إحصاء وراثي

قسم المحاصيل الحقلية / دراسات عليا

الفصل الربيعي 2025

مدرس المادة: أ.م.د.داود سلمان مدب

المصادر

. 1-الراوي، خاشع محمود وعبدالعزيز محمد خلف الله. (1980) تصميم وتحليل التجارب الزراعية.مطابع التعليم العالي، جامعة الموصل، العراق.

2-Singh, R. K. and B. D. Chaudhary (2007). Biometrical Methods in Quantitative Genetic Analysis. Kalyani Publishers, India.

وصف المقرر

اسم المقرر الدراسي: إحصاء وراثي

عدد الوحدات 3

الأسبوع	الجزء النظري	الجزء العملي
1	مقاييس التوسط والتمركز	تطبيقات في تقدير المتوسطات
2	مقاييس التشتت	تمارين حول إيجاد التباينات
3	المصفوفات الجبرية	التدريب في حل عمليات المصفوفات
4	متوسط المربعات المتوقع (EMS)	امثلة تطبيقية في إيجاد متوسط المربعات المتوقع
5	التضريب الاختباري للسلالات (Line Tester)	حل مثال تطبيقي حول تحليل الLine Tester
6	اجراء امتحان الشهر الاول	اجراء امتحان الشهر الاول
7	التضريب ثنائي الاليل (Diallel Crossing)	تكوين برنامج تحليل احصائي للتضريب ثنائي الاليل
8	التصميم المتشعب (Nested Design)	حل مثال تطبيقي حول التصميم المتشعب
9	التصميم العاملي (Factorial Design)	تطبيق عملي حول التصميم العاملي
10	تصميم North Carolina Design	اجراء تدريبات في حل المسائل حول تصميم North Carolina Design
11	التهجين الثلاثي Tri Crossing	التدريب في تحليل التهجين الثلاثي
12	اجراء امتحان ثاني	اجراء امتحان عملي ثاني

Tikrit University
College of Agriculture
Department of Field Crops
Course Description

SG	**Statistical Genetics**	**Higher studies**	**Mandatory (Essential)**
Number of units: 3	**(3) Practical Hours**	**(2) Hours of Theoretical Lectures**	**Planned Weekly Teaching Hours**
Illustration of using statistics in genetical studies			**Description of the Curriculum**
Introducing the student to the main principles of the genetic parameters calculation in plant breeding			**The objective of teaching the curriculum**
Increasing The ability of students in genetic improvement through estimation of gene behavior			**Learning Outcomes**
Singh, R. K. and B. D. Chaudhary (2007). Biometrical Methods in Quantitative Genetic Analysis. Kalyani Publishers, India.			**The methodological book**

Semester Estimates	Theoretical Midterm Exams	Practical Midterm Exams	Daily Theoretical Tests	Final Practical Exam	Final Theoretical Exam
	15	10%	5%		%70

المواضيع

Week	Theoretical Material	Number of hours	Practical Material	Number of hours
1	Central measurements	2	Application in means calculation	3
2	Divation measurements	2	Exercise in Variances calculation	3
3	Algebra Matrices	2	Practices in matrices exercises	3
4	Expected mean squares	2	Applicable examples in expected mean suquares	3
5	(Line Tester Analysis)	2	Application in Line Tester	3
6	Theoretical first exam	2	Practical First exam	3
7	(Diallel Crossing)	2	Programming analysis in Diallel through excel	3

	software			
8	Nested Design	2	Solving examination in Nested design	3
9	Factorial Design	2	Practical application in Factorial design	3
10	North Carolina Design	2	Solving problems in North Carolina Design	3
11	Tri Crossing	2	Training in solving problems in triple-test cross	3
12	Second exam	2	Second practical exam	3

المحاضرة الاول

المقاييس الاحصائية (مقاييس التوسط والتشتت)

تعد ذات اهمية كبيرة في فهم الطرق الاحصائية المختلفة والمتضمنة تحليل التباين او الاختلاف بين القيم الداخلة في الدراسة

اولا–مقاييس التوسط او التمركز

سميت بهذا الاسم لان جميع القيم تتمركز حول قيمة محورية مركزية وتشمل كل من

1–المتوسط الحسابي

(1) الوسط الحسابي *Arithmetic Mean* أو *Average*

الوسط الحسابي لمجموعة من القيم هو القيمة التي لو أعطيت لكل مفردة من مفردات المجموعة لكان مجموع القيم الجديدة مساويا لمجموع القيم الأصلية ويرمز له بالرمز $\overline{X}$ ويستخدم الوسط الحسابي في حالة البيانات الرقمية او الكمية فقط.

الوسط الحسابي لمجموعة قيم = $\frac{\text{مجموع هذه القيم}}{\text{عددها}}$

X_n تعني القيمة الخيرة رقم n

ومجموع هذه القيم هو : $X_1 + X_2 + X_3 + \ldots + X_n$

والذي يمكن أن يكتب اختصاراً $\sum X$ أي مجموع القيم حيث :

X ترمز للقيم

والرمز اللاتيني " سيجما " $\sum$ يرمز للمجموع

أي أن :

$$\sum X = X_1 + X_2 + X_3 + \ldots + X_n$$

وبالتالي فإن الوسط الحسابي والذي يرمز له بالرمز $\overline{X}$ (والذي ينطق X بار) هو

$$\overline{X} = \frac{\sum X}{n} \quad (1)$$

ومجموع هذه القيم هو :

$$\sum X = X_1 + X_2 + X_3 + \ldots + X_n$$

$$= 7+10+12+8+9 = 46$$

وبالتالي فإن الوسط الحسابي $\overline{X}$ هو :

$$\overline{X} = \frac{\sum X}{n} = \frac{46}{5} = 9.2$$

اما إذا كان لدينا عينتين، حجم الأولى n_1 وحجم الثاني n_2 وكان الوسط الحسابي لكل منهما $\overline{X}_1$ ، $\overline{X}_2$ فإنه يمكن حساب الوسط الحسابي للعينتين معاً بالاستفادة من متوسط كل منهما كما يلي :

$$\overline{X} = \frac{n_1 \cdot \overline{x}_1 + n_2 \overline{x}_2}{n_1 + n_2} \quad (2)$$

ويسمى المتوسط في هذه الحالة " المتوسط المرجح " **The weighted average** ونلاحظ أنه تم ترجيح كل متوسط بحجم العينة المحسوب منها، أي إعطاء كل متوسط وزن يساوي حجم العينة الخاصة به، ثم القسمة على مجموع حجم العينتين معاً. أي أنه لحساب الوسط الحسابي للعينتين معاً. لا نجمع الوسطين ونقسم على اثنين إلا في حالة واحدة فقط وهي إذا كان حجم العينتين متساويين.

(2) الوسيط ***Median***:

يعرف الوسيط لمجموعة من البيانات بأنه القيمة التي تقع في وسط المجموعة تماماً بعد ترتيبها تصاعدياً أو تنازلياً، أي هو القيمة التي تقسم مجموعة البيانات إلى قسمين بحيث يكون عدد القيم الأكبر منها مساوياً عدد القيم الأصغر منها ويرمز له بالرمز M_e. ويستخدم الوسيط في حالة البيانات الترتيبية.

ترتيب الوسيط = عدد القيم + 1 / 2 ، $\frac{n+1}{2}$

أو يساوي

ففي المثال رقم (1)، عدد القيم 5 لذا فإن ترتيب الوسيط هو $\frac{5+1}{2}=\frac{6}{2}=3$ فالوسيط هو القيمة رقم 3 (بعد ترتيب القيم تصاعدياً) والقيمة الثالثة في الترتيب هي 29 وهي تمثل قيمة الوسيط لأعمار الناخبين في المثال الأول.

وفي المثال رقم (2) عدد القيم المعطاة هي 6 لذا فإن ترتيب الوسيط هو : $\frac{6+1}{2}=\frac{7}{2}=3.5$ أي أنه يقع بين القيمتين الثالثة والرابعة، لذلك فقد جمعنا هاتين القيمتين وقسمنا على 2، أي أن قيمة الوسيط لدرجات الطلاب في هذا المثال هي :

$$\frac{14+15}{2}=14.5$$

(3) المنوال *Mode*:

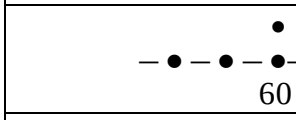
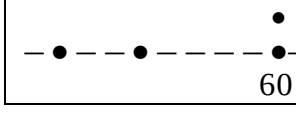
يعرف المنوال لمجموعة من البيانات بأنه القيمة الأكثر شيوعاً (تكراراً) في المجموعة ويرمز له بالرمز M_o. يفضل استخدام المنوال في حالة البيانات الوصفية والترتيبية.

ثانياً: مقاييس التشتت المطلق Measures of Dispersion

مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين مجموعات البيانات المختلفة. فقد تكون هناك مجموعات من البيانات لها نفس مقاييس النزعة المركزية (لها نفس الموضع) ولكنها تختلف في بعض الصفات الأخرى. فمثلاً المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتت البيانات.

مثال (1):

الجدول التالي يوضح مجموعتين لهما نفس المتوسط $\bar{x} = 60$ ونفس الوسيط $Med = 60$ ونفس المنوال $Mod = 60$ ولكنهما مختلفتان في طبيعة التشتت.

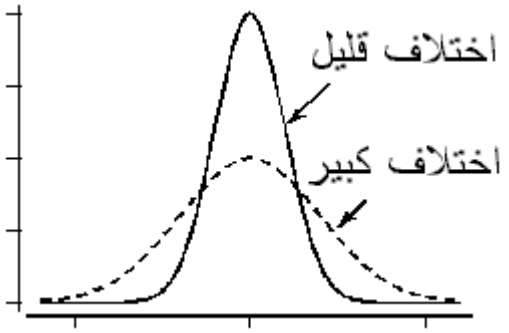
المجموعة	البيانات	شكل انتشار البيانات
الأولى	59, 61, 62, 60, 58, 60	
الثانية	50, 60, 66, 54, 60, 70	

بالرغم من تساوي مقاييس النزعة المركزية للمجموعتين إلا أن التشتت (أو الاختلاف) بين القيم في كل مجموعة غير متساو. فمن الواضح أن بيانات المجموعة الأولى أكثر تقاربًا فيما بينها (أقل تشتتًا وتباعدًا فيما بينها) من بيانات المجموعة الثانية. لذلك دعت الحاجة لإيجاد مقاييس تقيس طبيعة تشتت (أو تفرق أو اختلاف أو تباعد)

البيانات فيما بينها. هذه المقاييس تسمى مقاييس التشتت أو الاختلاف.

مثال (2):

الشكل التالي يوضح المضلعين التكراريين لمجموعتين من البيانات لهما نفس مقاييس النزعة المركزية ولكنهما مختلفتان في طبيعة التشتت.



شكل (1) المضلعان التكراريان لتوزيعين لهما نفس مقاييس النزعة المركزية ولكنهما مختلفين في التشتت

مقاييس التشتت هي مقاييس عددية تستخدم لقياس اختلاف أو تشتت البيانات. والاختلاف أو التشتت لمجموعة من البيانات هو مقدار تفرق أو تباعد أو انتشار البيانات فيما بينها. فتشتت البيانات يكون صغيرًا إذا كانت البيانات متقاربة فيما بينها والعكس بالعكس. وأما البيانات المتساوية فلا اختلاف ولا تشتت فيها. ومقاييس التشتت تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة إذ أن مقاييس النزعة المركزية لا تكفي وحدها لوصف مجموعة البيانات أو مقارنة مجموعات البيانات المختلفة. ومن أشهر مقاييس التشتت نذكر:

1. **المدى: Range**
2. **نصف المدى الربيعي Semi-Inter-quartile Range**
3. **التباين: Variance**
4. **الانحراف المعياري: Standard Deviation**
5. **معامل الاختلاف (أو التغير): Coefficient of Variation**

(2-4) المدى: Range

يعتبر المدى من أسهل مقاييس التشتت تعريفًا وحسابًا ويعطينا فكرة سريعة عن مدى تفرق البيانات.

تعريف (1):

نعرف المدى لمجموعة من البيانات على أنه الفرق بين أكبر قيمة وأصغر قيمة، أي أن المدى هو:

$$Range = X_{max} - X_{min} \qquad (1)$$

حيث نعرف X_{max} و X_{max} كما يلي:

(أ) للبيانات المفردة:

X_{max} = أكبر قيمة

X_{min} = أصغر قيمة

(ب) للبيانات المبوبة:

X_{max} = مركز الفترة العليا

X_{min} = مركز الفترة الدنيا

ملاحظة (1):

تعرف بعض الكتب أكبر قيمة وأصغر قيمة للبيانات المبوبة كما يلي:

X_{max} = الحد الأعلى للفترة العليا

X_{min} = الحد الأدنى للفترة الدنيا

مثال (3):

أوجد المدى للمشاهدات التالية والتي هي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص:

25, 30, 40, 45, 35, 55, 50

الحل:

$$X_{max} = 55$$

$$X_{min} = 25$$

$$Range = X_{max} - X_{min} = 55 - 25 = 30$$ (كيلوجرامًا)

مثال (4):

أوجد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f
12.95 – 13.95	13.45	3
13.9 5– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	16.45	16
16. 95– 17.95	17.45	10
17. 95– 18.95	18.45	1

X_{max} = مركز الفترة العليا = 18.45

X_{min} = مركز الفترة الدنيا = 13.45

$$Range = X_{max} - X_{min} = 18.45 - 13.45 = 5.00$$

بعض مميزات وعيوب المدى:

أ- يتميز المدى بسهولة التعريف والحساب

ب- يعيب المدى العيوب التالية:

1. يتأثر المدى بالقيم الشاذة أو المتطرفة.

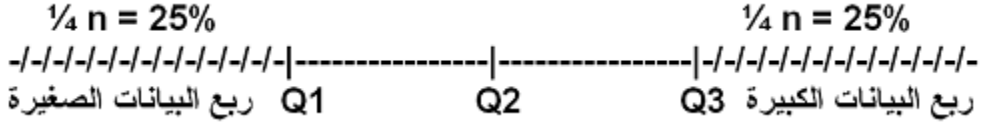
2. لا يأخذ المدى في الاعتبار جميع البيانات.

ملاحظة (2):

1. **وحدة المدى هي نفس وحدة البيانات الأصلية.**
2. **نظرًا لأن المدى يعتمد فقط على أكبر وأصغر قيمة ولا يأخذ في الاعتبار القيم الأخرى فهو مقياس غير جيد لقياس التشتت.**

(3-4) نصف المدى الربيعي Semi-Inter-quartile Range:

رأينا أن المدى يتأثر كثيرًا بالقيم الشاذة أو المتطرفة. ولذلك دعت الحاجة لإيجاد مقاييس أخرى للتشتت لا تتأثر بالقيم المتطرفة. وأحد هذه المقاييس هو نصف المدى الربيعي. وحيث أن القيم المتطرفة هي تلك القيم الصغيرة جدًا أو الكبيرة جدًا فإنه عند حساب نصف المدى الربيعي لا يؤخذ في الاعتبار ربع البيانات الصغيرة (25%) ولا ربع البيانات الكبيرة (25%). الشكل التالي يوضح موضع القيم الشاذة والمتطرفة.



يرمز لنصف المدى الربيعي بالرمز Q ويعرف بالصيغة التالية:

$$Q = \frac{Q_3 - Q_1}{2} \qquad (2)$$

حيث أن Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث وقد مر معنا كيفية إيجادهما للبيانات المبوبة بالطريقة الحسابية والبيانية.

مثال (5):

أوجد نصف المدى الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني باستخدام:

(أ) الطريقة الحسابية.

(ب) الطريقة البيانية.

الحل:

(أ) الطريقة الحسابية:

حساب الربيع الأول Q_1:

$$R = \frac{n}{4} = \frac{50}{4} = 12.5$$

$$A = 14.95,\quad L = 1.00,\quad F_1 = 8\ ,\ F_2 = 23$$

$$Q_1 = A + \left(\frac{R - F_1}{F_2 - F_1}\right) \times L = 14.95 + \left(\frac{12.5 - 8}{23 - 8}\right) \times 1.00 = 15.25$$

حساب الربيع الثالث Q_3:

$$R^* = \frac{3n}{4} = \frac{3 \times 50}{4} = 37.5$$

$$A^* = 15.95,\quad L^* = 1.00,\quad F_1^* = 23\ ,\ F_2^* = 39$$

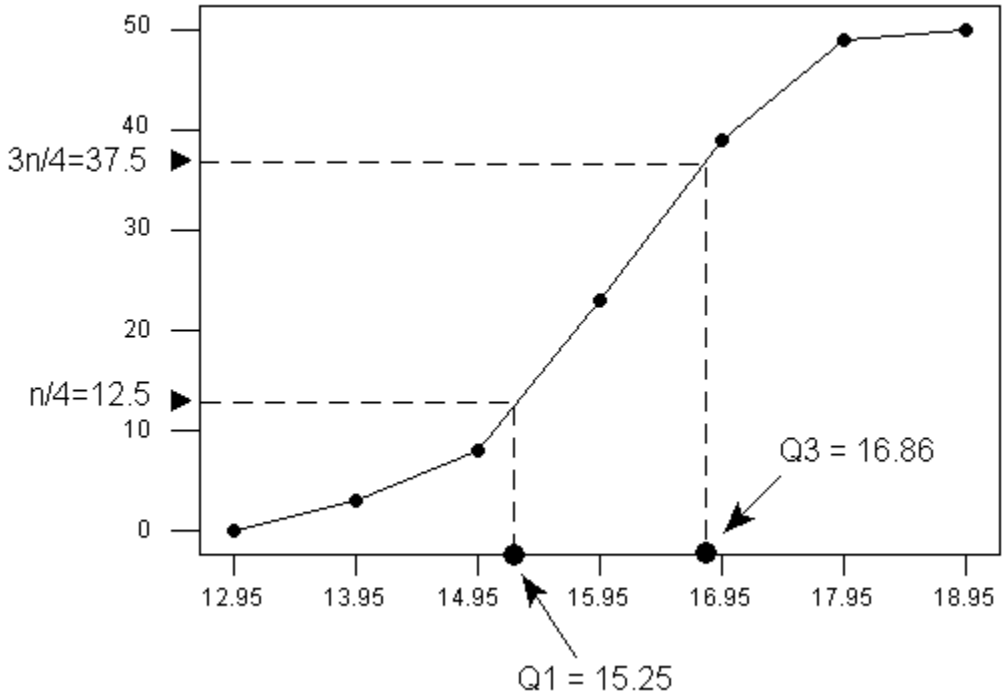
$$Q_3 = A^* + \left(\frac{R^* - F_1^*}{F_2^* - F_1^*}\right) \times L^* = 15.95 + \left(\frac{37.5 - 23}{39 - 23}\right) \times 1.00 = 16.86$$

وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = \frac{1.61}{2} = 0.805$$

	مستوى الهيموجلوبين		التكرار المتجمع الصاعد	
	12.95	أقل من	0	
	13.9 5	أقل من	3	
	14. 95= A	أقل من	8 = F_1	← $R = \frac{n}{4} = 12.5$
$Q_1 \Rightarrow$	15. 95= A^*	أقل من	23 = F_2 = F^*_1	← $R^* = \frac{3n}{4} = 37.5$
	16. 95	أقل من	39 = F^*_2	
$Q_3 \Rightarrow$	17. 95	أقل من	49	
	18.95	أقل من	50	

(ب) الطريقة البيانية:



وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = \frac{1.61}{2} = 0.805$$

بعض مميزات وعيوب نصف المدى الربيعي:

1. **من المميزات أنه لا يتأثر بالقيم الشاذة أو المتطرفة.**
2. **من العيوب أنه لا يأخذ في الاعتبار جميع البيانات.**

ملاحظة (3):

وحدة نصف المدى الربيعي هي نفس وحدة البيانات الأصلية.

(4–4) التباين (Variance) والانحراف المعياري (Standard Deviation):

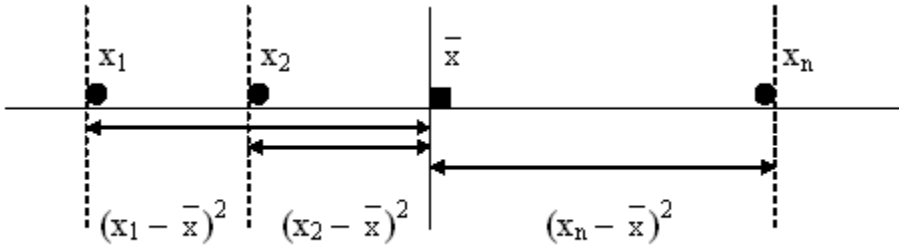
يعتبر التباين والانحراف المعياري من أهم وأفضل مقاييس التشتت ومن أكثرها شيوعًا واستخدامًا في التحليل

الإحصائي وذلك لما يتمتعان به من خصائص وصفات إحصائية جيدة.

التباين: Variance

فكرة التباين تعتمد على تشتت أو تباعد البيانات عن متوسطها. فالتباين يكون كبيرًا إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس.

ويعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي ويرمز له بالرمز S^2. الشكل التالي يبين مربعات انحرافات القيم عن الوسط الحسابي.



والجدول التالي يلخص طريقة حساب انحرافات ومربعات انحرافات القيم عن المتوسط.

القيم (البيانات)	x_n	...	x_2	x_1
انحرافات القيم عن المتوسط	$x_n - \bar{x}$	...	$x_2 - \bar{x}$	$x_1 - \bar{x}$
مربع انحرافات القيم عن المتوسط	$(x_n - \bar{x})^2$	...	$(x_2 - \bar{x})^2$	$(x_1 - \bar{x})^2$

تستخدم مربعات انحرافات القيم عن الوسط الحسابي لحساب التباين. فمن الواضح أن تشتت البيانات يزداد بزيادة مربعات الانحرافات والعكس بالعكس. والسؤال الذي يتبادر للأذهان هو أي واحد من هذه المربعات ينبغي أن يستخدم لقياس التشتت؟ إن من المنطقي أن نبحث عن قيمة نموذجية تمثل هذه المربعات لاستخدامها لقياس التباين. وبناءً على دراستنا السابقة لمقاييس النزعة المركزية فإن المتوسط من أفضل المقاييس التي تمثل مجموعة البيانات. وعليه فإن متوسط مربعات انحرافات القيم عن الوسط الحسابي $\sum_{i=1}^{n}(x_i - \bar{x})^2 / n$ يعتبر أحد المقاييس التي يمكن استخدامها لقياس التباين.

الانحراف المعياري: Standard Deviation

إن التباين من أهم وأفضل مقاييس التشتت ولكنه يقاس بوحدة البيانات الأصلية المربعة. وفي كثير من الأحيان نرغب في استخدام مقياس للتشتت يقاس بوحدة البيانات الأصلية ويتمتع بخصائص إحصائية جيدة مثل التباين. وأحد هذه المقاييس هو الانحراف المعياري. ويعرف الانحراف المعياري على أنه الجذر التربيعي الموجب للتباين ويرمز له بالرمز s.

حساب التباين والانحراف المعياري:

سنستعرض طرق حساب التباين والانحراف المعياري في حالة البيانات المفردة وفي حالة البيانات المبوبة.

أولاً: التباين والانحراف المعياري للبيانات المفردة (غير المبوبة):

إذا كانت $x_1, x_2, \ldots, x_n$ عينه حجمها n وكان متوسطها هو $\bar{x}$ فإن تباين العينة يعرف كما يلي:

$$s^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \qquad (3)$$

لاحظ أننا قسمنا على المقدار $(n-1)$، وهو ما يسمى بدرجات الحرية، بدلاً من القسمة على عدد البيانات n في الصيغة السابقة وذلك لكي نحصل على مقياس يتمتع بصفات إحصائية جيدة.

وأما الانحراف المعياري فإنه يعرف بالصيغة:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \qquad (4)$$

ملاحظة (4):

1. **$s^2 \geq 0$ (دائمًا) وكذلك $s \geq 0$ (دائمًا).**
2. **$s = 0 \Leftrightarrow s^2 = 0 \Leftrightarrow$ جميع قيم العينة متساوية (لا يوجد اختلاف بين القيم).**
3. **وحدة التباين، s^2، هي وحدة البيانات الأصلية المربعة.**
4. **وحدة الانحراف المعياري، s، هي نفس وحدة البيانات الأصلية.**
5. **يمكن حساب التباين بإحدى الصيغتين الحسابيتين التاليتين:**

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right\} \qquad (5)$$

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^{\,2} - n\left(\bar{x}\right)^2\right\} \qquad (6)$$

ولحساب تباين العينة باستخدام الصيغتين الحسابيتين السابقتين فإننا نحتاج إلى معرفة الكميات التالية فقط دون الحاجة لمعرفة البيانات الأصلية:

1. **حجم العينة =** n.
2. **مجموع البيانات =** $\sum_{i=1}^{n} x_i$.
3. **مجموع مربعات البيانات =** $\sum_{i=1}^{n} x_i^{\,2}$.

والصيغتين الحسابيتين السابقتين تستخدمان لحساب تباين العينة وذلك لسببين هما:

1. **لأنهما أكثر سهولة من صيغة التعريف.**
2. **لأنها أكثر دقة في الحساب من صيغة التعريف عندما يكون هناك تقريب في حساب متوسط العينة.**

مثال (6):

أوجد تباين العينة والانحراف المعياري لمجموعة الأوزان (بالكيلوجرام) التالية:

7.1, 2.5, 2.5, 5.4, 8.3

الحل:

نلخص الحل في الجدول التالي:

x_i	الانحراف $(x_i - \bar{x})$ = (x_i –5.16)	مربع الانحراف $(x_i - \bar{x})^2$	x^2
7.1	1.94	3.7636	50.41
2.5	–2.66	7.0756	6.25

2.5	**–2.66**	**7.0756**	**6.25**
5.4	**0.24**	**0.0576**	**29.16**
8.3	**3.14**	**9.8596**	**68.89**
$\sum_{i=1}^{n} x_i = 25.8$	**0.00**	$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.832$	$\sum_{i=1}^{n} x_i^2 = 160.96$

من هذا الجدول نوجد الكميات التالية:

$$\boldsymbol{n = 5}, \qquad \sum_{i=1}^{n} x_i = 25.8\ ,\ \sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.832\ ,\ \sum_{i=1}^{n} x_i^2 = 160.96$$

إن متوسط العينة هو:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25.8}{5} = 5.16 \quad \textbf{(كيلوجرامًا)}$$

حساب تباين العينة:

(أ) باستخدام التعريف:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1} = \frac{27.832}{5-1} = 6.958 \quad \textbf{(كيلوجرامًا مربعًا)}$$

(ب) باستخدام الصيغة الحسابية الأولى:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2 / n\right\} = \frac{1}{5-1}\left\{160.96 - (25.8)^2 / 5\right\}$$

$$= \frac{160.96 - 133.128}{4} = \frac{27.832}{4} = 6.958$$

(ج) باستخدام الصيغة الحسابية الثانية:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^{\,2} - n\ (\bar{x})^2\right\} = \frac{1}{5-1}\left\{160.96 - (5)(5.16)^2\right\}$$

$$= \frac{160.96 - 133.128}{4} = \frac{27.832}{4} = 6.958$$

الانحراف المعياري هو:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = s = \sqrt{6.958} = 2.6378 \text{ (كيلوجرامًا)}$$

بعض خصائص التباين والانحراف المعياري:

1. يخضع التباين والانحراف المعياري لبعض العمليات الجبرية. فإذا كان s^2 و s هما على الترتيب التباين والانحراف المعياري للبيانات $x_1, x_2, \ldots, x_n$ وكان a و b مقدارين ثابتين، فإن:

أ- تباين البيانات $x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$ هو s^2 وأما انحرافها المعياري فهو S. لذلك فإن التباين والانحراف المعياري لا يتأثران بإضافة أو طرح مقدار ثابت من جميع المشاهدات.

ب- تباين البيانات $ax_1, ax_2, \cdots, ax_n$ هو a^2s^2 وأما انحرافها المعياري فهو $|a|s$. حيث $|a|$ هي القيمة المطلقة للقيمة a.

ج- تباين البيانات $ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$ هو a^2s^2 وأما انحرافها المعياري فهو $|a|s$.

د- تباين المقدار الثابت يساوي الصفر.

ويمكن تلخيص هذه الخاصية في الجدول التالي:

المشاهدات	التباين	الانحراف المعياري
$x_1, x_2, \ldots, x_n$	s^2	s
$x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$	s^2	s
$ax_1, ax_2, \cdots, ax_n$	a^2s^2	$\|a\|s$
$ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$	a^2s^2	$\|a\|s$

مثال (7):

	المشاهدات	الانحراف المعياري	التباين
x :	2, 6, 4, 3, 5	$s = 1.581$	$s^2 = 2.5$
$x+5$:	7, 11, 9, 8, 10	1.581	2.5
$3x$:	6, 18, 12, 9,	$\|3\| \times 1.581 = 4.743$	$9 \times 2.5 = 22.5$

	15	
9×2.5=22.5	|3|×1.581=4.743	11, 23, 17, 14, 20 : 3x +5

مثال (8):

إذا كان التباين للمشاهدات $x_1, x_2, \ldots, x_n$ هو 36 فإن التباين للمشاهدات $\frac{x_1-10}{2}, \frac{x_2-10}{2}, \ldots, \frac{x_n-10}{2}$ هو $\left(\frac{1}{2}\right)^2 36 = \frac{36}{4} = 9$ وأما الانحراف المعياري فهو $\sqrt{9}=3$.

2. إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى n_1 ومتوسطها $\bar{x}_1$ وتباينها s_1^2 وكان عدد بيانات المجموعة الثانية n_2 ومتوسطها $\bar{x}_2$ وتباينها s_2^2 وإذا كان $\bar{x}_1 = \bar{x}_2$ (أي أن متوسطي المجموعتين متساويان) فإن تباين المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}$$

مثال (9):

أوجد تباين المجموعة الكلية المكونة من دمج المجموعتين التاليتين:

	المجموعة الأولى	المجموعة الثانية
حجم العينة	$n_1 = 4$	$n_2 = 6$
المتوسط	$\bar{x}_1 = 5$	$\bar{x}_2 = 5$
التباين	$s_1^2 = 3$	$s_2^2 = 3.5$

الحل:

أولاً نلاحظ أن متوسطي المجموعتين متساويان.

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}$$
$$= \frac{(4-1)(3)+(6-1)(3.5)}{4+6-1} = 2.944$$

ثانياً: التباين والانحراف المعياري للبيانات المبوبة:

إذا كان لدينا بيانات عددها n وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن:

أ- عدد الفترات هو k.

ب- مراكز الفترات هي $x_1, x_2, \ldots, x_k$.

ج- تكرارات الفترات هي $f_1, f_2, \ldots, f_k$.

بطريقة مشابهة لحسب المتوسط للتوزيع التكراري فإن التباين للتوزيع التكراري المبوب يمكن حسابه بشكل تقريبي بالصيغة التالية:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n-1} \qquad (7)$$

حيث أن:

$$n = \sum_{i=1}^{k} f_i \ , \quad \bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} \ ,$$

كما يمكن استخدام إحدى الصيغتين الحسابيتين التاليتين:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - \left(\sum_{i=1}^{k} x_i f_i\right)^2 / n\right\} \qquad (8)$$

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - n(\bar{x})^2\right\} \qquad (9)$$

ويكمن تلخيص عمليتي إيجاد المتوسط والتباين باستخدام الجدول التالي:

الفترة	مركز الفترة x	التكرار f	xf	$x^2 f$	$f(x-\bar{x})^2$
الفترة رقم 1	x_1	f_1	$x_1 f_1$	$x_1^2 f_1$	$f_1 (x_1-\bar{x})^2$
الفترة رقم 2	x_2	f_2	$x_2 f_2$	$x_2^2 f_2$	$f_2 (x_2-\bar{x})^2$

الفترة	مركز الفترة x	التكرار f	xf	$x^2 f$	$f(x-\bar{x})^2$
:	:	:	:	:	:
:	:	:	:	:	:
الفترة رقم k	x_k	f_k	$x_k f_k$	$x_k^2 f_k$	$f_k (x_k - ر)^2$
المجموع		$\sum f = n$	$\sum xf$	$\sum x^2 f$	$\sum_{i=1}^{k} f_i (x_i - \bar{x})^2$

مثال (10):

أوجد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

نلخص إيجاد التباين باستخدام الجدول التالي:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	xf	$x^2 f$	$f(x-\bar{x})^2$ $f(x-16.01)^2$
12.95 – 13.95	13.45	3	40.35	542.708	19.6608
13.9 5– 14.95	14.45	5	72.25	1044.013	12.1680
14. 95– 15.95	15.45	15	231.75	3580.538	4.7040
15. 95– 16.95	16.45	16	263.20	4329.640	3.0976
16. 95– 17.95	17.45	10	174.50	3045.025	20.7360
17. 95– 18.95	18.45	1	18.45	340.403	5.9536
المجموع		$\sum f = 50$	$\sum xf = 800.5$	$\sum x^2 f = 12882.33$	$\sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = 66.320$

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{800.5}{50} = 16.01$$

حساب التباين باستخدام صيغة التعريف:

$$s^2 = \frac{\sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2}{n-1} = \frac{66.320}{50-1} = 1.3535$$

حساب التباين باستخدام الصيغة الحسابية الأولى:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - \left(\sum_{i=1}^{k} x_i f_i\right)^2 / n\right\}$$

$$= \frac{1}{50-1}\left(12882.33 - \frac{(800.5)^2}{50}\right)$$

$$= \frac{1}{49}(12882.33 - 12816.005)$$

$$= \frac{66.325}{49} = 1.3536$$

حساب التباين باستخدام الصيغة الحسابية الثانية:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - n\left(\bar{x}\right)^2\right\}$$

$$= \frac{1}{50-1}\{12882.33 - (50)(16.01)^2\}$$

$$= \frac{1}{49}\{12882.33 - 12816.005\}$$

$$= \frac{66.325}{49} = 1.3536$$

حساب الانحراف المعياري:

$$s = \sqrt{s^2} = \sqrt{1.3536} = 1.163$$

(5-4) معامل الاختلاف (التغير): Coefficient of Variation:

ذكرنا سابقًا أن التباين والانحراف المعياري من المقاييس المفيدة لقياس التشتت لتوزيع متغير ما. ولكن في كثير من الأحيان نكون مهتمين بمقارنة التشتت والاختلاف لتوزيعي متغيرين مختلفين. وبما أن التباين والانحراف المعياري مقياسان يعتمدان على وحدة البيانات فإنه يصعب استخدامهما لمقارنة تجانس المجموعات المختلفة من

البيانات وذلك لاختلاف الوحدة المستخدمة. وبشكل عام فإن مقاييس التشتت التي ذكرناها آنفًا تكون غير مناسبة لمقارنة تجانس مجموعات البيانات المختلفة في الحالتين التاليتين:

1. إذا كانت وحدتا المتغيرين مختلفتين حيث لا نستطيع مقارنة الوحدات المختلفة ببعض.
2. إذا كان متوسطا المتغيرين مختلفين وذلك لأن تباين توزيع المتغير ذي المتوسط الصغير ينزع لأن يكون صغيرًا والعكس بالعكس.

لذلك دعت الحاجة إلى مقياس لا يعتمد على وحدة المتغير ويقيس ما يسمى بالتشتت النسبي. وأحد هذه المقاييس هو ما يسمى بمعامل الاختلاف أو معامل التغير. فمعامل الاختلاف هو أحد مقاييس التشتت النسبي وهو مقياس عديم الوحدة ويستخدم لمقارنة التشتت النسبي أو التجانس لمجموعات البيانات المختلفة. فمجموعة البيانات ذات معامل الاختلاف الأكبر يكون تشتتها النسبي أكبر أي أنها تكون أقل تجانسًا والعكس بالعكس. ويعرف معامل الاختلاف للعينة التي متوسطها $\overline{x}$ وانحرافها المعياري **s** بالصيغة التالية:

$$C.V. = \frac{s}{\overline{x}} \qquad (10)$$

مثال (11):

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس الوزن (بالكيلوجرام) والطول (بالسنتيمتر). أي البيانات أكثر تشتتًا نسبًا (أقل تجانسًا)؛ بيانات الأوزان أم بيانات الأطوال؟

رقم الشخص	1	2	3	4	5
الوزن	69	59	65	67	65
الطول	164	162	155	165	158

الحل:

أولاً نوجد المتوسط $\overline{x}$ والانحراف المعياري ***s*** لكل من بيانات الأوزان وبيانات الأطوال كما مر معنا سابقاً. نلخص الحسابات في الجدول التالي:

البيانات	المتوسط $\overline{x}$	الانحراف المعياري ***s***	معامل الاختلاف

			$C.V. = \frac{s}{\bar{x}}$
الأوزان	65.0 kg	3.7417 kg	0.0576
الأطوال	160.8 cm	4.2071 cm	0.026

بما أن معامل الاختلاف لبيانات الأوزان أكبر من معامل الاختلاف لبيانات الأطوال فإن التشتت النسبي لبيانات الأوزان أكبر من التشتت النسبي لبيانات الأطوال. أي أن بيانات الأوزان أقل تجانسًا من بيانات الأطوال.

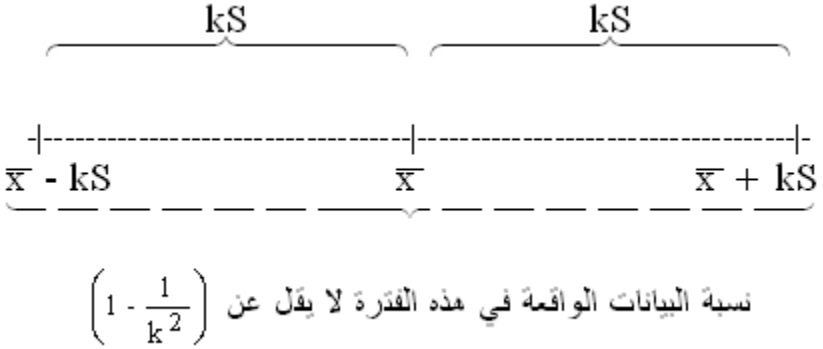
(6–4) نظرية (متراجحة) تشيبيشيف Chebychev Inequality:

إن نظرية تشيبيشيف من النظريات المفيدة إذا أنها تعطينا حدًا أدنى لنسبة البيانات الواقعة في فترة معينة عند معرفة متوسط البيانات وانحرافها المعياري دون الحاجة لمعرفة البيانات الأصلية أو التوزيع الذي أخذت منه العينة.

نظرية (1):

إذا كان لدينا عينة من البيانات متوسطها $\bar{x}$ وانحرافها المعياري s فإن نسبة البيانات الواقعة في الفترة $(\bar{x}-ks, \bar{x}+ks)$ لايقل عن $\left(1-\frac{1}{k^2}\right)$ حيث أن $k>1$.

الشكل التالي يوضح فكرة نظرية تشيبيشيف.



ملاحظة (5):

1. **في بعض الأحيان نكتب الفترة $(\bar{x}-ks, \bar{x}+ks)$ على الصورة $\bar{x} \pm ks$.**

2. تطبق نظرية تشيبيشيف للفترات التي منتصفها (مركزها) هو المتوسط $\bar{x}$.
3. تستخدم نظرية تشيبيشيف بطريقتين(في كلا الحالتين لابد من معرفة قيمة **k**):
   أ- تحديد النسبة (التقريبية) لعدد البيانات الواقعة في فترة معينة.
   ب- تحديد فترة يقع فيها ما لا يقل عن نسبة معينة.

مثال (12):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري s=5 فما هي نسبة البيانات الواقعة في الفترة (18 , 4-)؟

الحل:

أولاً نلاحظ أن منتصف الفترة المعطاة (4,18-) هو المتوسط $\bar{x} = 7$ لذلك نستطيع تطبيق نظرية تشيبيشيف. والآن:

$$(\bar{x} - ks , \bar{x} + ks) = (-4 , 18) \quad \Rightarrow \quad \bar{x} + ks = 18$$

$$\Leftrightarrow \quad 7 + k(5) = 18$$

$$\Leftrightarrow \quad 5k = 11$$

$$\Leftrightarrow \quad k = 11/5$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = \left(1 - \frac{1}{(11/5)^2}\right)$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = 0.7934$$

لذلك فإن نسبة البيانات الواقعة في الفترة (4,18-) لا تقل عن 79.34%.

مثال (13):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري s =5 فأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

الحل:

$$\left(1-\frac{1}{k^2}\right) = 0.75 \Leftrightarrow \frac{1}{k^2} = 1 - 0.75$$

$$\Leftrightarrow \frac{1}{k^2} = 0.25$$

$$\Leftrightarrow k = \sqrt{\frac{1}{0.25}}$$

$$\Leftrightarrow k = 2$$

وبالتالي فإن الفترة التي يقع فيها ما لا يقل عن 75% من البيانات هي:

$$(\bar{x} - ks, \bar{x} + ks) = (7 - 2\times5 , 7 + 2\times5)$$

$$= (7 - 10 , 7 + 10)$$

$$= (-3 , 17)$$

الدرجات (القيم) المعيارية : Standard Scores (Values)

نستطيع بكل يسر وسهولة استخدام قيمتي مشاهدتين في نفس المجموعة لمقارنتهما ببعض. فمثلاً، نستطيع أن نقول بأن أداء الطالب الحائز على الدرجة 85 في اختبار مقرر ما أفضل من أداء الطالب الحائز على الدرجة 80 في نفس الاختبار إذا كان الطالبان في نفس الشعبة. وفي المقابل، لا نستطيع القول بأن أداء الطالب الحائز على الدرجة 85 في اختبار مقرر ما في الشعبة التي يدرسها المدرس (أ) أفضل من أداء طالب آخر حائز على الدرجة 80 في نفس الاختبار ولكنه في شعبة أخرى يدرسها المدرس (ب). من هنا، نرى أنه من الضروري إيجاد قيم لا تعتمد على الوحدات ويمكن استخدامها لمقارنة البيانات في المجموعات المختلفة. هذه القيم التي لا تعتمد على الوحدات نسميها بالقيم (أو الدرجات) المعيارية.

تعريف (2):

لتكن $x_1, x_2, \ldots, x_n$ عينه من البيانات حجمها n ومتوسطها $\bar{x}$ وانحرافها المعياري s. نعرف الدرجة (القيمة) المعيارية للمشاهدة x_i بالصيغة التالية:

$$z_i = \frac{x_i - \bar{x}}{s} ; i = 1, 2, \ldots, n \qquad (11)$$

أي أن الدرجات المعيارية للبيانات $x_1, x_2, \ldots, x_n$ هي:

$$z_1 = \frac{x_1 - \bar{x}}{s} , \; z_2 = \frac{x_2 - \bar{x}}{s} , \; \ldots , \; z_n = \frac{x_n - \bar{x}}{s}$$

ملاحظة (6):

1. القيمة $z_i = \frac{x_i - \bar{x}}{s}$ هي الدرجة المعيارية للمشاهدة الأصلية x_i.
2. المشاهدة الأصلية للدرجة المعيارية z_i هي $x_i = \bar{x} + S z_i$.
3. الدرجات المعيارية هي قيم عديمة الوحدة ولذلك فإنها تستخدم للمقارنة بين المشاهدات المختلفة في المجموعات المختلفة للبيانات.
4. متوسط الدرجات المعيارية =0.
5. الانحراف المعياري للدرجات المعيارية يساوي =1.

مثال (14):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري s =5 فأوجد:

1. الدرجة المعيارية للقيمة $x = 9$.
2. القيمة الأصلية للدرجة المعيارية $z = 0.1$.

الحل:

1. الدرجة المعيارية للقيمة $x = 9$ هي:

$$z = \frac{x - \bar{x}}{s} = \frac{9-7}{5} = \frac{2}{5} = 0.4$$

2. القيمة الأصلية للدرجة المعيارية $z = 0.1$ هي:

$$x = \bar{x} + s\ z = 7 + 5 \times 0.1 = 7 + 0.5 = 7.5$$

مثال (15):

إذا كانت درجة أحد الطلاب في مقرر الإحصاء تساوي 82 ودرجته في مقرر الرياضيات تساوي 89، وإذا كان متوسط درجات الطلاب في مقرر الإحصاء يساوي 75 بانحراف معياري يساوي 10 ومتوسط درجات الطلاب في مقرر الرياضيات يساوي 81 بانحراف معياري يساوي 16، ففي أي المقررين كان أداء الطالب أفضل؟

الحل:

نلخص إيجاد الدرجات المعيارية في الجدول التالي:

المقرر	المتوسط $\bar{x}$	الانحراف المعياري s	الدرجة x	الدرجة المعيارية $z = \frac{x - \bar{x}}{s}$
الإحصاء	75	10	82	$z = \frac{82-75}{10} = 0.7$
الرياضيات	81	16	89	$z = \frac{89-81}{16} = 0.5$

بما أن الدرجة المعيارية لمقرر الإحصاء 0.7 أكبر من الدرجة المعيارية لمقرر الرياضيات 0.5 فإن أداء الطالب في مقرر الإحصاء أفضل من أدائه في مقرر الرياضيات بالرغم من أن درجته في مقرر الإحصاء أقل من درجته في مقرر الرياضيات.

من أهم مقاييس التشتت المطلق: المدى، نصف المدى، الانحراف المتوسط ، التباين والانحراف المعياري.

(1) المدى *Range*:

المدى هو أبسط مقاييس التشتت المطلق ويُعرف بأنه الفرق بين أكبر وأصغر قيمة في مجموعة البيانات ويرمز له بالرمز *R*.

(2) التباين والانحراف المعياري:

يعتبر التباين من أهم مقاييس التشتت المطلق ويعرف تباين مجموعة من القيم بأنه متوسط مجموع مربعات انحرافات هذه القيم عن وسطها الحسابي وبذلك فإن وحدات التباين هي مربع وحدات البيانات الأصلية. فإذا كانت وحدات القراءات الأصلية بالدينار فتكون وحدات التباين (الدينار)^2 وهكذا، ويرمز له بالرمز S^2.

والانحراف المعياري او القياسي لمجموعة من البيانات هو الجذر التربيعي الموجب للتباين، وبذلك فإن وحدات الانحراف المعياري هي نفس وحدات البيانات الأصلية ويرمز له الرمز S، وغالباً يفضل استخدام الانحراف

المعياري لأن مقياس التشتت المطلق يجب أن يكون له نفس وحدات القراءات الأصلية وهو متحقق في حالة الانحراف المعياري اي ان الصفات المختلفة ممكن ان تقارن بهذا المقياس.

ثالثاً: الالتواء Skewness

الالتواء هو بعد التوزيع عن التماثل، وقد يكون هذا التوزيع متماثلاً أو ملتوياً جهة اليمين أو ملتوياً جهة اليسار.

- **ففي حالة التوزيعات المتماثلة فإن الوسط الحسابي = الوسيط = المنوال**
- **إذا كان التوزيع ملتوياً جهة اليمين فإن:**

الوسط الحسابي > الوسيط > المنوال

ويسمى توزيع موجب الالتواء وفيه يكون الطرف الأيمن للمنحنى أطول من الأيسر.

- **إذا كان التوزيع ملتوياً جهة اليسار فإن:**

الوسط الحسابي < الوسيط < المنوال

ويسمى توزيع سالب الالتواء وفيه يكون الطرف الأيسر للمنحنى أطول من الأيمن.