احتمالات واحصاء

المحاضرة ٥

الفرقة الثالثه

شعب عامه

مقاييس التشتت ١

مقاييس التشتت

مقدمة:

في الباب السابق استخدمنا المتوسطات مثل الوسط الحسابي – الوسيط – المنوال وغيرها كمقاييس للنزعة المركزيه لمجموعه من البيانات وكان الغرض من هذه المتوسطات هو تمثيل مجموعه البيانات بطريقه بسيطه ودقيقه علي اساس انها قيم نموذجيه تمثل مجموعه البيانات الا ان هذا التمثيل لا يكون كاملا من بعض الاحيان ومثال ذلك:

مثال:

اذا كانت درجات مجموعتين A,B من الطلبه في احد الاختبارات علي النحو التالي:–

$$A: 40,43,44,45,48 \quad , \quad B: 20,31,42,55,72$$

فان:–

$$\bar{x}_A = \frac{40+43+44+45+48}{5} = \frac{220}{5} = 44$$

$$\bar{x}_B = \frac{20+31+42+55+72}{5} = \frac{220}{5} = 44$$

من هذا المثال نلاحظ ان :–

١– مجموعتي البيانات مختلفتان ولكن لهما نفس الوسط الحسابي

٢– بيانات المجموعه الاولي متقاربه فيما بينها بينما مجموعه البيانات الثانيه متباعده فيما بينها . لذلك عند المقارنه بين المجموعتين لا يجب الاكتفاء بدراسه الوسط الحسابي فقط ولكن يجب بالاضافه مراعاه تشتت قيم كلاً من المجموعتين

وتباعدها عن بعضهما

أهم مقاييس التشتت:

١- المدي
٢- نصف المدي الربيعي
٣- التباين والانحراف المعياري
٤- الانحراف المتوسط
٥- معامل الاختلاف والقيم العيارية

١- المدي

سبق تعريفه في حاله البيانات المفرده وهو عباره عن الفرق بين اكبر قيمه وأصغر قيمه في مجموعه البيانات المعطاه.
بالنسبه للبيانات المبوبه فإن المدي هو الفرق بين اصغر فئه والفئه التي تلي اخر فئه في الجدول.

٢- نصف المدي الربيعي

تعتمد فكره هذا المقياس علي التخلص من القيم المتطرفه سواء المتناهيه في الصغر والمتناهيه في الكبر علي النحو التالي :-

١- ترتيب البيانات ترتيباً تصاعدياً (وليكن عددها n)
٢- نرمز بالرمز Q_1 للربيع الادني وهو يمثل الحد الذي ترتيبه $\frac{n}{4}$
٣- نرمز بالرمز Q_3 للربيع الاعلي وهو يمثل الحد الذي ترتيبه $\frac{3n}{4}$
٤- نرمز ب الرمز Q الي نصف المدي الربيعي والذي يعرف بالعلاقه الرياضيه

التاليه:

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

إيجاد نصف المدي الربيعي للبيانات المفرده:

نفرض أن لدينا مجموعة من البيانات المفرده علي النحو التالي $x_1,x_2,x_3,...,x_n$ حيث n عدد البيانات فاذا كانت:

أ- n تقبل القسمة علي 4 فاننا نحصل علي Q_1,Q_3 مباشرة

ب- n لا يقبل القسمة علي 4 فان قيمة Q_1 تساوي الوسط الحسابي للحدين المتتاليين واللذان يحصران بينهما رتبه Q_1 وهي $\frac{n}{4}$ وبالمثل يمكن حساب Q_3

ت- نحسب Q مباشرة وذلك باستخدام الخطوات السابقة

مثال:

احسب نصف المدي الربيعي للبيانات التالية:

a- 8,12,5,6,4,10,13,11
b- 12,4,1,8,5,9,11,10,13,6

الحل:

أولا نرتب البيانات ترتيبا تصاعديا:

a- 4,5,6,8,10,11,12,13

$$\because n=8$$

رتبه $Q_1 = \frac{n}{4}=\frac{8}{4}=2$

$$\therefore Q_1=X_2=5$$

رتبه $Q_3 = \frac{3n}{4} = \frac{24}{4} = 6$

$$\therefore Q_3 = X_6 = 11$$

$$\therefore Q = \frac{Q_3 - Q_1}{2} = \frac{11-5}{2} = 3$$

b- 1,4,5,6,8,9,10,11,12,13

$\because n = 10$

رتبه $Q_1 = \frac{n}{4} = \frac{10}{4} = 2.25$

$$\therefore Q_1 = \frac{X_2 + X_3}{2} = \frac{4+5}{2} = 4.5$$

رتبه $Q_3 = \frac{3n}{4} = 30 = 7.5$

$$\therefore Q_3 = \frac{X_7 + X_8}{2} = \frac{10+11}{2} = 10.5$$

$$\therefore Q = \frac{Q_3 - Q_1}{2} = \frac{10.5-4.5}{2} = 3$$

نصف المدي الربيعي للبيانات المبوبه:

لحساب نصف المدي الربيعي في هذه الحالة نتبع الخطوات التالية:-

١- نكون الجدول التكراري المتجمع الصاعد أو الهابط.

٢- نحدد ترتيب (موقع) كلا من Q_1, Q_3 علي أنهما $\frac{n}{4}, \frac{3n}{4}$ حيث $n = \sum f_i$

٣- نعين فئة كلا من Q_1, Q_3

٤- نحسب Q_1 من خلال العلاقة الرياضية التالية:

$$Q_1 = A + \left(\frac{\frac{n}{4} - f_1}{f_2 - f_1} \right) . H$$

حيث:

A : بداية فئة الربيع الأدني Q_1

$\frac{n}{4}$: رتبه Q_1

f_1 : التكرار المتجمع الصاعد السابق لفئة Q_1

f_2 : التكرار المتجمع الصاعد اللاحق لفئة Q_1

H : طول الفئة

٥- نحسب Q_3 من خلال العلاقة الرياضية التالية:

$$Q_3 = A + \left(\frac{\frac{3n}{4} - f_1}{f_2 - f_1} \right) . H$$

حيث نفس التعريفات السابقة ولكن بالنسبة ل Q_3

٦- نحسب Q من خلال العلاقة الرياضية التالية:

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

مثال:

احسب نصف المدي الربيعي للبيانات التالية:

x_i	5-	9-	13-	17-	21-	25-	29-	33-	37-
f_i	3	4	8	9	15	8	8	6	3

الحل:

أولا : الجدول التكراري المتجمع الصاعد

x_i (فئات)	f_i (تكرار)	الحدود الفعلية للفئات	أقل من الحد الاعلي للفئة	تكرار متجمع صاعد
5-8	3	4.5-8.5	أقل من 8.5	3
9-12	4	8.5-12.5	أقل من 12.5	7
13-16	8	12.5-16.5	أقل من 16.5	15
17-20	9	16.5-20.5	أقل من 20.5	24
21-24	15	20.5-24.5	أقل من 24.5	39
25-28	8	24.5-28.5	أقل من 28.5	47
29-32	8	28.5-32.5	أقل من 32.5	55
33-36	6	32.5-36.5	أقل من 36.5	61
37-40	3	36.5-40.5	أقل من 40.5	64

$$n = \sum f_i = 64$$

لايجاد الربيع الاول Q_1 نستخدم العلاقة:

$$\therefore Q_1 = A + \left(\frac{\frac{n}{4} - f_1}{f_2 - f_1} \right) . H$$

موقع الربيع الاول $= \frac{n}{4} = \frac{64}{4} = 16$

$A = 16.5 \quad , \quad f_1 = 15 \quad , \quad f_2 = 24 \quad , \quad H = 4$

$$Q_1 = A + \left(\frac{\frac{n}{4} - f_1}{f_2 - f_1} \right) . H$$

$$= 16.5 + \left(\frac{16 - 15}{24 - 15} \right) . 4 = 16.5 + \frac{4}{9}$$

$$= 16.5 + 0.44 = 16.94$$

لايجاد الربيع الثالث Q_3 نستخدم العلاقة:

$$\therefore Q_3 = A + \left(\frac{\frac{3n}{4} - f_1}{f_2 - f_1} \right) . H$$

موقع الربيع الثالث $= \frac{3n}{4} = \frac{(3)(64)}{4} = 48$

$A = 28.5 \quad , \quad f_1 = 47 \quad , \quad f_2 = 55 \quad , \quad H = 4$

$$Q_3 = A + \left(\frac{\frac{3n}{4} - f_1}{f_2 - f_1} \right) . H$$

$$= 28.5 + \left(\frac{48 - 47}{55 - 47} \right) . 4 = 28.5 + \frac{4}{8}$$

$$= 28.5 + 0.5 = 29$$

$$\therefore Q = \frac{Q_3 - Q_1}{2} = \frac{29 - 16.94}{2} = \frac{12.06}{2} = 6.03$$

مميزات نصف المدي الربيعي:

١- سهل في فهمه.

٢- يصلح للتغلب علي القيم المتطرفه.

٣- أدق من المدي.

عيوب نصف المدي الربيعي:

١- لا يأخذ جميع البيانات عند حسابه.

٢- لا يسهل التعامل معه في التحاليل الاحصائي.

٣- التباين والانحراف المعياري

يعتبر التباين من أهم مقاييس التشتت ويعرف تباين مجموعة من القيم بأنه متوسط مجموع مربعات انحرافات هذه القيم عن وسطها الحسابي ويرمز له عاده بالرمز S^2 ويعرف الانحراف المعياري والذي يرمز له بالرمز S بأنه الجذر التربيعي للتباين.

التباين والانحراف المعياري للبيانات المفرده:

نفرض أن لدينا مجموعة من البيانات المفرده علي النحو التالي $x_1, x_2, x_3, ..., x_n$ حيث n عدد البيانات فيعرف التباين لهذه البيانات بالعلاقة التالية:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad n \geq 30$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad n < 30$$

ويعرف الانحراف المعياري :

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad n \geq 30$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad n < 30$$

مثال:

احسب التباين والانحراف المعياري للبيانات التالية:
7,5,9,7,8,6

الحل:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{7+5+9+7+8+6}{6} = 7$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad n < 30$$

$$S^2 = \frac{1}{5}\left[(7-7)^2 + (5-7)^2 + (9-7)^2 + (7-7)^2 + (8-7)^2 + (6-7)^2\right]$$

$$S^2 = \frac{0+4+4+0+1+1}{5} = \frac{10}{5} = 2$$

$$\therefore S = \sqrt{2}$$

ملاحظة:

بفرض وجود عينتين حجمهما n_1, n_2 وتباينهما S_1^2, S_2^2 علي الترتيب وكان لهما المتوسط نفسه فإن تباين العينة المكونة من دمج العينتين معا $\bar{x}$ يكون:

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

مثال:

اذا كان لدينا عينتين وأعطت النتائج التالية:

	I	II
حجم العينة	4	6
الوسط الحسابي	5	5
التباين	3	3.5

احسب التباين والانحراف المعياري الناتج من دمج العينتين.

الحل:

$$S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{(3)(3)+(5)(3.5)}{4+6-2} = \frac{26.5}{8} = 3.31$$

$$S = \sqrt{3.31}$$

اذا عدلت بيانات المجموعة الأولي حسب المعادلة $y = 2x-3$ فإن:

الوسط الحسابي بعد التعديل:

$$\bar{y} = 2\bar{x}-3 = 2(5)-3 = 7$$

التباين بعد التعديل:

$$S_y^2 = 4S_x^2 = 4(3) = 12$$

الانحراف المعياري بعد التعديل:

$$S_y = 2S_x = 2\sqrt{3} = 3.46$$

التباين والانحراف المعياري للبيانات المبوبة:

يعرف التباين في هذه الحالة بالعلاقة الرياضية التالية:

$$S^2 = \frac{1}{\sum f_i} \sum f_i (x_i^* - \bar{x})^2$$

حيث x_i^* مركز الفئة

$$\bar{x} = \frac{1}{\sum f_i} \sum f_i x_i^*$$

مثال:

احسب الانحراف المعياري للبيانات التالية:

x_i	5-	7-	9-	11-	13-	15-	17-
f_i	8	14	23	35	25	12	4

الحل:

x_i	f_i	x_i^*	$f_i x_i^*$	$(x_i^* - \bar{x})^2$	$f_i(x_i^* - \bar{x})^2$
5-6	8	5.5	44	30.25	242
7-8	14	7.5	105	12.25	171.5
9-10	23	9.5	218.5	2.25	51.75
11-12	35	11.5	402.5	0.25	8.75
13-14	25	13.5	337.5	6.25	156.25
15-16	12	15.5	186	20.25	243
17-18	4	17.5	70	42.25	169
	121		1363.5		1042.25

$$\bar{x} = \frac{1}{\sum f_i} \sum f_i x_i^* = \frac{1363.5}{121} = 11.3 \approx 11$$

$$S^2 = \frac{1}{\sum f_i} \sum f_i (x_i^* - \bar{x})^2$$

$$S^2 = \frac{1042.25}{121} = 8.61$$

$$S = \sqrt{8.61}$$

مميزات التباين والانحراف المعياري:

١- من أهم المقاييس في تعيين درجه التشتت.

٢- يتناول جميع القيم لمجموعة البيانات.

٣- أدق مقاييس التشتت المعروفه.

عيوب التباين والانحراف المعياري:

١- يتأثر بالقيم المتطرفة (الشاذه).

٢- يصعب حسابه في حالة البيانات الوصفية والبيانات الكميه ذات الجداول المفتوحه.

٤- الانحراف المتوسط

تعريف الانحراف المتوسط:

هو أحد مقاييس التشتت وهو عباره عن مجموع انحرافات البيانات عن وسطها الحسابي أي $\sum_{i=1}^{n}(x-\bar{x})$ وحيث أن هذا المجموع يساوي صفر دائما لأن مجموع الانحرافات الموجبة عن وسطها الحسابي= مجموع الانحرافات السالبة عن وسطها الحسابي ولكي يكون لهذا المجموع معني لابد من أخذ القيمة المطلقة لهذه الفروق.

الانحراف المتوسط للبيانات المفرده:

نفرض أن لدينا مجموعة من البيانات المفرده علي النحو التالي $x_1, x_2, x_3, \ldots, x_n$ حيث n عدد البيانات فيعرف الانحراف المتوسط لهذه البيانات بالعلاقة التالية:

$$M.D. = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right|$$

حيث $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

مثال:

احسب الانحراف المتوسط للبيانات: 4,5,6,7,8

الحل:

$$\bar{x} = \frac{4+5+6+7+8}{5} = 6$$

$$M.D. = \frac{1}{5}\left[\left|4-6\right| + \left|5-6\right| + \left|6-6\right| + \left|7-6\right| + \left|8-6\right|\right]$$

$$= \frac{1}{5}\left[2+1+0+1+2\right] = 1.2$$

الانحراف المتوسط للبيانات المبوبه:

يعرف الانحراف المتوسط في هذه الحالة بالعلاقة الرياضية التالية:

$$M.D. = \frac{1}{\sum f_i}\sum_{i=1}^{n}\left|x_i^* - \bar{x}\right| f_i$$

حيث x_i^* مركز الفئة

$$\bar{x} = \frac{1}{\sum f_i}\sum f_i x_i^*$$

مثال:

احسب الانحراف المتوسط للبيانات التالية:

x_i	2-	4-	6-	8-	10-	12-
f_i	8	5	4	3	2	8

الحل:

x_i	f_i	x_i^*	$f_i x_i^*$	$\left\lvert x_i^* - \bar{x} \right\rvert$	$f_i \left\lvert x_i^* - \bar{x} \right\rvert$
2-3	8	2.5	20	4.7	37.6
4-5	5	4.5	22.5	2.7	13.5
6-7	4	6.5	26	0.7	2.8
8-9	3	8.5	25.5	1.3	3.9
10-11	2	10.5	21	3.3	6.6
12-13	8	12.5	100	5.3	42.4
	30		215		106.8

$$\bar{x} = \frac{1}{\sum f_i} \sum f_i x_i^* = \frac{215}{30} = 7.2$$

$$M.D. = \frac{1}{\sum f_i} \sum_{i=1}^{n} \left| x_i^* - \bar{x} \right| f_i$$

$$M.D. = \frac{106.8}{30} = 3.56$$

٥- معامل الاختلاف

مقدمة:

عند مقارنة تشتت مجموعتين مختلفتين من البيانات فإنه يجب أن تتم علي أساس مقياس نسبي لا يعتمد علي وحدات المقياس في كلا من المجموتين لذلك نلجأ الي الجمع بين مقياس الموضع ومفاييس التشتت بدمجهما سويا لنحصل علي نسبة مئوية تصلح لغرض المقارنة وتسمي هذه النسبة بمعامل الاختلاف والذي يرمز له V بالرمز ويعرف بالعلاقة التالية:

$$V_x = \frac{S_x}{\bar{x}}.100$$

ويسمي معامل الاختلاف أيضا بالتشتت النسبي.

مثال:

مصنع لانتاج المكيفات ينتج نوعين منها هما A, B وكان متوسط أعمار النوعين $\bar{x}_A = 1495\,h\,, \bar{x}_B = 1875\,h$ والانحراف المعياري المناظر $S_A = 280\,h\,, S_B = 310\,h$ فأي من النوعين يكون له تشتت نسبي أكبر

الحل:

$$V_A = \frac{S_A}{\bar{x}_A}.100 = \frac{280}{1495}.100 = 18.7\%$$

$$V_B = \frac{S_B}{\bar{x}_B}.100 = \frac{310}{1875}.100 = 16.5\%$$

$$\because V_A > V_B$$

$\therefore$ النوع الأول يكون له تشتت نسبي أكبر (أو النوع الثاني يكون أكثر تجانسا من النوع الأول)

٦- القيم العيارية

مقدمة:

يكون من المطلوب أحيانا المقارنة بين قيمتين في مجموعتين مختلفتين من البيانات فمثلا إذا إفترضنا أن طالبا ما حصل علي درجتين في اختبارين لمقررين مختلفين ويكون مطلوبا منا أن تقرر في أي من المقررين يكون الطالب أكثر استيعابا فيه عن الأخر.

لإجراء مثل هذه المقارنة بطريقة صحيحه فإنه لا يمكن المقارنة علي أساس هاتين الدرجتين فقط بدون الأخذ في الاعتبار متوسط الدرجات والانحراف المعياري لدرجات كل من المقررين علي حده.

بفرض أن لدينا x من القيم فإنه يمكن أن تعرف الدرجة العيارية والتي يرمز لها بالرمز Z_X والتي تعرف بالعلاقة:

$$Z_X = \frac{x-\bar{x}}{S_X}$$

حيث $\bar{x}$ هو متوسط قيم x و S_X الانحراف المعياري لقيم x و x هي القيمة المراد ايجاد الدرجة العيارية لها.

مثال:

في الاختبار النهائي في مقرر الاحصاء كان متوسط الدرجات لمجموعة من 150 طالب هو 78 درجة والانحراف المعياري 8 درجات وفي مقرر الجبر كان متوسط الدرجات لنفس المجموعة هو 73 درجة والانحراف المعياري هو 7.6 درجة لأي من المقررين يكون التشتت النسبي لدرجاته أكبر. وإذا حصل الطالب علي 75 درجة في الاحصاء و 71 درجة في الجبر فأي من المقررين يكون الطالب أكثر استيعابا فيه عن الأخر.

الحل:

نرمز للاحصاء بالرمز x وللجبر بالرمز y أي أن:

$$\bar{x}=78 \qquad \bar{y}=73$$
$$S_x=8 \qquad S_y=7.6$$
$$x=75 \qquad y=71$$

$$\because V_x=\frac{S_x}{\bar{x}}.100=\frac{8}{78}.100=10.3\%$$

$$\because V_y=\frac{S_y}{\bar{y}}.100=\frac{7.6}{73}.100=10.4\%$$

$$\because V_y>V_x$$

∴مادة الجبر لها تشتت نسبي أكبر.

الأن نحسب الدرجة العيارية لكل من الجبر والإحصاء:

$$Z_x=\frac{x-\bar{x}}{S_x}=\frac{75-78}{8}=\frac{-3}{8}=\frac{-57}{152}$$

$$Z_y=\frac{y-\bar{y}}{S_y}=\frac{71-73}{7.6}=\frac{-2}{7.6}=\frac{-40}{152}$$

$$\because Z_y>Z_x$$

∴الطالب يكون أكثر استيعابا في مادة الجبر عنه في مادة الإحصاء.

تمرين:

١- أخذت عينتان من مجتمعين وأعطتا النتائج التالية:

I	II
$\sum_{i=1}^{50} x_i=300$	$\sum_{j=1}^{40} y_j=280$

$$\sum_{i=1}^{50} x_i^2 = 1950 \qquad \sum_{j=1}^{40} y_j^2 = 2100$$

احسب الوسط الحسابي لكل عينة – التباين والانحراف المعياري لكل عينة – الوسط الحسابي للعينتين معا – التباين الناتج من دمج العينتين معا – أي العينتين أكثر تجانس

٢- الجدول التالي يمثل نتائج امتحان شعبتين لطلاب الكليه في مقرر ما:

الشعبة	عدد الطلاب	المتوسط	الوسيط	المنوال	التباين
A	35	70	68	m	25
B	40	67	M	63	30

أ- احسب الوسط الحسابي والانحراف المعياري لنتائج الشعبتين معا

ب- مجموع درجات الشعبه الاولي

ت- احسب الوسيط والمنوال للشعبتين

ث- اذا عدلت درجات الشعبه الأولي حسب المعادلة $y = 1.2x - 20$ ودرجات الشعبه الثانيه حسب المعادلة $z = 0.8x + 20$ احسب الوسط الحسابي والوسيط والمنوال والانحراف المعياري للشعبتين بعد التعديل

٣- اذا كان لدينا مجموعتان X, Y من البيانات للظاهرتين علي النحو التالي:

$$\sum_{i=1}^{40} X_i = 200, \sum_{i=1}^{40} X_i^2 = 2960, \sum_{j=1}^{30} Y_j = 210, \sum_{j=1}^{30} Y_j^2 = 3390$$

دمجت العينتان معا فاحسب الوسط الحسابي والانحراف المعياري بعد الدمج – أي من المجموعتين لها تشتت نسبي أكبر

٤- حصل طالب علي الدرجه 84 في الامتحان النهائي لكرة السلة حيث كان متوسط الدرجات 76 درجه وانحرافها المعياري 10 درجات وفي الامتحان النهائي لكرة اليد كان متوسط الدرجات 82 درجه وانحرافها المعياري 16 درجه حصل الطالب علي الدرجه 90 أي المقررين له تشتت نسبي أكبر وكذلك أي المقررين يكون الطالب أكثر استيعابا فيه عن الاخر

٥- في تحليل للاجور الشهرية لشركتين من الشركات الخاصة وجدنا التالي:

الشركة الثانية	الشركة الاولي	
600	300	عدد العمال
64	36	تباينات الاجور
250	220	متوسط الاجور الشهرية

أ- أي من الشركتين تدفع مرتبات أكثر

ب- أي من الشركتين أكثر تجانس في اجور عمالها

ت- دمجت الشركتان معا فاحسب الوسط الحسابي والانحراف المعياري بعد الدمج