

מדדים:

ממוצע

$$\bar{x} = \sum_{i=1}^c \frac{x_i \cdot f_i}{n}$$

ממוצע משוקלל או ממוצע הממוצעים

$$\bar{\bar{x}} = \frac{\sum_{i=1}^c (\bar{x}_i \cdot n_i)}{\sum_{i=1}^c n_i}$$

סטית תקן

$$\hat{S} = \sqrt{\hat{S}^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^c (x_i - \bar{X})^2 * f_i}{n-1}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 - n\bar{X}^2}{n-1}}$$

$$C.V = \frac{\hat{S}}{\bar{x}}$$

מקדם ההשתנות

הסתברות

איחוד מאורעות:

כאשר המאורעות A ו-B אינם זרים זה לזה $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

הסתברות מותנה:

$$p(B/A) = \frac{p(B \cap A)}{p(A)}$$

המאורעות בלתי תלויים כאשר מתקיים:

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

או $p(B/A) = p(B)$

תוחלת של המשתנה המקרי הבדיד :

$$E_{(X)} = \sum x_i * p(x_i)$$

שונות של המשתנה המקרי הבדיד :

$$V_{(X)} = \sigma_{(X)}^2 = \sum (X_i - E(X))^2 * P(x_i)$$

$$V_{(X)} = \sigma_{(X)}^2 = \sum x_i^2 * P(x_i) - [E(X)]^2$$

סטיית תקן של המשתנה המקרי הבדיד:

$$\sigma_{(X)} = SD_{(X)} = \sqrt{\sum x_i^2 * P(x_i) - [E(X)]^2}$$

תוחלת של סכום משתנים מקריים:

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

שונות של סכום משתנים מקריים בלתי תלויים:

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

התפלגות בינומית: $x \sim Bin(n, p)$

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$E(X) = n \cdot p$$

$$V(X) = n \cdot p \cdot q$$

הסתברות משתנה מקרי הרציף והתפלגות הנורמלית: $X \sim N(\mu, \sigma^2)$

$$Z_t = \frac{x_t - \mu}{\sigma} \quad \text{ציון תקן:}$$

התפלגות דגימה

חוקי התפלגות הדגימה של הממוצע:

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

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

$$SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

דגימה מהתפלגות נורמלית:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{אם} \quad X \sim N(\mu, \sigma^2) \quad \text{אזי:}$$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

דגימה מהתפלגות כלשהי (משפט הגבול המרכזי):

אם $E(X) = \mu$ ו- $V(X) = \sigma^2$ אזי עבור n מספיק גדול: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ (נורמלי בקרוב).

רוח סמך לתוחלת ברמת סמך $1-\alpha$ שנות אוכלוסייה ידועה:

$$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$n \geq \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\frac{\varepsilon}{2}} \right)^2 = \left(\frac{2 \cdot z_{\frac{\alpha}{2}} \cdot \sigma}{L} \right)^2 \quad \varepsilon = \frac{L}{2}$$

רשימת הנושאים למבחן:

סוגי משתנים (סולמות מדידה) + הגרף המתאים לכל סוג משתנה (עוגה, מקלות, היסטוגרמה)
טבלת שכיחויות
צורות התפלגות (זנב ימני, זנב שמאלי, סימטרית חד שיאית, דו-שיאית, צורה סימטרית)
מדדי מרכז (שכיח, חציון, ממוצע)
מדדי פיזור (תחום, תחום בין רביעוני, עשירונים, ריבעונים, מאונים, סטיית תקן ושונות)
קריאת פלט
טרנספורמציה לינארית
מקדם ההשתנות CV
ממוצע משוקלל
הסתברות
בניית פונקציית הסתברות (תוחלת, סטיית תקן ושונות)
ציון תקן Z
התפלגות נורמלית

נושאים שירדו הסמסטר (בכלל לא יילמדו):

בינום (בינומית)

רווח סמך

מבנה המבחן: 20 שאלות אמריקאיות

ההמלצה הכי חשובה שלי:

לתרגל רק אמריקאיות: לא לתרגל שאלות פתוחות!

סוגי משתנים (סולמות מדידה):

איכותי = מילולי

1. איכותי שמי = מילים ללא חשיבות לסדר.
אין חשיבות לאיזו מילה נאמרה קודם. למשל, שמות ילדים: דן, חיים, קובי.
הגרף שמתאר איכותי שמי הוא רק עוגה.

שאלה ממבחן: תשובה א דווקא שגויה!!!

איזה גרף יתאים לתאר את המשתנה סוגי משרות של נשים בהייטק

לפי הנתונים שמצורפים בתמונה הבאה?

א. אסימטרית חיובית (זנב ימני)

ב. אסימטרית שלילית (זנב שמאלי)

ג. סימטרית

ד. עוגה



2. איכותי סדר: מילים כאשר יש חשיבות לסדר, להיררכיה (איזו מילה נאמרה קודם).
 תהיה מגמה שהולכת ועולה.
 למשל, דרגות בצבא: אלוף, רב סרן, סרן, סמל
 למשל, רמת שביעות רצון: גבוהה מאוד, גבוהה, בינונית, נמוכה.
 הגרף שמתאר משתנה מסוג איכותי סדר הוא **דיאגרמת מקלות**.

3. משתנה כמותי בדיד: משתנה כמותי-מספרי שמקבל ערכים שלמים בלבד.
 למשל, מספר ילדים במשפחה, מספר מכוניות בחניה.
 הגרף שמתאר משתנה מסוג כמותי בדיד הוא **דיאגרמת מקלות**.

4. משתנה כמותי רציף: משתנה כמותי-מספרי שערכיו יכולים להיות לא-שלמים, כלומר עשרוניים. למשל, שכר, משקל, גובה.
 אם הוא יופיע בטבלה אז יהיו מחלקות (קבוצות).
 הגרף שמתאר משתנה מסוג כמותי רציף הוא היסטוגרמה.

שאלה ממבחן: תשובה א נכונה:

לפניך טבלאות התפלגות:

3		2		1	
גיל הסטודנט	שכיחות	מספר ילדים במשפחה	שכיחות	מקצוע	שכיחות
21-24	65	1	10	עורך דין	60
24-26	55	2	40	מהנדס	40
26-29	30	3	35	אדריכל	35
29-31	20	4	20	רופא	20
31-35	20	5	25	פסיכולוג	15
35-37	10	6	30	רואה חשבון	30
סה"כ	200	סה"כ	160	סה"כ	200

הדיאגרמה המתאימה לכל אחת מהן:

- א.** דיאגרמת מעגל, 2- דיאגרמת מקלות, 3- היסטוגרמה
- ב.** דיאגרמת מקלות, 2- דיאגרמת מעגל, 3- היסטוגרמה
- ג.** דיאגרמת מעגל, 2- היסטוגרמה, 3- דיאגרמת מקלות
- ד.** דיאגרמת מקלות, 2- היסטוגרמה, 3- דיאגרמת מעגל

שאלה ממבחן מועד א 2023: תשובה א נכונה:

17. פורסמו תוצאות ההצבעה לאגודת הסטודנטים: 25% הצביעו למועמד א', 30% הצביעו למועמד ב', ו- 45% לא הצביעו. המשתנה המוצג הינו _____ מסולם _____ והגרף המתאים להצגתו הוא _____. השלם את המשפט המתאים:
- א. איכותי, שמי, עוגה.
 - ב. כמותי בדיד, שמי, מקלות.
 - ג. איכותי, סדר, מקלות.
 - ד. כמותי בדיד, סדר, היסטוגרמה

5. מחלק העיתונים בשכונת נווה-פרחים בתל אביב, החליט לערוך ניתוח סטטיסטי על מאפייני הלקוחות במסלול החלוקה שלו. לשם כך, עבור כל בית במסלול הוא מדד את: מספר הנפשות המתגוררות בו, סוג המכונית שעומדת בחניית הבית (אמריקאית, אירופאית או יפנית), רמת ההכנסה של משק הבית (פחות מ-8,000 ש"ח; 8,000-15,000 ש"ח; מעל 15,000 ש"ח), ואת גובה הגדר המקיפה את הגינה (בס"מ). אילו סוגי משתנים אסף מחלק העיתונים?

א. משתנה שמי, משתנה סדר ושני משתנים כמותיים

ב. שני משתנים שמיים ושני משתנים כמותיים

ג. משתנה שמי, שני משתני סדר ומשתנה כמותי

ד. משתנה שמי ושלושה משתנים כמותיים

צורות התפלגות עיקריות:

שאלה ממבחן:

טענה א: בהתפלגות א-סימטרית שמאלית הרבעון התחתון בהכרח מעל הממוצע

טענה ב: בהתפלגות א-סימטרית ימנית המאון ה 9 בהכרח מתחת לשכיח

אילו מהטענות נכונות?

א. אף אחת

ב. רק טענה א

ג. רק טענה ב

ד. שתי הטענות נכונות

לסדרה סטטיסטית בעלת צורת התפלגות א-סימטרית שמאלית (שלילית) אחוז המקרים הנמצאים מעל השכיח היא:

- א. קטן מאחוז המקרים הנמצאים מתחתיו
 ב. גדול מאחוז המקרים הנמצאים מתחתיו
 ג. שווה לאחוז המקרים הנמצאים מתחתיו
 ד. אין קשר בין אחוז המקרים לבין השכיח

לתרגל בבית, ממועד ב 2022: תשובה א נכונה:

4. בחברת תיווך נדל"ן ארצית, נאספו נתונים על מספר הדירות שמכר כל אחד מ- 50 עובדי החברה בשנה האחרונה. להלן נתונים על התפלגות מספר הדירות שנמכרו. נתון כי התפלגות מספר הדירות הינה **א-סימטרית שמאלית**. השתמשו בנתונים אלה כדי להשיב:

statistics	מס' דירות
N	50
Mean	1.42
Std. deviation	2

משפט 1: מספר הדירות החציוני קטן מ - 1.42

משפט 2: מספר הדירות השכיח גדול מ - 1.42

א. רק משפט 2 נכון

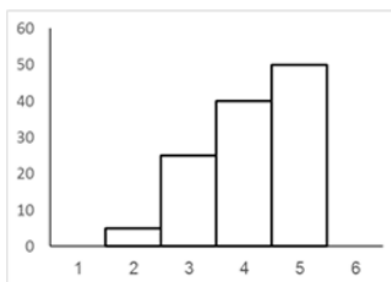
ב. רק משפט 1 נכון

ג. שני המשפטים נכונים

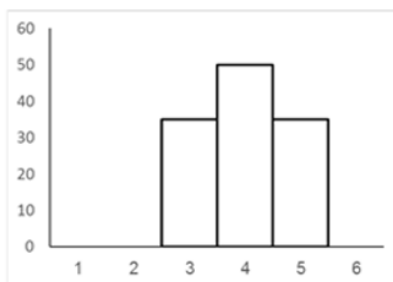
ד. אף משפט אינו בהכרח נכון

שאלה 4

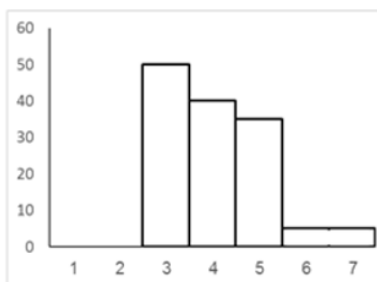
להלן 3 היסטוגרמות המתארות 3 התפלגויות:



(A)



(B)



(C)

לאיזו מההתפלגויות יש את השכיח הגדול/גבוה ביותר?

א. A

ב. B

ג. C

ד. זהה בכלם

בחברה העוסקת בטלמרקטינג, בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4, וכי סטיית התקן היא שנתיים. כיצד ישתנו הממוצע וסטיית התקן, אם יתווספו שני עובדים עם ותק של 3 ו- 5 שנים להתפלגות?

- א. הממוצע לא ישתנה, סטיית התקן תקטן
- ב. הממוצע יגדל, סטיית התקן לא תשתנה
- ג. הממוצע לא ישתנה, סטיית התקן תגדל
- ד. הן הממוצע והן סטיית התקן לא ישתנו

טרנספורמציה לינארית (אין צורך להכיר את השם הטיפשי הזה)

מה קורה כאשר מבצעים את אותו השינוי על כל המספרים?

לגבי הממוצע:

אם מוסיפים, מורידים, מכפילים, מחלקים או משנים באחוזים את כל המספרים אז הממוצע תמיד משתנה לפי השינוי שקרה.

למשל, אם הוספנו 100 שח לכל המשכורות, גם ממוצע השכר ייעלה ב 100.

למשל, אם נכפיל את כל המשכורות פי 3, גם הממוצע יוכפל פי 3.

לסיכום, הממוצע תמיד "זורם" לפי השינוי שקרה.

לגבי סטיית התקן:

אם נכפיל או נחלק או נשנה באחוזים את כל המספרים אז סטיית התקן תשתנה בדיוק לפי השינוי שקרה. לסיכום, כפל, חלוקה, אחוזים תמיד משנים את סטיית התקן.

X : **%**
בכפל, חילוק, אחוזים:
סטיית התקן "זורמת" לפי השינוי שקרה.

דוגמא:

10 30

אם נוסיף את אותו המספר לכולם אז סטיית התקן לא תשתנה.

אם נוריד את אותו המספר מכולם אז סטיית התקן לא תשתנה.

לסיכום,
+ הוספת מספר או הורדת מספר
- לא משנה את סטיית התקן.

דוגמא:

10 30

לפניכם שתי טענות :

טענה א: השכר הממוצע ברוטו של קבוצת עובדים הוא 3000 ₪ עם סטיית תקן 800 ₪.

אם כל העובדים ישלמו 20% משכרם מס הכנסה –

יהיה השכר הממוצע נטו 2400 ₪ וסטיית התקן תישאר 800 ₪.

טענה ב: הציון הממוצע של הנבחנים בסוציולוגיה היה 76 עם סטיית תקן 6.

גיא קיבל במבחן ציון 80.

המרצה החליט להוסיף לכל נבחן 5 נקודות.

לאחר תוספת , מיקומו היחסי של גיא בכיתה השתפר.

איזו/אילו מהטענות הנ"ל נכונה/ות ?

א. שתי הטענות שגויות.

ב. שתי הטענות נכונות.

ג. רק טענה ב

ד. רק טענה א

בקורס בסטטיסטיקה במכללת קשת בענן נבחנו 100 סטודנטים. ממוצע הציונים היה 50 ושוונות 36. המרצים לא היו מרוצים מהתוצאות והחליטו לתת פקטור של 10 אחוז ולאחר מכן תוספת של 15 נקודות לציוני כלל הסטודנטים.

חשבו את הממוצע וסטיית התקן של הציונים לאחר הבונוסים.

א. ממוצע 70, סטיית תקן 6.6
 ב. ממוצע 80, סטיית תקן 8.8
 ג. ממוצע 60, סטיית תקן 2.5
 ד. ממוצע 68, סטיית תקן 4

תזכורת

מה קורה כאשר מבצעים את אותו השינוי על כל המספרים?

לגבי הממוצע:

אם מוסיפים, מורידים, מכפילים, מחלקים או משנים באחוזים את כל המספרים אז הממוצע תמיד משתנה לפי השינוי שקרה.

למשל, אם הוספנו 100 שח לכל המשכורות, גם ממוצע השכר ייעלה ב 100.

למשל, אם נכפיל את כל המשכורות פי 3, גם הממוצע יוכפל פי 3.

לסיכום, הממוצע תמיד "זורם" לפי השינוי שקרה.

לגבי סטיית התקן:

אם נכפיל או נחלק או נשנה באחוזים את כל המספרים אז סטיית התקן תשתנה בדיוק לפי השינוי שקרה. לסיכום, כפל, חלוקה, אחוזים תמיד משנים את סטיית התקן.

X : **%**
 בכפל, חילוק, אחוזים:
 סטיית התקן "זורמת" לפי השינוי שקרה.

אם נוסיף את אותו המספר לכולם אז סטיית התקן לא תשתנה.
 אם נוריד את אותו המספר מכולם אז סטיית התקן לא תשתנה.
 לסיכום, **+** הוספת מספר או הורדת מספר **-** לא משנה את סטיית התקן.

CV מקדם ההשתנות:

מתי נשתמש?

המילים הומוגניות או הטרוגניות מעיפות אותנו ל CV

$$CV = \frac{S}{\bar{X}}$$

חוקים:

CV ו הומוגניות תמיד בכיוון הפוך:

ככל שה CV גדול, ההומוגניות קטנה.

ככל שה CV קטן, ההומוגניות גדולה.

הטרוגניות ו CV הם תמיד באותו כיוון:

ככל שה CV גדול, ההטרוגניות גדולה.

ככל שה CV קטן, ההטרוגניות קטנה.

וכעת הסבר עם הבנה למילה "הומוגניות" (לא חובה, ניתן להסתפק בחוקים):

כולנו יודעים מהחיים שהמילה הומו משמעותה אותו דבר...

ככל שסדרה היא יותר הומוגנית הכוונה היא שהמספרים שלה הם יותר דומים, יותר קרובים,

יותר צפופים וצמודים- יותר "אותו דבר": הפיזור של המספרים קטן יותר.

ולהפך:

ככל שסדרה היא יותר הטרוגנית אז המספרים שלה יותר שונים ורחוקים אחד מהשני-

הם מפוזרים יותר רחוק אחד מהשני (הפיזור נחשב לגדול יותר).

20. להלן פלט המציג מדדים שונים שחושבו על התפלגות הוותק והגיל בארגון מסוים:

Statistics			
		ותק	גיל
N	Valid	102	102
	Missing	0	0
Mean		4.32	27.255
Median		4.00	27.000
Mode		2	27.0
Std. Deviation			
Variance		5.607	8.192
Minimum		2	21.5
Maximum		9	31.0
Percentiles	25	2.00	27.000
	50	4.00	27.000
	75	6.00	30.000

איזו התפלגות, הגיל או הוותק, הטרוגנית יותר?

- א. הגיל
- ב. הוותק
- ג. שתי ההתפלגויות בעלות אותו הפיזור
- ד. אין מספיק נתונים על מנת לחשב

מכון כושר מעוניין לבדוק האם הדופק בזמן ריצה שונה אצל גברים ואצל נשים בקרב אנשים מעל גיל 35. לפניכם פלט המציג את הממצאים:

Statistics			
		נשים	גברים
N	Valid	16	15
	Missing	0	0
Mean		167.25	172.20
Median		168.00	174.00
Mode		162	162
Std. Deviation		11.287	8.670
Variance			
Range			
Minimum		146	156
Maximum		186	186
Percentiles	25	162.00	166.00
	50	168.00	174.00
	75	174.00	178.00

על סמך הפלט, אילו מהטענות הבאות נכונות:

טענה א': בקרב נשים קיימים ערכים נמוכים קיצוניים יותר מאשר בקרב גברים

טענה ב': הטווח (תחום) הבין-רבעוני שווה בשתי הקבוצות

טענה ג': ההטרוגניות של התפלגות הדופק בקרב גברים גבוהה מההטרוגניות של התפלגות הדופק בקרב נשים

- א. רק טענות א' ו- ב' נכונות
- ב. רק טענה א' נכונה
- ג. כול הטענות נכונות
- ד. רק טענות א' ו- ג' נכונות

חוקר רצה לבחון האם יש הבדלים בשכר בן בני זוג, לצורך כך דגם מדגם מקרי של 300 זוגות והתקבל ממוצע שכר גברים 15,700 ש"ח ושונות 122,500 וממוצע שכר נשים 10,110 ש"ח ושונות 90,000. באיזה מבין הקבוצות התפלגות המשכורות הומוגנית יותר?

- א. קבוצת הגברים
- ב. קבוצת הנשים
- ג. הפיזור זהה בשתי הקבוצות
- ד. לא ניתן לדעת

נוציא שורש לשונות הגברים 122500 ונקבל סטיית תקן גברים 350

$$CV \text{ גברים} = \frac{350}{15700} = 0.02229$$

נוציא שורש לשונות הנשים 90000 ונקבל סטיית תקן גברים 300

$$CV \text{ נשים} = \frac{300}{10110} = 0.02967$$

CV והומוגניות הולכים בכיוון הפוך תמיד.
CV גברים קטן יותר ולכן ההומוגניות אצל הגברים גדולה יותר
תשובה א נכונה

ממוצע משוקלל (ממוצע הממוצעים):

ממבחן מועד ב 2023

במאפיה מוכרים פיתות ולחמים, ביום מסויים נמכרו 210 לחמים וממוצע מחיר הלחם עמד על 20.2 ₪. כמו כן, נמכרו 400 חבילות פיתות, המחיר הממוצע עבור חבילת פיתות 10.5 ש"ח. מה יהיה ממוצע המחיר הכולל של המאפים?

א. 13.839

ב. 15.515

ג. 10.503

ד. 11.627

10. במסגרת תכנית לשיפור בריאות העובדים בחברה מסוימת, נאספו נתונים על מספר הסיגריות שעישנו העובדים במשך שעות העבודה ביום מסוים. התקבלו הנתונים הבאים בנוגע למחלקות הנותנות שרות לקהל:

מספר עובדים ממחלקות השירות	מספר סיגריות
58	0
62	1
37	2
23	3
15	4
5	5

בנוסף לנתונים הנ"ל התקבלו בשלב מאוחר יותר נתונים עבור 50 עובדים אחרים **ממחלקת הטיפעול והלוגיסטיקה**. ממוצע מספר הסיגריות שעישנו עובדים ממחלקת הטיפעול והלוגיסטיקה הוא 7.1, וסטיית התקן היא 5.1. מהו הממוצע הכולל של מספר הסיגריות בקרב שתי קבוצות אלו?

א. 2.58

ב. 4.27

ג. 3.75

ד. 5.24

פתרון:

מציאת סך הסיגריות שמעשנים מחלקת הטיפעול והלוגיסטיקה:

$$\bar{X} = \frac{(0 * 58) + \dots + (5 * 5)}{200} = \frac{290}{200} = 1.45$$

וכך הממוצע המשוקלל של כל עובדי המחלקות:

$$\bar{\bar{x}} = \frac{1.45 * (200) + 7.1 * (50)}{200 + 50} = \frac{645}{250} = 2.58$$