1- حصل عدد من الطلاب فى مادة الإحصاء على الدرجات التالية :

5	4	4	5	3	4	2	3	1	2
3	7	4	1	6	3	2	5	3	4
7	3	2	6	5	3	4	2	4	1

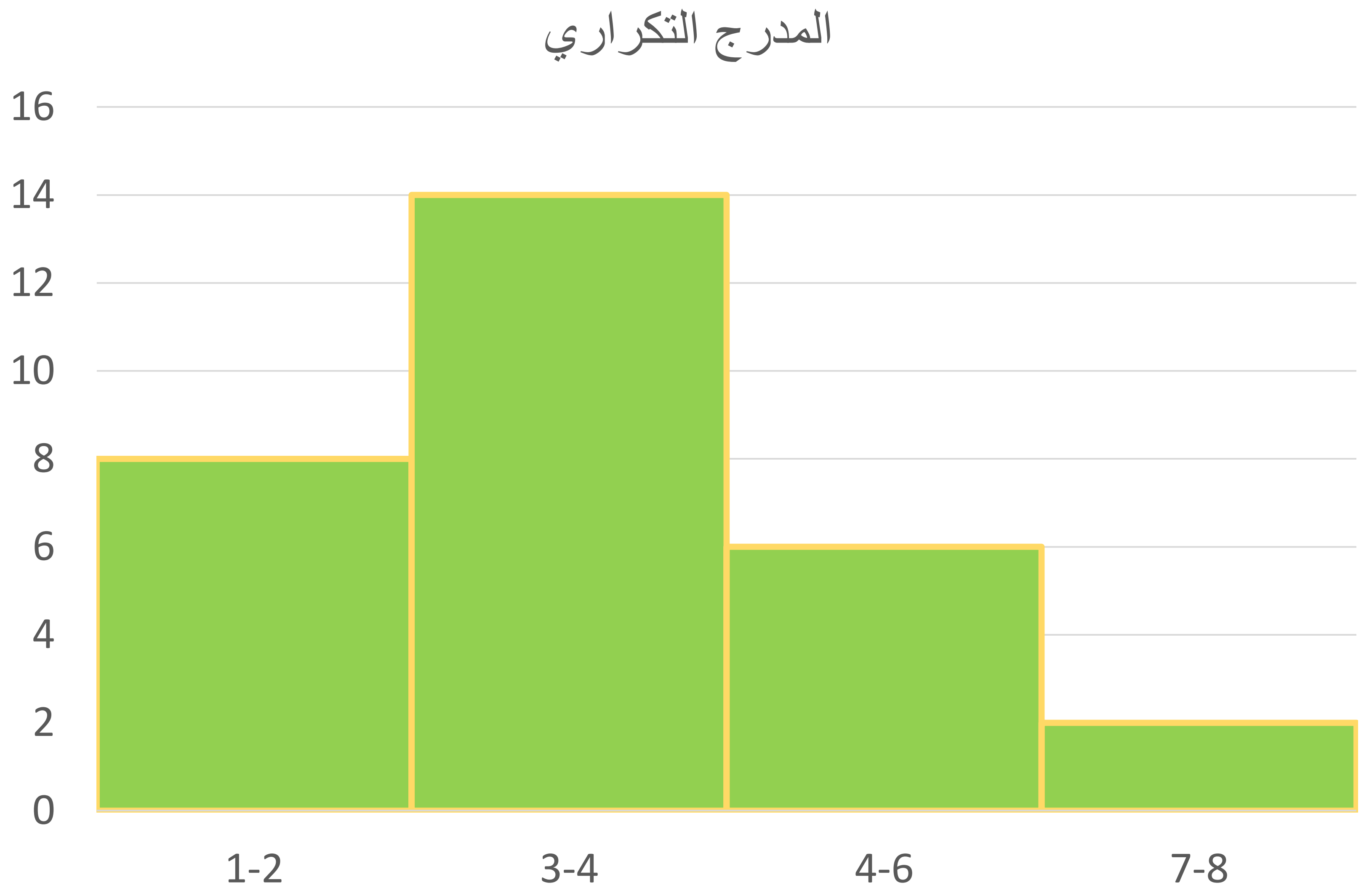
المطلوب : تكوين جداول تكرارية والمدرج ، المضلع، ومنحنيات التكرار والتجميعية الصاعد والنازل لهذه الدرجات.

Range= 7-1=6

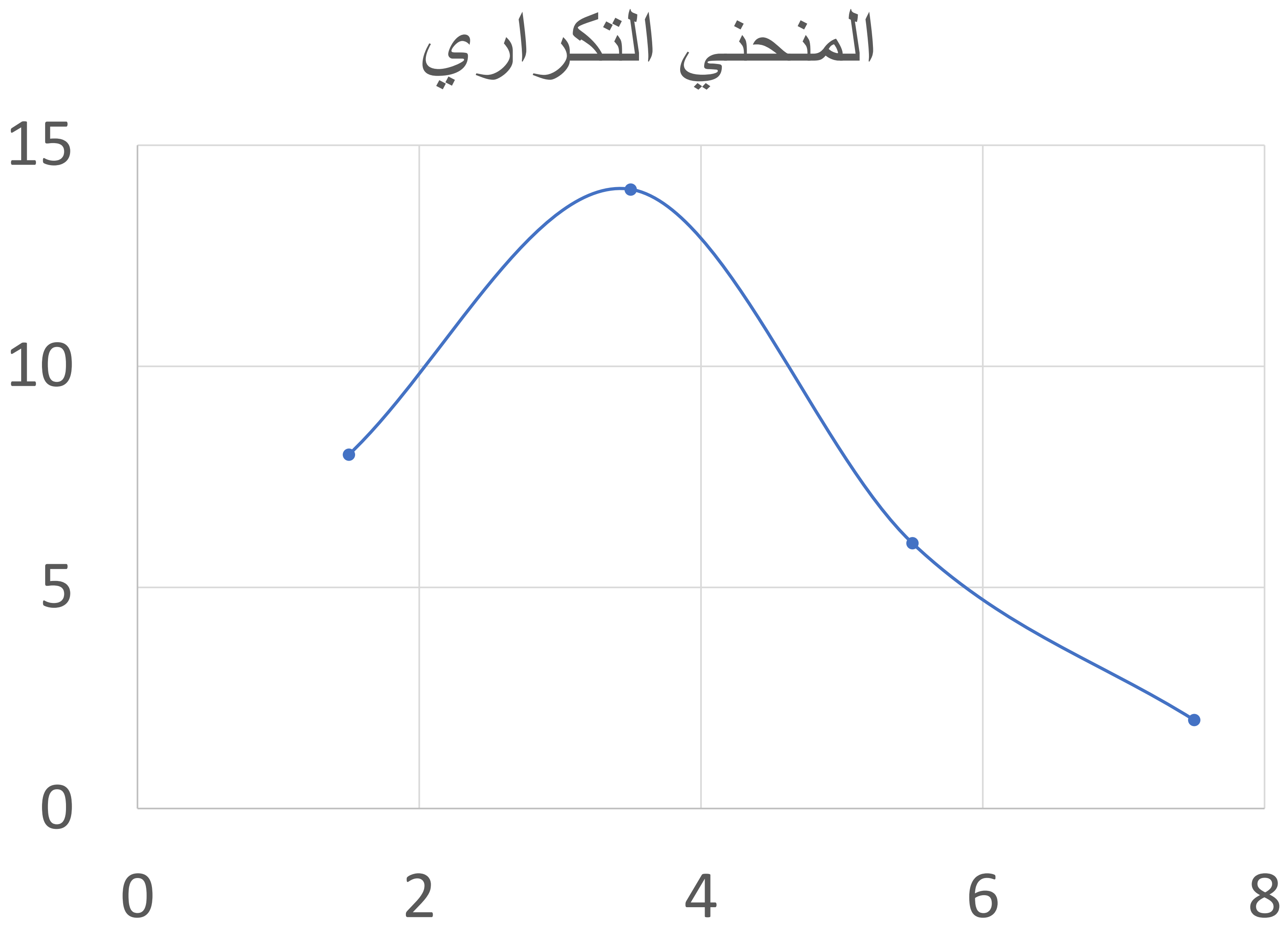
لنفرض عدد الفئات هو (4) اذا طول الفئة سيكون

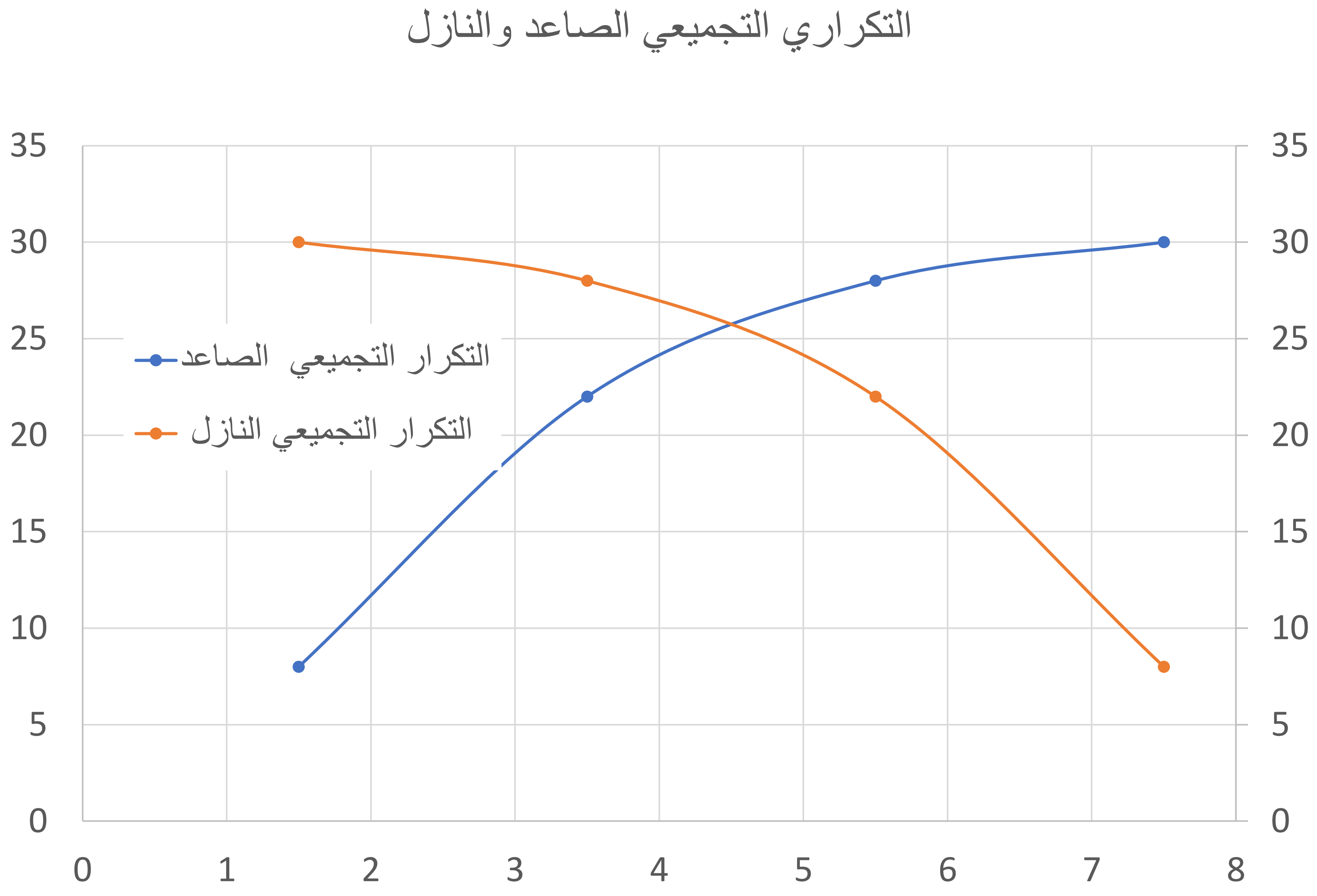
Class length= 6/4=1.5 ≈2

Range المدى	عدد الفئات	طول الفئة	طول الفئة		الفئات Classes	التكرار Frecuency	حدود الفئات الحقيقية	مركز الفئة - Class center	التكرار النسبي - Relative Frecuency	التكرار المئوي - Percentage Frecuency	التكرار التجميعي الصاعد	التكرار التجميعي النازل	مقلوب التكرار
6	4	1.5	2		1-2	8	0.5-2.5	1.5	0.266666667	26.66666667	8	30	2
					3-4	14	2.5-4.5	3.5	0.466666667	46.66666667	22	28	6
					4-6	6	4.5-6.5	5.5	0.2	20	28	22	14
					7-8	2	6.5-8.5	7.5	0.066666667	6.666666667	30	8	8
						30			1	100			


المدرج التكراري
16
14
12
10
8
6
4
2
0
1-2
3-4
4-6
7-8


المضلع التكراري
16
14
12
10
8
6
4
2
0
1
2
3
4


المنحني التكراري
15
10
5
0
0
2
4
6
8


التكراري التجميعي الصاعد والنازل
التكرار التجميعي الصاعد
التكرار التجميعي النازل
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6
7
8

مقاييس التمركز او التوسط (فص 4)

إن الأسلوب البياني في تحليل ودراسة الظواهر لتحديد الخصائص والاتجاهات والعلاقات ، يعتمد في دقته على دقة التمثيل البياني نفسه وبذلك ربما تختلف الخصائص من رسم إلى آخر لنفس الظاهرة، وعليه فإنه من الأفضل اللجوء إلى طرق القياس الكمي، حيث يستخدم الباحث الطريقة الرياضية في القياس.

ومن أهم مقاييس النزعة المركزية التي سنتعرض إليها بالدراسة الوسط الحسابي والوسيط والمنوال لكل من البيانات المفردة (غير المبوبة) ومن البيانات المبوبة .

1. **الوسط الحسابي (المتوسط)** يعرف بأنه مجموع قيم المشاهدات y_i مقسوماً على عددها n ويرمز له بالرمز ($\bar{y}$).

أ- **حساب الوسط الحسابي من البيانات غير المبوبة (المفردة)**

يحسب المتوسط الحسابي من البيانات غير المبوبة من العلاقة الاتية:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$; **Example:** y_i= 400, 380, 450, 350, 520

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{400+380+450+350+520}{5} = \frac{2100}{5} = 420$$

ب-حساب الوسط الحسابي من البيانات المبوبة ويحسب بكثر الطرق شيوعا من خلال استخدام مراكز الفئات وتكراراتها وفق العلاقة الاتية:

حيث ان $\bar{y} = \frac{\sum_{i=1}^{n} f_i y_i}{\sum_{i=1}^{n} f_i}$;

$y_i =$ مركز الفئة ذات التسلسل i ; $f_i =$ تكرار الفئة ذات التسلسل i

مثال (4): استخرج الوسط الحسابي لأطوال نباتات القطن المبوبة في الجدول التكراري الاتي:

الفئات	التكرار f	مركز الفئات y	f*y
40-31	1	35.5	35.5
50-41	2	45.5	91.0
60-51	5	55.5	277.5
70-61	15	65.5	982.5
80-71	25	75.5	1887.5
90-81	20	85.5	1710.0
100-91	12	95.5	1146.0
	$\sum f_i = 80$		$\sum f_i * y_i = 6130.0$

الحل:

$$\bar{y} = \frac{\sum_{i=1}^{n} f_i y_i}{\sum_{i=1}^{n} f_i} = \frac{6130.0}{80} = 76.62\ cm$$

خواص الوسط الحسابي

i. مجموع انحرافات القيم عن وسطها الحسابي = صفر
- للبيانات غير المبوبة $\Sigma(y_i - \bar{y})=0$
- للبيانات المبوبة $\Sigma f_i (y_i - \bar{y})=0$
- البرهان

$$\Sigma(y_i - \bar{y})=\Sigma y_i - \Sigma \bar{y} = \Sigma y_i - n\,\bar{y} = \Sigma y_i - \Sigma y_i =0$$

$$\Sigma f_i (y_i - \bar{y})= \Sigma f_i y_i - \bar{y}\, \Sigma f_i= \Sigma f_i y_i - \left(\frac{\Sigma fi\, yi}{\Sigma fi}\right) \Sigma f_i =\Sigma f_i y_i - \Sigma f_i y_i =0$$

ii. مجموع مربعات الانحرافات عن الوسط الحسابي اقل ما يمكن $\min\Sigma(y_i - \bar{y})^2$

iii. عند إضافة عدد ثابت (k) الى كل قيمة من المتغيرات فان وسطها الحسابي = الوسط الحسابي للقيم الاصلية + العدد الثابت k

$$A_i = y_i + k \rightarrow \bar{A} = \bar{y} +k$$

iv. اذا ضربت قيم المتغير بثابت k فان متوسط القيم الناتجة يساوي حاصل ضرب الثابت k بالوسط الحسابي للمتغير

$$A_i = y_i * k \rightarrow \bar{A} = \bar{y} *k$$

2. الوسيط Ṁe

يعرف الوسيط على أنه القيمة التي تتوسط مجموعة من القيم بحيث تكون التكرارات التي تسبقها مساوية (50%) لتلك التي تليها إذا رتبت مجموعة القيم ترتيباً تصاعدياً أو تنازلياً، ويعتمد حسابه على طبيعة ونوع البيانات.

أ- للبيانات غير المبوبة

i. **إذا كان عدد المفردات فردى (n فردية):** سيكون فيها مفردة واحدة تمثل الوسيط ويحسب ترتيبه من العلاقة: (n+1) / 2

مثال (5): احسب الوسيط من البيانات التالية (20 – 12 – 15 – 10 – 40 – 80 – 61)

الحل : نرتب المفردات ترتيبا تصاعديا أولاً : (10 12 15 20 40 61 80)

نحسب ترتيب الوسيط = (7 + 1) / 2 = 4 ، ترتيب الوسيط هو الرابع، فالوسيط **Ṁe** = 20 .

ii. **إذا كان عدد المفردات زوجي (n زوجيه):** سيكون فيها مفردتان محاذيتان للوسيط الذي يحسب عن طريق إيجاد الوسط الحسابي لهما، ويحسب ترتيبهما من العلاقة :

$$\left(\frac{n}{2}, \quad \frac{n+1}{2}\right)$$

مثال (6): احسب الوسيط من البيانات التالية : (15 – 12 – 15 – 14 – 18 – 20 – 33 – 40)

الحل :

نرتب المفردات ترتيبا تصاعديا أولاً : (12 14 15 15 18 20 33 40)

نحسب ترتيب الوسيط : $\left(\frac{8}{2}, \quad \frac{8+1}{2}\right)$ = (4 ، 5) ، ترتيب الوسيط بين المفردتان الرابعة والخامسة وقيمة الوسيط تمثل متوسط القيمتين اللتان ترتيبهما الرابع والخامس .

الوسيط **Ṁe** = (15 + 18) / 2 = 16.5 .

ب-للبيانات المبوبة

i. **باستخدام الجدول التكراري التجميعي الصاعد** وفقا للعلاقة الاتية

$$\dot{M}_e = L_m + \left(\frac{\sum_{i=1}^{n} \frac{f_i}{2} - F_i}{f_m}\right) * C$$

حيث ان:

L_m = الحد الأدنى الحقيقي للفئة الوسيطة (أي الفئة آلتي تضم المفردة الوسيطة للتوزيع)،

n = عدد المفردات في مجموعة البيانات.

Fi = مجموع التكرارات للفئات التي تسبق الفئة الوسيطة.

F_m= تكرار الفئة الوسيطة.

C = طول الفئة.

مثال (7): احسب الوسيط لأطوال نباتات القطن المبوبة في جدول المثال (4) أعلاه.

الحل:

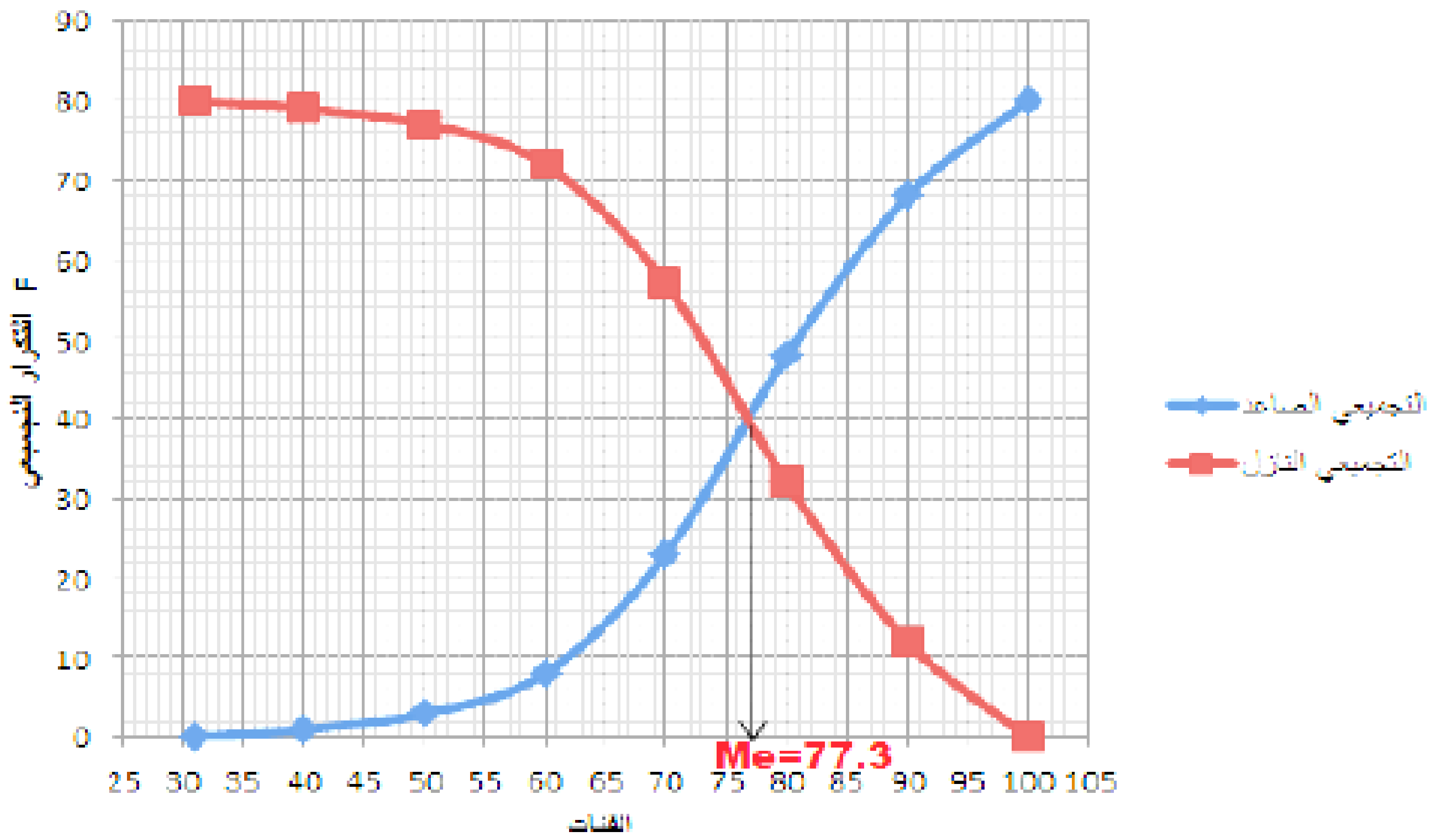
الفئات	التكرار f	تك ت ص F
31 >	0	0
40-31	1	1
50-41	2	3
60-51	5	8
70-61	15	23
80-71	25	48
90-81	20	68
100-91	12	80
	$\sum f_i = 80$	

1- التكرار الذي يمثل الوسيط = $\sum f/2 = 80/2 = 40$
2- الفئة التي يقع فيها الوسيط هي (80-71)
3- نحسب قيمة الوسيط وفق العلاقة

$$Me = L_m + \left(\frac{\sum f/_2 - F_l}{f_m}\right) * C$$

$$Me = 70.5 + \left(\frac{40-23}{25}\right) * 10 = 77.3\ cm$$

وهي نفس القيمة التي سنحصل عليها بطريقة رسم التكرارين التجميعيين الصاعد والنازل وكما موضح بالشكل رقم (م 7 – 1) ادناه



الشكل رقم (م 7-1) حساب الوسيط برسم التكرارين التجميعيين الصاعد والنازل

3. **المنوال** هو القيمة الأكثر تكراراً في مجموعة البيانات
أ- غير المبوبة .

مثال (8): احسب المنوال في كل من الحالات التالية :-

7 – 8 – 9 – 8 – 10 – 8 - 12 المنوال $\dot{M}o$ = 8

10- 12 – 10 – 15 – 12 – 10 المنوال $\dot{M}o$ = 10

15 – 16 – 15 – 20 - 16 – 30 المنوال $\dot{M}o$ = 15 ، 16

20 – 30 – 40 – 140 – 50 -60 المنوال $\dot{M}o$ = لا يوجد (Φ)

ب- أما بالنسبة **للبيانات المبوبة** يمكن حساب المنوال بأحد الطرق التالية:
i. وفق علاقة بيرسون الاتية:

$$\dot{M}o = L_m + \left(\frac{D_1}{D_1+D_2}\right) * C$$

حيث ان:

L_m = الحد الأدنى الحقيقي للفئة المنوالية (أي الفئة ذات التكرار الاكبر)،

D1 = الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي تسبقها.

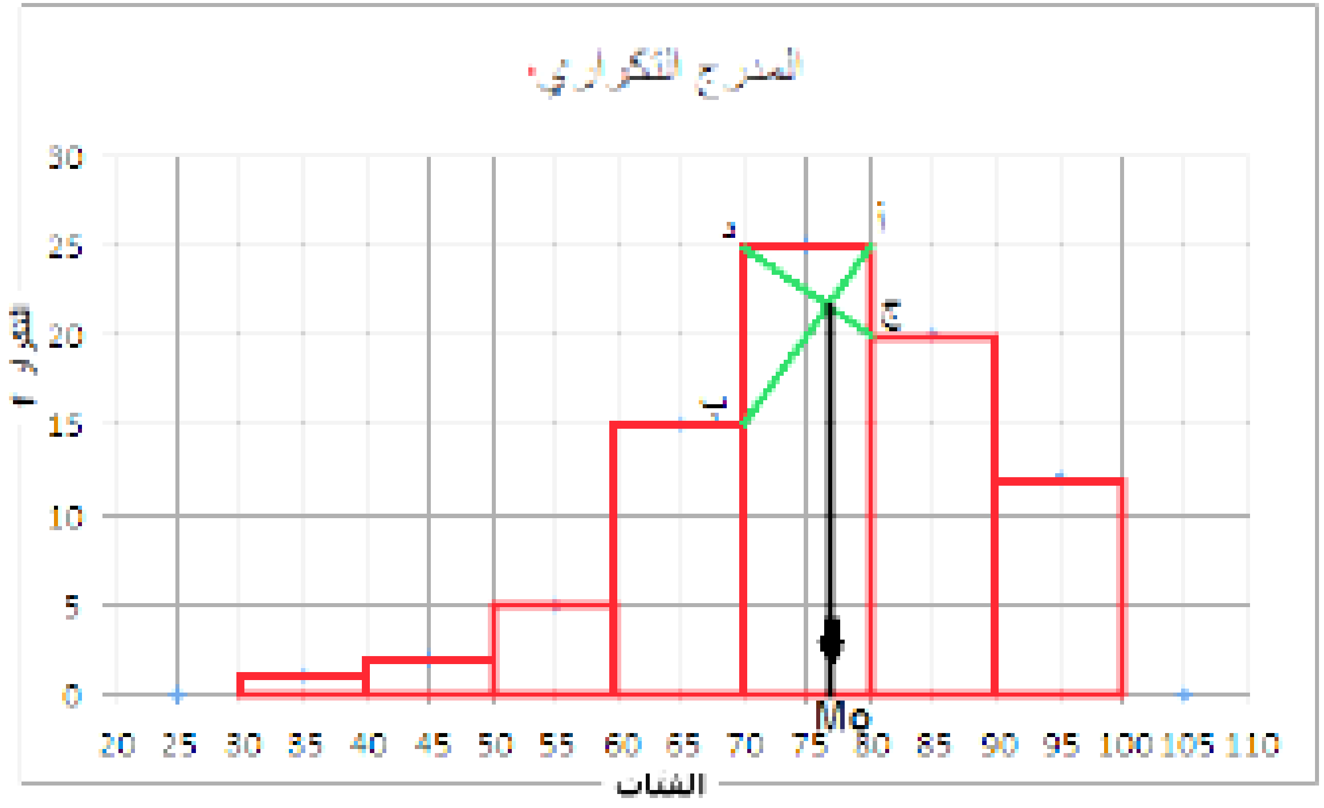
D2 = الفرق بين تكرار الفئة المنوالية والفئة التي تليها.

C = طول الفئة.

مثال (9): احسب المنوال لأطوال نباتات القطن المبوبة في جدول المثال (4) أعلاه.

$$\dot{M}o = L_m + \left(\frac{D_1}{D_1+D_2}\right) * C = 70.5 + \left(\frac{25-15}{25-15+25-20}\right) * 10 = 77.167\ cm$$

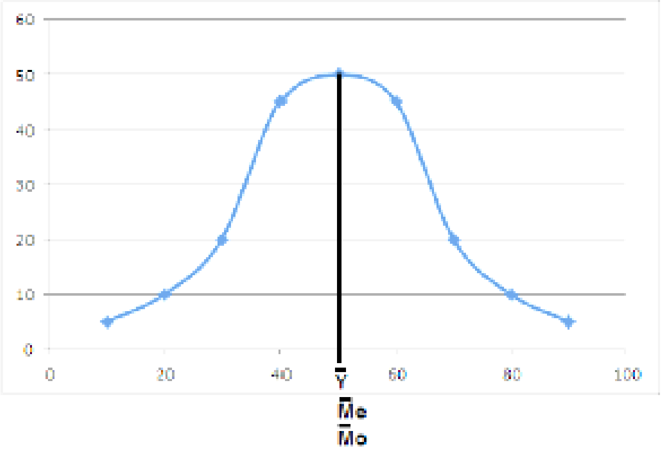
ii. بالطريقة البيانية من رسم المدرج التكراري وكما موضح بالشكل (م 9-1):



الشكل (م 9-1) حساب المنوال بالطريقة البيانية

4. العلاقة بين مقاييس التمركز للتوزيعات ذات المنوال الأحادي

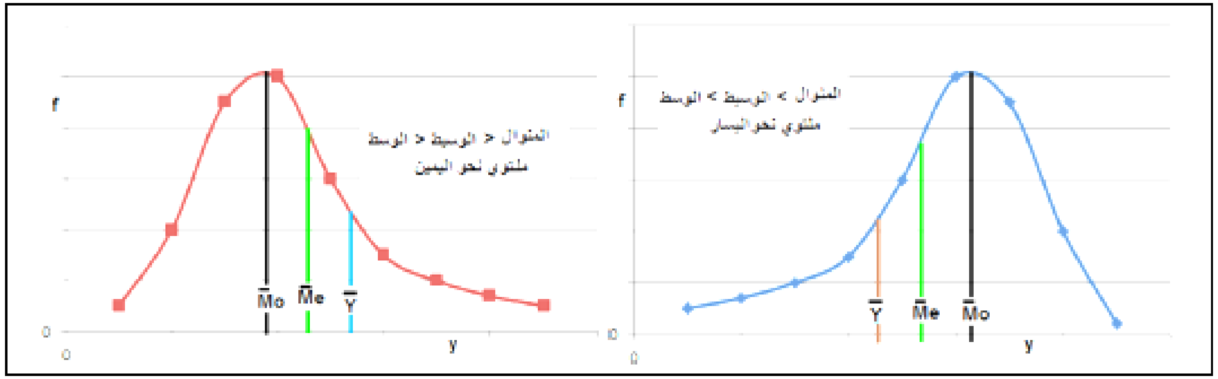
أ- في التوزيعات المتماثلة يكون الوسط = الوسيط = المنوال وكما موضح بالشكل الاتي:



ب- في التوزيعات غير المتماثلة (الملتوية) : لوحظ وجود علاقة خطية تقريبية بين مقاييس التمركز الثلاثة، وهي

الوسط – المنوال = 3 (الوسط – الوسيط)

ويكون شكل العلاقة بين تلك المقاييس في التوزيعات الملتوية بالشكلين الأتيين.



مقاييس التشتت او الاختلاف (فص5)

لا تعتبر مقاييس التمركز كافية لوصف مجموعة من البيانات وصفاً كاملاً فقد تتساوى بعض العينات في الوسط الحسابي بالرغم من اختلاف توزيع بياناتها حول مركزها (درجة تجانس البيانات)، مثل المجموعتين التاليتين من البيانات.

(150، 125، 100، 75 ، 50) ، (105، 103، 100 ، 97 ، 95)

حيث يلاحظ ان لهما نفس الوسط الحسابي والوسيط وهما 100 في حين المجموعتان مختلفتان من ناحية تشتتهما حول المركز او الوسط. إن الدرجة التي تتجه بها البيانات الرقمية للانتشار حول قيمة وسطى تسمى تشتت أو توزيع البيانات .

ومن أهم مقاييس التشتت المدى والتباين والانحراف المعياري والانحراف المتوسط .

1. المدى Range

المدى هو الفرق بين أكبر قيمة وأصغر قيمة .

مثال (1): لناخذ المجموعتين التاليتين من البيانات، ونحسب المدى لكل منهما :

Y	12	6	7	3	15	10	18	5
X	9	22	12	22	12	22	12	24

الحل:

أ- نرتب البيانات ترتيبا تصاعديا (تنازليا) ،
ب- نحسب المدى لكل منهما، وكما مبينا بالجدول الاتي

Y	18	15	12	10	7	6	5	3	R=18-3=15;	$\bar{y}$=9.5
A	9	12	12	12	22	22	22	24	R=24-9=15;	$\bar{A}$=16.875

يلاحظ ان مداهما متساويان مما يوحي ان المدى احيانا يكون مضللا كونه يعتمد على القيمتين الطرفيتين واللتين كثيرا ما تكونان شاذتين، لذلك ظهرت الحاجة الى مقاييس اكثر وضوحا للتشتت

2. الانحراف المتوسط The Mean Deviation

وهو متوسط الانحرافات المطلقة عن الوسط الحسابي، ويرمز له بالرمز M.D ويتم حسابه وفق:

أ- للبيانات غير المبوبة $M.D = \frac{\Sigma|y_i-\bar{y}|}{n}$

مثال(2): جد الانحراف المتوسط للقيم التالية:

y_i=9, 8, 6, 5, 7

الحل:

y_i	9	8	6	5	7	$\Sigma y_i=35$; $\bar{y}=7$
$Y_i - \bar{y}$	2	1	-1	-2	0	$\Sigma(y_i - \bar{y})=0$
$\|y_i - \bar{y}\|$	2	1	1	2	0	$\Sigma\|y_i - \bar{y}\|=6$; **M.D= 6/5=1.2**

ب-للبيانات المبوبة

$$M.D = \frac{\sum f_i |y_i - \bar{y}|}{\sum f_i}$$

مثال (3): جد الانحراف المتوسط لجدول التوزيع التكراري الاتي:

الفئات	60-62	63-65	66-68	69-71	72-74
f_i	5	18	42	27	8

الحل:

الفئات	fi	yi	fi*yi		|yi- ȳ|	fi |yi - ȳ |	
60-62	5	61	305	ȳ=6745/100=67.45	6.45	32.25	$M.D = \frac{\sum f_i \lvert y_i - \bar{y}\rvert}{\sum f_i}$ M.D=226.5/100 =2.265
63-65	18	64	1152		3.45	62.10	
66-68	42	67	2814		0.45	18.90	
69-71	27	70	1880		2.55	68.85	
72-74	8	73	584		5.55	44.40	
Σ	100		6745			226.50	

3. التباين والانحراف القياسي Variance and Standard Deviation

نلاحظ ان مجموع انحرافات عناصر (مفردات) العينة عن وسطها الحسابي $\Sigma(y_i - \bar{y})$ يساوي صفر ($\Sigma(y_i - \bar{y}) = 0$) ذلك بسبب كون قسم من الانحرافات موجبا بينما الاخر سالبا ، وللتغلب على هذه المشكلة فقد تم معالجتها بأخذ القيم المطلقة للانحرافات في الانحراف المتوسط أعلاه، ويمكن معالجتها بطريقة أخرى الا وهي تربيع قيم الانحرافات للحصول على مجموع مربعات الانحرافات (Squares Sum) والتي يرمز لها بالرمز SS .

$$SS=\Sigma(y_i - \bar{y})^2$$

ولكي نأخذ في الاعتبار حجم العينة حتى يمكن المقارنة بين العينات المختلفة الاحجام يتم تقسيم SS على درجات الحرية (n-1) مما يؤدي الى الحصول على التباين (Variance S^2)، ويحسب وفق؛

أ- للبيانات غير المبوبة في العينة

$$S^2 = \frac{\Sigma(y_i-\bar{y})^2}{n-1} = \frac{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}}{n-1}$$

اما في حالة المجتمع فتحسب وفق

$$\sigma^2 = \frac{\Sigma(y_i-\mu)^2}{N}$$ حيث ان

μ = وسط المجتمع الحسابي

N = عدد مفردات المجتمع

وسبب اخذ (n-1) في حالة العينة هو انه عند سحب عينة فان (n-1) من المشاهدات هي قيم حرة اما المشاهدة المتبقية (الأخيرة) لابد ان يكمل انحرافها مجموع الانحرافات عن الوسط الحسابي الى الصفر، لذلك فان عدد القيم الحرة في أي عينة هو (n-1) وهو ما سمي بدرجات الحرية.

ولاجل إعادة وحدات القياس لمفردات العينة الى اصلها يجب ان يؤخذ الجذر التربيعي للتباين للحصول على ($S=\sqrt{S^2}$) وهو ما يسمى بالانحراف القياسي.

$$\rightarrow \quad S=\sqrt{S^2}=\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n-1}}=\sqrt{\frac{\Sigma y_i^2-\frac{(\Sigma y_i)^2}{n}}{n-1}}$$

مثال (4): جد الانحراف القياسي للبيانات التالية باستخدام الطريقة المختصرة؛

Yi= 9, 8, 6, 5, 7

الحل: الطريقة المختصرة تتمثل بالعلاقة $S=\sqrt{\frac{\Sigma y_i^2-\frac{(\Sigma y_i)^2}{n}}{n-1}}$ بذلك لاجل اكمال الحل تأمين عناصر تلك العلاقة؛

y_i	9	8	6	5	7	$\Sigma y_i = 35$	$S=\sqrt{\frac{255-\frac{(35)^2}{5}}{4}}=\sqrt{\frac{255-245}{4}}=\sqrt{\frac{10}{4}}=1.58$
Y_i^2	81	64	36	25	49	$\Sigma Y_i^2 = 255$	

ب- وللبيانات المبوبة يحسب وفق العلاقة

$$S = \sqrt{S^2} = \sqrt{\frac{\sum f_i (y_i - \bar{y})^2}{\sum f_i - 1}} = \sqrt{\frac{\sum f_i . y_i^2 - \frac{(\sum f_i . y_i)^2}{\sum f_i}}{\sum f_i - 1}}$$

مثال (5): جد الانحراف القياسي للبيانات المبينة بالعمودين الأولين في الجدول التالي مستخدما الطريقة الطويلة:

الفئات	f_i	Solution الحل				
		y_i	$Y_i - \bar{y}$	$(y_i - \bar{y})^2$	$f_i * (y_i - \bar{y})^2$	$S = \sqrt{\frac{\sum f_i (y_i - \bar{y})^2}{\sum f_i - 1}}$
60 - 62	5	61	-6.45	41.6025	208.0125	
63 - 65	18	64	-3.45	11.9025	214.2450	$= \sqrt{\frac{852.75}{99}} = \sqrt{8.6}$
66 - 68	42	67	-0.45	0.2025	8.5050	
69 - 71	27	70	2.55	6.5025	175.5675	S = 2.9
72 - 74	8	73	5.55	30.8025	246.4200	
	Σ=100	$\bar{y}$=67.45		Σ	852.7500	

ت- اهم خواص التباين والانحراف القياسي

i. عند إضافة (او طرح) عدد ثابت k الى (من) قيم المشاهدات، فانه لا يؤثر على قيمة التباين والانحراف القياسي لتلك المشاهدات.

If $X_i = y_i + k$ or $X_i = y_i - k$ then $S_x^2 = S_y^2$ or $S_x = S_y$

مثال (6):

$Yi = 8, 3, 2, 12, 10;$ $Xi = yi + 3 = 11, 6, 5, 15, 13$

الحل:

Yi	8	3	2	12	10	$\sum y_i=35$
Yi^2	64	9	4	144	100	$\sum y_i^2=321$
Xi	11	6	5	15	13	$\sum X_i=50$
Xi^2	121	36	25	225	169	$\sum X_i^2=576$

$$S = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}} \; ; S_y = \sqrt{\frac{321-\frac{(35)^2}{5}}{4}} = \sqrt{19}$$

$$S_x = \sqrt{\frac{576-\frac{(50)^2}{5}}{4}} = \sqrt{19}$$

$$\rightarrow S_y = S_x$$

ii. اذا ضربت كل قيمة من قيم المشاهدات بعدد ثابت (k) فان:

1) تباين القيم الناتجة = تباين القيم الاصلية * مربع k

2) الانحراف القياسي للقيم الناتجة = الانحراف القياسي للقيم الأصلي * k

If $Xi = yi * k$ then $S_x^2 = S_y^2 * k^2$ or $S_x = k * S_y$

iii. اذا كان

If $Z_i = y_i + X_i$ then $S_Z^2 = S_y^2 + S_X^2$

iv. اذا كانت مجموعتان من القيم (n2, n1) ولهما تباين (S_2^2 , S_1^2) فان التباين المتجمع لجميع مشاهدات المجموعتان (Pooled Var. S_p^2) هو :

$$S_p^2 = \frac{SS_1+SS_2}{n_1+n_2-2} = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$

4. معامل الاختلاف Coeff. Of Var. (C. V.)

اذا كان S و $\bar{y}$ هما الانحراف القياسي والوسط الحسابي لمجموعة من القيم على التوالي فان معامل الاختلاف لها هو:

$$C.V. = \frac{s}{\bar{y}} * 100$$

مثال (7): نتائج الامتحانات النهائية لمادتين في الصف الأول كانتا كما موضحا بالجدول التالي، حدد في أي المادتين كان التشتت اكثر؟

	الرياضيات	الحاسوب
الوسط الحسابي	78	73
الانحراف القياسي	8	7.6

الحل:

بالرياضيات $C.V_M. = \frac{8}{78} * 100 = 10.25\%$

بالحاسوب $C.V_C. = \frac{7.6}{73} * 100 = 10.41\%$

يلاحظ ان التشتت في درجات الحاسوب كان اكثر من الرياضيات، ولكن لو قارنا من خلال الانحراف القياسي لوجدنا ان التشتت بدرجات الرياضيات اكثر من الكيمياء.

5. الدرجة القياسية Standardized Scores

عندما يتم المقارنة بين مفردتين من مجموعتين مختلفتين يستوجب تحويلهما الى وحدات قياسية لاجل ان تكون تلك المقارنة ذات مرجعية قياسية وذو دلالة حقيقية، ويتم ذلك من خلال التحويل الى الدرجة القياسية التي تحسب وفق العلاقة:

$$Z_i = \frac{y_i - \bar{y}}{s}$$

مثال (8): حصل طالب على درجة (84) في مادة الرياضيات التي متوسط درجات الطلبة فيها هو (76) وبانحراف قياسي قدره (10)، وفي مادة الفيزياء (90) التي متوسط درجات الطلبة فيها (82) وبانحراف قياسي (16)، ففي أي من الموضوعين كانت قابلية هذا الطالب اعلى؟

الحل:

$$Z_1 = \frac{84-76}{10} = 0.8$$

$$Z_2 = \frac{90-82}{16} = 0.5$$

يتضح ان قابلية الطالب في الرياضيات اعلى من الفيزياء