الفصل الرابع

مقاييس النزعة المركزية :

(مقاييس التوسط)

المقدمة :

لاحظنا في الفصلين والثاني والثالث استعراض لاهم اساليب جمع وتصنيف وتبويب البيانات وكيفية تمثيلها في جداول ورسوم هندسية وبيانية ، الهدف من ذلك اعطاء صورة سريعة توضح ما هي هذه البيانات . في هذا الفصل سوف نتطرق للحديث عن كيفية تمثيل مجموعة من البيانات بقيمة واحدة فقط من خلال مقياس يدعى بمقياس النزعة المركزية .

مفهوم المتوسطات والهدف من احتسابها :

يمكن تمثيل مجموعة من البيانات بقيمة واحدة فقط الهدف من ذلك اعطاء صورة سريعة عن ماهية تلك المجموعة من خلال ايجاد عدد يمثل قيمتها . ان المقياس الذي يختص بتحديد هذا العدد يسمى مقياس نزعة مركزية او مقياس توسط (متوسط) وان هذا العدد يميل بأن يقع في وسط تلك المجموعة من البيانات في حال ترتيبها حسب صغرها او كبرها ولهذا اطلق عليها مقياس النزعة المركزية والمتوسطات مهمة من ناحية

1- تقدير قيم عددية لبعض مؤشرات المجتمع .
2- دراسة خصائص المجتمعات من خلال العينة والمتوسط احد هذه الخصائص .
3- اهميته من الناحية العلمية في كثير من جوانب الحياة العملية مثل التخطيط الاقتصادي وخطط برامج التنمية ، فالخطط الاقتصادية للدولة التي من شأنها رفع مستوى رفاهية الشعب غالباً ما يكون الهدف منها رفع (متوسط) دخل الفرد الى حالة افضل من سابقتها .

انواع مقاييس النزعة المركزية :

1. الوسط الحسابي (Mean) :

ويسمى في بعض الاحيان بالمتوسط او الوسط او المعدل الحسابي (Average) وهو احد اهم مقاييس النزعة المركزية على الاطلاق لما يمتاز به من خصائص جيدة وسهولة في الحساب جعلته يقف في مقدمة مقاييس النزعة المركزية . فمثلاً لو كنا بصدد حساب ((معدل)) طالب تخرج من الدراسة الاعدادية سنقوم بجمع درجاته في جميع الدروس وقسمتها على عددها بهذا سنحصل على قيمة واحدة وهي المعدل ، وهذا يعني اننا تمكنا من تمثيل مجموعة من البيانات بقيمة واحدة وتمركزت في وسط هذه البيانات .

طرق حساب الوسط الحسابي :

1- احتساب الوسط الحسابي لبيانات غير مبوبة :

اولاً : الطريقة المباشرة (القيم الاصلية) :

افترض ان $x_1 , x_2 , ... , x_n$ تمثل بيانات او قياسات او قيم للمتغير العشوائي (X) لعينة حجمها (n) فيمكن حساب الوسيط الحسابي الذي عادةً ما يرمز له بالرمز $\overline{X}$ ، والذي يعرف على انه دالة بدلالة قياسات مفردات العينة اي ان :

$$\overline{X} = f (x_1 + x_2 + ... + x_n)$$

بحيث ان هذه الدالة تأخذ الشكل الاتي :

$$\overline{X} = \frac{1}{n} (x_1 + x_2 + ... + x_n)$$

ووفق هذا التعريف فأن الوسط الحسابي ما هو الا مجموع قياسات مفردات العينة مقسوماً على عددها اي ان :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

حيث ان :

$\overline{X}$: هو الوسط الحسابي

n : هو حجم العينة

مثال / جد الوسط الحسابي للبيانات الاتية باستخدام الطريقة المباشرة :

$$X_i = 4 , 3 , 7 , 2 , 3 , 5$$

$$\overline{X} = \frac{\sum_{i=1}^{6} x_i}{n}$$

$$\overline{X} = \frac{4+3+7+2+3+5}{6} = \frac{24}{6} = 4$$

ثانياً : الطريقة المختصرة (طريقة الانحرافات)

تستخدم هذه الطريقة في حالة كون قياسات العينة اعداد كبيرة يصعب التعامل معها عند ايجاد الوسيط الحسابي ، ولغرض تسهيل العينة نختار وسط فرضي ثم نجد انحرافات هذه القيم عند الوسط الفرضي باستخدام الصيغة التالية :

$$\overline{X} = a + \frac{\sum_{i=1}^{n} d_i}{n}$$

حيث ان

a : هو وسط فرضي

$d_i = X_i - a$ هو يمثل انحرافات القيم عن الوسط الفرضي .

مع الملاحظة انه ليس من الضروري ان تكون قيمة a المختارة من ضمن البيانات المتاحة الا انه يفضل ان تكون احدى قيم المجموعة القريبة من مركزها ففي ذلك تسهيل للعمل الحسابي .

مثال[1]/ اوجد الوسط الحسابي باستخدام طريقة الانحرافات للبيانات الاتية :

Xi = 4, 2, 3, 0, 6

الحل / نفرض ان a=2

Xi	di = Xi – a
4	4 – 2 = 2
2	2 – 2 = 0
3	3 – 2 = 1
0	0 – 2 = -2
6	6 – 2 = 4
	5

$$\overline{X} = a + \frac{\sum_{i=1}^{n} d_i}{n} = 2 + \frac{\mathbf{5}}{\mathbf{5}} = 2 + 1 = 3$$

مثال[2] / اوجد الوسط الحسابي للبيانات الاتية باستخدام الطريقة المباشرة وطريقة الانحرافات (الطريقة المختصرة) وافترض ان a = 5

Xi = 3 , 7 , 4 , 7 , 8 , 5 , 10 , 9 , 2 , 5 , 6 , 9

الحل /

الطريقة المباشرة :

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

$$\overline{X} = \frac{\mathbf{75}}{\mathbf{12}} = 6.25$$

طريقة الانحرافات :

$$\overline{X} = a = \frac{\sum_{i=1}^{n} d_i}{n}$$

d_i = -2 , 2 , -1 , 2 , 3 , 0 , 5 , 4 , -3 , 0 , 1 , 4 $\qquad \sum_{i=1}^{n} d_i = 15$

$$\overline{X} = 5 + \frac{\mathbf{15}}{\mathbf{12}} = 6.25$$

2- احتساب الوسط الحسابي لبيانات مبوبة :

اولاً : الطريقة المباشرة (القيم الاصلية) :

افترض ان X_m , ... , X_3 , X_2 , X_1 تمثل مراكز فئات توزيع تكراري عدد فئاته m ولتكن f_m , ... , f_2 , f_1 تمثل التكرارات المقابلة لفئات التوزيع ، عندئذٍ يستخرج الوسط الحسابي لهذا التوزيع باستخدام الصيغة التالية سواء كانت اطوال الفئات متساوية او غير متساوية :

$$\overline{X} = \frac{\sum_{i=1}^{m} f_i \quad X_i}{\sum_{i=1}^{m} f_i} \qquad , \quad \sum_{i=1}^{m} f_i = n$$

حيث ان :

X_i = يمثل مراكز الفئات

f_i = يمثل تكرار الفئات

مثال / الاتي توزيع تكراري لدرجات الحرارة في مدينة معينة مسجلة لمدة 95 يوماً متتالياً يطلب حساب متوسط درجة الحرارة في هذه المدينة خلال هذه الفترة .

درجات الحرارة	0-	1-	2-	3-	4-	5-	6-	7 – 8
عـــدد الايـــام	4	8	12	16	20	25	6	4

الحل /

Classes	fi	Xi	fiXi
0 –	4	0.5	2
1 –	8	1.5	12
2 –	12	2.5	30
3 –	16	3.5	56
4 –	20	4.5	90
5 –	25	5.5	137.5
6 –	6	6.5	39
7 – 8	4	7.5	30
	95		396.5

لذلك فان الوسط الحسابي سيكون :

$$\overline{X} = \frac{\sum_{i=1}^{m} f_i \quad X_i}{\sum_{i=1}^{m} f_i} = \frac{\mathbf{396.5}}{\mathbf{95}} = 4.174°$$

ثانياً : طريقة الانحرافات (الطريقة المختصرة) :

تستخدم هذه الطريقة عندما تكون مراكز الفئات اعداد كبيرة حيث نختاروسط فرضي (مركز الفئة المقابل لأكبر تكرار) ثم نجد انحرافات مراكز الفئات عن هذا الوسط الفرضي وذلك باستخدام الصيغة الاتية :

$$\overline{X} = a + \frac{\sum_{i=1}^{m} f_i \quad d_i}{\sum_{i=1}^{m} f_i}$$

حيث ان :

a : هو الوسط الفرضي

d_i : هو انحرافات مراكز الفئات عن الوسط الفرضي اي ان $d_i = X_i - a$

مثال / الاتي توزيع تكراري لأطوال عينة من الاشخاص قوامها 70 شخص يطلب ايجاد متوسط طول الشخص في هذه العينة ، افرض a = 168

فـــئات الطول	140-	148-	156-	164-	172-	180-	188 – 196
عدد الاشخاص	4	6	15	20	17	7	1

الحل /

Classes	fi	Xi	di	fidi
140 –	4	144	-24	-96
148 –	6	152	-16	-96
156 –	15	160	-8	-120
164 –	20	168	0	0
172 –	17	176	8	136
180 –	7	184	16	112
188 – 196	1	192	24	24
	70			-40

$$\overline{X} = a + \frac{\sum_{i=1}^{m} f_i \quad d_i}{\sum_{i=1}^{m} f_i}$$

$$= 168 + \frac{\mathbf{-40}}{\mathbf{70}} = 168 - 0.571$$

$$\overline{X} = 167.429$$

ثالثاً : طريقة الانحرافات المقسومة على طول الفئة (الطريقة القصيرة)

تستخدم هذه الطريقة في حالة تساوي اطوال الفئات للتوزيع التكراري فاذا فرضنا ان $x_1 , x_2 , \ldots\ldots , x_m$ تمثل مراكز فئات توزيع تكراري عدد فئاته m وان اطوال فئات التوزيع متساوية وتساوي L ولتكن $f_1 , f_2 , \ldots , f_m$ تمثل التكرارات المقابلة لفئات التوزيع عندئذٍ يمكن ايجاد قيمة الوسط الحسابي باستخدام الصيغة التالية :

$$\mathbf{\overline{X} = a + \frac{\sum_{i=1}^{m} f_i d_i^*}{\sum_{i=1}^{m} f_i} . L}$$

حيث ان : $\mathbf{d_i^* = \frac{x_i - a}{L}}$

تستخدم هذه الطريقة في حالة تساوي اطوال الفئات متساوية فقط ، اذا لم تكن الاطوال متساوية نستخدم الطريقات السابقتان .

مثال / لبيانات المثال السابق (49) احسب طول الشخص بالطريقة القصيرة

الحل /

Classes	fi	Xi	d_i^*	$f_i \quad d_i^*$
140 –	4	144	-3	-12
148 –	6	152	-2	-12
156 –	15	160	-1	-15
164 –	20	168	0	0
172 –	17	176	1	17
180 –	7	184	2	14
188 – 196	1	192	3	3
	70			-5

$$\bar{X} = a + \frac{\sum_{i=1}^{m} f_i \quad d_i^*}{\sum_{i=1}^{m} f_i} \times L = 168 + \frac{-5}{70} \quad (8)$$

$$\therefore \bar{X} = 168 - 0.571 = 167.429$$

وفيما يلي جدول يوضح خلاصة الطرق المختلفة لإيجاد الوسط الحسابي :

نوع البيانات	الطريقة المباشرة	الطريقة المختصرة	الطريقة القصيرة
بيانات غير مبوبة	$\frac{\sum_{i=1}^{n} X_i}{n}$	$a + \frac{\sum_{i=1}^{n} d_i}{n}$	ــــــــ
بيانات مبوبة	$\frac{\sum_{i=1}^{m} f_i \quad x_i}{\sum_{i=1}^{m} f_i}$	$a + \frac{\sum_{i=1}^{m} f_i \quad d_i}{\sum_{i=1}^{m} f_i}$	$a + \frac{\sum_{i=1}^{m} f_i \quad d_i^*}{\sum_{i=1}^{m} f_i} \times L$

مزايا وعيوب الوسط الحسابي :

المزايا :

1. يمتاز ببساطة فكرته .
2. سهولة حسابه .
3. خضوعه لعمليات جبرية بسيطة .
4. حسابه يستند الى كافة البيانات المتاحة .

العيوب :

1. لا يمكن تعيينه هندسياً ولا يمكن استخراج قيمته عن طريق التأمل .
2. لا يمكن ايجاد وسط حسابي لبيانات متغير وصفي (نوعي) اي المتغير الذي يقاس بالصفة (كالجنس واللون) الا في حالة الظواهر التي يمكن اعادة تقييس الصفات فيها على سبيل المثال تقديرات الامتحانات المعتمدة في التعليم العالي (مقبول ، متوسط ، جيد ، جيد جداً ، امتياز) حيث يمكن وصفها كالاتي :

التقدير	Classes
ضعيف	0 –
مقبول	50 –
متوسط	60 –
جيد	70 –
جيد جداً	80 –
امتياز	90 – 100

3. لا يمكن حساب الوسط الحسابي اذا فقدت احدى القيم الظاهرة الا في حالة تقدير القيمة المفقودة والاستعاضة عنها بقيمة مكملة لحجم العينة .
4. يتأثر الوسط الحسابي على نحو كبير بالقيم الشاذة المتطرفة وفي مثل هذه الحالة لا يجوز الاعتماد على هذا المقياس وانما يستوجب الاعتماد على مقياس اخر ونقوم باستبعاد البيانات الشاذة وفق اعتبارات خاصة والتعويض عنها بقيم مقدرة او بديلة . على سبيل المثال لو تم اختيار عينة عشوائية من طلبة صف معين قوامه 10 طلاب وتم ايجاد وزن كل منهم وكانت النتائج كما يلي : 57 ، 96 ، 61 ، 57 ، 61 ، 59 ، 92 ، 58 ، 52 ، 59 يلاحظ وجود طالبين وزنهما غير متناسقين مع اوزان بقية الطلاب وهما الطالب الرابع والتاسع . وعند حساب متوسط وزن الطالب في هذه العينة يلاحظ ان قيمة هذا المتوسط هي 65.2 كغم قد انحازت نحو القيمتين 96 ، 92 وهذا امر مخالف لإحدى اهم صفات الوسط الحسابي الا وهي صفة تمركزه وسط المجموعة او قريباً من وسطها ، في حين لو تم استعباد هاتين القيمتين واعادة احتساب قيمة الوسط الحسابي للبقية لكان الناتج مساوٍ الى 58 كغم .
5. يتأثر الوسط الحسابي بأخطاء المعاينة .
6. لا يمكن ايجاد قيمة الوسط الحسابي في حالة التوزيعات التكرارية المفتوحة من طرف واحد او طرفين وذلك لتعذر استخراج مراكز الفئات المفتوحة لذا يتم تقدير الوسط الحسابي من خلال العلاقة بأوساط اخرى .
7. لا يجوز الاعتماد على الوسط الحسابي للتوزيعات التكرارية ذات الالتواء الحاد سواء كانت القيمة سالبة او موجبة بسبب انحياز قيمة الوسط الحسابي نحو جهة الالتواء .

خصائص الوسط الحسابي :

1- مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر

$\sum_{i=1}^{n}(X_i - \overline{X}) = 0$ اي ان

في حالة البيانات الغير مبوبة

مثال[1] / اثبت بان مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر للبيانات التالية :

X_i	$(X_i - \overline{X})$
4	4 – 5 = -1
6	6 – 5 = 1
5	5 – 5 = 0
5	5 – 5 = 0
20	0

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\mathbf{20}}{\mathbf{4}} = 5$$

$$\therefore \sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

في حالة البيانات المبوبة :

مثال[2] / لبيانات الجدول التالي اثبت ان مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر :

Classes	fi	Xi	fi Xi	$X_i - \overline{X}$	$f_i (X_i - \overline{X})$
10 –	2	15	30	-30	-60
20 –	5	25	125	-20	-100
30 –	10	35	350	-10	-100
40 –	20	45	900	0	0
50 –	10	55	550	10	100
60 –	5	65	325	20	100
70 – 80	2	75	150	30	60
	54		2430		0

$$\overline{X} = \frac{\sum_{i=1}^{m} f_i \ X_i}{\sum_{i=1}^{m} f_i} = \frac{\mathbf{2430}}{\mathbf{54}} = 45$$

سنلاحظ ان مجموع العمود السادس مساوي الى صفر اي ان مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر .

2. مجموع مربعات انحرافات القيم عن وسطها الحسابي تكون اقل ما يمكن اي ان

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 < \sum_{i=1}^{n} (X_i - A)^2$$

ونفس الحالة في حالة البيانات المبوبة فإن :

$$\sum_{i=1}^{n} f_i(X_i - \overline{X})^2 < \sum_{i=1}^{m} f_i(X_i - A)^2$$

مثال / لبيانات المثال التالي اثبت ان مجموع مربعات انحرافات القيم عن وسطها الحسابي هو اقل ما يمكن على فرض ان قيمة $A = 5$

$X_i = 4 , 5 , 6 , 3 , 2$

الحل /

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 < \sum_{i=1}^{n} (X_i - A)^2$$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{4+5+6+3+2}{5} = \frac{20}{5} = 4$$

$$\sum_{i=1}^{5}(X_i - \overline{X})^2 = (4 - 4)^2 + (5 - 4)^2 + (6 - 4)^2 + (3 - 4)^2 + (2 - 4)^2$$

$$= 0 + 1^2 + 2^2 + (-1)^2 + (-2)^2$$

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = 10$$

$$\sum_{i=1}^{n} (X_i - A)^2 = (4 - 2)^2 + (5 - 5)^2 + (6 - 5)^2 + (3 - 5)^2 + (2 - \ )^2$$

$$= (-1)^2 + 0 + 1^2 + (-2)^2 + (-3)^2$$

$$= 1 + 0 + 1 + 4 + 9 = 15$$

$$\sum_{i=1}^{n} (X_i - A)^2 = 15$$

الوسط الحسابي المرجح :

فيما سبق تم التطرق الى الاساليب المختلفة لحساب الوسط الحسابي من بيانات مبوبة وغير مبوبة ، الا ان ذلك استند الى اعتبار قياسات مفردات العينة لها نفس الاهمية في المساهمة لإيجاد قيمة هذا المقياس حيث لم يتم تفضل مفردة على الاخرى ، الا انه ومن الناحية العملية

هناك الكثير من الحالات تكون بعض المفردات اكثر اهمية من الاخرى مما يستوجب ذلك اخذه بنظر الاعتبار لدى حساب الوسط الحسابي . فمثلاً عند حساب معدل درجات الطالب المتخرج من الكلية فان الامر يستوجب الاخذ بنظر الاعتبار عدد الساعات الاسبوعية المخصصة لكل درس من الدروس التي تدخل في حساب المعدل وهذا يعني ان هناك عملية ترجيح لقياسات المفردات بأوزان تمثل اهمية كل منها ، وهذه الاوزان غالباً ما تكون محددة سلفاً وعلى افتراض ان x_1 ، x_2 ، ..... ، x_n تمثل قياسات عينة من المفردات قوامها n وان W_1 ، W_2 ، ..... ، W_n تتمثل اوزان هذه المفردات عندئذً يعرف الوسط الحسابي المرجح بهذه الاوزان على النحو التالي :

$$\overline{X_w} = \frac{\sum_{i=1}^{n} W_i \quad X_i}{\sum_{i=1}^{n} W_i}$$

اما في حالة التوزيعات التكرارية المبوبة فان الوسط الحسابي المرجح هو :

$$\overline{X_w} = \frac{\sum_{i=1}^{n} W_i \quad f_i \quad X_i}{\sum_{i=1}^{n} W_i \quad f_i}$$

ويلاحظ مما تقدم ان الوسط الحسابي المرجح هو حالة اكثر عمومية من الوسط الحسابي الاعتيادي اي ما نعنيه هو ان الوسط الحسابي الاعتيادي هو حالة خاصة من الوسط الحسابي المرجح عندئذٍ ينظر الى كافة المفردات بنفس الاهمية .

مثال / كانت درجات احد طلبة الصف الثالث احصاء في الدروس المقررة في هذه المرحلة حسب الساعات الاسبوعية المحددة لكل درس هي كالاتي :

X_i الدرجات	W_i عدد الساعات
62	2
80	2
75	2
88	3
84	3
86	3
90	3
	18

يطلب حساب معدل درجات هذا الطالب .

الحل /

$$\overline{X_w} = \frac{\sum_{i=1}^{n} W_i \quad X_i}{\sum_{i=1}^{n} W_i}$$

$W_i X_i$

124
160
150
264
252
258
270
1478

$$\therefore \overline{X_w} = \frac{\mathbf{1478}}{\mathbf{18}} = 82.11$$

مثال[2] / في ما يلي توزيع تكراري لإنتاج مصنع معين المطلوب حساب متوسط انتاجه .

Classes	f_i	W_i
2 –	4	6
4 –	5	5
6 –	6	6
8 –	3	4
10 – 12	2	4

الحل /

$$\overline{X_w} = \frac{\sum_{i=1}^{n} W_i \quad f_i \quad X_i}{\sum_{i=1}^{n} f_i \quad W_i}$$

X_i	$W_i f_i$	$W_i f_i X_i$
3	24	72
5	25	125
7	36	252
9	12	108
11	8	88
	105	645

$$\overline{X_w} = \frac{\mathbf{645}}{\mathbf{105}} = 6.14$$

2. الوسط التوافقي (Harmonic mean) :

يعرف الوسط التوافقي بأنه مقلوب الوسط الحسابي لمقلوبات القيم . وصيغته الحسابية هي :

البيانات الغير مبوبة :

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}} \quad , \quad X_i \neq 0$$

اما في حالة البيانات المبوبة فان صيغته هي :

$$H = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{f_i}{X_i}}$$

حيث m هي عدد فئات التوزيع التكراري .

مثال[1] / جد الوسط التوافقي للبيانات التالية : (7 ، 8 ، 8 ، 4 ، 3 ، 5 ، 2)

الحل /

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

$$= \frac{\mathbf{7}}{\frac{\mathbf{1}}{\mathbf{2}}+\frac{\mathbf{1}}{\mathbf{5}}+\frac{\mathbf{1}}{\mathbf{3}}+\frac{\mathbf{1}}{\mathbf{4}}+\frac{\mathbf{1}}{\mathbf{8}}+\frac{\mathbf{1}}{\mathbf{8}}+\frac{\mathbf{1}}{\mathbf{7}}}$$

$$= \frac{\mathbf{7}}{\mathbf{0.5+0.2+0.33+0.25+0.13+0.13+0.14}} = \frac{\mathbf{7}}{\mathbf{1.68}} = 4.16$$

مثال[2] / اوجد الوسط التوافقي للتوزيع التكراري الاتي :

Classes	fi
50 –	8
60 –	10
70 –	16
80 –	14
90 –	10
100 –	5
110 – 120	2
	65

الحل /

$$H = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{f_i}{X_i}}$$

Xi	$\frac{f_i}{X_i}$
55	0.145
65	0.154
75	0.213
85	0.165
95	0.105
105	0.048
115	0.014
	0.847

$$H = \frac{\mathbf{65}}{\mathbf{0.847}} = 74.7$$

ملاحظة : يمكن حساب الوسط التوافقي سواء كانت اطوال الفئات متساوية او غير متساوية

مزايا وعيوب الوسط التوافقي :

مزايا الوسط التوافقي :

1- بساطة فكرته .
2- حسابه يستند الى كافة البيانات المتاحة .
3- خضوعه للعمليات الجبرية .
4- لا يتأثر كثيراً بأخطاء المعاينة .

عيوب الوسط التوافقي :

1- لا يمكن تحديد قيمته اذا كانت احدى قيم المتغير مساوية للصفر او ان احد مراكز فئات التوزيع التكراري كانت مساوية للصفر .
2- لا يمكن ايجاد قيمته في حالة التوزيعات التكرارية المفتوحة من طرف واحد او طرفين .
3- لا يمكن تعيينه هندسياً ولا يمكن حساب قيمته في حالة البيانات الوصفية .
4- لا يمكن ايجاد قيمته في حالة فقدان احدى القيم .

3. الوسط التربيعي (Quadratic mean) :

يعرف الوسط التربيعي بانه الجذر التربيعي الموجب لمتوسط مربعات قيم المتغير العشوائي X كما وان قيمة الوسط التربيعي هي قيمة موجبة ويتم حسابه استناداً الى نوع البيانات حيث ان :

في البيانات غير المبوبة :

$$\mathbf{Q} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}$$

في البيانات المبوبة :

$$\mathbf{Q} = \sqrt{\frac{\sum_{i=1}^{m} f_i \; X_i^2}{\sum_{i=1}^{m} f_i}}$$

مثال[1] / جد الوسط التربيعي للبيانات التالية : (6 ، 5 ، 4 ، 3 ، 2)

الحل /

$$\sum_{i=1}^{n} X_i^2 = (2)^2 + (3)^2 + (4)^2 + (5)^2 + (6)^2 = 90$$

$$Q = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}$$

$$= \sqrt{\frac{90}{5}} = 4.243$$

مثال[2] / جد الوسط التربيعي للجدول التكراري التالي :

Classes	f_i
10 –	2
20 –	3
30 –	4
40 –	10
50 –	36
60 – 70	14
	69

الحل /

X_i	X_i^2	$f_i X_i^2$
15	225	450
25	625	1875
35	1225	4900
45	2025	20250
55	3025	108900
65	4224	59150
		195525

$$Q = \sqrt{\frac{\sum_{i=1}^{m} f_i \quad X_i^2}{\sum_{i=1}^{m} f_i}}$$

$$Q = \sqrt{\frac{195525}{69}} = 53.23$$

ملاحظة : يمكن ايجاد الوسط التربيعي اذا كانت اطوال الفئات متساوية او غير متساوية .

مزايا وعيوب الوسط التربيعي :

المزايا :

1- بساطة فكرته .
2- خضوعه للعمليات الجبرية .
3- حسابه يستند الى كافة البيانات المتاحة .

العيوب :

1- لا يمكن ايجاد الوسط التربيعي في حالة التوزيعات التكرارية المفتوحة من طرف واحد او طرفين
2- لا يمكن حساب قيمته لبيانات متغير وصفي .
3- لا يمكن تعيينه هندسياً .
4- لا يمكن تحديد قيمته في حالة فقدان احدى قيم العينة .

5. المنوال (The Mode) (Mo)

يعرف المنوال بانه تلك القيمة التي تتكرر اكثر من غيرها من بين مجموعة القيم او انها القيمة الشائعة بين مجموعة القيم ويتم حساب المنوال استناداً الى نوع البيانات وكما يلي :

1. حساب المنوال في حالة البيانات غير المبوبة :

لتكن x_n ، ..... ، x_2 ، x_1 تتمثل قياسات عينة من المفردات قوامها n وافرض ان X_j قيمة من قيم هذه المجموعة لوحظ بانها تكررت اكثر من غيرها ، عندئذٍ وحسب تعريف المنوال فان X_j تمثل المنوال لهذه المجموعة .

مثال[1] / للبيانات التالية جد المنوال : (4 ، 8 ، 2 ، 8 ، 5 ، 2 ، 7 ، 4 ، 2 ، 3 ، 2)
الحل / نلاحظ ان العدد 2 قد تكررت اربع مرات وهو اكبر من تكرار اي عدد اخر وعليه فان المنوال لهذه المجموعة هو 2 .

مثال[2] / جد المنوال للبيانات التالية : (12 ، 10 ، 7 ، 8 ، 6 ، 3 ، 4 ، 2)
الحل / واضح من هذه المجموعة انه لا يوجد عدد متكرر اكثر من غيره ، وعليه فانه لا يوجد منوال لهذه المجموعة .
علماً انه من الممكن وجود اكثر من منوال واحد لمجموعة من البيانات في حالة تساوي التكرار .

مثال[3] / اوجد المنوال للبيانات التالية : (4 ، 6 ، 4 ، 2 ، 3 ، 2)
الحل / $Mo = 4$ ، $Mo = 2$

2. حساب المنوال في حالة البيانات المبوبة :

يعتمد حساب المنوال للبيانات المبوبة على نوع المتغير العشوائي اذا كان متقطع او مستمر :

اولاً : حساب المنوال اذا كان المتغير متقطع :

يمثل المنوال مركز الفئة المقابل لأكبر تكرار في التوزيع ، وفي حالة وجود فئتين او اكثر فان التوزيع من هذا النوع سيمتلك اكثر من قيمة واحدة للمنوال كل منها تمثل مركز تلك الفئة التي تقابل ذلك التكرار .

مثال / اوجد المنوال للجدول التكراري التالي :

Classes	f_i	
60 – 74	2	
75 – 89	6	
90 – 104	14	الفئة المنوالية
105 – 119	10	
120 – 134	8	

الحل / المنوال هو مركز الفئة المقابل لأكبر تكرار وبما انه اكبر تكرار هنا وهو 14 لذلك فإن المنوال هو

$$Mo = \frac{90+104}{2} = 97$$

ثانياً : حساب المنوال اذا كان المتغير مستمر :

اذا كان المتغير من النوع المستمر فيتم حساب المنوال وفق الصيغة التالية :

$$Mo = L_k + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} \times h_k$$

حيث ان :

L_k : هو الحد الادنى للفئة المنوالية .

f_k : تكرار الفئة .

f_{k-1} : التكرار السابق لفئة المنوال .

f_{k+1} : التكرار اللاحق لفئة المنوال .

h_k : هو طول فئة المنوال .

مثال / اوجد المنوال للتوزيع التكراري التالي :

Classes	f_i	
10 –	5	
20 –	7	
30 –	15	الفئة المنوالية
40 –	8	
50 – 60	4	

الحل /

$$Mo = L_k + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} \times h_k$$

$$= 30 + \frac{(\mathbf{15-7})}{(\mathbf{15-7})+(\mathbf{15-8})} \ (10)$$

$\therefore$ Mo = 180

مزايا وعيوب المنوال :

1- يمكن ايجاد المنوال لبيانات متغير وصفي فمثلاً لو كانت تقديرات طالب معين في مجموعة امتحانات هي متوسط ، مقبول ، متوسط ، جيد ، متوسط فان المنوال في هذه الحالة هو متوسط باعتباره قد تكرر اكثر من غيره .
2- يمكن تعيين المنوال هندسياً .
3- يمكن ايجاد المنوال في حالة التوزيعات التكرارية المفتوحة من طرف واحد او طرفين والمنوال لا يتأثر بالقيم الشاذة .

6. الوسيط (Median)

يعتبر الوسيط احد مقاييس النزعة المركزية ويعرف بانه تلك القيمة من المتغير العشوائي التي تقسم مجموعة القيم الى قسمين متساويين اي انها تلك القيمة التي تجعل عدد القيم التي قبلها مساوي الى عدد القيم التي بعدها .

طرق ايجاد الوسيط :

في حالة البيانات غير المبوبة :

لإيجاد قيمة الوسيط للبيانات الاصلية بحيث ترتب القيم ترتيباً تصاعدياً او ترتيباً تنازلياً فتكون قيمة الوسيط

أ- اذا كان عدد القيم فردي فان الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2}$

مثال / اوجد الوسيط للبيانات التالية (3 ، 4 ، 2 ، 25 ، 15 ، 12 ، 7)

الحل / نرتب القيم ترتيباً تصاعدياً :

2 , 3 , 4 , 7 , 12 , 15 , 25

$$\frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$$

∴ القيمة الرابعة هي الوسيط وهي 7 ∴ Me = 7

ب-اذا كان عدد القيم عدد زوجي فان الوسيط هو الوسط الحسابي للقيمتين اللتين ترتيبهما هو

$$\left(\frac{n}{2} , \frac{n}{2} + 1\right)$$

مثال / اوجد الوسيط للبيانات التالية : (90 ، 86 ، 44 ، 22 ، 56 ، 82)

الحل / نرتب القيم ترتيباً تصاعدياً :

22 , 44 , 56 , 82 , 86 , 90

$$\frac{n}{2} = \frac{6}{2} = 3$$

$$\frac{n}{2} + 1 = 3 + 1 = 4$$

∴ الوسيط هو الوسط الحسابي للقيميتين 82 و 56 وهو

$$\therefore Me = \frac{56+82}{2} = 69$$

مزايا الوسيط :

1. يمكن ايجاد الوسيط في حالة التوزيعات التكرارية ذات الفئات غير متساوية الاطوال.
2. يمكن ايجاد الوسيط اذا كان الجدول مفتوح من طرف واحد او من طرفين .
3. يمكن تعيينه هندسياً .
4. بساطة فكرته .
5. سهل الفهم والحساب .
6. يمكن تقدير قيمته عن طريق التخمين .
7. لا يتأثر بالقيم الشاذة .
8. يمكن تعيينه في حالة البيانات الوصفية .

عيوب الوسيط :

1. لا يخضع للعمليات الجبرية .
2. يتأثر بأخطاء المعاينة .
3. لا يستند في حسابه الى كافة البيانات المتاحة .

العلاقة بين الوسط الحسابي والوسيط والمنوال :

اذا فرضنا بان التوزيع التكراري كان متماثل او قريب من حالة التماثل عندئذٍ فان العلاقة بين هذه المقاييس تكون صحيحة اي ان :

$$\overline{X} = Me = \frac{1}{3}(\overline{X} - Mo)$$

وهذه العلاقة مهمة جداً حيث يتعذر في بعض الاحيان مثلاً حساب الوسط الحسابي من جدول تكراري مفتوح في حين يمكن ايجاد الوسيط والمنوال في مثل هذه الاحوال عندئذٍ يمكن الحصول على قيمة تقريبية للوسط الحسابي وحسب الصيغة التالية :

$$\overline{X} = \frac{3Me - Mo}{2}$$

مثال / اوجد الوسط الحسابي للجدول التكراري التالي :

Classes	fi	الحدود العليا	Fi
50 –	3	55	3
55 –	5	60	8
60 –	11	65	19
65 –	16	70	35
70 –	9	75	44
75 –	4	80	48
80 and more then	2	—	50
	50		

$$Mo = L_k + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} \times h_k$$

$$= 65 + \frac{(16-11)}{(16-11)+(16-9)} (5)$$

$\therefore Mo = 67.08$

$$Mo = \frac{50}{2} = 25$$

$19 < 25 < 35$

$$Me = L_k + \frac{h_k}{f_k} \left(\frac{\sum f_i}{2} - f_{k-1} \right) = 65 + \frac{5}{16} (25 - 19)$$

$\therefore Mo = 66.875$

$$\overline{X} \frac{3Me - Mo}{2} = \frac{(3)(66.875) - 67.08}{2} = 66.7725$$

مثال / اوجد قيمة Z التي تجعل الوسط الحسابي مساوي الى 30

20 , 25 , 28 , 31 , 37 , Z

الحل /

$$\overline{X} = \frac{\sum_{i=1}^{6} X_i}{n}$$

$$30 = \frac{141+Z}{6}$$

$$141 + Z = 180$$

$$\therefore \ Z = 180 - 141 = 39$$

الفصل الخامس

مقاييس التشتت

مفهوم التشتت والهدف من احتسابه

يعرف التشتت بانه تباعد او انتشار قيم مجموعة من المفردات عن بعضها البعض او عن قيمة معينة ثابتة (الوسط الحسابي مثلاً) .
ان الهدف من دراسة التشتت هو تكوين فكرة عن مدى تجانس قيم مجموعة من المفردات وهذا يعني ان دراسة التشتت امر مفيد في اجراء المقارنة بين قيم مجموعتين او اكثر من البيانات عن ظاهرة معينة .

انواع مقاييس التشتت :

هناك نوعين من مقاييس التشتت هما :

1. مقاييس التشتت المطلقة .
2. مقاييس التشتت النسبية

ان مقاييس التشتت المطلقة تبين درجة تجانس قيم مجموعة من البيانات بشكل مطلق وتكون مقاسة بنفس وحدات قياس المتغير العشوائي (وحدات الطول ، وحدات الوزن ، وحدات الزمن ، وحدان الكثافة ، ... الخ) وهذه المقاييس هي :

- المدى .
- الانحراف الربيعي .
- الانحراف المتوسط .
- الانحراف المعياري .

اما مقياس التشتت النسبية تبين درجة تجانس قيم مجموعة من البيانات بشكل نسبي وتكون خالية من وحدات قياس المتغير العشوائي وان هذه المقاييس هي معاملات التشتت .

مقاييس التشتت المطلقة :

1- المدى : (R) Range

يعتبر المدى ابسط انواع مقاييس التشتت المطلقة ويعرف المدى بانه الفرق ما بين اكبر قيمة في مجموعة البيانات واصغر قيمة فيها .

فاذا كانت X_n ، ..... ، X_3 ، X_2 ، X_1 تمثل قياسات عينة من المفردات قوامها n وان X_L تمثل اكبر قيمة فيها وان X_S تمثل اصغر قيمة فيها عندئذٍ فان المدى لهذه المجموعة هو

$$R = X_L - X_S$$

مثال / جد المدى للبيانات التالية (15 ، 12 ، 9 ، 10 ، 7 ، 8 ، 3 ، 5 ، 2)
الحل / اكبر قيمة في المجموعة هي 15
اصغر قيمة في المجموعة هي 2 عليه فإن :

$R = X_L - X_S$
$R = 15 - 2 = 13$

اما في حالة البيانات المبوبة في توزيع تكراري فإن المدى في هذه الحالة يمثل الفرق ما بين الحد الاعلى للفئة الاخيرة والحد الادنى للفئة الاولى .
ان هذا المقياس لا يعول عليه كثيراً في قياس التشتت لقيم مجموعة معينة نظراً لكونه يستند الى قيمتين متطرفتين فقط ويهمل بقية القيم الاخرى .

2- الانحراف الربيعي (Quartile Deviation)

يعرف الانحراف الربيعي بانه متوسط الفرق ما بين الربع الثالث والربع الاول لمجموعة من البيانات سواء كانت غير مبوبة او مبوبة في توزيع تكراري ، فاذا كان Q_1 يمثل الربع الاول و Q_3 يمثل الربع الثالث عندئذٍ فإن الانحراف الربيعي :

$$Q.D = \frac{Q_3 - Q_1}{2}$$

ويلاحظ مما تقدم انه يمكن ايجاد الانحراف الربيعي للتوزيعات التكرارية المفتوحة او المغلقة من طرف واحد او طرفين .
مثال / من جدول توزيع تكراري لأعمار عينة من تلاميذ احدى المدارس لوحظ ان قيمة الربع الاول 7.435 وقيمة الربع الثالث كانت 10.391 جد الانحراف الربيعي للعمر في هذا التوزيع .

الحل /

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$= \frac{\mathbf{10.391 - 7.435}}{\mathbf{2}} = 1.478$$

ان الانحراف الربيعي افضل من المدى كونه يستخدم 50% من البيانات المتاحة ويهمل 50% الاخرى

ان هذا المقياس هو الاخر قليل الاستخدام في التطبيق في حين انه يمتاز عن بقية مقاييس التشتت من حيث امكانية حسابه في حالة التوزيعات المفتوحة في حين يتعذر الحصول على بقية مقاييس التشتت الاخرى بالرغم من كونها افضل من الانحراف الربيعي .

3- الانحراف المتوسط (Mean Deviation)

يعرف الانحراف المتوسط بانه مجموعة الانحرافات المطلقة لقيم المتغير العشوائي عن نقطة اختيارية مثل A مقسوماً على عدد القيم وغالياً ما تختار النقطة A لان تكون احد مقاييس النزعة المركزية الثلاث ، الوسط الحسابي ، الوسيط ، المنوال .

طرق حساب الانحراف المتوسط :

1. حساب الانحراف المتوسط في حالة البيانات الغير مبوبة :

لتكن x_1 ، x_2 ، x_3 ، ..... ، x_n تمثل قياسات عينة من المفردات قوامها n وليكن A ثابت اختياري عندئذٍ يمكن حساب الانحراف المتوسط لهذه المجموعة وفق الصيغة التالية :

$$\mathbf{M.D} = \frac{\sum_{i=1}^{n} |X_i - A|}{n}$$

فاذا كانت $A = \bar{X}$ عندئذٍ

$$\mathbf{M.D} = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n}$$

فاذا كانت $A = Mo$ عندئذٍ

$$\mathbf{M.D} = \frac{\sum_{i=1}^{n} |X_i - Mo|}{n}$$

اذا كانت $A = Me$ عندئذٍ

$$\mathbf{M.D} = \frac{\sum_{i=1}^{n} |X_i - Me|}{n}$$

مثال / للبيانات التالية جد الانحراف المتوسط باستخدام $\bar{X}$ ، Mo ، Me

2 , 3 , 4 , 5 , 5 , 6 , 7 , 10 , 13 , 14 , 19

الحل /

واضح من البيانات وبعد الحساب ان قيمة $\bar{X} = 8$ و $Mo = 5$ و $Me = 6$ لذلك فان :

$$M.D = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n} = = \frac{\sum_{i=1}^{11} |X_i - 8|}{11}$$

$$= \frac{|2-8|+|3-8|+|4-8|+|5-8|+|5-8|+|6-8|+|7-8|+|10-8|+|13-8|+|14-8|+|19-8|}{11}$$

$$= \frac{6+5+4+3+3+2+1+2+5+6+11}{11} = \frac{48}{11} = 4.367$$

$$M.D = \frac{\sum_{i=1}^{n} |X_i - Mo|}{n} = = \frac{\sum_{i=1}^{11} |X_i - 5|}{11} = \frac{45}{11} = 4.091$$

$$M.D = \frac{\sum_{i=1}^{n} |X_i - Me|}{n} = \frac{\sum_{i=1}^{11} |X_i - 6|}{11} = \frac{44}{11} = 4$$

2. حساب الانحراف المتوسط في حالة البيانات المبوبة :

لتكن X_m ، ..... ، X_2 ، X_1 تمثل مراكز فئات توزيع تكراري عدد فئاته m وان f_m ، .... ، f_2 ، f_1 تمثل التكرارات المقابلة لهذه الفئات وافرض ان a ثابت اختياري عندئذٍ يمكن حساب الانحراف المتوسط لهذه الفئات وفق الصيغة التالية :

$$M.D = \frac{\sum_{i=1}^{m} f_i |X_i - A|}{\sum_{i=1}^{m} f_i}$$

فاذا كانت $A = \overline{X}$ فان الانحراف المتوسط :

$$M.D = \frac{\sum_{i=1}^{m} f_i |X_i - \overline{X}|}{\sum_{i=1}^{m} f_i}$$

واذا كانت $A = Mo$ فان الانحراف المتوسط :

$$M.D = \frac{\sum_{i=1}^{m} f_i |X_i - Mo|}{\sum_{i=1}^{m} f_i}$$

واذا كانت $A = Me$ فان الانحراف المتوسط :

$$M.D = \frac{\sum_{i=1}^{m} f_i |X_i - Me|}{\sum_{i=1}^{m} f_i}$$

مثال / الاتي توزيع تكراري لدرجات مجموعة من الطلبة في امتحان معين يطلب حساب الانحراف المتوسط باستخدام الوسط الحسابي اذا علمت ان قيمة $\overline{X} = 59.5$

Classes	fi
0 –	2
10 –	4
20 –	8
30 –	16
40 –	25
50 –	60
60 –	42
70 –	35
80 –	18
90 - 100	10
	220

الحل /

Xi	fi | Xi - $\overline{X}$ |
5	109
15	178
25	276
35	392
45	362.5
55	270
65	231
75	542.5
85	459
95	355
	3175

$$M.D = \frac{\sum_{i=1}^{m} f_i \,|\, X_i - \overline{X}|}{\sum_{i=1}^{m} f_i} = \frac{\mathbf{3175}}{\mathbf{220}} = 14.432$$

مزايا الانحراف المتوسط :

1. ان حسابه يعتمد على كافة البيانات المتاحة .
2. سهل الفهم والحساب .
3. خضوعه للعمليات الجبرية .
4. ان قيمته تكون اقل ما يمكن عند اختيار A = Me

عيوب الانحراف المتوسط :

1. اهمال الاشارات السالبة للفروق عن عملية حسابه .
2. لا يمكن حساب قيمته في حالة التوزيعات التكرارية المفتوحة من طرف واحد او طرفين .
3. لا يمكن حسابه في حالة البيانات الوصفية .
4. تتأثر قيمته في حالة وجود قيم شاذة متطرفة .
5. يتأثر بأخطاء المعاينة

4- الانحراف المعياري (Standard Deviation) :

ويسمى احياناً بالانحراف القياسي . يعتبر الانحراف المعياري احد اهم مقاييس التشتت المطلقة لما يمتاز به من ميزات مثلى ويعرف بالانحراف المعياري بانه الجذر التربيعي الموجب لمتوسط مجموع مربعات انحرافات قيم المتغير العشوائي عن وسطها الحسابي .

طرق حساب الانحراف المعياري :

1. حساب الانحراف المعياري في حالة البيانات غير المبوبة :

لتكن x_n ، ..... ، x_2 ، x_1 تمثل قياسات مفردات عينة بحجم n وليكن $\overline{X}$ الوسط الحسابي لهذه القياسات عندئذٍ فان الانحراف المعياري لهذا المجموعة هو :

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

مثال / البيانات التالية تمثل اوزان عينة من الطلبة بحجم 10طلاب . المطلوب حساب الانحراف المعياري اذا علمت ان $\overline{X} = 65$

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

الحل /

Xi	$Xi - \bar{X}$	$(Xi - \bar{X})^2$
56	-9	81
68	3	9
72	7	49
63	-2	4
65	0	0
68	3	9
71	6	36
69	4	16
62	-3	9
65	0	0
	9	213

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}} = \sqrt{\frac{\mathbf{213}}{\mathbf{10}}} = 4.615$$

ملاحظة : هناك صيغة اخرى مشتقة من الصيغة اعلاه لحساب الانحراف المعياري بعد تبسيطها ويطلق عليها الصيغة الحسابية وهي كالاتي :

الصيغة الحسابية :

$$\mathbf{S = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n} - (\frac{\sum_{i=1}^{n} X_i}{n})^2}}$$

مثال / لبيانات المثال السابق اوجد الانحراف المعياري باستخدام الصيغة الحسابية .

Xi	X_i^2
56	3136
68	4624
72	5184
63	3969
65	4225
68	4624
71	5041
69	4761
62	3844
65	4225
659	43633

$$S = \sqrt{\frac{\mathbf{43633}}{\mathbf{10}} - (\frac{\mathbf{659}}{\mathbf{10}})^2} = 4.527$$

2. حساب الانحراف المعياري في حالة البيانات المبوبة :

لتكن X_m ، ..... ، X_2 ، X_1 تمثل مراكز فئات توزيع تكراري عدد فئاته m وان f_m ، .... ، f_2 ، f_1 تمثل التكرارات المقابلة لهذه الفئات عندئذٍ وحسب التعريف فان المقياس للانحراف المعياري يكون بالشكل التالي :

$$S = \sqrt{\frac{\sum_{i=1}^{m} f_i(X_i - \bar{X})^2}{\sum_{i=1}^{m} f_i}} \quad , \quad \sum_{i=1}^{m} f_i = n$$

$$\mathbf{S} = \sqrt{\frac{\sum_{i=1}^{m} f_i X_i^2}{\sum_{i=1}^{m} f_i} - \left(\frac{\sum_{i=1}^{m} f_i X_i}{\sum_{i=1}^{m} f_i}\right)^2} \qquad \text{الصيغة الحسابية}$$

مثال / لبيانات المثال التالي اوجد الانحراف المعياري باستخدام الصيغة الثانية (الحسابية)

Classes	fi	Xi
0 –	2	5
10 –	4	15
20 –	8	25
30 –	16	35
40 –	25	45
50 –	60	55
60 –	42	65
70 –	35	75
80 –	18	85
90 – 100	10	95
	220	

الحل /

fi Xi	$f_i X_i^2$
10	50
60	900
200	5000
560	19600
1125	50625
3300	181500
2730	177450
2625	196875
1530	130050
950	90250
13090	852300

$$\mathbf{S} = \sqrt{\frac{\sum_{i=1}^{m} f_i X_i^2}{\sum_{i=1}^{m} f_i} - \left(\frac{\sum_{i=1}^{m} f_i X_i}{\sum_{i=1}^{m} f_i}\right)^2} = \sqrt{\frac{\mathbf{852300}}{\mathbf{220}} - \left(\frac{\mathbf{13090}}{\mathbf{220}}\right)^{\mathbf{2}}} = 18.271$$

مزايا الانحراف المعياري :

1. سهل الفهم والحساب .
2. حسابه يستند الى كافة البيانات المتاحة .
3. خضوعه للعمليات الجبرية .

عيوب الانحراف المعياري :

1. لا يمكن حساب قيمته في حالة التوزيعات التكرارية المفتوحة من طرف واحد او طرفين .
2. لا يمكن ايجاد قيمته في حالة البيانات الوصفية .
3. يتأثر بأخطاء المعاينة .
4. تتأثر قيمته في حالة وجود قيم شاذة او متطرفة .

العلاقة بين الانحراف المعياري والانحراف المتوسط :

في حالة التوزيعات التكرارية المتماثلة او القريبة جداً من التماثل لبيانات متغير عشوائي مستمر لوحظ وجود علاقة تربط ما بين الانحراف المعياري والانحراف المتوسط (المحسوب باستخدام الوسط الحسابي لقياسات عينة عشوائية قوامها n) .

$$M.D\ (\overline{X}) = \frac{4}{5} S_x$$

التباين (Variance) :

يعرف التباين بانه متوسط مجموع مربعات انحرافات القيم عن وسطها الحسابي اي انه مربع الانحراف المعياري ويرمز (S^2) كما ان وحدات قياس التباين تمثل مربعات وحدة قياس المتغير الاصلي فمثلاً اذا كان $S_x = 3 \Longrightarrow S^2 = 9$

ملاحظة : ان مزايا وعيوب التباين هي ذاتها مزايا وعيوب الانحراف المعياري وكذلك فان خصائص التباين هي ذاتها خصائص الانحراف المعياري من حيث المفهوم اي ان :

1- $S_x^2 \geq 0$.

2- If $Y_i = a X_i \Longrightarrow S_y^2 = a^2 S_x^2$.

3- If $Y_i = a X_i + b \Longrightarrow S_y^2 = a^2 S_x^2$

مقاييس التشتت النسبية :

1- معامل الاختلاف :

افرض ان $\overline{X}$ يمثل الوسط الحسابي ،S يمثل الانحراف المعياري لذا فان معامل الاختلاف :

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

يعتبر معامل الاختلاف افضل انواع التشتت النسبية وذلك لكونه يعتمد على افضل مقياس نزعة مركزية الا وهو $\overline{X}$ وافضل مقياس تشتت وهو الانحراف المعياري .

مثال / افترض ان متوسط درجات طلبة الصف الاول احصاء في مادة الرياضيات هي 69 درجة ، وبانحراف معياري قدره 19.3 في حين كان متوسط درجاتهم في مادة الاحصاء 75 درجة بانحراف معياري قدره 25.5 في اي من الامتحانين كان مستوى الاداء افضل من حيث التقارب ؟

الحل / نحتاج الى ايجاد معامل الاختلاف لكل امتحان :

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

$$C.V.\ (\text{stat.}) = \frac{\mathbf{25.5}}{\mathbf{75}}\ (\ 100\ ) = 34\%$$

لذا فان معامل الاختلاف في امتحان الرياضيات اقل منه في مادة الاحصاء ولهذا فان مستوى اداء الطلاب في مادة الرياضيات افضل منه في مادة الاحصاء .

2- الدرجة المعيارية :

في كثير من الاحيان نحتاج الى تحويل قيم المتغير العشوائي X الى شكل اخر يدعى بـ(الشكل المعياري) او الشكل القياسي لهذه القيم وخصوصاً في موضوعي الاحتمالات والاختيارات الاحصائية المتعلقة ببعض التوزيعات كالتوزيع الطبيعي مثلاً ، وكذلك نحتاج الى الشكل المعياري للقيم في حالات المقارنة بين مجموعتين او اكثر من القيم ذات الاوساط الحسابية والانحرافات المعيارية المختلفة اقيامها لذا فان الدرجة المعيارية (Z) لأي قيمة من قيم X تعرف على النحو التالي :

$$Z_i = \frac{X_i - \overline{X}}{S_x} \qquad , \ i = 1\ ,\ 2\ ,\ \ldots\ ,\ n$$

مثـال / الاتـي درجـات مجموعـة مـن الافـراد (عشـرة افـراد) . المطلـوب ايجـاد الدرجـة المعيارية المقابلة لهذه الدرجات ثم حقق خاصيتي الدرجة المعيارية اذا علمت ان

$S = 13.26$ ، $\bar{X} = 70.6$

Xi	Zi	Z_i^2
50	-1.55	2.4025
55	-1.17	1.3689
76	0.4	0.16
82	0.85	0.7225
60	-0.79	0.6241
69	-0.12	0.0144
62	-0.64	0.4096
95	1.84	3.3856
82	0.85	0.7225
75	0.33	0.1089
	0	10

الخاصية الاولى $\bar{Z} = 0$ $\therefore$ $\bar{Z} = \frac{\sum_{i=1}^{n} Z_i}{n} = \frac{0}{10} = 0$

الخاصية الثانية $S_z^2 = 1$ $\therefore$ $S_z^2 = \frac{\sum_{i=1}^{n} (Z_i - \bar{Z})^2}{n} = \frac{\sum_{i=1}^{n} Z_i^2}{n} = \frac{10}{10} = 1$

الفصل السادس

الارتباط والانحدار

الارتباط الخطي البسيط

ويعرف معامل الارتباط الخطي البسيط بانه القيمة العددية للعلاقة الخطية بين متغيرين ويرمز له بالرمز f_{xy} .

اما صيغة الارتباط الخطي البسيط في حالة البيانات غير المبوبة فهي :

$$f_{xy} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]\left[\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right]}}$$

قيمة الارتباط الخطي البسيط تكون محصورة بين 1- و1+ اي ان

$$-1 \leq f_{xy} \leq +1$$

عندما يكون $f_{xy} = 0$ فانه عديم الارتباط

عندما يكون $f_{xy} = +1$ فانه ارتباط قوي موجب (طردي)

عندما يكون $f_{xy} = -1$ فانه ارتباط قوي سالب (عكسي)

مثال / البيانات التالية تمثل الدخل الشهري والادخار لمجموعة من الاشخاص والمطلوب ايجاد معامل الارتباط الخطي البسيط .

الدخل	Xi	10	12	11	7	6	8	9	5	4	3
الادخار	Yi	4	5	5	3	2	3	3	1	1	0

الحل /

(الدخل)Xi	(الادخار)Yi	Xi Yi	X_i^2	Y_i^2
10	4	40	100	16
12	5	60	144	25
11	5	55	121	25
7	3	21	49	9
6	2	12	36	4
8	3	24	64	9
9	3	27	81	9
5	1	5	25	1
4	1	4	16	1
3	0	0	9	0
75	27	248	645	99

$$\overline{X} = \frac{\sum X_i}{n} = \frac{\mathbf{75}}{\mathbf{10}} = 7.5$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{\mathbf{27}}{\mathbf{10}} = 2.7$$

$$f_{xy} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right]\left[\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right]}}$$

$$= \frac{\mathbf{248-(10)(7.5)(2.7)}}{\sqrt{[\mathbf{645-(10)(7.5)^2}]\,[\mathbf{99-(10)(2.7)^2}]}} = 0.98$$

الانحدار الخطي البسيط :

يعرف الانحدار الخطي البسيط بانه عملية تقدير العلاقة الخطية بين متغيرين احدهما مستقل والاخر تابع (معتمد) من خلال تقدير معالم نموذج الانحدار حيث ان نموذج الانحدار العام هو

$$\hat{Y}_i = \hat{a} + \hat{b}\, X_i$$

نعني في هذا النموذج المتغير X هو متغير مستقل (قيمته لا تعتمد على قيمة Y) اما المتغير Y فهو متغير تابع (معتمد) لان قيمته تعتمد على قيمة X من خلال الثوابت a , b التي هي ثوابت حقيقية للمعادلة .

فمثلاً الانفاق يعتبر متغير معتمد على الدخل والذي يعتبر متغير مستقل فالعلاقة هنا ستكون علاقة طردية باتجاه واحد فكلما زاد الدخل زاد الانفاق والعكس غير صحيح فلا تعتمد قيمة زيادة الانفاق على زيادة الدخل .

في حالة انحدار Y على X (Y / X) يكون الانحدار العام هو :

$$\hat{Y}_i = \hat{a} + \hat{b}\, X_i$$

حيث :

$$\hat{a} = \overline{Y} - b\,\overline{X}$$

$$\hat{b} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

للمثال في ص[76] اوجد معادلة انحدار Y على X (Y / X)

الحل /

$\hat{Y}_i = \hat{a} + \hat{b} X_i$

$\hat{a} = \overline{Y} - b \overline{X}$

$\hat{b} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$

$\hat{b} = \frac{\mathbf{248-(10)(7.5)(2.7)}}{\mathbf{645-(10)(7.5)^2}} = \frac{\mathbf{45.5}}{\mathbf{82.5}} = 0.55$

$\hat{a} = 2.7 - (0.55) (7.5) = -1.425$

اذن معادلة الانحدار Y/X هي

$\hat{Y}_i = -1.426 + 0.55 X_i$