الفصل الرابع

مقاييس التشتت

4.1 – تعريف التشتت واستخداماته :

يعرف التشتت بأنه اختلاف قيم المفردات أو تباعدها أو انتشارها ، و يمكن التعبير عن هذا الاختلاف بطريقتين :

1 – بأنه الاختلاف بين قيم منتقاة أكبر و أصغر مفردة و هذا هو أسلوب المدى ، أو الاختلاف بين الربيع الأدنى و الأعلى حسب اسلوب الانحراف الربيعي .

2 – اختلاف القيم عن أحد مقاييس النزعة المركزية و ذلك حسب الانحراف المعياري و الانحراف المتوسط .

لنرى الآن ماهي استخدامات التشتت و كيف يستخدم التشتت في التحليل الاحصائي.

رأينا في الفصل السابق أن مقياس النزعة المركزية يساعد على مقارنة بيانات العينات أو المجتمعات و لكن هل يكفي مقياس النزعة المركزية لهذه المقارنة ؟

سنجيب على هذا السؤال من خلال المثال التالي :

لنفرض أنه لدينا عينتان صغيرتان من العمال :

الأجر الساعي لأفراد العينة الأولى :95 ، 80 ، 75 ، 70 ، 60 ليرة .

الآجر الساعي لأفراد العينة الثانية : 75 ، 75 ، 75 ، 75 ، 75 ليرة .

ان الوسط الحسابي لافراد العينة الاولى يبلغ 75 ليرة وكذلك للاجور الساعية لافراد العينة الثانية ، فهل نستطيع الحكم على مفردات العينتين بانها متساوية او مختلفة من خلال الوسط الحسابي فقط ،الجواب لا ولكن لو علمنا ان الوسط الحسابي لقيم العينة الثانية 75 ليرة واختلافها (تشتتها) يساوي الصفر والوسط الحسابي للعينة الثانية يبلغ 75 ليرة وتشتتها يبلغ قيمة معينة غير الصفر (سنرى كيفية حسابها لاحقا) فنستطيع عندها اعطاء وصفا افضل للبيانات وبالتالي مقارنة افضل . يتبين من هذا ان احد استخدامات التشتت هو استخدامه بجانب مقياس النزعة المركزية لوصف البيانات.

اذا كان التشتت معدوما كما في بيانات العينة الثانية بالمثال اعلاه فهذا يعني ان البيانات متساوية ومقياس النزعة المركزية يمثلها افضل تمثيل وعلى العكس كلما زاد

التشتت دل على اختلاف البيانات و على انخفاض القدرة التمثيلية لمقياس النزعة المركزية .

يبين اختلاف البيانات في بعض الحالات عدم استقرار الظاهرة مثل بيانات الاسعار أو بيانات عن مواصفات سلعة منتجة . . . فالتشتت هنا يدل على الاستقرار أو عدم الاستقرار في الظاهرة .

بالاضافة الى هذه الاستخدامات يستخدم التشتت بشكل واسع في اساليب التحليل الأخرى مثل الارتباط و الاستدلال كما سنرى .

يمكن استخدام مقاييس التشتت بشكل مطلق و هي بهذه الحالة تقاس بنفس وحدات البيانات ، و تستخدم بشكلها المطلق عند مقارنة بيانات لها نفس الوحدات و الفارق بين الأوساط الحسابية لكلا المجموعتين ليس كبيرا انظر المثال في الفقرة 4.2 ، أما في ما عدا ذلك فيجب استخدام مقاييس التشتت بشكل نسبي و ذلك بنسب التشتت المطلق الى مقاييس النزعة المركزية .

سندرس أدناه مقاييس التشتت التالية و بشكل مفصل :

1 – المدى.

2 –المدى الربيعي و الانحراف الربيعي .

3– الانحراف المتوسط .

4 – الانحراف المعياري .

البحث الاول

4.2 – المدى Range:

يعرف المدى المطلق بأنه الفرق بين أكبر مفردة و أصغر مفردة .

$$R = Max - Min$$

يتضح من هذا التعريف سهولة فهم هذا المقياس و سهولة حسابه و يستخدم بشكل واسع في خرائط مراقبة جودة الانتاج ، ولكن ما يؤخذ على هذا المقياس هو تركيزه على القيم الأكثر تطرفا و اهماله تشتت القيم بينهما ، كما أنه لايمكن حسابه للجداول التكرارية المفتوحة ، حيث يحسب بأنه الفرق بين الحد الاعلى الفئة الأخيرة و الحد الأدنى للفئة الاولى .

ويستخدم المدى المطلق بشكل نسبي اذا كانت وحدات البيانات مختلفة أو كان الفرق بين أوساطها كبيرا .

و يحسب المدى النسبي كما يلي :

$$PR = \frac{R}{\bar{X}} * 100$$

مثال:احسب لبيانات العينتين التاليتين المدى المطلق والنسبي وقارن النتائج:

عينة 1	2	4	5	6	15	28
عينة 2	100	116	118	120	130	136

$R_1=28-2=26$

$R_2=136-110=36$

$\bar{X} = 10$

$\bar{X} = 120$

$$PR_1 = \frac{26}{10} * 100 = 260\%$$

$$PR_2 = \frac{36}{120} * 100 = 30\%$$

نلاحظ أن المدى المطلق للعينة الأولى هو 26 وهو اصغر من المدى المطلق للعينة الثانية و البالغ 36 .أما اذا أخذنا المدى النسبي لأن الفرق كبير بين الأوساط الحسابية نرى أن تشتت العينة الثانية أصغر .

البحث الثاني

4.3 _ الانحراف الربيعي

للتخلص من أحد عيوب المدى المطلق باعتماده على القيم الأكثر تطرفا يتم انتقاء قيمتين غيرهما هما الربيع الأدنى و الربيع الأعلى و ندعو الفرق بينهما بالمدى الربيعي QR :

$$QR=Q_3-Q_1$$

و اذا ما أخذنا نصف المدى الربيعي نكون قد حصلنا على الانحراف الربيعيQD :

$$QD = \frac{QR}{2} = \frac{Q_3 - Q_1}{2}$$

و لأغراض المقارنة يمكن حساب الانحراف الربيعي النسبي و ذلك بنسب الانحراف الربيعي الى الوسيط :

$$PQD = \frac{QD}{Med} * 100$$

$$PQD = \frac{Q_3 - Q_1}{2Med} * 100$$

و اذا ما كان الوسيط مجهولا يمكن تقديره بالوسط الحسابي للربيع الأدنى و الأعلى مع العلم أن هذه

العلاقة محققة اذا كان التوزيع طبيعي $\{Med = \frac{Q_3 - Q_1}{2}\}$:

$$PQD = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100$$

يجب استخدام احدى العلاقتين لحساب الانحراف الربيعي النسبي لكلا البيانات الواجب مقارنتها .

مثال 1 :احسب الانحراف الربيعي والانحراف الربيعي النسبي لانتاج الحليب اليومي لعينة من الابقار حجمها 8 حيث كان انتاجها اليومي كما يلي(كيلو غرام):

20 25 30 31 32 35 40 42

الحل:1− نحسب الوسيط بايجاد موقعه اولا $Loc(Med)=\frac{n+1}{2}=\frac{9}{2}=4.5$ وبالتالي نكون قيمة الوسيط هي الوسط الحسابي للمفردتين الرابعة والخامسة 31.5 كيلو .

2−نوجد الربيع الادنى والاعلى بتحديد موقعهما اولا ومن ثم ايجاد قيمتيهما :

$Loc(Q1)=\frac{n}{4}+\frac{1}{2}=\frac{8}{4}+\frac{1}{2}=2.5$ وبالتالي تكون قيمة الربيع الاول عبارة عن الوسط الحسابي للمفردتين بالموقع الثاني والثالث أي:كيلو $\frac{25+30}{2}=27.5$.

$Loc(Q3)=\frac{3n}{4}+\frac{1}{2}=\frac{3*8}{4}+\frac{1}{2}=6.5$ وبالمثل تكون قيمة الربيع الاعلى هي الوسط الحسابي للمفردتين بالموقع السادس والسابع:كيلو $\frac{35+40}{2}=37.5$.

3−نحسب الانحراف الربيعي :كيلو $QD=\frac{Q3-Q1}{2}=\frac{37.5-27.5}{2}=5$.

4− نحسب الانحراف الربيعي النسبي:

$$QD=\frac{Q3-Q1}{2*Med}*100=\frac{37.5-27.5}{2*31.5}*100=15.87\%$$

مثال 2 :احسب الانحراف الربيعي والانحراف الربيعي النسبي لبيانات الجدول التالي(4.1) والذي يحوي بيانات عن المدة الزمنية التي استغرقها 100 عامل لانجاز عمل محدد :

التكرار التجميعي الصاعد F↑	عدد العمال Fi	الزمن المستغرق (ساعة)
8	8	1 -2
18	10	2 -3
33	15	3 -4
63	30	4 -5
83	20	5 -6
100	17	6 فما فوق
	100	

الجدول (4.1)

الحل : 1−لحساب الانحراف الربيعي يجب حساب الربيع الادنى والاعلى اولا :

-حساب الربيع الادنى :نحدد فئة الربيع الادنى وذلك بتدقيق عمود التكرار التجميعي الصاعد فنجد ان اول قيمة اكبر او تساوي n/4 هي مقابل الفئة الثالثة وبالتالي الفئة الثالثة هي فئة الربيع الاول ومن ثم نطبق العلاقة:

$$Q1 = L_{Q1} + \left[\frac{n/4 - CF}{F_{Q1}} \right] * C_{Q1}$$

$$Q1 = 3 + \left[\frac{100/4 - 18}{15} \right] * 1 = 3.46$$

من ثم نحدد فئة الربيع الاعلى بنفس الطريقة ولكن يجب ان نقارن التكرار الصاعد بالقيمة 3n/4 ونرى ان هذا يتحقق عند الفئة الخامسة ومنه:

$$Q3 = L_{Q3} + \left[\frac{3n/4 - CF}{F_{Q3}} \right] * C_{Q3}$$

$$Q3 = 5 + \left[\frac{300/4 - 63}{20} \right] * 1 = 5.66$$

2- نحسب الانحراف الربيعي من العلاقة :

$$QD = \frac{Q3 - Q1}{2} = \frac{5.66 - 3.46}{2} = 1.1$$

3-نحسب الانحراف الربيعي النسبي ومن اجل ذلك يجب حساب الوسيط :

$$Med = L_{Med} + \left[\frac{n/2 - CF}{F_{Med}} \right] * C_{Med}$$

$$Med = 4 + \left[\frac{100/2 - 33}{30} \right] * 1 = 4.566$$

$$PQD = \frac{Q3 - Q1}{2Med} * 100 = \frac{5.66 - 3.46}{2 * 4.566} * 100 = 22.9\%$$

مزايا الانحراف الربيعي:

1 _ يمكن حسابه اذا كان الجدول التكراري مفتوحا .

2 _ تخلص من أحد عيوب الانحراف المطلق الذي يعتمد على القيم الأكثر تطرفا و لكن لايزال هذا المقياس يهمل تشتت هذه القيم التي تم اسقاطها على طرفي التوزيع كما يهمل تشتت القيم الواقعة بين الربيع الأدنى و الأعلى.

البحث الثالث

4.4 _الانحراف المتوسط: Mean Deviation

يعرف الانحراف المتوسط بأنه الوسط الحسابي لانحراف قيم المفرادات عن أحد مقاييس النزعة الوسط الحسابي أو الوسيط مع اهمال الاشارات الجبرية . و غالبا ماتؤخذ هذه الانحرافات عن الوسيط Med ،ويتم اهمال الاشارات الجبرية كي لاتلغي هذه الانحرافات بعضها بعضا .

4.4.1 _ الانحراف المتوسط للبيانات غير المبوبة :

يحسب الانحراف المتوسط MD للبيانات غير المبوبة بالعلاقة التالية :

$$\boxed{MD = \frac{\sum |di|}{n}}$$

حيث: $di = |X_i - Med|$

ولأغراض المقارنة يجب ايجاد الانحراف المتوسط النسبي :

$$\boxed{PMD = \frac{MD}{Med} * 100}$$

مثال 1 : احسب الانحراف المتوسط و الانحراف المتوسط النسبي للنفقات الشهرية للاسر التالية:14 ، 27 ، 15 ، 18 ، 20 ، 17 ، 28 (ألف ليرة)

الحل : نرتب البيانات تصاعديا من أجل ايجاد قيمة الوسيط و من ثم نحسب الانحرافات و الانحراف المتوسط :

X_i	$di = \|X_i - Med\|$
14	4
15	3
17	1
18	0
20	2
27	9
28	10
المجموع	29

الوسيط هو المفردة الرابعة : Med = 18

$$MD = \frac{\sum |di|}{n} = \frac{29}{7} = 4.143$$

أي أن النفقات الشهرية تختلف وسطيا بمقدار 4.143 ألف ليرة عن الوسيط .

$$PMD = \frac{MD}{Med} * 100 = \frac{4.143}{18} * 100 = 23\%$$

و يفسر على أن الانحراف المتوسط يشكل ما قيمته 23 % من قيمة الوسيط .

مثال 2 :احسب الانحراف المتوسط لبيانات الجدول التالي(4.2) و الذي يحوي بيانات عن الدخل الشهري ل 100 اسرة .

فئات الدخل (الف ليرة)	F_i	X_i	F↑	$di = \lvert X_i - Med \rvert$	$\lvert d_iF_i \rvert$
10-12	3	11	3	7	21
12-14	7	13	10	5	35
14-16	10	15	20	3	30
16-18	30	17	50	1	30
18-20	35	19	85	1	35
20-22	10	21	95	3	30
22-24	5	23	100	5	25
المجموع	100			25	206

جدول (4.2)

الحل : 1 _ نحسب الوسيط:

$$Med = L_{Med} + \left[\frac{n/2 - CF}{F_{Med}} \right] * C_{Med}$$

$$Med = 16 + \left[\frac{100/2 - 20}{30} \right] * 2 = 18$$ الف ليرة

2 _ نحسب أوساط الفئات في الجدول .

3 _نحسب انحراف أوساط الفئات عن الوسيط $di = |X_i - Med|$ و من ثم نضرب بتكرار الفئة d_iF_i .

4 _نحسب الانحراف المتوسط:الف ليرة $MD = \frac{\sum |diF_i|}{\sum F_i} = \frac{206}{100} = 2.06$

5 _ نحسب الانحراف المتوسط النسبي :

$$PMD = \frac{MD}{Med} * 100 = \frac{2.06}{18} * 100 = 11.44\%$$

4.4.3 – مزايا و مساوئ الانحراف المتوسط :

1 – سهل، معناه واضح اذ هو متوسط انحراف القيم عن مقياس النزعة المركزية .

2 _ هو مفهوم وسطي تؤخذ قيم كافة المفرادات بعين الاعتبار عند حسابه .

3 _ لايمكن حسابه للجداول التكرارية المفتوحة .

4 _ ان اهمال الاشارات الجبرية عملية جبرية غير مرضية .

البحث الرابع

4.5− الانحراف المعياري Standard Deviation

4.5.1_ تمهيد :

يتشابه الانحراف المعياري مع الانحراف المتوسط بأنه متوسط انحراف القيم عن أحد مقاييس النزعة المركزية آلا و هو الوسط الحسابي حصرا بعد اجراء بعض العمليات الجبرية للتخلص من عيوب الانحراف المتوسط .تتلخص هذه الاجراءات بتربيع الانحرافات بدلا من اهمال الاشارات الجبرية فنحصل على مقياس جديد يدعى بالتباين ، فالتباين هو الوسط الحسابي لمربعات انحرافات القيم عن وسطها الحسابي و نرمز له بالرمز σ^2 للمجتمع الاحصائي و s^2 للعينة و بعد ذلك نجذر التباين للعودة الى القوة 1 فنحصل على الانحراف المعياري : فالانحراف المعياري هو جذر التباين .

و اعتمادا على هذا التعريف يمكن حساب التباين و الانحراف المعياري من العلاقات التي ستمر أدناه .

4.5.2_ الانحراف المعياري للبيانات غير المبوبة :

حسب التعريف يمكن حساب التباين و الانحراف المعياري من العلاقات التالية :

1− اذا كان حجم العينة اصغر من 30 $n < 30$

I

$$\boxed{S = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n-1}}} \qquad \boxed{S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1}}$$

2− اذا كان حجم العينة 30 او اكبر $n \geq 30$

$$S = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n}}$$

$$S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n}$$

2- اما للمجتمع فتعطى بالعلاقات التالي :

التباين : $\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$ حيث μ هي الوسط الحسابي لبيانات المجتمع.

الانحراف المعياري: $\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$

مثال: احسب الانحراف المعياري للقيم التالية:

Xi	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
3	-5	25
7	-1	1
8	0	0
10	2	4
12	4	16
40		46

الحل :

1 – نحسب الوسط الحسابي: $\bar{X} = \frac{\sum X_i}{n} = \frac{40}{5} = 8$

2 – نحسب انحراف القيم عن الوسط الحسابي : $(X_i - \bar{X})$

3 – نربع الانحرافات: $(X_i - \bar{X})^2$

4 – نحسب الانحراف المعياري من العلاقة:

$$\mathbf{S} = \sqrt{\frac{\mathbf{\Sigma(X_i} - \bar{X})^2}{\mathbf{n-1}}} = \sqrt{\frac{46}{4}} = 3.39$$

4.5.3 –الانحراف المعياري للبيانات المبوبة :

يمكن حساب الانحراف المعياري للبيانات المبوبة بالأساليب التالية :

1 – الاسلوب المباشر :

من علاقة حساب الانحراف المعياري للبيانات غير المبوبة بالاسلوب المباشر يمكن استنتاج العلاقات التالية و ذلك بأخذ التكرار F_i بعين الاعتبار :

I

$$S = \sqrt{\frac{\sum F_i(X_i - \bar{X})^2}{\sum F_i}}$$

$$S^2 = \frac{\sum F_i(X_i - \bar{X})^2}{\sum F_i}$$

مثال : أوجد الانحراف المعياري لبيانات الجدول(4.3) .

في الجدول بيانات عن عدد الأيام التي قضاها افراد عينة عشوائية مؤلفة من 100 سائح في أحد الفنادق.

الحل باستخدام العلاقة (انظر الجدول):

1 – نحسب متوسط مدة اقامة السائح: $\bar{X} = \frac{\sum F_i X_i}{\sum F_i} = \frac{846}{100} = 8.46$.

2 – نحسب الانحرافات المربعة لأوساط الفئات عن الوسط الحسابي $(X_i - \bar{X})^2$.

3 _ نضرب مربع الانحرافات لكل فئة بالتكرار F_i.

مدة الاقامة (يوم)	عدد السواح Fi	Xi	FiXi	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$F_i(X_i - \bar{X})^2$
1-3	5	2	10	-6.46	41.7316	208.65
3-5	7	4	28	-4.46	19.8916	139.24
5-7	13	6	78	-2.46	6.0516	78.670
7-9	35	8	280	-0.46	0.2116	7.408
9-11	20	10	200	1.54	2.3716	47.432
11-13	15	12	180	3.54	12.5316	187.974
13-15	5	14	70	5.45	30.6916	153.458
المجموع	100		846		113.4812	822.82

جدول (4.3)

4 _ نحسب الانحراف المعياري من العلاقة :

$$S = \sqrt{\frac{\sum F_i (X_i - \bar{X})^2}{\sum F_i}} = \sqrt{\frac{822.84}{100}} = 2.8685 \text{ يوم}$$

4.5.3 _ معامل الاختلاف (الانحراف المعياري النسبي) Coefficient Of Variation:

عند مقارنة تشتت البيانات لعينات مختلفة أو لمجتمعات مختلفة رأينا أن هناك حالات يجب استخدام التشتت بشكل نسبي . ننسب الانحراف المعياري الى الوسط الحسابي و نضرب بمائه للحصول على الانحراف المعياري النسبي وهو معامل الاختلاف :C.V

$$\boxed{C.V = \frac{S}{\bar{X}} * 100}$$

بهذا الشكل نحصل على التشتت بشكل مئوي .

مثال:نحسب معامل الاختلاف للمثال السابق:

$$C.V = \frac{S}{\bar{X}} * 100 = \frac{2.8685}{8.46} * 100 = 33.9\%$$

4.5.4 _خصائص و سمات الانحراف المعياري :

1 –هو مقياس محدد جبريا بدقة و يتمتع ببعض المزايا الجبرية تساعد على استخدامه في مقياييس أخرى .

2 – لايمكن حسابه للجداول التكرارية المفتوحة .

3 –هو مفهوم مجرد صعب الفهم .

4– اذا أضفنا أو طرحنا عدد ثابت c الى كافة القيم Xi فان قيمة الانحراف المعياري تبقى نفسها.

5–اذا ضربنا كافة القيم بعدد ثابت cفان التباين الناتج هو السابق مضروبا بـ C^2 و الانحراف المعياري هو السابق مضروبا ب c .

مسائل و تمارين الفصل الرابع

1 – تبين من عينة عشوائية حجمها 10 سواح تم سحبها بشكل عشوائي من أحد الفنادق ،البيانات التالية عن عدد الايام التي قضاها كل منهم في المدينة :

10 7 6 15 14 10 8 3 5 2

المطلوب :1–احسب المدى .

2–احسب الانحراف المتوسط والانحراف المتوسط النسبي.

2- بغية التنبؤ بنتيجة مباراة مقبلة بكرة السلة بين الفريقين A ,B تم اخذ اطوال اللاعبين في كلا الفريقين فكانت:

اطوال الفريق A	186	190	195	189	170	178	198
اطوال الفريق B	195	196	200	205	195	198	197

المطلوب :1-احسب متوسط الطول لكلا الفريقين وفسر النتائج.

2 - لاعبي أي من الفريقين اكثر تجانسا من حيث الطول برر ذلك بالحسابات اللازمة.

3- في دراسة من أجل وضع سعر موحد لقطع الصابون قامت وزارة التموين بأخذ عينة مؤلفة من 100 معمل للصابون و تم جمع البيانات التالية عن التكلفة :

عدد العمال	تكلفة القطعة (وحدة نقدية)
10	2-3
50	3-4
20	4-5
8	5-6
7	6-7
5	7-8
100	المجموع

المطلوب :1 - احسب متوسط قطعة الصابون من خلال بيانات العينة .

2 - احسب الانحراف المعياري مستخدما كل الطرق الممكنة .

3 - تبين من بيانات العينة أيضا أن متوسط استهلاك أحد المواد الداخلة في انتاج القطعة هو 10 غ بانحراف معياري 4 غرام فأي البيانات أكثر تشتتا بيانات الكلفة أم بيانات المادة المستهلكة في الانتاج .

4-لدينا البيانات التالية عن اطوال واوزان 10 راقصين في احدى الفرق الشعبية :

الطول	170	175	169	180	185	183	178	181	179	183
الوزن	70	69	71	72	68	68	72	69	72	70

المطلوب:1-احسب الانحراف المعياري للطول باستخدام اسلوب الانحرافات.

2- احسب الانحراف المعياري للطول باستخدام الاسلوب المباشر.

3-بين باي الصفات عناصر هذه الفرقة اكثر تجانسا.

5- من عينة عشوائية بحجم 200 قطعة أرض متساوية المساحة أخذت من مناطق مختلفة تم جمع البيانات التالية عن غلة القطعة من القمح :

الغلة (كغ)	عدد القطع
10-20	10
20-30	25
30-40	30
40-50	60
50-60	40
60-70	25
70فما فوق	10
المجموع	200

المطلوب 1- احسب وسيط الغلة و فسره .

2 - ماهو مقياس التشتت المناسب لبيانات هذه العينة ثم احسبه.

3- تبين من عينة أخرى مأخوذة من محافظة أخرى أن وسيط الغلة لقطع أرض مماثلة هو 50كغ بانحراف ربيعي نسبي 20% .فبيانات أي العينتين أكثر تجانسا.