<u>פיתרון תרגיל בית 2</u>

.1

x	f(x)	F(x)	נקודת האמצע	רוחב הקבוצה
5-7	10	10	6	2
7-9	18	28	8	2
9-11	34	62	10	2
11-13	22	84	12	2
13-15	16	100	14	2
15-17	10	110	16	2

א. ממוצע- $\overline{x} = \frac{\sum x_i f(x_i)}{n} = \frac{10\cdot 6+18\cdot 8+34\cdot 10+22\cdot 12+16\cdot 14+10\cdot 16}{10+18+34+22+16+10} = 10.8$

חציון- הקבוצה בה נמצא החציון היא הקבוצה בה נמצאים האיברים ה- 55 וה-56.
זוהי הקבוצה השלישית.

$F(x_{m-1}) = 28$

$\frac{N}{2} = 55 \quad , \frac{N}{2}+1 = 56$

$L_1 = 11$

$L_0 = 9$

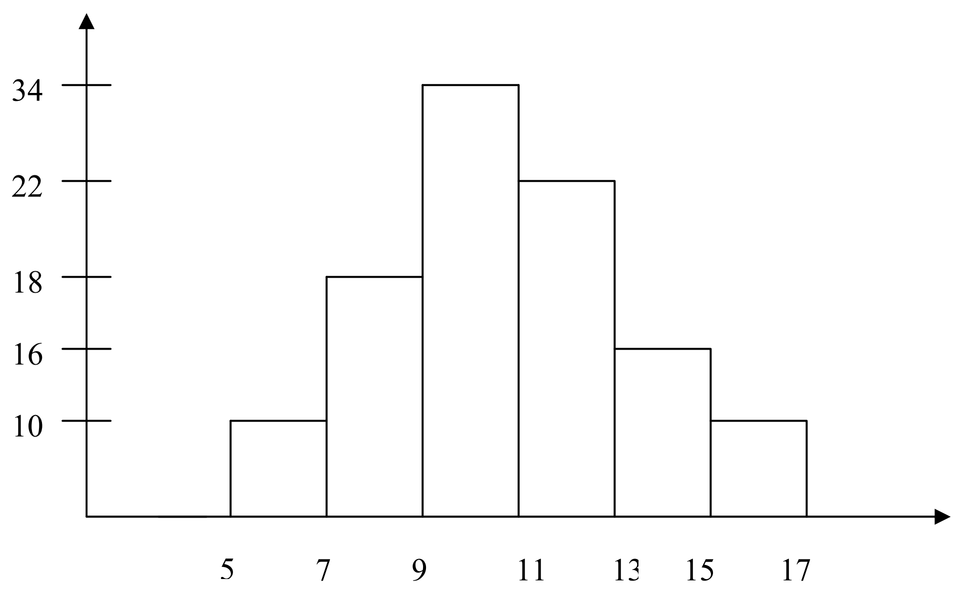
$f(x_m) = 34$

$$M_d = \frac{\left(\frac{55+56}{2}\right)-28}{34}\cdot(11-9)+9 = 10.618$$

מי שהשתמש ב-55 או 56 בלבד (ולא מיצע) מקבל גם את מלא הנקודות.

שכיח: הקבוצה (קטגוריה) השכיחה היא הקבוצה השלישית.

ב.



ג. ממוצע- לא נוכל לחשב.
שכיח (קטגוריה שכיחה)- לא ישתנה.
חציון- לא ישתנה.

.2

רוחב מחלקה	נקודת אמצע	F(x)	f(x)	x
0.5	0.75	5	5	[0.5-1.0)
0.5	1.25	15	10	[1.0-1.5)
0.5	1.75	35	20	[1.5-2.0)
0.5	2.25	85	50	[2.0-2.5)
0.5	2.75	155	70	[2.5-3.0)
0.5	3.25	255	100	[3.0-3.5)
0.5	3.75	335	80	[3.5-4.0)
0.5	4.25	375	40	[4.0-4.5)
0.5	4.75	400	25	[4.5-5.0)

א. ממוצע- $\bar{x} = \frac{\sum x_i f(x_i)}{n} = \frac{1270}{400} = 3.175$

חציון- הקבוצה בה נמצא החציון היא הקבוצה השישית.

$$\frac{N}{2} = 200 \ , \ \frac{N}{2} + 1 = 201$$

$$M_d = \frac{\frac{200+201}{2} - 155}{100}(3.5-3) + 3 = 3.2275$$

מי שהשתמש ב-201 או 200 בלבד (ולא מיצע) מקבל גם את מלא הנקודות.

שכיח (קטגוריה שכיחה)- הקבוצה השישית..

ב. ממוצע- לא ניתן לחשב.
חציון- לא ישתנה.
שכיח (קטגוריה שכיחה)- לא ישתנה.

.3 א.

נקודת האמצע	F(x)	f(x)	x
2.5	7	7	[0-5)
7.5	10	3	[5-10)
15	14	4	[10-20)
25	17	3	[20-30)
35	19	2	[30-40)

45	20	1	[40-50)
55	22	2	[50-60)
65	24	2	[60-70)
75	30	6	[70-80)

ב. ממוצע- $\bar{x} = \frac{980}{30} = 32.67$

רבעון תחתון- הקבוצה בה נמצא הרבעון התחתון היא הקבוצה השניה.

$$M_{0.25} = \frac{(7.5-7)}{3}(10-5)+5 = 5.83$$

רבעון עליון- הקבוצה בה נמצא הרבעון העליון היא הקבוצה השמינית.

$$M_{0.75} = \frac{(22.5-22)}{2}(70-60)+60 = 62.5$$

תחום בין רבעוני- $62.5-5.83 = 56.67$

ג. חציון- הקבוצה בה נמצא החציון היא הקבוצה הרביעית

$$M_d = \frac{\left(\frac{15+16}{2}\right)-14}{3}(30-20)+20 = 25$$

מי שהשתמש ב-15 או 16 בלבד (ולא מיצע) מקבל גם את מלא הנקודות.

החציון מסמל את הנקודה ש- 50% מהנתונים קטנים ממנה ו- 50% גדולים ממנה, במקרה שלנו החציון הוא 23.3. ואילו הממוצע גבוה יותר (32.67) וזאת משום שהוא מושפע מערכים קיצוניים בקצוות אשר משפיעים עליו לכיוון מעלה.

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 $MAD = \mathrm{Med}\{|x_k - Med\{x_k\}|\}$ (חציון ההפרשים המוחלטים מהחציון) האם הוא מדד לפיזור (מקיים את האקסיומות) דהיינו מקיים סימטריות , הומוגניות ואינוריואנטיות? (הוכיחו טענותיכם)

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

תשובה 4:

א. ממוצע ושונות מאוד רגישים לרעש. למעשה, בעזרת שינוי ערך של דגימה בודדת ניתן להגדיל את הממוצע והשונות כרצונכם. (שימו לב שבמדדים המופיעים בסעיפים הבאים שינוי של דגימה אחת לא ישפיע על המדד)

ב. עבור מדד פיזור יש לבדוק :

1. סימטריות: $f_n(x_1,\ldots,x_n) = f_n(x_{\sigma 1},\ldots,x_{\sigma n})$ לכל תמורה $\sigma \in S_n$.

2. הומוגניות: $f_n(cx_1,\ldots,cx_n) = cf_n(x_1,\ldots,x_n)$.

3. אינווריאנטיות: $f_n(a+x_1,\ldots,a+x_n) = f_n(x_1,\ldots,x_n)$.

כאן $\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n}) = \mathrm{Med}_{1\le k\le n}(\{|x_k - Med_{1\le k\le n}\{x_k\}|\}$

1. בחציון (Med) מסדרים תחילה את האיברים לכן סדר הופעתם כארגומנטים בפונקציה אינו משנה.

2.

$$\mathrm{f_n}(\mathrm{cx_1},\ldots,\mathrm{cx_n}) = \mathrm{Med}_{1\le k\le n}(\{|c\cdot x_k - Med_{1\le k\le n}\{c\cdot x_k\}|\} =$$
$$= \mathrm{Med}_{1\le k\le n}(\{|c\cdot x_k - c\cdot Med_{1\le k\le n}\{x_k\}|\} =$$
$$= \mathrm{Med}_{1\le k\le n}(\{|c\cdot(x_k - Med_{1\le k\le n}\{x_k\})|\} =$$
$$= |\mathrm{c}|\cdot \mathrm{Med}_{1\le k\le n}(\{|x_k - Med_{1\le k\le n}\{x_k\}|\} = |\mathrm{c}|\cdot \mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})$$

3.

$$\mathrm{f_n}(\mathrm{x_1}+\mathrm{a},,,\mathrm{x_n}+\mathrm{a}) = \mathrm{Med}_{1\le k\le n}(\{|(x_k+a) - Med_{1\le k\le n}\{x_k+a\}|\} =$$
$$= \mathrm{Med}_{1\le k\le n}(\{|x_k + a - Med_{1\le k\le n}\{x_k\} - a|\} =$$
$$= \mathrm{Med}_{1\le k\le n}(\{|x_k - Med_{1\le k\le n}\{x_k\}|\} = \mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})$$

שימו לב במהלך ההוכחות של התכונות, הסתמכנו על סימטריות הומוגניות ואינווריאנטיות החציון כמדד מרכז.

ג. עבור מדד מרכז יש לבדוק:

1. סימטריות: $f_n(x_1,\ldots,x_n)=f_n(x_{\sigma 1},\ldots,x_{\sigma n})$ לכל תמורה $\sigma\in S_n$.

2. עקביות: $f_{n+1}(x_1,\ldots,x_n,f_n(x_1,\ldots,x_n))=f_n(x_1,\ldots,x_n)$.

3. הומוגניות: $f_n(cx_1,\ldots,cx_n)=cf_n(x_1,\ldots,x_n)$.

4. אינווריאנטיות: $f_n(a+x_1,\ldots,a+x_n)=a+f_n(x_1,\ldots,x_n)$.

כאן

$$\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})=\frac{1}{n-2m}\sum_{\mathrm{i=m+1}}^{\mathrm{n-m}}\mathrm{x_i}$$

1. מאחר ובהגדרה נדרשים לסדר $x_1\le x_2\ldots\le x_n$ יש סימטריות.

2. עקביות: שימו לב $\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})$ הוא ממוצע של $n-2m$ האיברים עם הערכים "המרכזיים" ומתקיים $x_m\le\frac{1}{n-2m}\sum_{\mathrm{i=m+1}}^{\mathrm{n-m}}\mathrm{x_i}\le x_{n-m}$ לכן ב- $\mathrm{f_{n+1}}\left(\mathrm{x_1},\ldots,\mathrm{x_n},\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})\right)$ יש מיצוע של $n-2m$ האיברים שמוצעו ב- $\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})$ עם האיבר הנוסף שהוא $\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})$ עצמו. מאחר ו- $\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})$ הוא הממוצע של $n-2m$ האיברים האחרים נקבל שהעיקביות נגזרת מעיקביות הממוצע כמדד מרכז.

3. הומוגניות:

$$\mathrm{f_n}(c\mathrm{x_1},\ldots,\mathrm{cx_n})=\frac{1}{n-2m}\sum_{\mathrm{i=m+1}}^{\mathrm{n-m}}\mathrm{cx_i}=c\left(\frac{1}{n-2m}\sum_{\mathrm{i=m+1}}^{\mathrm{n-m}}\mathrm{x_i}\right)$$

$$=c\cdot\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})$$

4. הזזה:

$$\mathrm{f_n}(\mathrm{x_1}+\mathrm{a},\ldots,\mathrm{x_n}+\mathrm{a})=\frac{1}{n-2m}\sum_{\mathrm{i=m+1}}^{\mathrm{n-m}}(\mathrm{x_i}+\mathrm{a})=a+\left(\frac{1}{n-2m}\sum_{\mathrm{i=m+1}}^{\mathrm{n-m}}\mathrm{x_i}\right)$$

$$=\mathrm{a}+\mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})$$

תשובה 5:

$$\frac{1}{n}\sum_{\mathrm{k=1}}^{\mathrm{n}}(\mathrm{x_k}-\overline{x})^2=\frac{1}{n}\sum_{\mathrm{k=1}}^{\mathrm{n}}\left((\mathrm{x_k})^2-2\mathrm{x_k}\overline{x}+\overline{x}^2\right)=\frac{1}{n}\sum_{\mathrm{k=1}}^{\mathrm{n}}\mathrm{x_k^2}-2\overline{x}\cdot\left(\frac{1}{n}\sum_{\mathrm{k=1}}^{\mathrm{n}}\mathrm{x_k}\right)+$$
$$\frac{1}{n}\sum_{\mathrm{k=1}}^{\mathrm{n}}\overline{x}^2=\frac{1}{n}\sum_{\mathrm{k=1}}^{\mathrm{n}}\mathrm{x_k^2}-2\overline{x}\cdot\overline{x}+\frac{n\cdot\overline{x}^2}{n}=\left(\frac{1}{n}\sum_{\mathrm{k=1}}^{\mathrm{n}}\mathrm{x_k^2}\right)-\overline{x}^2$$