

סטודנטים יקרים

לפניכם ספר תרגילים בקורס מבוא לסטטיסטיקה והסתברות א'.
הספר הוא חלק מקורס חדשני וראשון מסוגו בארץ בנושא זה,
המועבר ברשת האינטרנט On-line.

הקורס באתר כולל פתרונות מלאים 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 [לחצו כאן](#).**

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
לחויית לימודים יוצאת דופן וחדשה לחלוטין, היכנסו עכשיו לאתר
www.gool.co.il.



אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

תוכן

פרק 1 - סטטיסטיקה תיאורית - הקדמה	3
פרק 2 - סטטיסטיקה תיאורית - הצגה של נתונים	6
פרק 3 - סטטיסטיקה תיאורית - סכימה	14
פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי	17
פרק 5 - סטטיסטיקה תיאורית - מדדי פיזור: הטוח, השונות וסטיית התקן	27
פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - טוח בין-רבעוני	32
פרק 7 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן	35
פרק 8 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים במחלקות	38
פרק 9 - מדדי קשר - מדד הקשר הלינארי (פירסון)	43
פרק 10 - מדדי קשר - רגרסיה ליניארית	51
פרק 11 - התפלגויות רציפות מיוחדות - התפלגות נורמלית	54
פרק 12 - התפלגות הדגימה	63
פרק 13 - אמידה נקודתית	70
פרק 14 - רווחי סמך	83
פרק 15 - טעויות ועוצמה בבדיקת השערות כללית	90
פרק 16 - בדיקת השערות - מבחנים פרמטרים	93

פרק 1 - סטטיסטיקה תיאורית - הקדמה

רקע:

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

בסטטיסטיקה תיאורית אנחנו נתבונן בקבוצה מסוימת ובתוך הקבוצה הזו נאסוף נתונים לגבי משתנה מסוים ונלמד להציג את הנתונים ולנתח אותם מכל מיני אספקטים.

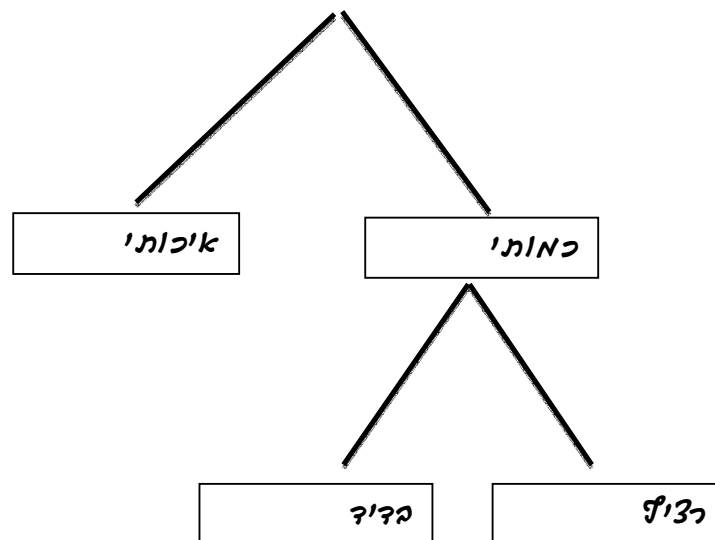
דוגמה:

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

מי הקבוצה הנחקרת?

מה גודל הקבוצה?

מה המשתנה הנחקר?

סוגי משתנים:

משתנה איכותי הוא משתנה שלערכיו אין משמעות של יותר או פחות, אין עניין כמותי לערכים המתקבלים.

כמו: מקום מגורים של אדם (רעננה, תל אביב, אשדוד..)

מין האדם (זכר, נקבה)

מצב משפחתי (רווק, נשוי, גרוש, אלמן)

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

את המשתנה הכמותי נסווג לשני סוגים:

משתנה בדיד: משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים. כמו: מספר ילדים למשפחה (1,2,3..)

ציון בבחינה (מ 0 ועד 100 בקפיצות של 1)
הערה:

משתנה רציף: משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים, הערכים מתקבלים ברצף וללא קפיצות של ערכים.

כמו: גובה בס"מ – אם למשל, הגובה הנמוך ביותר הוא 150 ועד 190 ס"מ בקבוצה הגבהים הם ברצף. גם בין 160 ל 161 ס"מ יש רצף אינסופי של ערכים אפשריים לגובה (160.33 ס"מ הוא גם גובה אפשרי)

משקל בק"ג, מהירות בקמ"ש וכולי.

תרגילים:

1. סווג את המשתנים הבאים לפי: איכותי / כמותי בדיד / כמותי רציף:
 - א. מספר הדירות בבניין.
 - ב. גיל אדם בשנים.
 - ג. אחוז האבטלה בעיר.
 - ד. מקצוע לימוד מועדף.
2. להלן התפלגות מספר האיחורים לעבודה בחודש של העובדים בחברת "סטאר".
בחברה 200 עובדים.

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

- א. מהו המשתנה הנחקר כאן?
- ב. האם מדובר במשתנה איכותי או כמותי? אם הוא כמותי האם הוא בדיד או רציף?
3. לפניכם רשימה של משתנים כמותיים. ציין ליד כל משתנה אם הוא רציף או בדיד.
 - א. שכר עובד בש"ח.
 - ב. ציון בחינת בגרות.
 - ג. תוצאה בהטלת קובייה.
 - ד. מהירות ריצה בתחרות.
 - ה. שיעור התמיכה בממשלה.

פרק 2 - סטטיסטיקה תיאורית - הצגה של נתונים

רקע:

דרכים להצגת נתונים שנאספו :

א. רשימה של תצפיות:

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

3 4 3 5 4

ב. טבלת שכיחויות בדידה:

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

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

למשל, להלן התפלגות הציונים בכיתה מסוימת :

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

שכיחות מצטברת – צבירה של השכיחויות : F_i - השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך .

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי : $\frac{f_i}{n}$ - איזה חלק מהתצפיות בקבוצה שוות לערך .

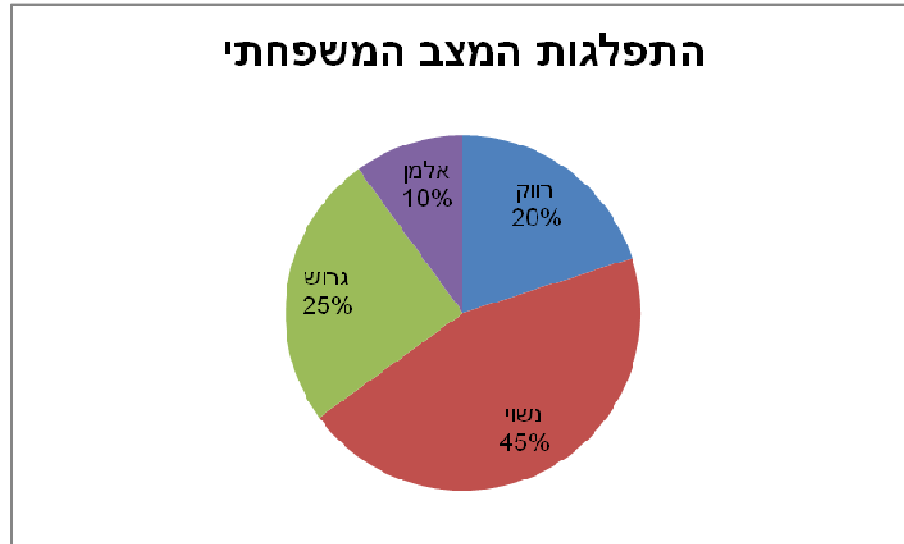
ג. טבלת שכיחויות במחלקות :

משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד וטבלת שכיחויות תהיה ארוכה מידי .
למשל, נתנו לקבוצת ילדים לבצע משימה מסוימת ובדקו את התפלגות זמן ביצוע המשימה בדקות.
להלן ההתפלגות שהתקבלה :

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

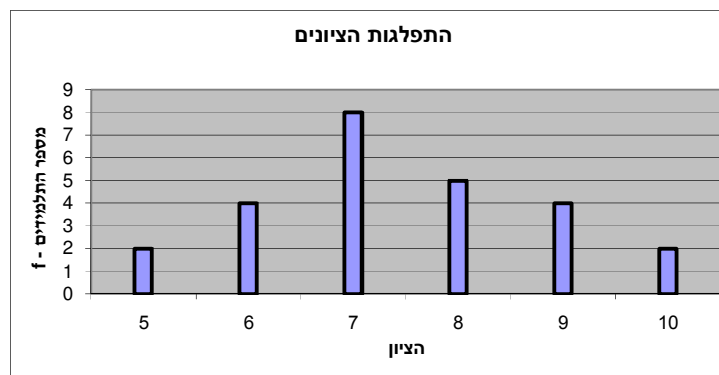
ד. דיאגרמת עוגה :

זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח" יחסי מהעוגה. הנתח בעוגה פרופורציוני לשכיחות היחסית של ערך המשתנה בנתונים.



ה. דיאגרמת מקלות :

הציר האופקי הוא הציר של המשתנה הציר האנכי של השכיחות – הגובה של המקל מעיד על השכיחות .
 רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף .
 כמו כן בסולמות מדידה עבור משתנה מסולם סדר.



1. היסטוגרמה:

ההיסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות.

רלבנטית למשתנה כמותי רציף.

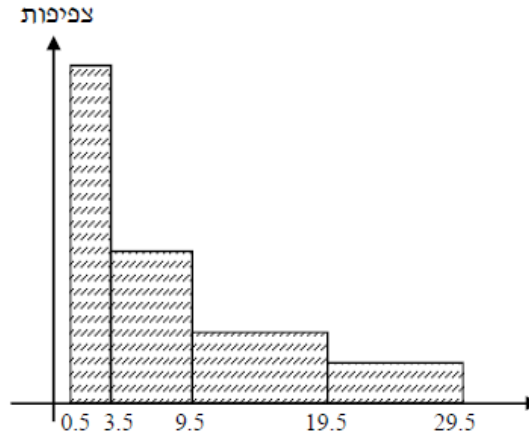
בהיסטוגרמה ציר האופקי הוא הציר של המשתנה וציר האנכי הוא הציר של הצפיפות.

הצפיפות מחושבת בכל מחלקה על ידי חלוקת השכיחות ברוחב של כל המחלקה והיא נותנת

את מספר התצפיות הממוצע בכל מחלקה ליחידה.

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

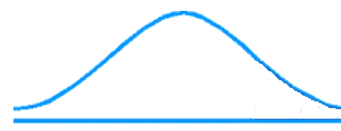
			X		
צפיפות	מצטברת	שכיחות	אמצע	רוחב	
6.6667	20	20	2	3	0.5 - 3.5
3	38	18	6.5	6	3.5 - 9.5
1.4	52	14	14.5	10	9.5 - 19.5
0.8	60	8	24.5	10	19.5 - 29.5



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

צורות התפלגות נפוצות

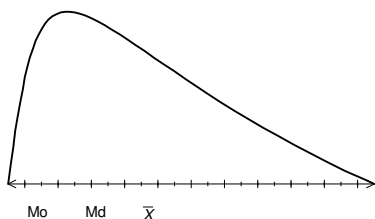
התפלגות סימטרית פעמונית- רוב התצפיות במרכז וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. למשל, ציוני IQ.



ישנן התפלגויות סימטריות שאינן פעמוניות:

התפלגות אסימטרית ימנית (חיובית) –רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. למשל, שכר במשק.

התפלגות א-סימטרית ימנית או חיובית



התפלגות אסימטרית שמאלית (שלילית) רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. למשל, אורך חיים.

התפלגות א-סימטרית שמאלית או שלילית



תרגילים:

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

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

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

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

3. להלן התפלגות ההשכלה במקום עבודה מסוים:

מספר העובדים	השכלה
60	נמוכה
120	תיכונית
20	אקדמאית

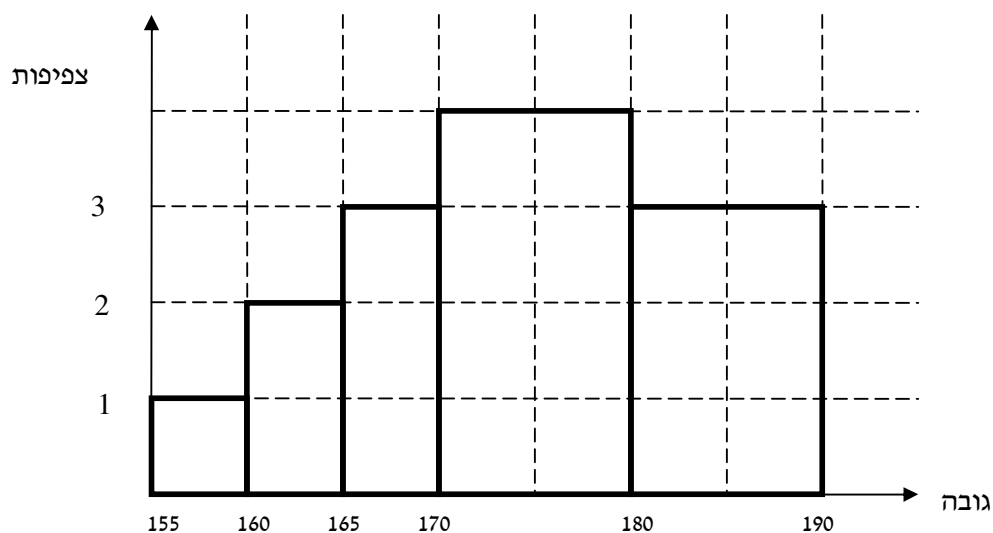
- א. מהו המשתנה הנחקר? מאיזה סולם הוא?
ב. תארו את הנתונים באופן גרפי.

4. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:

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

- א. מהו המשתנה? האם הוא בדיד או רציף?
ב. תאר את הרשימה בטבלת שכיחות.
ג. הוסף שכיחות יחסית לטבלה.
ד. תאר את הנתונים באופן גרפי.

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



- מהו המשתנה הנחקר? האם הוא בדיד או רציף?
- תאר את הנתונים בטבלת שכיחויות במחלקות.
- הוסף שכיחות יחסית לטבלה.
- הוסף את הצפיפות של כל מחלקה לטבלה.
- מהי צורת ההתפלגות של הגבהים?

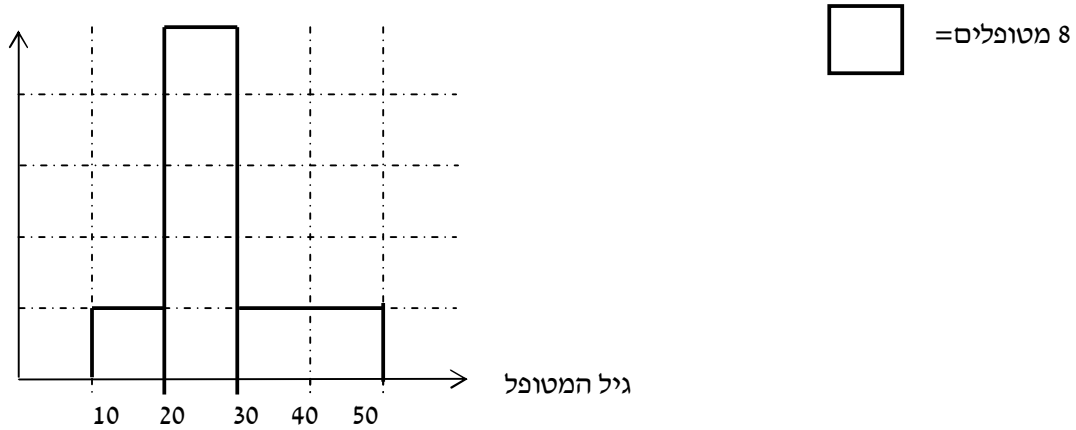
6. להלן התפלגות המשקל של קבוצה מסוימת בק"ג :

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

- תאר את ההתפלגות באופן גרפי.
- מה ניתן להגיד על צורת ההתפלגות?

7. להלן גיל המטופלים של ד"ר שוורץ בשנים :

קנה מידה :



א. מה המשתנה הנחקר? האם הוא בדיד או רציף?

ב. מהי הקבוצה הנחקרת?

ג. תרגמו את ההסיטוגרמה לטבלת שכיחות.

ד. מהי הפרופורציה של המטופלים של ד"ר שוורץ בגילאים 20-30?

פרק 3 - סטטיסטיקה תיאורית - סכימה

רקע:

בסטטיסטיקה ישנה צורת רישום מקובלת כדי לרשום סכום של תצפיות:

$$\sum_{i=1}^n X_i$$

נסביר את צורת הרישום על ידי הדוגמה הבאה:

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

תרגילים:

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

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

חשבו :

$$\sum_{i=1}^3 X_i$$

$$\sum_{i=1}^5 Y_i$$

$$\sum_{i=1}^4 X_i$$

$$\left(\sum_{i=1}^4 X_i\right)^2$$

$$\sum X_i$$

$$\sum X_i Y_i$$

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

2. נתון לוח ערכי המשתנים x_i ו y_i כאשר : $i=1,2,\dots,6$

i	1	2	3	4	5	6
x_i	3	2	4	-2	1	4
y_i	2	0	0	1	-5	2

ונתונים הקבועים : $b=5$ $a=2$ חשבו את הנוסחאות הבאות :

א. $\sum_{i=1}^4 y_i$

ב. $\sum_{i=1}^6 a$

ג. $\sum_{i=1}^6 x_i y_i$

ד. $\sum_{i=1}^6 (x_i + y_i)$

ה. $\sum_{i=1}^6 x_i + a$

3. קבע לכל זהות אם היא נכונה :

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

ב. $\sum_{i=1}^n a = a \cdot n$

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

4. נתון : $\sum_{i=1}^{10} X_i = 80$ $\sum_{i=1}^{10} X_i^2 = 1640$
 חשב : $\sum_{i=1}^{10} (X_i - 4)^2$ (פתרון : 1160)

פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי

רקע:

המטרה במדדי המיקום המרכזי למדוד את מרכז ההתפלגות של התצפיות.

השכיח – MODE

השכיח הוא הערך הנפוץ ביותר בהתפלגות.

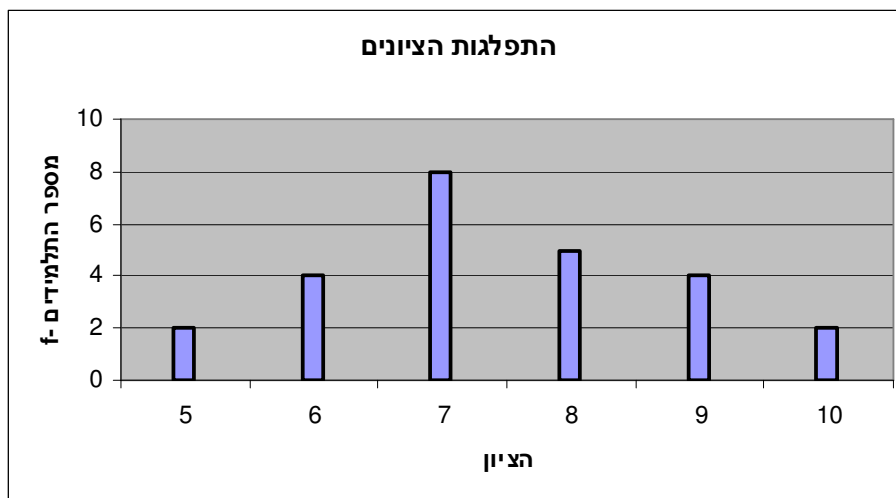
ברשימה : הערך החוזר על עצמו הכי הרבה פעמים .

7 9 4 8 4 10 6

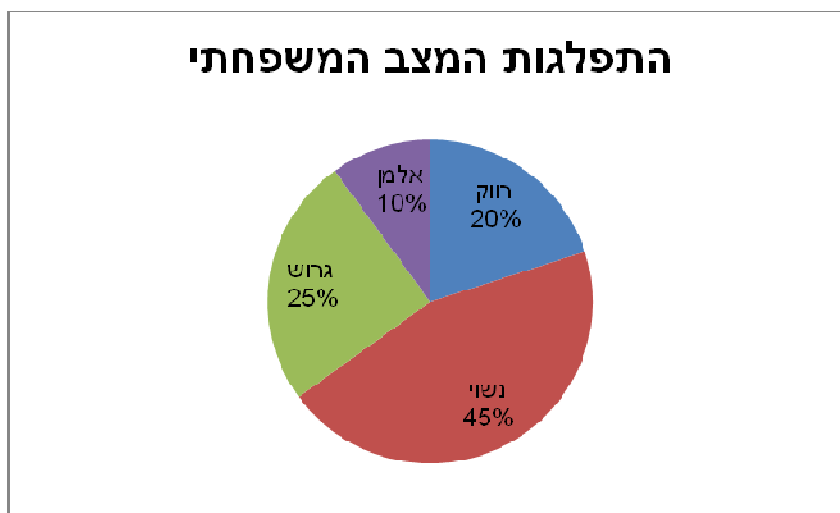
בטבלת שכיחויות בדידה : הערך שהשכיחות שלו היא הגבוהה ביותר.

# תכניות החיסכון	$f(x)$
0	100
1	75
2	25
3	25
4	25

בדיאגרמת מקלות : שיעור ה- X של המקל הגבוה ביותר.



בעוגה: הערך של הפלח הגדול ביותר.

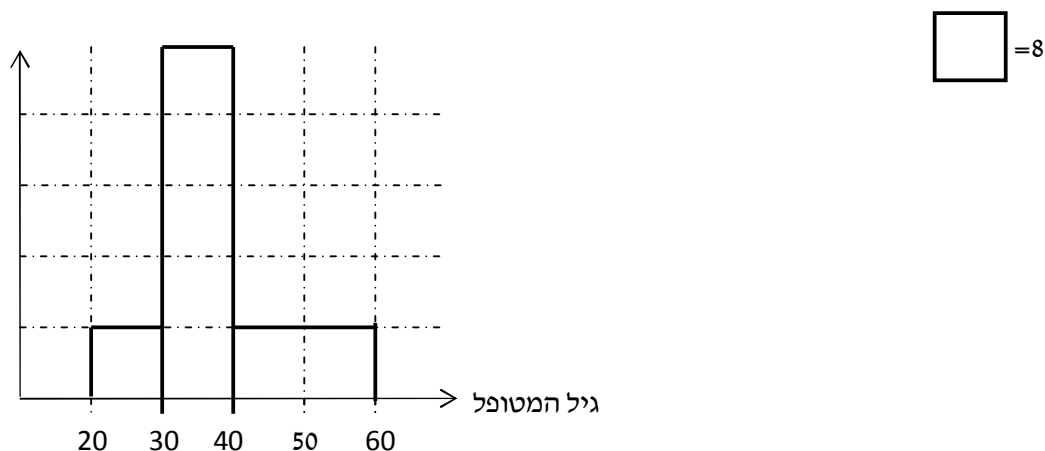


בטבלת שכיחויות במחלקות: אמצע המחלקה עם הצפיפות הגבוהה ביותר.
התפלגות הציונים בכיתה.

$f(x)$	X
20	0-60
10	60-70
18	70-80
15	80-90
15	90-100

בהיסטוגרמה: שיעור ה- X של אמצע המחלקה הגבוהה ביותר.

להלן גיל המטופלים של ד"ר שוורץ בשנים:



כללי: יתכן שלהתפלגות יותר משכיח אחד.

השכיח הוא מדד הרלבנטי לכל סוגי המשתנים.

MIDRANGE – (טווח)

הממוצע בין התצפית הגבוהה ביותר לתצפית הנמוכה ביותר.

$$MR = \frac{X_{\min} + X_{\max}}{2}$$

MEDIAN - החציון

החציון הוא ערך שמחצית מהתצפיות קטנות או שוות לו ומחצית מהתצפיות גדולות או שוות לו .
ברשימה : נסדר את התצפיות בסדר עולה.

אם יש מספר אי זוגי של איברים מקומו של החציון יהיה התצפית שמיקומה : $\frac{n+1}{2}$

אם יש מספר זוגי של איברים החציון יהיה הממוצע של האיבר ה- $\frac{n}{2}$ והאיבר ה- $\frac{n}{2} + 1$

כלומר שיש מספר אי-זוגי של תצפיות החציון יהיה : $md = X_{\frac{n+1}{2}}$

ושיש מספר זוגי של תצפיות החציון יהיה : $md = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

בטבלת שכיחויות בדידה : נעשה תהליך דומה אך נעזר בשכיחות המצטברת.

דיאגרמת מקלות : נמיר לטבלת שכיחויות בדידה במטרה למצוא את החציון.

בטבלת שכיחויות במחלקות :

שלב א : נימצא את המחלקה החציונית שמיקומה יהיה $\frac{n}{2}$.

שלב ב : נציב בנוסחה הבאה : $Md = L_0 + \frac{\frac{n}{2} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$

$F(x_{m-1})$ - שכיחות מצטברת של מחלקה אחת לפני המחלקה החציונית.

$f(x_m)$ - השכיחות של המחלקה החציונית.

L_0 - גבול התחתון של המחלקה.

L_1 - גבול העליון של המחלקה.

היסטוגרמה : החציון הוא הערך על ציר ה- X שמחלק את ההיסטוגרמה לשני חלקים שווים בשטח.

כללי : החציון אינו רלבנטי למשתנה מסולם שמי ולא רלבנטי למשתנה איכותי.

הממוצע

הנו מרכז הכובד של ההתפלגות.

ברשימה : $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

בטבלת שכיחויות : $\bar{x} = \frac{\sum x \cdot f}{n}$

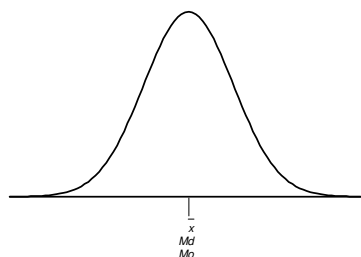
במחלקות : נשתמש באותה נוסחה רק נתייחס לאמצע המחלקה בתור ה- X . הממוצע הזה יהיה ממוצע מקורב.

כללי : הממוצע רלבנטי רק למשתנה כמותי.

מדדי המיקום המרכזי בהתפלגויות המיוחדות:

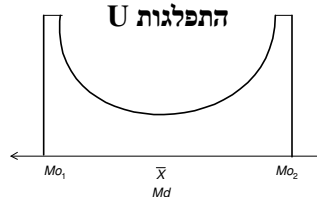
בהתפלגות סימטרית פעמונית כל מדדי המרכז שווים זה לזה :

התפלגות סימטרית



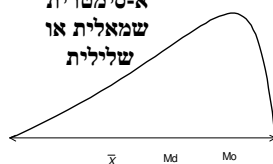
בהתפלגות סימטרית השכיח לא חייב להיות במרכז :

התפלגות U

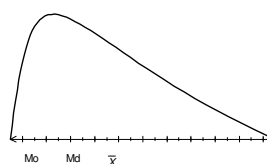


בהתפלגות אסימטרית

התפלגות א-סימטרית שמאלית או שלילית



התפלגות א-סימטרית ימנית או חיובית



תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6
חשב את החציון, השכיח, והממוצע של הציונים.

2. בדקו את מספר החדרים לדירה בבניין בן 5 דירות והתקבל ממוצע 3.8
לגבי 4 דירות נמצא מספר חדרים: 4, 3, 4, 5.
א. כמה חדרים יש בדירה החמישית?
ב. מהו השכיח ומהו החציון?

3. להלן התפלגות מספר מקלטי הטלויזיה שנספרו עבור כל משפחה בשוב מסוים:

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

ג. חשב את הממוצע, החציון והשכיח של ההתפלגות.
ד. הסבר ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא כולן) שלא היה להם עד היום טלוויזיה היו רוכשים מקלט אחד.

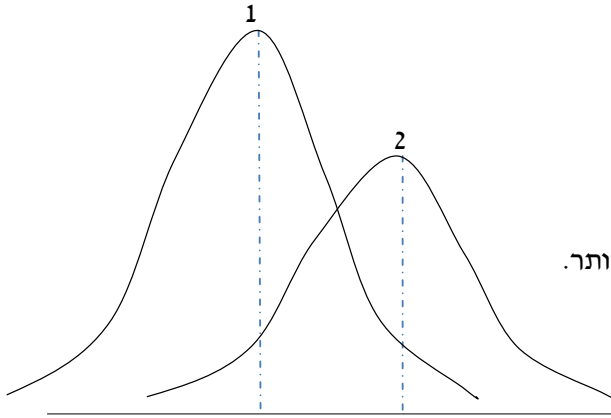
4. להלן התפלגות מספר המכוניות למשפחה בשוב "הגורן"

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

א. כמה משפחות יש בשוב?
ב. מה אחוז המשפחות בשוב עם לכל היותר 2 מכוניות?
ג. חשבו את הממוצע, החציון והשכיח.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

5. מורה לימד 2 כיתות, הוא תיאר באותה מערכת צירים את התפלגות הציונים בכל כיתה. בחר בתשובה הנכונה:



- א. בכיתה 1 השכיח גבוה יותר מכיתה 2.
- ב. בכיתה 2 השכיח גבוה יותר מכיתה 1.
- ג. בשתי הכיתות אותו שכיח.
- ד. לא ניתן לדעת באיזו כיתה השכיח גדול יותר.

6. ביישוב מסוים בדקו לכל משפחה את מספר הטלוויזיות שיש לה בבית. ביישוב גרות 200 משפחות. בממוצע יש למשפחה 1.5 טלוויזיות.

מספר טלוויזיות	מספר משפחות
0	28
1	62
2	
3	

- א. השלימו את הטבלה.
- ב. מהו השכיח, אמצע טווח והחציון.
- ג. חלק מהמשפחות להן הייתה טלוויזיה אחת בדיוק הוציאו את הטלוויזיה מביתם, כיצד כל מדד ישתנה (יגדל, יקטן או לא ישתנה) הסבירו ללא חישוב.

7. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

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

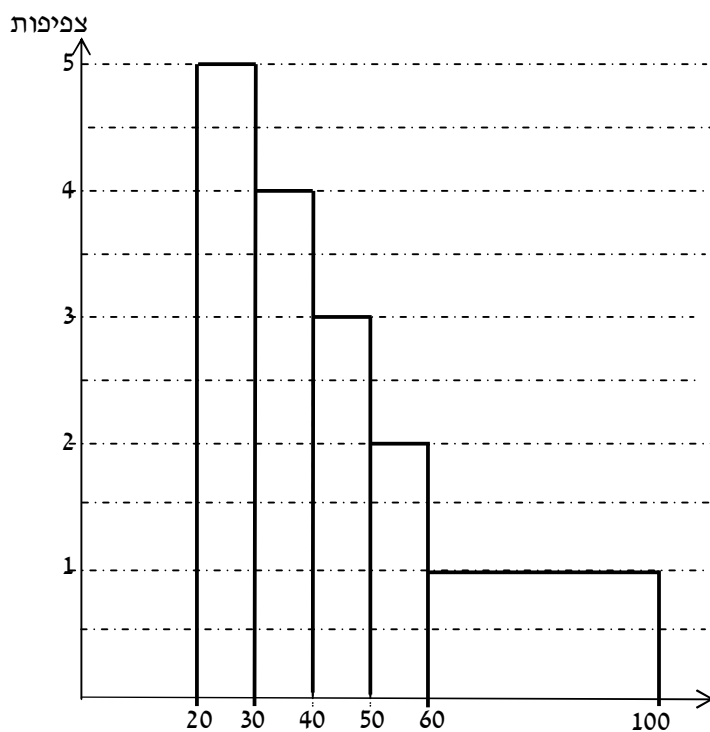
מה הממוצע והחציון של ההתפלגות?

8. להלן התפלגות הגבהים בס"מ בקבוצה מסוימת.

שכיחות	גובה בס"מ
30	150-160
40	160-170
60	170-175
70	175-180
40	180-190

חשב את הממוצע, החציון והשכיח של הגבהים בקבוצה זו.

9. בפקולטה מסוימת בדקו לסטודנטים העובדים בה את השכר לשעת עבודה. להלן התוצאות:



א. מצא את השכיח בהתפלגות.

ב. מצא את החציון בהתפלגות.

ג. הסבירו ללא חישוב האם הממוצע גדול/קטן/שווה לחציון.

ד. הסתבר שיש להוציא מספר תלמידים במחלקה בין 20-30 שקלים כיצד הדבר ישפיע על הממוצע, החציון והשכיח? הסבירו ללא חישוב.

פתרונות:**שאלה 1:**

החציון : 7

השכיח : 6

הממוצע : 6.9

שאלה 2:

א. 3

ב. שכיח : 3,4 חציון : 4

שאלה 3:

א. הממוצע : 1.7

החציון : 1.5

השכיח : 1

ב. הממוצע יגדל ויתר המדדים לא ישתנו.

שאלה 4:

א. 630

ב. 34.13%

ג. שכיח וחציון : 3

ממוצע : 2.952

שאלה 5:

תשובה : ב

שאלה 6:

ב חציון : 2 שכיח : 2 אמצע טווח : 1.5

שאלה 7:

חציון וממוצע : 55

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

רקע:

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

הטווחותחום RANGE:

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

שונות וסטיית תקן:

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

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

דוגמה : נחשב את השונות של סדרת המספרים הבאה : 5,4,9

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

להלן התפלגות הציונים בכיתה מסוימת בה ממוצע הציונים הוא 7.44

הציון X-	השכיחות f-	$x^2 \cdot f$
5	2	50
6	4	144
7	8	392
8	5	320
9	4	324
10	2	200

סה"כ

1430

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

$$s = \sqrt{s_x^2} = \sqrt{1.8464} = 1.3588$$

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6
חשבו את השונות, סטיית התקן והטווח של הציונים.

2. להלן התפלגות מספר המכוניות למשפחה ב"הגורן"

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

א. חשבו סטיית התקן.

ב. חשבו את הטווח של הנתונים.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

3. בחברה העוסקת בטלמרקטינג בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4 שנים וסטיית התקן היא שנתיים.

א. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים עם וותק של 4 שנים להתפלגות?

ב. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני

עובדים אשר אחד עם וותק של 0 שנים והשני עם וותק של 8 שנים להתפלגות?

4. נתונה רשימה של 5 תצפיות, אך רק עבור 4 מהן נרשמו הסטיות שלהן מהממוצע:

2, 3, 2, 1. חשב את השונות של חמש התצפיות.

5. בשכונה בדקו בכל דירה את מספר החדרים לדירה. בשכונה 200 דירות.

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

א. מה הממוצע של מספר החדרים לשכונה בדירה?

ב. חשבו את סטיית התקן של מספר החדרים לדירה.

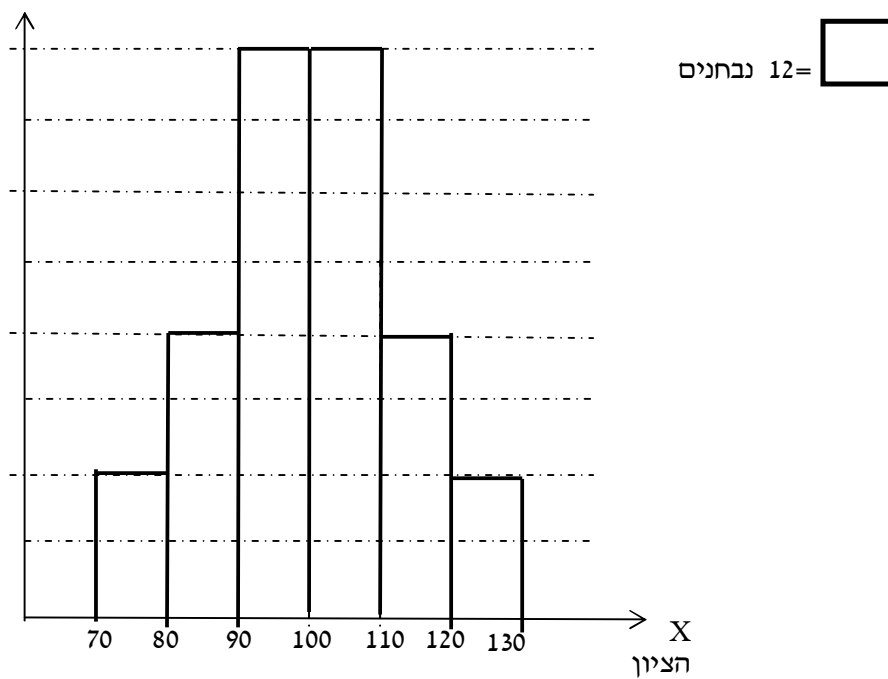
ג. חלק מבעלי הדירות בנו 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע (יקטין, יגדל, לא ישנה) כל מדד שחישבתם בסעיפים הקודמים.

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

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

מהי סטיית התקן של התפלגות המשקל?

7. להלן התפלגות הציונים במבחן אינטליגנציה:



- מה הממוצע ומה החציון של ההתפלגות?
- חשבו את סטיית התקן של הציונים.
- מסתבר שיש להוסיף 20 תצפיות לכל אחת משתי המחלקות 90-100 ו-100-110. כיצד הדבר ישתנה את כל אחד מהמדדים של הסעיפים הקודמים?

פתרונות :**שאלה 1:**

השונות : 2.19

סטיית תקן : 1.48

טווח : 6

שאלה 2:

א. סטיית תקן : 1.106

ב. טווח 4

שאלה 3:

א. ממוצע לא ישתנה, סטיית התקן תקטן.

ב. ממוצע לא ישתנה, סטיית התקן תגדל.

שאלה 4:

10.8

שאלה 5:

א. 3.05

ב. 1.16

שאלה 6:

7.73

שאלה 7 :

א. 100

ב. 12.96

פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - טווח בין-רבעוני

רקע:

הטווח הבין-רבעוני נותן את הטווח בין הרבעונים בו נמצאים 50% מהתצפיות המרכזיות.

שלבים במציאת טווח בין-רבעוני במחלקות:

F	f מספר עובדים (שכיחות)	$L_1 - L_0$ רוחב	מספר שנות ותק
56	56	4	0.5 – 4.5
106	50	5	4.5 – 9.5
154	48	2	9.5 – 11.5
190	36	3	11.5 – 14.5
200	10	5	14.5 – 19.5

שלב א: נימצא את הרבעון התחתון (האחוזון ה-25) והרבעון העליון (האחוזון ה-75).

מיקום הרבעון התחתון יהיה: $\frac{n}{4}$

מיקום הרבעון העליון יהיה: $\frac{3n}{4}$

נוסחאות הרבעונים יהיו:

$$Q_1 = L_0 + \frac{\frac{n}{4} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0) \quad ; \quad Q_3 = L_0 + \frac{\frac{3n}{4} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$$

שלב ב: נחסר את הרבעונים:

$$IQR = Q_3 - Q_1$$

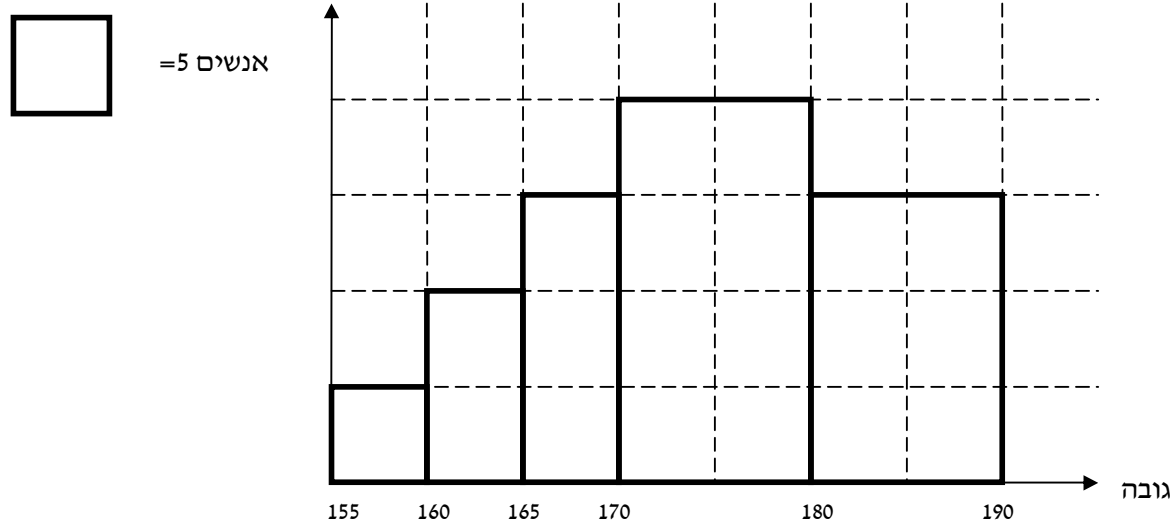
תרגילים:

1. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

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

מצא את הטווח הבין-רבעוני.

2. להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת:



מצא את הטווח הבין-רבעוני.

פתרונות:

שאלה 1:

13.75

שאלה 2:

13.33

פרק 7 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן

רקע:

המטרה למדוד איך תצפית ממוקמות יחסית לשאר התצפיות בהתפלגות.

ציון תקן:

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

ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע.

כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע.

ציון תקן חיובי אומר שהתצפית מעל הממוצע.

ציון תקן שלילי אומר שהתצפית מתחת לממוצע.

ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

דוגמה: (פתרון בהקלטה)

במקום עבודה מסוים ממוצע המשכורות 8 אלפי ₪ עם סטית תקן של 2 אלפי ₪ באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים עם סטית תקן של 1.5 שנים. ערן מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערן יותר באופן יחסי משכיל או משתכר ?

תרגילים

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

המקצוע	ממוצע	סטיית תקן
לשון	74	12
מתמטיקה	80	16

- עודד קיבל: 68 בלשון ו70 במתמטיקה.
א. באיזה מקצוע עודד טוב יותר באופן יחסי לשכבה שלו?
ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שיהיה שקול לציונו בלשון?

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

	תפוקה	מספר פועלים
ממוצע	48	15
סטיית תקן	10	2

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

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

פתרונות:

שאלה 1:

א. לשון

ב. 72

שאלה 2:

תשובה ב

שאלה 3:

א. משקל

ב. 70

פרק 8 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים במחלקות

רקע:

האחוזון (המאון) ה- p הוא הערך בנתונים המחלק את הנתונים בצורה כזאת שעד אליו יש $p\%$

מהנתונים. מסמנים את האחוזון ה- p ב- X_p .

למשל, המאון ה-25 הוא האחוזון ה-25 או הרבעון התחתון: ערך ש-רבע מהתצפיות קטנות

ממנו והשאר גבוהות ממנו. מסומן: $X_{0.25}$

מציאת מאון במחלקות:

שלב א: נימצא את המחלקה הרלבנטית שמיקומה יהיה $\frac{np}{100}$.

שלב ב: נציב בנוסחה הבאה: $x_p = L_0 + \frac{\frac{n \cdot p}{100} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$

$F(x_{m-1})$ - שכיחות מצטברת של מחלקה אחת לפני המחלקה הרלבנטית.

$f(x_m)$ - השכיחות של המחלקה הרלבנטית.

L_0 - גבול התחתון של המחלקה.

L_1 - גבול העליון של המחלקה.

אם רוצים לחלץ את אחוז התצפיות שמתחת לערך מסוים נשתמש בנוסחה הבאה:

$$P_x = \left[\frac{(x - L_0)}{(L_1 - L_0)} \cdot f(x_m) + F(x_{m-1}) \right] \cdot \frac{100}{n}$$

דוגמה: (פתרון בהקלטה)

להלן התפלגות השכר של עובדים בחברה מסוימת :

שכר בש"ח	$f(x)$
4000-6000	140
6000-10000	128
10000-15000	60
15000-20000	54
20000-40000	18

א. מצאו את המאון ה-40.

ב. מהו אחוז העובדים שמשכרים מתחת ל-5,000 ₪?

תרגילים:

1. להלן התפלגות השכר (באלפי שקלים) בחברה:

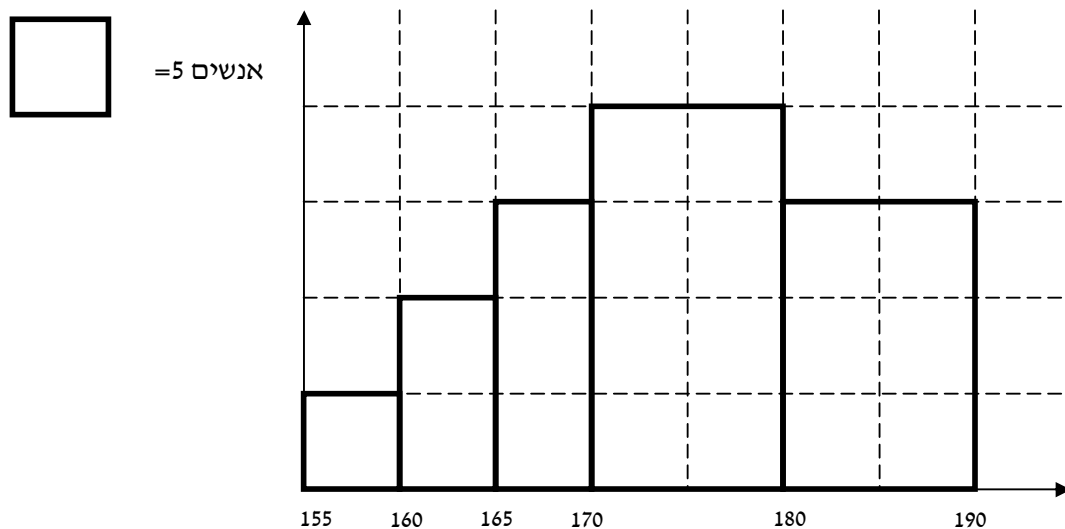
שכר X	שכיחות מצטברת
6-10	48
10-15	100
15-20	120
20-30	132
30-60	136

- א. חשבו את המאון ה-60.
- ב. מהו העשירון העליון?
- ג. 20% מהמשכורות הגבוהות ביותר הן משכורות של הבכירים, מהי המשכורת המינימאלית לבכיר?
- ד. מה אחוז האנשים שמשתכרים מתחת ל-7000 ₪?
- ה. איזה אחוז מהעובדים משתכרים מעל ל-25,000 ₪?
- ו. איזה אחוז מהעובדים משתכרים בין 7000 ל-25,000 ₪?

2. למבחן ניגשו 400 נבחנים. נתון שהעשירון התחתון הוא הציון 60. הרבעון העליון הוא הציון 80. כמו כן ההתפלגות של הציונים היא סימטרית. מלאו את השכיחות החסרות.

ציון X	$f(X)$
50-60	
60-70	
70-80	
80-90	
90-100	

3. להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת:



א. העשירון התחתון.

ב. האחוזון ה-30.

ג. הגובה ש-20% מהתצפית גדולות ממנו.

ד. את אחוז התצפיות מתחת לגובה 158 ס"מ.

ה. את אחוז התצפיות מעל לגובה 185 ס"מ.

ו. את אחוז התצפיות בין גובה 170 ס"מ ל-185 ס"מ.

פתרונות:**שאלה 1 :**

א. 13.23

ב. 22

ג. 17.2

ד. 8.82%

ה. 7.36%

ו. 83.82%

שאלה 3:

א. 162.5

ב. 170

ג. 183.33

ד. 3%

ה. 15%

ו. 55% ת

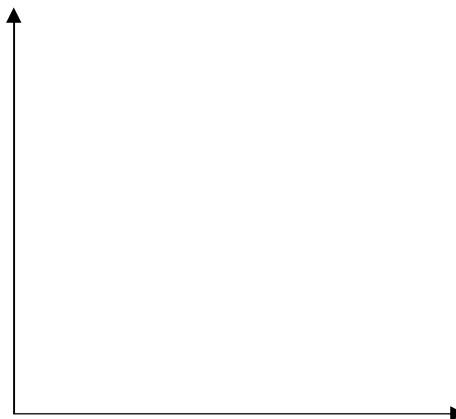
פרק 9 - מדדי קשר - מדד הקשר הלינארי (פירסון)

רקע:

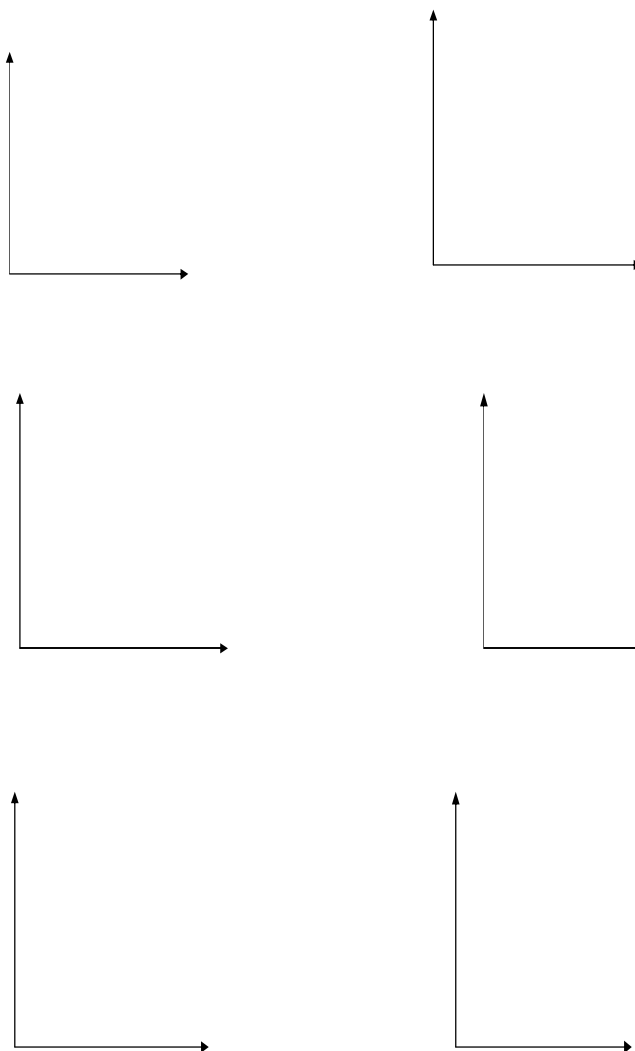
המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו Y הוא המשתנה המוסבר (התלוי). למשל, נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד X מסבירה את ההכנסה שלו Y . במקרה זה שנות ההשכלה 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) שבדוק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שנראה בשלב הראשון רק בעין.

המדד בודק את כיוון הקשר (חיובי או שלילי).

ואת עוצמת הקשר (חלש עד חזק).

מקדם מתאם זה מקבל ערכים בין -1 ל 1.

מקדם מתאם -1 או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי

$$y = bx + a$$

מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר לינארי מלא בו השיפוע b יהיה חיובי ואילו

מתאם שלילי מלא אומר שקיים קשר לינארי מלא בו השיפוע b שלילי (מקדם מתאם -1).

מתאם חיובי חלקי אומר שככל שמשנתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת

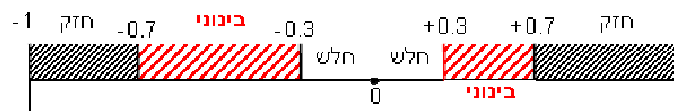
נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל

שמשנתנה 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.

$$COV(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} = \frac{\sum xy}{n} - \bar{x} \cdot \bar{y} : \text{שונות משותפת}$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 : \text{שונות של המשתנה X}$$

$$S_Y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2 : \text{שונות המשתנה Y}$$

$$r_{xy} = \frac{COV(x, y)}{s_x \cdot s_y} : \text{מקדם המתאם הלינארי}$$

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.

מספר חיסורים	ציון
2	80
1	90
0	90
2	70
3	70
4	50

- א. שרטט דיאגרמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר בין 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:

x	y
10	12
14	15
15	15
18	17
20	21

- א. מה הממוצע של כל רמת הורמון?
- ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

3. נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות :

$$\sum_{i=1}^{20} Y_i = 200 \quad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 76 \quad \sum_{i=1}^{20} (X_i - \bar{X})^2 = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y_i - \bar{Y}) = 60.8$$

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 :

$$\sum_{i=1}^{20} Y_i = 200 \quad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} Y_i^2 = 2080 \quad \sum_{i=1}^{20} X_i^2 = 2960$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

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 הייתה 2. מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם ?

6. להלן רשימת טענות, לגבי כל טענה קבע נכון/לא נכון ונמק!

א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.

ב. לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$ $S_x = S_y = 1$ לכן מדד הקשר של פירסון יהיה 1.

ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות אמריקאיות:

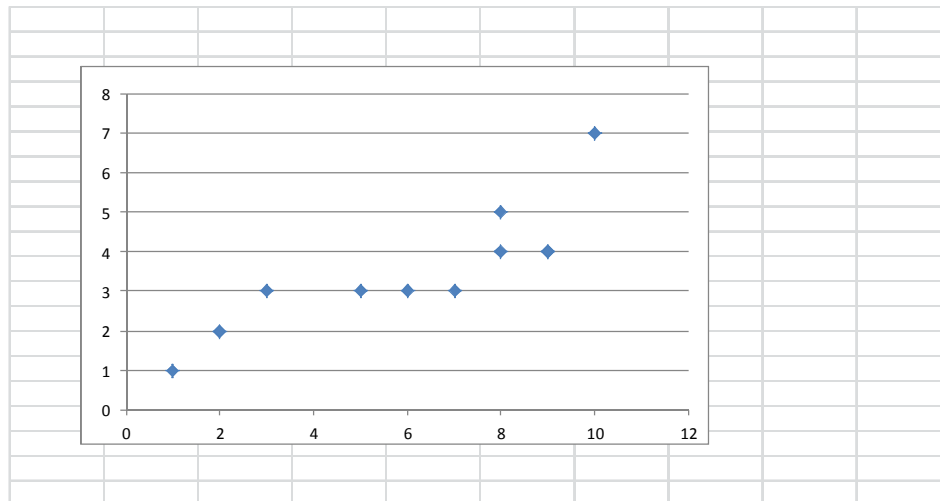
7. נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן :

- 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 דיאגרמת 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. $\bar{y} = 16$ $\bar{x} = 15.4$ 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 10 - מדדי קשר - רגרסיה ליניארית

רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבויי 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:

$$\sum_{i=1}^{20} Y_i = 200 \quad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 76 \quad \sum_{i=1}^{20} (X_i - \bar{X})^2 = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y_i - \bar{Y}) = 60.8$$

- 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?

2. נסמן ב- X את ההשכלה של אדם בשנות למוד. נסמן ב- Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_y = 5 \quad S_x = 2$$

$$\bar{Y} = 8 \quad \bar{X} = 14$$

$$COV(X, Y) = 7.5$$

- 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. על פי משוואת הרגרסיה תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון?
 ג. מהו קו הרגרסיה לניבוי הציון לפי שעות ההכנה?

4. נתונים 2 משתנים Y, X . כמו כן נתון: X ממוצע = 1.5, שונות X = שונות Y = 4, וכן שקו הרגרסיה של Y על בסיס X הינו $Y = -0.2X + 0.5$. חשב מהו מקדם המתאם בין X ל- Y ?

פתרונות:**שאלה 1:**

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. $y = 1.2x + 29$

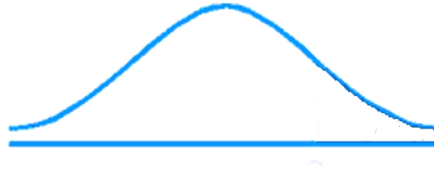
שאלה 4:

-0.2

פרק 11 - 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.

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

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

כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלבנטיים 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:

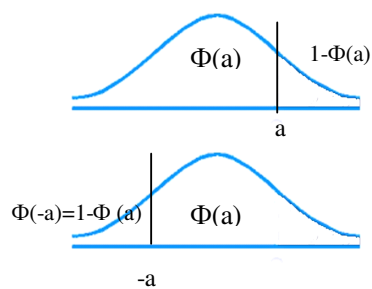
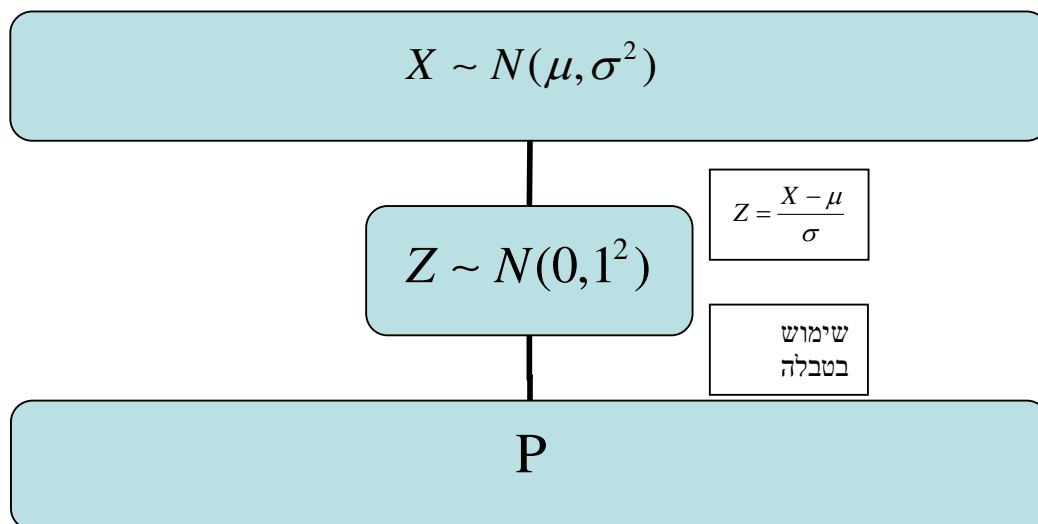
$$Z = \frac{X - \mu}{\sigma}$$

אחרי תקנון מקבלים ערך הנקרא ציון תקן.

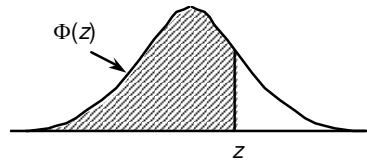
ציון התקן משמעו בכמה סטיות תקן הערך סוטה מהממוצע.

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 $\Phi(z)$



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

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

דוגמה: (הפתרון בהקלטה)

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

- 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 ש-90% מהחפיסות בקו הייצור שוקלים פחות מהם?

תרגילים:

1. הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטיית 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.

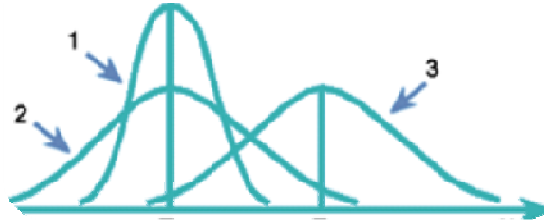
- א. מה אחוז האנשים שמשקלם נמוך מ- 55 ק"ג?
- ב. מהי פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
- ג. מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל- 70 ק"ג?
- ד. לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ- 4 ק"ג?
- ה. מה הסיכוי שאדם אקראי ישקול מתחת ל- 140 ק"ג?

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 95.
- ג. מצאו את העשירון התחתון.

5. ציוני מבחן אינטליגנציה מתפלג נורמלית עם ממוצע 100 ושונויות 225 .
- א. מה העשירון העליון של הציונים במבחן האינטליגנציה?
 - ב. מה העשירון התחתון של ההתפלגות?
 - ג. מהו הציון ש- 20% מהנבחנים מקבלים מעליו?
 - ד. מהו האחוזון ה- 20?
 - ה. מהו הציון ש- 5% מהנבחנים מקבלים מתחתיו?
6. נפח משקה בבקבוק מתפלג נורמלית עם סטיית תקן של 20 מ"ל, נתון ש-33% 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 אורך חיי מכשיר?
 - ה. 1% מהמכשירים בעלי אורך החיים הקצר ביותר נשלח למעבדה לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשליחת מכשיר למעבדה?

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



א. לאיזו התפלגות הממוצע הגבוה ביותר?
ב. במה מבין המדדים הבאים התפלגות 1 ו 2 זהות?

- 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 "טריירה" מתפלגת נורמלית עם ממוצע של 2000 דולר וסטיית 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 12 - התפלגות הדגימה

א. התפלגות ממוצע המדגם ומשפט הגבול המרכזי

1. להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב מסוים:

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

נגדיר את x להיות מספר המקלטים של משפחה אקראית.

א. בנו את פונקציית ההסתברות של x .

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

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

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

3. 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 שביום מסוים נולדו 4 תינוקות.

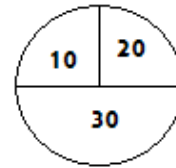
ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?

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?

4. הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועיות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך פתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלי.
- א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
- ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה פחות ממנו?
- ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
5. נפח היין בבקבוק מתפלג נורמאלי עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.
- א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ"ק?
- ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ"ק?
- ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ"ק?
- ד. בקבוקי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?
6. משתנה מתפלג נורמאלי 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.

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



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.

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

8. 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 ₪?

9. מטילים קובייה 50 פעמים בכל פעם מתבוננים בתוצאה של הקובייה. מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72 ב-50 ההטלות?

10. 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.

התפלגות מספר ההצלחות במדגם ופרופורציות ההצלחות במדגם וקירוב נורמאלי

להתפלגות הבינומית

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. הסיכוי שמבוגר יחלה בשפעת בחורף מוערך ב-0.4. נדגמו 30 מבוגרים אקראיים.
 - א. חשבו את הסיכוי שבדיוק 15 יחלו בשפעת (1) לפי ההתפלגות הבינומית ; (2) לפי הקירוב הנורמאלי.
 - ב. חשבו את הסיכוי לכל היותר 14 יחלו בשפעת.
 - ג. חשבו את הסיכוי שלפחות 11 יחלו בשפעת.
3. במפעל 10% מהמוצרים פגומים. נלקחו 100 מוצרים באקראי מקו הייצור.
 - א. מה ההסתברות שנדגמו לכל היותר 6 מוצרים פגומים?
 - ב. מה ההסתברות שמספר המוצרים הפגומים יהיה לכל היותר 11 במדגם?
 - ג. מה ההסתברות שיהיו בדיוק 11 פגומים?
4. 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?
5. במדינה יש 10% של אבטלה. 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?
6. נניח ש-30% מהאוכלוסייה תומכת 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.

פתרונות:**פרק א' - התפלגות ממוצע מדגם ומשפט הגבול המרכזי****שאלה 1**

א.

$$E(\bar{X}) = 3.5$$

$$\sigma(\bar{X}) = 1.208$$

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

$$\sigma = 0.973 \quad \sigma^2 = 0.9475 \quad \mu = 2.05 \quad \text{ב.}$$

$$V(\bar{X}) = 0.237 \quad E(\bar{X}) = 2.05 \quad \text{ג.}$$

$$\sigma(\bar{X}) = 0.487$$

שאלה 3

א. 0.8413

ב. 0.0013

ג. 0

ד. 0.1974

שאלה 4

א. 0.0465

ב. 32.29

ג. 0.2628

שאלה 5

א. 0

ב. 0.1587

ג. 0.1587

ד. 0.5

שאלה 6

א. 0.5468

ב. 0.6826

שאלה 7

א.

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

ב. התוחלת: 22.5

השונות: 68.75

ג. התוחלת: 22.5

השונות: 13.75

ד. 0.8997

שאלה 8

0.0475

שאלה 9

0.1814

שאלה 10

א. 0.9772

ב. 0.0228

ג. 271

**פרק ב' - התפלגות מספר ההצלחות במדגם ופרופורציות ההצלחות במדגם וקירוב
נורמאלי להתפלגות הבינומית**

<u>שאלה 1</u>	<u>שאלה 2</u>
א. 0.201	א. 0.0783 (1) ; 0.0794 (2)
ב. 0.3758	ב. 0.8238
ג. התוחלת : 2	ג. 0.7123
סטית התקן : 1.2649	
<u>שאלה 3</u>	<u>שאלה 4</u>
א. 0.121	0.1611
ב. 0.6915	
ג. 0.124	
<u>שאלה 5</u>	<u>שאלה 6</u>
א. התוחלת : 0.1	א. 0.0618
השונות : 0.00064	ב. 0.0618
ב. 0.5	ג. 0.8238
ג. 0.3446	

פרק 13 - אמידה נקודתית

א. אומדים חסרי הטיה

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

$$T_1 = \frac{X_1 + \dots + X_5}{5} \quad \text{חוקר א' הציע:}$$

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

$$T_3 = \frac{2X_1 + X_3}{2} \quad \text{חוקר ג' הציע:}$$

- (א) איזה מן האומדים הוא חסר הטיה?
 (ב) במדגם התקבלו הציונים הבאים: 65, 78, 58, 82, 100. חשבו את האומדנים המתקבלים עבור האומדים חסרי הטיה.
 (ג) איזה מבין שני האומדים חסרי ההטיה עדיף? נמקו.

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

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

- (א) בדקו לגבי כל אומד אם הוא בלתי מוטה.
 (ב) איזה אומד עדיף? נמקו.
3. $X \approx B(n, p)$ 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- $E(X)$.

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

פונקציית ההסתברות של X – מספר המניות שיעלו ביום מסוים :

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

- א. מצאו 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 התלוייה בפרמטר θ באופן הבא :

הסיכוי שמטפלת תטפל בתינוק אחת 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 שתיים מהם מטפלות בתינוק אחד בלבד, אחת מהן בשתיים 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.

ב. אומדי נראות מכסימלית

1. הסיכוי של שחקן לנצח במשחק הוא p (לא ידוע). השחקן משחק במשחק עד אשר הוא מנצח בפעם הראשונה. נתון שהשחקן ניצח לראשונה רק במשחק השני.
 - (א) חשבו את פונקציית הנראות של p וציירו גרף שלה.
 - (ב) מצאו אומדן נראות מכסימלית עבור p .
 - (ג) מצאו אומדן נראות מכסימלית ל- p אם ביום אחד הוא נאלץ לשחק 4 פעמים וביום אחר הוא נאלץ לשחק 5 פעמים עד אשר ניצח.
2. מספר הלקוחות שנכנסים לחנות מסוימת, מתפלג פואסונית עם תוחלת של λ לקוחות ביום.
 - (א) מצאו אומד נראות מכסימלית ל- λ על סמך מספר הלקוחות שנכנסים ביום מסוים.
 - (ב) מצאו אומד נראות מכסימלית ל- λ על סמך מספר הלקוחות שנכנסים ב- n ימים מסוימים.
3. הזמן שלוקח לאדם לחכות בתור מתפלג מעריכית עם פרמטר λ . דגמו 4 אנשים מקריים שחיכו בתור ומדדו את זמני ההמתנה שלהם. התוצאות שהתקבלו בדקות הן: 3, 7, 5 ו-3.
 - (א) פתחו אומד נראות מכסימלית לפרמטר זה על סמך n תצפיות כלשהן.
 - (ב) מהו האומדן לפרמטר?
4. משך זמן הכנת שיעורי הבית (בשעות) של בני נוער ביום אחד מתפלג אחיד $U(0, \theta)$. כדי לאמוד את θ , נשאלו ביום מסוים מספר בני נוער כמה שעות הם הכינו שיעורי בית באותו יום.
 - (א) אלעד הכין ביום מסוים שיעורי בית במשך שעה שלמה. חשבו את פונקציית הנראות של θ המתבססת על תצפית זו, וציירו את הגרף שלה.
 - (ב) מצאו אומדן נראות מכסימלית ל- θ על סמך התצפית.
 - (ג) משכי הכנת שיעורי בית (שעות) של 3 בני נוער היו 1.5, 3, 1. מצאו אומדן נראות מכסימלית ל- θ על סמך המדגם הזה.
 - (ד) מצאו באופן כללי אומד נראות מכסימלית ל- θ על סמך מדגם של n בני נוער - $X_1, \dots, X_n$.

5. הגובה של אוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת ידועה של 170 ס"מ ושונות σ^2 לא ידועה.

- א. מצאו אומד נראות מכסימלית עבור השונות על סמך מדגם $X_1, \dots, X_n$ של n תצפיות מהאוכלוסייה.
- ב. נדגמו 5 אנשים בלתי תלויים בעלי הגבהים: 170, 182, 174, 165, 174. מהו האומדן לשונות הגבהים באוכלוסייה?

6. פתחו אומד נראות מכסימלי לפרמטר P בהתפלגות הבינומית על סמך מדגם בגודל n בו X הוא מספר ההצלחות במדגם.

7. X הוא משתנה מקרי בעל פונקצית הצפיפות:

$$f(x) = \begin{cases} 2\theta x e^{-\theta x^2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

- א. מצאו אומד נראות מכסימלית ל- θ על סמך n תצפיות בלתי תלויות $X_1, \dots, X_n$.
- ב. מצאו אומד נראות מכסימלית ל- θ^2 .

8. בכד א 10 כדורים שחורים ו 10 לבנים בכד ב 5 כדורים שחורים ו- 15 לבנים. דוגמים באקראי כדור אך אינך יודע מאיזה כד.

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

ב) מהו האומדן אם הצבע הוא שחור?

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

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

ב) מה אומדן הנראות המכסימלי לסיכוי שייקח לו לפחות חצי שעה לפתור את התשבץ הבא?

10. מספר הלקוחות הממתינים בתור במוקד טלפוני הוא משתנה מיקרי X בעל התפלגות התלויה בפרמטר θ באופן הבא :

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

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

- (א) מצאו אומדן בשיטת הנראות המכסימלית עבור הפרמטר θ על-סמך המדגם הנתון.
- (ב) מצאו אומדן בשיטת הנראות המכסימלית לסיכוי שלא יהיו לקוחות בתור.

11. אדם מחזיק בידו שני מטבעות : מטבע הוגן ומטבע שאינו הוגן שהסיכוי בו לתוצאה עץ הוא 0.2. האדם מטיל את אחד המטבעות פעמיים ומודיע לך כמה פעמים הוא קיבל עץ. אתה צריך לנחש איזה מטבע הוא הטיל : את ההוגן או זה שאינו הוגן.
- (א) מצא אומדן בשיטת הנראות המכסימלית לסוג המטבע שהוטל.
- (ב) מהו האומדן אם האדם קיבל פעמיים עץ?

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

- (א) מצא אומדן נראות מכסימלית לשיעור המובטלים באוכלוסייה.
- (ב) מצא אומדן לשיעור העובדים באוכלוסייה.
- (ג) מצא אומדן ליחס בין שיעור העובדים לשיעור המובטלים באוכלוסייה.

ג. שאלות מסכמות לאומדים חסרי הטיה ונראות מכסימלית

1. במפעל מייצרים מוצרים בשלוש מכונות שונות ובלתי תלויות. במכונה הראשונה הסיכוי שמוצר יהיה תקין הוא P , במכונה השנייה ההסתברות שמוצר יהיה תקין הוא P^2 ובמכונה השלישית הסיכוי הוא $2P$. דוגמים 20 מוצרים מכל מכונה. נסמן ב- X את מספר המוצרים התקינים שיוצרו במכונה א'. נסמן ב- Y את מספר המוצרים התקינים שיוצרו במכונה השנייה וב- Z את מספר המוצרים התקינים שיוצרו במכונה השלישית.
 - (א) מהם הערכים האפשריים של הפרמטר P .
 - (ב) מצאו אומד בלתי מוטה עבור הפרמטר P על סמך X ו- Z .
 - (ג) אם התקבל ש- $Y=3$, $X=6$ מהו אומדן נראות מכסימלית ל- P .
2. מספר תאונות הדרכים בקטע כביש א' מתפלג פואסונית עם קצב של λ תאונות בחודש. מספר תאונות הדרכים בקטע כביש ב' מתפלג פואסונית עם קצב של 2λ תאונות בחודש. הוחלט לספור את מספר התאונות בחודש בכל אחד מקטעי הכביש.
 - (א) נסמן ב- X את מספר התאונות בחודש בקטע א' ו ב- Y בקטע ב'. מצאו אומד נראות מכסימלי לפרמטר λ על סמך X ו- Y .
 - (ב) מצאו אומד נראות מכסימלי לסיכוי שבקטע כביש א' תהיה לפחות תאונה אחת בחודש?
 - (ג) האם האומד שמצאת בסעיף א' הוא חסר הטיה ל- λ ?
3. זמן הייצור של מוצר מסוים בתהליך ייצור מתפלג נורמאלי עם תוחלת ושונות שאינן ידועות.
 - (א) הציעו אומדים חסרי הטיה לתוחלת והשונות של זמן הייצור של המוצר.
 - (ב) הציעו אומדי נראות מכסימלי לתוחלת ולשונות של זמן הייצור של המוצר.
 - (ג) הציעו אומד נראות מכסימלי לריבוע התוחלת של זמן הייצור.
 - (ד) האם האומד מהסעיף הקודם הוא גם חסר הטיה?

4. בקזינו משחק בו 4 תאים ממוספרים מ 1 עד 4. מפעיל המשחק שם כסף באחד מארבעת התאים והאדם המשתתף צריך לנחש באיזה תא הכסף מוחבא. מפעיל הקזינו מודיע שהסיכוי להחביא את הכסף בכל אחד משלושת התאים הראשונים שווה אך לא בהכרח שווה לסיכוי להחביא אותו בתא הרביעי.

יש לאמוד את הסיכוי להחביא את הכסף בתא הראשון : P .

(א) מצא את תחום ההגדרה של הפרמטר P .

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

(ב) מצאו אומדן ל- P על סמך התוצאות הללו בשיטת הנראות המכסימלית.

(ג) מצאו אומדן חסר הטיה ל- P מהו האומדן לפי התוצאות של יעל?

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

ד. קריטריון MSE - תוחלת ריבוע הטעות

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 $P=0.1$?

תשובות סופיות - אמידה נקודתית

פרק א' - תשובות לאומדים חסרי הטייה

שאלה 1

ה. $T_2 T_1$

ו. $T_2 = 110 T_1 = 76.6$

ז. T_1

שאלה 4

א. $\frac{3x}{2}$

ב. $\frac{3\bar{x}}{2}$

שאלה 3

א. $\frac{x}{n}$

ב. $1 - \frac{x}{n}$

ג. x

שאלה 5

א. $1 - \frac{x}{2}$

ג. 0.125

ה. לשונות 0.917

פרק ב' - תשובות לאומדי נראות מכסימליתשאלה 2א. X ב. $\bar{X}$ שאלה 4

א. 1

ג. 3

ד. $X_{\max}$ שאלה 6

$$\frac{x}{n}$$

שאלה 8

ב. כד א

שאלה 10

א. 0.45

ב. 0.81

שאלה 12

א. 0.08

ב. 0.92

ג. 11.5

שאלה 1

ב. 0.5

ג. $\frac{2}{9}$ שאלה 3א. $\frac{1}{\bar{X}}$ ב. $\frac{2}{9}$ שאלה 5

$$\frac{\sum_{i=1}^n (x_i - 170)^2}{n}$$

ב. 40.2

שאלה 7

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

$$\left(\frac{n}{\sum X_i^2}\right)^2$$

שאלה 9

א. 32

ב. 0.3916

שאלה 11

ב. הוגן

פרק ג' - תשובות לשאלות המסכמות לאומדים חסרי הטיה ונראות מכסימלית.

שאלה 1

א. $0.5 \leq P \leq 1$

ג. 0.345

שאלה 4

א. $0 \leq P \leq \frac{1}{3}$

ב. $\frac{1}{3}$

ג. 0.389

שאלה 2

א. $\frac{x+y}{3}$

ג. כן

פרק ד' - תשובות ל-MSE

שאלה 1

זה עם השלוש תצפיות

שאלה 3

א. T_1

ב. T_2

ג. T_1

שאלה 2

א. $2x$

ג. סעיף ב

נספח : אומדי נראות מכסימלית ואומדים חסרי הטיה בהתפלגויות השונות

מודל בינומי

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

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

מודל אחיד (בדיד)

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

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

מודל פואסוני

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

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

מודל גיאומטרי

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

א.נ.מ עבור p הוא $\hat{p} = \frac{1}{\bar{X}}$ אינו א.ח.ה. וא.נ.מ עבור התוחלת $\frac{1}{p}$ הוא $\bar{X}$ והנו א.ח.ה.

מודל נורמלי

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

א.נ.מ עבור μ הוא $\hat{\mu} = \bar{X}$

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

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

אומד חסר-הטיה עבור σ^2 :

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

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ כאשר } \mu \text{ לא-ידוע}$$

מודל מעריכי

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

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

מודל אחיד (רציף)

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

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

בכל התפלגות:

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

אומד חסר-הטיה עבור σ^2 :

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

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ כאשר } \mu \text{ לא-ידוע}$$

פרק 14 - רווחי סמך

א. אמידה נקודתית (אומדים חסרי הטיה)

1. 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.
2. נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים. להלן התוצאות שהתקבלו :
 $\sum_{i=1}^{20} X_i^2 = 1502.2$ $\sum_{i=1}^{20} X_i = 162$.
 א. אמדו את השכר הממוצע של השכירים במשק.
 ב. אמדו את סטיית התקן של שכר השכירים במשק.

ב. רווח סמך לתוחלת

ב.1. שונות האוכלוסיה ידועה

1. 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.
 ד. מהי טעות התקן של האומד?
2. 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.

3. מעוניינים לאמוד את המשקל הממוצע של רכיב מסוים. ידוע שהמשקל מתפלג נורמאלי עם סטיית תקן 12.
- א. מהו גודל המדגם המינימאלי שיש לקחת אם מעוניינים ששגיאת האמידה המקסימאלית תהיה 2 גרם ברמת סמך של 95%.
- ב. ביצעו מדגם שאת גודלו מצאתם בסעיף א והתקבל ממוצע של 23 גרם. בנו רווח סמך ברמת סמך של 95% למשקל הממוצע של הרכיב.
4. 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.

5. משתנה מקרי מתפלג נורמאלי עם סטיית תקן ידועה 12. מה צריך להיות גודל המדגם כדי לבנות רווח סמך ברמת סמך של 98% שאורכו לא יעלה על 2?

2.ב. שונות האוכלוסיה לא ידועה

1. זמן התגובה מתפלג נורמאלי. במטרה לאמוד את תוחלת זמן התגובה נדגמו 4 תצפיות. להלן התוצאות בשניות: 3, 4, 5, 6, 7, 4. בנו רווח סמך ברמת סמך של 95% לממוצע זמן התגובה באוכלוסיה.
2. ציוני מבחן אינטליגנציה מתפלגים נורמאלי. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 ו סטיית תקן מדגמית 13.
- א. בנו רווח סמך לממוצע הציונים באוכלוסיה ברמת ביטחון של 95%.
- ב. חזרו על סעיף א' אם סטיית התקן הנתונה היא סטיית התקן האמיתית של כלל הנבחנים.
- ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.

3. 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.

ג. רווח סמך לפרופורציה

1. במטרה לאמוד את אחוז המובטלים במשק נדגמו 200 אזרחים. מתוכם התקבל ש 24 היו מובטלים.
 - א. בנו רווח סמך לאחוז המובטלים באוכלוסיה ברמת סמך של 95%.
 - ב. מהו האומד לטעות התקן?
 - ג. מה יקרה לאורך רווח הסמך אם הוא יבנה ברמת סמך של 99%?
2. נבנה רווח סמך לאחוז בעלי רישיון נהיגה באוכלוסיה (69%, 79%) הוא התבסס על מדגם שכלל 400 נשאלים.
 - א. לכמה נשאלים לא היה רישיון נהיגה במדגם?
 - ב. באיזו רמת סמך נבנה רווח הסמך?
3. מעוניינים לאמוד את הרייטינג של תוכנית מסוימת. 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 שאחוז הרייטינג של התוכנית לא אמור לעלות על 20%.
4. 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 באוכלוסיה.
 - ב. מצאו רווח סמך ברמה של 99% לאחוז הלא נשואות באוכלוסיה.
5. הממשלה אומדת מדי חודש את אחוז התמיכה בה. מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמיתי באוכלוסיה ביותר מ-3%, וזאת בביטחון של 95%?

ד. רווח סמך לשונות וסטיית תקן

1. זמן התגובה מתפלג נורמאלי. במטרה לאמוד את תוחלת זמן התגובה נדגמו 4 תצפיות. להלן התוצאות בשניות: 3.5, 4.6, 5.2, 4.7. בנו רווח סמך, ברמת סמך של 95% לשונות זמן התגובה באוכלוסיה.
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 נורמאלי:
 - א. בנו רווח סמך לתוחלת הטמפ' בחודשים אלה בצהרי ת"א ברמת סמך של 95%.
 - ב. בנו רווח סמך לסטיית התקן של הטמפ' בחודשים אלה בצהרי ת"א ברמת סמך של 95%.

ה. רווח סמך להפרשים

ה.1. הפרשי פרופורציות

1. 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?
2. 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.

ה.2. הפרשי תוחלות (שונות ידועות)

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. 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%.

תשובות סופיות - רווחי סמך

פרק א' - אמידה נקודתית

שאלה 2

א. 8.1

ב. 3.16

שאלה 1

ג. 177.9

ד. 64.1

ה. 0.4

פרק ב' - רווח סמך לתוחלת

שונות האוכלוסיה ידועה

<u>שאלה 2</u>	<u>שאלה 1</u>
א. 102	א. $223.42 < \mu < 236.58$
ב. 3	ב. $222.16 < \mu < 237.84$
ג. 0.9544	ד. 4
<u>שאלה 4</u>	<u>שאלה 3</u>
א. $3.58 < \mu < 4.42$	א. 139
ב. יקטן פי 2	ב. $21 < \mu < 25$
ג. גדל	
	<u>שאלה 5</u>
	780

שונות האוכלוסיה לא ידועה

<u>שאלה 2</u>	<u>שאלה 1</u>
א. $96.63 < \mu < 107.37$	$4.39 < \mu < 5.51$
ב. $96.90 < \mu < 107.10$	
	<u>שאלה 3</u>
	$3.149 < \mu < 3.351$

פרק ג' - רווח סמך לפרופורציה

<u>שאלה 1</u>	<u>שאלה 2</u>
א. $7.5\% < p < 16.5\%$ ב. 2.29% ג. יגדל	א. 104 ב. 0.9774
<u>שאלה 3</u>	<u>שאלה 4</u>
א. 423 ב. 271	א. $22.5\% < p < 30.9\%$ ב. $45.91\% < p < 60.72\%$
<u>שאלה 5</u>	
1068	

פרק ד' - רווח סמך לשונות וסטיית תקן

<u>שאלה 1</u>	<u>שאלה 2</u>
$0.039 < \sigma^2 < 1.708$	א. $30.285 < \mu < 31.315$ ב. $0.837 < \sigma < 1.607$

פרק ה' - רווח סמך להפרשים**הפרשי פרופורציות**

<u>שאלה 1</u>	<u>שאלה 2</u>
(33%, 47%)	א. (4.93%, 14.93%) ב. (22.5%, 31.8%)

הפרשי תוחלות

<u>שאלה 1</u>	<u>שאלה 2</u>
(-20, 90)	א. (-372, 772) ב. (6467, 7133)

פרק 15 - טעויות ועוצמה בבדיקת השערות כללית

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

$$p(X = k) = \frac{1}{n} \quad \text{עבור } k = 1, 2, \dots, n$$

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 "בינוני" וגרוע 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:

$$H_0: X \sim t(5) \quad (\text{התפלגות } t \text{ עם 5 דרגות חופש})$$

$$H_1: X \sim Z \quad (\text{התפלגות נורמאלית סטנדרטית})$$

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.

תשובות סופיות - טעויות ועוצמה בבדיקת השערות כללית

<u>שאלה 1</u>	<u>שאלה 2</u>
	$\beta = \frac{2}{5} \quad \alpha = \frac{1}{3}$
<u>שאלה 3</u>	<u>שאלה 4</u>
$\beta = 0.5 \quad \alpha = 0.25$	א. $\beta = 0.8 \quad \alpha = 0.2$ ב. $\beta = 0.3 \quad \alpha = 0.2$ ג. כלל ב'.
<u>שאלה 5</u>	<u>שאלה 6</u>
ב. 0.00781 ג. 0.1678	א. 0.05 ב. 0.022
<u>שאלה 7</u>	
א. 0.0228 ב. 0.0918 ג. 0.9082	

פרק 16 - בדיקת השערות - מבחנים פרמטרים

א. בדיקת השערות על תוחלת

שונות האוכלוסיה ידועה

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%?

שונות האוכלוסיה לא ידועה

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. ציוני מבחן IQ מתפלגים נורמלית בארצות הברית עם ממוצע 100. במדגם על 20 נבחנים ישראלים התקבלו התוצאות הבאות.

$$\bar{x} = 102$$

$$\hat{S} = 16$$
 האם בישראל ממוצע הציונים הוא כמו בארצות הברית? בדקו ברמת מובהקות של 2% הניחו הנחות אם יש צורך.

ב. בדיקת השערות על פרופורצייה

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%?

ג. בדיקת השערות להפרש פרופורציות

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?

ד. בדיקת השערות להפרש תוחלות במדגמים בלתי תלויים

שוניות האוכלוסייה ידועות

1. 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%.

2. 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. האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית בדקו ברמת מובהקות של 10%. מה ההנחה הדרושה לצורך פתרון?

שוניות האוכלוסייה לא ידועות (בהנחה שהן שוות)

1. במדגם של 10 ישראלים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות:

$$\begin{aligned}n_1 &= 10 \\ \sum x_i &= 1020 \\ \sum x_i^2 &= 105120\end{aligned}$$

במדגם של 14 אמריקאים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות:

$$\begin{aligned}n_1 &= 14 \\ \sum x_i &= 1386 \\ \sum x_i^2 &= 138644\end{aligned}$$

בדקו האם קיים הבדל בממוצע הציונים בבחינת ה-IQ בין ישראל לארה"ב. רשמו את כל ההנחות הדרושות. ברמת מובהקות של 5%.

2. להלן תוצאות מדגם הבדק אורך חיים של נורות מסוג W60 ומסוג W100. אורך החיים נמדד בשעות.

1-100W	2-60W	הקבוצה
956	1007	$\bar{x}$
72	80	$\hat{S}$
15	13	n

בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפיתרון.

ה. בדיקת השערות על מדגמים מזווגים

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.

ו. בדיקת השערות על שונות

1. 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. ברמת מובהקות של 5% בדקו האם סטיית התקן של זמן החלמה של התרופה החדשה נמוכה מהתרופה המקורית.

2. הגובה של אוכלוסייה מסוימת נחשב כמתפלג נורמלית על 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.

ז. בדיקת השערות על שתי שונויות

1. מעוניינים להשוות בין נשים וגברים מבחינת השונות בזמנים שלהם לבצע משימה מסוימת. במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמני ביצוע המשימה:

$$\sum (y_i - \bar{y})^2 = 204$$

במדגם של 13 נשים התקבלו התוצאות הבאות:

$$\sum (x_i - \bar{x})^2 = 200$$

בדקו ברמת מובהקות של 2% האם קיים הבדל בין השונויות? מה יש להניח?

2. ציוני בחינת הבגרות במתמטיקה מתפלגים נורמלית עם שונות 150. במדגם של 16 נבחנים מתל אביב התקבלה שונות חסרת הטיה- 190. במדגם של 25 ירושלמים התקבלה שונות חסרת הטיה- 118.

- א. בדקו ברמת מובהקות של 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.

ח. הגישה המעשית - מובהקות התוצאה - P-VALUE - $\alpha_{\min}$

1. השכר הממוצע במשק בשנת 2006 היה 7800 ₪ עם סטיית תקן 2000 ₪. במדגם כיום של מאה עובדים התקבל שכר ממוצע 7500 ₪. מטרת המחקר היא לבדוק האם כיום חלה שחיקה בשכר כיום. עבור אילו רמות מובהקות שיבחר החוקר יוחלט לקבל את טענת המחקר?

$$H_0 : \mu = 70$$

$$H_1 : \mu > 70$$

2. לפניך השערות של מחקר :

המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות :

$$n = 100$$

$$\bar{x} = 74$$

מהי האלפא המנימלית לדחיית השערת האפס?

3. 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%?

4. במטרה לבדוק האם מטבע הוא הוגן הוא הוטל 100 פעמים. התקבל 58 פעמים "עץ".
האם יתכן שחוקר אחד יחליט שהמטבע הוגן ואחר יחליט שהמטבע לא הוגן?

5. 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 גבוהה מ-174 ס"מ באופן מובהק. מהי בקרוב מובהקות התוצאה ועל פיה מה תהיה המסקנה ברמת מובהקות של 6% ?

ט. טעויות ועוצמה

1. 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 העוצמה של הכלל מסעיף א?

2. 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 העוצמה אם התוחלת שווה ל-195?
- ג. הסבר ללא חישוב איך העוצמה תשתנה אם רמת המובהקות תהיה 5%?
3. 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%?
- ב. אם המדגם יתבסס על יותר תצפיות כיצד הדבר ישנה את התשובה לשל סעיף א'?
4. 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%. חשבו את הסיכוי לבצע טעות מסוג שני בהנחה והיום יש 28% אקדמאים.

5. 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 ככל 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.

י. הקשר בין רווח סמך לבדיקת השערות

1. 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 מ"ג. נלקח מדגם של 100 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרם סוכר היא 112.6 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 113 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א שהדיאטה משפיעה על רמת הסוכר בדם? הסבר.

2. 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.

3. במטרה לבדוק האם קיים הבדל בין חברת X לחברת Y מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלגים נורמלית בנו רווח סמך לממוצע ההפרשים שיבדוק את השערות המחקר ברמת מובהקות של 5%

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

תשובות סופיות - בדיקת השערות

פרק א' - בדיקת השערות לתוחלת

שונות האוכלוסיה ידועה

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	נדחה H_0
<u>שאלה 3</u>	
נדחה H_0	

שונות האוכלוסיה לא ידועה

<u>שאלה 1</u>	<u>שאלה 2</u>
נדחה H_0	נקבל H_0

פרק ב' - בדיקת השערות על פרופורציה

<u>שאלה 1</u>	<u>שאלה 2</u>
נדחה H_0	נדחה H_0
<u>שאלה 3</u>	
נקבל H_0	

פרק ג' - בדיקת השערות להפרש פרופורציות

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	א. נדחה H_0
	ב. נדחה H_0

פרק ד' - בדיקת השערות להפרש תוחלותשונויות האוכלוסייה ידועות

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	נדחה H_0

שונויות האוכלוסייה לא ידועות

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	נדחה H_0

פרק ה' - בדיקת השערות במדגמים מזווגים

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	נקבל H_0

פרק ו' - בדיקת השערות על שונות

<u>שאלה 1</u>	<u>שאלה 2</u>
נדחה H_0	א. נדחה H_0

פרק ז' - בדיקת השערות לשתי שונות

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	א. נקבל H_0
	ב. נקבל H_0

פרק ח' - מובהקות התוצאה

<u>שאלה 1</u>	<u>שאלה 2</u>
מעל 0.0668	0.0228
<u>שאלה 3</u>	<u>שאלה 5</u>
א. 0.0455	נקבל H_0

פרק ט' - טעויות ועוצמה

<u>שאלה 1</u>	<u>שאלה 2</u>
ב. נדחה H_0	ב. 0.8051
ג. 1	
<u>שאלה 3</u>	<u>שאלה 4</u>
א. 0.4404	0.1446

פרק י' - הקשר בין בדיקת השערות לרווח סמך

<u>שאלה 1</u>	<u>שאלה 3</u>
(110.64, 114.56)	(-0.241, 0.041)