

מבוא להסתברות וסטטיסטיקה

תרגיל 2:

הגשה תוך שבועיים בתרגול לידי המתרגל בלבד. (13.3 ו-15.3 לפי המתרגל)

שאלה 1:

חיי מדף של מוצר מסוים היו כדלהלן (בימים)

שכיחות	ימים
10	[5-7)
18	[7-9)
34	[9-11)
22	[11-13)
16	[13-15)
10	[15-17)

מצאו: א. ממוצע, חציון, שכיח.

ב. ציירו היסטוגרמה

ג. מה ישתנה בתשובותכם לסעיף א' אם השורה התחתונה היא $+15$ (15 יום או יותר)?

שאלה 2:

לפניכם מדגם של כמות אשפה יומית לנפש, שמייצר אדם באזור ת"א:

שכיחות	כמות אשפה בק"ג
5	[0.5-1.0)
10	[1.0-1.5)
20	[1.5-2.0)
50	[2.0-2.5)
70	[2.5-3.0)
100	[3.0-3.5)
80	[3.5-4.0)
40	[4.0-4.5)
25	[4.5-5.0)

מצאו: א. ממוצע, חציון, שכיח.

ב. מה ישתנה בתשובותיכם לסעיף א' אם השורה התחתונה היא $+4.5$ (4.5 ק"ג או יותר)?

שאלה 3:

חוקרים מצאו בסקר על תאונות בבית שגילם של 30 נפגעים הוא (בשנים שלמות):

0,1,2,2,3,4,4,5,5,7,10,12,15,18,20,24,29,34,37,42,51,58,64,68,70,72,74,75,78,79

א. צרו טבלה המתארת את התפלגות הגילאים לפי הקבוצות הבאות:

[0-5), [5-10), [10-20), [20-30), [30-40), [40-50), [50-60), [60-70), [70-80)

ב. חשבו ממוצע, תחום בין רבעוני.

ג. איזו מדידה גבוהה יותר: הממוצע או החציון? ממה נובע ההבדל?

ד. צירו היסטורמה עם מצולע שכיחויות.

שאלה 4:

א. בעולם האמיתי נתונים לעיתים קרובות מכילים "רעש" דהיינו חלק מהדגימות אינן משקפות את הנתונים האמיתיים, לדוגמא, בגלל תקלה במכשיר המדידה. לפעמים מתעוררת בעיה כאשר משתמשים בממוצע או בשונות כמדדים במקרה שכזה, מדוע? (האם אתם יכולים חלסום את גודל ההטיה מהממוצע או השונות "האמיתית" של הנתונים בהינתן שמדדת דגימת "רעש").

ב. נגדיר $MAD = Med\{|x_k - Med\{x_k\}|\}$ (חציון הפרשים המוחלטים מהחציון) האם הוא מדד לפיזור (מקיים את האקסיומות) דהיינו מקיים סימטריות, הומוגניות ואינוריאנטיות? (הוכיחו טענותיכם)

ג. נגדיר $T = \frac{1}{n-2m} \sum_{i=m+1}^{n-m} x_i$ באשר $x_1 \leq x_2 \leq \dots \leq x_n$ מסודרים. (דהיינו מוציאים את m הדגימות הקטנות ביותר והגדולות ביותר). האם זהו מדד מרכזי? (הוכיחו טענותיכם)

שאלה 5 (מתנה):

הוכיחו את המשוואה הבאה:

$$\frac{1}{n} \sum_{k=1}^n x_k^2 - \bar{x}^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$$

באשר $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$.