממדעי הרוח?

   ד. כמה בחירות יש לו אם 2 ממדעי הרוח, 2 ממדעי החברה ו-1 ממתמטיקה?

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

   א. מה ההסתברות שהמשלחת תורכב רק מבנות?

   ב. מה ההסתברות שבמשלחת תהיה רק בת אחת?

   ג. מה ההסתברות שבמשלחת תהיה לפחות בת אחת?

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

   א. מה ההסתברות שבמשחק הזוכה כל המספרים הם זוגיים?

   ב. מה ההסתברות שבמספר הזוכה יש לכל היותר מספר זוגי אחד?

   ג. מה ההסתברות שבמספר הזוכה לפחות פעם אחת יש מספר זוגי?

   ד. מה ההסתברות שבמספר הזוכה כל המספרים גדולים מ-30?

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

א. מה ההסתברות שעודד יקבל את כל המלכים וערן את כל המלכות?

ב. מה ההסתברות שאחד השחקנים יקבל את הקלף אס-לב?

ג. מה ההסתברות שערן יקבל קלפים שחורים בלבד ועודד יקבל שני קלפים שחורים בדיוק?

ד. מה ההסתברות שערן יקבל לפחות 3 קלפים שהם מספר (אס אינו מספר)?

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

א. כל המזכירות בוועד יהיו ממסלול "מדעי ההתנהגות".

ב. כל המזכירות בוועד יהיו מאותו המסלול.

ג. מכל מסלול תבחר לפחות מזכירה אחת.

7. הוכח 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.

פתרונות:

שאלה 1

א. 53,130

ב. 20,475

ג. 3003

שאלה 2

א. 252

ב. 100

ג. 100

ד. 60

שאלה 3

א. 0.1117

ב. 0.1445

ג. 0.9819

שאלה 4

א. 0.02

ב. 0.187

ג. 0.972

ד. 0.00246

שאלה 6

א. $6.45 \cdot 10^{-5}$

ב. $2.58 \cdot 10^{-4}$

ג. 0.3225

שאלה 8

א. $\binom{2n}{n}^2$

ב. $\sum_{i=1}^{n} \binom{2n}{i}^2$

פרק 8 - קומבינטוריקה- שאלות מסכמות

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

3. 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 ע"י אימו לסדר 6 משחקים בארונית. הילד מכניס את המשחקים באקראי למגירות השונות. כל מגירה יכולה להכיל גם 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 שה"דומינו" יוכנס למגירה העליונה ויתר המשחקים לשאר המגירות.
   ד. מה ההסתברות שה"דומינו" לא יוכנס למגירה העליונה?

5. בעיר מסוימת מתמודדות למועצת העיר 4 מפלגות שונות: "הירוקים", "קדימה", "העבודה" ו"הליכוד". 6 אנשים אינם יודעים למי להצביע, ולכן בוחרים באקראי מפלגה כלשהי.
   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 הם רצו לראות סרט. באפשרותם ספריה המונה 8 סרטים שונים. כל אחד התבקש לבחור סרט באקראי.
   א. מה ההסתברות שכולם ייבחרו 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, השנייה 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. בוחרים באקראי 5 מספרים. מה ההסתברות שהמספר 7 הוא השני בגודלו מבין המספרים שנבחרו?

10. 6 אנשים עלו לאוטובוס שעוצר ב-10 תחנות. כל אדם בוחר באופן עצמאי ואקראי באיזו תחנה לרדת.

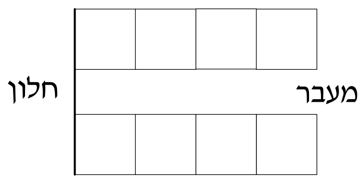
א. מה ההסתברות שכל אחד יורד בתחנה אחרת?

ב. מה ההסתברות שבדיוק 3 ירדו בתחנה החמישית?

ג. מה ההסתברות שרונית תרד בתחנה השנייה והשאר לא?

ד. מה ההסתברות שכולם ירדו בתחנות 5,6 ולפחות אחד בכל אחת מהתחנות הללו?

11. ברכבת 4 מקומות ישיבה עם כיוון הנסיעה 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.
בטא באמצעות n את הסיכוי שלכל אחד מהם אותו מספר פעמים של התוצאה "ראש".

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

16. מטילים זוג קוביות 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?

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 ?

פתרונות:

שאלה 1

ד. 102,400,000

ה. 78,960,960

ו. 658,0088

שאלה 2

ג. 810,000

ד. 657,720

ה. 27,405

שאלה 3

ד. 14,040,000

ה. 1,404,000

ו. 5,616,000

ז. 8,424,000

שאלה 4

א. 0.00024

ב. 0.00098

ג. 0.05933

ד. 0.75000

שאלה 5

ד. 0.00098

ה. 0.17798

ו. 0.02929

ז. 0.02197

<u>שאלה 6</u>

א. $\frac{1}{4096}$

ב. $\frac{1}{32,768}$

ג. 0.205

ד. 0.795

ה. 0.0105

ו. 0.5129

ז. 0.1071

<u>שאלה 7</u>

א. 4200

ב. 50,400

ג. 604,800

<u>שאלה 8</u>

א. 604,800

ב. 2,880

ג. 2,880

<u>שאלה 9</u>

0.238

<u>שאלה 10</u>

א. 0.1512

ב. 0.014

ג. 0.059

ד. $\frac{62}{10^6}$

<u>שאלה 11</u>

א. 40,320

ב. 0.1071

ג. 0.2142

ד. 0.0357

ה. 0.5714

ו. 0.1429

ז. 0.0143

ח. 0.0095

<u>שאלה 14</u>

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

<u>שאלה 16</u>

לפחות 25

<u>שאלה 17</u>

א. 0.35721

ב. 0.1759

<u>שאלה 18</u>

א. 0.75

ב. 0.075

ג. 0.375

ד. 0.15

<u>שאלה 19</u>

א. 0

ב. $\frac{450}{729}$

ג. $\frac{90}{729}$

<u>שאלה 20</u>

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

<u>שאלה 21</u>

א. 720

ב. 360

ג. 432

<u>שאלה 22</u>

א. $\dfrac{j^k-(j-1)^k}{6^k}$

ב. $\dfrac{(7-i)^k-(6-i)^k}{6^k}$

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

פרק 9 - 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 אשר ידוע שמאורע אחר התרחש/ לא התרחש.

ההסתברות של A בהינתן ש B- כבר קרה:

$$P\left(A|B\right)$$

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

למשל, (פתרון בהקלטה)

מטילים קובייה.
נגדיר:
A – התוצאה זוגית.
B – התוצאה גדולה מ-3 .
נרצה לחשב את :
$P\left(A|B\right)$

תרגילים:

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

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

5. אדם הטיל זוג קוביות והתקבל שהתוצאות זהות. מה הסיכוי שלפחות אחת התוצאות 5?

6. אדם הטיל זוג קוביות והתקבל לפחות פעם אחת 4. מה הסיכוי שאחת התוצאות 5?

7. נבחרה משפחה בת שני ילדים. ידוע שאחד הילדים בן. מה ההסתברות שבמשפחה שני בנים בקרב הילדים?

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

השאלות הבאות משלבות קומבינטוריקה:

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

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

פתרונות:

שאלה 1

0.2

שאלה 2

1/3

שאלה 3

0.5

שאלה 4

0

שאלה 5

1/6

שאלה 6

2/11

שאלה 7

1/3

שאלה 8

3/4

שאלה 9

2/3

שאלה 10

1/4

פרק 10 - הסתברות מותנית - מרחב לא אחיד

רקע:

הסיכוי שמאורע A יתרחש בהינתן ש – מאורע 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 שנתון שהתרחש :

למשל,

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.

אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית? (פתרון בהקלטה)

תרגילים:

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 והסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9. הסיכוי 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 שני המבחנים?

2. במדינה שתי חברות טלפון סלולארי "סופט" ו"בלי". 30% מהתושבים הבוגרים רשומים אצל חברת "בלי". 60% מהתושבים הבוגרים רשומים אצל חברת "סופט".
ל-15% מהתושבים הבוגרים אין טלפון סלולארי בכלל.
   א. איזה אחוז מהתושבים הבוגרים רשומים אצל שתי החברות?
   ב. נבחר אדם שרשום אצל חברת "סופט", מה ההסתברות שהוא רשום גם אצל חברת "בלי"?
   ג. אם אדם לא רשום אצל חברת "בלי", מה ההסתברות שהוא כן רשום בחברת "סופט"?
   ד. אם אדם רשום אצל חברה אחת בלבד, מה ההסתברות שהוא רשום בחברת "סופט"?

3. במכללה שני חניונים: חניון קטן וחניון גדול. בשעה 08:00 יש סיכוי של 60% שבחניון הגדול יש מקום, סיכוי של 30% שבחניון הקטן יש מקום וסיכוי של 20% שבשני החניונים יש מקום.
   א. מה ההסתברות שיש מקום בשעה 08:00 רק בחניון הגדול של המכללה?
   ב. ידוע שבחניון הקטן יש מקום בשעה 08:00, מה הסיכוי שבחניון הגדול יש מקום?
   ג. אם בשעה 08:00 בחניון הגדול אין מקום, מה ההסתברות שבחניון הקטן יהיה מקום?
   ד. נתון שלפחות באחד מהחניונים יש מקום בשעה 08:00, מה ההסתברות שבחניון הגדול יש מקום?

4. נלקחו 200 שכירים ו-100 עצמאים, מתוך השכירים 20 הם אקדמאיים, מתוך העצמאיים 30 הם אקדמאיים.
   א. בנו טבלת שכיחות משותפת לנתונים.
   ב. נבחר אדם אקראי מהי 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 הבן אדם שנבחר הוא לא אקדמאי, מה ההסתברות שהוא שכיר?

5. 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, 5% מחזיקים בכל שלושת הכרטיסים הנ"ל.

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

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

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

פתרונות:

שאלה 1

א. 0.833

ב. 0.9375

ג. 0.0625

ד. 0.5

ה. 0.789

שאלה 2

א. 5%

ב. 0.0833

ג. 0.786

ד. 0.6875

שאלה 3

א. 0.4

ב. $\frac{2}{3}$

ג. 0.25

ד. $\frac{6}{7}$

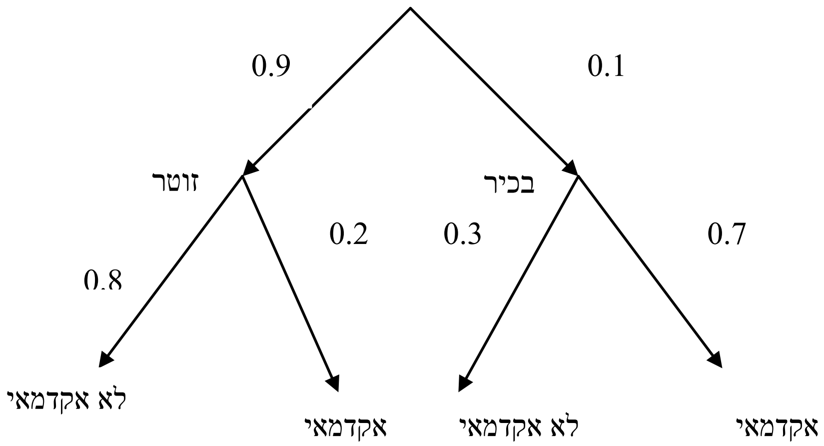
פרק 11 - דיאגרמת עצים, נוסחת בייס ונוסחת ההסתברות השלמה

רקע:

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

למשל,
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.

א. מה הסיכוי שהוא בכיר אקדמאי ?

$0.1*0.7=0.07$

ב. מה הסיכוי שהוא זוטר לא אקדמאי ?

$0.9*0.8=0.72$

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

ג. מה הסיכוי שהוא אקדמאי ?

0.1*0.7+0.9*0.2=0.25

ד. נבחר אקדמאי מה ההסתברות שהוא עובד זוטר?

מדובר כאן על שאלה בהסתברות מותנה ולכן נשתמש בעיקרון של הסתברות מותנה

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

<u>נוסחת ההסתברות השלמה</u>

B מאורע כלשהו, $A_1,\dots,A_n$ חלוקה ממצה של Ω.

אזי: $P(B) = \sum_{i=1}^{n} P(A_i)\cdot P(B/A_i)$

<u>נוסחת בייס</u>

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

תרגילים:

1. בשקית סוכריות 4 סוכריות תות ו-3 לימון . מוציאים באקראי סוכרייה אם היא בטעם תות אוכלים אותה ומוציאם סוכרייה נוספת , אך אם היא בטעם לימון מחזירים אותה לשקית ומוציאים סוכרייה נוספת.
   א. מה ההסתברות שהסוכרייה הראשונה שהוצאה בטעם תות והשנייה בטעם לימון ?
   ב. מה ההסתברות שהסוכרייה השנייה בטעם לימון?

2. באוכלוסיה מסוימת 30% הם ילדים, 50% בוגרים והיתר קשישים. לפי נתוני משרד הבריאות הסיכוי שילד יחלה בשפעת במשך החורף הוא 80%, הסיכוי שמבוגר יחלה בשפעת במשך החורף הוא 40% והסיכוי שקשיש יחלה בשפעת במשך החורף הוא 70%.
   א. איזה אחוז מהאוכלוסייה הינו קשישים שלא יחלו בשפעת במשך החורף?
   ב. מה אחוז האנשים שיחלו בשפעת במשך החורף?
   ג. נבחר אדם שחלה במשך החורף בשפעת, מה ההסתברות שהוא קשיש?
   ד. נבחר ילד, מה ההסתברות שהוא לא יחלה בשפעת במשך החורף?

3. בכד א' 5 כדורים כחולים ו-5 כדורים אדומים. בכד ב' 6 כדורים כחולים ו-4 כדורים אדומים. בוחרים באקראי כד, מוציאים ממנו כדור ומבלי להחזירו מוציאים כדור נוסף.
   א. מה ההסתברות ששני הכדורים שיוצאו יהיו בצבעים שונים?
   ב. אם הכדורים שהוצאו הם בצבעים שונים, מה ההסתברות שהכדור השני שהוצא יהיה בצבע אדום?

4. חברת סלולר מסווגת את לקוחותיה לפי 3 קבוצות גיל: נוער, בוגרים ופנסיונרים. נתון כי: 10% מהלקוחות בני נוער, 70% מהלקוחות בוגרים והיתר פנסיונרים. מתוך בני הנוער 90% מחזיקים בסמארט-פון, מתוך האוכלוסייה הבוגרת ל 70% יש סמארט-פון ומתוך אוכלוסיית הפנסיונרים 30% מחזיקים בסמארט-פון.
   א. איזה אחוז מלקוחות החברה הם בני נוער עם סמארט-פון?
   ב. נבחר לקוח אקראי ונתון שיש לו סמארט-פון. מה ההסתברות שהוא פנסיונר?
   ג. אם ללקוח אין סמארט-פון, מה ההסתברות שהוא לא בן נוער?

5. כדי להתקבל למקום עבודה יש לעבור שלושה מבחנים. המבחנים הם בשלבים, כלומר אם נכשלתם במבחן מסוים אינכם ניגשים 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 ש האנייה 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%. במחלות B,C הסימפטום מתגלה בסיכוי של 90% בכל מחלה.

א. מה ההסתברות שאדם הגיע לקליניקה וגילו אצלו את סימפטום X?

ב. אם התגלה אצל אדם סימפטום X, מה ההסתברות שהוא חולה במחלה A?

ג. אם לאדם יש את סימפטום X, מה ההסתברות שהוא חולה במחלה A?

ד. אם לא גילו אצל אדם את סימפטום X, מה ההסתברות שהוא בריא?

9. סטודנט ניגש למבחן אמריקאי. הסיכוי שהוא יודע לשאלה מסוימת את התשובה הוא p , אם 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 :

א. דולב, רונית, דולב

ב. 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?

פתרונות:

שאלה 1

א. 2/7

ב. 23/49

שאלה 2

א. 6%

ב. 58%

ג. 0.241

ד. 0.2

שאלה 3

א. 0.544

ב. 0.5

שאלה 4

א. 9%

ב. 0.09375

ג. 0.9722

שאלה 8

א. 0.0886

ב. 0.2889

ג. 0.3137

ד. 0.8778

שאלה 9

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

שאלה 10

אפשרות א עדיפה

פרק 12 - 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 A אינו תלוי גם ב- B.

כשהמאורעות בלתי תלויים מתקיים ש : $P(A\bigcap B)=P(A)\cdot P(B)$.

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 :

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

ב. מה הסיכוי להיכשל בשני הניסויים ?

באופן דומה : $p(\bar{A}\bigcap \bar{B})=P(\bar{A})\cdot P(\bar{B})=(1-0.7)(1-0.4)=0.18$

<u>הרחבה</u> : <u>אי תלות בין n מאורעות</u>

n מאורעות $A_1,\ldots,A_n$ הם בלתי תלויים אם ורק אם :

$$P(\bigcap_{i=1}^{n}A_i)=\prod_{i=1}^{n}P(A_i)$$

תרגילים:

1. נתון:

$$p(A)=0.2$$
$$P(B)=0.5$$
$$P(A\cup B)=0.6$$

האם המאורעות הללו בלתי תלויים?

2. תלמיד ניגש לשני מבחנים שהצלחתם לא תלויה זו בזו. הסיכוי שלו להצליח במבחן הראשון הוא 0.7 והשני 0.4 .
   א. מה הסיכוי להצליח בשני המבחנים יחדיו?
   ב. מה הסיכוי שניכשל בשני המבחנים ?

3. במדינה מסוימת 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. מדינה מסוימת 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. הוכח אם

$$P(A/B)=P(B/A)$$

אז :

$$P(A)=P(B)$$

8. קבע אילו מהטענות הבאות נכונות. נמק!

א. אם $p(A\bigcup B)=p(A)\bullet p(B)$ אזי המאורעות בלתי תלויים.

ב. מאורע A כלול במאורע B. $P(A)>0$, $0<p(B)<1$ 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 A‏ו- B מאורעות זרים.

ה. $P(\overline{A}\bigcap \overline{B})=1-P(A)-P(B)$ לכן A ו- B מאורעות זרים.

פתרונות :

שאלה 1

כן

שאלה 2

א. 0.28

ב. 0.18

שאלה 3

א. 0.0064

ב. 0.1536

שאלה 4

א. 0.5904

ב. 0.9984

שאלה 8

א. לא נכון

ב. לא נכון

ג. נכון

ד. לא נכון

ה. נכון

פרק 13 - שאלות מסכמות בהסתברות

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 נימצא 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

בהסתברות של 0.7 הם יצאו לפחות לאחד מהם, באולינג/קפה.

א. מה ההסתברות שהם יצאו רק לבאולינג?

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

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

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

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

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. באוכלוסיה 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, ב- 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, הוא 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 יראו את 2 המודעות?

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

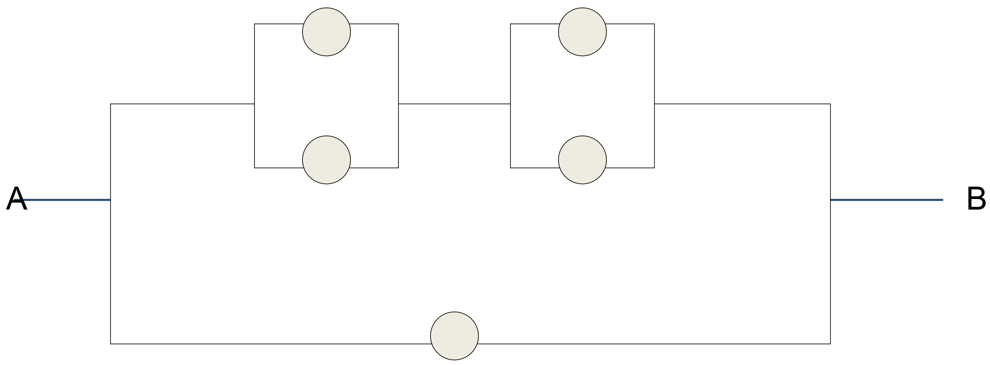
ג. האם המאורעות: "לראות את המודעה באינטרנט" ו"לראות את המודעה בעיתון" בלתי תלויים?

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.

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

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

8. נתונה המערכת החשמלית הבאה:



כל יחידה עובדת באופן בלתי תלוי ובהסתברות p.

כדי שהמערכת תפעל צריך לעבור זרם מהנקודה A לנקודה B.

הוכח שהסיכוי שהמערכת תפעל:

$$P+(1-P)(2P-P^2)^2$$

9. ליאת מעוניינת לתרגל לבחינה בהסתברות. היא מצאה באינטרנט מאגר הכולל 25 שאלות מבחינות. השאלות ממוספרות כאשר 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 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 המספור 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- $\bar{B}$ בלתי תלויים.

12. משחקים משחק מזל פעמיים בכל 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 - משחק הראשון היה ניצחון. מה ערכו של P עבורו A ו- B יהיו מאורעות בלתי תלויים?

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

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

פתרונות:

שאלה 1

א. 0.25

ב. 0.75

ג. 0.6

ד. 0.5

שאלה 2

א. 0.2

ב. 0.267

ג. 0.524

ד. המאורעות תלויים.

שאלה 3

א. 0.2

ב. המאורעות אינם זרים.

ג. המאורעות הללו תלויים.

ד. 0.06

שאלה 4

א. 0.94

ב. 0.255

ג. 0.03

ד. 0.168

שאלה 5

א. 15%

ב. 0.692

ג. לא זרים ותלויים.

שאלה 6

א. 0.65

ב. A ו- B תלויים

ג. 0.5384

שאלה 7

א. 8%

ב. 0.733

ג. תלויים

ד. 1. 0.478

2. 0.15

שאלה 8

הוכחה

שאלה 9

א. $\frac{19}{480,700}$

ב. $\frac{27,132}{480,700}$

ג. 0.4015

שאלה 10

הדוגמה בהקלטה

שאלה 11

הוכחה

שאלה 12

$\frac{1}{2}$

שאלה 13

הוכחה

שאלה 14

הוכחה

פרק 14 - המשתנה המקרי הבדיד - פונקציית ההסתברות

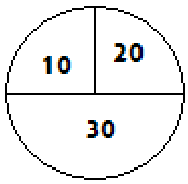
רקע:

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

פונקצית הסתברות : פונקציה המתאימה לכל ערך אפשרי של המשתנה את ההסתברות שלה.

סכום ההסתברויות על פונקציית ההסתברות חייב להיות 1.

למשל, בקזינו יש רולטה כמוראה בשרטוט :



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה בש״ח.
בנו את פונקציית ההסתברות של סכום הזכייה במשחק בודד (פתרון בהקלטה).

תרגילים:

1. ידוע שביישוב מסוים התפלגות מספר המכוניות למשפחה הוא :

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 C,B,A 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 והסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9. הסיכוי לעבור את שני המבחנים הנו 0.75. יהי X מספר המבחנים שהסטודנט עבר. בנה את פונקצית 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. בנה את פונקצית 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. הסיכוי שפרויקט ג' יצליח הינו 0.9. נתון שהצלחת כל פרויקט בלתי תלויה זו בזו. נגדיר את X להיות מספר הפרויקטים שיצליחו. בנה את פונקצית ההסתברות של X.

6. להלן פונקציית הסתברות של משתנה מקרי כלשהו :

$$P(X=k)=\frac{k}{A}$$
$$k=1,2...4$$

מצא את ערכו של A.

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

8. בסקר שנערך בדקו בקרב אנשים האם הם צופים במהדורת החדשות של ערוצים 1,2,10

להלן הנתונים :

20% צופים בערוץ 2.

8% צופים בערוץ 1.

10% צופים בערוץ 10.

כמו כן נתון ש 1% צופים בשלושת המהדורות גם יחד.

10% צופים בשתי המהדורות מתוך השלושה.

נגדיר את X להיות מספר המהדורות מבין 3 המהדורות המדוברות שאדם אקראי צופה. בנו את פונקציות ההסתברות של X.

פתרונות

שאלה 3

x	0	1	2
P(x)	0.05	0.20	0.75

שאלה 4

x	1	2	3	4
P(x)	0.3	0.21	0.147	0.343

שאלה 5

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

שאלה 6

10

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

רקע:

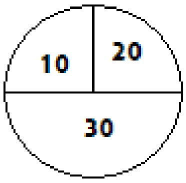
$$E(X) = \sum_i x_i P(x_i) = \mu$$
$$V(X) = \sum_i (x_i - \mu)^2 P(x_i) = \sum_i x_i^2 P(x_i) - \mu^2 = \sigma^2$$

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

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

סטיית תקן- שורש של השונות. – הפיזור הממוצע הצפוי סביב התוחלת.

למשל, בקזינו רולטה כמוראה בשרטוט :



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה בש"ח.

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 :

$$\sigma_x = \sqrt{V(X)} = \sqrt{68.75} = 8.29$$

תרגילים:

1. אדם משחק במשחק מזל. נגדיר את X להיות סכום הזכייה. להלן פונקצית 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 של הישוב. ל-40% חשבון בנק בסניף הלאומי של הישוב. ל20% מהתושבים הבוגרים אין חשבון בנק בישוב. יהי X מס' סניפי הבנק שלבוגר בישוב יש חשבון. חשב 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 תקן 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 לא להשתמש בו שוב. נסמן ב-X את מספר הניסיונות עד שהדלת תפתח.

א. בנה את פונקצית ההסתברות של X.

ב. חשב את התוחלת והשונות של X.

5. נתונה פונקצית ההסתברות של המשתנה המקרי X:

x	2	4	6	8
P(x)		0.3		0.2

כמו כן נתון ש: $E(X)=4.2$

א. מצא את ההסתברויות החסרות בטבלה.

ב. חשב את $V(X)$.

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

7. להלן ההתפלגות של משתנה מקרי X .

P	X
¼	1
½	3
¼	K

מהו הערך K שייתן ערך מינימלי לשונות של X.

פתרונות:

שאלה 1

תוחלת : 2   שונות : 796

שאלה 3

ב . תוחלת : 3.36   סטיית תקן : 1.603

שאלה 4

א.

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

ב. תוחלת : 3

שונות 2

שאלה 5

א.

x	2	4	6	8
P(x)	0.4	0.3	0.1	0.2

ב. 5.16

שאלה 6

x	-5	0	5
P(x)	0.2	0.6	0.2

שאלה 7

2.33

פרק 16 - המשתנה המקרי הבדיד - טרנספורמציה לינארית

רקע

מצב שבו מבצעים הכפלה של קבועה ו או הוספה של קבוע על המשתנה המקורי. (כולל גם חלוקה של קבוע והחסרה של קבוע)

אם $Y = aX + b$

אזי :

$$E(Y) = aE(X) + b$$

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

$$\sigma_Y = |a|\sigma_x$$

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 :

$$E(X) = 22.5 = \mu$$

$$V(X) = 68.75 = \sigma^2$$

תרגילים:

1. סטודנט ניגש ל- 5 קורסים הסמסטר. נניח שכל קורס שסטודנט מסיים מזכה אותו ב-4 נקודות אקדמאיות. חשב 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 הינו 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 לעלות 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 $Y=7-X$.
חשב את: $E(Y)$ ו- $V(Y)$.

5. אדם החליט לבטח את רכבו, שווי רכבו 100,000 ₪.
להלן התביעות האפשריות והסתברותן:
בהסתברות של 1/1000 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 פונקצית 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 שגוייה, מופחתת נקודה. מהי התוחלת ומה השונות של הציון בבחינה?

7. להלן פונקצית 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: $\frac{X}{2}-4$

פתרונות :

שאלה 1:

תוחלת : 14 שונות : 32

שאלה 2:

תוחלת : 8 שונות : 12

שאלה 3:

תוחלת : 13.2

סטיית תקן : 5.5

שאלה 4:

תוחלת: 3

שונות : 3

שאלה 6:

ב. $V(X)=1.8275$

שאלה 7 :

א. A=10

ב. $E(X)=3$
$V(X)=1$

ג. $E(X^3)=35.4$
$V(X^3)=616.84$

ד. $E(y)=-2.5$
$V(y)=0.25$

פרק 17 - תוחלת ושונות של סכום משתנים מקריים

רקע:

אם $X_1, X_2, \ldots\ldots\ldots, X_n$ משתנים מקרים אזי:

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

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

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

למשל,

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?

תרגילים:

1. הרווח ממניה א' הוא עם תוחלת של 5 ושונות 10. הרווח ממניה ב' הוא עם תוחלת של 4 ושונות 5. 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 הזכיה במשחק הראשון.

Y= סכום הזכייה במשחק השני.

נתון:

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

מהי התוחלת ומהי סטיית התקן של סכום הזכייה בשני המשחקים?

4. ברולטה הסיכוי לזכות ב- 30 ₪ הוא חצי וב-10 ₪ רבע כך גם ב- 20 ₪. מה היא התוחלת והשונות של סכום הזכייה הכולל לאדם המשחק ברולטה 4 פעמים.

5. נתון משתנה מקרי בעל פונקציית ההסתברות הבאה:

$$P(X=K)=\begin{cases} \dfrac{A}{K-1} & K=2,3,4,5 \\ 0 & \text{אחרת} \end{cases}$$

א. מצא את ערכו של A.

ב. חשב את התוחלת והשונות של X.

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

פתרונות:

שאלה 1

תוחלת: 9

שונות : 15

שאלה 3

תוחלת: 22

סטיית תקן: 5

שאלה 4

תוחלת : 90

שונות : 275

שאלה 5

א. $$A=\frac{12}{25}=0.48$$

ב. תוחלת 2.92

שונות 1.1136

ג. תוחלת 2.92n

שונות 1.1136n

פרק 18 - התפלגויות בדידות מיוחדות - התפלגות בינומית

רקע:

נגדיר את המושג ניסוי ברנולי:

ניסוי ברנולי הנו ניסוי שיש לו שתי תוצאות אפשריות: " הצלחה" ו" כישלון " כמו: 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: $X \sim B(n, p)$.

<u>פונקציית ההסתברות של X</u>:

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

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

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

<u>תוחלת</u>: $E(X) = np$

<u>שונות</u>: $V(X) = npq$

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 <u>n</u> פעמים.
3) X– מוגדר <u>כמספר ההצלחות המתקבלות</u> בסך הכול.

<u>דוגמה :</u> (פתרון בהקלטה)

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

א. מהי 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?

תרגילים:

1. במדינה 10% מהאוכלוסייה מובטלת. נבחרו 5 אנשים באקראי מאותה אוכלוסיה.
נגדיר את Xלהיות מספר המובטלים שהתקבלו במדגם.
א. מהי התפלגות 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 סמארט-פון. נבחרו 10 אנשים באקראי. נגדיר את X כמספר האנשים שנדגמו עם סמארט-פון.

א. מהי התפלגות 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 האוכלוסיה 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 מהמדינה הנ"ל.
א. מה ההסתברות ש-5 מהם אקדמאים?
ב. מה התוחלת של מסי' בעלי ההשכלה הנמוכה?

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

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

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

6. במבחן אמריקאי 20 שאלות. סטודנט ניגש למבחן והסיכוי שהוא יודע שאלה היא 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. מטילים מטבע הוגן 5 פעמים. נגדיר את X – מספר הפעמים שהתקבל עץ. חשבו את $E(x^2)$.

פתרונות :

שאלה 2:

ב. 0.2335

ג. 0.1493

ד. תוחלת : 7

סטיית תקן : 1.449

שאלה 3 :

א. 0.5314

ב. 0.0984

ג. 0.1143

ד. תוחלת : 18-

סטיית תקן : 14.697

שאלה 4:

א. 0.1789

ב. 2

שאלה 5 :

א. 0.1956

ב. 0.4253

שאלה 6 :

א. 0.85

ב. 0.182

ג. תוחלת : 82 נקודות

שונות : 91.8 נקודות

שאלה 7 :

א. 0.401

ב. תוחלת : 8.025

סטיית תקן : 2.193

שאלה 8 :

7.5

פרק 19 - 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:

$$P(X=k)=pq^{k-1} \qquad k=1,2,\ldots\infty$$

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, $P(X=n+k)/X>k)=P(X=n)$.

$$P(X>k)=q^k$$

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?

תרגילים:

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. הוא 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 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?

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 עוגת גבינה ועוגת שוקולד שנארזות באריזות אטומות. 40% מהעוגות הן עוגות גבינה והיתר עוגות 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 לייצרן 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?

<u>פתרונות :</u>

<u>שאלה 2 :</u>

א. 0.009

ב. 0.0001

ג. תוחלת : 1.111

שונות: 0.1234

ד. תוחלת : 44.4

שונות : 308.5

<u>שאלה 3 :</u>

א. 0.999

ב. 0.875

ג. 0.03125

ד. 1

<u>שאלה 4:</u>

א. 0.1029

ב. 9.72

<u>שאלה 5:</u>

א. 0.06

ב. 0.7176

ג. 0.0729

ד. 9.487משחקים

<u>שאלה 6 :</u>

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

פרק 20 - התפלגויות בדידות מיוחדות - התפלגות אחידה

רקע:

התפלגות זו הנה התפלגות שבה לכל תוצאה יש את אותה הסתברות.

הערכים המתקבלים בהתפלגות הם החל מ- a ועד b בקפיצות של אחד.

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

פונקציית ההסתברות:

$$P(X=K)=\frac{1}{b-a+1}$$

K= a, a+1,..,b

תוחלת: $E(X)=\frac{a+b}{2}$

שונות: $V(X)=\frac{(b-a+1)^2-1}{12}$

דוגמה: (פתרון בהקלטה)

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

תרגילים :

1. במשחק הלוטו 45 כדורים ממוספרים מ- 1 ועד 45 .
נתבונן במשתנה X המספר של הכדור הראשון שנשלף על ידי המכונה.
א. חשבו את $P(X=2)$
ב. חשבו את $P(X\le 30)$
ג. חשבו את $P(X>4|X\le 10)$
ד. חשבו את $P(X=k)$

2. קוסם מבקש לבחור מספר שלם אקראי בין 1 ל- 100. בהנחה שאין כאן מניפולציות של הקוסם.
א. מהי התוחלת ומהי סטיית התקן של המספר שיבחר?
ב. הקוסם ביקש משישה אנשים לבחור מספר :
1. מה ההסתברות ששלושה מהם יבחרו מספר הגדול מ 80?
2. מה התוחלת ומהי סטיית התקן של סכום המספרים שהאנשים בחרו?

3. יהי X התוצאה בהטלת קובייה.
א. מהי ההתפלגות של X?
ב. מה התוחלת של X?
ג. קובייה הוטלה 4 פעמים . מה התוחלת ומה השונות של סכום התוצאות ב- 4 ההטלות?

4. בכד 10 כדורים שרק אחד צבע אדום. אדם מוציא כדור ללא החזרה עד אשר מתקבל הכדור האדום. מה התוחלת ומהי השונות של מספר הכדורים שהוצאו?

5. יש לבחור מספר אקראי בי 1 ל- 50 כולל.
א. מה הסיכוי שהמספר 4 יבחר?
ב. מה הסיכוי שהמספר שיבחר גדול מ- 20 ?
ג. אם נבחר מספר גדול מ- 20 מה ההסתברות שהוא קטן מ- 28?

6. הוכח שאם $X\sim U\left(a,b\right)$ אז מתקיים ש : $E(X)=\dfrac{a+b}{2}$.

<u>פתרונות :</u>

<u>שאלה 1 :</u>

א. תשובה : $\frac{1}{45}$

ב.תשובה : $\frac{30}{45}$

ג. תשובה : 0.6

<u>שאלה 2</u>

א. תוחלת : 50.5

סטיית התקן : 28.87

ב. 1. תשובה : 0.08192

ב. 2 תוחלת : 303 סטיית תקן :70.71

<u>שאלה 4 :</u>

תוחלת 5.5

שונות :8.25

פרק 21 - ההתפלגויות 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 בממוצע אירועים קורים ביחידת זמן.

$$X \sim pois(\lambda)$$

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

פונקציית ההסתברות של ההתפלגות הפואסונית נתונה:

$$P(X=K)=\frac{e^{-\lambda}\cdot\lambda^{K}}{K!}$$

$$K=0,1,2,\ldots\infty$$

התוחלת והשונות של ההתפלגות:

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

תכונות מיוחדות של ההתפלגות:

- בהתפלגות הזו הפרמטר λ פורפורציונלי 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?

<u>תרגילים:</u>

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 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 (הנח שבחודש יש 4 שבועות) יהיה בדיוק שבוע אחד בו יהיו 3 תאונות דרכים קטלניות?

4. לחנות AMPM 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 מסוים מתפלג פואסונית עם תוחלת של 8 לידות ביום.
   א. מה ההסתברות שביום א' נולדו 10 תינוקות וביום ב' נולדו 7 תינוקות?
   ב. מיילדת עובדת במשמרות של 8 שעות. מה ההסתברות שבמשמרת שלה נולדו 3 תינוקות?
   ג. מהי התוחלת של מספר הימים בשבוע בהם נולדים ביום עשרה תינוקות?

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

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

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

פתרונות :

שאלה 1:

א. 0.0337

ב. 0.9933

ג. 0.1246

ד. 5

שאלה 2:

א. 0.084

ב. 0.099

ג. 0.151

שאלה 3:

א. 4

ב. 0.407

שאלה 5 :

א. 0.0139

ב. 0.2196

ג. 0.6948

שאלה 6 :

א. 16.7

ב. 0.0708

פרק 22 - 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~H(N,D,n).**

פונקציית ההסתברות של ההתפלגות:

$$P(X=k)=\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

התוחלת של ההתפלגות:

$$E(X)=n\cdot\frac{D}{N}$$

השונות של ההתפלגות:

$$V(X)=n\cdot\frac{D}{N}\cdot(1-\frac{D}{N})\cdot\frac{N-n}{N-1}$$

דוגמה : (הפתרון בהקלטה)

בכתה 40 תלמידים מתוכם 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?

תרגילים:

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 ולא בטבלה.
   ב. מה התוחלת וסטיית התקן של X?
   ג. חשבו את ההסתברות הבאה: $P(X=2\mid X>1)$

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

פתרונות:

שאלה 1

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

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

שאלה 2

ב. תוחלת: 1.5

סטיית תקן: 0.748

ג. 0.9

פרק 23 - 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 ית .

X- מספר החזרות עד שהתקבלו r הצלחות.

$X \sim NB(r, p)$

פונקציית ההסתברות :

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

התוחלת:

$$E(X)=\frac{r}{p}$$

השונות:

$$V(X)=\frac{r(1-p)}{p^2}$$

דוגמה :(פתרון בהקלטה)

מטילים קובייה עד אשר מקבלים 3 פעמים תוצאה שהיא גדולה מ-4.

א. מה הסיכוי להטיל את הקובייה 6 פעמים?
ב. מה תוחלת ושונות מספר הפעמים שנטיל את הקובייה?

תרגילים:

1. בכד 4 כדורים שחורים ו-6 כדורים לבנים. אדם מוציא כדור באקראי פעם אחר פעם ומחזיר בין הוצאה להוצאה את הכדור. נסמן ב-X את מספר הכדורים שהוא הוציא עד אשר הוא קיבל 2 כדורים לבנים בסך הכול אך 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. הראה שההתפלגות הגאומטרית היא מקרה פרטי של ההתפלגות הבינומית השלילית.

4. מטילים מטבע שוב ושוב עד אשר מקבלים שלוש פעמים עץ בסך בכול.

   א. בנו את פונקצית 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_i} מספר החזרות עד ההצלחה הראשונה בניסיונות ברנוליים בלתי תלויים זה בזה כאשר i=1,2,…n .

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

פתרונות:

שאלה 1

א. 0.36

ב. 0.288

שאלה 4

ב. תוחלת: 6 שונות: 6

ג. 0.1886

פרק 24 - קירוב פואסוני להתפלגות הבינומית

רקע:

אם $X \sim B(n,p)$ עבור n גדול ו-p קטן ניתן לקרב את ההתפלגות להיות פואסונית כאשר הפרמטר $\lambda = np$

כאשר פונקציית ההסתברות של ההתפלגות הפואסונית כזכור היא : $p(X=k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$.

הערה : יש הטוענים, כי n 'גדול' ו-p 'קטן' משמעו : $np \geq 10$ ו- $p \leq 0.1$.

דוגמה : (פתרון בהקלטה)

בקו ייצור המוני 10% מהמוצרים כחולים . בוחרים באקראי 20 מוצרים מקו הייצור.
חשבו את ההסתברות שמתוך המוצרים שיבחרו בדיוק 1 יהיה כחול.
פעם לפי ההתפלגות הבינומית ופעם לפי הקירוב הפואסוני.

תרגילים:

1. במדינת שומקום 10% מהאוכלוסייה מובטלת . נדגמו 10 תושבים אקראיים מאותה מדינה. חשבו את הסיכוי שבמדגם יהיה לכל היותר מובטל אחד.
השוו את התוצאה לקירוב הפואסוני.

2. מקו ייצור המוני נדגמו 1000 מוצרים. ידוע ש- 5% מהמוצרים בקו הייצור פגומים. מה הסיכוי שבמדגם יתקבלו 45 מוצרים פגומים?

3. 1% מהתושבים באוכלוסייה גדולה חולים במחלה מסוימת . בסניף קופת חולים נרשמו 2000 תושבים אקראיים. חשב לפי הקרוב הפואסוני שבדיוק 18 מהם יהיו חולים.

4. בעיר ניו יורק ישנם כתשעה מיליון תושבים מתוכם 900 אלף אסיאתיים . מה בקירוב ההסתברות שמתוך 100 תושבים אקראיים לפחות שני אסיאתיים?

פתרונות :

שאלה 1:

ללא קירוב 0.7361 עם קירוב : 0.7358

שאלה 2:

0.0458

שאלה 3:

0.0844

שאלה 4:

0.9995

פרק 25 - המשתנה המקרי הבדיד – שאלות מסכמות

תרגילים:

1. נתון ש:

$$X \sim B(4, \frac{1}{2})$$

$$Y \sim B(10, \frac{1}{4})$$

א. חשב את התוחלת וסטיית התקן של X.

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

ג. $T = X + Y$, חשב את התוחלת של T. האם ניתן לדעת מה סטיית התקן של T?

2. ערן משחק בקזינו בשתי מכונות הימורים. משחק אחד בכל מכונה (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. בנה את פונקצית ההסתברות של X.

ב. חשב 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 "בערוץ 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?

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 (וכולל) שלראשונה יימכר מוצר מתוצרת הארץ. מהי פונקצית 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?

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 $V(X)=1$. חשב $E(X-5)^2$.

10. הסיכוי לעבור מבחן נהיגה הינו P. בוחרים באקראי ארבעה נבחנים . ההסתברות ששניים מהם יעברו את מבחן הנהיגה גבוה פי 8/ 3 מהסיכוי שכל הארבעה יעברו את המבחן.
א. חשבו את ערכו של P.

תלמיד ניגש לבחינה עד אשר הוא עובר אותה.

ב. 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. לאילו ערכי P המשחק כדאי?

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

14. לפניכם N מגירות ממוספרות מ-1 ועד N . ברשותכם n חולצות . עליכם באופן אקראי לבחור לכל חולצה מגירה . כל מגירה יכולה להכיל גם את כל החולצות שברשותכם. נגדיר את X_1 - מספר החולצות שהונחו במגירה מספר 1. נגדיר את X_N - מספר החולצות שהונחו במגירה מספר N.
   חשבו את $V(X_1+X_N)$.

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 $E(\sqrt{x+7})$.

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 שנשתל.

א. מצאו את ההתפלגות של X.

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

פתרונות :

שאלה 1:

א. תוחלת : 2

סטיית תקן : 1

ב. תוחלת : 0

סטיית תקן : 2

ג. תוחלת : 4.5

סטיית תקן : לא ניתן

שאלה 2:

א. 0.03

ב. תוחלת : 0.15, שונות 0.1875

ג. 0.1328

שאלה 3:

א.

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

ב. תוחלת : 3

שונות : 2

ג. תוחלת : 1.5

שונות : 0.5

שאלה 4 :

א. 0.9975

ב. 0.0172

ג. 1.0025

<u>שאלה 5:</u>

א. 0.375

ג. 0.6

ד. 0.282

<u>שאלה 6:</u>

ב. תוחלת : 1.5

שונות 0.61

ג. תוחלת : 0

סטיית תקן : 4.68

<u>שאלה 7:</u>

א. 0.4348

ב. 0.0923

<u>שאלה 8 :</u>

א. 2.013

ב. 0.5256

ג. תוחלת : 8

שונות : 1.6

ד. תוחלת : 10.5

שונות 3.475

<u>שאלה 9 :</u>

10

<u>שאלה 10:</u>

א. 0.6

ב. 0.0384

ג. 0.0256

ד. תוחלת: 0.67

שונות: 1.11

ה. 0.24

שאלה 12:

ב. $\frac{1-2p^2}{p}$

ג. $0<p<\sqrt{\frac{1}{2}}$

שאלה 13:

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

שאלה 14:

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

שאלה 15:

$$\frac{2^n}{2n}$$

שאלה 16:

x	1	2	3
P(x)	0.0008	0.2893	0.7099

שאלה 17:

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

שאלה 18 :

2.675

שאלה 19 :

א. $P(X=k)=\frac{n-k-1}{\binom{n}{2}} \qquad k=0,1,\ldots,\text{n}-2$

ב. הוכחה

פרק 26 - 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.

פונקציית התפלגות מצטברת:

$$\boxed{F(t)=p(X\le t)=\int_{-\infty}^{t} f(x)dx}$$

כמו כן :

$$\boxed{p(X>t)=1-F(t)} \qquad \boxed{p(a<X<b)=F(b)-F(a)}$$

תוחלת של משתנה רציף :

$$\boxed{E(X)=\int_{-\infty}^{\infty} X\cdot f(x)dx=\mu}$$

שונות של משתנה רציף:

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

תוחלת של פונקציה של X:

$$E\left(g(x)\right)=\int_{-\infty}^{\infty} g(x)\,f\,(x)\,dx$$

אחוזונים :

האחוזון ה- P הוא ערך (נסמן אותו : x_p) שהסיכוי ליפול מתחתיו הוא P. כלומר :

$$\boxed{p(X\le x_p)=p}$$

ריענון מתמטי:

נוסחאות לחישוב שטחים:

שטח משולש: גובה (h) כפול הבסיס (a) חלקיי 2: $S_{triangle} = \dfrac{h \cdot a}{2}$

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

משוואת קו ישר:

y=mx+n

m = שיפוע.

n = נקודת החיתוך עם ציר הy.

שיפוע של ישר העובר דרך שתי נקודות : $(X_1, Y_1), (X_2, Y_2)$: $m = \dfrac{Y_2 - Y_1}{X_2 - X_1}$

משוואת ישר שעובר דרך נקודה ספציפית (X_1, Y_1) ושיפועו ידוע m:

$$y - Y_1 = m(x - X_1)$$

נוסחאות - אינטגרלים

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

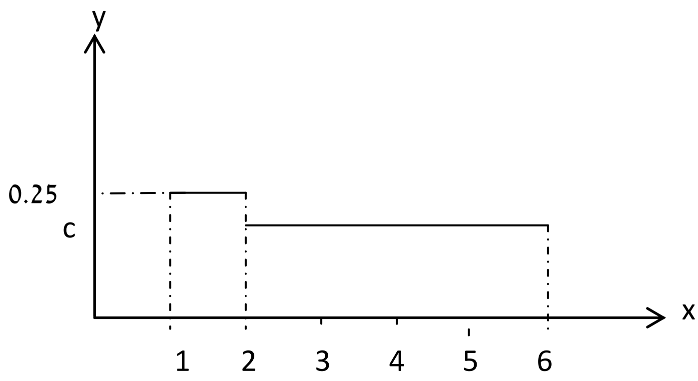
$$\int e^f \cdot f' dx = e^f + c \qquad \int \cos f \cdot f' dx = \sin(f) + c$$

$$\int \sin f \cdot f' dx = -\cos(f) + c \qquad \int \frac{f'}{\sqrt{f}} dx = 2\sqrt{f} + c$$

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

תרגילים:

1. X הינו משתנה רציף עם פונקצית צפיפות כמוצג בשרטוטו:



א. מצא את ערכו של c.

ב. בנה את פונקצית ההתפלגות המצטברת.

ג. חשבון את ההסתברויות הבאות:

1. $P(x<4)$
2. $P(x>1.5)$
3. $P(1.5<x<5)$
4. $P(5<x<10)$

ד. מצא את החציון של המשתנה.

2. נתון משתנה מקרי רציף X שפונקצית הצפיפות שלו היא:

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

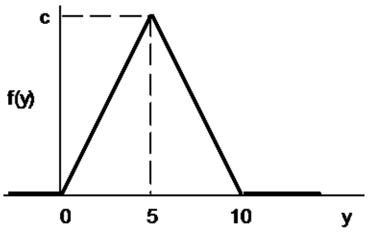
ידוע ש- $P(0 < X < 1) = 1/4$.

א. מצאו במפורש את פונקצית הצפיפות של X.

ב. מצאו את החציון של X.

ג. מה הסיכוי ש- X קטן מ- 0.5 ?

3. נתונה פונקצית צפיפות של משתנה מקרי Y :



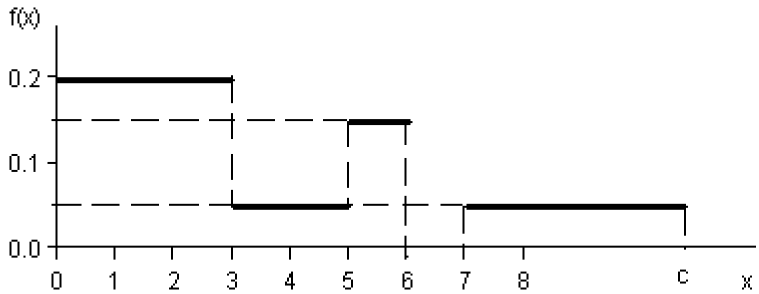
א. מצאו את c.

ב. מצאו את פונקצית ההתפלגות המצטברת של Y .

ג. חשבו את ההסתברויות : $P(Y=7.0)$, $P(Y\le 3.0)$, $P(7.5\le Y\le 15.5)$ $P(Y>4)$.

ד. מצאו את העשירון התחתון $y_{0.1}$, הרבעון התחתון $y_{0.25}$ והחציון של Y. הסיקו מהו העשירון עליון $y_{0.9}$.

4. נתונה פונקצית צפיפות של משתנה מקרי X :

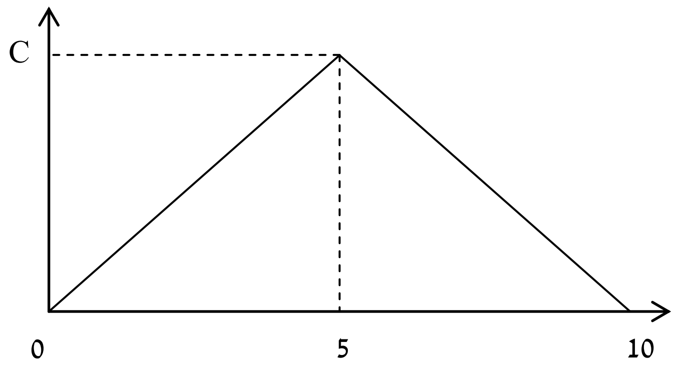


א. מצאו ערך c שעבורו תתקבל פונקצית צפיפות.

ב. מצאו את פונקצית ההתפלגות המצטברת .

ג. חשבו את ההסתברויות הבאות : $P(X\ge 4)$, $P(X\ge -2.0)$, $P(1.0<X\le 5.0)$

5. נתונה פונקצית הצפיפות הבאה :



א. מה ערכו של C?

ב. מצא אינטרוול (תחום) סימטרי סביב הערך 5 שהסיכוי ליפול בו הינו 0.5

6. נתונה פונקצית צפיפות $f(X)=\frac{2}{x}$ פונקציה זו מוגדרת מ-1 ועד K.

א. מצא את ערכו של K.

ב. בנה את פונקציית ההתפלגות המצטברת.

ג. חשב את הסיכוי ש X לפחות 1.5.

ד. מצא את העשירון התחתון של ההתפלגות.

ה. מה התוחלת של X?

7. נתונה פונקצית צפיפות הבאה : $f(X)=AX^2(10-X)$ 0<X <10 A הינו קבוע חיובי.

א. מצא את A.

ב. חשב את $P(x>5\,|\,x>2)$.

ג. מה התוחלת ומהי השונות של X ?

8. פונקצית הצפיפות של משתנה מקרי רציף X :

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

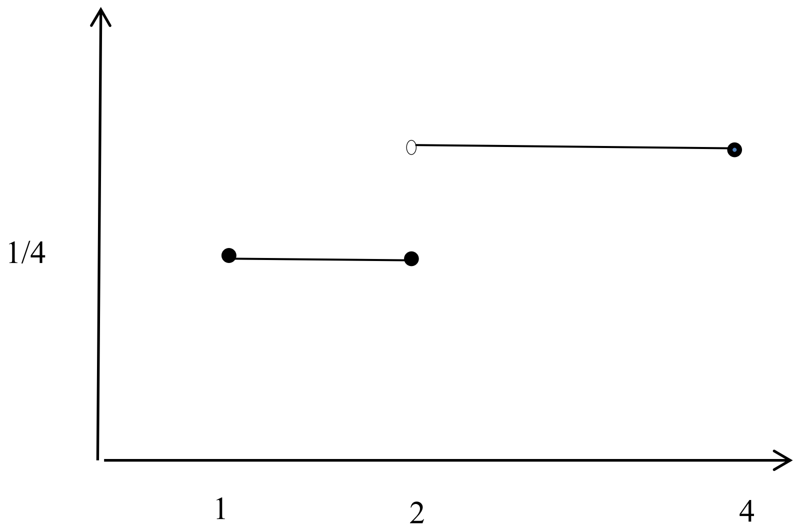
א. מצא את ערכו של c.

ב. מצא את פונקציית ההתפלגות המצטברת של ההתפלגות.

ג. חשב $P(X > 0)$.

ד. מהו הרבעון העליון של ההתפלגות?

9. נתונה פונקצית הצפיפות הבאה של משתנה מקרי X :



א. רשום את נוסחת פונקציית הצפיפות .

ב. בנה את פונקציית ההתפלגות המצטברת.

ג. מצא את החציון של ההתפלגות.

ד. חשב את התוחלת והשונות של המשתנה.

ה. חשב את $E(X^3)$

10. במפעל מייצרים מוצר A. זמן תהליך הייצור של המוצר בשעות הוא בעל פונקציית הצפיפות הבאה:

$$f(x)=6x(1-x) \qquad 0\le x\le 1$$

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. שרטט 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 ש90% מהלקוחות מחכים מתחתיו?

12. פונקצית 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. מצא את פונקציית ההתפלגות המצטברת.

ב. חשב את התוחלת והשונות של ההתפלגות.

ג. מצא את התוחלת של $\frac{1}{X}$.

פתרונות :

שאלה 1 :

א. $\frac{3}{16}$ ד. $3\frac{1}{3}$

שאלה 2 :

א. $b=2$ $c=0.5$

ב. 1.41

ג. 0.0625

שאלה 3 :

א. 0.2

ג. 0, 0.18, 0.125, 0.32

ד. העשירון התחתון : 2.24

הרבעון התחתון: 3.54

החציון : 5

העשירון העליון : 7.76

שאלה 4 :

א. 10

שאלה 5:

א. C=0.2

ב. 5 ± 1.46

שאלה 6 :

א. $e^{\frac{1}{2}}$

ג. 0.189

ד. 1.051

ה. 1.297

שאלה 7 :

א. 0.0012

ב. 0.7067

ג. תוחלת : 6, שונות : 4

שאלה 8 :

א. 2

ג. 0.75

ד. 0.549

שאלה 9 :

ג. $2\frac{2}{3}$

ד. תוחלת : 2.625 שונות : 0.6927

ה. 23.4375

שאלה 10 :

א. $\frac{7}{27}$

ב. 0

ג. 3.704

שאלה 11 :

ב. 0.0498

ג. 0.6321

ד. 11.51

שאלה 12 :

א. $\frac{2}{3}$

ב. 5.22

ג. $\frac{2}{9}$

שאלה 13:

א. $\frac{1}{6}$

ג. 0.229

שאלה 14 :

ב. תוחלת :

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

השונות :

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

פרק 27 - 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 היא :

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

פונקציית ההתפלגות המצטברת היא :

$F(t) = p(x \leq t) = 1 - e^{-\lambda t}$

התוחלת:

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

השונות:

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

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

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?

תרגילים:

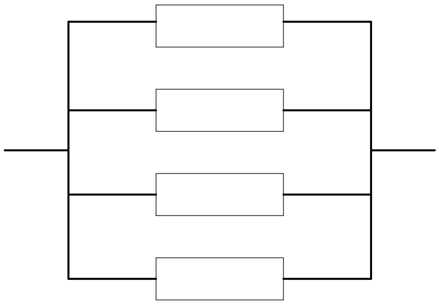
1. הזמן שלוקח במערכת עד שתקלה מתרחשת מתפלג מעריכית עם תוחלת של 0.5 שעה.
   א. מה הסתברות שהתקלה הבאה תתרחש תוך יותר מ-0.5 שעה?
   ב. מה ההסתברות שהתקלה הבאה תתרחש תוך פחות משעה?
   ג. מצא את הזמן החציוני להתרחשות תקלה במערכת.

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 השני לשלישי יש להמתין פחות מרבע שעה?

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 100שעות.

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?

פתרונות:

שאלה 1 :

א. 0.368

ב. 0.865

ג. 0.347

שאלה 2:

א. 24 שעות

ב. 0.632

ג. 0.135

שאלה 3:

א. 0.393

ב. 0.239

ג. 0.513

ד.69.08

שאלה 4:

א. 0.264

ב. 0.368

ג. 0.233

שאלה 5:

א. 0.8403

ב. A0.0588>K

פרק 28 - התפלגויות רציפות מיוחדות - התפלגות אחידה

רקע:

זו התפלגות שפונקציית הצפיפות שלה קבועה בין a לבין b.

X~U (a,b)

פונקציית הציפות :

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

פונקציית ההתפלגות המצטברת:

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

התוחלת :

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

השונות:

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

דוגמה : (הפתרון בהקלטה)

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

א. מה הסיכוי ש- X קטן מ-25?

ב. מה התוחלת והשונות של X?

תרגילים:

1. משך (בדקות) הפסקה בשיעור, X, מתפלג $U(13, 16)$.
   א. מהי התוחלת ומהי סטית התקן של משך ההפסקה?
   ב. מהי ההסתברות שהפסקה תמשך יותר מ- 15 דקות?
   ג. מהי ההסתברות שמשך ההפסקה יסטה מהתוחלת בפחות מדקה?

2. רכבת מגיעה לתחנה בשעות היום כל עשר דקות . אדם הגיע לתחנה בזמן אקראי.
   א. הסבר כיצד מתפלג זמן ההמתנה לרכבת?
   ב. אם זמן ההמתנה לרכבת ארך יותר מ-5 דקות , מהי ההסתברות שבסך הכל האדם ימתין לרכבת פחות מ- 8 דקות?
   ג. מה תוחלת מספר הימים שיעברו עד הפעם הראשונה שהאדם ימתין לרכבת יותר מ-9 דקות?

3. מכונה אוטומטית ממלאת גביעי גלידה. משקל הגלידה לגביע מתפלג אחיד בין 100-110 גרם (המשקל הוא של גלידה ללא הגביע).
   א. מה ההסתברות שמשקל הגלידה בגביע יהיה מעל 108 גרם?
   ב. נתון שהגלידה בגביע עם משקל נמוך מ-107 גרם. מה ההסתברות שמשקל הגלידה יהיה מעל 105 גרם?
   ג. מה העשירון העליון של משקל הגלידה בגביע?

פתרונות:

שאלה 1:

א. תוחלת: 14.5

שונות: 0.866

ב. $1/3$

ג. $2/3$

שאלה 2:

א. $X \sim U(0,10)$

ב. 0.6

ג. 10

שאלה 3:

א. 0.2

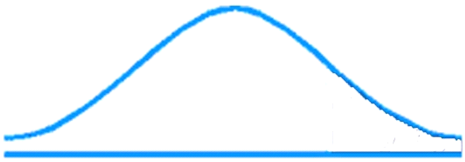
ב. $\frac{2}{7}$

ג. 109

פרק 29 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

רקע:

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



לעקומה זו קוראים גם עקומת גאוס ועקומה אחת נבדלת מהשנייה באמצעות הממוצע וסטיית התקן שלה. אלה הם הפרמטרים שמאפיינים את ההתפלגות.

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

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

כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלבנטים שמתחת לעקומה.
כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון.
התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת והיא תסומן באות Z.

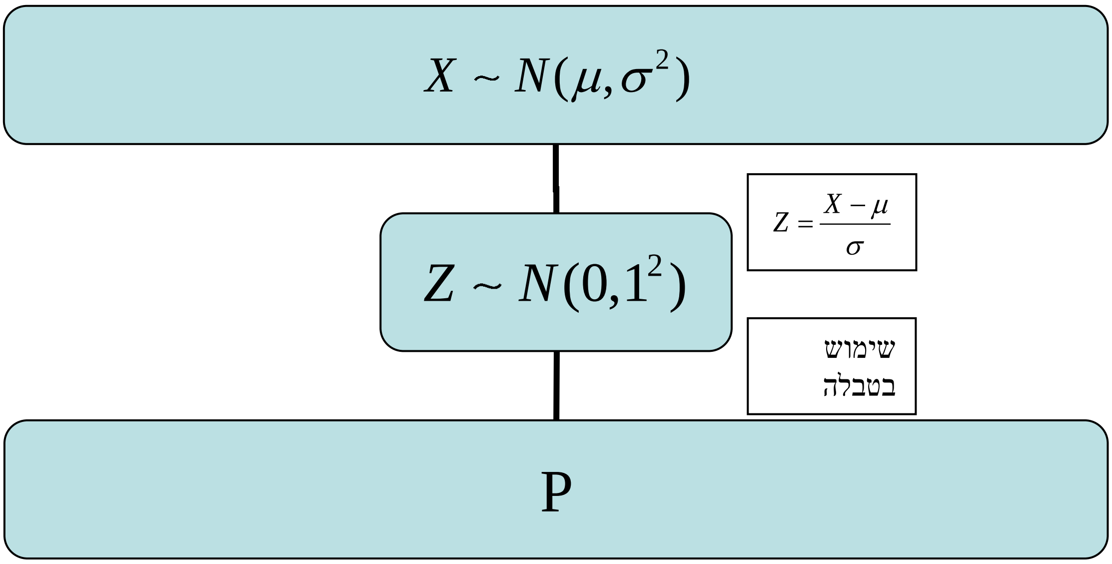
$Z \sim N(0,1^2)$

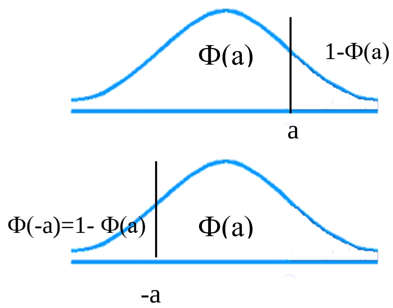
תהליך התקנון מבוצע על ידי הנוסחה הבאה :

$$Z=\frac{X-\mu}{\sigma}$$

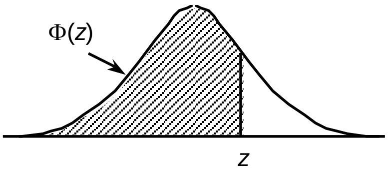
אחרי תקנון מקבלים ערך הנקרא ציון תקן.
ציון התקן משמעו בכמה סטיות תקן הערך סוטה מהממוצע.
לאחר חישוב ציון התקן של ערך מסוים נעזרים בטבלה של ההתפלגות הנורמלית הסטנדרטית לחישוב השטח הרצוי.

ובאופן כללי נתאר את הסכמה הבאה :





טבלת ההתפלגות המצטברת הנורמלית סטנדרטית – ערכי Φ(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

<u>דוגמה</u> : (הפתרון בהקלטה)

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

א. מה אחוז חפיסות השוקולד ששוקלות מתחת ל- 110 גרם?

ב. מה אחוז חפיסות השוקולד השוקלות מעל 110 גרם?

ג. מה אחוז חפיסות השוקולד השוקלות מתחת ל 92 גרם?

ד. מהו המשקל ש90% מהחפיסות בקו הייצור שוקלים פחות מהם?

תרגילים:

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?

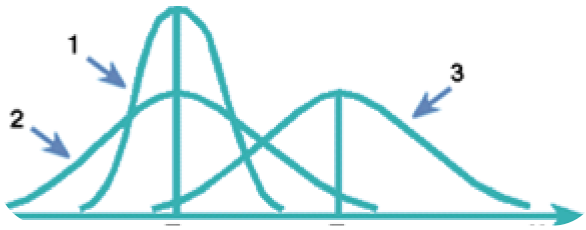
4. משקל תינוקות ביום היוולדם מתפלג נורמלית עם ממוצע של 3300 גרם וסטיית תקן 400 גרם.
   א. מצאו את העשירון העליון.
   ב. מצאו את האחוזון ה95.
   ג. מצאו את העשירון התחתון.

5. ציוני מבחן אינטיליגנציה מתפלג 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, נתון ש33% מהבקבוקים הם עם 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 המאון 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 בינהן.



א.לאיזו התפלגות הממוצע הגבוה ביותר?

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

א. בעשירון העליון.

ב. בממוצע.

ג. בשונות.

ג. לאיזו התפלגות סטיית התקן הקטנה ביותר?

א. 1

ב. 2

ג. 3

ד. אין לדעת.

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

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

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

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

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

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

א. מה ערכו של T?

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

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

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

פתרונות :

שאלה 1

א. 89.25%

ב. 2.28%

ג. 0

ד. 50%

שאלה 3

א. 26.43%

ב. 89.44%

ג. 39.44%

ד. 0.383

ה. 100%

שאלה 5

א. 119.2

ב. 80.8

ג. 112.6

ד. 87.4

שאלה 7

א. 500

ב. 100

ג. 0.3446

ד. 733

ה. 267

שאלה 8

א. 3

ב. בממוצע.

ג. 1

שאלה 9

א. 0.1587

ב. 0.0228

ג. 0.8563

ד. 0.3975

שאלה 10

א. 1925

ב. 0.2266

ג. 0.1587

שאלה 11

א. 0.1359

ב. 0.675

ג. 100

ד. 0.25

פרק 30- משתנה דו מימדי בדיד - פונקציית הסתברות משותפת

רקע:

התפלגות דו ממדית הינה התפלגות שדנה בשני משתנים.

נרצה כעת לבנות פונקציית הסתברות דו ממדית.

בפונקציה שכזו יש התפלגות של שני משתנים בו זמנית : X ו Y .

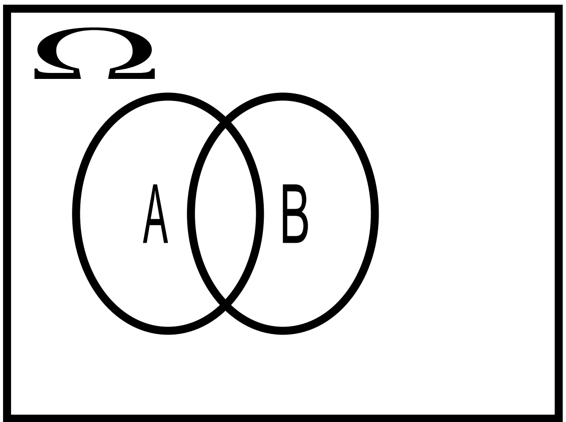
דוגמה:

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

יהי X- מספר הקורסים שהסטודנט עבר.

יהי Y- משתנה אינדיקטור המקבל את הערך אחד אם הסטודנט עבר את הבחינה בכלכלה ואפס אחרת.

בנו את פונקציית ההסתברות המשותפת של X ו Y .



נחשב את כל ההסתברויות המשותפות :

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

$Y \backslash X$	0	1	2
0	0.05	0.15	0
1	0	0.05	0.75

שימו לב שסכום כל ההסתברויות בפונקציית ההסתברות המשותפת הוא 1.

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

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

משתנים בלתי תלויים:

X ו Y יהיו משתנים בלתי תלויים אם עבור כל X ו- Y אפשריים התקיים הדבר הבא :

$$\boxed{p(x=k, y=l) = p(x=k) \cdot p(y=l)}$$

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

למשל , בדוגמה הזאת:

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

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

תרגילים :

1. אדם נכנס לקזינו עם 75 דולר . הוא ישחק במכונת מזל בה יש סיכוי של 03 לנצח. במקרה של ניצחון במשחק הוא יקבל מהקזינו 25 דולר ובמקרה של הפסד הוא ישלם 25 דולר . אותו אדם החליט שיפסיק לשחק ברגע שיהיה לו 100 דולר , אך בכל מקרה לא ישחק יותר מ – 3 משחקים. נגדיר את X להיות הכסף שברשות האדם בצאתו מהקזינו ואת Y מספר המשחקים שהאדם שיחק.

א. בנו את פונקצית ההסתברות המשותפת והשוליות.

ב. מה תוחלת מספר המשחקים שישחק האדם?

ג. אם האדם יצא מהקזינו שברשותו 100 דולר , מה התוחלת ומהי השונות של מספר המשחקים ששיחק?

2. להלן פונקצית ההסתברות המשותפת והשוליות של שני משתנים מקריים בדידים :

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

א. השלם את ההסתברויות החסרות בטבלה.

ב. האם X ו- Y תלויים ?

ג. מצא את הסתברות ש-Y=3 , אם ידוע ש- X=1 .

3. מפעל משווק מוצר הנארז בחבילות בגדלים שונים. ישנו מספר שווה של חבילות בנות שני מוצרים ושלושה מוצרים. ההסתברות שמוצר מסוים יהיה פגום היא 1/10. מהנדס הייצור בוחר באקראי חבילת מוצרים לשם בקורת איכות. יהיו : $X-$ מספר המוצרים בחבילה, $Y-$ מספר המוצרים הפגומים בחבילה.

א. מה ההתפלגות של המשתנה Y בהינתן X הינו 3.

ב. מה ההתפלגות של המשתנה Y בהינתן X הינו K כלשהו.

ג. מהי תוחלת מספר המוצרים הפגומים בחבילות בנות 3 מוצרים? נמקו.

ד. בנה את פונקצית ההסתברות המשותפת.

4. מתוך כד עם שלושה כדורים ממוספרים במספרים 2, 4, 8 שולפים באקראי שני כדורים **ללא החזרה**. נגדיר: X - המספר הקטן מבין השניים; Y– המספר הגדול מבין השניים.

א. חשבו את ההתפלגות של (X, Y).

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

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

5. ביישוב שני סניפי בנק. סניף פועלים וסניף לאומי. להלן הנתונים לגבי האוכלוסייה הבוגרת המתגוררת ביישוב:

ל-60% יש חשבון בסניף פועלים של היישוב.

ל-50% יש חשבון בסניף לאומי של היישוב.

ל-95% יש חשבון בלפחות אחד מהסניפים.

יהי X- מספר הסניפים בישוב אשר לתושב בוגר יש בהם חשבון.

יהי Y- משתנה אינדיקטור:

1- אם יש לתושב חשבון בסניף פועלים.

0- אחרת.

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

ב.הוסיפו את פונקציית ההסתברות השולית.

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

פתרונות:

שאלה 1:

ב. 2.4

ג. התוחלת 1.348 השונות 0.575

שאלה 2:

ב. תלויים

ג. 0.125

שאלה 4:

ב. 0.5

ג. תוחלת 2

שאלה 5:

ג. 0.75

פרק 31 - משתנה דו מימדי בדיד - מתאם בין משתנים

רקע:

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

-1 0 1

מקדם מתאם 1- או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי הנוסחה : $y = ax + b$.

מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר לנארי מלא בו השיפוע a יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לנארי מלא בו השיפוע a שלילי (מקדם מתאם 1-).

מתאם חיובי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

חישוב מקדם המתאם :

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

השונות המשותפת:

$$\text{cov}(x,y) = E[(x-\mu_x)(y-\mu_y)] = E(x\cdot y) - E(x)\cdot E(y)$$

תכונות של השונות המשותפת :

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

משתנים בלתי מתואמים :

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

השפעת טרנספורמצייה לינארית על מקדם המתאם

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

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

דוגמה : (פתרון בהקלטה)
נחזור לדוגמה שהוצגה בפרק הקודם :
תלמיד ניגש בסמסטר לשני מבחנים מבחן בכלכלה ומבחן בסטטיסטיקה.
כמו כן נתון שהסיכוי לעבור את המבחן בכלכלה הנו 0.8 והסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9.
הסיכוי לעבור את שני המבחנים הנו 0.75
יהי X- מספר הקורסים שהסטודנט עבר.
יהי Y- משתנה אינדיקטור המקבל את הערך אחד אם הסטודנט עבר את הבחינה בכלכלה ואפס אחרת.
נחשב את מקדם המתאם :

$Y \backslash X$	0	1	2	P_Y
0	0.05	0.15	0	0.2
1	0	0.05	0.75	0.8
P_X	0.05	0.2	0.75	1

2	1	0	x
0.75	0.20	0.05	P(x)

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

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

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

y	P_Y
0	0.2
1	0.8

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

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

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

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

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

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

כל קורס שהסטודנט מסיים מזכה אותו ב- 3 נקודות אקדמאיות.

מה יהיה מקדם המתאם בין נקודות הזכות שיצבור למשתנה Y? ?

תרגילים:

1. הסיכוי שסטודנט יעבור את מועד א בסטטיסטיקה הוא 0.8. אם הוא נכשל במועד א' הוא ניגש למועד ב' שם הסיכוי לעבור את המבחן מוערך להיות 0.9 (סטודנט שעובר את א' לא ניגש לב'). במידה והסטודנט נכשל במועד ב' הוא מגיש בקשה למועד ג' אותה מאשרים בסיכוי של 0.2. ואז הסיכוי שלו לעבור את מועד ג' הוא 0.7.

נגדיר את X להיות מספר המבחנים אליהם ניגש הסטודנט.

נגדיר את Y להיות מספר הנבחנים שנכשל בהם.

א. בנו את פונקצית ההסתברות המשותפת ואת פונ' ההסתברות השולית.

ב. האם המשתנים הינם בלתי תלויים?

ג. ידוע שהסטודנט ניגש ליותר ממבחן אחד, מה ההסתברות שהוא נכשל בפחות משלושה מבחנים?

ד. האם המתאם בין X ל-Y מלא או חלקי? חיובי או שלילי? הסבר ללא חישוב.

ה. חשבו את מקדם המתאם בין X לבין Y.

ו. האם המשתנים הם בלתי מתאומים?

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

א. בנו את פונקציית ההסתברות המשותפת של X ו Y ואת פונקציות ההסתברות השוליות.

ב. האם X ו- Y הם משתנים בלתי תלויים?

ג. מהו מקדם המתאם בין X ל-Y . האם המשתנים מתואמים?

ד. אם בשתי ההטלות הראשונות יצא בדיוק עץ אחד, מה ההסתברות שבשתי ההטלות האחרונות יצאו שני עצים ?

ה. אם בשתי ההטלות האחרונות יצא לפחות פעם אחת עץ, מה ההסתברות שבשתי ההטלות הראשונות יצא עץ אחד?

3. מפזרים שלושה כדורים שונים בשלושה תאים.

נגדיר את המשתנים הבאים:

X= מספר הכדורים בתא הראשון.

Y= מספר הכדורים בתא השני.

א. בנו את פונקציית ההסתברות המשותפת.

ב. האם המשתנים בלתי מתאאמים?

4. מטילים קובייה הוגנת פעמיים.

יהי : X= ההטלה הגדולה מבין שתי התוצאות

Y= מסי ההטלות בהן יצאה תוצאה זוגית.

א. מצא את פונקצית ההסתברות המשותפת של X ו-Y.

ב. חשבו את מקדם המתאם של X ו-Y.

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

5. בבניין בן 5 דירות. דירות מספר אחת ושלוש הן דירות משופצות והשאר אינן. הוחלט לבחור שתי דירות באקראי מבין הדירות בבניין. נגדיר את המשתנים הבאים :

X- מספר הדירות המשופצות שנבחרו.

Y- מספר הדירות האי זוגיות שנדגמו.

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

ב. האם המשתנים מתואמים?

ג. מה מקדם המתאם בין X לבין Y?

ד. מה יהיה מקדם המתאם :

1. בין מספר הדירות המשופצות למספר הדירות הזוגיות שנדגמו.
2. בין מספר הדירות הזוגיות לדירות האי זוגיות שנדגמו.

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

פתרונות :

שאלה 1 :

ג . 0.994

ה . 0.963

שאלה 2 :

ב. תלויים.

ג. מקדם המתאם : 0.5. מתואמים

ד. 0.25

ה. 0.5

שאלה 3:

ב. מתואמים

שאלה 4 :

ב. 0.252

שאלה 5:

ב. X ו-Y מתואמים.

ג. $\frac{2}{3}$

ד.1. $-\frac{2}{3}$

ד.2. (-1)

ה. $-\frac{2}{3}$

פרק 32 - המשתנה המקרי הדו ממדי - קומבינציות לנאריות

רקע:

תוחלת ושונות של סכום משתנים :

$$E(X+Y)=E(X)+E(Y)$$

$$V(X+Y)=V(X)+V(Y)+2\cdot COV\left(X,Y\right)$$

תוחלת ושונות של הפרש משתנים :

$$E(X-Y)=E(X)-E(Y)$$

$$V(X-Y)=V(X)+V(Y)-2\cdot COV\left(X,Y\right)$$

קומבינציות ליניאריות:

יוצרים משתנה חדש שהוא קומבינציה ליניארית של שני משתנים אחרים :

$$W=(aX+b)+(cY+d)$$

$$COV\left[\left(aX+b\right),\left(cY+d\right)\right]=a\cdot c\cdot COV\left(X,Y\right)$$

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

$$V(W)=V\left((aX+b)+(cY+d)\right)=a^2V\left(X\right)+c^2V\left(Y\right)+2\cdot a\cdot c\cdot COV\left(X,Y\right)$$

דוגמה : (פתרון בהקלטה)

עבור שני משתנים מקריים נתון :

$$\mu_X=80$$
$$\sigma_X=15$$
$$\mu_Y=70$$
$$\sigma_Y=20$$
$$COV(X,Y)=200$$

- מצא את התוחלת והשונות של סכום המשתנים.
- מצא את התוחלת והשונות של Y-X.
- מצא את השונות ומה התוחלת של המשתנה $W=2X+3Y$

תרגילים:

1. נתונה פונקצית ההסתברות המשותפת הבאה:

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

א. השלם את ההסתברויות החסרות.

ב. האם המשתנים תלויים?

ג. האם המשתנים בלתי מתואמים?

ד. חשב את השונות המשותפת.

ה. חשב את התוחלת והשונות של סכום המשתנים.

ו.חשב את התוחלת והשונות של הפרש המשתנים.

2. מבחן בנוי מחלק כמותי וחלק מילולי. תוחלת הציון בחלק הכמותי היא 100 עם סטיית תקן 20. תוחלת הציונים בחלק המילולי 90 עם סטיית תקן 15. מקדם המתאם בין הציון הכמותי לציון המילולי הוא 0.8.

א. חשבו את השונות המשותפת בין הציון הכמותי לציון המילולי.

ב. חשבו את התוחלת והשונות של סכום הציונים בחלק הכמותי ובחלק המילולי.

ג. חשבו את התוחלת והשונות של הפרש הציונים בין החלק הכמותי לחלק המילולי.

ד. עלות הבחינה 2000 שקלים. הוחלט לזכות שקל עבור כל נקודה שנצברה בחלק המילולי ושני שקלים עבור כל נקודה שנצברה בחלק הכמותי. מהי התוחלת ומהי השונות של עלות הבחינה נטו (העלות לאחר הזיכוי)?

3. נתון: $Var(X-2Y)=2$. $Var(X+2Y)=3$. חשבו: $Cov(X,Y)$.

4. מטילים קובייה n פעמים.

נגדיר את המשתנים הבאים:

X=מספר הפעמים שהתקבלה התוצאה 6.

Y=מספר הפעמים שהתקבלה התוצאה 5

בטאו את השונות המשותפת באמצעות n.

פתרונות :

שאלה 1:

ב. תלויים

ג. מתואמים.

ד. ‎-0.1

ה. תוחלת : 4.4, שונות : 0.84

ו. תוחלת : ‎-0.4, שונות : 1.24

שאלה 2:

א. 240

ב. תוחלת: 190 שונות: 1105

ג. תוחלת: 10 שונות: 145

ד. תוחלת: 1710 שונות: 2785

שאלה 3:

‎-0.125

שאלה 4

$\frac{-n}{36}$

פרק 33 - משתנה דו ממדי בדיד – שאלות מסכמות

רקע:

משתנים בלתי תלויים:

X ו Y יהיו משתנים בלתי תלויים אם עבור כל X ו- Y אפשריים מתקיים :

$$\boxed{p(x=k,y=l)=p(x=k)\cdot p(y=l)}$$

מקדם המתאם :

$$\rho=\frac{\mathrm{cov}(x,y)}{\sigma_x\cdot\sigma_y}$$

השונות המשותפת:

$$\mathrm{cov}(x,y)=E[(x-\mu_x)(y-\mu_y)]=E(x\cdot y)-E(x)\cdot E(y)$$

תכונות של השונות המשותפת :

3. $\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$
4. $\mathrm{cov}(X,X)=Var(X)$
5. $COV\left[(aX+b),(cY+d)\right]=a\cdot c\cdot COV(X,Y)$

משתנים בלתי מתואמים :

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

השפעת טרנספורמציה ליניארית על מקדם המתאם

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

תוחלת ושונות של סכום משתנים :

$$E(X+Y)=E(X)+E(Y)$$

$$V(X+Y)=V(X)+V(Y)+2\cdot COV(X,Y)$$

קומבינציות ליניאריות: $W=(aX+b)+(cY+d)$

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

$$V(W)=V\left((aX+b)+(cY+d)\right)=a^2V(X)+c^2V(Y)+2\cdot a\cdot c\cdot COV(X,Y)$$

תרגילים :

1. יש ליצור סיסמא בת 3 תווים . כל תו יכול להיבחר רק מתוך כלל התווים הבאים : 1 2 C B A.
יהי X מספר הפעמים שהספרה 1 מופיעה בסיסמא. יהי Y מספר הפעמים שהספרה 1 מופיעה בקצות הסיסמא (ראשון או האחרון).
   א. זהו את ההתפלגויות השוליות של X ו-Y כהתפלגויות מיוחדות.
   ב. מצאו את ההתפלגות המשותפת של X ושל Y.
   ג. מצאו את מקדם המתאם בין X ל-Y.
   ד. מהו המתאם בין $2X$ ל- $3Y+5$?

2. במסיבת סוף שנה ישנו ארגז קרח ובתוכו 7 בקבוקי בירה : 4 "מכבי" , 2 "גולדסטאר" ו- 1 "טובורג" . קרן לקחה 3 בקבוקי בירה באקראי מתוך ארגז הקרח.
נסמן ב- X את מספר בקבוקי "מכבי " שנלקחו על ידי קרן.
נסמן ב-Y את מספר בקבוקי "טובורג" שנלקחו על ידי קרן.
   א. בנו את פונקציית ההסתברות המשותפת של X ושל Y.
   ב. חשבו את התוחלת והשונות של X ושל Y .
   ג. מצאו את השונות המשותפת של X ושל Y.
   ד. נגדיר את W – מספר בקבוקי ה"גולדסטאר" שנלקחו על ידי קרן. בטאו את W באמצעות X ו- Y וחשבו את התוחלת והשונות של W על סמך התוצאות שהתקבלו בשני הסעיפים הקודמים בלבד.
   ה. מהו מקדם המתאם בין מספר בקבוקי "מכבי" שנלקחו על ידי קרן למספר בקבוקים שאינם "מכבי" שנלקחו על ידי קרן?

3. במגירה 6 זוגות נעלים . יהודה הוציא מהמגירה 4 נעלים (לא בהכרח זוגות) באקראי. נסמן את W– מספר זוגות הנעליים שהוציא יהודה. נסמן את R - מספר הנעליים השמאליות שהוציא יהודה.
   א. מצא את ההתפלגות המשותפת של המשתנים שהוצגו .
   ב. האם המשתנים שהוצגו בלתי תלויים?
   ג. מצא את התפלגות מספר הנעליים השמאליות שהוצאו אם בסך הכול הוצא זוג נעלים יחיד על ידי יהודה.
   ד. אם ידוע שהוצאו לפחות 3 נעליים שמאליות מה הסיכוי שהוצא לכל היותר זוג אחד?

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

X- משתנה המקבל את הערך 1 אם נבחר לפחות כדור אחד כחול ו-0 אחרת.

Y – מספר הכדורים הלבנים שנבחרו.

א. חשב את $P(X=1)$.

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

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

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

5. ביום ההולדת 4 של טל הוא מחלק שלושה פרסים שונים באקראי ל- 5 ילדים. בכל פעם שטל מחלק פרס הוא בוחר באקראי ילד מתוך ה- 5 באופן אקראי ובלתי תלוי לבחירות הקודמות. נגדיר את המשתנים הבאים :

X- מספר הפרסים שקיבלה יוליה.

Y- מספר הילדים שלא קיבלו פרס.

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

ב. האם X ו- Y הם משתנים בלתי מתואמים?

ג. מצאו את התוחלת של $X\cdot Y^2$.

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

6. קבעו אילו מהטענות הבאות נכונות . נמקו.

א. אם שני משתנים הם מתואמים אזי הם תלויים.

ב. אם שני משתנים הם תלויים אזי הם מתואמים.

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

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

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

נסמן X_i - מספר הגברים שבחרו במתנה i.

נסמן Y_i - מספר הנשים שבחרו במתנה i.

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

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

ג. מהי ההתפלגות של X_1+X_2 ?

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

8. הוכח את הזהות הבאה עבור שלושת המשתנים :

$COV(X+Y,Z)=COV(X,Z)+\mathrm{COV}(\mathrm{Y},\mathrm{Z})$.

9. מספר העלים שנושרים בסתיו מהעץ בגינה מתפלג פואסונית עם תוחלת של 50 עלים בדקה.

נסמן Y - מספר העלים שנושרים בסתיו בין 12:00 ל- 12:10 .

נסמן Q - מספר העלים שנושרים בין 12:05 ל- 12:30 .

א. חשבו את $COV(4Y,Q+6)$.

ב. מה המתאם בין Y ל- Q?

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

11. נתון ש : $Y\sim B(1,p)$ כאשר $0<p<1$ הוכח שאם מתקיים :

$P(X=x\,|\,Y=0)=P(X=x\,|\,Y=1)$ לכל X , אזי X ו- Y הם משתנים בלתי תלויים.

12. נתון ש $X\sim B(n,p)$ ו- $Y\sim B(m,p)$ והם בלתי תלויים זה בזה. הוכח שמתקיים

$X\,|\,\mathrm{X+Y=k}\sim \mathrm{HG(n+m,n,k)}$.

פתרונות :

שאלה 1 :

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

ב.

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

ג. 0.816

ד. 0.816

שאלה 2 :

א.

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

ב. $V(Y)=\frac{12}{49}$ $E(Y)=\frac{3}{7}$ $V(X)=\frac{24}{49}$ $E(X)=\frac{12}{7}$

ג. $-\frac{8}{49}$

ד. $V(W)=\frac{20}{49}$ $E(W)=\frac{6}{7}$

ה. ‎-1

שאלה 3:

א.

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

ב. המשתנים תלויים.

ד. 1

שאלה 4 :

א. $\frac{185}{220}$

ג. 1.714

ד. 0.071

שאלה 5:

ב. X ו-Y בלתי מתואמים.

ג. 4.128

ד.0

<u>שאלה 6 :</u>

א. נכון

ב. לא נכון

ג. נכון

ד. לא נכון

<u>שאלה 7 :</u>

א. בלתי תלויים

ב. תלויים

ד. חלקי ושלילי.

<u>שאלה 8 :</u>

הוכחה

<u>שאלה 9:</u>

א.1000

ב. 0.316

<u>שאלה 10 :</u>

-0.5

<u>שאלה 11 :</u>

הוכחה

<u>שאלה 12 :</u>

הוכחה

פרק 34 - קומבינציות לינאריות להתפלגות נורמאלית

רקע:

כל קומבינציה לינארית של משתנים המתפלגים נורמאלית מתפלגת נורמאלית בעצמה.

דוגמה: (פתרון בהקלטה)

הגובה של גברים במדינת ישראל מתפלג נורמלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ, כמו כן הגובה של נשים במדינה מתפלג נורמלית עם תוחלת של 165 ס"מ וסטיית תקן של 8 ס"מ.

מה הסיכוי שגבר אקראי מהמדינה יהיה גבוה מאישה אקראית? (0.7823)

תרגילים:

1. המשקל של גברים במדינת ישראל מתפלג נורמלית עם תוחלת של 75 ק״ג וסטיית תקן של 10 ק״ג.
כמו כן המשקל של נשים במדינה מתפלג נורמלית עם תוחלת של 65 ק״ג וסטיית תקן של 8 ק״ג.
מה הסיכוי שאישה אקראית תהיה בעלת משקל גבוה יותר מגבר אקראי?

2. ההוצאה השנתית על ביגוד לאדם מתפלג נורמלית עם תוחלת של 3000 ₪ וסטיית תקן של 1000 ₪ . ההוצאה השנתית על בילויים מתפלגת נורמלית עם תוחלת של 4000 ₪ וסטיית תקן של 1500 ₪. מקדם המתאם בין ההוצאה השנתית על ביגוד וההוצאה השנתית על בילויים הינו 0.6.
א. מה התוחלת ומהי סטיית התקן של התפלגות ההוצאה השנתית הכוללת על ביגוד ובילוי?
ב. מה הסיכוי שההוצאה השנתית הכוללת על ביגוד ובילוי תעלה על 8000 ₪?
ג. מהו העשירון העליון של ההוצאה השנתית הכוללת על ביגוד ובילוי?

3. צריכת הירקות היומית במסעדה מתפלג נורמלית עם תוחלת של 50 ק״ג וסטיית תקן של 4 ק״ג. נתון שמחיר ק״ג ירק הוא 6 ₪ לקילו.
א. מה התוחלת ומהי השונות של העלות היומית של ירקות למסעדה?
ב. מה ההסתברות שהעלות היומית על ירקות תהיה נמוכה מ-290 ₪?
ג. מהו האחוזון ה-40 של התפלגות העלות היומית של המסעדה על ירקות?

4. 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?

5. לדוד משה הייתה חווה. בחווה פרה ועזה. תנובת החלב של הפרה מתפלג נורמאלית עם ממוצע של 20 ליטר ביום וסטיית תקן של 5 ליטר ותנובת החלב של העזה מתפלג גם כן נורמאלית עם ממוצע של 10 ליטר וסטיית תקן של 2 ליטר. כל ליטר חלב פרה נימכר ב- 2 ₪ וליטר חלב עזה נימכר ב-3 ₪.
א. מה הסיכוי שהפדיון היומי של דוד משה מחלב יהיה לפחות 62 ₪?
ב. מה הסיכוי שמתוך 5 ימים יהיו לפחות 4 ימים בהם תנובת החלב מהפרה והעזה ביחד תהיה מתחת ל- 30 ליטר ?
ג. מה הסיכוי שביום מסוים תנובת הפרה תהיה נמוכה מתנובת העזה?

פתרונות:

שאלה 1

0.2177

שאלה 2

א. תוחלת 7000, סטיית תקן 2247.

ב. 0.3264

ג. 9881

שאלה 3

א. תוחלת 300, שונות 576.

ב. 0.3372

ג. 294

שאלה 4

א. תוחלת 3000 מ"ל וסטיית תקן 40 מ"ל.

ב. 0.0062

פרק 35 - התפלגות הדגימה

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.

א. 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. מי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטית התקן של ממוצע המדגם?

ב. דגימה מהתפלגות נורמאלית

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע μ ושונות σ^2 ממוצע המדגם גם יתפלג נורמאלית :

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$$

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

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?

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 נורמאלית $\bar{x} \sim\succ N(\mu, \frac{\sigma^2}{n})$.

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?

תרגילים :

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?

ה. מהו תוחלת ממוצע המדגם?

ו. מהי טעות התקן?

2. להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב מסוים :

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

נגדיר את x להיות מספר המקלטים של משפחה אקראית.

א. בנו את פונקצית ההסתברות של x.

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

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

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

4. משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם
א. מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?

נתון כי ביום מסוים נולדו 4 תינוקות.
ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק״ג ?
ג. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק״ג?
ד. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
ה. הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?

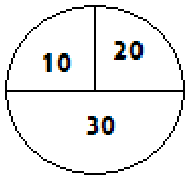
5. הגובה של המתגייסים לצה״ל מתפלג נורמאלית עם תוחלת של 175 ס״מ וסטיית תקן של 10 ס״מ. ביום מסוים התגייסו 16 חיילים.
א. מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס״מ?
ב. מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס״מ?
ג. מה ההסתברות שהגובה הממוצע שלהם יסטה מתחולת הגבהים בפחות מ-5 ס״מ?
ד. מהו הגובה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

6. הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך פתרון 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 מ755 סמ״ק?
ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ״ק?
ד. בקבוקיי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?

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

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



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.
א. בנו את פונקצית 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. מטילים קובייה 50 פעמים בכל פעם מתבוננים בתוצאה של הקובייה. מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72 ב- 50 ההטלות?

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. העזר במשפט הגבול המרכזי.

13. נתון משתנה מקרי בדיד בעל פונקצית ההסתברות הבאה :

2	4	6	8	X
¼	¼	¼	¼	P(X)

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

14. נתון ש $X \sim N(\mu,\sigma^2)$ דגמו 5 תצפיות מאותה התפלגות והתבוננו בממוצע המדגם : $\bar{X}$

לכן $P(\bar{X} > \mu)$ יהיה : (בחר 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 ש :

(בחר בתשובה הנכונה)

א. $X \sim N(\mu,\frac{\sigma^2}{200})$

ב. $\mu \sim N(\mu,\frac{\sigma^2}{200})$

ג. $\bar{X} \sim N(\mu,\sigma^2)$

ד. $\bar{X} \sim N(\mu,\frac{\sigma^2}{200})$

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

(בחר בתשובה הנכונה)

א. μ ו- $\bar{X}$ יהיו משתנים מקריים.

ב. μ יהיה משתנה מקרי ו $\bar{X}$ קבוע.

ג. $\bar{X}$ יהיה משתנה מקרי ו μ קבוע.

ד. μ ו $\bar{X}$ יהיו קבועים.

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

א. מהי סטיית התקן של משקל חפיסת שוקולד בודדת?

ב. מה הסיכוי שמתוך 4 קרטונים בדיוק קרטון אחד יהיה עם משקל ממוצע לחפיסה הנמוך מ-100 גרם?

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

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

20. הוכיחו שאם משתנה מתפלג כלשהו עם תוחלת μ ושונות σ^2 ומבצעים מדגם בגודל n של תצפיות בלתי תלויות מהמשתנה , אזי מתקיימות התכונות הבאות לגבי ממוצע המדגם :

$$E(\bar{x}) = \mu$$

$$V(\bar{x}) = \frac{\sigma^2}{n}$$

פתרונות:

שאלה 2

א.

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

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

ג. $\mu_{\bar{x}} = 2.05 \ \sigma_{\bar{x}}^2 = 0.2369$

$\sigma(\overline{X}) = 0.486$

שאלה 3

$\mu_{\bar{x}} = 3.5$

$\sigma(\overline{X}) = 1.21$

שאלה 4

א. 0.8413

ב. 0.0013

ג. 0

ד. 0.1974

שאלה 6

א. 0.0465

ב. 27.71

ג. 0.2628

שאלה 7

א. 0

ב. 0.1587

ג. 0.1587

ד. 0.5

שאלה 8

א. 0.5468

ב. 0.6826

שאלה 9

א.

	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

שאלה 10

0.0475

שאלה 11

0.1814

שאלה 12

א. 0.9772

ב. 0.0228

ג. 271

שאלה 14

התשובה ב

שאלה 15

התשובה ד

שאלה 16

התשובה ג

שאלה 17

א. 2.429

ב. 0.25

התפלגות סכום תצפיות המדגם ומשפט הגבול המרכזי

רקע:

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

כאשר כל התצפיות נדגמו באקראי מאותה אוכלוסייה.

כלומר, היו $X_1,...,X_n$ - משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה μ ושונותה σ^2 אזי:

א. התוחלת והשונות של סכום התצפיות:

$$E(T)=n\mu$$
$$V(T)=n\sigma^2$$

ב. דגימה מתוך התפלגות נורמלית:

אם $X \sim N(\mu,\sigma^2)$ אזי

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

ג. משפט הגבול המרכזי :

אם x מתפלג כלשהו וידוע

$$E(X)=\mu$$
$$V(X)=\sigma^2$$

אזי עבור מדגם מספיק גדול (לפחות 30)

$$T \sim\succ N(n\mu, n\sigma^2)$$

דוגמה: (פתרון בהקלטה)

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

א. מה התוחלת וסטיית התקן של סך המשכורות שיופקדו לסניף הבנק על ידי העובדים הללו?

ב. מה ההסתברות שלסניף יופקד פחות מ-780 אלף ₪ ע״י אותם עובדים? (0.1587)

תרגילים:

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 פונקצית ההסתברות של מספר המכוניות ליחידת דיור:

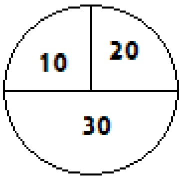
x	1	2	3	4	5
$P(X=x)$	0.1	0.2	0.3	0.25	0.15

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

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

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

5. בקזינו ישנה רולטה עליה מסומנים המספרים הבאים :



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.

א. אם האדם משחק את המשחק 50 פעמים מה ההסתברות שבסך הכול יזכה בסכום של 1050 שקלים ומעלה?

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

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

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

7. אורך חיי סוללה בשעות הוא בעל פונקצית הצפיפות הבאה :

$$f(x)=2x \quad 0<x<1$$

ברגע שסוללה מתרוקנת מחליפים אותה במידית בסוללה אחרת. כמה סוללות יש להחזיק במלאי אם רוצים שבסיכוי של 90% לפחות המלאי יספיק עבור 35 שעות לפחות?

פתרונות:

שאלה 1

א. 0.6915

ב. 0.8413

ג. 0.5

שאלה 2

א. תוחלת 3000 מ"ל וסטיית תקן 40 מ"ל

ב. 0.0062.

שאלה 4

א. 0.883

שאלה 5

א. 0.8997

ב. תוחלת : 1.111 שונות 0.1239

שאלה 7

56

התפלגות מספר ההצלחות במדגם - הקרוב הנורמלי להתפלגות הבינומית

רקע:

תזכורת על התפלגות בינומית

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

מספר ההצלחות במדגם נסמן ב –Y.

מחלקים כל תצפית במדגם להצלחה או כישלון.

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

תצפית

הצלחה כישלון

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 : $V(y) = npq$

<u>קירוב נורמלי עבור התפלגות בינומית</u>

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 : $Z_y = \dfrac{y-np}{\sqrt{npq}}$

<u>תיקון רציפות:</u>

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

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

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

הערות:

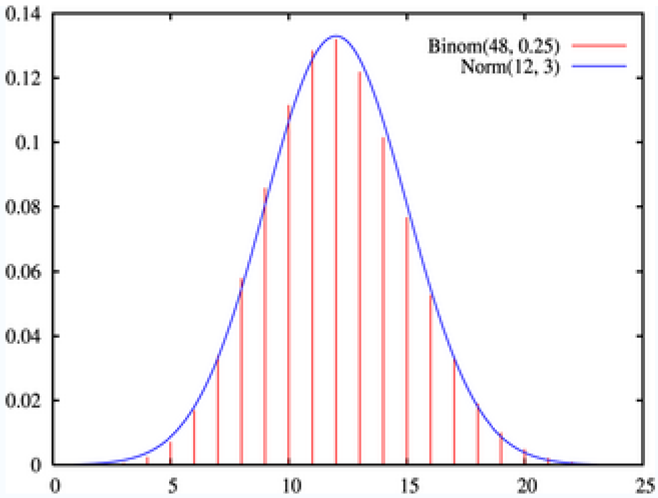
- התנאים למעבר מבינומי לנורמלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:
  1. $n \cdot p \geq 5$
  2. $n \cdot (1-p) \geq 5$
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
  1. $n \cdot p \geq 10$
  2. $n \cdot (1-p) \geq 10$
- וישנם מרצים שפשוט התנאי שהם נותנים הוא: $(n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

דוגמה: (הפתרון בהקלטה)

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

א. מה הסיכוי שבדיוק 14 מתוכם יהיו זקוקים למשקפיים?

ב. מה הסיכוי שלכל היותר 13 מתוכם זקוקים למשקפיים?



תרגילים:

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 ש הסיכוי 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 ?

<u>פתרונות</u> :

<u>שאלה 1</u>

א. 0.201

ב. 0.3758

ג. התוחלת : 2, סטיית התקן : 1.2649

<u>שאלה 2</u>

א. 0.9332

ב. 0.6915

<u>שאלה 3</u>

0.1611

<u>שאלה 4</u>

א. 0.9406

<u>שאלה 5</u>

א. 0.015

<u>שאלה 6</u>

0.9996

<u>שאלה 7</u>

0.8643

התפלגות פרופורציית ההצלחות במדגם

רקע:

בפרק זה נדון על התפלגות הדגימה של פרופורציית המדגם.

Y- מספר ההצלחות במדגם (למשל , מספר המובטלים במדגם)

$\hat{p}=\frac{y}{n}$ - פרופורציית ההצלחות במדגם (למשל , שיעור המובטלים במדגם)

למשל,

$n=200$

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:

$$E(\hat{P})=p$$

$$V(\hat{P})=\frac{pq}{n}$$

$$\sigma(\hat{p})=\sqrt{\frac{pq}{n}}$$

משפט הגבול המרכזי עבור הפרופורציה המדגמית :

אם $np\geq 5\,\&\,nq\geq 5$ אזי $\hat{p}\sim N(p,\frac{pq}{n})$

$$Z_{\hat{p}}=\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$$

הערות:

- התנאים לקרוב הנורמאלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:
  1. $n \cdot p \geq 5$
  2. $n \cdot (1-p) \geq 5$
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
  1. $n \cdot p \geq 10$
  2. $n \cdot (1-p) \geq 10$
- וישנם מרצים המשתמשים בתנאי: $(n \geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור לנורמלית.
- כיוון שפרופורציה אינה חייבת להיות מספר שלם בהכרח לא נהוג לבצע כאן תיקון רציפות.

דוגמה: (פתרון בהקלטה)

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

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

תרגילים:

1. במדינה מסוימת 10% מכלל האוכלוסייה הינם מובטלים. נדגמו באקראי 140 אנשים מהמדינה.
   א. מה התוחלת ומהי השונות של פרופורציות המובטלים שנדגמו?
   ב. מה ההסתברות שבמדגם לפחות 10% יהיו מובטלים?
   ג. מה ההסתברות שלכל היותר 9% מהמדגם יהיו מובטלים?

2. נניח כי 30% מהאוכלוסייה תומכים בהצעת חוק מסוימת. אם נדגום מהאוכלוסייה 200 איש. חשבו את ההסתברויות הבאות:
   א. לפחות 35% יתמכו בהצעת החוק במדגם.
   ב. לכל היותר 25% יתמכו בהצעת החוק במדגם.
   ג. יותר מ – 27% יתמכו בהצעת החוק במדגם.

3. לפי נתוני משרד התקשורת 40% מהאוכלוסייה מחזיקים בטלפון נייד מסוג "סמארטפון". נדגמו 400 אנשים מהאוכלוסייה.
   א. מה ההסתברות שבמדגם לכל היותר ל 40% יש סמארטפון?
   ב. מה ההסתברות שבמדגם לרוב יש 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 האמתית ביותר
   מ-4%?
   ג. כמה בתי אב יש לדגום כדי שהסטייה בין הפרופורציה המדגמית לפרופורציה האמתית 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. ל"מועדון ה-700" נכללים נבחנים שמקבלים ציון מעל 700 בפסיכומטרי. מה הסיכוי שבמועד בו נבחנו 2000 נבחנים אקראיים יהיו לפחות 3% המשתייכים למועדון?

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

א. הוכיחו ש :

$$E(\hat{P}) = p$$
$$V(\hat{P}) = \frac{P(1-P)}{n}$$

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

פתרונות:

שאלה 1

א. התוחלת: 0.1, השונות: 0.00064

ב. 0.5

ג. 0.3446

שאלה 2

א. 0.0618

ב. 0.0618

ג. 0.8238

שאלה 3

א. 0.5

ב. 0

ג. 0.8968

ד. גדלה

שאלה 4

א. 0.0062

ב. 0.0456

פרק 36 - אמידה נקודתית

אומד חסר הטיה

רקע:

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

דוגמה : (פתרון בהקלטה)

המשתנה X הוא בעל פונקציית ההסתברות הבאה :

X	1	2	3
הסתברות	2θ	$1-6\theta$	4θ

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

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

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

ב. מהי ההטיה של האומד T_1.

ג. תקנו את T_1 כך שיהיה אומד חסר הטיה.

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

ד. מוצא האומד הבא : $T_3=1.5X_1-X_2-1$. האם הוא עדיף על האומד שהצעת בסעיף ג?

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

ה. מצאו אומד חסר הטיה ל: $P(X=3)$.

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

ו. מצאו אומד חסר הטיה לשונות של X.

<u>תזכורות חשובות:</u>

- אם $Y=aX+b$ אזי:

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

- אם $X_1, X_2, \ldots\ldots, X_n$ משתנים מקרים אזי:

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

אם $X_1, X_2, \ldots\ldots, X_n$ משתנים מקריים <u>בלתי תלויים</u> בזוגות, אזי:

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

תרגילים:

1. הציון במבחן מסוים של תלמידי כתה ח׳ הנו משתנה מקרי בעל תוחלת μ וסטיית תקן 10. כדי לאמוד את התוחלת - μ, נלקח מדגם של 5 ציונים $X_1, \ldots, X_5$. שלושה חוקרים הציעו אומדים לתוחלת על סמך מדגם זה:

חוקר א׳ הציע: $$T_1 = \frac{X_1 + \ldots + X_5}{5}$$

חוקר ב׳ הציע: $$T_2 = \frac{2X_1 - X_3 + X_4}{2}$$

חוקר ג׳ הציע: $$T_3 = \frac{2X_1 + X_3}{2}$$

א. איזה מן האומדים הוא חסר הטיה?
ב. הצע תיקון לאומד המוטה כך שיהיה חסר הטיה.
ג. במדגם התקבלו הציונים הבאים: 65, 78, 58, 82, 100. חשבו את האומדנים המתקבלים עבור האומדים חסרי ההטיה.
ד. איזה מבין שני האומדים חסרי ההטיה עדיף? נמקו.

2. כדי לאמוד את המשקל הממוצע של הנשים בארה״ב, נבחר מדגם של 2n נשים. נסמן את שונות הגובה ב- σ^2. הוצעו שני אומדים לממוצע המשקל על סמך מדגם זה:

$$T_2 = \frac{1}{2n}\sum_{i=1}^{2n} X_i \qquad\qquad T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

א. בִּדקו לגבי כל אומד אם הוא בלתי מוטה.
ב. איזה אומד עדיף? נמקו.

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

א. פתחו אומד חסר הטיה ל-P.
ב. מהו אומד חסר הטיה לסיכוי לכישלון בניסיון בודד.
ג. מהו אומד חסר הטיה ל- $E(X)$.
ד. מצאו אומד חסר הטיה ל- $E(X^2)$

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

פונקציית ההסתברות של X – מספר המניות שיעלו ביום מסוים :

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

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

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

5. בקרב המטפלות בת"א מספר התינוקות שבטיפולן הוא משתנה מיקרי בעל התפלגות התלויה
בפרמטר θ באופן הבא :

הסיכוי שמטפלת תטפל בתינוק אחד בלבד הוא 3θ,

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

הסיכוי שמטפלת תטפל ב- 3 תינוקות הוא θ .

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

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

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

ג. מהו האומדן לפרמטר θ על סמך תוצאות המדגם .

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

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

6. קבע אילו מהטענות הבאות נכונות :

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

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

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

א. $\frac{X}{20}$

ב. $\frac{Y}{20}$

ג. $\frac{X+Y}{60}$

ד. $\frac{2X+Y}{80}$

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

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

ב. מצא אומד חסר הטייה ל- $\theta(1-\theta)$ המתבסס על T_1 ו- T_2 .

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

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

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

10. $X_i \sim N(\mu,1)$ כאשר $i = 1, 2..., n$

נתון שהתצפיות הינן בלתי תלויות זו בזו. מצא אומד חסר הטיה ל- μ^2 .

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

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

א. הראה כי האומד $3\bar{X}$ הנו אומד בלתי מוטה ל β.

ב. מצא את השונות של האומד מהסעיף הקודם.

12. $X_1, X_2, ..., X_n$ הינם משתנים מקריים רציפים בלתי תלויים בעל פונקצית הצפיפות הבאה :

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

א. בטא את ערכו של A באמצעות θ כדי שפונקצית הצפיפות תהיה לגיטימית.

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

פתרונות:

שאלה 1

א. T_1 ו- T_2

ב. $\frac{2}{3}T_3$

ג. $T_1 = 76.6$ $T_2 = 110$

ד. T_1

שאלה 2

ב. T_2

שאלה 3

א. $\frac{x}{n}$

ב. $1-\frac{x}{n}$

ג. x

שאלה 4

א. $\frac{3x}{2}$

ב. $\frac{3\overline{x}}{2}$

שאלה 5

א. $1-\frac{x}{2}$

ג. 0.125

ה. לשונות 0.917

שאלה 6

א. נכון.

ב. לא נכון.

שאלה 7

תשובה : ב

שאלה 8

א. $T_1 \cdot T_2$

ב. $T_1 - T_1 \cdot T_2$

שאלה 9

הוכחה

שאלה 10

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

שאלה 11

ב. $V(3\overline{X}) = \frac{3-\beta^2}{n}$

שאלה 12

א. $A = \frac{2}{\theta^2}$

ב. $\theta = \frac{3}{2}\overline{X}$

קריטריון MSE - תוחלת ריבוע הטעות

רקע

הקריטריון הנפוץ ביותר כדי לבדוק את טיב האומד הוא קריטריון MSE. תוחלת ריבוע טעות האמידה.

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

$V(\hat{\theta})$ - הינה שונות האומד.

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

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

דוגמה: (הפתרון בהקלטה)

נתון משתנה X המתפלג אחיד רציף באופן הבא: $X \sim U(3,\theta)$. מוצעים שני אומדים לפרמטר θ

על סמך תצפית בודדת $T_1 = 2X-3$ ו- $T_2 = \frac{3X-3}{2}$

איזה אומד עדיף לאמידת הפרמטר θ?

תרגילים :

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

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

   א. הצע אומד חסר הטיה ל- θ על סמך X.

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

   ג. איזה אומד מבין האומדים של סעיף א או ב עדיף?

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

   א. מי מבין האומדים הללו הוא חסר הטיה?

   ב. מי מבין האומדים עדיף אם P=0.5?

   ג. מי מבין האומדים עדיף אם P=0.1?

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

$$\hat{\lambda}_1 = \frac{\sum_{i=1}^{10} X_i}{10}$$

$$\hat{\lambda}_2 = \frac{2\cdot\sum_{i=1}^{5} X_i + 2\cdot\sum_{i=6}^{10} X_i}{10}$$

X_i = מספר השריפות בחודש אוקטובר ה- i.

איזה מהאומדים עדיף לצורך אמידת הפרמטר λ ?

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

פתרונות :

שאלה 1

זה עם השלוש תצפיות.

שאלה 2

א. $2x$

ג. סעיף ב

שאלה 3

א. T_1

ב. T_2

ג. T_1

שאלה 5

הוכחה

אומד חסר הטיה יעיל ביותר - MVUE

(Minimum- variance unbiased estimator)

רקע:

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

דוגמה: (פתרון בהקלטה)

לרשת חנויות ישנם שני סניפים. מספר הלקוחות הנכנסים לכל סניף ביום מתפלג פואסונית עם קצב של λ בסניף A וקצב של 2λ בסניף B. נדגמו n ימים מכל סניף ונבדק בכל יום:

X_i -מספר הלקוחות שנכנסו לסניף A ביום i.

Y_j -מספר הלקוחות שנכנסו לסניף B ביום j.

על מנת לאמוד את λ מוצע האומד: $\alpha\bar{X} + \beta\bar{Y}$.

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

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

תרגילים:

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

כמו כן נגדיר : $T = aT_1 + bT_2$

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

ב. σ_1^2 ו σ_2^2 הם השונויות של T_1 ו- T_2 בהתאמה. מצאו את a ו- b כך ש-T יהיה אומד חסר הטיה ל θ ובעל שונות מינימלית.

2. במפעל 3 מכונות המייצרות את אותו חלק. תוחלת הקוטר של החלקים המיוצרים בכל מכונה זהה אומנם השונויות של כל מכונה שונות ומקיימות :

$\sigma_2^2 = 2\sigma_1^2 \qquad \sigma_3^2 = 3\sigma_1^2$

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

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

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

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

ב. נניח ש $a_1 = a_2$. מה במקרה זה המשקלים המביאים את האומד להיות MVUE?

פתרונות :

שאלה 1:

א. $a+b=1$

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

שאלה 2:

א. $\sum_{i=1}^{3} a_i = 1$

ב. $a_1=a_2=0.4$
$a_3=0.2$

<u>פרק 37 - מושגים בסיסיים באמידה</u>

<u>רקע:</u>

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 : $|\hat{\theta}-\theta|$ - ההפרש בין האומד לאמת(הפרמטר).

<u>דוגמה:</u> <u>(פתרון בהקלטה)</u>

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.

מה הפרמטר בדוגמה זו?

מהי טעות האמידה של ערוץ 10?

- $\hat{\theta}$ יהיה אומד חסר הטיה ל θ אם התוחלת של $\hat{\theta}$ תהיה שווה ל θ : $E(\hat{\theta})=\theta$
- טעות התקן של אומד היא סטיית התקן שלו , כלומר : $\sigma(\hat{\theta})=S.E$

<u>להלן פרמטרים מרכזיים והאומדים שלהם</u>:

<u>ממוצע האוכלוסייה</u>: 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: $\sigma(\bar{x})=\frac{\sigma}{\sqrt{n}}=SE$

<u>פרופורציה באוכלוסייה</u>: 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: $\sigma(\hat{P})=\sqrt{\frac{p\cdot(1-p)}{n}}$

<u>שונות האוכלוסייה</u>: 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.

$$S^2=\frac{\sum(x_i-\bar{x})^2}{n-1}=\frac{\sum x_i^{\,2}-n\bar{x}^2}{n-1}$$

<u>הערה:</u> אומד הוא הנוסחה הכללית לאמידת הפרמטר ואומדן הוא הערך הספציפי שהתקבל במדגם מסוים.

<u>דוגמה: (פתרון בהקלטה)</u>

נדגמו 10 משפחות בתל אביב ונבדק עבור כל משפחה מספר הילדים שלה. להלן התוצאות שהתקבלו:

2,1,3,2,1,4,5,2,1,3

אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים:

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

תרגילים:

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. מצא אומדן חסר הטיה לגובה הממוצע של מתגייסי צה״ל.

ב. מצא אומדן חסר הטיה לשונות הגבהים של מתגייסי צה״ל.

ג. מצא 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 :

$$\sum_{i=1}^{20} X_i = 162 \quad \sum_{i=1}^{20} X_i^2 = 1502.2 .$$

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.

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. $X_1, X_2, \ldots\ldots, X_{16}$ 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.

פתרונות:

שאלה 3

א. 177.9
ב. 64.1
ג. 0.4

שאלה 4

א. 8.1
ב. 3.16

שאלה 5

התשובה היא ד.

שאלה 6

התשובה היא ג.

שאלה 7

התשובה היא ד.

שאלה 8

התשובה היא ד.

שאלה 9

התשובה היא ג.

פרק 38 - רווח סמך לתוחלת (ממוצע האוכלוסייה)

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%.

מהי אוכלוסיית המחקר?

מה המשתנה באוכלוסייה?

מה הפרמטר שהחוקר רצה לאמוד?

מהו רווח הסמך?

מה אורך רווח הסמך?

מהי רמת הביטחון של רווח הסמך?

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

הפרמטר שנרצה לאמוד : μ

האומד נקודתי : $\overline{x}$

התנאים לבניית רווח הסמך :

1 $X \sim N$ או $n \geq 30$

2 σ^2 (שונות האוכלוסייה) ידועה

הנוסחה לרווח הסמך :

$$\overline{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

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.

<u>שגיאת האמידה המקסימלית :</u>

$$\varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

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

<u>דוגמה</u> : (פתרון בהקלטה)

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

<u>קשרים מתמטיים ברווח הסמך :</u>

- 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 $z_{1-\alpha/2}$ יותר גבוה, ורווח הסמך יותר ארוך.

תרגילים :

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. בנה 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. הסבר 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 ב50%?
   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?
5. בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (99,105). 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%, מי מהמשפטים הבאים אינו יהיה נכון.
א. אורך רווח הסמך החדש יהיה קטן יותר.
ב. גודל המדגם יהיה כעת קטן יותר.
ג. המרחק בין ממוצע המדגם לקצות רווח הסמך יהיו קטנים יותר ברווח הסמך החדש.
ד. רמת הביטחון לבנות רווח הסמך החדש תהיה קטנה יותר.

9. חוקר בנה רווח סמך ל- μ וקיבל $48 < \mu < 54$ מה נכון בהכרח:
א. $\mu = 51$
ב. $\overline{X} = 6$
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? (בחר בתשובה הנכונה)
א. רמת הביטחון.
ב. סטיית התקן באוכלוסייה.
ג. מספר המשתתפים.
ד. סטיית התקן במדגם.

11. חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא : $63 < \mu < 83$.
נתון שסטיית התקן בהתפלגות הייתה ידועה לו ושהמדגם התבסס על 40 תצפיות.
א. אם החוקר היה רוצה לבנות רווח סמך באורך 10. כמה תצפיות עליו היה לדגום?
ב. רווח הסמך שנבנה על ידי החוקר היה ברמת סמך של 95%. בנה את רווח הסמך שהיה מתקבל ברמת סמך של 98%.

12. נתון משתנה מקרי רציף מתפלג אחיד : $X_i \sim U(\mu - 0.5, \mu + 0.5)$. נרצה לאמוד את μ . מצאו רווח סמך ל-μ ברמת-בטחון של 0.95 אם במדגם של 45 תצפיות התקבל: $\overline{x} = 74$.

(תזכורת על השונות בהתפלגות אחידה רציפה : $Var(X_i) = \frac{(b-a)^2}{12}$)

<u>פתרונות :</u>

<u>שאלה 2</u>

$4920.6 < \mu < 4979.4$

<u>שאלה 3</u>

א. $223.42 < \mu < 236.58$

ב. $222.16 < \mu < 237.84$

<u>שאלה 5</u>

א. 102

ב. 3

ג. 0.9544

<u>שאלה 6</u>

א. $83.5 < \mu < 4.42$

ב. יקטן פי 2

ג. גדל

<u>שאלה 7</u>

א. 87

ב. 5

ג. 0.9544

<u>שאלה 8</u>

א. 139

ב. $21 < \mu < 25$

<u>שאלה 9</u>

התשובה היא : ב

<u>שאלה 10</u>

התשובה היא : ג

<u>שאלה 11</u>

התשובה היא : ד

<u>קביעת גודל מדגם באמידת תוחלת עם שונות אוכלוסייה ידועה</u>

<u>רקע:</u>

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.

<u>דוגמה:</u> (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)

תרגילים:

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?

פתרונות :

שאלה 1

780

שאלה 2

א. 139

ב. הדבר יקטין את ε פי 2.

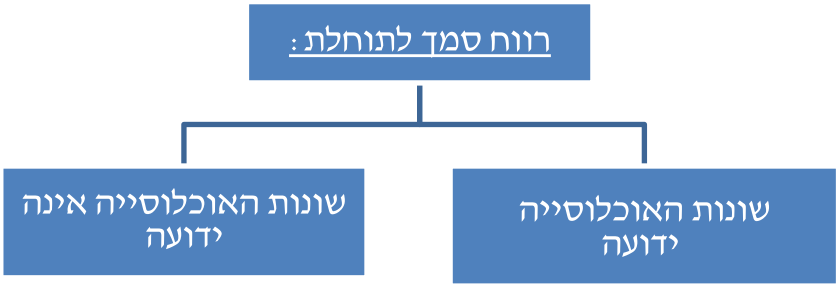
שאלה 3

$n=62$

רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונות האוכלוסייה אינה ידועה

רקע:

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



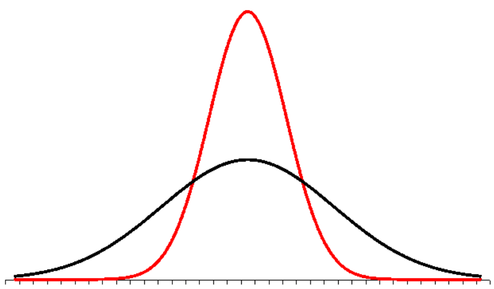
בפרק זה נעסוק במקרה ששונות האוכלוסייה (σ^2) אינה ידועה לנו .מקרה יותר פרקטי .

התנאי: $X \sim N$ או שהמדגם גדול

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

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

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.

<u>דוגמה :</u> (פתרון בהקלטה)

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

במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח' . להלן התוצאות שהתקבלו בדקות : 4.7,5.2,4.6,5.3.

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

<u>פתרון :</u>

$4.39< \mu <5.51$

תרגילים:

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: 27,34,32,40,30 .
א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?
ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים ?

4. ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.
א. בנו רווח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.
ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמתית של כלל הנבחנים.
ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.

5. נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$ $\sum_{i=1}^{60} X_i^2 = 643.19$. בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

6. נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו.
להלן התוצאות שהתקבלו: $\bar{x} = 13.8$, $S = 2$
בנו רווח סמך ברמת סמך של 96% לממוצע ההשכלה של אזרחים מעל גיל 50.

7. שני סטטיסטיקאים בנו רווח בר-סמך לאותו פרמטר **μ**. לכל אחד מהסטטיסטיקאים מדגם אחר, אך באותו גודל 10 . שניהם קבעו אותה רמת סמך.

סטטיסטיקאי א : הניח $\sigma = 20$

סטטיסטיקאי ב : חישב לפי המדגם וקיבל **S =20**

למי משני הסטטיסטיקאים יהיה רווח סמך ארוך יותר? (בחר בתשובה הנכונה)

א. סטטיסטיקאי א

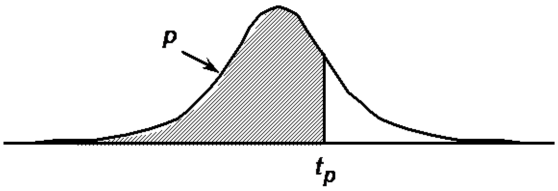
ב. סטטיסטיקאי ב

ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.

ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

8. נתון ש : $X \sim N(\mu,\sigma^2)$ ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגמית 10. אורך רווח הסמך שהתקבל הוא : 8.765 . מהי רמת הביטחון של רווח הסמך?

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

<u>פתרונות:</u>

<u>שאלה 1</u>

א. $79.88< \mu <89.72$

<u>שאלה 4</u>

א. $96.63< \mu <107.37$

ב. $96.90< \mu <107.10$

<u>שאלה 5</u>

$3.149< \mu <3.351$

<u>שאלה 8</u>

90%

פרק 39 - רווח סמך לפרופורציה

רקע:

מטרה : לאמוד את P – פרופורציה באוכלוסייה.

האומד הנקודתי : $\hat{p}=\frac{y}{n}$ (Y- מספר ההצלחות שבמדגם)

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

התנאי לבנות את רווח הסמך הינו 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 : $\hat{P}=\frac{A+B}{2}$ $\quad L=2\varepsilon$

דוגמה : (פתרון בהקלטה)

1. 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?

פתרון:

א. 7.5%<p<16.5%

ב. 2.29%

תרגילים:

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 היי-טק).
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?

פתרונות:

שאלה 3

א. 52

ב. 0.997

שאלה 5

א. $22.5\%<p<30.9\%$

ב. $45.91\%<p<60.72\%$

שאלה 6

200

קביעת גודל מדגם באמידת פרופורציה

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 ($\hat{p}$).

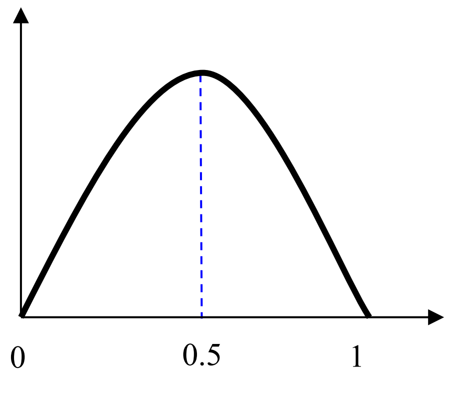
$\varepsilon = z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

נקבל ש: $n \ge \left(\frac{2\cdot Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\hat{p}(1-\hat{p})}}{L}\right)^2$

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 $\hat{p}=0.5$

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

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

<u>דוגמה:</u> (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%.

<u>פתרון:</u>

א. 423

ב. 271

תרגילים:

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. השאלות הבאות מתייחסות לסעיף 4 :
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 תקבל 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 מ3% ?

פתרונות:

שאלה 1

1068

שאלה 3

א. 601

ב. 108000 ₪.

פרק 40 - רווח סמך להפרש פרופורציות

רקע:

המטרה : לאמוד את $p_1 - p_2$: הפרש פרופורציות בין שתי אוכלוסיות שונות.

האומד הנקודתי: $\hat{p}_1 - \hat{p}_2$

התנאי לבניית רווח הסמך : כל מדגם מעל 30 או לבדוק שמספר ההצלחות ומספר הכישלונות בכל מדגם לפחות 5 בכל מדגם (יש כאלה שבודקים לפחות 10).

רווח סמך :

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

רק שאפס נופל בתחומי רווח הסמך להפרש הפרופורציה נאמר שלא ניתן לקבוע שקיים הבדל מובהק בין הפרופורציות באוכלוסיות.

דוגמה: (פתרון בהקלטה)

במטרה להשוות בין שתי תרופות נדגמו 200 איש שלקחו תרופה x. מתוכם 180 טענו שהתרופה עזרה להם. כמו כן נלקחו 300 איש שלקחו את תרופה y. מתוכם 150 טענו שהתרופה עזרה להם. בנו רווח סמך להפרש אחוזי ההצלחה של התרופות ברמת סמך של 95%. מה ניתן לומר על סמך רווח הסמך על ההבדלים בין התרופות?

פתרון :

(33% ,47%)

תרגילים:

1. מתוך 150 נשים שנדגמו באקראי 30% תמכו בהצעת חוק מסוימת. מתוך 200 גברים שנדגמו באקראי 25% תמכו בהצעת החוק.
א. בנו רווח סמך לפער בין אחוזי התמיכה של הנשים לעומת הגברים ברמת סמך של 96%.
ב. בנו רווח סמך ברמת סמך של 95% לאחוז התמיכה בהצעת החוק .

2. במחקר רפואי השתתפו 200 אנשים הסובלים מכאבים כרוניים. הם חולקו באקראי ל-2 קבוצות שוות בגודלן.
קבוצה 1 קיבלה את תרופה A וקבוצה שנייה קיבלה את תרופה B.
בקרב לוקחי תרופה A 90 טענו שמצבם השתפר. בקרב לוקחי תרופה B 70 טענו שמצבם השתפר.
א. בנו רווח סמך ברמת סמך של 95% להפרש בין שיעורי ההצלחה של שתי התרופות.
ב. האם על סמך סעיף א ניתן לקבוע שקיים הבדל בין התרופות מבחינת שיעורי ההצלחה?

3. נדגמו 200 משפחות מגוש דן. ל-70% מתוכן מכשיר DVD בבית.
נדגמו 300 משפחות מאזור הצפון ל- 65% מתוכן מכשיר DVD בבית.
א. בנו רווח סמך ברמת סמך של 98% לפרופורציות המשפחות בגוש דן עם DVD בבית.
ב. בנו רווח סמך ברמת סמך של 95% להפרש בין פרופורציות המשפחות בגוש דן עם DVD לבין פרופורציות המשפחות בצפון עם DVD.

פתרונות:

שאלה 2

א. $0.093 < P_A - P_B < 0.307$

שאלה 3

א. $0.625 < p < 0.7754$

פרק 41 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים

כששונויות האוכלוסייה ידועות

רקע:

מטרה: 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. $\sigma^2{}_1,\sigma^2{}_2$ ידועות.

2. $X_1,X_2\sim N$ או $n_1,n_1>30$

3. שני מדגמים בלתי תלויים.

רווח סמך:

$$(\bar{x}_1-\bar{x}_2)\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}$$

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

דוגמה: (פתרון בהקלטה)

נדגמו 100 תושבים מאזור a והמשכורת הממוצעת הייתה שם 9200 ₪.

כמו כן נדגמו 120 תושבים מאזור b וממוצע המשכורות שהתקבל שם 8700 ₪.

לצורך פתרון נניח שסטיית התקן של המשכורות באוכלוסיית שני האזורים היא 1800 ₪.

אמדו ברמת סמך של 90% את הפרש השכר הממוצע בין אזור a לאזור b .

תרגילים:

1. מעוניינים לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. ידוע שציוני הפסיכומטרי מתפלגים נורמאלית עם סטיית תקן 100. במדגם של 16 נבחנים חיילים התקבל ממוצע 543. במדגם של 20 תלמידי תיכון התקבל ממוצע 508. בנו רווח סמך לפער תוחלות הציונים בין חיילים לתלמידי תיכון ברמת סמך של 90%. מה ניתן להסיק מרווח סמך זה?

2. ציוני .I.Q מתוכננים כך שיתפלגו נורמאלית עם סטיית תקן של 15. במדגם של 20 נבחנים ישראלים התקבל ממוצע ציונים 104. במדגם של 23 נבחנים אמריקאיים התקבל ממוצע ציונים 99.

א. בנו רווח סמך ברמת סמך של 95% לפער בין ישראל לארה״ב בממוצע הציונים במבחן ה-IQ.

ב. האם קיים הבדל בין ישראלים לאמריקאים מבחינת ממוצע הציונים?

3. חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. ידוע שרמת הקשיות של ברגים מתפלגת נורמלית עם סטיית תקן של 4 יחידות. במדגם של 15 ברגים מסוג א׳ התקבל רמת קשיות ממוצעת של 28 יחידות ובמדגם של 12 ברגים מסוג ב׳ התקבל רמת קשיות ממוצעת של 25 . עבור אילו רמות בטחון יקבע שאין הבדל בין שני סוגי הברגים מבחינת ממוצע רמת הקשיות שלהם?

פתרונות :

שאלה 1

(-20,90)

שאלה 3:

רמות בטחון הגבוהות מ : 0.9476

כששונויות האוכלוסייה אינן ידועות אך שוות והמדגמים בלתי תלויים

רקע:

מטרה: 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.

פתרון : (1357,1643)

תרגילים:

1. נדגמו 15 ישראלים ו-15 אמריקאים.

כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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

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

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

25	21	20	22	X
12	17	25	18	Y

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

פרק 42 - רווח סמך לתוחלת ההפרש במדגם מזווג

רקע:

<u>מדגם מזווג</u>: 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

<u>התנאים לבניית רווח הסמך:</u>

- $x, y \sim N$
- המדגם מזווג

<u>נוסחת רווח הסמך:</u>

$$\bar{D} \pm t^{n-1}_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

כאשר דרגות החופש: $d.f = n-1$

<u>דוגמה:</u> (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.

מְצאו רווח סמך של 95% להפרש תוחלת הזמן בין שתי התוכנות.

תרגילים:

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

סמסטר א	סמסטר ב
74	80
68	84
90	87
75	76
82	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. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

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

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

פרק 43 - רווח סמך לשונות וסטיית תקן

רקע:

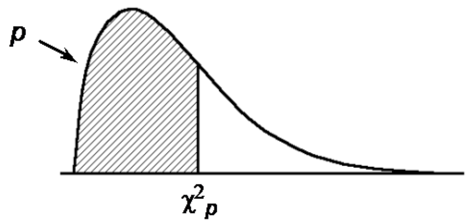
בפרק זה נדון על בניית רווח סמך לשונות האוכלוסייה.

התנאי לבניית רווח הסמך : המשתנה הנחקר מתפלג נורמלית , למרות שנהוג לא לדרוש את התנאי הזה אם המדגם מספיק גדול.

רווח הסמך יתבסס על התפלגות הנקראת חי בריבוע.

התפלגות זו היא התפלגות אסימטרית חיובית המתחילה מהערך אפס ותלויה בדרגות חופש.

דרגות החופש במקרה זה יהיו : n-1



רווח הסמך לשונות : $\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2,\,n-1}} \le \sigma^2 \le \dfrac{(n-1)S^2}{\chi^2_{\alpha/2,\,n-1}}$

כאשר $S^2 = \dfrac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n-1} = \dfrac{\sum_{i=1}^{n} X^2_i - n\cdot\bar{X}^2}{n-1}$ אומד לשונות הלא-ידועה.

אם נרצה לבנות רווח סמך לסטיית תקן אז נוציא שורש לרווח סמך לשונות.

דוגמה : (פתרון בהקלטה)

זמן התגובה מתפלג נורמאלית. במטרה לאמוד את שונות זמן התגובה נדגמו 4 תצפיות. להלן התוצאות בשניות : 4.7,5.2,4.6,5.3. בנו רווח סמך, ברמת סמך של 95% לשונות זמן התגובה באוכלוסייה.

פתרון :

$0.039<\sigma^2<1.708$

<u>תרגילים :</u>

1. חמישה מטופלים קבלו תרופה מסוימת. בדקו לכל מטופל את זמני התגובה שלו. להלן הזמנים שהתקבלו בדקות : 18,17,21,26,28.
בהנחה וזמני התגובה מתפלגים נורמאלית, בנו רווח סמך ברמת סמך של 95 % לשונות זמן התגובה.

2. נדגמו 20 ימים אקראיים מחודשי יולי-אוגוסט ונמדדה בהם הטמפ׳ במעלות צלזיוס בת״א.
במדגם התקבל טמפ׳ ממוצעת 30.8 וסטיית תקן מדגמית 1.1. בהנחה והטמפ׳ מתפלגת נורמאלית :
א. בנו רווח סמך לתוחלת הטמפ׳ בחודשים אלה בת״א ברמת סמך של 95%.
ב. בנו רווח סמך לסטיית התקן של הטמפ׳ בחודשים אלה בת״א ברמת סמך של 95%.

3. ציוני IQ בארה״ב מתפלגים נורמאלית עם ממוצע 100 וסטיית תקן 5. נבחנו 20 נבחנים ישראלים במבחן ה-IQ. להלן התוצאות שהתקבלו :

$$\sum_{i=1}^{20} X_i = 2080$$

$$\sum_{i=1}^{20} X_i^2 = 218,220$$

נניח שגם בישראל הציונים מתפלגים נורמאלית.
א. מצאו אומדנים לממוצע הציונים בישראל ולשונות הציונים בישראל באמצעות אומדנים חסרי הטיה.
ב. אמדו ברמת ביטחון של 95% את תוחלת הציונים של נבחנים בישראל.
ג. אמדו ברמת סמך של 90% את סטיית התקן של הציונים של נבחנים ישראלים.
ד. על סמך הסעיפים הקודמים, האם בישראל ממוצע הציונים וסטיית התקן של הציונים שונה מבארה״ב? הסבירו.

4. באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

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

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

נתון ש $X_i \sim N(\mu, \sigma^2)$

א. בנו רווח סמך ל- μ ברמת סמך של 95%.

ב. בנו רווח סמך ל- σ^2 ברמת סמך של 95%.

פתרונות :

שאלה 2

א. $30.285 < \mu < 31.315$

ב. $0.837 < \sigma < 1.607$

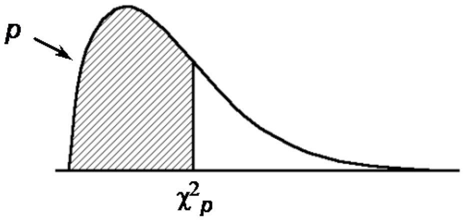
תשובה 3

א. לממוצע 104, לשונות 100.

ב. $99.32 < \mu < 108.68$

ג. $7.94 < \sigma < 13.7$

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p



p

df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
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

פרק 44 - תרגול מסכם ברווחי סמך

1. מהירות הגלישה באינטרנט במקום מסוים מתפלגת נורמאלית . בדקו את מהירות הגלישה ב- 30 זמנים אקראיים. מהירות הגלישה נמדדה ב-Mbps . מהירות מתחת ל- 10 Mbps מוגדרת על ידי החברה כנמוכה.
התוצאות שהתקבלו במדגם : ממוצע היה 87 עם סטיית תקן 17 ו-12 פעמים המהירות הייתה נמוכה.בנו רווחי סמך ברמת סמך של 95% לפרמטרים הבאים :
א. תוחלת מהירות הגלישה.
ב. סטיית תקן של מהירות הגלישה.
ג. הסיכוי שמהירות הגלישה תהיה נמוכה.

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 95% 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 הינו מגדילים את רמת הסמך שלו?

5. להלן תוצאות מדגם שבדק עבור כל משפחה האם יש לה בבית מכשיר טאבלט.

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

א. בנו רווח סמך להבדל בין אחוז המשפחות עם טאבלט בגוש דן ואחוז המשפחות בעלי טאבלט בשאר חלקי הארץ. ברמת סמך של 98%.

ב. בנו רווח סמך לפרופורצית משפחות בעלות טאבלט בכלל הארץ ברמת סמך של 95%.

6. הגובה של מתגייסים לצה"ל מתפלג נורמלית במדגם של 25 מתגייסים התקבלו התוצאות הבאות:

$$\bar{x} = 176.2cm$$

$$\sum (x_i - \bar{x})^2 = 2832cm^2$$

א. אמדו את הגובה הממוצע של המתגייסים ברמת סמך של 98%.

ב. אמדו ברמת סמך של 90% את סטיית התקן של הגובה של מתגייסים של צה"ל.

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

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

ב. בנו רווח סמך לתוחלת המשכורת באזור A ברמת סמך של 95%.

8. להלן מדגם של שכר הדירה בש״ח של 5 דירות שלושה חדרים בשכונת בבלי בתל אביב :

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

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

פתרונות:

שאלה 1

א. $80.65 \le \mu \le 93.35$

ב. $13.5 \le \sigma \le 22.9$

ג. $0.225 \le p \le 0.575$

שאלה 2

א. $1.21 \le \mu \le 1.65$

ב. $10.85\% \le p \le 19.15\%$

שאלה 3

א. 70

ב. 99.88%

ג. $83\% \le p \le 89\%$

שאלה 4

א. $97.4 \le \mu \le 106.6$

ב. לא

ג. יגדל

שאלה 5

א. $0.5\% < p_1 - p_2 < 19.5\%$

ב. $0.704 \le p \le 0.768$

שאלה 6

א. $170.8 \le \mu \le 181.6$

ב. $8.8 \le \sigma \le 14.3$

<u>שאלה 7</u>

א. $-372 \le \mu_A - \mu_B \le 772$

ב. $6467 \le \mu \le 7133$

<u>שאלה 8</u>

$-21 \le \mu_{2013} - \mu_{2012} \le 821$

פרק 45 - בדיקת השערות כללית

רקע:

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

השערת האפס המסומנות ב- H_0

והשערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1.

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

למשל ,

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

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

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

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :
הכלל יוצר אזור שניקרא אזור דחייה (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) ואזור 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. נשנה את כלל ההכרעה אנחנו יכולים לקבל מסקנה אחרת . נבצע מדגם חדש אנחנו עלולים לקבל תוצאה אחרת.

לכן יתכנו טעויות במסקנות שלנו :

		הכרעה	
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

<u>הגדרת הטעויות:</u>

<u>טעות מסוג ראשון</u>- להכריע לדחות את H_0 למרות שבמציאות H_0 נכונה.

<u>טעות מסוג שני</u>- להכריע לקבל את H_0 למרות שבמציאות H_1 נכונה.

מה הן הטעויות האפשריות במחקר של התרופות? (בהקלטה)

<u>נגדיר את ההסתברויות הבאות:</u>

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}(לקבלH_0)

עוצמה:

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

<u>**דוגמה:**</u> (פתרון בהקלטה)

בכד יש 10 כדורים. יתכן ש- 5 מהם לבנים והיתר שחורים (כד א- השערת האפס) או ש- 7 מהם לבנים והיתר שחורים (כד ב- השערה אלטרנטיבית).

כדי להחליט איזה מהכדים ברשותנו, הוחלט להוציא כדור ולהשתמש בכלל ההחלטה הבא: אם הכדור שהוצא הוא לבן שזהו כד ב' (H_1).

א. חשבו את רמת המובהקות ואת רמת הביטחון של המבחן המוצע.

ב. חשבו את הסיכוי לטעות מסוג שני והעוצמה של המבחן המוצע.

תרגילים:

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

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

3. יהי X מספר שלם הנבחר באקראי מבין המספרים השלמים. הסיכוי ש- X יקבל ערך כלשהו נתון על ידי הנוסחה: $p(X=k)=\frac{1}{n}$ עבור $k=1,2,......,n$

   נתונות ההשערות הבאות לגבי התפלגות של X:

   $H_0: n=4$

   $H_1: n=6$

   כמו כן נתון כלל ההכרעה הבא: נדחה את השערת האפס אם X>3.

   חשבו את הסיכוי לטעות מסוג ראשון וטעות מסוג שני ואת העוצמה?

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

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

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

א. להלן כלל החלטה: אם מדובר במוצר שטיבו "טוב" נכריע שהמוצר בא ממפעל "שמשון", מהן ההסתברויות לסוגי הטעויות השונים?

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

ג. איזה כלל החלטה עדיף? נמק!

5. במטרה לבדוק האם מטבע תקין הטילו אותו 8 פעמים. הוחלט שאם מספר העצים יהיה בין 1 ל 7 כולל יוחלט שהמטבע תקין, אחרת נחליט שהמטבע מזויף.
   א. רשמו את השערות המחקר.
   ב. מה ההסתברות לטעות מסוג ראשון?
   ג. מהי עצמת המבחן אם במציאות אכן המטבע אינו תקין כי הסיכוי לעץ בו הוא 20%.

6. להלן השערות:

$H_0 : X \sim t(5)$ (התפלגות t עם 5 דרגות חופש)

$H_1 : X \sim Z$ (התפלגות נורמאלית סטנדרטית)

כלל החלטה: נדחה את השערת האפס אם X גדול מ-2.015.
   א. מהי רמת המובהקות של כלל ההחלטה?
   ב. מהי העוצמה של כלל ההחלטה?

7. במפעל מסוים נפלטים לאוויר חומרים רעילים. במצב שיגרה העוצמה הממוצעת של החומר הרעיל אמורה להיות 6,000 יחידות עם סטיית תקן 900. במצב חירום העוצמה הממוצעת היא 7,000 עם סטיית תקן 900. במפעל מערכת התראה נתמכת על ידי 9 חיישנים. אם ממוצע העוצמה של החומר הרעיל לפי תשעת החיישנים עולה על 6600 יחידות מופעלת מערכת ההתראה. נתון שעוצמת הזיהום מתפלגת נורמאלית.
   א. מה הסיכוי להתראת שווא? (באיזה סוג טעות מדובר)?
   ב. מה הסיכוי שבמצב חירום מערכת ההתראה לא תפעל? (באיזה סוג טעות מדובר)?
   ג. מה ההסתברות שאם המצב הוא מצב חירום מערכת ההתראה תפעל? (איך קוראים להסתברות זו)?
   ד. בסעיפים הבאים נשנה בכל סעיף נתון מסוים. כל סעיף עומד בפני עצמו, כיצד השינוי ישנה את הסיכוי לטעות מסוג ראשון ושני?
      1. המפעל יקנה עוד 4 חיישנים.
      2. מצב חרום מוגדר כעת בתוחלת של 7500 יחידות.
      3. מערכת ההתראה תופעל אם ממוצע של תשעת החיישנים יהיה מעל 6700.

8. במטרה לבדוק האם במקום עבודה מסוים פרופורציית הבנים נמוכה מפרופורציית הבנות נדגמו באקראי 10 עובדים. הוחלט שאם מספר הבנים במדגם יהיה לכל היותר 2 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות.
   א. מה רמת המובהקות של כלל ההכרעה הנ"ל ?
   ב. מהי העצמה בהנחה ובחברה 30% בנים?

9. זמן ההשפעה של משכך הכאבים "אופטלנוס" מתפלג נורמאלית עם תוחלת של 40 דקות וסטיית תקן של 12 דקות.
חברת התרופות המייצרת את התרופה מנסה לשפר את התרופה כך שתוחלת הזמן עד להשפעה תתקצר. לצורך כך, דגמו 25 מטופלים שיקבלו את התרופה "אופטלנוס פורטה", ממוצע זמן התגובה של המטופלים היה 34.5 דקות. חברת התרופות החליטה מראש שאם ממוצע הזמן עד להשפעה יהיה נמוך מ 35 דקות, היא תמשיך בתהליך שיווק "אופטלנוס פורטה".
א. מהי רמת המובהקות של המבחן המוצע?
ב. על סמך תוצאות המדגם. מהי המסקנה ומהי הטעות האפשרית במסקנה?
ג. מהי עצמת המבחן המוצע אם במציאות התרופה "אופטלנוס פורטה" מפחיתה את התוחלת לכדי 32דקות?
ד. כיצד תשתנה התשובה לסעיף ג' אם החברה הייתה מחליטה שהיא תמשיך בתהליך שיווק התרופה החדשה כאשר ממוצע המדגם יהיה נמוך מ-36 דקות?

10. ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 120.
מכון טוען שלימודים אצלו מעלים את ממוצע הציונים ביותר מ-30 נקודות. נלקחו 20 שלמדו במכון ו-20 שניגשו לבחינה בלמידה עצמית. הוחלט במשרד פרסום לקבל את טענת המכון רק אם במדגם ממוצע הציונים של אלה שלמדו במכון יהיה גבוהה בלפחות 50 נקודות מאלה שלא היו.
א. מהי רמת המובהקות של המחקר?
ב. מה הסיכוי לעשות טעות מסוג שני II בהנחה שהמכון מעלה את ממוצע הציונים ב-60 נקודות ?
ג. כיצד התשובות לסעיף א ו ב' היו משתנות אם מסתבר שסטיית התקן בציוני הפסיכומטרי הינה 100. הסבירו ללא חישוב.

11. קו ייצור נחשב תקין אם יש בו לכל היותר 4% פגומים , ונחשב שאינו תקין אחרת. מנהל האיכות דוגם בכל יום מקו הייצור 500 מוצרים. אם במדגם יהיה לפחות 30 מוצרים פגומים יפסיקו באותו היום את קו הייצור.
א. מה ההסתברות להפסיק את קו הייצור כשהוא תקין. איך קוראים להסתברות זאת?
ב. מה ההסתברות להמשיך ביום מסוים את קו הייצור למרות שאינו תקין כי היו 8% פגומים בקו הייצור. איך קוראים להסתברות זאת?

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

13. מספר המכוניות הנכנסות לחניון "עזרים" מתפלג פואסונית. בשנה שעברה המכוניות נכנסו לחניון בקצב של 2 מכוניות לדקה. בעקבות תלונות על עומס יתר בכניסה לחניון מעוניין מנהל החניון לבדוק האם קצב כניסת המכוניות לחניון גדל השנה. מנהל החניון החליט לספור את מספר המכוניות שיכנסו לחניון בדקה אקראית. אם מספר המכוניות שיספרו יהיה לפחות 4 יפתח מנהל החניון שער נוסף לחניון.
א. רשום את השערות מנהל החניון ואת כלל ההחלטה שלו. האם כלל ההכרעה הגיוני?
ב. מהי רמת המובהקות של כלל ההכרעה ?
ג. מהי העוצמה של כלל ההחלטה , אם כיום קצב כניסת המכוניות לחניון גדל ל- 4 מכוניות בדקה?

14. עודד עובד במפעל שבו מתחילים לעבוד בשעה 8:00. עודד בדרך כלל מאחר לעבודה והמנהל החליט לרשום את שעת בואו לעבודה. המנהל טוען שמשך האיחור של עודד (דקות), X, היא משתנה אחיד $U(0, 60)$. עודד טוען שהוא לא מגיע באיחור כה גדול, אלא שהתפלגות X היא בעלת התפלגות מעריכית עם תוחלת איחור של 20 דקות.
לבדיקת טענת המנהל (H_0) כנגד טענת עודד(H_1), המבוסס על משך האיחור של חגי ביום אחד.
מוצאים שני כללי הכרעה :
כלל 1 : דחה את השערת האפס אם משך האיחור יהיה לפחות 40 דקות.
כלל2 : דחה את השערת האפס אם משך האיחור יהיה לכל היותר 20 דקות.
חשב את הסיכוי לטעות מסוג ראשון ושני לכל אחת מכללי ההכרעה. מי עדיף?

פתרונות:

שאלה 2

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

שאלה 3

$\alpha = 0.25$ $\beta = 0.5$

שאלה 4

א. $\alpha = 0.2$ $\beta = 0.8$

ב. $\alpha = 0.2$ $\beta = 0.3$

ג. כלל ב'

שאלה 5

ב. 0.00781

ג. 0.1678

שאלה 6

א. 0.05

ב. 0.022

שאלה 7

א. 0.0228

ב. 0.0918

ג. 0.9082

שאלה 8

א. 0.055

ב. 0.383

שאלה 10

א. 0.2981

ב. 0.3974

ג. קטן

שאלה 11

א. 0.0113

ב. 0.0495

<u>שאלה 12</u>

חוקר א

<u>שאלה 13</u>

ב. 0.1428

ג. 0.566

פרק 46 - בדיקת השערות על פרמטרים

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 שניקרא **כלל הכרעה** :

הכלל יוצר אזור שניקרא **אזור דחייה** (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) **ואזור קבלה** (קבלה של השערת האפס ודחייה של האלטרנטיבה). כלל ההכרעה מתבסס על איזשהו סטטיסטי .

אזור הדחיה מוכתב על ידי סיכון שלוקח החוקר מראש שנקרא רמת מובהקות ומסומן ב- α.

שלב ה:

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

שלב ו :

להסיק מסקנה בהתאם לתוצאות המדגם.

דוגמה: (פתרון בהקלטה)

משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום 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:

$n = 20$

$\overline{X} = 3120$

$S = 280$

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

תרגילים:

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.

   א. מהי אוכלוסיית המחקר?

   ב. מה המשתנה הנחקר?

   ג. מה הפרמטר הנחקר?

   ד. מהן השערות המחקר?

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. מהי אוכלוסיית המחקר?

   ב. מה המשתנה הנחקר?

   ג. מה הפרמטר הנחקר?

   ד. מהן השערות המחקר?

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.

   א. מהי אוכלוסיית המחקר?

   ב. מה המשתנה הנחקר?

   ג. מה הפרמטר הנחקר?

   ד. מהן השערות המחקר?

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.

   א. מהי אוכלוסיית המחקר?

   ב. מה המשתנה הנחקר?

   ג. מה הפרמטר הנחקר?

   ד. מהן השערות המחקר?

טעויות בבדיקת השערות

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא **כלל הכרעה** :

הכלל יוצר אזור שניקרא **אזור דחייה** (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) **ואזור 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. נשנה את כלל ההכרעה אנחנו יכולים לקבל מסקנה אחרת . נבצע מדגם חדש אנחנו עלולים לקבל תוצאה אחרת.

לכן יתכנו טעויות במסקנות שלנו :

	הכרעה		
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	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?

תרגילים:

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?

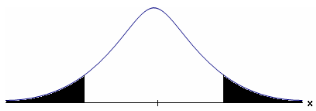
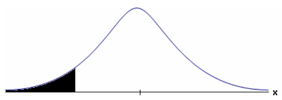
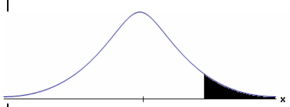
2. במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
   א. האם ניתן לדעת אם בוצע טעות במחקר?
   ב. מה סוג הטעות האפשרית?

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?

פרק 47 - בדיקת השערות על תוחלת (ממוצע)

כאשר שונות האוכלוסיה ידועה

רקע:

השערת האפס : **השערת 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 :	$Z_{\bar{x}} < -Z_{1-\frac{\alpha}{2}}$ 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 :

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

חלופה אחרת לכלל הכרעה :

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

<u>דוגמה :</u> (פתרון בהקלטה)

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. בדוק את ההשערה ברמת מובהקות של 1%.

<u>תרגילים:</u>

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 תהייה 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? בחר בתשובה הנכונה.
   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.

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: (בחר בתשובה הנכונה)

   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.

פתרונות :

שאלה 1:

נקבל H_0

שאלה 2:

נדחה H_0

שאלה 3:

נדחה H_0

שאלה 4:

נדחה H_0

שאלה 5:

נקבל H_0

שאלה 6:

א

שאלה 7:

ג

שאלה 8:

א. אותה מסקנה

ב. לא ניתן לדעת.

סיכוי לטעויות ועוצמה כאשר שונות האוכלוסייה ידועה

רקע:

	הכרעה		
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות)

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

הסיכוי לבצע טעות מסוג 2:

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

רמת בטחון:

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

עוצמה:

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

התהליך לחישוב סיכוי לטעות מסוג שני:

השערת האפס : **השערת 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:	3. σ ידועה 4. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$\bar{X} > \mu_0 + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$
חישוב β :	$P_{H_1}(\mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}})$	$P_{H_1}(\bar{X} > \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}})$	$P_{H_1}(\bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}})$

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

התקנון : $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

דוגמה : (פתרון בהקלטה)

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'?

תרגילים:

1. נתון ש $X \sim N(\mu,\sigma^2=1)$

להלן השערות של חוקר לגבי הפרמטר μ :

$$H_0 : \mu = 5$$
$$H_1 : \mu = 7$$

מעוניינים ליצור כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.

א. עבור אילו ערכים של X שידגם נדחית השערת H_0 ?

ב. מה הסיכוי לבצע טעות מסוג שני?

ג. אם במדגם התקבל ש $X = 6.9$ מה תהיה המסקנה ומה הטעות האפשרית?

2. לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. מעוניינים לבדוק אם כיום ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה.

א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי ברמת מובהקות של 5%.

ב. בהמשך לסעיף א מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?

ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העצמה של הכלל מסעיף א?

3. להלן נתונים על תהליך של בדיקת השערות על תוחלת :

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

א. רשום כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.

ב. בהמשך לסעיף א מהי העצמה אם התוחלת שווה ל- 195?

ג. הסבר ללא חישוב איך העצמה תשתנה אם רמת המובהקות תהייה 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. רשום 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. הסבר ללא חישוב כיצד התשובה לסעיף ב תשתנה אם רמת המובהקות תגדל.

ד. הסבר ללא חישוב כיצד התשובה לסעיף ב תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ״מ.

5. להלן השערות של מחקר

$$H_0 : \mu = 50$$
$$H_1 : \mu = 58$$

מעוניינים לדגום 100 תצפיות . ידוע שסטיית התקן של ההתפלגות הינה 20.

א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10% . מהי רמת המובהקות?

ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו) ?

1. סטיית התקן הייתה יותר גדולה .
2. הסיכוי לטעות מסוג שני גדול יותר.

<u>השאלות שלהלן הן שאלות רב ברירתיות. בחר בכל שאלה את התשובה הנכונה ביותר:</u>

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.

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.

פתרונות :

שאלה 1:

א. מעל 6.645

ב. 0.3632

שאלה 2:

א. נדחה H_0 אם $\bar{X} < 2.24$

ב. נדחה H_0

ג. 1

שאלה 3:

א. נדחה H_0 אם $\bar{X} > 203.29$ או $\bar{X} < 196.71$

ב. 0.8051

ג. תקטן.

שאלה 4:

א. נדחה H_0 אם $\bar{X} < 48.9$

ב. 0.0885

ג. תקטן.

ד. תקטן.

שאלה 6:

ד

שאלה 7:

ג

שאלה 8:

ג

שאלה 9:

א

שאלה 10:

ב

קביעת גודל מדגם כששונות האוכלוסיה ידועה

רקע:

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

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

הנוסחה הבאה נותנת את גודל המדגם הרצוי :

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

דוגמה : (פתרון בהקלטה)

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?

תרגילים:

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.

הוכח שגוגל המדגם הרצוי לכך יהיה:

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

פתרונות :

שאלה 1:

41

שאלה 2 :

78

שאלה 3:

הוכחה

מובהקות התוצאה (p-value) בבדיקת השערות על תוחלת עם שונות 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.

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

אם ההשערה היא דו צדדית :

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס : **השערה אלטרנטיבית:**	$H_0: \mu = \mu_0$ $H_1: \mu \ne \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
תנאים:	5. σ ידועה 6. $X \sim N$ או מדגם מספיק גדול		
p-value	אם $\bar{x} > \mu_0 \Leftarrow 2 \cdot P_{H_0}(\bar{X} \ge \bar{x})$ אם $\bar{x} < \mu_0 \Leftarrow 2 \cdot P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \ge \bar{x})$

כאשר בהנחת השערת האפס : $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$

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

<u>דוגמה:</u> (פתרון בהקלטה)

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%?

תרגילים:

1. לפניך השערות של מחקר :

$$H_0 : \mu = 70$$
$$H_1 : \mu > 70$$

המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות:

$$n = 100$$
$$\bar{x} = 74$$

מהי מובהקות התוצאה?

2. השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪ . מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?

3. אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ – 100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטית 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 מאוד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון ? לא נכון? נמק.

6. 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.

7. מובהקות התוצאה (PV) היא גם : (בחר בתשובה הנכונה)
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. בבדיקת השערות מסוימת התקבל p value=0.0254 לכן (בחר 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.

פתרונות :

שאלה 1:

0.0228

שאלה 2:

עבור כל רמת מובהקות סבירה.

שאלה 3:

ב. 0.1056

ג. 0.1056

ד. נכריע שיש עמידה בהתחייבות של החברה.

שאלה 4:

א. 0.0006

ב. יקטן.

ג. נכריע שאין כיול.

שאלה 5:

נכון

שאלה 6:

תשובה :א

שאלה 7:

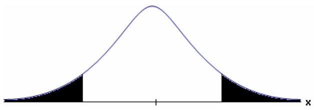
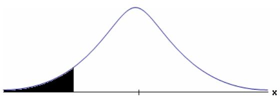
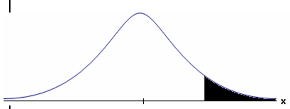
תשובה : א

שאלה 8:

תשובה : ג

בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה

רקע:

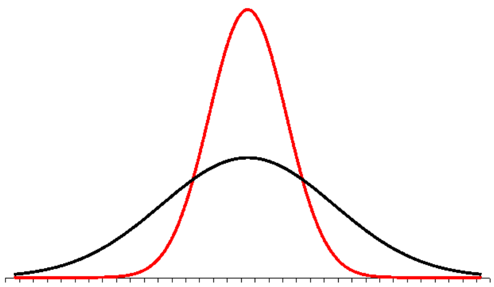
השערת האפס : **השערה אלטרנטיבה:**	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$
תנאים:	7. σ אינה ידועה 8. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$t_{\bar{x}} > t^{(n-1)}_{1-\frac{\alpha}{2}}$ או $t_{\bar{x}} < -t^{(n-1)}_{1-\frac{\alpha}{2}}$  $-t_{1-\frac{\alpha}{2},n-1}$ $t_{1-\frac{\alpha}{2},n-1}$ - דוחים את H_0	$t_{\bar{x}} < -t^{(n-1)}_{1-\alpha}$  $-t_{1-\alpha,n-1}$ - דוחים את H_0	$t_{\bar{x}} > t^{(n-1)}_{1-\alpha}$  $t_{1-\alpha,n-1}$ - דוחים את H_0
חלופה לכלל הכרעה : נדחה H_0 אם מתקיים:	$\bar{X} > \mu_0 + t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ א ו $\bar{X} < \mu_0 - t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 - t^{n-1}_{1-\alpha} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t^{n-1}_{1-\alpha} \cdot \frac{S}{\sqrt{n}}$

סטטיסטי המבחן :

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

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

<u>התפלגות T:</u>



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.

<u>דוגמה:</u> (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%.

תרגילים:

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

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

$$n = 20$$
$$\bar{x} = 3120$$
$$S = 280$$

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

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

4. באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות :

$$\sum_{i=1}^{10} X_i = 750$$
$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$

נתון שההתפלגות היא נורמלית.
בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80 .

5. ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם.

ליאור השתמש בטבלה של התפלגות Z.

רוני השתמשה בטבלה של התפלגות t.

מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.

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:

$$H_0: \mu = \mu_0$$
$$H_0: \mu < \mu_0$$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות.

בחר בתשובה הנכונה:

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

<u>פתרונות:</u>

<u>שאלה 1:</u>

נדחה H_0

<u>שאלה 2:</u>

נדחה H_0

<u>שאלה 3:</u>

נקבל H_0

<u>שאלה 4:</u>

נקבל H_0

<u>שאלה 5:</u>

התשובה היא : ב

<u>שאלה 6:</u>

התשובה היא : ג

מובהקות התוצאה (p-value) כאשר שונות האוכלוסייה לא ידועה

רקע:

נזכיר שהמסקנה 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.

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

אם ההשערה היא דו צדדית:

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

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:	9. σ אינה ידועה 10. $X \sim N$ או מדגם מספיק גדול		
p-value	אם $\bar{x} > \mu_0 \Leftarrow 2 \cdot P_{H_0}(\bar{X} \ge \bar{x})$ אם $\bar{x} < \mu_0 \Leftarrow 2 \cdot P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \ge \bar{x})$

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

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

$$d.f = n-1$$

<u>דוגמה :</u> (פתרון בהקלטה)

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 : 27,34,32,40,30 . הנח שזמן הנסיעה מתפלג נורמלית.

א. רשום את השערות המחקר.

ב. מצא חסמים למובהקות התוצאה.

ג. מה המסקנה ברמת מובהקות של 5% ?

תרגילים :

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 :

1008,1024,996,1005,997

א. רשמו את השערות המחקר.

ב. מהי מובהקות התוצאה? הצג חסמים.

ג. מה המסקנה ברמת מובהקות של 5%?

2. חוקר בדק את הטענה כי פועלים העובדים במשמרת לילה איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. במדגם מיקרי של 25 פועלים שעבדו במשמרת לילה נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטית תקן של 3 שעות.

מהי ה- α המינימלית שלפיה ניתן להחליט שאכן העובדים במשמרת לילה איטיים יותר ?

3. הגובה של מתגייסים לצה״ל מתפלג נורמלית . במדגם של 25 מתגייסים מדדו את הגבהים שלהם בס״מ והתקבלו התוצאות הבאות :

$$\bar{x} = 176.2$$

$$\sum (x_i - \bar{x})^2 = 2832$$

מטרת המחקר היא לבדוק האם תוחלת הגבהים של המתגייסים גבוה מ-174 ס״מ באופן מובהק. מהי בקרוב מובהקות התוצאה ועל פיה מה תהיה המסקנה ברמת מובהקות של 6% ?

פתרונות :

שאלה 3:

נקבל H_0

הקשר בין רווח סמך לבדיקת השערות על תוחלת

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: (פתרון בהקלטה)

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?

<u>תרגילים :</u>

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 10.4מ״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? הסבר.

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 של 5 מ״ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.
א. בנה רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.
ב. בדוק ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

<u>פתרונות :</u>

<u>שאלה 1:</u>

1. נקבל השערת H_0

<u>שאלה 2:</u>

א. $112.87 \le \mu \le 118.13$

ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.

<u>שאלה 3:</u>

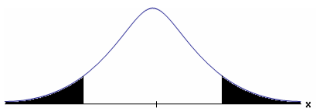
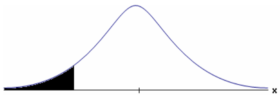
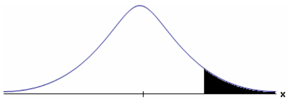
א. $191.8 \le \mu \le 200.2$

ב. נכריע שיש אמת בפרסום.

פרק 48 - בדיקת השערות על פרופורציה

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 :	$Z_{\hat{p}} < -Z_{1-\frac{\alpha}{2}}$ או $Z_{\hat{p}} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $\quad Z_{1-\frac{\alpha}{2}}$ ■ - דוחים את H_0	$Z_{\hat{p}} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ ■ - דוחים את H_0	$Z_{\hat{p}} > Z_{1-\alpha}$  $Z_{1-\alpha}$ ■ - דוחים את H_0

סטטיסטי המבחן :

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

חלופה אחרת לכלל הכרעה:

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

<u>דוגמה:</u> (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.

תרגילים:

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. בדוק 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.

פתרונות :

שאלה 1:

נדחה H_0

שאלה 2:

נדחה H_0

שאלה 3:

נקבל H_0

שאלה 4:

ב. נקבל H_0

ג. המסקנה לא תשתנה.

שאלה 5:

א. נדחה H_0

ב. המסקנה לא תשתנה.

שאלה 6:

התשובה היא : ג.

סיכוי לטעויות ועוצמה

רקע:

	הכרעה	
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

α=P(H_0 לדחות את | נכונה H_0)= P_{H_0}(H_0 לדחות)

הסיכוי לבצע טעות מסוג 2:

β =P(H_0 לקבל את | נכונה H_1)=P_{H_1}(H_0 לקבל)

רמת בטחון:

(1-α) =P(H_0 לקבל את | נכונה H_0)= P_{H_0}(H_0לקבל)

עוצמה :

π=(1-β) =P(H_0 לדחות את | נכונה H_1) =P_{H_1}(H_0 לדחות)

התהליך לחישוב סיכוי לטעות מסוג שני:

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

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

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.

תרגילים:

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.
   א. 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 (בחר בתשובה הנכונה)
   א. השערת האפס נכונה.
   ב. השערת האפס נדחתה.
   ג. השערת האפס לא נדחתה.
   ד. אף אחת מהתושבות לא נכונה בהכרח.

6. קבע אם הטענה הבאה נכונה:
"בבדיקת השערות לא ניתן לבצע בו זמנית טעות מסוג ראשון וטעות מסוג שני"

פתרונות:

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

שאלה 1:

ב. 0.9015

ג. תגדל

ד. טענה לא נכונה.

שאלה 2:

0.4404

שאלה 3:

א. 0.1446

ב. תקטן.

שאלה 4:

חוקר א.

שאלה 5:

התשובה הנכונה היא ג.

שאלה 6:

נכונה.

קביעת גודל מדגם

רקע:

השערות המחקר הן :
$$H_0 : p = p_0$$
$$H_1 : p = p_1$$

מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β.

הנוסחה הבאה נותנת את גודל המדגם הרצוי :

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

דוגמה: (פתרון בהקלטה)

רוצים לבדוק האם אחוז האנשים השוהים בשמש ללא הגנה ירד בעקבות הפרסומת על נזקי השמש. בעבר 60% מהאוכלוסייה שהתה בשמש ללא הגנה. מה גודל המדגם המינימלי שיש לקחת כדי לבדוק שהאחוז הנ״ל ירד ל48% אם מעוניינים שהסיכוי לטעות מסוג ראשון יהיה 5% והסיכוי לטעות מסוג שני יהיה 1%?

<u>תרגילים:</u>

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 מביניהם צודק, מה צריך להיות גודל המדגם שיענה על שני התנאים הבאים:
   - אם משרד התמ"ת צודק, נדחה את טענתו בסיכוי של 10%.
   - 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%.

פתרונות:

שאלה 1:

891

שאלה 2:

224

מובהקות התוצאה

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.

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

אם ההשערה היא דו צדדית:

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

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 $\hat{p} > p_0 \Leftarrow 2 \cdot P_{H_0}(\hat{P} \ge \hat{p})$ אם $\hat{p} < p_0 \Leftarrow 2 \cdot P_{H_0}(\hat{P} \le \hat{p})$	$P_{H_0}(\hat{P} \le \hat{p})$	$P_{H_0}(\hat{P} \ge \hat{p})$

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

והתקנון:

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

דוגמה: (פתרון בהקלטה)

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

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

תרגילים:

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 מ60% 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-value=0.02* .
   מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 5% : (בחר בתשובה הנכונה)
   א. יקבל את השערת האפס
   ב. ידחה את השערת האפס.
   ג. לא ניתן לדעת כי אין מספיק נתונים.

5. קבע 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?

פתרונות :

שאלה 1:

א. 0.0455

שאלה 2:

א. 0.0548

ב.0.0548

ג. מעל 0.0548

ד. נכריע לטובת טענת המחקר.

שאלה 3:

$p_v = 0$

שאלה 4:

התשובה הנכונה : ב

שאלה 5:

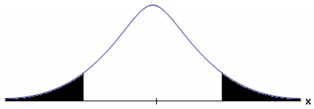
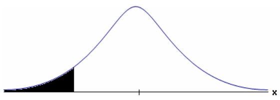
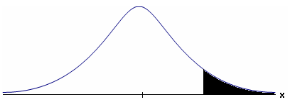
הטענה נכונה

שאלה 6:

0.7486

פרק 49 - בדיקת השערות על הפרש פרופורציות

רקע:

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

סטטיסטי המבחן :

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

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

חלופה אחרת לכלל הכרעה:

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

התפלגות של $\hat{p}_1 - \hat{p}_2$:

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

תקנון:

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

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

דוגמה : (פתרון בהקלטה)

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

תרגילים:

1. במדגם של 200 גברים. 8% מהם היו מובטלים. המדגם של 180 נשים 10% מהן היו מובטלות האם קיים הבדל מובהק בין פרופורציית המובטלים לפרופורציית המובטלות. בדוק ברמת מובהקות של 5%.

2. אחוז בעלי רישיון נהיגה בקרב האוכלוסייה הבוגרת הינו 60%. במדגם של 300 בוגרים מתל אביב 204 היו בעלי רישיון נהיגה. במדגם של 220 בוגרים מירושלים 100 היו בעלי רישיון נהיגה.
   א. ברמת מובהקות של 5% האם תקבלו את הטענה שאחוז בעלי הרישיון בתל אביב גבוה מהאחוז הארצי?
   ב. ברמת מובהקות של 10% האם תקבלו את הטענה שאחוז בעלי הרישיון נהיגה בתל אביב גבוה מאחוז בעלי רישיון הנהיגה בירושלים?

3. נדגמו 500 בוגרים מתוכם 200 גברים והיתר נשים. במדגם התקבל : מתוך הגברים ל-48% תעודת בגרות. מתוך הנשים ל-58% תעודת בגרות. מטרת המחקר היא לבדוק האם שיעור הזכאיות לבגרות גבוה משיעור הזכאים.
   א. מהי מובהקות התוצאה?
   ב. מה תהיה המסקנה ברמת מובהקות של 8%?

4. במדגם שנערך על 100 פרות מחוות בדרום הארץ התקבל כי 20 פרות נושאות וירוס מסוים. במדגם שנערך על 200 פרות מחוות בצפון הארץ התקבל כי 10 מתוכן נושאות וירוס גם כן.
   א. בנו מבחן ברמת מובהקות של 5% לבדיקת הטענה כי הווירוס תקף את פרות הדרום באופן משמעותי יותר מאשר את הפרות בצפון הארץ.
   ב. מהי המסקנה לבדיקת הטענה של סעיף א ומהי הטעות האפשרית במסקנה?
   ג. מהי עוצמת המבחן אם שיעור הפרות בדרום עם הווירוס גבוה ב10% משיעור הפרות בצפון עם
   הווירוס?
   ד. כיצד העוצמה תשתנה אם נגדיל את רמת המובהקות?

<u>פתרונות :</u>

<u>שאלה 1:</u>

לא נדחה את H_0

<u>שאלה 2:</u>

א. נדחה H_0

ב. נדחה H_0

<u>שאלה 3:</u>

א. 0.0139

ב. נדחה H_0

<u>שאלה 4:</u>

ב. נדחה H_0

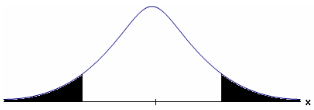
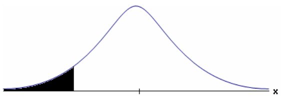
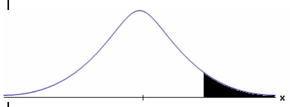
ג. 0.8238

ד. תגדל

פרק 50 - 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:	1. מדגמים בלתי תלויים 2. $\sigma_1,\ \ \sigma_2$ ידועות 3. $X_1,X_2\sim N$ 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 :

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

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 :

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

<u>דוגמה</u> : (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%?

<u>תרגילים :</u>

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 אלו 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?

פתרונות:

שאלה 1:

נדחה H_0

שאלה 2:

לא נדחה את H_0

שאלה 3:

א. 0

ב. נדחה H_0

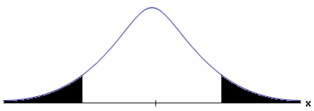
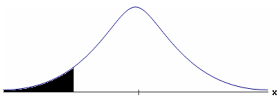
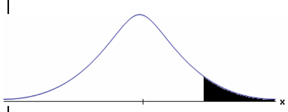
שאלה 4:

א. נדחה H_0 אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותר מ-10.72 ס״מ.

ב. 0.6331

<u>כששונויות האוכלוסיה לא ידועות ומניחים שהן שוות</u>

<u>רקע:</u>

השערת האפס : **השערת 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:	4. מדגמים בלתי תלויים 5. $\sigma_1, \quad \sigma_2$ לא ידועות אך שווֹת 6. 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 $t_{\bar{x}_1-\bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  $-t_{1-\frac{\alpha}{2},n_1+n_2-2}$ $\quad$ $t_{1-\frac{\alpha}{2},n_1+n_2-2}$ ■ - 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 : $$S_p{}^2=\frac{(n_1-1)S_1{}^2+(n_2-1)S_2{}^2}{n_1+n_2-2}$$

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** $\bar{x}_1-\bar{x}_2<c-t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$	$\bar{x}_1-\bar{x}_2<c-t_{1-\alpha}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$	$\bar{x}_1-\bar{x}_2>c+t_{1-\alpha}^{(n_1+n_2-2)}\cdot\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$

<u>דוגמה :</u> (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.

<u>להלן תוצאות המדגם:</u>

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%.

תרגילים:

1. להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (מטרים רבועים):

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	2-60W	1-100W
$\bar{x}$	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 60דולקות בממוצע יותר מאשר נורות מסוג W 100?

ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W 60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.

פתרונות :

שאלה 1:

לא נדחה H_0

שאלה 2:

שאלה 3:

א. נדחה H_0

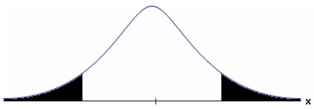
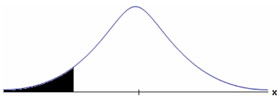
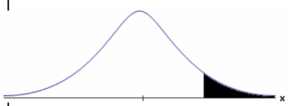
ב. רמות מובהקות של לפחות 5%

ג. לא נדחה H_0

פרק 51 - 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:	7. σ_D אינה ידועה 8. $D \sim N$ 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 את 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 :

$$t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

$$S_D{}^2 = \frac{\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} D^2{}_i - n\bar{D}^2}{n-1}$$

<u>דוגמה</u> : (פתרון בהקלטה)

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה נמוכים מהמחירים מרשת השיווק "שופרסל".

לצורך בדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות.

להלן המחירים :

המוצר	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".

תרגילים:

1. במטרה לבדוק האם קיים הבדל בין חברת X לחברת Y מבחינת המחירים לשיחות בינ״ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה.
להלן התוצאות:

המדינה	X	Y
ארה״ב	1.5	1.4
קנדה	2.1	2
הולנד	2.2	1.9
פולין	3	3.1
מצרים	3.5	3.2
סין	3.2	3.2
יפן	4.2	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 מתפלג נורמלית.

3. נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן הציונים שלהם בסמסטר א' ו-ב':

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

פורסם שתלמידים שמסיימים את סמסטר ב משפרים בממוצע את הציונים ב-5 נקודות לעומת סמסטר א' . הנח 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.

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):

5.2	7.0	6.9	5.4	משקל במכשיר 1
5.0	7.1	6.9	5.3	משקל במכשיר 2

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 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.

פתרונות:

שאלה 1:

לא נדחה H_0

שאלה 2:

לא נדחה H_0

שאלה 3:

א. $0.25 \le pv \le 0.5$

ב. 0.5

ג. לא נדחה H_0

שאלה 4:

התשובה היא ד.

שאלה 5:

התשובה היא ד.

שאלה 6:

התשובה היא ג.

פרק 52 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע:

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α על $\mu_1-\mu_2$:

$$H_0:\mu_1-\mu_2=C$$
$$H_1:\mu_1-\mu_2\neq C$$

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 :

$$H_0:\mu_D=80$$
$$H_1:\mu_D\neq 80$$
$$\alpha=5\%$$

החוקר בנה רווח סמך ברמה של 90%

$$78<\mu_D<83$$

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

תרגילים:

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

סטטיסטיקה א	סטטיסטיקה 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 אמריקאיות:

3. סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים.
הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח $-2 < \mu_1 - \mu_2 < 4.5$.
אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות:
$H_0 : \mu_1 - \mu_2 = 0 \ ; \ H_1 : \mu_1 - \mu_2 \neq 0$, ברמת מובהקות 0.05 מסקנתו תהיה:
א. לדחות את השערת האפס.
ב. לא לדחות את השערת האפס.
ג. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05 .
ד. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

4. במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה

והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו :

$-0.0293 < \mu_D < 0.2145$ רווח הסמך הוא ברמת סמך של 95%.

לכן מסקנת המחקר היא :

א. ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.

ב. ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.

ג. לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D.

פתרונות:

שאלה 1:

א. $-3.8 \le \mu_D \le 19$

ב. נכריע שיש אמת בפרסום.

שאלה 2:

א. $-8.5 \le \mu_X - \mu_Y \le 26.5$

ב. נכריע שאין הבדל.

שאלה 3:

התשובה היא ג.

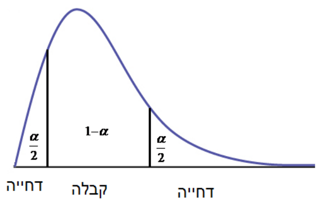
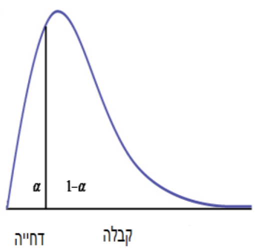
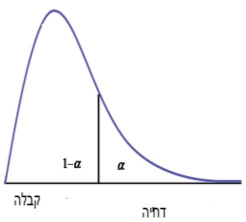
שאלה 4

התשובה היא א.

פרק 53- בדיקת השערות על שונויות

בדיקת השערות על שונות האוכלוסייה כאשר התוחלת לא ידועה

רקע:

השערת האפס : **השערת אלטרנטיבית:**	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$
תנאים :	$X \sim N$		
נדחה את השערת האפס אם:	 או $\chi^2 < \chi^{2(n-1)}_{\frac{\alpha}{2}}$ $\chi^2 > \chi^{2(n-1)}_{1-\frac{\alpha}{2}}$	 $\chi^2 < \chi^{2(n-1)}_{\alpha}$	 $\chi^2 > \chi^{2(n-1)}_{1-\alpha}$

סטטיסטי המבחן : $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

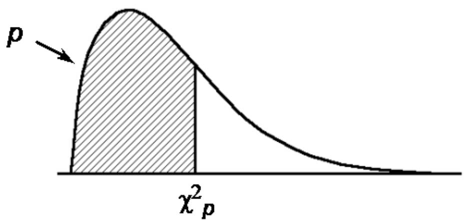
התפלגות חי בריבוע :

אם דדן $X_i \sim N(\mu,\sigma^2)$ והפרמטר μ אינו ידוע מתקיים ש : $\dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^{2(n-1)}$

התפלגות זו היא התפלגות אסימטרית חיובית המתחילה מהערך אפס וערכיה שואפים לאינסוף.

התפלגות זו תלויה בדרגות החופש. אם μ אינו ידוע אז :

$d.f = n-1$

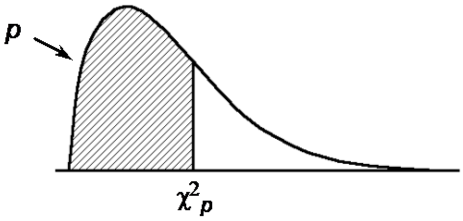


דוגמה : (פתרון בהקלטה)

ציוני IQ לפי סטנדרטים אמריקאים מתפלגים נורמאלית עם $\sigma=15$. מעוניינים לבדוק האם שונות הציונים של נבחנים ישראלים שונה מאמריקה. במדגם של 20 ישראלים התקבל :

$\sum_{i=1}^{20}(x_i-\bar{x})^2=3420$.מה המסקנה ברמת מובהקות של 5%?

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p



df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
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

תרגילים:

1. חברה אורזת סוכר במשקל עם סטיית תקן 20 גרם. משקל הסוכר באריזה מתפלג נורמאלית. החברה החליפה את מכונות האריזה במטרה לדייק יותר במשקל הנארז. (רוצים שסטיית התקן תהיה קטנה יותר).
לצורך בדיקה דגמו 5 אריזות סוכר ולהלן משקלן בגרמים : 1008,1024,996,1005,997.
מה המסקנה ברמת מובהקות של 5%?

2. זמן ההחלמה ממחלה מסוימת כאשר משתמשים בטיפול מסוים מתפלג נורמלית עם סטיית תקן של 80 שעות. תרופה חדשה נוסתה על 5 חולים. זמני ההחלמה שלהם בשעות היו : 38,72,90,110,50 .
א. ברמת מובהקות של 5% בדקו האם סטיית התקן של זמן החלמה של התרופה החדשה נמוכה מהתרופה המקורית.
ב. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נגדיל את רמת המובהקות ?
ג. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נקטין את רמת המובהקות?
ד. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נוסיף תצפית שערכה 70 ?

3. הגובה של אוכלוסייה מסוימת נחשב כמתפלג נורמלית עם ממוצע של 174 ס"מ וסטיית תקן 12. במדגם של 20 אנשים מהאוכלוסייה התקבל ממוצע 171 וסטיית תקן מדגמית 23.
א. בדקו ברמת מובהקות של 5% האם חל שינוי בשונות הגבהים באוכלוסייה.
ב. בדקו ברמת מובהקות של 5% האם חל שינוי בתוחלת הגבהים באוכלוסייה , בבחירת המבחן המתאים הסתמך על המסקנה מסעיף א'.

4. השערות המחקר הן : $H_0:\sigma^2=100$

$$H_1:\sigma^2>100$$

מתכננים לבצע מדגם בגודל 10 תצפיות. רמת המובהקות היא 5%.
א. מה תהה עוצמת המבחן אם $\sigma_1^2=150$?
ב. איזו השערה אלטרנטיבית תיתן עוצמה של 90%?

5. השערות המחקר הן : $H_0:\sigma=2$

$$H_1:\sigma<2$$

במדגם של 21 תצפיות התקבל סטיית תקן 1.143. תן הערכה למובהקות התוצאה.

פתרונות:

שאלה 1

לא נדחה H_0

שאלה 2

א. נדחה H_0

ב. לא תשתנה

ג. לא ניתן לדעת

ד. לא תשתנה

שאלה 3

א. נדחה H_0

ב. לא נדחה H_0

שאלה 4

א. בין 25% ל- 50%

ב. 405.3

שאלה 5

$$0 < P_V < 0.005$$

פרק 54- שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות מסכמות בבדיקת השערות על פרמטרים

1. שני חוקרים נתבקשו לבדוק את ההשערות הבאות:

$$H_0: \mu = 520$$
$$H_1: \mu > 520$$

כל חוקר בדק מדגם של 225 נחקרים . ידוע ש- $\sigma = 20$.

חוקר א' קבע את כלל ההכרעה לפי $\alpha = 0.05$.

חוקר ב' מחליט לדחות H_0 אם $\bar{X} > 522$.

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 ה"לייק"ים היומי של דנה היא 12 עם סטיית תקן 5.
דני טוען שהוא יותר פופולארי מדנה בכך שהוא מקבל יותר "לייק"ים מדנה ביום. על-מנת לבדוק זאת ספר דני כמה "לייק"ים הוא קיבל בכל יום במהלך 7 שבועות (כלומר, ב – 49 ימים) וקיבל סך-הכול 637 "לייק"ים. נניח כי סטיית התקן של מספר ה"לייק"ים שדני מקבל ביום זהה לסטיית התקן של דנה.

א. מהי רמת המובהקות שכדאי לדני לדרוש, כדי שדנה תשתכנע בצדקת טענתו (שדני פופולארי יותר בכך שהוא מקבל יותר "לייק"ים 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	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.

4. במדגם של 10 ישראלים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות :

$$n = 10$$
$$\sum X_i = 1020$$
$$\sum X_i^2 = 105120$$

במדגם של 14 אמריקאים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות :

$$n = 14$$
$$\sum X_i = 1386$$
$$\sum X_i^2 = 138644$$

נתון שציוני הבחינה מתפלגים נורמלית בכל מדינה.

א. בדוק ברמת מובהקות של 10% האם קיים שוויון שונויות בין אוכלוסיית אמריקה לאוכלוסיית ישראל?
ב. בדקו האם קיים הבדל בממוצע הציונים בבחינת ה-IQ בין ישראל לארה״ב. ברמת מובהקות של 5%?

5. 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. נסח את ההשערות ובדוק אותן ברמת מובהקות של 5%. רשום את כלל ההכרעה ואת ההנחות הדרושות לביצוע המבחן הפרמטרי.

ב. חשב את P-value .

ג. ישנה טענה שממוצע זמן ההשקעה בלימודים במכללות הוא 3.5 שעות ביום. בדוק את הטענה כאשר רמת המובהקות הינה 5%.

6. במדינת טרפפו המשכורות במשק מתפלגות נורמלית עם ממוצע של 1 אלף דולר וסטיית תקן של 0.2 אלף דולר.

בוצע מדגם מקרי בו השתתפו 5 נשים ו 5 גברים במדינת שומקום שבה המשכורות מתפלגות נורמאלית גם כן. להלן משכורותיהם באלפי דולר:

גברים	2	0.9	0.7	1.2	1.1
נשים	1.4	1.1	1.9	1.8	1.2

א. בדוק את הטענה שממוצע משכורותיהם של אזרחי שומקום גבוה מאשר ממוצע משכורותיהם של אזרחי טרפפו ברמת מובהקות של 5%. בהנחה שסטיית התקן זהה בשתי המדינות.

ב. חזור על הסעיף הקודם ללא ההנחה הנ"ל.

ג. ישנה טענה שסטיית התקן במדינת שומקום גבוהה מזו של טרפפו. בדוק ברמת מובהקות של 5%.

7. במטרה להשוות בין אחוזי הצפייה של גברים ונשים בתוכנית טלוויזיה מסוימת בוצע סקר ובו התקבלו תוצאות הבאות:

	צופים	לא צופים
נשים	320	42
גברים	72	120

א. האם יש הבדל בין אחוזי הצפייה של גברים ונשים ברמת מובהקות של 1%?

ב. עבור רמת מובהקות של 5% בדוק טענה שמבין הצופים בתוכנית הטלוויזיה אחוז הנשים גדול פי 2 מאחוז הגברים.

8. 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.

ג. מהי הטעות האפשרית במסקנה מהסעיף הקודם. האם ניתן לדעת את הסתברותה?

9. 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. בדוק 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%.

10. להלן התפלגות מספר הנסיעות לחופשה השנתית במדגם של משפחות ישראליות :

מספר הנסיעות	0	1	2	3	4
מספר המשפחות	84	102	26	20	12

בדוק ברמת מובהקות של 5% :

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?

11. נתון כי :

$$X \sim N(\mu, \sigma^2 = 10^2)$$

מעוניינים לבדוק את ההשערות :

$$H_0 : \mu = 40$$
$$H_1 : \mu > 40$$

דגמו 25 תצפיות מהאוכלוסייה והתקבל $\overline{X} = 45$.

א. חשבו את P-value (מובהקות התוצאה).

ב. חזור על סעיף א אם ההשערה האלטרנטיבית הייתה :

$$H_1 : \mu < 40$$

ג. חזור על סעיף א אם ההשערה האלטרנטיבית הייתה :

$$H_1 : \mu \neq 40$$

12. ציוני בחינת הבגרות במתמטיקה מתפלגים נורמלית עם שונות 150. במדגם של 16 נבחנים מתל אביב התקבלה שונות מדגמית- 190. במדגם של 25 ירושלמים התקבלה שונות מדגמית 118.

א. בדקו ברמת מובהקות של 2.5% האם שונות הציונים במתמטיקה בקרב נבחני תל אביב גבוהה מהשונות בכלל הארץ.

ב. בדקו ברמת מובהקות של 5% האם שונות ציונים במתמטיקה בקרב תלמידי תל אביב גבוהה מאשר בקרב תלמידי ירושלים.

פתרונות:

שאלה 1:

א. חוקר א

ב. 0.0122

ג. גדלה

ד. נדחה H_0

שאלה 2:

א. לפחות 0.0808

ב. 0.7995

שאלה 3:

א. לא נדחה H_0

ב. לא נדחה H_0

שאלה 4:

א. לא נדחה H_0

ב. לא נדחה H_0

שאלה 5:

א. לא נדחה H_0

ב. בין 5% ל- 10%

ג. נדחה H_0

שאלה 6:

א. נדחה H_0

ב. נדחה H_0

ג. נדחה H_0

שאלה 7:

א. נדחה H_0

ב. נדחה H_0

שאלה 8:

ב. נדחה H_0

ג. טעות מסוג ראשון בסיכוי של 0.05

שאלה 9:

א. לא נדחה H_0

ב. 0.8643

ג. 0.8749

שאלה 10:

א. נדחה H_0

ב. נדחה H_0

שאלה 11:

א. 0.0062

ב. 0.9938

ג. 0.0124

שאלה 12:

א. לא נדחה H_0

ב. לא נדחה H_0

שאלות מסכמות בסגנון רב ברירה (אמריקאיות) 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 הנה 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 :

לפני הנישואין-X	69	93	82	68
לאחר הנישואין-Y	80	88	84	71

מהן ההשערות הנבדקות?(ההפרש חושב X-Y)

א. $H_0: \mu_d = 0, \quad H_1: \mu_d < 0$

ב. $H_0: \mu_X - \mu_Y = 0, \quad H_1: \mu_X - \mu_Y < 0$

ג. $H_0: \mu_X - \mu_Y = 0, \quad H_1: \mu_X - \mu_Y < 0$

ד. $H_0: \mu_d = 0, \quad H_1: \mu_d > 0$

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. התפלגות Z ללא דרגות חופש.

ב. התפלגות T ו- 3 דרגות חופש.

ג. התפלגות T ו- 6 דרגות חופש.

ד. התפלגות χ^2 ו- 3 דרגות חופש.

9. שני סטטיסטיקאים בודקים השערות ברמת מובהקות $\alpha = 0.05$, על סמך אותו מדגם.

סטטיסטיקאי א' בודק את ההשערה: $H_0 : \mu = 20$ כנגד האלטרנטיבה $H_1 : \mu \neq 20$ ומחליט לא לדחות את השערת האפס.

סטטיסטיקאי ב' בודק את ההשערה $H_0 : \mu \leq 20$ כנגד האלטרנטיבה $H_1 : \mu > 20$

מה יחליט סטטיסטיקאי ב'?

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 האמיתית אצל הצמחוניים.

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 H0 כאשר היא הייתה נכונה.

ב. החוקר דחה את השערת H1 כאשר היא הייתה נכונה.

ג. החוקר לא דחה את השערת H0 כאשר היא הייתה לא נכונה.

ד. המדגם של החוקר שייך בפועל להתפלגות הדגימה של H1.

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.

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:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu < \mu_0$$

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.

24. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :

א. השערת האפס נכונה.

ב. השערת האפס נדחתה.

ג. השערת האפס לא נדחתה.

ד. אף אחת מהתשובות לא נכונה בהכרח.

25. מה המצב הרצוי לחוקר המבצע בדיקת השערה :

	α	$1-\beta$
ה.	גדולה	גדולה
ו.	גדולה	קטנה
ז.	קטנה	גדולה
ח.	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- $p\text{-}value=0.02$.
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה.

ה. יקבל את השערת האפס בכל מקרה.

ו. ידחה את השערת האפס מקרה.

ז. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.

ח. לא ניתן לדעת כי אין מספיק נתונים.

29. מובהקות התוצאה (PV◌◌◌◌◌◌) היא גם :

א. רמת המובהקות המינימאלית לדחות השערת האפס.

ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.

ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.

ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.

30. בבדיקת השערות מסוימת התקבל p value=0.0254 לכן :

א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0.

ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0.

ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0.

ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0.

31. רמת המובהקות במחקר הייתה 2% לכן.

א. בסיכוי של 2% נדחה את השערת האפס.

ב. בסיכוי של 2% לא נדחה את השערת האפס.

ג . בסיכוי של 2% השערת האפס לא נכונה.
ד. אף תשובה לא נכונה.

32. נתון ש $X \sim N(\mu,\sigma^2)$ כמו כן נתונות ההשערות הבאות :

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu < \mu_0$$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% .אם הוא היה מגדיל את רמת המובהקות ל- 10% אזי :

א. כעת בברור הוא ידחה את השערת האפס.
ב. כעת הוא דווקא יקבל את השערת האפס.
ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

33. לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית.

מספר דרגות החופש במבחן הוא :

א. 9
ב. 19
ג. 18
ד. 8

34. בתחנת טיפת חלב מסוימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשיים נשקל בכל אחד מהמשקלים. להלן תוצאות השקילה (בק״ג) :

משקל במכשיר 1	5.4	6.9	7.0	5.2

משקל במכשיר 2	5.3	6.9	7.1	5.0

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:

ה. לדחות את השערת האפס.

ו. לא לדחות את השערת האפס.

ז. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05 .

ח. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

38. במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו:

$-0.0293 < \mu_D < 0.2145$ רווח הסמך הוא ברמת סמך של 95%.

לכן מסקנת המחקר היא:

ד. ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.

ה. ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.

ו. 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 : $\begin{matrix} H_0 : \mu = 10 \\ H_1 : \mu \neq 10 \end{matrix}$. 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

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-α).

ג. אף תשובה לא נכונה.

ד. כל התשובות נכונות.

43. איזה מהמשפטים הבאים נכון תמיד?

א. $POWER + \beta + \alpha = 1$

ב. $\beta - POWER = .05$

ג. $POWER + \alpha = 1$

ד. $\beta + \alpha = 1$

ה. הכול לא נכון

44. מה נכון לומר לגבי הנחת שיוויון השונויות במבחן T למדגמים בלתי תלויים?

א. היא אומרת שהשונויות המדגמיות שוות.

ב. בלעדיה אין שום דרך לבדוק השערה על הפרש בין תוחלות.

ג. היא חשובה הן עבור מדגמים מזווגים והן עבור מדגמים בלתי תלויים.

ד. אף תשובה אינה נכונה.

45. חוקר החליט לא לדחות השערה ברמת מובהקות של α . במידה וחוקר זה היה בודק השערה זו ברמת מובהקות של 2α על סמך אותם נתונים האם ההשערה תדחה?

א. ההשערה תדחה.
ב. ההשערה לא תדחה.
ג. התשובה תלויה בעוצמת המבחן.
ד. לא ניתן לדעת בוודאות אם ההשערה תדחה או לא.

46. חוקרת שיערה, כי בגילאי הגן בנות יותר תקשורתיות מבנים. אם החוקרת תדגום אקראית 30 בנים ו-30 בנות, ובמדגם יתקבל אותו ממוצע של ציון תקשורת . סטטיסטי המבחן יהיה :

א. אפס

ב. חיובי

ג. שלילי

ד. לא ניתן לדעת

47. עוצמה שווה ל-1 פרושה :

א. לעולם לא לדחות את השערת האפס כאשר היא נכונה.

ב. תמיד לדחות את השערת האפס כאשר היא נכונה.

ג. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

48. מה מהבאים **נכון** לגבי מבחן t מדגמים מזווגים?

א. כל התצפיות במחקר אינן תלויות זו בזו.
ב. כל התצפיות במחקר תלויות זו בזו.
ג. כל הצמדים של תצפיות במחקר אינם תלויים זה בזה.
ד. התצפיות בתוך כל צמד אינן תלויות זו בזו.

49. לבדיקת ההשערה החד צדדית על התוחלת של התפלגות נורמלית $\begin{matrix} H_0: \mu \geq 10 \\ H_1: \mu < 10 \end{matrix}$ נלקח מדגם והתקבלה רמת מובהקות מינימאלית לדחיית השערת האפס 0.058. לו רצינו לבדוק את ההשערה הדו צדדית $\begin{matrix} H_0: \mu = 10 \\ H_1: \mu \neq 10 \end{matrix}$, 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 $\begin{matrix} H_0 : \mu = 55 \\ H_1 : \mu = 65 \end{matrix}$ נלקח מדגם מקרי בגודל n

מאוכלוסייה בעלת התפלגות נורמלית ושונות σ^2 . רמת המובהקות היא 5%. נמצא שהעוצמה היא 0.9. להלן 3 טענות:

1) עבור מדגם בגודל n וברמת מובהקות 5% לבדיקת ההשערות: $\begin{matrix} H_0 : \mu = 55 \\ H_1 : \mu = 60 \end{matrix}$ העוצמה תהיה גדולה מ- 0.9.

2) עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות: $\begin{matrix} H_0 : \mu = 55 \\ H_1 : \mu = 65 \end{matrix}$ העוצמה תהיה גדולה מ- 0.9.

3) עבור מדגם בגודל n ורמת מובהקות 10% לבדיקת ההשערות: $\begin{matrix} H_0 : \mu = 55 \\ H_1 : \mu = 65 \end{matrix}$ 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.

פתרונות :

1	א	2	א
2	ד	2	ב
3	ד	2	א
4	א	2	א
5	ג	3	ג
6	ג	3	ד
7	א	3	א
8	ב	3	א
9	ג	3	ד
1	א	3	ג
1	א	3	א
1	ג	3	ג
1	א	3	א
1	א	3	ג
1	א	4	א
1	א	4	א
1	א	4	א
1	ג	4	ה
1	א	4	ד
2	ג	4	ד
2	ב	4	א
2	ג	4	ד
2	ד	4	ג
2	ג	4	ב
2	ג	5	ד

פרק 55- מבחני חי בריבוע

מבחן טיב התאמה

רקע:

מבחן זה הוא מבחן הבא לבדוק האם אוכלוסייה מסוימת מתפלגת לפי התפלגות נתונה.

המשתנה הנחקר מחולק למספר קטגוריות ויש לבדוק האם תוצאות המדגם תואמות להתפלגות הנתונה.

מבנה המבחן:

השערות:

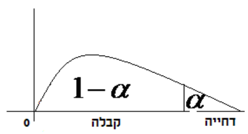
H_0: המשתנה מתפלג לפי התפלגות מסוימת

H_1: אחרת

כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש.

$d.f=K-1$, כאשר K - מספר הקטגוריות.



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

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

סטטיסטי המבחן:

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

O_i - השכיחות שנצפתה במדגם בקטגוריה i.

p_i - הסתברות לקטגוריה i לפי השערת האפס.

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

הערה: תנאי כדי לבצע את המבחן הוא $E_i\geq 5$ לכל i. במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.

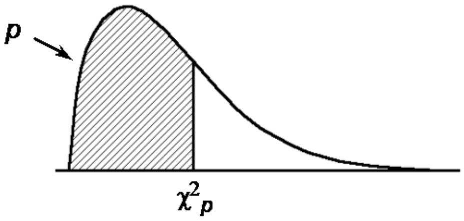
<u>דוגמה</u>: (פתרון בהקלטה)

במדינה מסוימת שלוש מפלגות. בפרלמנט הנוכחי התפלגות מספר המושבים היא 30% למפלגה A, 60% למפלגה B ו-10% למפלגה C.

לקראת הבחירות המתוכננות בשבוע הבא נעשה סקר שכלל 300 אזרחים.

בסקר התקבל ש40% יצביעו למפלגה A, 50% למפלגה B ו-10% למפלגה C. האם תוצאות הסקר תואמות להתפלגות המושבים בפרלמנט הנוכחי? בדקו ברמת מובהקות של 5%.

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p



p

df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
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

תרגילים:

1. במטרה לבדוק האם קובייה הוגנת, מטילים אותה 120 פעמים. התקבל 17 פעמים 1, 23 פעמים 2, 20 פעמים 3, 25 פעמים 4, 18 פעמים 5 ו- 17 פעמים 6. מה מסקנתכם ברמת מובהקות של 5%?

2. מפעל מייצר סוכריות בצבעים כחול, אדום, ירוק וכתום. מעוניינים לבדוק שפרופורציית הסוכריות הכחולות גדולה פי 2 מכל צבע אחר. לצורך כך נדגמו באקראי 200 סוכריות והתקבל: 70 כחולות, 50 אדומות, 40 ירוקות והיתר כתומות. מה מסקנתכם ברמת מובהקות של 5%?

3. משרד החינוך טוען שבקרב השכירים במשק היחס בין השכירים בעלי השכלה נמוכה, תיכונית ואקדמאית הוא 1:2:1 בהתאמה. במדגם של 200 שכירים התקבלו 56 אנשים בעלי השכלה נמוכה, 105 בעלי השכלה תיכונית והיתר בעלי השכלה גבוהה.
   א. על סמך תוצאות המדגם האם התפלגות ההשכלה היא כמו שמשרד החינוך מפרסם? בדוק ברמת מובהקות של 5%.
   ב. בנו רווח סמך ברמת סמך של 95% לפרופורציית השכירים במשק בעלי השכלה אקדמאית.

4. 200 איש נתבקשו לבחור ספרה באקראי והנה התוצאות שהתקבלו:
   18 איש בחרו בספרה 0
   24 איש בחרו בספרה 1
   17 איש בחרו בספרה 2
   19 איש בחרו בספרה 3
   20 איש בחרו בספרה 4
   18 איש בחרו בספרה 5
   22 איש בחרו בספרה 6
   היתר בחרו בספרות 7-9
   א. על סמך התוצאות הללו האם בחירת הספרות אקראית? רמת מובהקות של 2.5%.
   ב. תנו הערכה למובהקות התוצאה.
   ג. אם נגדיל את גודל המדגם פי 2 ונשמור על אותם יחסים של כמות האנשים במדגם שבחרו בספרות, כיצד הדבר ישפיע על ערכו של הסטטיסטי χ^2? מה תהיה המסקנה במקרה זה?

5. מעוניינים לבדוק האם קובייה היא הוגנת. הטילו את הקובייה פעמיים והתבוננו בסכום הוצאות. חזרו על התהליך 72 פעמים. להלן התוצאות שהתקבלו במדגם :

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

מה המסקנה ברמת מובהקות של 5%.

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

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

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

7. מטילים מטבע עד שלראשונה מתקבל "ראש". חוזרים על התהליך 120 פעמים. נסמן ב-X את מספר ההטלות עד קבלת הראש. להלן התוצאות שהתקבלו :

x	1	2	3	4	5	6
מספר החזרות על התהליך	54	20	16	22	6	2

א. בהנחה והמטבע הוגן, מהי ההתפלגות של X?

ב. בדוק האם המטבע הוגן, על סמך תוצאות המדגם ברמת מובהקות של 5%.

.

8. להלן השערות מחקר:

$$H_0 : X \sim N(40, 2^2)$$
$$H_1 : else$$

תוצאות המדגם הן:

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

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

פתרונות:

שאלה 1 :

לא נדחה H_0

שאלה 2 :

לא נדחה H_0

שאלה 3 :

א. לא נדחה H_0 ב. $(0.14,0.25)$

שאלה 4 :

א. לא נדחה H_0 ב. בין 0.95 ל- 0.975 ג. יגדל פי 2. מסקנה לא תשתנה.

שאלה 5 :

נכריע שהקובייה אינה הוגנת.

שאלה 6 :

0.005

שאלה 7 :

א. $X \sim G(0.5)$ ב. נסיק שהמטבע לא הוגן.

שאלה 8 :

14

הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה

רקע:

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

דוגמה: (פתרון בהקלטה)

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

א. באמצעות מבחן טיב התאמה.

ב. באמצעות מבחן Z לפרופורציה בודדת.

תרגילים:

1. בסקר שנעשה על 320 נשאלים, 43.75% טענו שהחיה המועדפת עליהם היא כלב. עד היום היה נהוג לחשוב ש40% מהאנשים מעדיפים כלבים.

בדקו ברמת מובהקות של 5% האם הסקר יישנה את הסברה שהייתה נהוגה עד היום לגבי העדפת כלב.

א. באמצעות מבחן טיב התאמה.
ב. באמצעות מבחן על פרופורציה.

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

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

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

3. בתחרות ריצה בית ספרית שלושה מסלולי ריצה. ב-50 תחרויות בדקו באיזה מסלול היה הניצחון. התוצאות שהתקבלו מסוכמות בטבלה הבאה:

המסלול	1	2	3
מספר הניצחונות	20	15	15

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

ב. בדקו ברמת מובהקות של 5% האם הסיכוי לנצח במסלול מספר 1 גבוה מ- $\frac{1}{3}$.

פתרונות:

שאלה 1 :

לא נדחה H_0

שאלה 2 :

א. נדחה H_0 ב. לא נדחה H_0

שאלה 3 :

א. לא נדחה H_0 ב. לא נדחה H_0

מבחן חי בריבוע לאי תלות בין משתנים

רקע:

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים.

שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות .

מבנה המבחן:

השערות:

H_0: אין תלות בין המשתנים

H_1 : יש תלות בין המשתנים

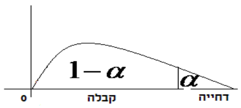
כלל הכרעה :

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:

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

O_i - השכיחות נצפית במדגם בתא i.

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

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

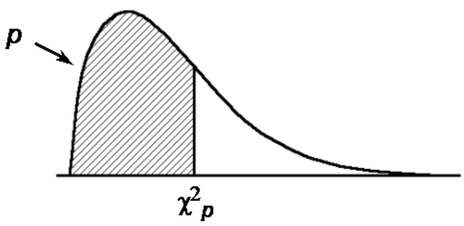
הערה: תנאי כדי לבצע את המבחן הוא $E_i\geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים. תנאי חלופי : אין E קטן מ- 1 וגם אין ביותר מ 20% מהתאים E קטן מ- 5.

דוגמה : (הפתרון בהקלטה)

האם יש תלות בין המגדר לבין דעה מסוימת ? יש לבדוק ברמת מובהקות של 5% על סמך תוצאות הסקר:

דעה / המגדר	בעד	נגד	נמנע	סה"כ
גברים	50	40	10	
נשים	20	60	20	
סה"כ				

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p



p

df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
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

תרגילים :

1. נבדק 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 ? בדוק ברמת בטחון של 95%.

4. במטרה לבדוק האם השתנו דפוסי ההצבעה למפלגות השונות בין שבוע שעבר לשבוע נלקחו שני סקרים אחד מהשבוע שעבר והאחר מהשבוע . להלן דפוסי ההצבעה שהתקבלו בסקרים אלה :

	מפלגה א	מפלגה ב	מפלגות אחרות	סה"כ
שבוע שעבר				
השבוע	143		253	550
סה"כ	243	314		1050

א. מהי רמת המובהקות המינמלית עבורה ניתן להחליט שהשתנו דפוסי ההצבעה משבוע שעבר לשבוע באופן מובהק?

ב. כיצד הייתה התשובה לסעיף א משתנה אם כל השכיחויות בטבלה של תוצאות המדגם היו מוכפלות פי 2?

ג. בנו רווח סמך לשיעור המצביעים למפלגה א השבוע ברמת סמך של 95%.

5. בחנות בגדים A בדקו את התפלגות הצבעים של הבגדים הנמכרים ביום מסוים:

צבע	שחור	לבן	אדום	כחול
מספר הפריטים	15	20	15	50

כמו כן בדקו את התפלגות הצבעים בחנות שכנה B:

צבע	שחור	לבן	אדום	כחול
מספר הכדורים	60	20	10	20

א. בדוק ברמת מובהקות של 5% האם התפלגות הצבעים בחנות A היא ביחס של 3:1:1:1 לטובת הכחול.

ב. בדוק ברמת מובהקות של 2.5% האם קיים הבדל בין החניות מבחינת התפלגות הצבעים של הפריטים הנמכרים.

6. סטודנט קיבל בבדיקת השערות ערך χ^2 (chi-square) השווה לאפס. הסטודנט הסיק כי לא קיימת תלות, בין שני המשתנים שבדק, בכל רמות מובהקות. **נכון ?**

לא נכון? נמק/י

7. להלן טבלת O של שני משתנים שהתקבל במדגם כלשהו:

	Y_1	Y_2	Y_3	Y_4	$f(x)$
X_1					200
X_2					200
$f(y)$	60	60	120	160	

מה צריכות להיות השכיחויות בתוך הטבלה כדי שמובהקות התוצאה (PV) תהיה 100%?

פתרונות:

שאלה 1:

נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.

שאלה 2:

נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.

שאלה 3:

נסיק שאין קשר בין היות מוצר לייצוא למועד שבו הוא יוצר.

שאלה 4:

א. 10%

ב. קטן

ג. (0.223,0.297)

שאלה 5:

א. נסיק שהתפלגות הצבעים בחנות היא כמו שמצוין.

ב. נסיק שיש הבדל בין החנויות מבחינת התפלגות הצבעים.

שאלה 6:

נכון

הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות

רקע:

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

דוגמה: (פתרון בהקלטה)

בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים?

האם אפשר לבדוק זאת גם על ידי מבחן לאי תלות וגם על ידי בדיקת השערות לשתי פרופורציות?

<u>תרגילים:</u>

1. בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? בדוק ברמת מובהקות של 5%.

   א. על ידי מבחן לאי תלות.

   ב. על ידי בדיקת השערות לשתי פרופורציות.

2. נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.

   א. בנה טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.

   ב. האם ניתן להגיד שהסיכוי להימצא במצב בריאותי תקין גבוה יותר כאשר עוסקים בפעילות גופנית סדירה לעומת המצב שלא עוסקים בפעילות גופנית סדירה? בדוק ברמת בטחון של 90%

   .

<u>פתרונות:</u>

<u>שאלה 1:</u>

נדחה את השערת האפס.

<u>שאלה 2:</u>

ב. נדחה את השערת האפס.

פרק 56 - ניתוח שונות חד כיוונית

רקע תיאורטי:

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות ($\mu_1,......,\mu_k$) של k אוכלוסיות שונות.

ולכן בניתוח שונות השערות המחקר הן:

$H_0 : \mu_1 = \mu_2 = . . . = \mu_k$ (התוחלות של כל האוכלוסיות שוות)

$H_1 :$ אחרת (לפחות שתיים מהתוחלות שונות)

ההנחות הדרושות לביצוע התהליך הן:

1. בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.
2. כל האוכלוסיות הן עם אותה שונות σ^2.
3. המדגמים בלתי תלויים זה בזה.

ישנו משתנה המבדיל בין הקבוצות השונות, הוא המשתנה הבלתי תלוי הנקרא גורם (factor)

משתנה זה הוא קטגוריאלי עם k רמות (levels).

כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה:

נסמן ב- n_i – את גודל המדגם בקבוצה i.

$n = \sum_{1}^{k} n_i$ - 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 $SS_W = \sum_{i=1}^{k} [n_i - 1] \bullet \hat{S}_i^2$

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

SST = SSB + SSW

יש למלא את טבלת ניתוח השונות הבאה :

טבלת ניתוח שונות

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

$$F=\frac{SS_B/(k-1)}{SS_W/(n-k)}\sim F(k-1,n-k)$$

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

תרגילים:

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 :

בית הספר	"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 ?

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

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

האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות? בדקו ברמת מובהקות של 5%.

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.

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			

מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

6. להלן טבלת ANOVA בטבלה הושמטו חלקים. השלם את החלקים בטבלה שהושמטו ומסומנים באותיות.

ANOVA

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

7. חברת תרופות לקחה 15 אנשים ברמת בריאות דומה. החברה חילקה את האנשים ל שלוש קבוצות שוות בגודלן. לכל קבוצה ניתנה אותה תרופה במינון שונה (dosage) . המינונים שניתנו הם: 10 מ"ג , 20 מ"ג ו- 30 מ"ג. לאחר שעה מזמן לקיחת התרופה 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			

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.

Homogeneous Subsets

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.

א. בדוק 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. נמק!

ד. לטבלת ה ANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "טוקי". ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

8. בעיר מסוימת רצו לבדוק האם קיים הבדל ברמה של התלמידים בין בתי הספר השונים בעיר. ביצעו מדגם מכל בית ספר ונתנו מבחן זהה לכל הנדגמים.
לאחר מכן ריכזו את הנתונים בתוכנה סטטיסטית והפעילו ניתוח שונות. מצורפים הפלטים שהתקבלו.
ענו על הסעיפים הבאים:
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			

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

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

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

Homogeneous Subsets

grade

Scheffe[a]

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

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

פתרונות סופיים חלקיים - ניתוח שונות חד כיוונית

2. אם חישבת נכון ה F הסטטיסטי יוצא: 0.58

3. נדחה את השערת האפס.

4.להלן טבלת הניתוח השונות המתקבלת:

	Sum of Squares	df	Mean Square	F
Between Groups	126.000	2	63.000	.756
Within Groups	750.000	9	83.333	
Total	876.000	11		

5.נקבל את השערת האפס.

6.א. 165.5 ב. 2 ג. 178.725 ד. 9.375 ה. 18.36

7.א. נדחה את השערת האפס. ב. לא משתנה. ג. כן

8.א. 5 ב. 25 ג. כן

פרק 57 - 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).למשל,

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.

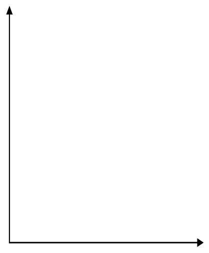
בשלב הראשון, נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים.

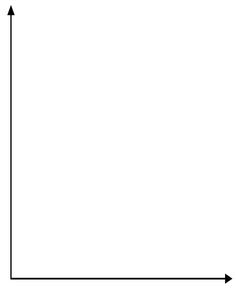
למשל, בבניין של 5 דירות בדקו את הנתונים הבאים: X - מס' חדרים בדירה. Y- מס' נפשות הגרות בדירה. להלן התוצאות שהתקבלו :

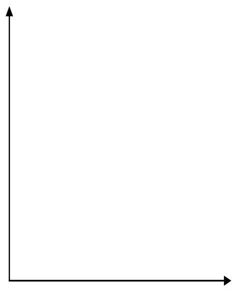
מס' דירה	X	Y
1	3	2
2	2	2
3	4	3
4	3	3
5	5	4

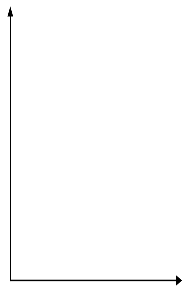
נשרטט מנתונים הללו דיאגרמת פיזור :

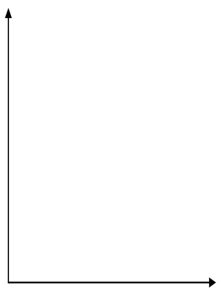
נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן :











בשלב השני, 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).

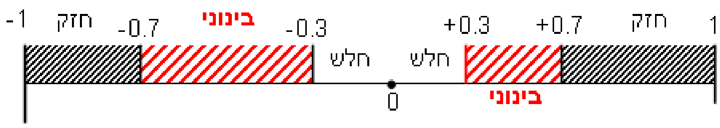
מקדם מתאם זה מקבל ערכים בין 1- ל 1.

מקדם מתאם1- או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי הנוסחה : $y = bx + a$.

מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר לנארי מלא בו השיפוע b יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לנארי מלא בו השיפוע b שלילי (מקדם מתאם 1-).

מתאם חיובי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

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 : $COV(x,y)=\frac{\sum(x-\overline{x})(y-\overline{y})}{n}=\frac{\sum xy}{n}-\overline{x}\cdot\overline{y}$

שונות של המשתנה X: $s_x^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}=\frac{\sum_{i=1}^{n}x_i^2}{n}-\overline{x}^2$

שונות המשתנה Y: $S_Y^2=\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{n}=\frac{\sum_{i=1}^{n}y_i^2}{n}-\overline{y}^2$

מקדם המתאם הלינארי : $r_{xy}=\frac{\mathrm{cov}(x,y)}{s_x\cdot s_y}$

תרגילים:

1. להלן נתונים לגבי שישה תלמידים שנגשו למבחן . בדקו לגבי כל תלמיד את הציון שלו בסוף הקורס וכמו כן את מספר החיסורים שלו מהקורס.

ציון	מספר חיסורים
80	2
90	1
90	0
70	2
70	3
50	4

א. שרטט דיאגראמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר ביו מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?

ב. חשב את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?

ג. הסבר 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:

y	x
12	10
15	14
15	15
17	18
21	20

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 :

$$\sum_{i=1}^{20} Y_i = 200 \qquad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 76 \quad \sum_{i=1}^{20} (X_i - \bar{X})^2 = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y - \bar{Y}) = 60.8$$

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 :

$$\sum_{i=1}^{20} Y_i = 200 \qquad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} Y_i^2 = 2080 \qquad \sum_{i=1}^{20} X_i^2 = 2960$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

חשב 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 הייתה2.
מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם ?

6. להלן רשימת טענות, לגבי כל טענה קבע נכון/לא נכון ונמק!

א. מתוך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתוך 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.

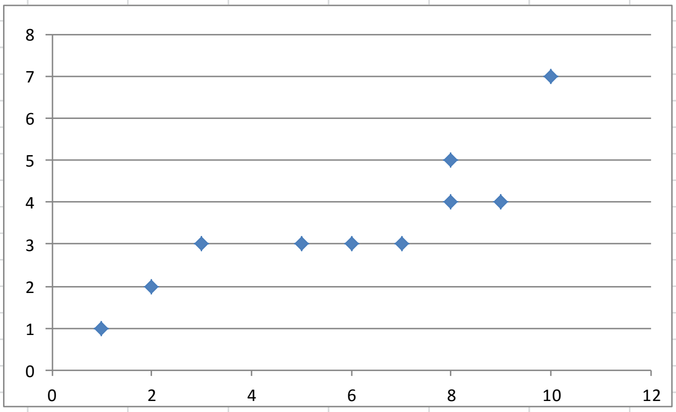
ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות אמריקאיות:

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.

8. נלקחו 20 מוצרים וניבדק 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

פתרונות:

שאלה 1:

א. בהקלטה

ב. ‎-0.9325

שאלה 2:

א. $\bar{y}=16$ $\quad$ $\bar{x}=15.4$

ב. $r_{xy}=0.96$

שאלה 3:

א: 0.8

שאלה 4:

0.8

שאלה 5:

1

שאלה 6:

א. נכון

ב. לא נכון

ג. נכון

שאלה 7:

התשובה: ג

שאלה 8:

התשובה: א

שאלה 9:

התשובה: ב

פרק 58 - מדדי קשר - השפעת טרנספורמציה לינאריות על מדד הקשר של פירסון

רקע:

טרנספורמציה לינארית בין אם נעשית על X ובין אם נעשית על y, או בין אם נעשית על שניהם, אינה משנה את עוצמת הקשר . היא עלולה רק לשנות את כיוונו אם השיפועים של שתי הטרנספורמציות שוני סימן.

$$r_{[(aX+b),(cY+d)]}=\begin{cases} r_{x,y}\_\_\_if\_\_a\cdot c>0 \\ -r_{x,y}\_\_if\_\_a\cdot c<0 \end{cases}$$

תרגילים:

1. מבחן בנוי מחלק כמותי ומילולי.
מקדם המתאם בין שני הציונים של שני החלקים הוא 0.9.
א. אם יעלו את כל הציונים בחלק המילולי ב- 20% , מה יהיה מקדם המתאם בין הציון המילולי החדש לציון הכמותי ובין הציון המילולי הישן לציון המילולי החדש?
ב. נגדיר משתנה חדש W להיות המרחק של הציון בחשיבה מילולית מהציון המקסימאלי בבחינה- 150. מצא את מקדם המתאם בין הציון המילולי ל- W ובין W ל-לציון הכמותי.

2. מקדם המתאם בין ההכנסה לבין ההוצאה של 10 משפחות חושב והתקבל 0.7 . אם חל גידול של 5% בהכנסת האוכלוסייה כולה וגידול של 7% בהוצאה שלה, אז מה יהיה מקדם המתאם בין ההכנסה החדשה להוצאה החדשה?

3. חברת "לקי" המייצרת גלידה החליטה לערוך מחקר לבדיקת הקשר בין מספר חבילות הגלידה הנמכרות ביום לבין הטמפרטורה באותו יום. נבדקו 10 ימים והתקבל מתאם לינארי 0.85.
חברת "לקי" דואגת להתחיל כל יום עם מלאי של 150 חבילות גלידה. בנוסף, מעוניינים כי הטמפרטורה תבוטא במעלות פרנהייט במקום במעלות צלסיוס. מה ערכו של מקדם המתאם בין מספר חבילות הגלידה שנשארות בסוף היום לבין הטמפרטורה במעלות פרנהייט?
הקשר בין מעלות צלסיוס $(C°)$ למעלות פרנהייט $(F°)$ נתון ע"י $F = \frac{9}{5}C + 32$.

<u>בחר בתשובה הנכונה:</u>
א. 0.85
ב. 0.85-
ג. 1
ד. לא ניתן לדעת.

4. מקדם המתאם בין X ל- Y הנו 0.4 כל ערכי ה-X הוכפלו ב- 2 לכן מקדם המתאם החדש בין שני המשתנים יהיה :
<u>בחר בתשובה הנכונה:</u>
א. 0.8
ב. 0.4
ג. -0.4
ד. לא ניתן לדעת.

פתרונות :

שאלה 1:

א. בין הציון המילולי הישן לחדש: 1

בין הציון המילולי החדש לכמותי: 0.9

ב. בין W ל ציון המילולי : 1-

בין W לציון הכמותי: 0.9-

שאלה 2:

0.7

שאלה 3:

התשובה : ב

שאלה 4:

התשובה : ב

פרק 59 - 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 - בעצם נותן את ערך Y כאשר X הנו אפס על גבי קו הניבויים. הוא ניקרא 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:

$$\tilde{Y} = bX + a$$

$$b = r\frac{S_Y}{S_X}$$

$$a = \bar{Y} - b\bar{X}$$

אם נרצה לבנות קו ניבויים לניבוי X על סמך Y נצטרך לעדכן את הנוסחאות בהתאם.

תרגילים:

1. נסמן ב-X את ההכנסה של משפחה באלפי ₪. נסמן ב-Y את ההוצאות של משפחה באלפי ₪.
נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} Y_i = 200 \qquad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20}(Y_i - \bar{Y})^2 = 76 \quad \sum_{i=1}^{20}(X_i - \bar{X})^2 = 76$$

$$\sum_{i=1}^{20}(X_i - \bar{X})(Y - \bar{Y}) = 60.8$$

א. חשב את מדד הקשר הלינארי בין X ל-Y. מיהו המשתנה התלוי?

ב. מצא את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבר את משמעות הפרמטרים של קו הרגרסיה.

ג. משפחת כהן הכניסה 15,000 ₪, מה ההוצאה הצפויה שלה?

1. נסמן ב-X את ההשכלה של אדם בשנות למוד. נסמן ב-Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_Y = 5 \qquad S_X = 2$$

$$\bar{Y} = 8 \qquad \bar{X} = 14$$

$$COV(X,Y) = 7.5$$

א. חשב 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 הרגרסייה לניבוי הציון לפי שעות ההכנה?

4. נתונים 2 משתנים Y,X. כמו כן נתון: X ממוצע = 1.5, שונות X = שונות Y = 4,וכן שקו הרגסיה של Y על בסיס X הינו $Y = -0.2X + 0.5$. חשב מהו מקדם המתאם בין X ל Y?

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

פתרונות:

שאלה1:

א. 0.8

ב. $\tilde{Y} = 0.8X + 0.4$

ג. 12.4

שאלה 2:

א. 0.75

ב. 4.25 אלפי ש"ח

ג. 14.6 שנים

שאלה 3:

א. 1.2

ב. 29

ג. y=1.2x+29

שאלה 4:

-0.2

פרק 60 - מדדי קשר - רגרסיה - שונות מוסברת ושונות לא מוסברת

רקע:

המטרה ברגרסיה הנה להסביר את השונות של המשתנה התלוי.
למשל, להסביר את השונות של המשכורות באמצעות הוותק או להסביר את השוני בציונים באמצעות כמות החיסורים.

r^2 -נותן בעצם איזה חלק מהשונות של המשתנה התלוי מוסבר.
השונות המוסברת נקראת גם שונות ניבויים.
השונות הלא מוסברת נקראת גם שונות טעויות.

תרגילים :

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 קבע נכון/לא נכון ונמק!

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 הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות אמריקאיות:

בשאלות הבאות יש לבחור בתשובה הנכונה.

3. בקשר בין שני משתנים התקבל $\boldsymbol{r^2 = 0.64}$ 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 מה ניתן לומר?

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

5. בקורס מבוא לכלכלה ניתנו במשך השנה שני מבחנים : מבחן בסוף סימסטר א (X) ומבחן בסוף סימסטר 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