إحصاء 1

الفصل الثالث

مقاييس التشتت

إعداد			
الاسم و اللقب	الرتبة	الكلية	البريد الالكتروني
سليم العمراوي	محاضر أ	العلوم الاقتصادية	**salimlamraoui@gmail.com**

الفئة المستهدفة		
الكلية	القسم	السنة
العلوم الاقتصادية	جذع المشترك	الأولى LMD

المرجع الأساسي المستخدم : السعدي رجال، الاحصاء الوصفي، مؤسسة الرجاء للطباعة و النشر، قسنطينة، الجزائر، 2013.

تمهيد:

عرفنا في الفصل السابق أن مقاييس النزعة المركزية (من متوسط ووسيط ومنوال) تسمح لنا بالحصول على القيم المتوسطة للبيانات أو على تجمعها، غير أن هذه المقاييس لا تكفي لوحدها لمعرفة الصفات الإحصائية اللازمة لوصف الظواهر.

ومن الناحية العلمية ثبت أنه لا يمكننا الاعتماد فقط على مقاييس النزعة المركزية من أجل أخذ فكرة شاملة ودقيقة وأقرب إلى الواقع عن الظاهرة محل الدراسة، لأن الفروق بين قيم الظواهر قد تزداد أو تنقص رغم تساوي المتوسطات لهذه الظواهر، ولتوضيح ما سبق نفترض أن طالبين تحصلا على النتائج التالية في خمس مواد دراسية:

الطالب	إحصاء	رياضيات	اقتصاد	قانون	تاريخ	المجموع	المعدل
x	40	35	30	50	45	200	40
y	10	20	70	35	65	200	40

إذا انصب تركيزنا على مقارنة الطالبين فإننا نلاحظ أن لهما نفس المعدل غير أن التمعن الجيد في الدرجات التي تحصل عليها الطالبين. حيث أن الفروق في علامات الطالب y كبيرة والفروق في علامات الطالب x صغيرة.

وعليه يمكن القول أن علامات الطالب x أكثر تناسقا وتجانسا من علامات الطالب y، أي أقل اختلافا وتشتتا. إن هذه الحقيقية تبين أن مقاييس النزعة المركزية لا تعطي فكرة وافية عن مدى تجانس البيانات، ولا تحقق كل الأغراض التي نرغب الوصول إليها من دراستنا لذلك فإن مقاييس النزعة المركزية لا بد أن تكون مصحوبة بمقاييس أخرى لقياس مدى تباعد أو تقارب البيانات من بعضها البعض أو من متوسطها، تسمى هذه المقاييس بمقاييس التشتت.

ما معنى التشتت؟.

- هو الدرجة التي تتجه بها البيانات الرقمية للإنتشار حول قيمة وسطى.
- هو مدى تباعد وتناثر قيم مفردات العينة أو المجتمع محل الدراسة عن بعضها البعض.

مقاييس التشتت: يقاس تشتت البيانات بعدة مقاييس وهي نوعان:

1. مقاييس التشتت المطلق:

يقصد بمقاييس التشتت المطلق تلك المقاييس التي تقيس مقدار التشتت حول المتوسط مقدرا بوحدات تقيس قيم الظاهرة نفسها.

1.1. المدى (المطلق):

يعتبر المدى الوسيلة المباشرة لمعرفة مدى تقارب القيم أو تباعدها في أي توزيع، وهو أبسط مقاييس التشتت مفهوما وتطبيقا وحسابا، حيث أنه يمكننا من الحصول على فكرة عامة وبسيطة عن تشتت البيانات.

المدى لمجموعة من البيانات هو الفرق بين أكبر قيمة وأصغر قيمة لها ويرمز له بالرمز E.

المدى= أكبر قيمة – أصغر قيمة.

ويعبر عنه بما يلي:

$$E=X_{max}-X_{min}$$

مثال 1:

أوجد المدى للبيانات التالية: 18،30،28،22،12.

الحل:

لإيجاد المدى نقوم دوما بترتيب البيانات تصاعديا (أو تنازليا) على الشكل التالي:

30،28،22،18،12

نقوم بالتعويض في المعادلة لحساب المدى كالتالي:

$$E=Xmax-Xmin= 30-12=18$$

أما في حالة البيانات المبوبة، يمكن حساب المدى بإحدى الطريقتين:

- مركز آخر فئة مطروحا منه مركز أول فئة؛
- الفرق بين الحد الأعلى للفئة العليا وبين الحد الأدنى للفئة الدنيا.

مثال 2:

الجدول التكراري التالي يبين توزيع 60 مزرعة حسب المساحة المزروعة بالقمح بمئات الهكتارات:

المساحة	15-20	20-25	25-30	30-35	35-40	40-45	المجموع
عدد المزارع	3	9	15	18	12	3	60

المطلوب: أحسب المدى للمساحة المزروعة بالقمح في هذه المزارع.

الحل:

الطريقة رقم 1:

المدى= مركز الفئة الأخيرة–مركز الفئة الأولى

- مركز الفئة الأخيرة: (45+40)/2=42,5
- مركز الفئة الأولى: (20+15)/2=17,5

ومنه:

$E=42,5-17,5=25$ أي: المدى (E) 2500 هكتار،

الطريقة رقم 2:

المدى = الحد الأعلى للفئة العليا – الحد الأدنى للفئة الدنيا

- الحد الأعلى للفئة العليا: 45.

- الحد الأدنى للفئة الدنيا: 15.
- المدى (E) 3000 هكتار، أي: $E=45-15=30$

▪ **خواص المدى:**

1 – يتصف المدى بسهولة حسابه.

2 – هو مجال (فترة) يحوي جميع البيانات، يعتمد في حسابه على قيمتين فقط هما القيمة الكبرى والقيمة الصغرى.

3– بسبب الخاصية الثانية فإن المدى شديد التأثر بالقيم المتطرفة.

4– لا يمكن حسابه للتوزيعات المفتوحة.

6– غير قابل للعمليات الجبرية.

2.1. الانحراف المتوسط المطلق:

بما أن التشتت هو اختلاف القيم عن بعضها البعض، يكون كبيرا إذا كانت القيم متباعدة ويكون التشتت صغيرا إذا كانت القيم متقاربة وبالتالي فإن التشتت هو مقياس لقوة تجمع البيانات حول بعضها. بما أن التجمع يكون حول القيمة المتوسطة، فيكون منطقيا لو عملنا على إيجاد مقدار الاختلاف بين القيم ومتوسطها.

وحيث أن مجموع الاختلافات (الانحرافات) عن المتوسط يساوي صفرا، فإنه لو حسبنا القيم المطلقة لمقدار الاختلاف عن المتوسط يكون متوسط هذه الاختلافات مقياسا مناسبا لمقدار التشتت، يسمى هذا المقياس بالانحراف المتوسط المطلق.

ويعرف الانحراف المتوسط المطلق بأنه المتوسط الحسابي للقيم المطلقة لانحرافات القيم عن متوسطها الحسابي، وسوف نرمز للانحراف المتوسط في دراستنا بالرمز e_x وعليه:

أ. **حالة البيانات المفردة:** إذا كانت لدينا القيم التالية: $x_1, x_2, x_3, x_4 .... x_n$ فإن الانحراف المتوسط لها هو:

$$e_{\overline{X}} = \frac{\left|x_1 - \overline{X}\right| + \left|x_2 - \overline{X}\right| + \left|x_3 - \overline{X}\right| + ... + \left|x_n - \overline{X}\right|}{n}$$

أو

$$e_{\overline{X}} = \frac{\sum_{i=1}^{n} \left|x_i - \overline{X}\right|}{n}$$

حيث أن: $\left|x_i - \overline{X}\right|$ تعني القيمة الموجبة للإنحرافات.

فالمهم هنا هو مقدار الاختلاف وليس اتجاهه.

مثال 3:

لنفرض أنه لدينا مجموعة الأعداد التالية: 1، 2، 3، 4، 5، 6، 7، 8، 9، 10.

ولنحسب الانحراف المتوسط المطلق لها.

الحل:

حساب الانحراف المتوسط المطلق يتطلب أولا حساب المتوسط الحسابي:

$$\overline{X} = \frac{\sum x_i}{n} = \frac{1+2+3+4+5+6+7+8+9+10}{10} = 5,5$$

ومنه فالانحراف المتوسط المطلق هو:

$$e_{\bar{x}} = \frac{\sum |x_i - \overline{X}|}{n} = \frac{|1-5,5| + |2-5,5| + |3-5,5| ... + |9-5,5| + |10-5,5|}{10} = \frac{25}{10} = 2,5$$

ب. حالة البيانات المبوبة:

إذا تعلق الأمر ببيانات مرجحة أو مبوبة فإن الصيغة المقابلة للانحراف المتوسط المطلق تصبح كالتالي:

$$e_{\overline{X}} = \frac{\sum_{i=1}^{n} ni |x_i - \overline{X}|}{N}$$

مثال4:

الجدول التالي يبين توزيع الوزن بالكيلوجرام لعينة من الطلاب:

فئات الوزن	40-50	50-60	60-70	70-80	80-90	90-100	المجموع
عدد الطلاب	3	4	3	8	4	2	24

المطلوب: أوجد الانحراف المتوسط المطلق.

الحل:

كخطوة أخرى نقوم بحساب المتوسط الحسابي، مع العلم أننا بصدد بيانات مبوبة:

$$\overline{X} = \frac{\sum n_i x_i}{n} = \frac{1680}{24} = 70$$

لندرج الجدول المساعد الذي يمكننا من حساب الانحراف المتوسط مع العلم أننا بصدد بيانات مبوبة:

$$e_X = \frac{\sum_{i=1}^{n} ni |x_i - \overline{X}|}{N}$$

يكون الجدول المساعد كما يلي:

الوزن	التكرار n_i	x_i	$n_i x_i$	$\lvert x_i - \overline{X} \rvert$	$n_i \lvert x_i - \overline{X} \rvert$
50−40	3	45	135	25	75
60−50	4	55	220	15	60
70−60	3	65	195	5	15
80−70	8	75	600	5	40
90−80	4	85	340	15	60
100−90	2	95	190	25	50
المجموع	24		1680		300

بصب نتائج العمليات الحسابية في قانون الانحراف المتوسط نجد:

$$e_x = \frac{\sum n_i \lvert x_i - \overline{X} \rvert}{\sum n_i} = \frac{300}{24} = 12{,}5$$

ويعتبر الانحراف المتوسط أفضل من المدى لأنه أقل تأثرا بالقيم المتطرفة.

- **خواص الانحراف المتوسط:**

1- يعتمد في حسابه على جميع القيم وليس على القيمة الكبرى والصغرى فقط.

2- لا يمكن حسابه في حالة التوزيعات التكرارية المفتوحة.

3- يتأثر بالقيم المتطرفة، لأن انحرافها عن المتوسط الحسابي يكون كبيرا.

4- غير قابل للعمليات الجبرية (يصعب التعامل معه رياضيا).

3.1. الانحراف المعياري والتباين:

يعتبر الانحراف المعياري من أهم المقاييس الإحصائية للتشتت، وهو الأكثر استخداما في القوانين والنظريات الإحصائية، لأنه يعطي فكرة سليمة ومنطقية عن ظاهرة التشتت.

أ. حالة البيانات المفردة:

الانحراف المعياري لمجموعة من القيم في مجتمع معين هو الجذر التربيعي لمتوسط مجموع مربع انحراف القيم عن متوسطها الحسابي، ويرمز له بالرمز: δ، في حالة المجتمع، أي:

$$\delta = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

حيث أن: μ هي متوسط المجتمع و N حجمه.

كما يرمز له بالرمز S في حالة العينات(**) ذات الحجم n والمتوسط $\bar{X}$، أي:

$$S = \sqrt{\frac{\sum (x_i - \overline{X})^2}{n}}$$

(**) في حالة العينات الصغيرة (n<30) فإنه يمكن استخدام العلاقة $S = \sqrt{\frac{\sum (x_i - \overline{X})^2}{n-1}}$

وقد سمي مربع الانحراف المعياري بالتباين، وفي حالة العينة (وهي الحالة العامة التي سنعتمدها) يكون كما يلي:

$$S^2 = \frac{\sum (x_i - \overline{X})^2}{n}$$

إن التباين مقدار موجب دوما $S^2 \geq 0$.

مثال5:

أوجد الانحراف المعياري والتباين للبيانات التالية: 3،7،6،1،11،5،6،9.

الحل:

لحساب الانحراف المعياري والتباين نتبع ما يلي:

1. نجد المتوسط الحسابي.
2. نقوم بطرح المتوسط الحسابي من قيم المفردات x_i من خلال إدراج جدول تلخيصي (جدول مساعد).

$(x_i - \overline{X})^2$	$(x_i - \overline{X})$	x_i
9	−3	3
1	1	7
0	0	6
25	−5	1
25	5	11
1	−1	5
0	0	6
9	3	9
70		48

المتوسط الحسابي: $\overline{X} = \frac{\sum xi}{n} = \frac{48}{8} = 6$

3. نطبق قانون الانحراف المعياري والتباين لحالة بينات غير مبوبة (مفردة).

الانحراف المعياري: $S = \sqrt{\frac{\sum (x_i - \overline{X})^2}{n}} = \sqrt{\frac{70}{8}} = 2{,}95$

التباين: $S^2 = \frac{\sum (x_i - \overline{X})^2}{n} = \frac{70}{8} = 8.75$

ب. حالة البيانات المبوبة:

بطبيعة الحال إذا كنا بصدد المتوسط الحسابي لبيانات مبوبة أو مرجحة فإن التباين يجب أن يأخذ الصورة التالية:

$$S^2 = \frac{\sum n_i (x_i - \bar{X})^2}{n}$$

ومنه فالانحراف المعياري يكون بالشكل التالي:

$$S = \sqrt{\frac{\sum n_i (x_i - \bar{X})^2}{n}}$$

ومنه فإن الانحراف المعياري هو الجذر التربيعي الموجب للتباين.

مثال 6:

الجدول الموالي يبين توزيع الأجور اليومية لعينة من العمال:

فئات الأجر	65-75	75-85	85-95	95-105	105-115	المجموع
عدد العمال	5	10	20	10	5	50

المطلوب: أوجد قيمة كل من: التباين والانحراف المعياري.

الحل:

خطوات إيجاد التباين من بيانات مبوبة أي من توزيع تكراري كالتالي:

1. نحسب عمود مراكز الفئات x_i.
2. نوجد حاصل ضرب عمودي: x_i, n_i (التكرار في مركز الفئة) ثم أوجد المجموع.
3. نوجد قيمة المتوسط الحسابي.
4. أطرح قيمة المتوسط الحسابي من كل قيمة من قيم مراكز الفئات x_i لتحصل على العمود $(x_i - \bar{X})$.
5. ربع قيم العمود السابق أي: $(x_i - \bar{X})^2$.
6. اضرب عناصر العمود السابق في قيم التكرارات المقابلة له ثم أوجد المجموع $\sum n_i (x_i - \bar{X})^2$.
7. طبق القانون التالي:

$$S^2 = \frac{\sum n_i (x_i - \bar{X})^2}{n}$$

$n_i(x_i - \bar{X})^2$	$(x_i - \bar{X})^2$	$(x_i - \bar{X})$	$n_i x_i$	x_i	التكرار n_i	فئات الأجر
2000	400	-20	350	70	5	75-65
1000	100	-10	800	80	10	85-75
0	0	0	1800	90	20	95-85
1000	100	10	1000	100	10	105-95
2000	400	20	550	110	5	115-105
6000			4500		10	المجموع

بصب هذه النتائج في قانون التباين، نجد:

$$S^2 = \frac{\sum n_i (x_i - \overline{X})^2}{n} = \frac{6000}{50} = 120$$

أما الانحراف المعياري S فهو الجذر التربيعي للتباين، أي:

$$S = \sqrt{120} = 10{,}95$$

- **طريقة مختصرة لحساب التباين:**

إذا انطلقنا من صيغة التباين السابقة وقمنا بفكها وإجراء الاختصارات المناسبة فإننا نحصل على صيغة أكثر بساطة، وهي الصيغة الأكثر استخداما في العمليات الحسابية وتسمى بالصيغة المختصرة (وتسمى صيغة كونينق نسبة إلى مكتشفها).

– حالة البيانات المفردة: (*)

$$\boxed{S^2 = \frac{1}{n}\sum x_i^2 - \overline{X}^2}$$

– حالة البيانات المبوبة: فإن العلاقة تصبح:

$$\boxed{S^2 = \frac{1}{n}\sum n_i x_i^2 - \overline{X}^2}$$

مثال 7:

لنحسب الانحراف المعياري للقيم التالية: 5، 8، 12، 15.

الحل:

المتوسط الحسابي: $\overline{X} = \frac{\sum xi}{n} = \frac{5+8+12+15}{4} = 10$

ولنعتمد الصيغة المختصرة لحساب التباين، أي:

$$S^2 = \frac{1}{n}\sum x_i^2 - \overline{X}^2$$

بالتعويض نجد:

$$S^2 = \frac{1}{4}\sum(25+84+144+225) - (10)^2 = 14{,}5$$

ومنه فالانحراف المعياري:

$$S = \sqrt{14{,}5} = 3{,}81$$

(*) يفضل العلاقة المختصرة لسهولة الحسابات ولأنها تتعامل مع القيم الأصلية بدلا من انحرافاتها عن المتوسط الحسابي.

مثال 6:

أوجد التباين والانحراف المعياري بالصيغة المختصرة للبيانات المبوبة في الجدول الإحصائي التالي:

الفئة	8-4	13-9	18-14	23-19	28-24	المجموع
التكرار	3	4	6	2	4	19

الحل:

سنعتمد على الجدول التلخيصي المساعد التالي:

الفئة	التكرار	x_i	n_ix_i	x_i^2	$n_ix_i^2$
8-4	3	6	18	36	108
13-9	4	11	44	121	484
18-14	6	16	96	256	1536
23-19	2	21	42	441	882
28-24	4	26	104	676	2704
المجموع	19		304		5714

المتوسط الحسابي:

$$\overline{X} = \frac{\sum n_i x_i}{\sum n_i} = \frac{304}{19} = 16$$

التباين بالصيغة المختصرة:

$$S^2 = \frac{1}{n}\sum n_i x_i^2 - \overline{X}^2$$

$$= \frac{5714}{19} - (16)^2 = 300.74 - 256 = 44.74$$

الانحراف المعياري:

$$S = \sqrt{\frac{5714}{19} - 256} = \sqrt{44.73} = 6.69$$

- **خصائص الانحراف المعياري:**

1- قابل للعمليات الجبرية (يسهل التعامل معه رياضيا)؛ حيث:

أ - إنه إذا أضيفت أو طرحت قيمة ثابتة (a) إلى أو من جميع القيم فإن الانحراف المعياري لا يتغير. أي أنه إذا كان $y_i=a \pm x_i$ ، فإن:

$$S_x = S_y$$

ب- إذا ضربت كل قيمة بالمقدار a (أو قسمت كل قيمة على المقدار a) فإن الانحراف المعياري يتأثر بالمقدار نفسه أي أنه إذا كان $y_i=ax_i$ ، فإن:

$$S_y = aS_x$$

2 – بالنسبة للتوزيع الطبيعي فإن:

* 68.27% من البيانات تقع في المجال $\overline{X} \pm S$.

* 95.45% من البيانات تقع في المجال $\overline{X} \pm 2S$.

* 99.73% من البيانات تقع في المجال $\overline{X} \pm 3S$.

3 – كثير الدقة وهو أكثر مقاييس التشتت استخداما.

4– يأخذ كل القيم في الاعتبار.

5–يتأثر بكل القيم وخاصة المتطرفة منها (الشاذة).

6– يأخذ الانحراف المعياري نفس وحدة القياس للمتغير الأصلي (كلغ، متر، لتر ....) لذلك لا يمكن استخدامها كأساس للمقارنة بين تشتت توزيعين لهما وحدات قياس مختلفة.

7 – بما أن الانحراف المعياري يتأثر بالمتوسط الحسابي لبيانات الظاهرة فإنه لا يمكن استخدامه للمقارنة بين تشتت بيانات توزيعين لهما متوسط حسابي مختلف ولو كان هذين التوزيعين من نفس النوعية.

8 – لا يمكن إيجاده بالنسبة للتوزيعات التكرارية المفتوحة من البداية أو النهاية.

9 – يتطلب أحيانا عملا شاقا في العمليات الحسابية.

2. مقاييس التشتت النسبي:

إن ما كنا بصدده سابقا من مقاييس التشتت إنما يعطي لنا نتائج مطلقة تقدر بوحدات العد أو وحدات القياس المستعملة مثل: الدينار، الكلغ،... غير أنه في كثير من الحالات نحتاج إلى المقارنة بين تشتتي مجموعتين من البيانات مختلفة في وحدات قياسها أو مختلفة في متوسطاتها الحسابية أو الاثنين معا. في هذه الحالة فإن استخدام مؤشرات التشتت المطلقة لا تستقيم في عمل المقارنات بين المجموعات المختلفة للبيانات.

وعلى ذلك فإنه يجب تحويل مقاييس التشتت المطلقة إلى مقاييس نسبية وذلك عن طريق نسبة مقياس التشتت إلى أحد مقاييس المتوسطات والذي يلائم هذا المقياس.

أهم المقاييس التشتت النسبية نجد:

– **معامل الاختلاف:**

معامل الاختلاف = $\dfrac{\text{الانحراف المعياري}}{\text{المتوسط الحسابي}} \times 100$

$$C_V = \frac{S}{\overline{X}} \times 100$$

مثال 7:

إذا كان متوسط درجات مجموعة من الطلبة في مادة ما هو 15 بانحراف معياري 3 ومتوسط درجاتهم في مادة أخرى هو 8 بانحراف معياري 2، فأي الدرجات في نظرك أكثر تشتتا؟

الحل:

إذا اعتمدنا على الانحراف المعياري فإننا نحكم على أن درجات المادة الأولى أكثر تشتتا (S =3) من درجات المادة الثانية (S =2)، وهذا غير صحيح لأننا إذا أدخلنا المتوسط الحسابي لدرجات الطلبة في المادتين في الحسبان سنحصل على النتائج التالية:

$$C_{V1} = \frac{3}{15} \times 100 = 20\%$$

$$C_{V2} = \frac{2}{8} \times 100 = 25\%$$

أي أن درجات المادة الثانية أكثر تشتتا.

مثال 8:

لنفترض أننا بصدد قياس أوزان وأطوال طلاب السنة الأولى ثانوي، وكانت النتائج بعد الفرز والتبويب وإجراء العمليات الحسابية المقابلة، كما يبرزها الجدول التالي:

الانحراف المعياري	المتوسط الحسابي	
10 كلغ	40 كلغ	الأوزان
14 سم	140 سم	الأطوال

المطلوب: أي الظاهرتين أكثر تشتتا (الأوزان أم الأطوال)؟

الحل:

من خلال تطلعنا إلى القيم الناتجة لا نستطيع القول، باعتماد الإنحراف المعياري وحده، بأن التشتت في الأطوال أكبر من التشتت في الأوزان، وذلك لاختلاف وحدات القياس (بالإضافة إلى اختلاف المتوسطات) ويصبح من الضرورة استخدام معامل الاختلاف لأغراض المقارنة كما يلي:

- معامل الاختلاف للأوزان: $C_{V1} = \frac{10}{40} \times 100 = 25\%$
- معامل الاختلاف للنوع للأطوال: $C_{V2} = \frac{14}{140} \times 100 = 10\%$

وعلى ذلك نستطيع القول بأن التشتت في الأوزان أكبر من التشتت في الأطوال.