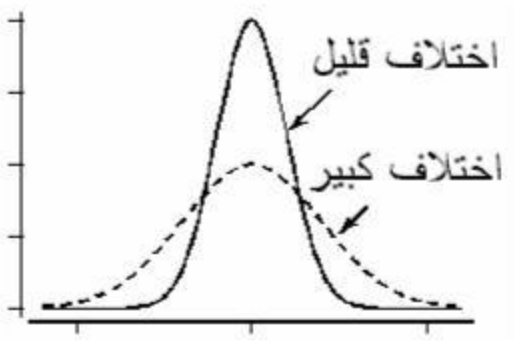
مقاييس التشتت أو الإختلاف Measures of Dispersion

لقد ذكرنا سابقاً بعض مقاييس النزعة المركزية والتي تمثل مقاييس عددية لموضع أو مكان تركز البيانات لظاهرة ما . وقد ذكرنا بأن هذه المقاييس تستخدم لمقارنة مجموعات البيانات المختلفة . وفي الحقيقة فإن 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 . فمثلا المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتتهما.

مثال :

المجموعة	البيانات	المتوسط	
الأولى	59, 61, 62, 58, 60	60	----------●--●--●--●--●----------
الثانية	50, 60, 66, 54, 70	60	----------●-----●-----●-----●------●--- 60

بالرغم من أن المتوسط يساوي 60 للمجموعتين إلا أن التشتت (أو الاختلاف) بين القيم في كل مجموعة غير متساو . فمن الواضح أن بيانات المجموعة الأولى أكثر تقارباً فيما بينها (أقل تشتتاً وتباعداً) 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.



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 صغيراً إذا كانت البيانات متقاربة فيما بينها والعكس صحيح . 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 أهم مقاييس التشتت هي :

١- المدى Range

وهو ابسط مقاييس التشتت ويعرف بأنه الفرق بين أعلى قيمة وأصغر قيمة من البيانات، فإذا كان المدى صغيراً كانت البيانات محصورة في فترة قصيرة ومعنى هذا ان مجموعة البيانات المراد دراستها متقاربة أي متجانسة وعلى العكس إذا كان المدى كبيراً فإنه يدل على إن مفردات المجموعة منتشرة على مساحة واسعة ومبعثرة ومشتتة ومتباعدة عن بعضها :

$$\text{Range} = X_{max} - X_{min}$$

مثال:

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)

وكذلك يعرف المدى في البيانات المبوبة اي في حالة جداول التوزيع التكراري بأنه الفرق بين الحد الاعلى للفئة العليا والحد الادنى للفئة الدنيا ، ومن هذا التعريف يتضح لنا بأن المدى لا يعتمد على جميع البيانات بل على أكبر قيمة وأصغر قيمة فقط. وهذا يقلل من أهميته العلمية إذ قد يحدث أن تكون القيمتان المتطرفتان (أكبر قيمة وأصغر قيمة) قيمتين شاذتين عندئذ يكون المدى كبيرا بينما مفردات البيانات ليست متباعدة عن بعضها البعض كما في المثال الاتي:

مثال : أوجد المدى للبيانات الاتية :

70, 66, 72, 71,70, 30, 75, 66, 79 ,68, 73, 65, 74, 100

الحل :

$Range = X_{max} - X_{min} = 100\text{-}30=70$

معظم القيم واقعة بين ٦٥-٧٩ اي إنها متقاربة ، بينما قيمة المدى (٧٠) وهي قيمة كبيرة بالنسبة لتقارب البيانات (وهذه من أهم عيوب المدى) ، وإذا ما أريد التخلص من هذا العيب فيجب حذف القيم المتطرفة (100, 30) بإعتبارها قيم شاذة فيصبح المدى للبيانات الجديدة :

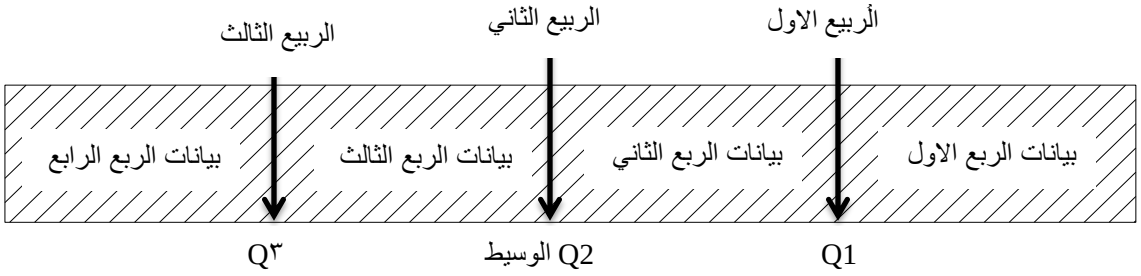
$$Range = X_{max} - X_{min} = 79\text{-}65=14$$

يمتاز المدى كمقياس للتشتت بسهولة حسابه ويعاب عليه ان لا يدخل في إستخراجه إلا قيمتين.

٢ – المدى الرُبيعي ونصف المدى الرُبـــيعي (الانحراف الرُبيعي)

The Quartile Range (QR) and The Half Quartile Range (HQR)

للتخلص من العيب الذي يلازم المدى وهو تأثره بالقراءات المتطرفة وذلك بأن تستبعد الرُبيع الاول من البيانات والرُبيع الاخير منها ونحسب المدى للبيانات المتبقية ، بمعنى أن نجد الرُبيع الادنى(الاول) والرُبيع الاعلى(الثالث) ، حيث إن الرُبيع الادنى يقع ربع البيانات الاول قبله (بعد ترتيب البيانات تصاعدياً) وكذلك الرُبيع الاعلى بحيث يقع ثلاث أرباع البيانات قبله (بعد ترتيبها تصاعدياً) ، فالبيانات بشكل عام تقسم الى أربعة أجزاء بواسطة ثلاثة مقاييس تدعى الرُبيعات وهي الرُبيع الاول والرُبيع الثاني (الوسيط) والرُبيع الثالث كما هو موضح في الشكل التالي:



فمن الممكن أن نرمز للرُبع الاول بـ (Q1) ويقصد به إن رُبع البيانات تقل عن(Q1) ، والرمز (Q2) يقصد به الرُبيع الثاني وهو يعني إن نصف البيانات تكون أقل من (Q2) ، وكذلك (Q3) يقصد به إن ثلاث أرباع البيانات تقل عن قيمة (Q3) ومن الممكن توضيحها كالآتي :

١– الرُبيع الاول (Q1) : وهو القيمة التي يسبقها ربع البيانات ويليها ثلاث أرباع البيانات.

٢- الرُبيع الثاني (Q2) : وهي القيمة التي يسبقها نصف البيانات ويليها النصف الاخر ، وهو نفسه قيمة الوسيط.

٣- الرُبيع الثالث (Q3) : وهو القيمة التي يسبقها ثلاث أرباع البيانات ويليها الربع الاخير.

ثم بعد ذلك نجد نصف المسافة بين الرُبيع الاول والثالث والتي تعتبر كمقياس للتشتت ويسمى هذا المقياس بنصف المدى الرُبيعي أو الانحراف الرُبيعي ويكون كما يلي :

$$HQR = \frac{Q_3 - Q_1}{2}$$

اما المدى الرُبيعي فيكون :

$$QR = Q_3 - Q_1$$

مثال : إذا كانت أكبر قيمة في البيانات ١٠٨ وأقل قيمة ٢٠ وكان الرُبيع الاول Q=45 والرُبيع الثالث Q3=87، فما هو المدى (من المحاضرة السابقة) وما هو المدى الرُبيعي وما هو نصف مدى الرُبيعي ؟

الحل :

المدى :

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

$$Range = 108 - 20 = 88$$

المدى الرُبيعي :

$$QR = Q_3 - Q_1 = 87 - 45 = 42$$

نصف المدى الرُبيعي :

$$HQR = \frac{Q_3 - Q_1}{2}$$

$$HQR = \frac{87 - 45}{2}$$

$$HQR = \frac{42}{2} = 21$$