תרגול כיתה 1 – מבוא להסתברות וסטטיסטיקה
סטטיסטיקה תיאורית – תיאור טבלאי וגרפי של נתונים

שאלה 1

א) סווג את המשתנים הבאים לבדידים/רציפים:

(1). גובה מעל פני הים. (2). צבע עיניים. (3). קצב המכוניות שנכנסות לחניון בשעה.

ב). סווג את המשתנים הבאים לפי סולמות מדידה (שמי/סדר/מרווחי/יחס)

(1). דרגה צבאית. (2). ציון במבחן כלשהו. (3). כמות דלק במיכל. (4). התשובות הבאות בשאלון טעימת גלידה חדשה: 1=אוהב; 2=אדיש; 3=לא אוהב.

פתרון:

[הערה: הסיווג לא תמיד חד-משמעי ולעתים משתנה לפי הגדרת המשתנים והבעיה]

א). בדידים/רציפים:

(1). גובה מעל פני הים = רציף (לא נתון קנה המידה למדידה. אם היתה השאלה למשל 'במטרים שלמים' אז המשתנה בדיד)

(2). צבע עיניים = בדיד (מקובל)

(3). קצב = רציף (מס' מכוניות חלקי יח' הזמן).

ב). סולמות מדידה:

(1). דרגה צבאית = סדר.

(2). ציון במבחן כלשהו = מרווחי (ניתן לקבוע ציונים: 200-800, 0-100 וכו').

(3). כמות דלק במיכל = יחס (הכמות 0 = אין נוזל במיכל).

(4). התשובות בשאלון טעימת גלידה = שמי.

שאלה 2

בקבוצה של 230 משפחות של נשים עובדות נערך סקר על מספר הילדים במשפחה. התוצאות מסוכמות להלן:

x	$f(x)$
8	1
7	2
6	4
5	8
4	20
3	38
2	60
1	60
0	37

א. חשב את השכיח (Mo).

ב. חשב את החציון (Me).

ג. חשב את אמצע הטווח (MR).

ד. חשב את הממוצע ($\bar{x}$).

ה. חשב את סטיית התקן המדגמית (s).

ו. מה תוכל לומר מתוך התוצאות הקודמות על צורת ההתפלגות?

פתרון:

א. *שכיח:* במקרה שלנו יש 2 שכיחים, הערכים 1 ו-2 שלהם שכיחות מקסימלית – 60.

ב. חציון: (נוסיף לטבלה עמודה של שכיחות מצטברת להקל על מציאת החציון)

x	*f(x)*	*F(x)*
8	1	230
7	2	229
6	4	227
5	8	223
4	20	215
3	38	195
2	60	157
1	60	97
0	37	37

(נשים לב שמדובר ברשימת נתונים, ולא במחלקות).

n=230 הוא זוגי, ולכן החציון הוא ממוצע שני הערכים במקומות n/2 ו- (n/2)+1, כלומר ממוצע הערך

ה-115 וה- 116. שני ערכים אלה שווים ל- 2, לכן: $Md = \frac{2+2}{2} = 2$.

ג. אמצע הטווח: $MR = \frac{X_{\max} - X_{\min}}{2} = \frac{8+0}{2} = 4$.

ד. ממוצע:

$$\bar{x} = \frac{\sum_i x_i f_i}{n} = (8\cdot 1 + 7\cdot 2 + 6\cdot 4 + 5\cdot 8 + 4\cdot 20 + 3\cdot 38 + 2\cdot 60 + 1\cdot 60 + 0\cdot 37) = \frac{460}{230} = 2$$

ה. סטיית התקן המדגמית:

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2 \cdot f_i}{n-1}} = \sqrt{\frac{(8-2)^2\cdot 1 + ... + (2-2)^2 \cdot 60 + ... + (0-2)^2 \cdot 37}{230-1}} = \sqrt{2.393} = 1.547$$

ו. ההתפלגות היא אסימטרית חיובית. אמצע הטווח הוא הגדול ביותר, כלומר, יש ערכים ששכיחותם נמוכה, הסוטים לכיוון החיובי. שאר המדדים מתרכזים באותה סביבה.

<u>שאלה 3</u>

נתונה התפלגות ציונים בכיתה בת 60 תלמידים. ידוע שהתפלגות הציונים סימטרית.

x	f(x)
50-60	
60-70	
70-80	20
80-90	
90-100	5
סה"כ	60

א. השלם את טבלת השכיחויות, הוסף שכיחות מצטברת ושכיחות מצטברת יחסית.

ב. חשב ממוצע וחציון.

ג. חשב את הרבעון התחתון והרבעון העליון וכן את התחום הבין רבעוני.

ד. מצא את הציון ש- 90% מהציונים קטנים ממנו.

<u>פתרון:</u>

א. על סמך העובדה שההתפלגות סימטרית ניתן למלא את החסר בתא הראשון. כעת נותרו 30 תצפיות אותן יש לחלק שווה בשווה בין התאים השני והרביעי.

x	f(x)	נקודת אמצע	F(x)	F(x)/n
50-60	5	55	5	05/60 = 1/12
60-70	15	65	20	20/60 = 1/3
70-80	20	75	40	40/60 = 2/3
80-90	15	85	55	55/60 = 11/12
90-100	5	95	60	60/60 = 1
סה"כ	60			

ב. ממוצע: $\bar{x} = \frac{1}{n}\sum_i x_i \cdot f(x_i) = \frac{55\cdot5+65\cdot15+75\cdot20+85\cdot15+95\cdot5}{60} = 75$

הנוסחה לחציון- $Md = Me = \frac{n/2 - \mathrm{F}(x_{m-1})}{f(x_m)}\cdot(L_1 - L_0) \; + \; L_0$

<u>כאשר:</u>

n: מספרן הכולל של התצפיות.

x_m : המחלקה בה נמצא החציון.

x_{m-1} : המחלקה הקודמת ל- x_m .

$F(x_{m-1})$: השכיחות המצטברת עד למחלקה x_{m-1} .

$f(x_m)$: שכיחות המחלקה x_m .

L_0 : גבול תחתון אמיתי של המחלקה שבה נמצא החציון.

L_1 : גבול עליון אמיתי של המחלקה שבה נמצא החציון.

חציון: $\frac{n}{2}=30 \Rightarrow X_m=(70,80) \Rightarrow Md=L_0+\frac{\frac{n}{2}-F(x_{m-1})}{f(x_m)}\left(L_1-L_0\right)=70+\frac{30-20}{20}(80-70)=75$

ג. רבעון תחתון:

$$\frac{n}{4}=15 \Rightarrow X_m=(60,70) \Rightarrow Q_1=L_0+\frac{\frac{n}{4}-F(x_{m-1})}{f(x_m)}\left(L_1-L_0\right)=60+\frac{15-5}{15}(70-60)=66.67$$

רבעון עליון:

$$\frac{3n}{4}=45 \Rightarrow X_m=(80,90) \Rightarrow Q_3=L_0+\frac{\frac{3n}{4}-F(x_{m-1})}{f(x_m)}\left(L_1-L_0\right)=80+\frac{45-40}{15}(90-80)=83.33$$

התחום הבין רבעוני: $IQR=Q_3-Q_1=83.33-66.67=16.67$ (50% מהתצפיות נמצאות בו)

ד. האחוזון ה- m מחושב ע"י: $X_m=L_0+\frac{\frac{m\cdot n}{100}-F(x_{m-1})}{f(x_m)}\left(L_1-L_0\right)$.

$$\frac{m\cdot n}{100}=\frac{90\cdot 60}{100}=54 \Rightarrow X_m=(80,90) \Rightarrow X_{90}=80+\frac{54-40}{15}(90-80)=89.33$$

<u>שאלה 4</u>

להלן ציוניו של סטודנט בשני קורסים וכן הממוצע וסטיית התקן של הציונים במבחנים האלה:

קורס	ממוצע	סטיית תקן	ציון הסטודנט
טופולוגיה	76	16	60
מד"ר	72	10	82

איזה ציון צריך הסטודנט לקבל בטופולוגיה על מנת שציונו יהיה שווה ברמתו, ביחס לכל הנבחנים, לציונו במד"ר?

<u>*פתרון:*</u>

ציון התקן במד"ר: $Z_1=\frac{82-72}{10}=1$.

כדי לקבל ציון תקן זהה בטופולוגיה נפתור את המשוואה: $Z_2=1=\frac{x-76}{16} \Rightarrow x=92$.

<u>שאלה 5</u>

בטבלת השכיחויות הבאה רוכזו נתוני השכר לשעה (בשקלים) של 100 עובדים בחברה מסויימת:

f	X (שכר לשעה)
20	30 – <50
40	50 – <70
15	70 – <80
5	80 – <90
20	90 – <120

הערה: 50> – 30 מסמל: $30 \le X < 50$ וכו'.

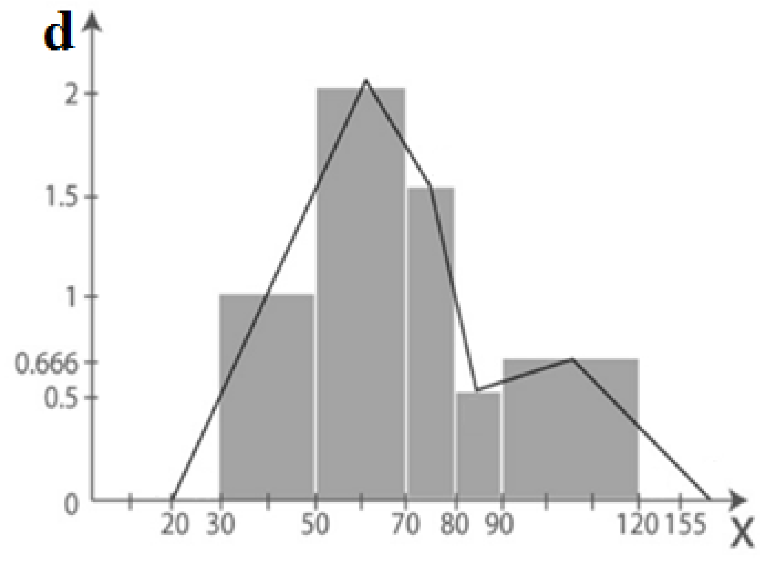
א. תאר את הנתונים בעזרת היסטוגרם וטבלת עזר מתאימה.
ב. הוסף את מצולע השכיחויות על ההיסטוגרם.

<u>פתרון:</u>

א. לבניית ההיסטוגרם נצרף לטבלה עמודת צפיפויות (d):
(הערה: ניתן לצרף לטבלה המקורית מיותר לבנותה מחדש)

d	f	X (שכר לשעה)
20/20 = 1	20	30 – <50
40/20 = 2	40	50 – <70
15/10 = 1.5	15	70 – <80
5/10 = 0.5	5	80 – <90
20/30 = 0.667	20	90 – <120

<u>ההיסטוגרם המבוקש:</u>



ב. מצולע השכיחויות צויר על ההיסטוגרם ע"י חיבור אמצעי המלבנים.