الإحصاء وتصميم التجارب

Statistics and experimental design

علم الإحصاء Statistics: فرع من العلوم يختص بطرق جمع وتنظيم وتلخيص وعرض وتحليل وتفسير البيانات.... ويُقسم علم الإحصاء إلى:

الإحصاء الوصفي Descriptive Statistics: عبارة عن الطرق الخاصة بتنظيم وتلخيص المعلومات والغرض من التنظيم هو المساعدة على فهم المعلومات، والطرق الوصفية تحتوي على توزيعات تكرارية (الجداول التكرارية) ورسوم بيانية وطرق حساب مقاييس النزعة المركزية ومقاييس التشتت ومختلف القياسات الأخرى.

الإحصاء الاستدلالي Inferential Statistics: عبارة عن الطرق العلمية التي تستخدم للاستدلال عن معالم المجتمع بناءً على المعلومات التي تم الحصول عليها من العينة المأخوذة منه، وذلك وفق الطرق الإحصائية المعلومة، وبمعنى آخر فإن الإحصائيات الاستدلالية تُتيح لنا اختيار فكرة أو فرضية وسيكون هناك دائماً درجة من عدم اليقين لأنها لا يُمكن أن تخبرك إلا عن مدى احتمالية حدوث شيء من عدمه.

مفاهيم أساسية في علم الإحصاء:

المجتمع الإحصائي Statistical Population: جميع الوحدات الإحصائية التي نرغب في دراستها، وكل فرد من المجتمع يُدعى وحدة إحصائية Statistical Unit، وقد يكون المجتمع الإحصائي محدود مثل عدد الأشجار في مزرعة ما، أو غير محدود كعدد النجوم في السماء.

العينة Sample: جزء من المجتمع الإحصائي يجب ألا يقل عدد أفرادها عن 2– 10% من عدد أفراد المجتمع، تؤخذ بطريقة عشوائية، من أنواعها:

1– العينة العشوائية البسيطة Simple Random Sample: هي العينة التي يكون لكل وحدة إحصائية من المجتمع الإحصائي المدروس نفس الفرصة بأن تكون مُمثَلة في هذه العينة، وهي أكثر أنواع العينات الإحصائية شيوعاً، وتستخدم عندما يكون المجتمع متجانساً.

2- العينة العشوائية الطبقية Stratified Random Sample: في هذه الحالة يكون المجتمع الإحصائي مقسم إلى طبقات أو أقسام، يُؤخذ من كل قسم أو طبقة عينة متجانسة بطريقة عشوائية، قد تؤخذ على أساس جغرافي كأن تُقسم المدينة إلى مناطق جغرافية، أو تُقسم المصانع حسب أنواع الصناعات..

3- العينة العشوائية المنتظمة Systematic Random Sample: تُؤخذ بطريقة عشوائية ولكن من أماكن أو فترات منتظمة، مثال: قد نختار من الحقول التي تبعد ميلاً واحداً عن بعضها(مع اختيار الحقل الأول عشوائياً في الميل الأول، وذلك عند معاينة محصول معين)، أو قد نقوم بقياس درجات الحرارة كل يوم أو كل ساعة....

4- العينة العشوائية العنقودية Cluster Random Sample: تؤخذ من مجتمع مرتب بطريقة عنقودية أو تنظيمات هرمية، مثل التنظيمات الإدارية في المحافظات (محافظة، مدينة، قسم، منطقة، ناحية، قرية)

البيانات Data: مجموعة المشاهدات التي نحصل عليها أثناء إجراء دراسة ما وتصنف إما بيانات كمية مثل الدخل الشهري أو بيانات وصفية مثل الحالة الاجتماعية، وتستخدم كلمة الإحصاءات لتعبر عن البيانات التي توجد في الجداول.

أنواع البيانات: تُقسم البيانات بشكل أساسي إلى البيانات الوصفية والبيانات الكمية

1- البيانات الوصفية Qualitative Data : بيانات غير رقمية، لا يُمكن قياسها عددياً، تُقسم بدورها لإلى نوعين:

- بيانات قابلة للترتيب حسب رتبة معينة تصاعدياً أو تنازلياً مثل المستوى التعليمي.
- بيانات غير قابلة للترتيب مثل الحالة العائلية أو الجنسية.

2- البيانات الكمية Quantitative Data: بيانات يعبر عنها بأرقام عددية تمثل القيمة الفعلية للظاهرة أو الصفة المدروسة، مثل الإنتاج- الاستهلاك- الوزن- الطول....، تُقسم أيضاُ إلى نوعين:

- بيانات كمية منقطعة: تأخذ قيماً صحيحة لا يُمكن تجزئتها مثل عدد أفراد الأسرة، عدد الغرف...
- بيانات كمية مستمرة: تأخذ كل القيم الممثلة لمجال الدراسة، ونظراً للعدد غير المتناهي لهذه القيم يُقسم مجال الدراسة إلى مجالات جزئية تسمى الفئات مثل الوزن ، الطول.

طرق عرض البيانات:

الخطوة التالية بعد جمع البيانات في مجال الإحصاء الوصفي، هو تبويب البيانات وعرضها بصورة يمكن الاستفادة منها في وصف الظاهرة محل الدراسة، من حيث تمركز البيانات، ودرجة تجانسها، وهناك طريقتين لعرض البيانات هما:

1- عرض البيانات جدولياً

2- عرض البيانات بيانياً

جدولياً: يمكن عرض البيانات في صورة جدول تكراري ويختلف شكل هذا الجدول طبقاً لنوع البيانات، وعدد المتغيرات، ويتم توزيع البيانات في أعمدة وأسطر تتضمن حدود الفئة ومركز الفئة وتكرار الفئة وغيرها...

خطوات إنشاء جدول التوزيع التكراري:

مثال: فيما يلي أرقام تمثل إنتاجية أبقار حلوب (24 بقرة):

2، 5، 7، 4، 2، 3، 5، 6، 4، 2، 3، 5، 7، 9، 11، 12، 10، 12، 11، 12، 10، 9، 8، 6.

1- نرتب الأرقام تصاعدياً:

2-2-2-3-3-4-4-5-5-5-6-6-7-7-8-9-9-10-10-11-11-12-12-12

2- نقوم بحساب المدى العام: أكبر قيمة- أصغر قيمة

$Range= X_{max}- X_{min}$

12- 2= 10

3- حساب عدد الفئات (Classes):

نستخدم إحدى الطريقتين:

طريقة يول Yule:

$C= 2.5 \sqrt[4]{n}$

طريقة ستورجس Sturgis:

$C= 1+ 3.32 \log n$

حيث n: عدد أفراد العينة

$C= 1+3.32 \log 24$

$C= 5.58 \approx 6$

أي لدينا 6 فئات.

4- حساب مدى الفئة (طول الفئة Length):

مدى الفئة= المدى العام/ عدد الفئات

$$L=R/C= 10/6= 1.7\approx 2$$

(ملاحظة هامة: يجب تقريب رقم مدى الفئة **لأعلى** رقم صحيح، مثال: مدى الفئة 4.2 نقرب مدى الفئة لل5 وليس 4).

5- حساب مركز الفئة= الحد الأدنى للفئة+ الحد الأعلى للفئة/2

6- تكرار الفئة: هو عدد مرات ظهور الرقم ويتضمن الحد الأدنى وكافة الأرقام الواقعة ضمن حدود الفئة دون إضافة الحد الأعلى (لأنه يدخل في الفئة التالية)

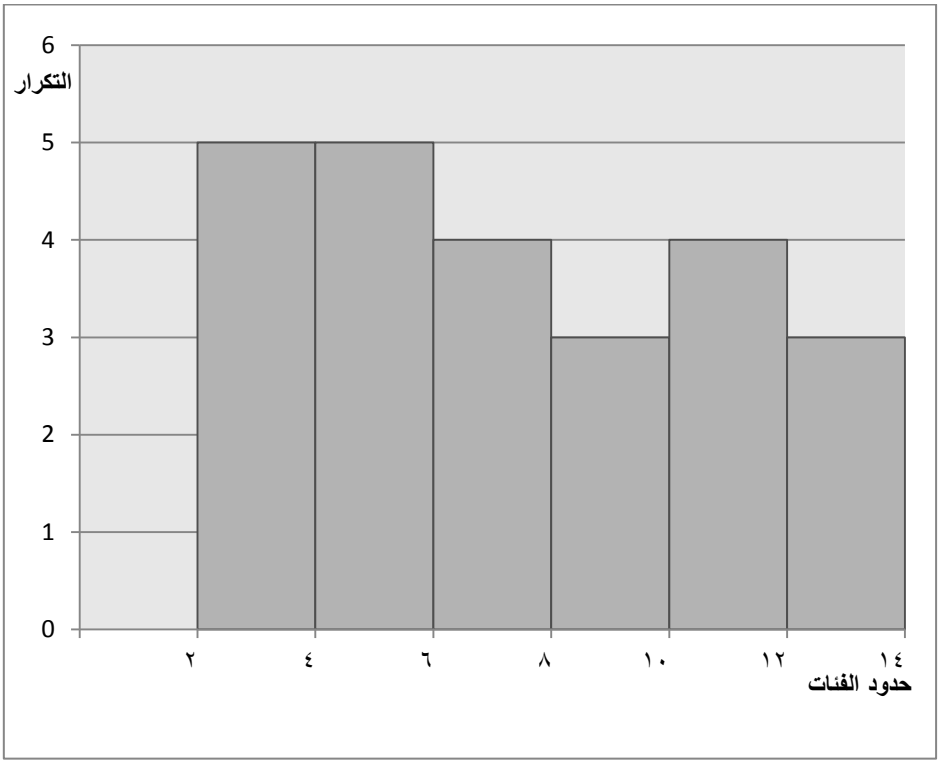
7- التكرار التجميعي الصاعد: هو مجموع تكرار الفئة ذاتها مع تكرار جميع الفئات السابقة.

8- التكرار الشكلي: إما على شكل خماسيات أو على شكل مظروف.

9- رسم المضلع أو المنحني التكراري: عبارة عن الأعمدة التي تمثل بيانياً نقاط تقاطع حدود الفئات مع تكراراتها.

التكرار الشكلي	التكرار على شكل خماسيات	تكرار متجمع صاعد	تكرار الفئة	مركز الفئة	حدود الفئة
	𝍸	5	5	3	2-4
	𝍸	10	5	5	4-6
	\\\\	14	4	7	6-8
	\\\	17	3	9	8-10
	\\\\	21	4	11	10-12
	\\\	24	3	13	12-14

المضلع التكراري:

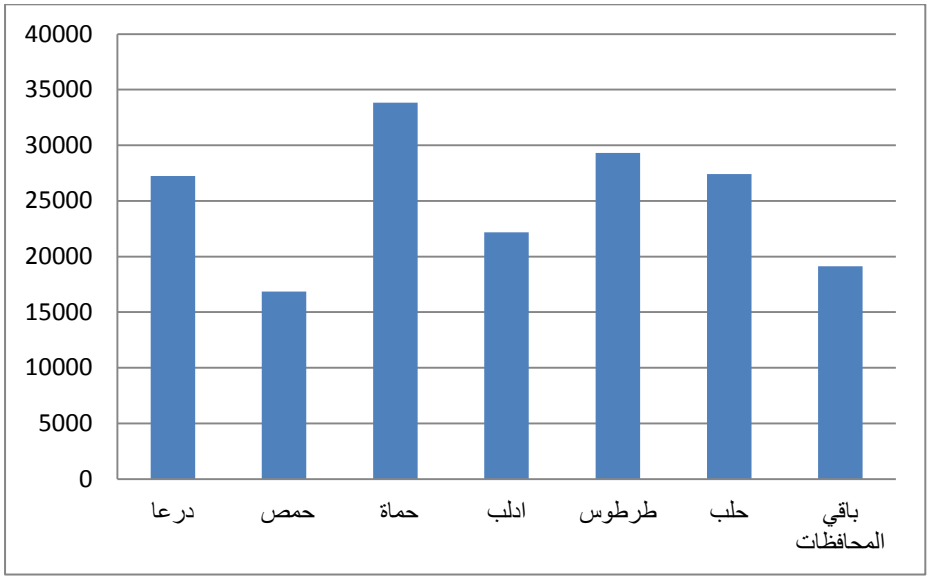


بيانياً: استخدم الإحصائيون تقنيات الرسم البياني لوصف البيانات بطرقة أفضل، إذ أن العرض البياني يساعد في إعطاء صورة سريعة لوصف ظاهرة معينة..

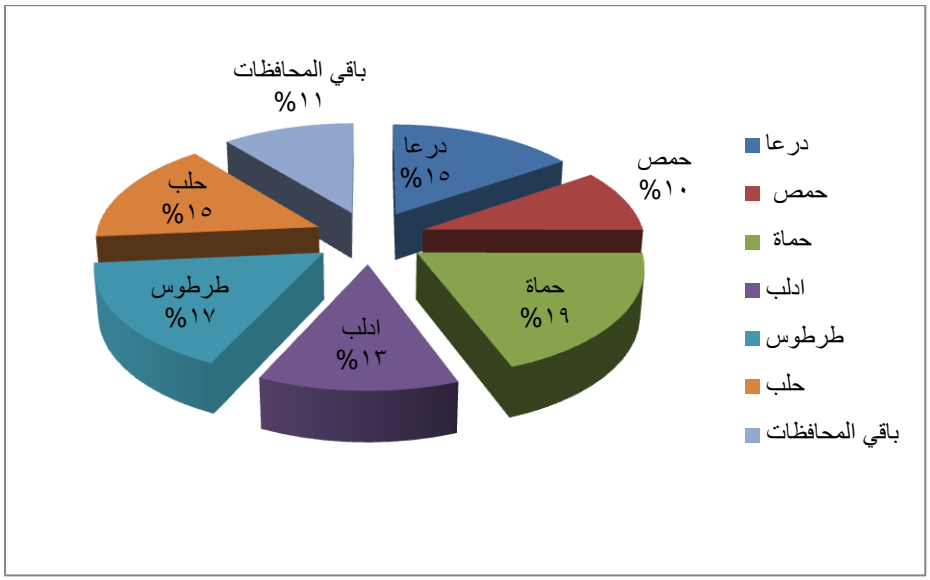
1- العرض البياني لبيانات متغير وصفي:

أمثلة عن أهم طرق العرض البياني لبيانات متغير وصفي:

- طريقة الأعمدة البيانية:



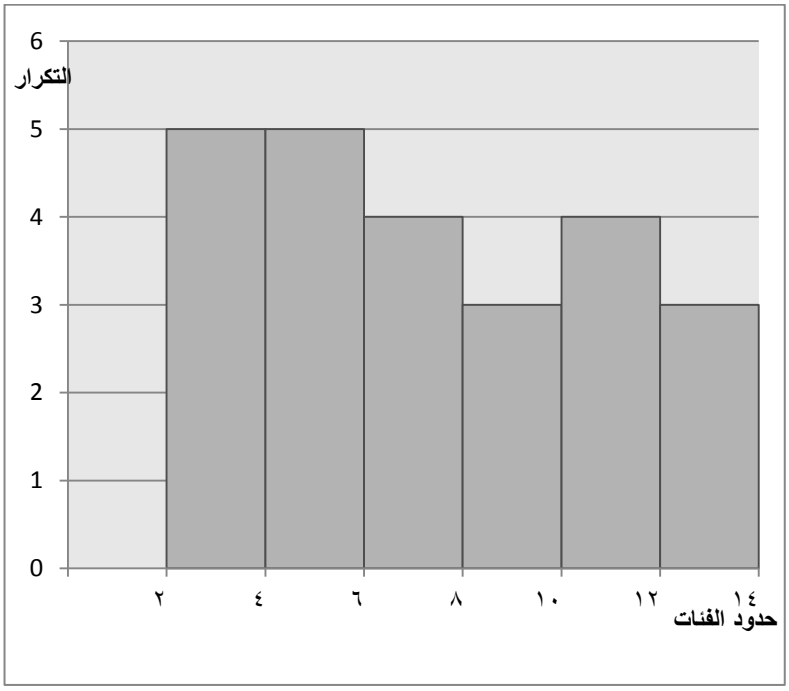
- طريقة الدائرة البيانية:



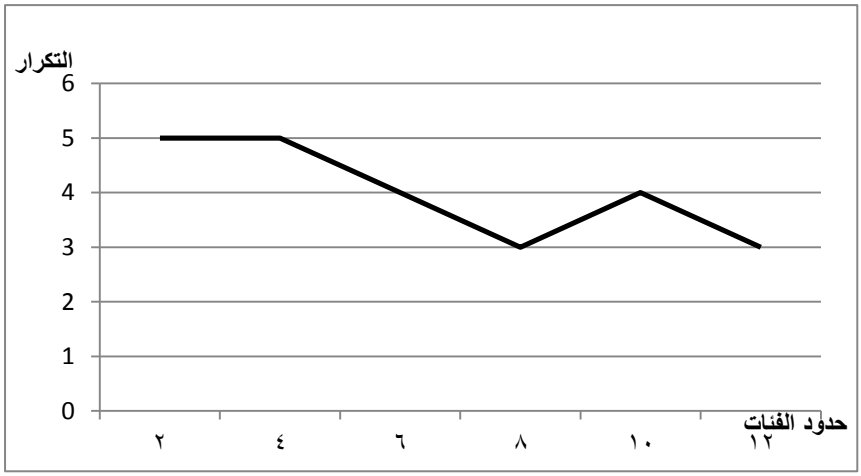
2- العرض البياني لبيانات متغير كمي:

أمثلة أهم طرق العرض البياني لبيانات متغير كمي:

المدرج التكراري:



الخطوط البيانية:



أهم المقاييس الإحصائية:

أولاً: مقاييس النزعة المركزية Measures of Central Tendency:

لأي بيانات إحصائية خواص تساعد على إعطاء فكرة ومدلول عن وضع هذه البيانات ومن هذه الخصائص إحصائيات النزعة المركزية وتسمى بالمتوسطات وأهم هذه المتوسطات: المتوسط الحسابي، الوسيط، المنوال وتتمكن بواسطتها من تحديد موقع النقطة التي تتمحور حولها كافة القيم.

1- **الوسط الحسابي Arithmetic Mean:** يعتبر من أهم مقاييس النزعة المركزية والأكثر استخداماً في النواحي التطبيقية، ويُستخدم عادةً في المقارنات بين الظواهر المختلفة، ويُحسَب الوسط الحسابي رياضياً للبيانات المبوبة وغير المبوبة كما يلي:

1-1- حالة البيانات غير المبوبة (غير المدرجة ضمن جدول تكراري):

يُحسب من العلاقة:

$$\bar{X} = \sum_{i=1}^{n} x_i \,/\, n$$

حيث: x تمثل بيانات الظاهرة المدروسة، n تمثل عدد قيم الظاهرة.

مثال: احسب المتوسط الحسابي لإنتاج حليب عدد من الأبقار الحلوب خلال يوم واحد في مزرعة إذا كان إنتاجها كالتالي 30- 20- 25- 15- 40 كغ

$$\bar{X} = \sum_{i=1}^{n} x_i \,/\, n$$

$$= (40+15+20+25+30)/\ 5$$

$$= 26 \text{ kg}$$

1-2- حالة البيانات المبوبة:

يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة X_i مع توفر التكرارات وبذلك يكون المتوسط الحسابي يمثل مجموع ضرب مراكز الفئات في التكرارات مقسوماً على مجموع التكرارات.

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{\sum_{i=1}^{n} f_i}$$

مثال: احسب متوسط إنتاج حليب الأبقار في الجدول الآتي:

فئات الإنتاج	32-34	34-36	36-38	38-40	40−42	42-44
عدد الأبقار	4	7	13	10	5	1

الحل:

فئات الإنتاج	التكرارات f_i	مراكز الفئات x_i	$x_i \cdot f_i$
32-34	4	33	132
34-36	7	35	245
36-38	13	37	481
38-40	10	39	390
40−42	5	41	205
42-44	1	43	43
∑	**40**		**1496**

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

= 1496/ 40

= 37.4

2- **الوسيط Median:** هو أحد مقاييس النزعة المركزية والذي يأخذ في الاعتبار رتب القيم، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم (50% من القيم أقل منه، و50% من القيم أعلى منه)، وفيما يلي كيفية حساب الوسيط في حالة البيانات غير المبوبة والبيانات المبوبة.

2-1- البيانات غير المبوبة:

نقوم باتباع الخطوات الآتية:

- نرتب القيم تصاعدياً
- نحدد رتبة الوسيط وهي $\frac{n+1}{2}$
- إذا كان عدد القيم زوجي فإن الوسيط يقع بين القيمة رقم $\frac{n}{2}$ والقيمة رقم $\frac{n}{2}+1$ ومن ثم يحسب الوسيط بتطبيق المعادلة الآتية:

الوسيط= {العينة رقم $\frac{n}{2}$ + العينة رقم $\frac{n}{2}+1$ } /2

مثال: لدينا القيم التالية احسب الوسيط لها: 50- 70-90- 60- 80

نرتب القيم تصاعدياً:

50- 60- 70- 80- 90

نحدد رتبة الوسيط:

= (5+1)/2 =3

أي الوسيط 70

بعد إضافة قيمة أخرى للمثال السابق، احسب الوسيط

50- 70- 90- 100- 60- 80

نرتب القيم تصاعدياً:

50- 60- 70- 80- 90- 100

الوسيط= {العينة رقم $\frac{n}{2}$ + العينة رقم $\frac{n}{2}+1$ } /2

$$= \frac{80+70}{2}$$

=75

2-2- حالة البيانات المبوبة:

يتم تحديد الوسيط باتباع الخطوات التالية:

1- نكون الجدول التكراري الصاعد ثم نحدد ترتيب الوسيط من العلاقة:

$$C_1 = \frac{\Sigma_{i=1}^{n} f_i}{2}$$

2- نقوم بحساب قيمة الوسيط من العلاقة الآتية:

$$Me = L + \frac{C_1 - C_2}{C_3} . h$$

حيث:

L: الحد الأدنى للفئة الوسيطية، h: طول الفئة، C_1: ترتيب الوسيط (الفئة الوسيطية)، C_2: التكرار المتجمع الصاعد السابق للفئة الوسيطية ، C_3: تكرار الفئة الوسيطية

مثال:

فئات الإنتاج	التكرارات f_i	مراكز الفئات x_i	التكرار المتجمع الصاعد
32-34	4	33	4
34-36	7	35	11
36-38	13	37	24
38-40	10	39	34
40-42	5	41	39
42-44	1	43	40
∑	**40**		

$$C_1 = \frac{\Sigma_{i=1}^{n} f_i}{2}$$

= 40/ 2 =20

إذاً الفئة الوسيطية هي الفئة 36- 38 لأن تكرارها المتجمع الصاعد هو 24 والذي قبلها 11 وال20 تقع بين 11- 24

نقوم بتطبيق القانون: $$Me = L + \frac{C_1 - C_2}{C_3} . h$$

$$Me= 36+ \frac{20-11}{13} . 2$$

$$Me= 37.38$$

(يُلاحظ أن قيمة الوسيط لا بد وأن تقع داخل حدود فئة الوسيط، أي لا تقل عن بداية فئة الوسيط ولا تزيد عن نهايتها).

3- **المنوال Mode:** تعبر قيمة المنوال عن المشاهدة الأكثر تكراراً فهو بمثابة القيمة الشائعة، وقد يكون للبيانات منوال واحد ويمكن أن يكون لها أكثر من منوال، كما يمكن ألا يوجد منوال لمجموعة من البيانات، ويُعتبر المنوال أفضل مقياس لوصف البيانات النوعية.

طرق حساب المنوال:

3-1- المنوال للبيانات الأولية (غير المبوبة): في البيانات الأولية هو القيمة أو الصفة الأكثر تكراراً مقارنةً بباقي القيم أو الصفات، وعلى ضوء هذا التعريف فإن المنوال لمجموعة من البيانات قد لا يكون قيمة أو صفة واحدة.

مثال: البيانات الآتية تمثل التقديرات التي حصل عليها 10 طلاب:

ممتاز، جيد، جيد جداً، جيد، متوسط، فوق المتوسط، جيد، ضعيف، جيد جداً، جيد.

المطلوب إيجاد المنوال لهذه البيانات؟

نلاحظ أن الصفة جيد هي الصفة الأكثر تكراراً من بين الصفات وعليه فإن المنوال هو الصفة جيد.

3-2- المنوال للبيانات المبوبة يُحسب من العلاقة:

$$D= L_{mo}+ \frac{d_1}{d_{1+}\, d_2}. h$$

حيث:

L_{mo}: الحد الأدنى للفئة المنوالية، d_1 الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة، d_2 الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة، h: طول الفئة.

فئات الإنتاج	التكرارات f_i	مراكز الفئات x_i
32-34	4	**33**
34-36	7	**35**
36-38	13	**37**
38-40	10	**39**
40−42	5	**41**
42-44	1	**43**
∑	**40**	

$$D= L_{mo}+ \frac{d_1}{d_{1+}\, d_2}. \, h$$

الفئة المنوالية هي الفئة الأكثر تكراراً 36−38، فيكون المنوال:

$$D= 36+ \frac{6}{6+3} .2= 37.33$$

مثال غير محلول:

حدود الفئات	تكرار الفئة f_i
4−6	**2**
6−8	**6**
8−10	**7**
10−12	**9**
12−14	**5**
14−16	**1**

قم بحساب المتوسط الحسابي والوسيط والمنوال.