مقاييس التشتت

مقدمة:

إن مقاييس النزعة المركزية لا تكفي لوحدها لوصف البيانات وإجراء المقارنات بين التوزيعات التكرارية، لأنها لا تعطينا فكرة عن مدى تجانس أو عدم تجانس البيانات، 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.

مثال1:

إذا كان لدينا مجموعتين من الموظفين، وكل مجموعة تتكون من خمسة أشخاص حيث الأجور الشهرية (دج) لكل فرد في المجموعتين هي كالآتي:

- المجموعة الأولى: 30000 ، 31000 ، 32000 ، 33000 ، 34000.
- المجموعة الثانية: 16000 ، 20000 ، 24000 ، 44000 ، 56000.

لو قمنا بحساب المتوسط الحسابي لكل مجموعة، نجد أنه متساوي في كل منهما ويساوي 32000، ومع ذلك أجور المجموعة الأولى أكثر تجانسا من أجور المجموعة الثانية، من أجل ذلك لجأ الإحصائيون إلى استخدام مقاييس أخرى لقياس مدى تجانس البيانات أو مدى انتشار البيانات حول مقياس النزعة المركزية، ويمكن استخدامها في المقارنة بين مجموعتين أو أكثر من البيانات، ومن هذه المقاييس، مقاييس التشتت و مقاييس الشكل، وسنتطرق في هذا الفصل إلى مقاييس التشتت على أن نتطرق إلى مقاييس الشكل في الفصل الخامس.

أولا- مقاييس التشتت المطلقة:

	تعريفه	القانون في حالة سلسلة إحصائية	القانون في حالة توزيع تكراري
المدى E	المدى: هو الفرق بين أكبر قيمة وأقل قيمة في البيانات، فهو من أبسط مقاييس التشتت وأسهلها في الحساب، ومن عيوبه أنه يعتمد على القيمتين المتطرفتين فقط.	$E = Max(X_i) - Min(X_i)$,	$E = Max(X_i) - Min(X_i)$,
المدى الربيعي IQ	هو الفرق بين الربيع الثالث والربيع الأول، ويعطينا فكرة عن المجال الذي تنتشر فيه نصف عدد البيانات متوسطة القيمة فقط.	$IQ = Q_3 - Q_1$.	$IQ = Q_3 - Q_1$.
الانحراف المتوسط عن المتوسط الحسابي $EM_{\bar{X}}$	هو البعد المتوسط بالقيمة المطلقة لقيم المتغير الاحصائي عن المتوسط الحسابي وهو الأكثر استعمالا، من عيوبه أنه لا يفرق بين القيم التي تكون أكبر وأقل من المتوسط الحسابي.	$EM_{\bar{X}} = \frac{\sum_{i=1}^{n}\lvert X_i-\bar{X}\rvert}{n}$,	$EM_{\bar{X}} = \frac{\sum_{i=1}^{k}n_i\lvert X_i-\bar{X}\rvert}{n}$,

الانحراف المتوسط عن الوسيط EM_{M_e}	هو البعد المتوسط بالقيمة المطلقة لقيم المتغير الاحصائي عن الوسيط، من عيوبه أنه لا يفرق بين القيم التي تكون أكبر وأقل من الوسيط.	$EM_{M_e} = \frac{\sum_{i=1}^{n}\lvert X_i - M_e\rvert}{n}$,	$EM_{M_e} = \frac{\sum_{i=1}^{k} n_i\lvert X_i - M_e\rvert}{n}$,
التباين $V(X)$	يعتبر من أهم مقاييس التشتت و هو كثير الاستعمال في مجال الاحصاء $V(X) = \left(\delta(X)\right)^2$	$V(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$.	$V(X) = \frac{\sum_{i=1}^{k} n_i(X_i - \bar{X})^2}{n}$.
الانحراف المعياري $\delta(X)$	يعتبر من أهم مقاييس التشتت و هو كثير الاستعمال في مجال الاحصاء $\delta(X) = \sqrt{V(X)}$	$\delta(X) = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$,	$\delta(X) = \sqrt{\frac{\sum_{i=1}^{k} n_i(X_i - \bar{X})^2}{n}}$,

ملاحظة: إذا كان المتغير الإحصائي مستمر فإننا نعوض $\mathbf{X_i}$ بمراكز الفئات $\mathbf{C_i}$ في كل المعادلات السابقة.

ثانيا- مقاييس التشتت النسبية:

إذا كنا بصدد إجراء مقارنة بين توزيعات تكرارية ليست لها نفس وحدة القياس أو ليس لهما نفس المتوسط، فمن الضروري هنا استخدام مقاييس التشتت النسبية والتي من أهمها:

المقياس	القانون
المدى النسبي $E\%$	$\boldsymbol{E\% = \frac{E}{\bar{X}} \times 100}$,
المدى الربيعي النسبي $IQ\%$	$\boldsymbol{IQ\% = \frac{IQ}{M_e} \times 100}$,
الانحراف المتوسط عن المتوسط الحسابي النسبي $EM_{\bar{X}}\%$	$\boldsymbol{EM_{\bar{X}}\% = \frac{EM_{\bar{X}}}{\bar{X}} \times 100}$,
الانحراف المتوسط عن الوسيط النسبي $EM_{M_e}\%$	$\boldsymbol{EM_{M_e}\% = \frac{EM_{M_e}}{M_e} \times 100}$.
الانحراف المعياري النسبي (معامل الاختلاف) CV	$\boldsymbol{CV = \frac{\delta(X)}{\bar{X}} \times 100}$,

مثال2:

البيانات التالية تمثل أوزان 60 طالبا بالكيلوغرام في أحد أقسام الـــــــــــ: LMD بكلية العلوم الاقتصادية وعلوم التسيير بجامعة سطيف:

أوزان الطلبة X_i	عدد الطلبة n_i
$]55-50]$	2
$]60-55]$	5
$]65-60]$	12
$]70-65]$	16
$]75-70]$	14
$]80-75]$	8
$]85-80]$	3
المجموع $\sum n_i$	60

المطلوب:

1- حساب المدى المطلق والنسبي؟

2- حساب المدى الربيعي المطلق والنسبي؟

3- الانحراف المتوسط عن المتوسط الحسابي المطلق والنسبي؟

4- الانحراف المتوسط عن الوسيط المطلق والنسبي؟

5- التباين، الانحراف المعياري ؟

6- في دراسة مماثلة عن أوزان الطلبة بجامعة أخرى تحصلنا على النتائج التالية: $\bar{X}=75$ ، $\delta(X)=12$

- قارن بين تشتت الأوزان في الدراستين؟

الحل:

أوزان الطلبة X_i	n_i	C_i	$N_i^{\uparrow}$	$.n_i\times C_i$	$n_i\lvert C_i-\bar{X}\rvert$	$n_i\lvert C_i-M_e\rvert$	$n_i(C_i-\bar{X})^2$
$]55-50]$	2	52,5	2	105	31.84	31.88	506,8928
$]60-55]$	5	57,5	7	287.5	54.6	54.7	596,232
$]65-60]$	12	62,5	19	750	71,04	71.28	420,5568
$]70-65]$	16	67,5	35	1080	14,72	15.04	13,5424
$]75-70]$	14	72,5	49	1015	57,12	56.84	233,0496
$]80-75]$	8	77,5	57	620	72,64	72.48	659,5712
$]85-80]$	3	82,5	60	247,5	42,24	42.18	594,7392
المجموع $\sum n_i$	60	/	/	4105	344,2	344.4	3024,584

1- حساب المدى:

أ- المدى المطلق:

$$\bar{X}=\frac{\sum_{i=1}^{7}n_ic_i}{n}=\frac{4105}{60}=68,42$$

$$E=Max(X_i)-Min(X_i)=85-50=35$$

طول المجال الذي تنتشر فيه أوزان الطلبة هو: 35 كلغ.

ب- المدى النسبي:

$$\boldsymbol{E\%=\frac{E}{\bar{X}}\times 100=\frac{35}{68,42}\times 100=51,15\%}$$

2- حساب المدى الربيعي:

أ- المدى الربيعي المطلق:

$$\boldsymbol{IQ\ =Q_3-Q_1}$$

نقوم بحساب Q_1 كمايلي:

تحديد الفئة الربيعية الأولى: وهي أول فئة تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{n}{4}$ ، أي:

$$N_{Q_1}^{\uparrow}\geq\left(\frac{n}{4}=\frac{60}{4}=15\right)$$

ومنه الفئة الربيعية الأولى هي: $]65-60]$

- حساب الربيع الأول بطريقة المد الداخلي:

$$Q_1=Lim_{Q_1}+\left[\frac{\frac{n}{4}-N_{Q_1-1}^{\uparrow}}{n_{Q_1}}\right]\times A_{Q_1}=60+\left[\frac{15-7}{12}\right]\times 5=63,33$$

نقوم بحساب Q_3 كمايلي:

تحديد الفئة الربيعية الثالثة: وهي أول فئة تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{3n}{4}$ ، أي:

$$N_{Q_3}^{\uparrow} \geq \left(\frac{3n}{4} = \frac{3(60)}{4} = 45\right)$$

ومنه الفئة الربيعية الثالثة هي: $[70 - 75[$

- حساب الربيع الثالث بطريقة المد الداخلي:

$$Q_3 = Lim_{Q_3} + \left[\frac{\frac{3n}{4} - N_{Q_3-1}^{\uparrow}}{n_{Q_3}}\right] \times A_{Q_3} = 70 + \left[\frac{45-35}{14}\right] \times 5 = 73{,}57$$

ومنه: $\boldsymbol{IQ = 73{,}57 - 63{,}33 = 10{,}24}$

الشرح: طول المجال الذي تنتشر فيه الأوزان المتوسطة لنصف الطلبة هو: 10,24 كلغ.

ب- المدى الربيعي النسبي:

نقوم بحساب الوسيط في حالة البيانات المبوبة فنجده يساوي: 68,44 كلغ، أي: $\mathrm{M_e} = 68.44$

$$\boldsymbol{IQ\% = \frac{IQ}{M_e} \times 100 = \frac{10.24}{68{,}44} \times 100 = 14{,}96\%}$$

3- الانحراف المتوسط عن المتوسط الحسابي:

$$\bar{X} = \frac{\sum_{i=1}^{7} n_i c_i}{n} = \frac{4105}{60} = 68{,}42$$

أ- المطلق: $EM_{\bar{X}} = \frac{\sum_{i=1}^{k} n_i |X_i - \bar{X}|}{n} = \frac{344{,}2}{60} = 5{,}74$

يقدر الانحراف المتوسط عن المتوسط الحسابي لأوزان الطلبة بـــــــــ: 5,74 كلغ.

ب- النسبي: $\boldsymbol{EM_{\bar{X}}\% = \frac{EM_{\bar{X}}}{\bar{X}} \times 100 = \frac{5{,}74}{68{,}42} \times 100 = 8{,}39\%}$

4- الانحراف المتوسط عن الوسيط:

$$M_{\mathrm{e}} = 68{,}44$$

أ- المطلق: $EM_{M_e} = \frac{\sum_{i=1}^{k} n_i |X_i - M_e|}{n} = \frac{344{,}4}{60} = 5{,}74$

يقدر الانحراف المتوسط عن الوسيط لأوزان الطلبة بـــــــــ: 5,74 كلغ.

ب- النسبي: $EM_{M_e}\% = \frac{EM_{M_e}}{M_e} \times 100 = \frac{5{,}74}{68{,}42} \times 100 = 8{,}39\%$

5-حساب التباين والانحراف المعياري:

أ- التباين: $V(X) = \frac{\sum_{i=1}^{k} n_i (X_i - \bar{X})^2}{n} = \frac{3024{,}584}{60} = 50{,}41$

ب- الانحراف المعياري: $\delta(X) = \sqrt{\frac{\sum_{i=1}^{k} n_i (C_i - \bar{X})^2}{n}} = \sqrt{V(X)} = \sqrt{50.41} = 7.1$

يقدر متوسط تشتت أوزان الطلبة بــــــ: 7,1 كلغ.

6- المقارنة بين تشتت الأوزان في الدراستين:

أ- الدراسة الأولى:

الانحراف المعياري النسبي (معامل الاختلاف): $CV_1 = \frac{\delta(X)}{\bar{X}} \times 100 = \frac{7,1}{68,42} \times 100 = 10,38\%$

ب- الدراسة الثانية:

الانحراف المعياري النسبي (معامل الاختلاف): $CV_2 = \frac{\delta(X)}{\bar{X}} \times 100 = \frac{12}{75} \times 100 = 16\%$

نلاحظ أن تشتت الأوزان في الدراسة الثانية أكبر منه في الدراسة الأولى.

- مشتقات الوسيط:

1- الربيعيات: وهي الناتجة من تقسيم البيانات إلى أربع أقسام متساوية، وبالتالي كل قسم يمثل 25% من البيانات.

1-1- الربيع الأول: تقسم البيانات إلى 25% من القيم أقل من قيمة الربيع الأول و 75% من القيم أكبر من قيمة الربيع الأول، ونرمز له بالرمز $_1$.

أ- إذا كانت البيانات في شكل سلسلة إحصائية أو توزيع تكراري لمتغير إحصائي متقطع:

$$Q_1 = X_{\left(\frac{n+1}{4}\right)}$$

إذا كانت $\frac{n+1}{4}$ طبيعية (دون فواصل) نأخذ القيمة مباشرة.

إذا كانت $\frac{n+1}{4}$ غير طبيعية (وجود فواصل) نأخذ متوسط القيمتين، مثلا:1، 3، 5، 7، 9، 11، 13، 15.

$\frac{n+1}{4} = \frac{9}{4} = 2,25$.

الربيع الأول هنا: متوسط القيمة التي رتبتها 2 والقيمة التي رتبتها3 أي $\boldsymbol{Q_1} = \frac{3+5}{2} = 4$

ب- في حالة توزيع تكراري لمتغير إحصائي مستمر:

- تحديد الفئة الربيعية الأولى: وهي أول فئة تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{n}{4}$، أي: $N_{Q_1}^{\uparrow} \geq \frac{n}{4}$

- حساب الربيع الأول بطريقة المد الداخلي: $$Q_1 = Lim_{Q_1} + \left[\frac{\frac{n}{4} - N_{Q_1-1}^{\uparrow}}{n_{Q_1}}\right] \times A_{Q_1}$$

2-1- الربيع الثالث: تقسم البيانات إلى 75% من القيم أقل من قيمة الربيع الثالث و 25% من القيم أكبر من قيمة الربيع الثالث، ونرمز له بالرمز $_3$.

أ- إذا كانت البيانات في شكل سلسلة إحصائية أو توزيع تكراري لمتغير إحصائي متقطع:

$$Q_3 = X_{\frac{3(n+1)}{4}}$$

إذا كانت $\frac{3(n+1)}{4}$ طبيعية (دون فواصل) نأخذ القيمة مباشرة.

إذا كانت $\frac{3(n+1)}{4}$ غير طبيعية (وجود فواصل) نأخذ متوسط القيمتين، مثلا:1، 3، 5، 7، 9، 11، 13، 15.

$\frac{3(n+1)}{4} = \frac{27}{4} = 6,75$.

الربيع الثالث هنا: متوسط القيمة التي رتبتها 6 والقيمة التي رتبتها 7 أي $\boldsymbol{Q_3} = \frac{11+13}{2} = 12$

ب- في حالة توزيع تكراري لمتغير إحصائي مستمر:

- تحديد الفئة الربيعية الثالثة: وهي أول فئة تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{3n}{4}$، أي: $N_{Q_3}^{\uparrow} \geq \frac{3n}{4}$

- حساب الربيع الثالث بطريقة المد الداخلي: $$Q_3 = Lim_{Q_3} + \left[\frac{\frac{3n}{4} - N_{Q_3-1}^{\uparrow}}{n_{Q_3}}\right] \times A_{Q_3}$$

2- العشيريات: وهي الناتجة من تقسيم البيانات إلى عشر أقسام متساوية، وبالتالي كل قسم يمثل 10% من البيانات.

2-1- العشير الأول: تقسم البيانات إلى 10% من القيم أقل من قيمة العشير الأول و 90% من القيم أكبر من قيمة العشير الأول، ونرمز له بالرمز $\boldsymbol{D_1}$.

أ- إذا كانت البيانات في شكل سلسلة إحصائية أو توزيع تكراري لمتغير إحصائي متقطع:

$$D_1 = X_{\left(\frac{n+1}{10}\right)}$$

إذا كانت $\frac{n+1}{10}$ طبيعية (دون فواصل) نأخذ القيمة مباشرة.

إذا كانت $\frac{n+1}{10}$ غير طبيعية (وجود فواصل) نأخذ متوسط القيمتين.

مثلا:1، 3، 5، 7، 9، 11، 13، 15، 17، 19، 21.

$\frac{n+1}{10} = \frac{11}{10} = 1{,}1$.

العشير الأول هنا: متوسط القيمة التي رتبتها 1 والقيمة التي رتبتها2 أي $2 = \frac{1+3}{2} = \boldsymbol{D_1}$

ب- في حالة توزيع تكراري لمتغير إحصائي مستمر:

- تحديد الفئة العشيرية الأولى: وهي أول فئة تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{n}{10}$، أي: $N_{Q_1}^{\uparrow} \geq \frac{n}{10}$

- حساب العشير الأول بطريقة المد الداخلي:

$$D_1 = Lim_{D_1} + \left[\frac{\frac{n}{10} - N_{D_1-1}^{\uparrow}}{n_{D_1}}\right] \times A_{D_1}$$

2-2- العشير الثامن: تقسم البيانات إلى 80% من القيم أقل من قيمة العشير الثامن و 20% من القيم أكبر من قيمة العشير الثامن، ونرمز له بالرمز $\boldsymbol{D_8}$.

أ- إذا كانت البيانات في شكل سلسلة إحصائية أو توزيع تكراري لمتغير إحصائي متقطع:

$$D_8 = X_{\frac{8(n+1)}{10}}$$

إذا كانت $\frac{8(n+1)}{10}$ طبيعية (دون فواصل) نأخذ القيمة مباشرة.

إذا كانت $\frac{8(n+1)}{10}$ غير طبيعية (وجود فواصل) نأخذ متوسط القيمتين، مثلا:

مثلا:1، 3، 5، 7، 9، 11، 13، 15، 17، 19، 21.

$\frac{8(n+1)}{10} = \frac{96}{10} = 9{,}6$.

العشير الثامن هنا: متوسط القيمة التي رتبتها 9 والقيمة التي رتبتها 10 أي $18 = \frac{17+19}{2} = \boldsymbol{D_8}$

وهكذا بالنسبة لباقي العشيريات.

3- المئويات: وهي الناتجة من تقسيم البيانات إلى مائة قسم متساوي، وبالتالي كل قسم يمثل 1% من البيانات.

3-1- المئوي الأول: تقسم البيانات إلى 1% من القيم أقل من قيمة المئوي الأول و 99% من القيم أكبر من قيمة المئوي الأول، ونرمز له بالرمز $\boldsymbol{P_1}$.

أ- إذا كانت البيانات في شكل سلسلة إحصائية أو توزيع تكراري لمتغير إحصائي متقطع:

$$P_1 = X_{\left(\frac{n+1}{100}\right)}$$

إذا كانت $\frac{n+1}{100}$ طبيعية (دون فواصل) نأخذ القيمة مباشرة.

إذا كانت $\frac{n+1}{100}$ غير طبيعية (وجود فواصل) نأخذ متوسط القيمتين.

مثلا إذا وجدنا: $\frac{n+1}{100} = 6{,}3$، المئوي الأول هنا: متوسط القيمة التي رتبتها 6 والقيمة التي رتبتها 7.

ب- في حالة توزيع تكراري لمتغير إحصائي مستمر:

- تحديد الفئة المئوية الأولى: وهي أول فئة تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{n}{100}$، أي: $N^{\uparrow}_{P_1} \geq \frac{n}{100}$

- حساب المئوي الأول بطريقة المد الداخلي: $$P_1 = Lim_{P_1} + \left[\frac{\frac{n}{100} - N^{\uparrow}_{P_1-1}}{n_{P_1}}\right] \times A_{P_1}$$

3-2- المئوي السابع والستون: تقسم البيانات إلى 67% من القيم أقل من قيمة المئوي السابع والستون و 33% من القيم أكبر من قيمة المئوي السابع والستون ، ونرمز له بالرمز $\boldsymbol{P_{67}}$.

أ- إذا كانت البيانات في شكل سلسلة إحصائية أو توزيع تكراري لمتغير إحصائي متقطع:

$$D_8 = X_{\frac{67(n+1)}{100}}$$

إذا كانت $\frac{67(n+1)}{100}$ طبيعية (دون فواصل) نأخذ القيمة مباشرة.

إذا كانت $\frac{67(n+1)}{100}$ غير طبيعية (وجود فواصل) نأخذ متوسط القيمتين.

مثلا إذا وجدنا: $\frac{67(n+1)}{100} = 31,6$.

المئوي السابع والستون هنا: متوسط القيمة التي رتبتها 31 والقيمة التي رتبتها 32.

ب- في حالة توزيع تكراري لمتغير إحصائي مستمر:

- تحديد الفئة المئوية السابعة والستون: وهي أول فئة تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{67n}{100}$،

أي: $N^{\uparrow}_{P_1} \geq \frac{67n}{100}$

- حساب المئوي السابع والستون بطريقة المد الداخلي:

$$P_{67} = Lim_{\quad 67} + \left[\frac{\frac{67n}{100} - N^{\uparrow}_{P_{67}-1}}{n_{P_{67}}}\right] \times A_{P_{67}}$$

وهكذا بالنسبة لباقي المئويات.