ناقشنا في الوحدة السابقة مقاييس النزعة المركزية التي تشير إلى نقطة تمركز البيانات، غير أن التوزيعات الإحصائية قد تتشابه فيما بينها من حيث هذه المقاييس؛ فقد نجد توزيعين أو أكثر لهما نفس المتوسط الحسابي مثلاً، أو نفس الوسيط أو المنوال. فهل ذلك يعني أن هناك تماثلاً في توزيع المفردات البيانية (المعطيات) لتلك التوزيعات؟ الجواب لا، لأن مقاييس النزعة المركزية لا تفصح عما يجري بالضبط داخل التوزيع نفسه، بمعنى أن لشكل إنتشار البيانات ، من تقارب أو تباعد عن بعضها البعض وحول مقياس التمركز قد يختلف من توزيع لآخر.

في هذه الوحدة (الفصل) نحاول التطرق لأهم مقاييس التشتت:المدى، الإنحرافالمتوسط، التباين و الإنحراف المعياري ، نصف المدى الربيعي...

و جدير بالذكر مسبقاً أن جميع قيم مقاييس التشتت لا يمكن أن تكون سالبة. كما يمكن التمييز بين مقاييس التشتت المطلقة و مقاييس التشتت النسبية.

I-مقاييس التشتت المطلق:

1-المدى : L'Etendue

أول مقياس لتشتت توزيع معين هو المدى، و هو أبسط مقاييس التشتت ، يعطي إنطباعاً سريعاً عن طبيعة توزيع المشاهدات الإحصائية. نرمز له بالرمز (E) و يعرف على أنه " الفرق بين أكبر قيمة و أصغر قيمة في المشاهدات" في حالة البيانات المفردة (غير المبوبة) ، أو الفرق بين الحد الأعلى للفئة الأخيرة و الحد الأدنى للفئة الأولى في حال فئات. يتم حسابه بالصيغة التالية :

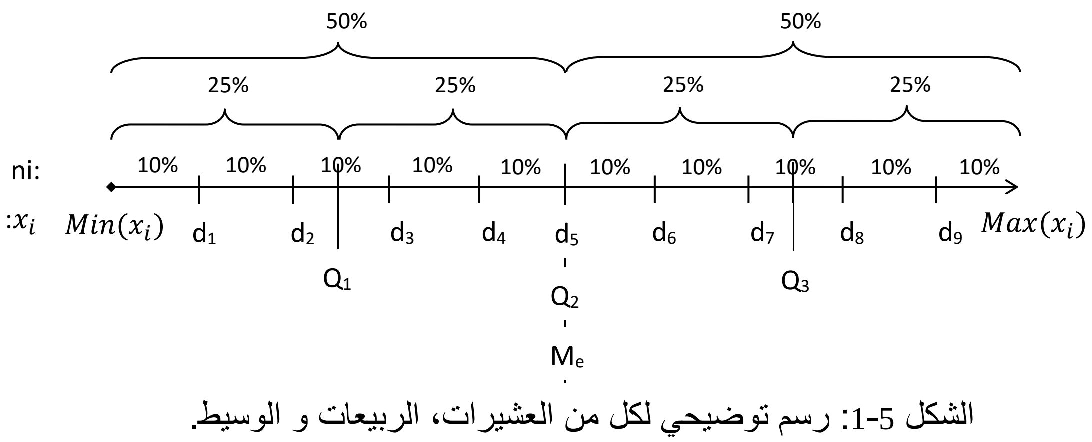
$$E = MAX(x_i) - Min(x_i)$$

2-الربيعات،العشيرات و المؤينات: Quartiles,Percentiles et Déciles

تعتبر كل من الربيعات، العشيرات و المؤينات كمقاييس مماثلة للوسيط لأنها تبنى على نفس المنطق و التفكير و تتشابه في طريقة الحساب حتى أن البعض يُدرجها تحت بند مقاييس النزعة المركزية. فقد رأينا في السابق كيف أن الوسيط يمكّن من تقسيم السلسلة الإحصائية المرتبة إلى قسمين متساويين بحيث يكون مقدار عدد المشاهدات على طرفي الوسيط هو 50% من إجمالي المشاهدات.

ماذا الآن لو أردنا تقسيم مشاهدات السلسلة الإحصائية بعد ترتيبها إلى أربعة أجزاء متساوية يتضمن كل جزء 25% من هذه المشاهدات فنحصل بالتالي على ما يعرف بالربيعات (Q)،أو إلى عشرة أجزاء يتضمن كل واحد $\frac{1}{10}$ عشر المشاهدات (10%) و هو ما يصطلح عليه بالعشيرات (D) ، بل و أحيانا إذا كان عدد المشاهدات كبير جداً، نحتاج لحساب المؤينات (P) و هي التي تقسم السلسلة الإحصائية إلى مائة جزء بالتساوي.

الشكل الموالي يلخص فكرة كل من الربيعات ، العشيرات ، المؤينات و كذلك الوسيط:



الشكل 5-1: رسم توضيحي لكل من العشيرات، الربيعات و الوسيط.

☚في هذا الشكل إكتفينا بتمثيل كل من الوسيط ، الربيعات و العشيرات بينما لم نرغب فيتمثيل المئينات حفاظاً على وضوح الشكل من جهة، نظراً لكثرتها (هناك مائة جزء مئيني) و لأنها تقوم على نفس المبدأ من جهة أخرى.

يمكن تحليل هذا الرسم التوضيحي على النحو التالي :

أولاً: الربيعات:

يرمز لها بالرمز Q.لدينا ثلاثة ربيعات هي :Q_1 , Q_2 , Q_3تقسّم توزيع البيانات إلى أربعة أجزاء متساوية حيث :

-الربيع الأول:نرمز له بالرمز Q_1و هو يُغطي ربع المشاهدات الأولى من السلسلة، و عليه فإن رتبته (موقعه) هو : $\mathrm{Rg}_{Q1} = \frac{\sum ni}{4}$ (25% أو 0.25 كتكرار نسبي)، بحيث يكون لدينا 25% من المشاهدات تقل (أو تساوي) عن هذه القيمة (Q_1) و الباقي (بمعنى 75%) فوق هذه القيمة.

-الربيع الثاني:يعرف بالرمز Q_2 و هو يُغطي رُبعين إثنين من المشاهدات، و بالتالي النصف، من إجمالي المشاهدات كما نلاحظ.رتبته (موقعه) هي:$\mathrm{Rg}_{Q2} = \frac{2}{4}\sum ni = \frac{\sum ni}{2}$ (50% أو0.5)؛ فهو إذن يقسم السلسلة إلى قسمين متساويين مما يجعله ينطبق تماماً مع الوسيط،(Q_2=Me).

-الربيع الثالث:يرمز له ب Q_3 وهو نظير Q_1 إذ يُغطي ثلاثة أرباع من السلسلة ($\frac{3}{4}$) و الربع الأخير يفوق قيمته، و عليه فإن ترتيبه هو:$\mathrm{Rg}_{Q3} = \frac{3}{4}\sum ni$

لإيجاد أي من الربيعات نتبع نفس أسلوب إيجاد الوسيط و نفس الخطوات في الحل.

أ-حالة البيانات المفردة :

مثال5-4:نعتبر السلسلة الإحصائية التالية :

0 , 1 , 2 , 4 , 6 , 5 , 3 , 7 , 6 , 8 , 9 , 12 , 11 , 12 , 14

-أوجد كل من الربيع الأول ، الثاني و الثالث.

الحل: لإيجاد الربيعات نقوم أولاً بترتيب السلسلة تصاعدياً على النحو التالي :

0 , 1 , 2 , 3 , 4 , 5 , 6 , 6 , 7 , 8 , 9 , 11 , 12 , 12 , 14

نلاحظ كذلك أن عدد المشاهدات هو خمسة عشر (فردي n=15)

بإستخدام نفس منهجية إستخراج الوسيط في حال المشاهدات الفردية، بإختلاف في الرتبة فقط، يكون لدينا:

- رتبة الربيع الأول:$\mathrm{Rg}_{Q1} = \frac{n+1}{4} = \frac{15+1}{4} = \frac{16}{4} = 4$ وعليه فإن الربيع الأول Q_1 هو القيمة ذات الرتبة الرابعة أي: $Q_1=3$
- رتبة الربيع الثاني:$\mathrm{Rg}_{Q2} = \frac{2(n+1)}{4} = \frac{15+1}{2} = \frac{16}{2} = 8 = \mathrm{Rg}_{Me}$ وعليه فإن الربيع الثاني (Q_2) الذي يمثل الوسيط كذلك هو القيمة ذات الترتيب الثامن في السلسلة أي : $Q_2=Me=6$

- رتبة الربيع الثالث: $Rg_{Q3} = \frac{3(n+1)}{4} = \frac{3(16)}{4} = 12$ و بالتالي فإن الربيع الثالث Q_3 هو القيمة صاحبة الترتيب الثاني عشر في السلسلة و هي: $Q_3=11$

ملاحظة:

إذا كان عدد المشاهدات زوجياً يمكن إتباع نفس خطوات إيجاد الوسيط في حال البيانات المفردة و n زوجي مع مراعات ترتيب كل ربيع فقط. (بالنسبة لكل ربيع نحصل على قيمتين ثم نأخذ متوسطهما).

ب-حالة البيانات المكررة (جدول تكراري):

إذا كانت البيانات موزعة في جداول توزيعات تكرارية نحتاج حينئذ لحساب التكرار المتجمع الصاعد ثم نقوم بإستخراج كل ربيع بناءاً على رتبته التي نقرأها من عمود التكرار الصاعد حيث يكون الربيع عبارة عن القيمة التي تقابل الرتبة بالضبط إن وجدت أو التي هي أكبر منها مباشرة كما في حالة الوسيط.

جـ-حالة البيانات المبوبة في فئات :

إذا كانت البيانات مبوبة في فئات و للحصول على قيمة الربيعات نقوم بإتباع الخطوات التالية :

1-إيجاد التكرار المتجمع الصاعد $ni\nearrow$ (يمكن كذلك النازل)؛

2-تحديد رتبة الربيع و هي:$\frac{1}{4}\sum ni$بالنسبة للربيع الأول Q_1 أو $\frac{2}{4}\sum ni$ بالنسبة للربيع الثاني Q_2 أو $\frac{3}{4}\sum ni$ بالنسبة للربيع الثالث؛

3-بحسب الربيع المطلوب، و في عمود التكرار المتجمع الصاعد نبحث عن رتبة الربيع أو القيمة الأكبر منها مباشرة ثم نقوم باستخراج الفئة التي تقابل ذلك و هي التي تعرف بفئة الربيع؛

4-أخيراً نقوم بتطبيق الصيغة التالية وفق ما هو مطلوب:

$$Q_K = L_{i1} + \frac{Rg_{Q_k} - n\nearrow_{(i-1)}}{n_i}.\,ai avec k = 1,2,3$$

حيث:

- Q_K: الربيع مع $K=1,2,3$. (Q_3,Q_2,Q_1).
- L_{i1}:الحد الأدنى لفئة الربيع.
- Rg_{Q_k}:رتبة الربيع المبحوث عنه $k=1,2,3$.
- $n\nearrow_{(i-1)}$: التكرار المتجمع الصاعد للفئة ما قبل فئة الربيع.
- n_i : التكرار المطلق لفئة الربيع.
- ai: طول فئة الربيع.

- **ثانياً: العشيرات: Déciles :**
- إن العشيرات تعني تقسيم المعطيات إلى عشرة أجزاء ، كل جزء يشتمل على عدد متساوي من البيانات. هناك تسعة عشيرات يمكن حسابها نرمز لها بالرمز D_Kحيث $k=1,2,3,\ldots,9$. (أنظر الشكل5-1)
- و كما هو الحال عن حساب الوسيط أو الربيعات يجب معرفة أولاً رتبة كل عشير نرغب في حسابه؛ فالعشير الأول هو القيمة التي تقع عند العُشر الأدنى ($\frac{1}{10}$ أو10%) من القيم، أي أنه

القيمة التي يقع قبلها$\{\frac{1}{10}\}$ من المشاهدات و $\{\frac{9}{10}\}$ بعدها و هكذا الحال بالنسبة لباقي العشيرات مع ملاحظة أن العشير الخامس له نفس رتبة كل من الربيع الثاني و الوسيط و بالتالي تتساوى قيم هذه المقاييس الثلاثة مع بعضها البعض. $\left(\mathrm{Rg}_{\mathrm{D}_5} = \frac{5\sum \mathrm{ni}}{10} = \frac{2\sum \mathrm{ni}}{4} = \frac{1\sum \mathrm{ni}}{2}\right)$ et $\mathrm{D}_5 = \mathrm{Q}_2 = \mathrm{Me}$

- يمكن تحديد موقع العشير بتطبيق الصيغة العامة التالية:
- $Rg_{D_k} = \frac{k\sum ni}{10} aveck = \{1, 2, \ldots, 9\}$
- **ثالثاً: المئينات (أو المئويات): Percentiles**
- هي التي تقسم السلسلة الإحصائية إلى مائة جزء متساوي من حيث عدد المشاهدات. نرمز للمئينات بـ P و يمكن حساب 99 مُئين (P_k ; k=1,2,...,99) مع ملاحظة أنه عند كل عشرة مئينات نُصادف عُشيراً، فمثلاً:" $P_{10}=D_1$; $P_{30}=D_3$..." كما أن المُئين الخمسين تتطابق قيمته مع كل من الوسيط، الربيع الثاني إضافةً إلى العشير الخامس، فكل هذه المقاييس متساوية. كذلك يمكننا ملاحظة أن المئين الخامس و العشرون هو نفسه الربيع الأول ($P_{25}=Q_1$) و المئين الخامس و السبعون يساوي الربيع الثالث ($P_{75}=Q_3$).
- و حيث أن تقسيم السلسلة الإحصائية إلى مائة جزء يبدو غير مُجد و غير ذي أهمية مع المعطيات و البيانات قليلة العدد، فسنتعرض تحت هذا العنوان إلى المئينات في حال البيانات المبوبة في فئات فقط؛ و طريقة حسابها تتماثل مع طرق حساب المقاييس السابقة، فبعد معرفة رتبة المئين و فئته نُطبق الصيغة العامة التالية:
- $P_K = L_{i1} + \frac{Rg_{P_K} - n_{(i-1)}^{\nearrow}}{n_i} \times a_i \qquad K = 1,2, \ldots, 99.$
- حيث:
- P_k-: هو المئين المطلوب مع $\mathrm{k} = \{1,2, \ldots, 99\}$ ؛
- L_{i1}-: الحد الأدنى لفئة المئين ؛
- Rg_{P_K} -: رتبة المئين k و يتم تحديدها من خلال الصيغة: $Rg_{P_k} = \frac{\mathrm{K}\sum \mathrm{n_i}}{100}$؛
- $n_{(i-1)}^{\nearrow}$-: التكرار المتجمع الصاعد للفئة ما قبل فئة المئين ؛
- n_i-: التكرار المطلق لفئة المئين ؛
- a_i-: طول فئة المئين.

3-المدى الرُبيعي، العُشيري و المُئيني:

يوجد ثلاثة مديات يمكن حسابها، إضافة إلى المدى العام، لأي سلسلة إحصائية هي:

1-3: المدى الرُبيعي [Q_1;Q_3] :

يُرمز له بـ"IQ" (Interquartile)، و كما يبينه الشكل 5-1 السابق فإن هذا المدى يضم 50% من القيم الوسطى للمشاهدات مهما كان التوزيع الإحصائي و يُعرف كما يلي: $IQ = Q_3 - Q_1$. و نظراً لبساطته فإن إستعمالاته محدودة ، و هو يُستعمل خاصة في المقارنة بين توزيعين إحصائيين أو أكثر.

2-3: المدى العشيري:

أحياناً يتم إستبعاد أول و آخر عُشُر (1/10) من البيانات لنحصل على مجال يُعرف بالمدى العُشيري: [D_1;D_9] و هو عبارة عن الفرق بين أكبر عشير (العشير التاسع) و أدنى عشير (الأول)؛ يُرمز له بـ" ID" Inter déciles و يحسب بالصيغة التالية: $ID = D_9 - D_1$. يتضمن هذا المجال 80% من المشاهدات الوسطى و الباقي 20% موزعة مناصفةً على طرفيه (10% أقل من حده الأدنى D_1 و 10% أعلى من حده الأعلى D_9) و هذا مهما كان المجتمع الإحصائي المدروس.

3-3: المدى المئيني: $[P_1; P_{99}]$ Inter-Centile

هو عبارة عن الفرق بين أعلى و أدنى مئيني. يضم هذا المدى 98% من المشاهدات التي تتوسط السلسلة و 2% المتبقية تتوزع بالتساوي على طرفيه. نحصل عليه من خلال الصيغة: $IC= P_{99}-P_1$

4-نصف المدى الربيعي (Semi Interquartile "SIQ"):

يعرف نصف المدى الربيعي لمجموعة من المشاهدات (القيم) بأنه الفرق بين الربيع الأعلى و الربيع الأدنى مقسوماً على إثنين. نرمز له بـ"SIQ" و هو سهل الحساب حيث يساوي: $SIQ = \frac{Q_3-Q_1}{2} = \frac{IQ}{2}$.

5-الإنحراف المتوسط المطلق: L'écart absolu moyen.

الآن نُقدم مقياس أكثر حساسية لقياس التشتت؛ مقياس يستخدم كل البيانات ألا و هو الإنحراف المتوسط المطلق. و نرمز له بـ "EAM".

يعرف الإنحراف المتوسط على أنه متوسط القيم المطلقة لإنحرافات القيم Xi عن متوسطها الحسابي. و يتم إزالة أثر الإشارة عند حساب الإنحرافات بأخذ القيمة المطلقة لكل إنحراف عن المتوسط $|X_i - \overline{X}|$تفادياً لإشكال خاصية المتوسط الحسابي التي تقول أن مجموع إنحرافات القيم عن متوسطها الحسابي يساوي "الصفر": $\sum(X_i - \overline{X}) = 0$.

أ-حالة البيانات المفردة:

إذا كانت $x_1,x_2,...,x_n$مجموعة من المشاهدات متوسطها الحسابي $\overline{X}$، فإن الإنحراف المتوسط المطلق لهذه القيم يعرف بالعلاقة التالية:

$$EAM = \frac{\sum_{i=1}^{n}|X_i - \overline{X}|}{n}$$

حيث:

- X_i: المشاهدات أو القيم؛
- $\overline{X}$: المتوسط الحسابي لهذه القيم X_i ؛
- $|X_i - \overline{X}|$: إنحرافات القيم X_i عن متوسطها الحسابي $\overline{X}$ بالقيمة المطلقة (i=1→n)؛
- n: عدد المشاهدات.

ب-حالة البيانات المتكررة (جدول توزيع تكراري):

إذا كانت المشاهدات $x_1, x_2, x_3, \dots, x_n$ تتكرر بـعدد مرات قدره $n_1, n_2, n_3, \dots, n_n$ مرة على الترتيب، فإن الإنحراف المتوسط المطلق يُعرف بالصيغة التالية:

$$EAM = \frac{\sum_{i=1}^{n} n_i|X_i - \overline{X}|}{\sum_{i=1}^{n} n_i}$$

وهذا يعني أننا نقوم بضرب كل إنحراف مطلق نحصل عليه في التكرار المقابل له قبل حساب المجموع الكلي في البسط؛ ثم نقوم بقسمة ذلك على مجموع التكرارات.

جـ-حالة البيانات المبوبة في فئات:

إذا كانت البيانات مبوبة في فئات فإننا نستخدم نفس القانون السابق بحيث تُمثل Xi مركز الفئة i

6-التباين و الإنحراف المعياري: La Variance et l'écart type.

التباين و الإنحراف المعياري مفهومان ومقياسان متلازمان مع بعضهما البعض، و هما من أهم المفاهيم الإحصائية المستخدمة في المجال العلمي و العملي و من ركائز الإحصاء الوصفي و الإستدلالي.

الإنحراف المعياري يعتبر أهم مقاييس التشتت و أكثرها إستعمالاً، تم إقتراح فكرته " Standard Deviation" سنة 1893 من قبل العالم الرياضي و الإحصائي "كارل بيرسون Karl Pearson" بينما التباين، الذي ما هو إلا مربع الإنحراف المعياري، تم إستعماله في الإحصاء من طرف عالم الإحصاء الإنجليزي "رونالد فيشر Ronald Fisher".

يُعرف التباين بأنه" متوسط مربعات إنحرافات القيم عن متوسطها الحسابي" و نرمز له بالرمز V(X) أو σ^2. بينما يُعرف الإنحراف المعياري الذي نرمز له بـ" σ" على أنه "الجذر التربيعي الموجب للتباين". و تجدر الإشارة إلى أن التباين موجب دائماً لأنه ناتج عن مجموع مربعات، أي مجموع كميات موجبة، و يكون التباين صفراً إذا و فقط إذا كانت القياسات (البيانات) جميعها متساوية.

أ-البيانات المفردة:

إذا كانت $x_1,x_2,...,x_n$مجموعة من المشاهدات متوسطها الحسابي $\overline{X}$، فإن تباينها و إنحرافها المعياري يعرفان على التوالي بالعلاقة التالية:

- $$V(x) = \sigma^2_{(x)} = \frac{1}{n}[(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + \cdots + (x_n-\overline{x})^2]$$
$$= \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

- $$\sigma_{(x)} = \sqrt{V(x)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

ب-البيانات المتكررة:

إذا كانت المشاهدات $x_1, x_2, x_3, \dots, x_n$ تتكرر بــ $n_1, n_2, n_3, \dots, n_n$ مرة على الترتيب،متوسطها الحسابي $\overline{X}$، فإن التباين هو:

- $$V(x) = \sigma^2_{(x)} = \frac{1}{\sum n_i}[n_1(x_1-\overline{x})^2 + n_2(x_2-\overline{x})^2 + \cdots + n_n(x_n-\overline{x})^2]$$
$$= \frac{1}{\sum_{i=1}^{n} n_i}\sum_{i=1}^{n} n_i(x_i-\overline{x})^2$$

و الإنحراف المعياري هو:

- $$\sigma_{(x)} = \sqrt{V(x)} = \sqrt{\frac{1}{\sum_{i=1}^{n} n_i}\sum_{i=1}^{n} n_i(x_i-\overline{x})^2}$$

جـ-البيانات المبوبة في فئات:

إذا كانت البيانات مبوبة في فئات فإننا نستخدم نفس العلاقة السابقة (بيانات متكررة) حيث تدل Xi هنا على مركز الفئة i.

$$V(x) = \sigma_{(x)}^2 = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} n_i (x_i - \overline{x})^2 \;\; et \; \sigma_{(x)} = \sqrt{V(x)}$$

ملاحظة هامة:

عند حسابنا للتباين (ومنه الإنحراف المعياري) في الصيغ السابقة قمنا بقسمة مجموع مربعات إنحرافات القيم على n (حجم العينة)؛ في حين في الإحصاء الإستدلالي نقوم بالقسمة على عدد المشاهدات مطروحاً منها واحد، بمعنى n−1 بدلاً من n في المقام، وذلك لأننا نقوم في الإحصاء الإستدلالي بتقدير تباين المجتمع $\sigma_{(x)}^2$ إذا كان غير معلوماً إعتماداً على التباين المحسوب من بيانات العينة (الذي يُرمز له بـ"S^2") أين يجب توفر بعض الشروط منها عدم التحيز.[1]

و عليه فإننا في هذا الملخص سنفترض و كأننا نتعامل مع مجتمع عند دراسة التباين و بالتالي فإن المقام سيكون دائماً يساوي n (أو $\sum_{i=1}^{n} n_i$ بوجود تكرارات) نظراً لأن هذا الإجراء سوف يسهل من العملية الحسابية و يساعد الطالب المبتدئ على متابعة المادة العلمية و إستيعابها بيُسر و ليس له تأثير كبير جداً هنا على النتائج و لهذا سوف نُهمل العدد واحد من المقام.

☚ الصيغة المختصرة لحساب التباين و الإنحراف المعياري:

يُمكن تبسيط معادلة التباين (ومنه الإنحراف المعياري) بغرض تسهيل عملية حسابه؛ و ذلك من خلال حساب مفكوك عبارة الجداء الشهير المتواجد في البسط بالصورة التالية:

$$V(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum(x_i^2 - 2x_i\overline{x} + \overline{x}^2)$$

$$= \frac{1}{n}\left(\sum x_i^2 - \sum 2x_i\overline{x} + \sum \overline{x}^2\right)$$

$$= \frac{1}{n}\left(\sum x_i^2 - 2\overline{x}\sum x_i + n\overline{x}^2\right)$$

$$= \frac{\sum x_i^2}{n} - \frac{2\overline{x}\sum x_i}{n} + \frac{n\overline{x}^2}{n}$$

$$= \frac{\sum x_i^2}{n} - 2\overline{x}\overline{x} + \overline{x}^2$$

$$= \frac{\sum x_i^2}{n} - 2\overline{x}^2 + \overline{x}^2$$

---

[1] يُبرهن في الإحصاء الإستدلالي و نظرية التقدير أن تباين العينة S^2 سيكون مقدراً أفضل لتباين المجتمع إذا قمنا بقسمة مجموع مربعات الإنحرافات عن المتوسط $\sum_{i=1}^{n}(x_i - \overline{x})^2$ على (n-1) بدلاً من n. أي أن تباين العينة هو: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ و إنحرافها المعياري هو: $S = \sqrt{S^2}$.

$$= \frac{\sum x_i^2}{n} - \overline{x}^2$$

فنلاحظ كيف أن التباين أصبح يساوي متوسط مربعات القيم Xi مطروحا منه مربع المتوسط$\overline{x}$؛ و هذه الصيغة المختصرة تُسهل من العملية الحسابية و تمكن من إختزالها، خاصة و أنها تجنب القيم بعد الفاصلة عند حساب الإنحرافات $x_i\overline{x}$- في صيغة التعريف،فهي بذلك أسرع و أدق.

هذا و يمكننا البرهنة كذلك أن التباين في حال وجود تكرارات أو بيانات مبوبة في فئات يساوي:

$$\sigma^2_{(x)} = \frac{\sum_{i=1}^{n} n_i (x_i - \overline{x})^2}{\sum_{i=1}^{n} n_i} = \frac{\sum_{i=1}^{n} n_i\, x_i^2}{\sum_{i=1}^{n} n_i} - \overline{x}^2$$

و الإنحراف المعياري هو الجذر التربيعي الموجب للتباين مهما كانت الصيغة المستعملة في حساب هذا الأخير.

7-خصائص في حساب التباين و الإنحراف المعياري:

التباين و الإنحراف المعياري هما مقياسان ليسا خطيان كالمتوسط الحسابي، إلاّ أنهما يمتلكان خصائص مهمة جداً نذكر منها:

- تباين العدد الثابت يساوي "الصفر"، فإذا كان b عدداً ثابتاً فإن: V(b) = 0
- V(X±b) = V(X) ⇒σ(X±b) = σ(X):

و هذا يعني أنه إذا أضفنا (أو طرحنا) مقداراً ثابتاً إلى (من) كل قيمة من قيم المفردات فإن ذلك لن يؤثر و لن يغير من تشتت هذه القيم، و الإنحراف المعياري بعد الإضافة أو الطرح هو نفسه الإنحراف المعياري للقيم الأصلية قبل التعديل.

- نعتبر السلسلة الإحصائية Xi حيث i=1,2,...,n وYi سلسلة إحصائية أخرى حيث : ∀ i : $y_i = x_i \pm b$ ؛ فإن الإنحراف المعياري للسلسلة Yi يساوي الإنحراف المعياري لـ Xi، بمعنى: V(Y) = V(X) ⇒σ(Y) = σ(X)
- إذا كانت لدينا السلسلة الإحصائية Xi حيث i=1,2,...,n و قمنا بضرب قيم هذه السلسلة في عدد حقيقي ثابت a فإن التباين الجديد سوف يساوي التباين الأصلي (تباين السلسلة قبل التعديل) مضروباً في مربع هذا العدد و منه الإنحراف المعياري مضروباً في القيمة المطلقة للعدد:$V(aX) = a^2V(X)$ et σ(aX) = │a│σ(X).
- نعتبر السلسلتين الإحصائيتين Xi و Yi حيث i=1,2,...,n و : ∀ i : $y_i = ax_i + b$ فإن : $V(Y) = V(ax_i + b) = a^2V(X)$ ⇒σ(Y) = │a│σ(X).

- **8-متباينة "تشيبتشيف":**
- إن العلاقة التجريبية السابقة التي تحدد لنا نسب المشاهدات التي يحتويها كل مجال بوحدات إنحراف معياري إبتعاداً عن المتوسط الحسابي تُستعمل في حال ما إذا كانت التوزيعات متماثلة تأخذ الشكل المعتدل على هيئة الجرس؛ لكن ماذا لو كان توزيع البيانات غير متماثل، بمعنى ملتوي؟. في مثل هذه الحالات لا يمكن تطبيق القانون التجريبي السابق لأن تلك القاعدة التجريبية تصبح غير دقيقة؛ لكنه يمكننا في هذا الصدد أن نلجأ إلى إستخدام متباينة أو نظرية "تشيبتشيف" التي تُعد تعميماً للقاعدة التجريبية على كل أشكال التوزيعات.
- تنص هذه النظرية على أنه، و بصرف النظر عن شكل التوزيع، فإن نسبة المشاهدات التي لا تبعد عن المتوسط بأكثر من k إنحراف معياري هي **على الأقل**$1 - \frac{1}{k^2}$حيث (k>1).

- بمعنى أن المجال $[\overline{x}-k\sigma\,;\,\overline{x}+k\sigma]$ مثلاً سوف يحتوي على نسبة من المشاهدات مقدارها $1-\frac{1}{k^2}$ **على الأقل**. و لا بد من الإشارة إلى أن هذه النسبة $1-\frac{1}{k^2}$ التي تنص عليها نظرية تشيبتشيف هي الحد الأدنى للنسبة الفعلية من قيم التوزيع التي تقع في المجال $[\overline{x}-k\sigma\,;\,\overline{x}+k\sigma]$؛ و تعتبر هذه النظرية متحفظة جداً كونها تنطبق على كل التوزيعات الإحصائية بصرف النظر عن شكلها[1].
- فمثلاً، إذا كان k=2 فإن هذا يدل على أن المجال $[\overline{x}-2\sigma\,;\,\overline{x}+2\sigma]$سوف يحتوي على 75%$(1-\frac{1}{2^2}=\frac{3}{4}=0.75)$ على الأقل من قيم البيانات.[2] أما إذا كان k=3 فإن ذلك يعني، وفق نظرية تشيبتشيف، أن المجال $[\overline{x}-3\sigma\,;\,\overline{x}+3\sigma]$ يحتوي على $(1-\frac{1}{3^2}=\frac{8}{9}\approx 0.89)$ حوالي 89% على الأقل من مجموعة المشاهدات.

II-مقاييس التشتت النسبي:

الإنحراف المعياري يُشكل مقياساً كمياً ناجحاً لقياس التشتت داخل سلسلة إحصائية معينة؛ و كلما كان الإنحراف صغيراً في قيمته إستنتجنا أن البيانات أكثر تجانساً فيما بينها و أقل تشتتاً، إضافةً إلى أنه يُقاس بنفس وحدة قياس البيانات المحسوب منها. لكن إذا أردنا المقارنة بين مجموعتين من القيم (البيانات) أو كانت وحدات القياس تختلف من مجموعة إلى أخرى فإننا نلجأ إلى إستخدام "مقاييس التشتت النسبية".

فمقاييس التشتت النسبية تُستخدم إذن للمقارنة بين مجموعتين أو أكثر من البيانات، كالمقارنة بين علامات فوجين من الطلبة مثلاً، أو إذا كانت وحدات القياس في المجموعات تختلف، كمقارنة تشتت وحدات الدخل مع وحدات الوزن، أو وحدات الدخل بين بلدين بعملات مختلفة.... و بالتالي فإن مقاييس التشتت النسبية تعتبر مجردة من وحدة القياس و يعبر عنها بنسبة مئوية نحصل عليها عن طريق قسمة أحد مقاييس التشتت المطلق على مقياس من مقاييس النزعة المركزية.

و من أهم هذه المقاييس:

1-معامل الإختلاف: Coefficient de Variation

يُعرف كذلك بمعامل التغير و نرمز له بـ"CV". تم إقتراح هذا المقياس من قبل "كارل بيرسون" Karl Pearson، و هو أهم مقاييس التشتت النسبي و أكثرها إستخداماً حيث يُستعمل لإزالة أثر وحدة القياس و المقارنة بين متغيرين مُقاسين بوحدتين مختلفتين أو كذلك إذا تساوى الإنحراف المعياري لمجموعتين فأكثر؛ و مجموعة القيم التي يكون فيها معامل الإختلاف هو الأكبر تكون أكثر تشتتاً و أقل تجانساً و إتساقاً بين مفرداتها.

معامل الإختلاف هو عبارة عن نسبة الإنحراف المعياري σ إلى المتوسط الحسابي $\overline{x}$ لمجموعة من البيانات. و يُحسب بالصيغة التالية:

$$C.V=\frac{\text{المعياريالإنحراف}}{\text{الحسابيالمتوسط}}\times 100=\frac{\sigma}{\overline{x}}\times 100$$

---

[1] لأن عبارة "ما لا يقل عن أو على الأقل" تجعل نظرية تشيبتشيف متحفظة، و في معظم الحالات تكون النسبة الفعلية أكبر من المقدار $1-\frac{1}{k^2}$ خاصة إذا كانت البيانات الإحصائية أقرب للتناظر.

[2] لاحظ أنه حسب القاعدة التجريبية فإن هذا المجال يتضمن حوالي 95% من المشاهدات عندما يكون التوزيع معتدلاً ذو شكل جرسي.

من هذا القانون نستنتج أنه بالنسبة لمجموعتين من البيانات:

- إذا كان الوسطان الحسابيان متساويين فإن المجموعة ذات الإنحراف المعياري الأكبر تكون أكثر تشتتاً؛
- إذا كان الإنحرافان المعياريان متساويين فإن المجموعة ذات الوسط الحسابي الأكبر تكون أقل تشتتاً؛
- إذا كان المتوسط الحسابي في إحدى المجموعتين يساوي الصفر، كما في المجموعة النظامية مثلاً، فلا يُمكن حساب معامل الإختلاف لتلك المجموعة و لا يُمكن مقارنتها بغيرها.

2-الدرجة المعيارية (القياسية):

إذا كان لدينا مجموعة من المشاهدات (البيانات)$x_1,x_2,\dots,x_n$متوسطها الحسابي $\overline{X}$ و إنحرافها المعياري σ، فإن الدرجة المعيارية، التي نرمز لها بـ"Z"، لأي قيمة من مجموعة المشاهدات تُعرف بالعلاقة التالية:

$$Z_i = \frac{x_i - \overline{x}}{\sigma} \quad avec\ i = 1,2,\dots,n.$$

فنلاحظ من خلال التعريف بأن الدرجة المعيارية هي إنحراف قيمة معينة عن المتوسط الحسابي (المسافة على يمين أو يسار $\overline{x}$ معبراً عنه بوحدات الإنحراف المعياري.

كما نشير إلى أن تحويل مجموعة من القيم إلى الشكل المعياري يجعل متوسطها الحسابي مساوياً للصفر و إنحرافها المعياري مساوياً للواحد الصحيح و هو ما يعطينا مجتمعاً معيارياً (مجموعة نظامية) له نفس شكل توزيع القيم الخام الأصلية.

و لبرهان أن المجتمع المعياري (النظامي) متوسطه يساوي الصفر و انحرافه المعياري هو الواحد الصحيح فإننا نحسب متوسط القيم المعياريةZ_iو إنحرافها المعياري فنجد:

لدينا: $Z_i = \frac{x_i - \overline{x}}{\sigma} \quad avec\ i = 1,2,\dots,n.$ إذن:

- $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{x_i-\overline{x}}{\sigma}\right)$

$$= \frac{1}{n\sigma}\sum_{i=1}^{n}(x_i - \overline{x})$$

$$= \mathbf{0}$$

تذكر أن من خصائص المتوسط الحسابي : $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$

- $V(Z_i) = \frac{\sum\left(Z_i - \overline{Z}\right)^2}{n}$

$$= \frac{\sum Z_i^2}{n}$$

$$= \frac{1}{n}\sum\left(\frac{x_i - \overline{x}}{\sigma}\right)^2$$

$$= \frac{1}{n\sigma^2} \sum (x_i - \overline{x})^2$$

{$\sigma^2_{(x)} = \frac{\sum (x_i - \overline{x})^2}{n}$:تذكر أن ☚}

$$= \frac{}{\sigma^2} = 1$$

$$\Rightarrow \sigma_{Z_i} = \sqrt{1} = \mathbf{1}$$

و هذا يعني أن معايرة مجموعتين (أو أكثر) يجعل لهما نفس المتوسط، و هو الصفر، و نفس الإنحراف المعياري و هو الواحد؛ لكن هذا لا يعني بالطبع أن كل المجموعات المعيارية (النظامية) متشابهة من حيث الشكل فيما بينها.