جامعة محمد بن أحمد – وهران

كلية العلوم الاقتصادية، التجارية و علوم التسيير

السنة الأولى ليسانس السداسي الثاني 2019 - 2020

المادة : الإحصاء 2

1- مقاييس التشتت

1- Caractéristiques de dispersion

أستاذ المادة :

أستاذة المحاضرة :

- شنوف ص. منسق المادة
- سليماني ر.
- قنصاب ح. م.
- كيحل م.ر.

أستاذة التطبيق:

– صقال - محجوب ح.	– مرحوم س.	– بوكروس ج.
– جنان - شنوف ح.	– بلعباسي ن.	– خليفة ح.
– بن عودة ا.	– صايم ط.	– بوعلي ر.
– شايمي ي.	– كمال و.	– زاوي ل.

السنة الجامعية : 2019 - 2020

مقاييس التشتت

أولا: **مقاييس التشتت**

I. **امقاييس تشتت مطلقة.**
II. **اامقاييس نشتت نسبية.**

ثانيا: **تمارين محلولة**

مقاييس التشّتت

تمهيد:رأينا أن الهدف من مقاييس النزعة المركزية($\bar{x}$، M_O, M_e) هو قياس مستوى الظاهرة المدروسة، فمثلا نقيس مستوى الأجور في مؤسسة اقتصادية بواسطة الأجر المتوسط، ونقيس مستوى الدخل في بلد ما بواسطة الدخل المتوسط، ونقيس مستوى الطالب في آخر السنة بواسطة المعدل وهو متوسط حسابي...إلخ ، إلا أن هذه المقاييس تشوبها مساوئ عدة، أهمها إخفاء الفروق الموجودة بين القيم، فعند إجراء مقارنة بين ظاهرتين يمكن أن يتساوى متوسطهما الحسابي، ورغم ذلك نجد أن انتشار البيانات في الظاهرتين مختلف كثيرا لأن البيانات غير متجانسة،لهذا وجدت مقاييس أخرى تعطينا فكرة عن مدى تباعد البيانات عن بعضها البعض، تسمى هذه المقاييس بمقاييس التشتت.

مثال : إذا كان لدينا توزيعين A و B يمثلان مجموعة من المصابيح الكهربائية مع فترات الإضاءة لكل مصباح بالساعة.

السلسلة A : 780 -790 -790 -800 -800 -800 -810 -810 -820

السلسلة B : 400 -600 -600 -800 -800 -800 -1000 -1000 -1200

نلاحظ أن التوزيعين لهما نفس مقاييس النزعة المركزية.

$$\bar{x}_A = \bar{x}_B = 800$$

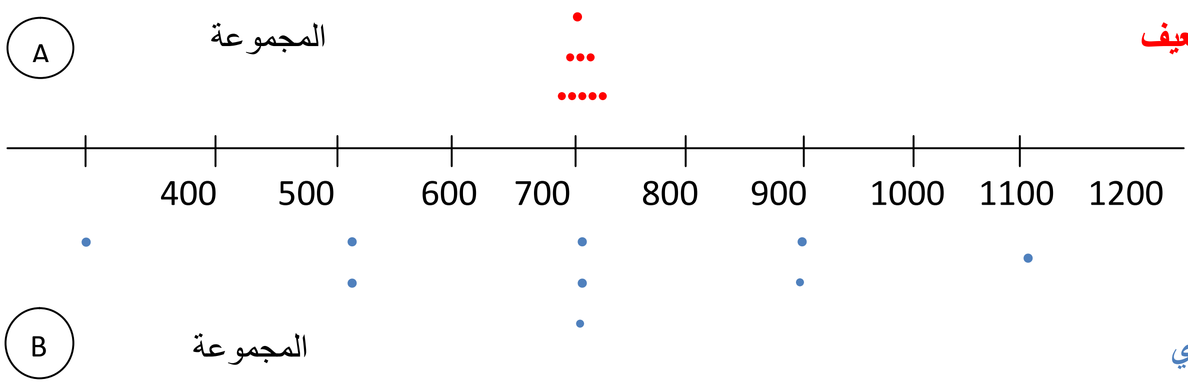
$$Me_A = Me_B = 800$$

$$Mo_A = Mo_B = 800$$

إذن إن الخصائص الثلاثة للنزعة المركزية متساوية في التوزيعين $\bar{x} = M_e = M_O$، غير أن طول مجال الدراسة في التوزيعين (المدى العام E) يختلف بين التوزيعين. معنى ذلك أن التوزيعين يختلفان إختلافا كبيرا من حيث انتشار و توزيع قيمهما على مجال الدراسة.

نضع معطيات السلسلتين على خط :

تشتت ضعيف



تشتت قوي

إن قيم المجموعة B منتشرة و متباعدة حول متوسطها، بينما قيم المجوعة **A** متجمعة حول متوسطها عندئذ يقال ⟸ إن المجموعة **A** **أقل تشتت** من المجموعة **B**، بعبارة أخرى إن قيم المجموعة **A** **أكثر تجانس** من قيم المجموعة **B**.

خلاصة : لمقارنة مجموعتين إحصائيتين أو أكثر، لا نكتفي بمقاييس النزعة المركزية ($\bar{x}$، M_e وM_O) و حتى نستكمل دراسة التوزيع الإحصائي، نتطرق إلى مقاييس التشتت.

سنكتفي بالمقاييس الأكثر استعمالا.

سنتطرق إلى مقاييس التشتت حسب درجة أهميتها، فمن الأقل أهمية إلى الأكثر أهمية.

تقسم مقاييس التشتت إلى قسمين :

I - **مقاييس تشتت مطلقة.**

II - **مقاييس نشتت نسبية.**

I) مقاييس التشتت المطلقة :

تنقسم بدورها إلى مجموعتين :

A– مقاييس التشتت التى تقيس تقارب أو تباعد القيم عن بعضها البعض.

1 – المدى : Etendu

هو الفرق بين أكبر و أصغر قيمة من مفردات السلسلة.

$$\boxed{E = Max_{(xi)} - Min_{(xi)}}$$

في السلسلتين A و B :

السلسلة B أ

$$E_A = 820 - 780 = 40$$

$$E_B = 1200 - 400 = 800$$

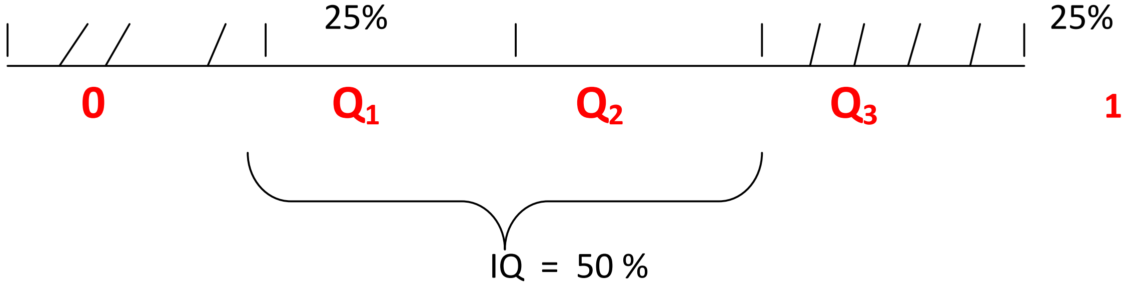
تشتت، نستنتج

أن السلسلة A أ تجانسا.

من مميزات المدى E أنه سهل الحساب، و من عيوبه أنه يعتمد على القيمتين المتطرفتين فقط و التي من الممكن أن تكون قيم شاذة **valeurs aberrantes**.

2 – المدى (أو المجال الربيعي) : Intervalle interquartile

$$\boxed{I_Q = Q_3 - Q_1}$$



يضم المدى الربيعي 50% من المجتمع مهما كان التوزيع الإحصائي. للتخلص من القيم المتطرفة نحسب I_Q .

كلما كان I_Q صغيرا كلما كان التشتت ضعيفا و التمركز قويا.

I_Q استعمالاته محدودة نظرا لبساطته، غير أنه أحسن من المدى العام.

يستعمل I_Q في المقارنة بين توزيعين أو أكثر. من مزايا هذا المقياس أنه لا يعتمد على القيم المتطرفة، و لكن من عيوبه أنه يأخذ بعين الاعتبار فقط 50% من قيم السلسلة.

ملاحظة : في بعض الأحيان نحسب المدى **(المجال) العشيري Intervalle interdécile** و الذي يضم 80% من المعطيات :

$$\boxed{I_D = D_9 - D_1}$$

أو **المدى المئيني Intervalle interpercentile** و الذي يضم 98% من المعطيات:

$$\boxed{I_P = P_{99} - P_1}$$

B - مقاييس التشتت التي تقيس قرب أو بعد القيم من قيمة مركزية غالبا الوسط الحسابي :

مثال : لدينا نقاط (علامات) لستة طلبة في الإحصاء :

2 -17 -7 -18 - 3 -13

متوسط النقاط هو : $\bar{x} = \frac{\sum x_i}{N} = \frac{60}{6} = 10$

هل يمكن استنتاج أن هذه المجموعة متجانسة ؟

بمعنى هل الطلبة لديهم نفس المستوى ؟ الجواب لا، لأن 50% من الطلبة فقط لديهم المعدل.

لقياس التشتت حول $\bar{x}$ نحسب مختلف الانحرافات بين القيم و $\bar{x}$ نحصل على :

x_i	2	17	7	18	3	13	Σ

$(x_i - \bar{x})$	- 8	+7	-3	+8	-7	+3	$\boxed{0}$
$(x_i - \bar{x})^2$	64	49	9	64	49	9	**244**

مجموع انحرافات القيم عن $\bar{x} = 0$:
$$\sum(x_i - \bar{x}) = 0$$

هناك **حل أول** يتمثل في أخذ **القيم المطلقة valeurs absolues** لهذه الانحرافات $|x_i - \bar{x}|$ لكن بما أن القيم المطلقة لا تخضع للحسابات الجبرية، لذلك نفضل **حل ثاني** و هو أن نقوم بتربيع الانحرافات. $(x_i - \bar{x})^2$.

عند حساب متوسط الانحرافات المربعة نحصل على التباين (Variance) .

1 – التباين variance :

تعريف : التباين هو عبارة عن الوسط الحسابي لمربعات الفروق بين قيم المتغير الإحصائي و الوسط الحسابي و يرمز له بالرمز **V(x).**

a– علاقة التعريف : Formule de définition

D.N.G بيانات غير مبّوبة :

$$V(x) = \frac{\sum(x_i - \bar{x})^2}{N}$$

D.G بيانات مبّوبة :

$$V(x) = \frac{\sum\left[n_i(x_i - \bar{x})^2\right]}{\sum n_i}$$

التباين باستخدام التكرار النسبي

$$V(x) = \sum f_i (x_i - \bar{x})^2$$

ملاحظة : إن وحدة التباين هي مربع وحدة البيانات. مثال تباين المتغير الأجور وحدته DA^2 إذا أردنا نفس وحدات البيانات نأخذ الجذر التربيعي للتباين فنحصل على الانحراف المعياري.

2 – الانحراف المعياري : Ecart-type

$$\boxed{\delta_x = \sqrt{\text{التباين}} = \sqrt{v(x)}}$$

<u>ملاحظة</u> : وحدة الانحراف المعياري هي نفس وحدة البيانات. يستعمل الانحراف المعياري لمقارنة توزيعين تكراريين أو أكثر.
كلما كانت قيمة الانحراف المعياري كبيرة، كلما كان تشتت القيم حول وسطها الحسابي كبيرا. (قيم متباعدة) و العكس صحيح.

<u>مثال</u> : أحسب الانحراف المعياري للمثال السابق :

$$\delta_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{N}} = \sqrt{\frac{(-8)^2+7^2+(-3)^2+8^2+(-7)^2+3^2}{6}}$$

$$\Longrightarrow \quad \delta_x = \sqrt{40,66} = \boxed{6,37}$$

<u>التفسير</u> : بعض الطلبة (الممتازين) يكون لديهم العلامة (بالتقريب) المتوسطة 10 زائد 6,37، الآخرين (les mauvais) لديهم العلامة 10 ناقص 6,37.

$$\bar{x} - \delta_x \text{ et } \bar{x} + \delta_x$$

$$(10 - 6,37) \quad \text{et} \quad (10 + 6,37)$$

العلامات محصورة بين 3,63 و 16,37

<u>مثال 2</u> : لدينا النقاط التالية في الإحصاء لمجموعة ثانية من الطلبة : 11 – 9 – 12 – 8
المطلوب : أحسب الانحراف المعياري ثم قارنه مع الانحراف المعياري للمجموعة الأولى .

$$\bar{x}_2 = \frac{\Sigma xi}{N} = \frac{40}{4} = \boxed{10}$$

<u>تباين المجموعة الثانية</u> :

$$V(x)_2 = \frac{\Sigma(xi - \bar{x})^2}{4} = \frac{1}{4}[(8-10)^2 + (12-10)^2 + (9-10)^2 + (11-10)^2]$$

$$\Longrightarrow \quad V(x)_2 = \frac{10}{4} = 2,5.$$

الانحراف معياري للمجموعة 2 :

$$\delta_{x2} = \sqrt{2,5} = \boxed{1,581}$$

المقارنة :

$\delta_{x1} = 6,37$ et $\delta_{x2} = 1,581$

$$\delta_{x1} < \delta_{x2}$$

المجموعة الثانية أكثر تجانس من المجموعة الأولى، لأن تشتت النقاط في المجموعة 1 أكثر من تشتت النقاط في المجموعة 2 بأربع مرات.

القانون المنشور : علاقة كونينغ Formule développée : Relation de Köning :

بيانات غير مبوّبة D.N.G : $V(x) = \delta_x^2 = \frac{\Sigma xi^2}{N} - \bar{x}^2$

بيانات مبوّبةD.G : $V(x) = \delta_x^2 = \frac{\Sigma xi.xi^2}{\Sigma xi} - \bar{x}^2$

باستخدام التكرارات النسبية : $V(x) = \delta_x^2 = \Sigma\, Fi.x^2 - \bar{x}^2$

Remarque : si $\delta_x = \sqrt{V(x)} \Longrightarrow V(x) = \delta_x^2$

مثال 1 : لتكن المعطيات التالية : 15- 11- 10- 9- 5

أحسب الانحراف المعياري بواسطة :

1) علاقة التعريف 2) القانون المنشور

xi	Xi - $\bar{x}$	(xi – $\bar{x}$)^2	Xi²
5	-5	25	25
9	-1	1	81

10	0	0	100
11	+1	1	121
15	+5	25	225
$\boxed{50}$		$\boxed{52}$	$\boxed{552}$

$$\bar{x} = \frac{\Sigma xi}{N}$$

$$\Longrightarrow \quad \bar{x} = \boxed{10}$$

1) <u>علاقة التعريف :</u>

$$:\delta_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}} \qquad \delta_x = \sqrt{\frac{52}{5}}$$

$$\Longrightarrow \quad \delta_x = \sqrt{10,} \quad \boxed{\delta_x = 3,224}$$

2) <u>القانون المنشور</u> :

$$\delta_x = \sqrt{\frac{\Sigma . x_i{}^2}{N} - \bar{x}^2} = \sqrt{\frac{552}{5} - 10^2} \quad \sqrt{110,4 - 100}$$

نفس النتيجة السابقة :

$$\Longrightarrow \quad \delta_x = \sqrt{10,4} \qquad \boxed{\delta_x = 3,224}$$

<u>مثال 2</u> : لدينا الجدول التالي :

الفئات	n_i	x_i	$n_i.x_i$	x_i-$\bar{x}$	$(x_i - \bar{x})^2$	$n_i.(x_i$-$\bar{x})^2$	$n_i.x_i{}^2$
[10-20[	2	15	30	-11	121	242	450
[20-30[	5	25	125	-1	1	5	3125
[30-40[	3	35	105	+9	81	243	3675
Σ	$\boxed{10}$		$\boxed{260}$			$\boxed{490}$	$\boxed{7250}$

<u>نفس السؤال</u> même question

1) علاقة التعريف :

$$\bar{x} = \frac{\Sigma ni.xi}{\Sigma ni} = \frac{260}{10} \boxed{26}$$

$$\delta_x = \sqrt{\frac{\Sigma[ni.(xi-\bar{x})^2]}{\Sigma ni}} = \sqrt{\frac{490}{10}} = \sqrt{49} = \boxed{7}$$

2) القانون المنشور : Koening

$$\delta_x = \sqrt{\frac{\Sigma n_i.x_i{}^2}{\Sigma ni} - \bar{x}^2} = \sqrt{\frac{7250}{10} - 26^2} = \boxed{7}$$

نفس النتيجة السابقة

ملاحظة : يعترض تطبيق علاقة التعريف أحيانا حسابات طويلة و شاقة خاصة عندما تكون $\bar{x}$ قيمة عشرية valeur décimale لذلك من الأفضل استعمال القانون المنشور لحساب التباين و الانحراف المعياري و ذلك لسهولة العمليات الحسابية.

II– **مقاييس التشتت النسبية :**

إن المقاييس التي قمنا بدراستها لحد الآن هي مقاييس تشتت مطلقة و نستعملها عندما نريد مقارنة مجموعتين لهما نفس وحدات القياس، كأن نقارن أجور مجموعة A أجورهم بـ مع مجموعة ثانية B أجورهم معّبر عنها كذلك بالدج.

أما إذا تعلق الأمر بمجموعة نعّبر عن أجورها بالدج مع مجموعة ثانية نعّبر عن أجورها بالأورو (عملة أخرى) و نطلب المقارنة.

في هذه الحالة مقياس التشتت المطلق (الانحراف المعياري δ_x) لا يصلح للمقارنة يجب أن نستعمل مقياس تشتت نسبي حتى نحصل على رقم خال من وحدات القياس.

يوجد عدة معاملات للتشتت النسبي، نذكر :

1) **معامل المدى الربيعي النسبي** **Coefficient interquartile relatif**

$$C_Q = \frac{Q_3 - Q_1}{Q_2} \times 100$$

2) معامل الإختلاف (المعياري) Coefficient de variation

$$\boxed{CV = \frac{\delta x}{\bar{x}} \times 100}$$

ملاحظة : إن معامل الاختلاف CV يستعمل بشكل واسع في التحليل الإحصائي لمقارنة توزيعين أو أكثر.

Le CV est utilisé le plus fréquemment dans la comparaison de deux ou plusieurs séries statistiques.

مثال: نفترض أننا نريد مقارنة أجور عمال قطاع معّين في الجزائر و فرنسا، نفترض أننا حسبنا كل من $\bar{x}$ و δ_x لهذه الأجور في كل بلد و وجدناهم حسب الجدول التالي :

الجزائر : $\delta_x = 350$ DA ، $\bar{x} = 25000$ DA

فرنسا : $\delta_x = 7000$ Euros, $\bar{x} = 25$ Euros

قارن بين الأجور في البلدين .

الحل : فضلا على أنه لا يمكننا أن نقارن DA مع Euros نحسب معامل التشتت النسبي و ⇦ معامل الاختلاف المعياري في كلتا الحالتين و ذلك حتى نحصل على رقم خال من وحدات القياس :

$$CV_{\text{Algérie}} = \frac{\delta x}{\bar{x}} \times 100 = \frac{350}{25000} \times 100 = \underline{1,4\ \%}$$

$$CV_{\text{France}} = \frac{25}{7000} \times 100 = \underline{0,357\ \%}$$

$\Longrightarrow$ $\mathbf{CV_{France} < CV_{Algérie}}$

نلاحظ أن الاختلاف و التشتت في الأجور عند العمال في فرنسا أقل منه تشتت الأجور في الجزائر ⇦ الأجور في فرنسا أكثر تجانس.

Nous pouvons en conclure qu'il ya davantage d'uniformité et d'homogénéité en ce qui concerne les salaires des français qu'en ce qui concerne les salaires des algériens.