<u>פיתרון תרגיל בית 2</u>

<u>תשובה 1:</u>

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

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

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

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

$$L_1 = 11$$

$$L_0 = 9$$

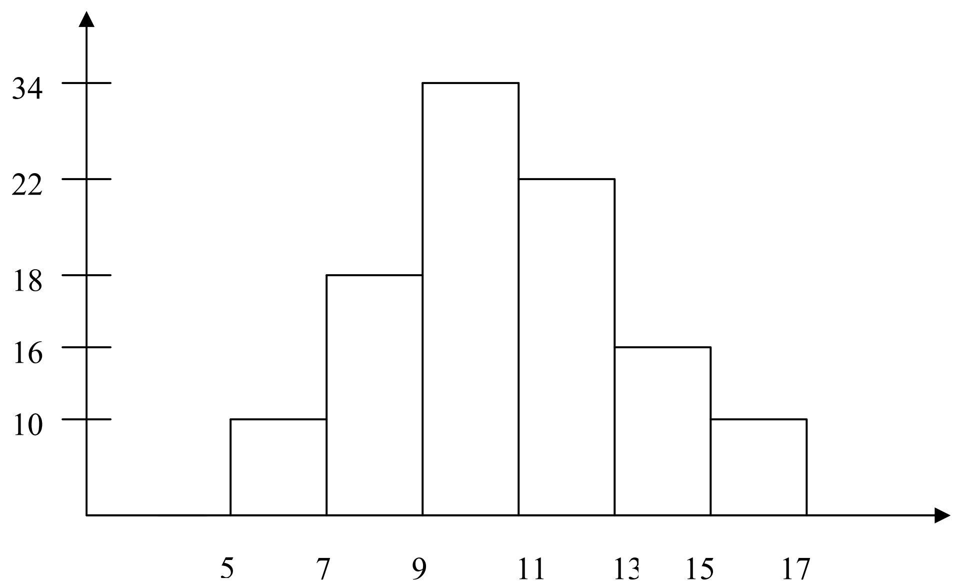
$$f(x_m) = 34$$

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

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

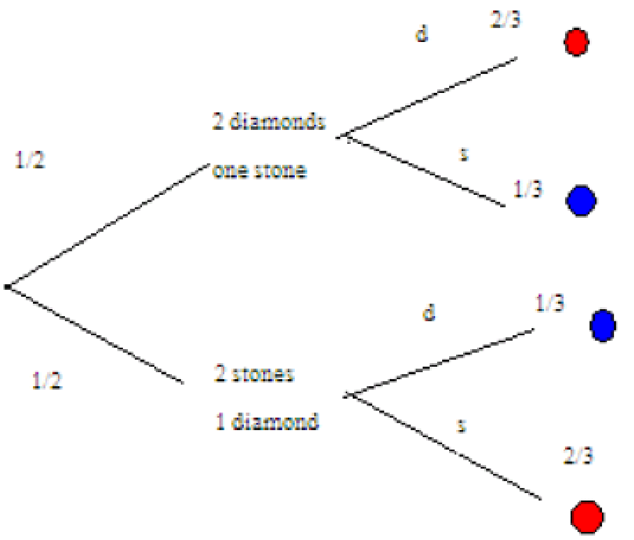
שכיח: הקבוצה (קטגוריה) השכיחה היא הקבוצה השלישית.

ב.



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

<u>תשובה 2:</u>



באדום – סיגל מקבלת את הקופסא הטובה. בכחול- את הפחות טובה.
ההסתברות לטובה : $\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)+\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)=\left(\frac{2}{3}\right)$
הסבר: יש הסתברות שווה שסיגל תבחר כל אחת מהקופסאות. אם בחרה את הקופסא עם 2 היהלומים (העליונה בתרשים) יש לה סיכוי של שני שלישים להציא יהלום (ואז היא בוחרת את הקופסא הזו) וסיכוי של שליש להוציא אבן (ולהחליף קופסא). באופן אנלוגי מפרשים את החלק התחתון בתרשים.

<u>תשובה 3:</u>
בהחלט יתכן. להלן דוגמא :

הפקולטה למדעים מדוייקים		הפקולטה לחוכמת הרחוב		
ניגשו	התקבלו	ניגשו	התקבלו	
1000	200	10	9	נשים
10	1	1000	500	גברים

סה"כ ניגשו 1010 גברים ו- 1010 נשים . התקבלו 501 גברים ורק 209 נשים. לכן ההסתברות לקבלת אישה לאוניברסיטה קטנה יותר. עם זאת 20 אחוז מהנשים שנגשו למדעים התקבלו, מול 10 אחוז מהגברים. גם בפקולטה לחוכמת הרחוב יש 90% קבלה לנשים מול 50% בלבד לגברים.
העניין הוא פשוט. הגברים ברובם הלכו לפקולטה לחכמת הרחוב, בה סיכויי הקבלה גבוהים יותר מבפקולטה למדעים, לעומתם רוב הנשים העדיפו ללכת לפקולטה למדעים, אליה כמובן סיכויי הקבלה קטנים בהרבה. דהיינו אם ניתבונן בנוסחאת ההסתברות השלמה לקבלת אישה מול מקבילתה לקבלת גבר (כאשר החלוקה היא לפקולטות) ההסתברויות המותנות (בפקולטות) גדולות יותר אצל נשים, אבל החלוקה לפקולטות שהיא בעצם שיקלול ההסתברויות המותנות לקבלת הסתברות הקבלה לאוניברסיטה גורמת לגברים ליהיות בעלי סיכוי קבלה גבוה יותר.

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

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

2.

$$\begin{aligned}
&\mathrm{f_n}(\mathrm{cx_1},\ldots,\mathrm{cx_n}) = \mathrm{M}ed_{1\le k\le n}(\{|c\cdot x_k - Med_{1\le k\le n}\{c\cdot x_k\}|\} =\\
&= \mathrm{M}ed_{1\le k\le n}(\{|c\cdot x_k - c\cdot Med_{1\le k\le n}\{x_k\}|\} =\\
&= \mathrm{M}ed_{1\le k\le n}(\{|c\cdot (x_k - Med_{1\le k\le n}\{x_k\})|\ \} =\\
&= |\mathrm{c}|\cdot \mathrm{M}ed_{1\le k\le n}(\{|x_k - Med_{1\le k\le n}\{x_k\}|\} = |\mathrm{c}|\cdot \mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})
\end{aligned}$$

3.

$$\begin{aligned}
&\mathrm{f_n}(\mathrm{x_1}+\mathrm{a},,,\mathrm{x_n}+\mathrm{a}) = \mathrm{M}ed_{1\le k\le n}(\{|(\ x_k + a) - Med_{1\le k\le n}\{\ x_k + a\ \}|\} =\\
&= \mathrm{M}ed_{1\le k\le n}(\{|\ x_k + a - Med_{1\le k\le n}\{\ x_k\} - a|\} =\\
&= \mathrm{M}ed_{1\le k\le n}(\{|\ x_k - Med_{1\le k\le n}\{\ x_k\}|\} = \mathrm{f_n}(\mathrm{x_1},\ldots,\mathrm{x_n})
\end{aligned}$$

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

3. הומוגניות:

$$f_n(cx_1, \dots, cx_n) = \frac{1}{n-2m}\sum_{i=m+1}^{n-m} cx_i = c\left(\frac{1}{n-2m}\sum_{i=m+1}^{n-m} x_i\right)$$

$$= c \cdot f_n(x_1, \dots, x_n)$$

4. הזזה:

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

$$= a + f_n(x_1, \dots, x_n)$$

<u>תשובה 5:</u>

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