المحاضره الثامنه " الجزء الاول "

المقاييس الإحصائية للبيانات المبوبة

يقصد بالبيانات المبوبة تلك البيانت التى تم وضعها فى صورة جداول تكرارية.

والجداول التكرارية للمتغير الكمى المتقطع يمكن تحويلها لتكون بيانات غير مبوبة و نتعامل معها كما سبق توضيح ذلك في المحاضرة السابقة، الا أن الأمر يختلف بالنسبة للمتغير الكمى المتصل حيث يصعب ذلك ولابد من التعامل معها كما هي على صورتها الجدولية وهذا ما سوف نتناوله فى هذه المحاضرة إن شاء الله

وسيتم عرض لكيفية حساب كلا من مقاييس النزعة المركزية ومقاييس التشتت فى ثلاث حالات للجداول التكرارية وهى :

- الجداول المنتظمة
- الجداول غبر المنتظمة
- الجداول المفتوحة

الجداول المنتظمة:

- وهى تلك الجداول التى تكون فيها أطوال الفئات جميعها متساوية .

اولا- الوسط الحسابي والتشتت حولة:

الوسط الحسابى كما سبق أن تم تعريفة فى الفصل السابق هو القيمة التى إذا أخذها جميع المفردات لكان مجموعها يساوى مجموع القيم الأصلية، ويمكن حساب الوسط الحسابى او المتوسط الحسابى كما يلي:

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

$\bar{x}$ الوسط الحسابي

x_i مركز الفئة i وهى تساوى (الحد الأعلى للفئة + الحد الأدنى للفئة) ÷ 2

f_i تكرار الفئة i

l عدد الفئات

ويتم حساب التشتت حول المتوسط الحسابى من خلال الآتي:

أ- متوسط الانحرافات المطلقة *AAD*:

وهو يقيس إنحراف القيم عن وسطها الحسابى بغض النظر عن اشارة ذلك الانحراف حيث يتم حسابه من خلال المعادلة التالية :

$$AAD = \frac{\sum_{i=1}^{l} |x_i - \bar{x}| f_i}{\sum_{i=1}^{l} f_i}$$

ب – التباين σ^2 :

وهو متوسط مجموع مربع إنحرافات القيم عن وسطها الحسابى. ويتم حسابه من خلال المعادلة التالية:

$$\sigma^2 = \frac{\sum_{i=1}^{l} (x_i - \bar{x})^2 f_i}{\sum_{i=1}^{l} f_i}$$

جـ - الانحراف المعيارى σ :

هو الجذر التربيعى للتباين، ويمكن حسابه من خلال المعادلة التالية:

$$\sigma = \sqrt{\sigma^2}$$

مثال:

البيانات التالية توضح توزيع مجموعة من المدرسين العاملين في مجال التربية وفقا لفئات أعمارهم فكانت النتائج كما يلى:

فئات العمر	20 –	30 –	40–	50 – 60
عدد العمال	10	30	50	20

المطلوب: حساب التالي:

- الوسط الحسابى
- التباين
- الانحراف المعياري
- متوسط الانحرافات المطلقة

الحل:

يتم إعداد الجدول التالي حتى يمكن حساب كلاً من الوسط الحسابي والتباين والانحراف المعيارى:

فئات العمر	التكرار f	مركز الفئة x	xf	$x^2 f$
20 -	10	25	250	6250
30 -	30	35	1050	36750
40 -	50	45	2250	101250
50 - 60	20	55	1100	60500
المجموع	110		4650	204750
	$\sum f$		$\sum xf$	$\sum x^2 f$

- الوسط الحسابي:

يمكن إيجاد الوسط الحسابى بتطبيق المعادلة الخاصة بذلك كما يلى:

$$\bar{x} = \frac{\sum_{i=1}^{l} x_i f_i}{\sum_{i=1}^{l} f_i} = \frac{4650}{110} = 42.2727$$

- التباين:

يمكن الحصول على التباين باستخدام المعادلة الخاصة بذلك كما يلى:

$$\sigma^2 = \frac{\sum_{i=1}^{l} x_i^2 f_i}{\sum_{i=1}^{l} f_i} - \bar{x}^2$$

$$\sigma^2 = \frac{204750}{110} - (42.2727)^2 = 74.3801$$

- الانحراف المعياري:

يمكن حسابة باستخدام المعادلة الخاصة بذلك كما يلى :

$$\sigma = \sqrt{74.3801} = 8.62439$$

- متوسط الانحرافات المطلقة *AAD*:

حتى يمكن إيجاد متوسط الانحرافات المطلقة لابد اولاً من إيجاد الانحرافات عن الوسط الحسابى ثم استخدامها فى الحساب كما يتضح ذلك من الجدول التالى:

$\lvert x-\bar{x}\rvert f$	$(x-\bar{x})f$	$x-\bar{x}$	مركز الفئةx	التكرار f	فئات العمر
172.7273	-172.727	-17.2727	25	10	20 -
218.1818	-218.182	-7.27273	35	30	30 -
136.3636	136.3636	2.727273	45	50	40 -
254.5455	254.5455	12.72727	55	20	50 - 60
781.8182	0			110	المجموع
$\sum\lvert x-\bar{x}\rvert f$	$\sum(x-\bar{x})f$				

ويمكن الحصول على متوسط الانحرافات المطلقة *AAD* بتطبيق المعادلة الخاصة بذلك كما يلى:

$$AAD=\frac{\sum_{i=1}^{l}\lvert x_i-\bar{x}\rvert f_i}{\sum_{i=1}^{l} f_i}=\frac{781.8182}{110}=7.1074$$

فيتضح لنا من الجدول السابق أن:

مجموع انحرافات القيم عن وسطها الحسابي تســاوى صفر حيث أن

$$\sum(x-\bar{x})f=0$$

كما يمكن الاعتماد على هذه الانحرافات فى حساب التباين بتطبيق المعادلة الخاصة بذلك.

تابع ... المقاييس الإحصائية للبيانات المبوبة

ثانيا – الوسيط و التشتت حوله:

الوسيط هو القيمة التى يصغرها عدد من القيم يتساوى مع العدد الذى يكبر هذه القيمة، و هو يعتبر أحد مقاييس النزعة المركزية التى نلجأ إليها لتحليل الظواهر وفقا لشكل التوزيع الإحصائى محل الدراسة.

ولحساب الوسيط من البيانات المبوبة هناك ثلاث خطوات يجب إتباعها و هى:

- إيجاد الجدول التكرارى المتجمع الصاعد
- إيجاد ترتيب الوسيط من خلال المعادلة التالية:

$$k_{Med} = n / 2$$

- إيجاد قيمة الوسيط من خلال المعادلة التالية:

$$Med = \mathrm{L}_{\mathrm{Med}} + \frac{k_{Med} - F_a}{F_b - F_a} \times I$$

حيث أن :

Med	قيمة الوسيط
L_{Med}	الحد الادنى لبداية الفئة الوسيطة
k_{Med}	ترتيب الوسيط
F_a	التكرار المتجمع السابق للفئة الوسيطة
F_b	التكرار المتجمع اللاحق للفئة الوسيطة
I	طول الفئة الوسيطة

مثال :ـ

فى بيانات المثال السابق توزيع مجموعة من المدرسين العاملين فى مجال التربية وفقا لفئات أعمارهم، أحسب قيمة الوسيط؟

فئات العمر	20 -	30 -	40-	60 – 50
عدد العمال	10	30	50	20

الحل :ـ

- إيجاد الجدول التكرارى المتجمع الصاعد كما يلي:

الحدود العليا للفئات	التكرار المتجمع الصاعد
أقل من 20	0
أقل من 30	10
أقل من 40	40
أقل من 50	90
أقل من 60	110

- إيجاد ترتيب الوسيط كالتالي:

$$k_{Med} = 110/2 = 55$$

- إيجاد قيمة الوسيط من خلال التالي:

نلاحظ أن ترتيب الوسيط = 55 ، مما يعنى أن ترتيب الوسيط يقع بين التكرار المتجمع الصاعد (40 F_a) وهو المقابل للحد الأعلى للفئة 40 والتكرار المتجمع الصاعد (90 F_b) وهو المقابل للحد الأعلى للفئة 50

أي أن الحد الادنى للفئة هو L Med = 40

وبالتالى يكون طول الفئة الوسيطية هو:

$$I = 50 - 40 = 10$$

وعلى ذلك يمكن حساب قيمة الوسيط كما يلي:

$$Med = \mathrm{L}_{\mathrm{Med}} + \frac{k_{Med} - F_a}{F_b - F_a} \times I$$

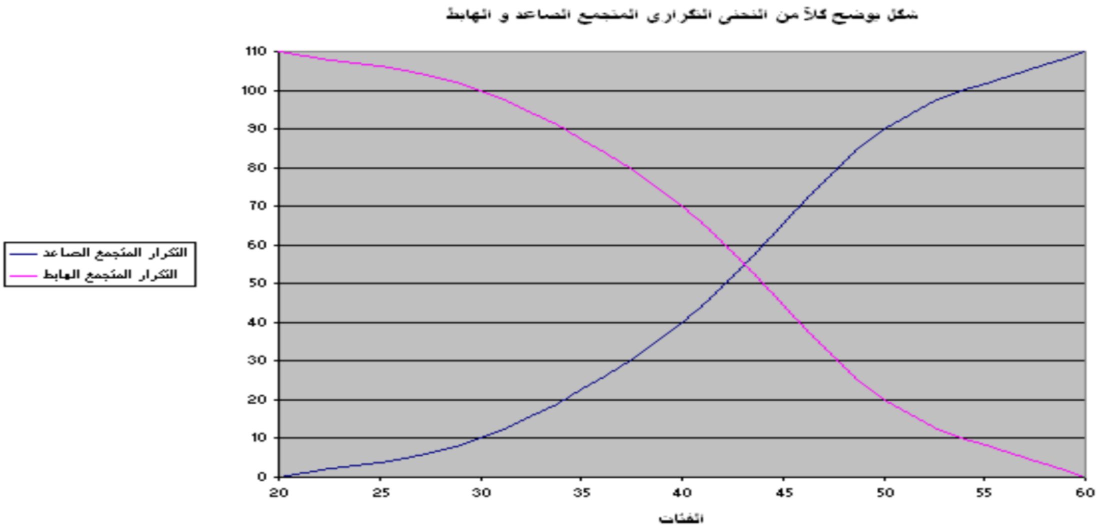
$$Med = 40 + \frac{55 - 40}{90 - 40} \times 10 = 43$$

كما يمكن إيجاد الوسيط عن طريق رسم كلا من المنحنى التكراري المتجمع الهابط والمنحنى التكراري المتجمع الصاعد كما يلى :

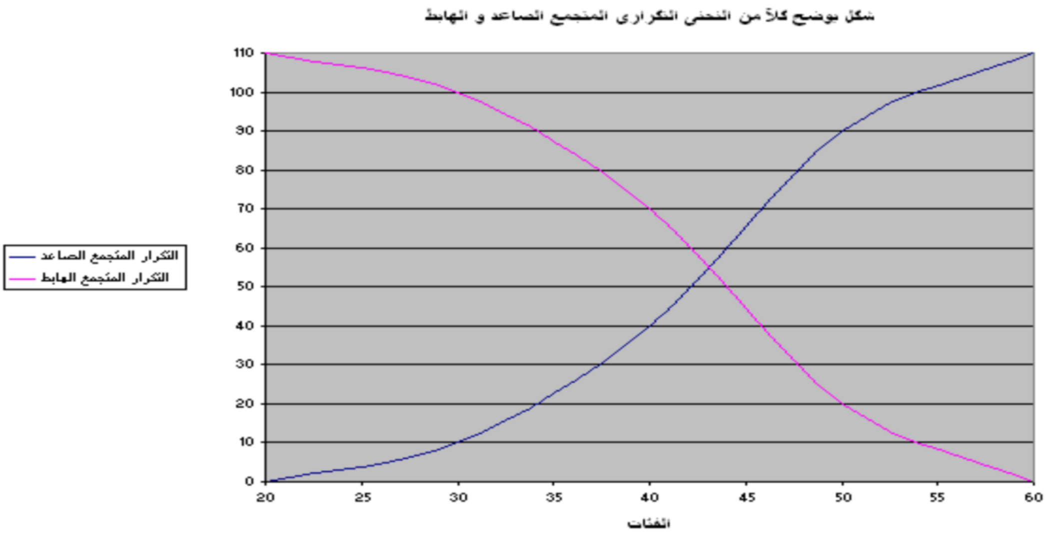
لقد قمنا في أثناء حل المثال السابق بحساب التكرار المتجمع الصاعد، ونقوم الآن بإيجاد التكرار المتجمع الهابط كما يلى:

الحدود الدنيا للفئات	التكرار المتجمع الهابط
20 فأكثر	110
30 فأكثر	100
40 فأكثر	70
50 فأكثر	20
60 فأكثر	صفر

ثم بعد ذلك نقوم برسم كلا من المنحنى التكراري المتجمع الصاعد والمنحنى التكرارى المتجمع الهابط معا كما يلى:



ويمكن الحصول على قيمة الوسيط من خلال الرسم بأن نسقط عمود رأسي من نقطة تقاطع كلا من المنحنى التكراري المتجمع الصاعد مع المنحنى التكراري المتجمع الهابط على المحور الرأسى لنقرأ قيمة الوسيط كما يتضح مما يلى:



و يتضح لنا من الشكل السابق أن الوسيط تبلغ قيمتة 43 تقريبا

الرُبيع الادنى (الأول):

يُعبر الرُبيع الأول Q1 عن القيمة التى يكون قبلها عدد المشاهدات ربع العدد الكلى للمشاهدات والمشاهدات بعده تمثل ثلاث ارباع العدد الكلى للمشاهدات محل الدراسة.

لذلك يتم حسابة كما فى حالة الوسيط مع اختلاف أن ترتيب الرُبيع الاول Q1 هو (n / 4)

$$Q_1 = L_{Q_1} + \frac{\frac{n}{4} - F_a}{F_b - F_a} \times I_{Q_1}$$

الرُبيع الاعلى (الثالث):

يُعبر الرُبيع الثالث Q3 عن القيمة التى يكون قبلها عدد المشاهدات ثلاث ارباع العدد الكلى للمشاهدات والمشاهدات بعده تمثل ربع العدد الكلى للمشاهدات محل الدراسة.

لذلك يتم حسابه كما فى حالة الوسيط مع اختلاف أن ترتيب الرُبيع الثالث Q3 هو (3 n / 4)

$$Q_3 = L_{Q_3} + \frac{\frac{3(n)}{4} - F_a}{F_b - F_a} \times I_{Q_3}$$

ويمكن إيجاد كلا من الرُبيع الادنى (الاول) Q1 و الرُبيع الاعلى (الثالث) Q3 بنفس خطوات حساب الوسيط الا أن الامر المختلف هنا هو الترتيب حيث يكون كالتالي:

Q3	Q1	
$k_{Q3} = 3n/4$	$k_{Q1} = n/4$	الترتيب

مثال :-

من بيانات المثال السابق أحسب كلا من الرُبيع الاول والرُبيع الثالث؟

حساب الرُبيع الادنى (الاول) Q1 :

الجدول المتجمع الصاعد تم إعداده سابقا

- إيجاد ترتيب الرُبيع الاول كالتالي:

$$k_{Q1} = n/4 = 110/4 = 27.5$$

- إيجاد قيمة الرُبيع الادنى (الاول) Q1 كالتالي:

نلاحظ أن ترتيب الرُبيع الادنى (الاول) Q1 27.5 مما يعنى أن الرُبيع الادنى (الاول) Q1 يقع بين التكرار المتجمع الصاعد (10 F_a) وهو المقابل للحد الأعلى للفئة 30 و التكرار المتجمع الصاعد (40 F_b) وهو المقابل للحد الأعلى للفئة 40 والحد الادنى للفئة هو = (30 L_{Q1}) .

وبالتالى يكون طول فئة الرُبيع الادنى (الاول) Q1

$$I_{Q_1} = 40 - 30 = 10$$

وعلى ذلك يمكن حساب قيمة الرُبيع الادنى (الاول) Q1 من خلال المعادلة التالية كما يلى :

$$Q_1 = L_{Q_1} + \frac{\frac{n}{4} - F_a}{F_b - F_a} \times I_{Q_1}$$

$$Q1 = 30 + \frac{27.5 - 10}{40 - 10} \times 10 = 35.8333$$

حساب الرُبيع الاعلى (الثالث) Q3 :

الجدول المتجمع الصاعد تم إعداده سابقا

نقوم بإيجاد ترتيب الرُبيع الاعلى (الثالث) Q3 كالتالي:

$$k_{Q3} = 3(n)/4 = (3)110/4 = 82.5$$

إيجاد قيمة الرُبيع الاعلى (الثالث) Q3 ، ونلاحظ أن ترتيب الرُبيع الاعلى (الثالث) Q3 82.5 مما يعنى أن الرُبيع الاعلى (الثالث) Q3 يقع بين التكرار المتجمع الصاعد (40 F_a) وهو المقابل للحد الأعلى للفئة 40 والتكرار المتجمع الصاعد (90 F_b) وهو المقابل للحد الأعلى للفئة 50 والحد الادنى للفئة هو = (40 L_{Q3}) .

وبالتالى يكون طول فئة الربيع الاعلى (الثالث)

I_{Q_3} = 50 – 40 = 10

وعلى ذلك يمكن حساب قيمة الربيع الاعلى (الثالث) Q3 كما يلي:

$$Q_3 = \mathrm{L}_{Q_3} + \frac{\frac{3(n)}{4} - F_a}{F_b - F_a} \times I_{Q_3}$$

$$Q3 = 40 + \frac{82.5 - 40}{90 - 40} \times 10 = 48.5$$

حساب قيمة العُشير :

وبنفس الطريقة السابقة يمكن الحصول على العشير $P_{0.10}$ وهو القيمة التى يكون قبلها 10 % من مفردات المجتمع و 90 % منها أكبر منه. والاختلاف يكون فقط فى الترتيب حيث أن ترتيب العشير هو:

$$k_{P_{0.10}} = n/10$$

ففى مثالنا هذا يكون ترتيب العشير هو :

$$k_{P_{0.10}} = n/10 = 110/10 = 11$$

ونلاحظ أن ترتيب العُشير $P_{0.10}$ هو 11 مما يعنى أن العُشير $P_{0.10}$ يقع بين التكرار المتجمع الصاعد (10 F_a) وهو المقابل للحد الأعلى للفئة 30 والتكرار المتجمع الصاعد (40 F_b) وهو المقابل للحد الأعلى للفئة 40 والحد الادنى للفئة هو = (30 $L_{P_{0.10}}$) .

وبالتالى يكون طول فئة العُشير :

$I_{P_{0.10}}$ = 40 – 30 = 10

وعلى ذلك يمكن حساب قيمة العُشير $P_{0.10}$ كما يلى :

$$P_{0.10} = L_{P_{0.10}} + \frac{\frac{n}{10} - F_a}{F_b - F_a} \times I_{P_{0.10}}$$

$$P_{0.10} = 30 + \frac{11 - 10}{40 - 10} \times 10 = 30.333$$

حساب قيمة المئويين أو المئين $P_{0.01}$:

وبنفس الطريقة السابقة يمكن الحصول على المئويي $P_{0.01}$ وهو القيمة التى يكون قبلها 1 % من مفردات المجتمع و 99 % منها أكبر منه، والاختلاف بينه وبين ما سبق حسابه من الوسيط والربيع الأول أو الربيع الثالث أو العُشير يكون فقط فى الترتيب حيث أن ترتيب المئويين هو :

$$k_{P_{0.01}} = n/100$$

ففى مثالنا هذا يكون ترتيب المئويين $P_{0.01}$ هو :

$$k_{P_{0.01}} = n/100 = 110/100 = 1.1$$

ونلاحظ أن ترتيب المئويين $P_{0.01}$ هو 1.1 مما يعنى أن المئويين $P_{0.01}$ يقع بين التكرار المتجمع الصاعد (صفر F_a) وهو المقابل للحد الأعلى للفئة 20 والتكرار المتجمع الصاعد (10 F_b) وهو المقابل للحد الأعلى للفئة 30 والحد الادنى للفئة هو = (20 $L_{P_{0.01}}$).

وبالتالى يكون طول فئة المئويين

$I_{P_{0.01}}$ = 30 – 20 = 10

وعلى ذلك يمكن حساب قيمة المئويين كما يلى :

$$P_{0.01} = \mathrm{L}_{P_{0.01}} + \frac{\frac{n}{100} - F_a}{F_b - F_a} \times I_{P_{0.01}}$$

$$P_{0.01} = 20 + \frac{1.1 - 0}{10 - 0} \times 10 = 21.1$$

وعلى ذلك نكون قد حصلنا على مقاييس النزعة المركزية التى تصف تركز البيانات عند أي نسبة من مفردات البيانات محل الدراسة فى حالة البيانات المبوبة والتى كانت كما يلي:

المقياس	$P0.01$	$P_{0.10}$	Q1	Med	Q3
القيمة	21.1	30.333	35.8333	43	48.5

نصف المدى الرُبيعي Inter Quartile Range

بسبب العيب الموجود فى مقياس التشتت (المدى) وتأثرة بالقيم الشاذة أدى ذلك للجوء إلى مقياس آخر يسمى (نصف المدى الربيعى IQR) والذى يستبعد القيم المتطرفة من الطرفين، حيث يعتمد فى حسابه على كلا من الرُبيع الأول Q1 والرُبيع الثالث Q3 ويتم حسابه من خلال المعادلة التالية:

$$IQR = \frac{Q3 - Q1}{2}$$

وبالتالى يكون نصف المدى الرُبيعي فى مثالنا هو:

$$IQR = \frac{48.5 - 35.8333}{2} = 6.33335$$

ثالثا: المنوال :

المنوال هو القيمة الأكثر شيوعا أو تكرارا. وفى حالة البيانات المبوبة يمكن حسابة باستخدام المعادلة التالية:

$$Mod = L_{Mod} + \frac{D1}{D1 + D2} \times I$$

Mod قيمة المنوال

L_{Mod} الحد الأدنى لفئة المنوال

$D1$ يساوى تكرار فئة المنوال – تكرار الفئة السابقة

$D2$ يساوى تكرار فئة المنوال – تكرار الفئة اللاحقة

I طول الفئة المنوالية

أحسب المنوال للأعمار مجموعة من المدرسين العاملين فى مجال التربية فى المثال السابق؟

الحل:

بيانات توزيع العمال وفقا لفئات العمر كانت كما يلى :

الحد الادنى L_{Mod}

فئات العمر	20 -	30 -	40-	50 – 60
عدد العمال	10	30	50	20

D1 = 30

الفئه المنواليه

D2 = 20

نلاحظ أن أكبر تكرار هو 50 و يكون مقابل للفئة " 40 – 50 " لذلك يطلق عليها الفئة المنوالية و من ثم فأن الحد الادنى لها هو (40 L_{Mod}) وطول الفئة هو

.(I 10)

كما يمكن حساب كلا D1 و D2 كالتالي:

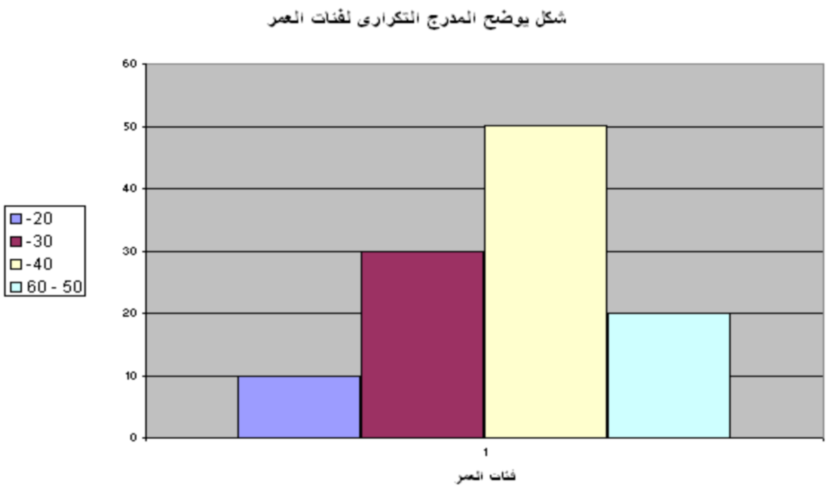
$$D1 = 50 - 30 = 20$$

$$D2 = 50 - 20 = 30$$

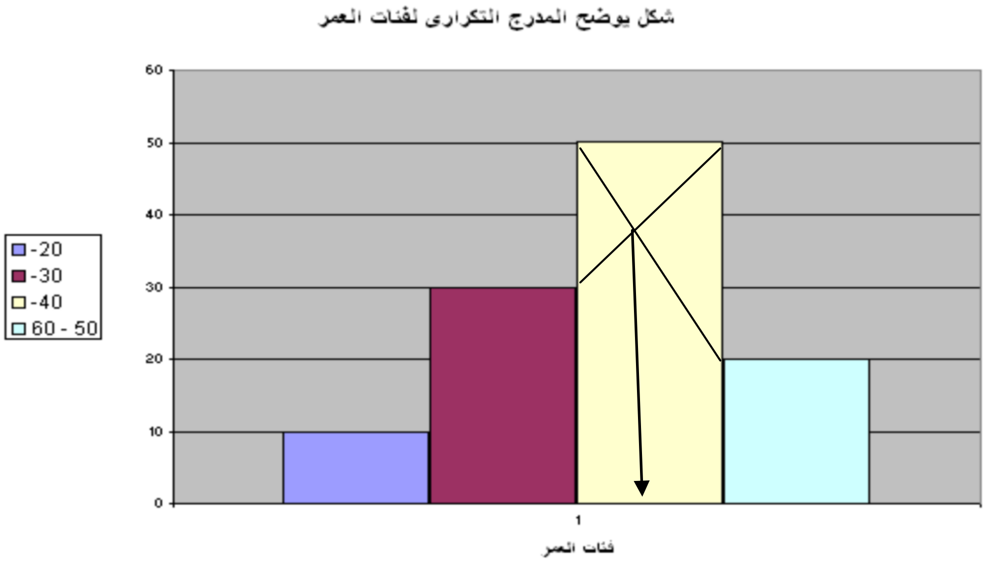
وبالتالى يمكن حساب قيمة المنوال Mod كالتالي:

$$Mod = 40 + \frac{20}{20+30} \times 10 = 44$$

كما يمكن إيجاد المنوال بيانيا، ويتم ذلك عن طريق رسم المدرج التكرارى كما يلي:



ثم نأتى على أعلى مستطيل الذى يمثل أكبر تكرار ونصل بداية الفئة المنوالية ببداية الفئة التالية ونهاية الفئة المنوالية بنهاية الفئة السابقة عليها فيتقاطع الخطان عند نقطة نسقط منها عامودا على المحور الأفقى فيلتقى عند نقطة تكون هي قيمة المنوال كما يتضح من الشكل التالى:



الجداول غير المنتظمة:

وهي الجدوال التى يكون فيها أطوال الفئات غير متساويه ويكفى وجود فئة واحدة فقط طولها غير متساوى مع باقى الفئات لجعل الجدول غير منتظم.

ويتم حساب المقاييس الإحصائية التى سبق عرضها فى حالة الجداول المنتظمة بنفس الطريقة فيما عدا المنوال،

ويتعين علينا عند حساب المنوال تعديل التكرارات قبل حسابة وكذلك قبل رسم المدرج التكرارى وذلك لأن حجم التكرارات فى تلك الحالة قد يسبب اتساع أو ضيق في أعمدة فئات التوزيع ولذلك يتم التخلص من تأثير طول الفئة بإيجاد التكرار المعدل، ويتم ذلك من خلال المعادلة التالية:

التكرار المعدل = التكرار الأصلى للفئة ÷ طول الفئة

مثال:

البيانات التالية توضح توزيع مجموعة من الموظفين وفقا لفئات دخلهم الشهري بالألف ريال فكانت كما يلي:

فئات الدخل	3 -	5 -	8-	15 – 10
عدد الموظفين	20	50	15	15

المطلوب حساب:

1- الوسط الحسابي
2- متوسط الانحرافات المطلقة
3- التباين
4- الانحراف المعياري
5- الوسيط
6- الرُبيع الأول
7- الرُبيع الثالث
8- العُشير
9- المئوين
10- نصف المدى الرُبيعى
11- المنوال

الحل

يمكن حساب المطلوب من 1 إلى 10 بنفس طريقة حسابها فى حالة الجداول المنتظمة بدون أي تعديل. اما المطلوب رقم 11 فيطلب حساب المنوال، وهو الذي طريقته تحتاج إلى تعديل في الحساب في حالة الجداول غير المنتظمة، ويتم ذلك وفق التالي:

ولحساب المنوال فى هذه الحالة فإنه لا يتم الاعتماد على بيانات الفئات الاصلية وإنما يتم إيجاد التكرار المعدل بقسمة تكرار كل فئة على طولها كما يلي:

فئات الدخل	التكرار f	طول الفئة	التكرار المعدل
3 -	20	2	10
5 -	50	3	16.6667
8 -	15	2	7.5
10 -15	15	5	3
المجموع			

نلاحظ أن أكبر تكرار معدل هو 16.6667 و يكون مقابل للفئة " 5 - 8 " لذلك يطلق عليها الفئة المنوالية و من ثم فإن الحد الادنى L_{Mod} لها هو 5 طول الفئة I هو 3.

كما يمكن حساب كلا من D1 و D2 كالتالي:

$$D1 = 16.66667 - 10 = 6.66667$$

$$D2 = 16.66667 - 7.5 = 9.16667$$

وبالتالى يمكن حساب قيمة المنوال Mod من خلال تطبيق معادلة حساب المنوال مع الأخذ في الاعتبار التكرار المعدل كالتالي:

$$Mod = 5 + \frac{6.66667}{6.66667 + 9.16667} \times 3 = 6.263158$$

الجداول المفتوحة :

وهى ذلك النوع من الجداول التى يكون فيها الحد الادنى للفئة الاولى غير محدد أو الحد الأعلى للفئة الأخيره غير محدد أو كلاهما. وفى هذا النوع من الجداول يصعب حساب الوسط الحسابى والتباين والانحراف المعيارى، حيث لا يمكن تحديد مركز الفئة للفئات المفتوحة، لذا فيعبير من أنسب المقاييس الإحصائية فى تلك الحالة هى المقاييس الوسيطية والتى يقصد بها الوسيط والرُبيع الادنى والرُبيع الاعلى والعُشير والمئويين وكذلك لقياس التشتت يتم من خلال نصف المدى الربييعى.

مثال:

البيانات تعبر عن أوزان مجموعة من الطلاب بالكيلوجرام في المرحلة الجامعية فكانت كما يلي:

فئات الوزن	أقل من 50	50 -	60 -	70 -	80 فأكثر
عدد الطلاب	5	10	35	15	10

المطلوب:

حساب مقاييس النزعة المركزية والتشتت المناسبة ؟

الحل :

موجود في الكتاب صفحة 144 و 145