مقاييس التشتت

Measures of Dispersion

١مقدمة:

تناولنا في الفصل السابق المتوسطات أو مقاييس النزعة المركزية والتي تعبر عن المستوى العام للظاهرة السياسية أو الاجتماعية محل البحث. ولكن ترى هل هذا يعتبر كافياً لوصف البيانات وتحليلها كمياً لنصل بالتالي إلى فهم أكثر وضوحاً للظاهرة محل الدراسة ؟ للإجابة على هذا السؤال نسوق المثال التالي :

مثال ١ :

نفترض أن لدينا شريحتين من شرائح المجتمع تعيشان في منطقتين مختلفتين وكانت دخولهم الأسبوعية (بالدولار الأمريكي) هي كما يلي :

المجموعة الأولى A :

70	75	71	75	74	76	73	78

المجموعة الثانية B :

99	56	80	100	29	70	65	93

وبحساب الوسط الحسابي للشريحتين المذكورتين في كل من المجموعتين – حسب ما أوضحناه في الفصل السابق – فإن الوسط الحسابي لدخول المجموعة الأولى :

$$\overline{X}_A = \frac{70+75+71+75+74+76+73+78}{8}$$

$$\overline{X}_A = \frac{592}{8} = 74 \text{ دولاراً}$$

أي أن الوسط الحسابي لدخل المجموعة A هو 74 دولاراً.

وكذلك بالنسبة للمجموعة B فإن الوسط الحسابي لدخولها هو :

$$\overline{X}_B = \frac{99+56+80+100+29+70+65+93}{8}$$

$$\overline{X}_B = \frac{592}{8} = 74 \text{ دولاراً}$$

والوسط الحسابي لدخل المجموعة B هو أيضاً 74 دولاراً.

ومعنى هذا أن المستوى العام لدخل الأفراد في المجموعتين واحد ويساوي 74 دولاراً. ولكن بإمعان النظر في دخول المجموعة الأولى نجد أنها متجانسة إلى حدٍ كبير أي أنها قريبة جداً من بعضها أو من الوسط الحسابي والذي يساوي 74 دولاراً. وهنا نقول أن تشتت الدخول قليل أو صغير. بينما دخول المجموعة الثانية غير متجانسة فهي بعيدة عن بعضها، أو عن الوسط الحسابي بشكل كبير. وهنا نقول أن تشتت دخول المجموعة الثانية كبير فهي أقل تجانساً (أو أكثر تشتتاً) من المجموعة الأولى.

(الفرق بين أكبر دخل وأصغر دخل في المجموعة الأولى: ٧٨-٧٠ =٨ دولارات فقط، بينما الفرق بين أكبر دخل وأصغر دخل في المجموعة الثانية: ١٠٠- ٢٩-٧١ دولاراً)

نستنتج من هذا المثال بأن المتوسط ليس كافياً لتوصيف البيانات أو تحليلها كمياً. فها هو المتوسط واحد في المجموعتين ورغم ذلك فإن البيانات تختلف تماماً في مدى تشتتها (أو تجانسها). وهنا تبرز الحاجة إلى مقاييس كمية أو إحصائية ليقاس مدى تشتت البيانات ولتعطي للباحث بالتالي صورة أكثر وضوحاً وصدقاً للظاهرة السياسية محل الدراسة. وفيما يلي نتناول بعض مقاييس التشتت التي تخدم الغرض من هذا الكتاب وهي :

المدى، والتباين والانحراف المعياري، ومعامل اختلاف

٢,٥ المدى The Range

يعرّف المدى بأنه " الفرق بين أكبر قيمة وأصغر قيمة ".

فمدى المجموعة A في المثال رقم (١) هو : $78 - 70 = 8$

ومدى المجموعة B في نفس المثال هو : $100 - 29 = 71$

وواضح تماماً أن مدى المجموعة الثانية أكبر بكثير جداً من مدى المجموعة الأولى مما يعني أن المجموعة الثانية أكثر تشتتاً من المجموعة الأولى.

ومن تعريف المدى يتضح أنه مقياس بسيط جداً وأن كل المطلوب معرفته هو أكبر قيمة وأصغر قيمة فقط. ومن هذا التعريف نخرج بالملاحظات التالية :

١ – أن المدى يعتمد في حسابه على قيمتين فقط (أكبر قيمة وأصغر قيمة) وأنه يهمل بالتالي باقي القيم.

٢ – أنه لا يقيس تشتت البيانات عن متوسطها، فهو لا يشير – من قريب أو من بعيد – إلى متوسط البيانات (أو مركزها).

٣ – أنه حساس جداً لأي قيمة شاذة أو متطرفة والمثال التالي يوضح هذه النقطة بشيء من التفصيل.

<u>مثال (٢) :</u>

إذا كانت أعمار أعضاء السلطة التشريعية في بلد ما هي :

72 <u>78</u> <u>20</u> 74 73 75 70

فإن المدى في هذه الحالة هو : $78 - 20 = 58$

وهنا نلاحظ وجود قيمة شاذة بالنسبة لباقي القيم وهي <u>20</u> وهي أصغر قيمة. وإذا أهملت هذه القيمة (أو لم تكن موجودة أصلاً) لكان المدى :

$$78 - 70 = 8$$

وهذا يعني أن وجود قيمة شاذة <u>(20)</u> رفعت قيمة المدى من 8 سنوات إلى 58 سنة. وهذا يوضح مدى حساسية هذا المقياس للقيم الشاذة (أو المتطرفة).

ولكل هذه الأسباب – أو الملاحظات – فإن كثيراً من الإحصائيين والباحثين لا يعتمدون كثيراً على المدى كمقياس للتشتت. ويستخدم فقط إذا كان المطلوب فكرة سريعة أو عامة (وليست دقيقة) عن مدى تشتت البيانات.

٣,٥ التباين The Variance :

يعتبر التباين أحد مقاييس التشتت المهمة لأنه من ناحية يأخذ جميع القيم في الاعتبار عند حسابه، ومن ناحية أخرى لأنه يقيس التشتت عن الوسط الحسابي للقيم، هذا بالإضافة إلى أنه تسهل معالجته رياضياً، وأنه يدخل في تكوين عدد من المقاييس والاختبارات الإحصائية المهمة.

والفكرة الأساسية للتباين هي حساب إنحرافات جميع القيم عن وسطها الحسابي (أي حساب الفرق بين كل قيمة والوسط الحسابي)، وسوف نجد أن بعض القيم أكبر من الوسط فتكون الفروق (أو الإنحرافات) بالموجب، والبعض الأخر أصغر من الوسط فتكون الفروق (أو الإنحرافات) بالسالب. ودائماً يكون مجموع هذه الإنحرافات مساوياً للصفر. ويكون الحل هنا إما إهمال الإشارات السالبة أو تربيع هذه الإنحرافات. وإهمال الإشارات السالبة ليس له مبرر رياضي، فيكون الحل هو تربيع هذه الإنحرافات، ثم نحسب متوسط الانحرافات المربعة فنحصل على التباين. أي أن التباين يعرف كما يلي:

التباين : هو متوسط مربعات إنحرافات القيم عن وسطها الحسابي. والمثال التالي يوضح كيفية حساب التباين.

مثال (٣) :

أحسب تباين دخول الشريحية الاجتماعية التالية : مثال رقم (١) :

70 75 71 75 74 76 76 78

الحل :

١ – نحسب أولاً الوسط الحسابي للدخول كما يلي :

$$\overline{X} = \frac{592}{8} = 74$$

٢ – نحسب إنحرافات القيم عن الوسط الحسابي (أي الفروق بينها وبين الوسط الحسابي) كما يلي :

- 4 , + 1 , - 3 , + 1 , 0 , + 2 , - 1 , + 4

(لاحظ أن مجموع الانحرافات يساوي صفر)

٣ – نربع هذه الانحرافات كما يلي :

16 1 9 1 0 4 1 16

٤ – التباين يساوي متوسط هذه المربعات، أي يساوي :

$$\frac{16+1+9+1+0+4+1+16}{8}=\frac{48}{8}=6$$

٤,٥ الإنحراف المعياري : Standard Deviation :

نلاحظ أن تمييز التباين سيكون " دخول مربعة " وبصفة عامة " وحدات مربعة " لأنه يتم تربيع الانحرافات (أو الفروق). لذلك فإنه يؤخذ الجذر التربيعي الموجب للتباين فنحصل على ما يسمى " الإنحراف المعياري " والذي سيكون تمييزه هو نفس الوحدات. ففي هذا المثال يكون الانحراف المعياري للدخول هو :

$\sqrt{6}=2.449$ دولاراً

والانحراف المعياري – عادة – يستخدم بدلاً من التباين كأهم مقياس للتشتت. فالتباين ما هو إلا مربع الانحراف المعياري، والانحراف المعياري ما هو إلا الجذر التربيعي الموجب للتباين.

تعريف الانحراف المعياري : الانحراف المعياري هو. الجذر التربيعي الموجب للوسط الحسابي لمربعات انحرافات القيم عن وسطها السحابي. أي هو الجذر التربيعي الموجب للتباين. وعادة يرمز للانحراف المعياري للعينة بالرمز S ولتباين العينة بالرمز S^2 وللانحراف المعياري للمجتمع بالرمز اللاتيني σ (سيجما) ولتباين المجتمع بالرمز σ^2.

ويمكن التعبير عن كل من التابين والانحراف المعياري بالرموز كما يلي :

إذا فرضنا أن قيم العينة هي :

$X_1 , X_2 , \dots , X_n$ (أي عددها n) فإن الخطوات تكون كما يلي :

١ – حساب الوسط الحسابي أي حساب $\overline{X}$ حيث :

$$\overline{X}=\frac{\sum X}{n}$$

٢ – حساب إنحرافات القيم عن الوسط الحسابي، أي :

$$X_1-\overline{X}, X_2-\overline{X}, \ldots, X_n-\overline{X}$$

٣ – تربيع هذه الانحرافات، أي :

$$(X_1-\overline{X})^2, (X_2-\overline{X})^2, \ldots, (X_n-\overline{X})^2$$

٤ – التباين هو الوسط الحسابي لهذه المربعات، أي :

$$S^2=\frac{(X_1-\overline{X})^2+,(X_2-\overline{X})^2,+\ldots+,(X_n-\overline{X})^2}{n}$$

والتي يمكن كتابته (بإختصار البسط) كما يلي :

تباين العينة

$$\boxed{S^2=\frac{\sum_{i=1}^{n}(X_I-\overline{X})^2}{n}}$$

٥ – الانحراف المعياري هو الجذر التربيعي الموجب للتباين، أي :

الانحراف المعياري للعينة

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

وبالطريقة نفسها إذا كانت لدينا قيم المجتمع الإحصائي. فإذا كانت قيم المجتمع هي : $X_1 , X_2 , \ldots , X_N$ أي عددها N فإن قانوني التباين والانحراف المعياري للمجتمع – بافتراض أن الوسط الحسابي للمجتمع هو μ (ميو) – يمكن كتابتها كما يلي:

تباين المجتمع

$$\boxed{\sigma^2=\frac{\sum (x-\mu)^2}{N}}$$

الانحراف المعياري للمجتمع

$$\boxed{\sigma=\sqrt{\frac{\sum (x-\mu)^2}{N}}}$$

<u>ملاحظة مهمة :</u>

يمكن حساب كل من التباين والانحراف المعياري بطريقة مختصرة – وتعطي النتائج نفسها بطبيعة الحال – كما يلي :

* تباين العينة بالطريقة المختصرة :

$$S^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

* الانحراف المعياري للعينة بالطريقة المختصرة :

$$S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

وفي الطريقة المختصرة فإن كل المطلوب معرفته لحساب التباين أو الإنحراف المعياري هو $\sum x$ (أي مجموع القيم)، $\sum x^2$ (أي مجموع مربعات القيم) ثم التعويض في القانون.

مثال (٤) :

أحسب التباين والإنحراف المعياري بالطريقة المختصرة لبيانات المثال رقم (٣).

الحل :

يمكن تنظيم الحل في الجدول التالي :

X^2	X
4900	70
5625	75
5041	71
5625	75
5476	74
5776	76
5329	73
6084	78
$\sum x^2 = 43856$	$\sum x = 592$

ويكون التباين (حيث n = 8) هو :

$$S^2 = \frac{\sum x^2}{n} = (\frac{\sum x}{n})^2$$
$$= \frac{43856}{8} = \left(\frac{592}{8}\right)^2$$
$$= 5482 - 5476$$
$$S^2 = 6$$

وبالتالي فإن الإنحراف المعياري هو : $S = \sqrt{6} = 2.449$ وهي النتائج نفسها التي حصلنا عليها سابقاً ولكن بطريقة أبسط.

<u>بعض خصائص الإنحراف المعياري :</u>

١ – قيمة الانحراف المعياري دائماً موجبة أو أكبر من أو تساوي صفر. فأقل قيمة تساوي الصفر (وذلك عندما تكون جميع القيم متساوية، وفي هذه الحالة لا توجد فروق أو إنحرافات بينها وبين الوسط الحسابي وبالتالي لا يوجد أي تشتت بين القيم، وبالتالي فإن قيمة الإنحراف المعياري في حالة تساوي جميع القيم تساوي الصفر).

٢ – كلما كان التشتت كبيراً حول الوسط كلما كان الانحراف المعياري كبيراً، والعكس صحيح.

٣ – إذا أضفنا وطرحنا مقداراً ثابتاً من كل القيم فإن قيمة الانحراف المعياري (أو التباين) لا تتغير (أي لا تتأثر قيمة الانحراف المعياري بالطرح أو الجمع). ولتوضيح هذه الخاصية نأخذ المثال التالي :

<u>مثال (٥) :</u>

حل المثال السابق رقم ٤ بعد طرح 70 (على سبيل المثال من كل القيم لاحظ أن هذا سوف يسهل الحسابات).

الحـل :

المربعات (بعد الطرح)	القيمة بعد طرح (70)
X2	X
0	0
25	5
1	1
25	5
16	4
36	6
9	3
64	8
$\Sigma_X{}^2 = 176$	$\Sigma_X = 176$

ويكون التباين

$$S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2$$

$$= \frac{176}{8} - \left(\frac{32}{8}\right)^2$$

$$= 22-16$$

$$= 6$$

والانحراف المعياري $S = \sqrt{6} = 2.449$ وهي النتائج نفسها مع ملاحظة أن العمليات الحسابية أسهل في هذه الحالة. وتجدر الإشارة إلى أنه لو طرحنا أي قيمة أخرى سنحصل على النتائج نفسها.

٤- إذا ضربنا كل قيمة في مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب القسمة على هذا المقدار الثابت. وإذا قسمنا كل قيمة على مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب الضرب في هذا المقدار الثابت.

٥,٥ معامل الاختلاف : The Coefficien of Variation

لمقارنة تشتت مجموعتين (أو أكثر) من البيانات وكانت البيانات تختلف في مستواها العام (أي في أوساطها الحسابية) و / أو تختلف في وحدات القياس (مثلاً مقارنة بيانات الدخل حيث تقاس بالريال ببيانات العمر حيث تقاس بالسنوات) فإن المقارنة لا تتم مباشرة بمقارنة الانحراف المعياري لكل منهما بل تتم من خلال مقياس آخر هو "معامل الاختلاف" أو ما يسمى أحياناً مقياس التشتت النسبي حيث ينسب الانحراف المعياري لكل مجموعة إلى وسطها الحسابي والضرب في 100 فنحصل على مقياس نسبي أو مئوي (وبدون تمييز) أي تتم المقارنة بحساب معامل الاختلاف لكل منهما، والمجموعة التي لها معامل اختلاف أكبر تكون أكبر تشتتاً والعكس صحيح أي أن :

$$\text{معامل الاختلاف} = \boxed{\frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100}$$

وإذا رمزنا لمعامل الاختلاف بالرمز C. V فإن :

$$\boxed{C.V = \frac{S}{\overline{X}} \times 100}$$

مثال (٦) إذا كان الوسط الحسابي والانحراف المعياري لدخول عينة من الناخبين بالريالات هو : $\overline{X}_1 = 1500,\ S_1 = 152$

وكان الوسط الحسابي والانحراف المعياري لأعمارهم (بالسنوات) هو :

$$\overline{X}_2 = 42,\ S_2 = 9.2$$

فأيهما أكثر تشتتاً الدخل أم العمر ؟

الحل :

لمقارنة التشتت نحسب معامل الاختلاف لكل من الدخل والعمر كما يلي :

١- معامل اختلاف الدخل : $c.v_1 = \frac{s_1}{\bar{x}_1} \times 100 = \frac{152}{1500} \times 100 = 10.13\%$

٢- معامل اختلاف العمر : $c.v_2 = \frac{s_2}{\bar{x}_2} \times 100 = \frac{9.2}{42} \times 100 = 21.9\%$

٣- بما أن معامل اختلاف العمر أكبر من معامل اختلاف الدخل فإن بيانات العمر تكون أكثر تشتتا من بيانات الدخل.

حساب التباين والانحراف المعياري للبيانات المبوبة :

إذا كانت البيانات مبوبة على شكل قيم وتكرارات، فإن التباين والانحراف المعياري بالطريقة المختصرة يمكن حسابهما كما يلي :

التباين في حالة البيانات المبوبة : $\boxed{S^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$

الانحراف المعياري للبيانات المبوبة $\boxed{S = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}}$

حيث X تمثل القيم، f تمثل التكرارات. وإذا كانت البيانات على شكل فئات تحسب مراكز هذه الفئات ويرمز لها بالرمز X، ويتم حساب $\sum Xf$ ، $\sum X^2 f$

مثال (٧) : الجدول التالي يمثل توزيع مجموعة من الناخبين حسب أعمارهم وفي وقت معين، والمطلوب حساب الانحراف المعياري لأعمار الناخبين ؟

الأعمار X	عدد الناخبين f
20	3
25	4
30	5
35	2
المجموع	14

الحل :

ينظم الحل كما في الجدول التالي حيث يتم حساب كل من Xf، X^2f للحصول على ΣXf، $\Sigma X^2 f$

الأعمار X	أعداد الناخبين f	حاصل ضرب Xf	حاصل ضرب X^2f
20	3	60	1200
25	4	100	2500
30	5	150	4500
35	2	70	2450
المجموع	$\Sigma f = 14$	$\Sigma xf = 380$	$\Sigma x^2 f = 10650$

ويحسب الانحراف المعياري باستخدام الصيغة التالية :

$$S = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$$

$$=\sqrt{\frac{10650}{14}-\left(\frac{380}{14}\right)^2}$$

$$=\sqrt{760.7-736.6}$$

$$=\sqrt{24.1}=4.9$$

أي أن الانحراف المعياري لأعمار الناخبين يساوي ٤,٩ سنة.

مثال (٨) شامل على التشتت :

البيانات في الجدول الخاص بالمثال الشامل رقم (٦) على المتوسطات التي تمثل أهم الحروب التي شهدها العالم خلال الفترة من عام 1945، وحتى عام 1980م. احسب المدى، الانحراف المعياري، ومعامل الاختلاف لعدد الحروب خلال الفترة ؟

الحل :

يمكن تلخيص بيانات الجدول رقم (١) في الجدول التالي :

الحروب X	التكرارات f
0	4
1	8
2	7
3	9
4	6
6	2
المجموع	36

والجدول يوضح أن هناك أربع سنوات لم تحدث فيها حروب، وثمان سنوات حدثت في كل منها حرب واحدة، وسبع سنوات حدثت في كل منها حربان، وهكذا....

أولاً: المدى

المدى هو الفرق بين أكبر قيمة وأصغر قيمة. وأكبر قيمة هنا تساوي 6 حروب وأصغر قيمة تساوي صفر (لا توجد حروب)، وبالتالي فالمدى هو :

(6 = 0 − 6) أي أن المدى يساوي 6 حروب.

ثانياً: الانحراف المعياري :

يمكن تلخيص الحسابات كما في الجدول التالي :

الحروب X	التكرارات f	حاصل ضرب X. f	حاصل ضرب X^2f
0	4	0	0
1	8	8	8
2	7	14	28
3	9	27	81
4	6	24	96
6	2	12	72
المجموع	$\Sigma f = 36$	$\Sigma xf = 85$	$\Sigma x^2 f = 285$

$$S = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$$

$$= \sqrt{\frac{285}{36} - \left(\frac{85}{36}\right)^2}$$

$$= \sqrt{7.917 - 5.569}$$

$$= \sqrt{2.348} = 1.53$$

أي أن الانحراف المعياري لعدد الحروب خلال تلك الفترة يساوي 1.53 حرباً

(لاحظ أن التباين يساوي 2.348 فهو القيمة قبل استخراج الجذر)

ثالثاً: معامل الاختلاف :

معامل الاختلاف يساوي خارج قسمة الانحراف المعياري على الوسط الحسابي مضروباً في (100) أي أن :

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

١- الوسط الحسابي (كما حسبناه سابقا) :

$$\overline{X} = \frac{\sum xf}{\sum f} = \frac{85}{36} = 2.36$$

٢- الانحراف المعياري (كما حسبناه أيضا) $S = 1.53$

٣- معامل الاختلاف هو:

$$CV = \frac{1.53}{2.36} \times 100$$
$$= 64.8\ \%$$