f_i : يمثل تكرار الفئة i.

4. التباين Variance:

يعرف التباين بأنه متوسط مجموع مربعات انحرافات القيم عن وسطها الحسابي. أي أن التباين ما هو إلا مربع الانحراف القياسي لتلك المجموعة من القيم ويرمز له بالرمز S^2.

طرق حساب التباين:

أ. في حالة البيانات الغير مبوبة:

لتكن $X_1, X_2, ..., X_n$ تمثل قياسات عينة من المفردات حجمها n وليكن $\overline{X}$ الوسط الحسابي لهذه القيم عندئذٍ يمكن حساب التباين وفق الصيغة التالية:

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}$$

ب. في حالة البيانات المبوبة:

لتكن $X_1, X_2, ...X_m$ تمثل مراكز فئات توزيع تكراري عدد فئاته m وإن $f_1, f_2, ...f_m$ تمثل التكرارات المقابلة لهذه الفئات. عندئذٍ يمكن حساب التباين وفق الصيغة التالية:

$$S^2 = \frac{\sum_{i=1}^{m} f_i \left(X_i - \overline{X}\right)^2}{\sum_{i=1}^{m} f_i}$$

إذ أن:

X_i : يمثل مركز الفئة i.

$\overline{X}$: يمثل الوسط الحسابي للتوزيع.

f_i : يمثل تكرار الفئة i.

مثال (1):-

البيانات التالية تمثل أوزان عينة من الطلبة قوامها عشرة طلاب جد التباين والانحراف القياسي لهذه البيانات:

56, 62, 69, 71, 68, 65, 63, 72, 68, 56

الحل:-

$$X_i = 56, \quad 62, \quad 69, \quad 71, \quad 68, \quad 65, \quad 63, \quad 72, \quad 68, \quad 56$$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{650}{10} = 65$$

$$\left(X_i - \overline{X}\right) = -9, \quad -3, \quad 4, \quad 6, \quad 3, \quad 0, \quad -2, \quad 7, \quad 3, \quad -9$$

$$\left(X_i - \overline{X}\right)^2 = 81, \quad 9, \quad 16, \quad 36, \quad 9, \quad 0, \quad 4, \quad 49, \quad 9, \quad 81$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}} = \sqrt{\frac{294}{10}} = \sqrt{49.4} = 7.0285$$

$S^2 = 49.4$ $= \frac{\Sigma (x_i - \bar{x})^2}{n}$

مثال (2):- من جدول التوزيع التكراري التالي جد الانحراف القياسي والتباين.

الفئات	التكرار f_i
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
المجموع	100

الحل:-

الفئات	التكرار f_i	مركز الفئة X_i	$f_i X_i$	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$	$f_i\left(X_i - \overline{X}\right)^2$
60-62	5	61	305	-6.45	41.6025	208.0125
63-65	18	64	1152	-3.45	11.9025	214.2450
66-68	42	67	2814	-0.45	0.2025	8.5050
69-71	27	70	1890	2.55	6.5025	175.5675
72-74	8	73	584	5.55	30.8025	246.4200
المجموع	100		6745			852.7500

$$\overline{X} = \frac{\sum_{i=1}^{m} f_i X_i}{\sum_{i=1}^{m} f_i} = \frac{6745}{100} = 67.45$$

$$S = \sqrt{\frac{\sum_{i=1}^{m} f_i \left(X_i - \overline{X}\right)^2}{\sum_{i=1}^{m} f_i}} = \sqrt{\frac{852.7500}{100}} = \sqrt{8.5275} = 2.92$$

$$S^2 = 8.5275 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}$$

ب- مقاييس التشتت النسبية:

1. معامل التشتت المستند إلى الانحراف القياسي (معامل الاختلاف) :Coefficient of Variation

إذا كان $\overline{X}$ يمثل الوسط الحسابي لمجموعة قيم و S يمثل الانحراف القياسي لهذه القيم عندئذٍ يعرف معامل الاختلاف (يرمز له بالرمز C.V) بالشكل الأتي:

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

إن معامل الاختلاف يعتبر أفضل أنواع معاملات التشتت كونه يعتمد على أفضل مقياس نزعة مركزية وهو $\overline{X}$ وأفضل مقياس تشتت S. إن هذا المعامل يوضح نسبة حصة كل وحدة من وحدات الوسط الحسابي من الانحراف المعياري. وعليه وعند إجراء مقارنة بين قيم مجموعتين تتم مقارنة معامل اختلاف الأولى مع معامل اختلاف الثانية وعندئذٍ يقال عن المجموعة بأنها أكثر تجانساً إذا كان معامل اختلافها اقل من الأخرى.

مثال:-

إذا كان متوسط درجات طلبة الصف الأول إحصاء في امتحان الرياضيات 69 درجة وبانحراف قياسي قدره 19.3 في حين كان متوسط درجاتهم في امتحان الإحصاء 75 درجة وبانحراف قياسي قدره 25.5. في أي من الامتحانين كان مستوى أداء الطلبة أكثر تقارباً.

الحل:-

نجد معامل الاختلاف لكل مادة

$$C.V(\text{رياضيات}) = \frac{S}{\overline{X}} \cdot 100$$
$$= \frac{19.3}{69} .100$$
$$= 28\%$$

$$C.V(\text{احصاء}) = \frac{S}{\overline{X}} \cdot 100$$
$$= \frac{25.5}{75} .100$$
$$= 34\%$$

نلاحظ أن معامل الاختلاف في امتحان الرياضيات اقل من معامل اختلاف امتحان مادة الإحصاء لذا فان مستوى أداء الطلبة في امتحان الرياضيات كان أكثر تقارباً.

الدرجة المعيارية (القياسية) Standard Score:

في كثير من الأحيان نحتاج إلى تحويل قيم المتغير X إلى شكل أخر يدعى بالشكل القياسي أو الشكل المعياري. حيث أننا نحتاج إلى هذا التحويل للقيم في حالات المقارنة بين مجموعتين أو أكثر من القيم ذات الأوساط الحسابية والانحرافات المعيارية المختلفة قيمها.

فإذا كانت $X_1, X_2, ..., X_n$ تمثل قياسات عينة من المفردات حجمها n وان $S, \overline{X}$ يمثلان على التوالي الوسط الحسابي والانحراف المعياري لقيم هذه العينة عندئذٍ تعرف الدرجة القياسية Z لأية قيمة من قيم X على النحو التالي:

$$Z_i = \frac{X_i - \overline{X}}{S} \quad , \quad i = 1,2,....,n$$

ففي حالة البيانات الغير مبوبة فان:

Z_i : تمثل الشكل القياسي لقيم المتغير العشوائي.

X_i : تمثل القيم الأصلية للمتغير العشوائي.

$\overline{X}$: الوسط الحسابي للقيم الأصلية للمتغير العشوائي.

S : الانحراف المعياري للقيم الأصلية للمتغير العشوائي.

أما في حالة التوزيعات التكرارية (البيانات المبوبة) فان:

Z_i : تمثل الشكل القياسي لقيم المتغير العشوائي.

X_i : تمثل مركز الفئة i.

$\overline{X}$: الوسط الحسابي للتوزيع.

S: الانحراف المعياري للتوزيع.

ملاحظات:

1- إن الدرجة القياسية خالية من وحدات قياس المتغير الأصلي.

2- إن الوسط الحسابي للدرجة القياسية يساوي صفر وتباينها يساوي واحد.

مثال (1):-

كانت درجات خمسة طلاب في مادتي الرياضيات والإحصاء هي:

تسلسل الطالب	درجة الرياضيات X	درجة الإحصاء Y
1	219.04 50	62
2	163.84 52	77
3	17.64 69	82
4	51.84 72	85
5	262.44 81	86

يطلب حساب الدرجات القياسية لكل مادة.

142.96 $\frac{714.8}{5}$

الحل:-

نجد الوسط الحسابي والانحراف القياسي لكل من درجة الرياضيات ودرجة الإحصاء لذا فان:

$$\overline{X} = 64.8, \quad S_x = 11.96$$

$$\overline{Y} = 78.4, \quad S_y = 8.78$$

$$Z_i = \frac{X_i - \overline{X}}{S}$$

تسلسل الطالب	درجة الرياضيات X	درجة الإحصاء Y	Z_x	Z_y
1	50	62	-1.24	-1.87
2	52	77	-1.07	-0.16
3	69	82	0.35	0.41
4	72	85	0.60	0.75
5	81	86	1.36	0.87

نلاحظ هنا مثلاً ولإغراض المقارنة أن الطالب الخامس كانت درجته في الإحصاء 86 أفضل منها في الرياضيات 81 كمقارنة مطلقة. إلا انه من حيث الأهمية فان درجته في الرياضيات هي في الحقيقة أفضل منها في الإحصاء حيث أن الدرجة القياسية في الرياضيات كانت أعلى من الدرجة القياسية في الإحصاء.