مقاييس النزعة المركزية

Measures of Central Tendency
(Location Measures)

الإحصاء والاحتمالات (١٢٠١ إحص)

الفصل الصيفي ١٤٣٧/١٤٣٨هـ

مقاييس النزعة المركزية

- بعض مقاييس النزعة المركزية تمثل مقاييس عددية لموضع أو مكان تركز البيانات لظاهرة ما.
- هذه المقاييس تستخدم لمقارنة مجموعات البيانات المختلفة.
- مقاييس النزعة المركزية:
  - الوسط الحسابي (المتوسط) **Arithmetic Mean (Mean)**
  - الوسيط **Median**
  - المنوال **Mode**

Arithmetic Mean (Mean) الوسط الحسابي (المتوسط)

- المتوسط للبيانات المفردة (غير المبوبة)
- المتوسط للبيانات المبوبة

المتوسط للبيانات المفردة (غير المبوبة)

- إذا كان عدد البيانات (حجم العينة) هو n وكانت قيم أو مشاهدات العينة هي $x_1, x_2, \ldots, x_n$ ، فإن المتوسط (الوسط الحسابي) يرمز له $\bar{x}$ بالرمز ويعرف بالصيغة التالية:

$$\text{الوسط الحسابي} = \frac{\text{مجموع البيانات}}{\text{عدد البيانات}}$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

المتوسط للبيانات المفردة (غير المبوبة)

مثال:

أوجد المتوسط (الوسط الحسابي) للمشاهدات التالية والتي هي عبارة عن أوزان (با لكيلوجرام) مجموعة مكونة من سبعة أشخاص:

٢٥, ٣٠, ٤٠, ٤٥, ٣٥, ٥٥, ٥٠

الحل:

$x_1=25,\ x_2=30,\ x_3=40,\ x_4=45,\ x_5=35,\ x_6=55,\ x_7=50$

$n = 7$

المتوسط هو :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_7}{7}$$

$$= \frac{25+30+40+45+35+55+50}{7} = \frac{280}{7} = 40 \quad \text{(كيلوجرامًا)}$$

المتوسط للبيانات المبوبة

- البيانات الأصلية غير معروفة.
- عدد البيانات في كل فترة (تكرار الفترة) معروف.
- يستخدم مركز الفترة كقيمة تقريبية لجميع البيانات في الفترة.
- إذا كان لدينا بيانات عددها **n** وكانت هذه البيانات ملخصة في جدول تكراري.

الفترة	مركز الفترة x	التكرار f	$x\,f$
الفترة رقم 1	x_1	f_1	$x_1\,f_1$
الفترة رقم 2	x_2	f_2	$x_2\,f_2$
:	:	:	:
:	:	:	:
الفترة رقم k	x_k	f_k	$x_k\,f_k$
المجموع		$\sum f = n$	$\sum x\,f$

- لذلك فإن المتوسط للتوزيع التكراري المبوب يمكن حسابه بشكل تقريبي بالصيغة التالية:

$$\overline{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{x_1f_1 + x_2f_2 + \cdots + x_kf_k}{f_1 + f_2 + \cdots + f_k}$$

المتوسط للبيانات المبوبة

مثال:

أوجد المتوسط لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم.

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	xf
12.95 – 13.95	13.45	3	40.35
13.9 5– 14.95	14.45	5	72.25
14. 95– 15.95	15.45	15	231.75
15. 95– 16.95	16.45	16	263.20
16. 95– 17.95	17.45	10	174.50
17. 95– 18.95	18.45	1	18.45
المجموع		$n = \sum f = 50$	$\sum xf = 800.5$

المتوسط (الوسط الحسابي) هو:

$$\bar{x} = \frac{\sum x f}{\sum f} = \frac{800.5}{50} = 16.01$$

متوسط المجموعة الكلية

إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى هو n_1 ومتوسطها هو $\bar{x}_1$ وعدد بيانات المجموعة الثانية هو n_2 وكان متوسطها هو $\bar{x}_2$ ، فإن متوسط المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$\text{متوسط المجموعة الكلية} \quad \bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

مثال:

إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى هو ١٠ ومتوسطها هو ٥ وكان عدد بيانات المجموعة الثانية هو ٢٠ ومتوسطها هو ٢ ، فإن متوسط المجموعة الكلية المكونة من دمج هاتين المجموعتين هو:

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2} = \frac{10 \times 5 + 20 \times 2}{10 + 20} = \frac{90}{30} = 3$$

مميزات وعيوب الوسط الحسابي (المتوسط)

مميزات المتوسط :

إن المتوسط يعتبر من أفضل مقاييس النزعة المركزية ومن أكثرها شيوعا وذلك لما يتمتع به من صفات جيدة. ومن مميزات المتوسط نذكر ما يلي:

١ـ المتوسط سهل التعريف والحساب ويخضع للعمليات الجبرية بسهولة.

٢ـ المتوسط وحيد لمجموعة البيانات الواحدة.

٣ـ يأخذ المتوسط في الاعتبار جميع البيانات.

عيوب المتوسط :

بالرغم من أن المتوسط يعتبر من أفضل مقاييس النزعة المركزية إلا أن له بعض العيوب نذكر منها ما يلي:

١ـ يتأثر المتوسط بالقيم الشاذة أو المتطرفة.

٢ـ المتوسط غير معرف للبيانات الوصفية (النوعية) إذ يمكن حسابه للبيانات الكمية فقط.

ملاحظة:

وحدة المتوسط هي نفس وحدة البيانات الأصلية . فإذا كانت وحدة البيانات هي الكيلوجرام فإن وحدة المتوسط هي الكيلوجرام.

الوسط المرجح (الموزون) Weighted Mean

في بعض الأحيان تكون المشاهدات $x_1, x_2, \ldots, x_n$ مقرونة بالأوزان $w_1, w_2, \ldots, w_n$ علـــى التوالي. وفي هذه الحالة نعرف الوسط المرجح كما يلي:

$$\bar{x}_w = \frac{\sum x\,w}{\sum w} = \; = = \frac{x_1 w_1 + x_2 w_2 + \cdots + x_n w_n}{w_1 + w_2 + \cdots + w_n}$$

Weighted Mean (الموزون) الوسط المرجح

مثال:

أوجد الوسط المرجح لدرجات الطلاب باعتبار أن الوزن هو عدد الساعات للمقرر فيما يلي:

المقرر	عدد الساعات (w)	الدرجة (x)
إحص	2	40
فيز	4	65
ريض	3	70

الحل:

x	w	wx
40	2	80
65	4	260
70	3	210
	$\sum w = 9$	$\sum x\,w = 550$

$$\bar{x}_w = \frac{\sum x\,w}{\sum w} = \frac{550}{9} = 61.11 \quad \text{(درجة)}$$

الوسيط Median

- الوسيط هو أحد مقاييس النزعة المركزية المشهورة. ويعرف الوسيط لمجموعة من البيانات على أنه تلك القيمة التي تتوسط البيانات عند ترتيبها تصاعدياً (أو تنازلياً) أي أنه تلك القيمة التي تقسم البيانات بعد ترتيبها إلى جزأين متساويين فتكون البيانات في الجزء الأول تقل عن أو تساوى الوسيط والبيانات في الجزء الثاني تزيد عن أو تساوى الوسيط . أي أن ٥٠% من البيانات تساوي أو تقل عن الوسيط و ٥٠% من البيانات تساوي أو تزيد عن الوسيط . يرمز للوسيط بالرمز (Med).

- الوسيط للبيانات المفردة (غير مبوبة)

- الوسيط للبيانات المبوبة

الوسيط للبيانات المفردة (غير مبوبة)

إذا كانت قيم العينة هي $x_1, x_2, \ldots, x_n$ وحجم العينة هو n فإن الوسيط يعرف كما يلي:

- **أولاً: إذا كان حجم العينة n عدداً فردياً:**

الوسيط = القيمة التي في منتصف البيانات بعد ترتيبها وهي القيمة المرتبـــة ذات الترتيب $\frac{n+1}{2}$.

البيانات مرتبة	$X_{(1)}$	$X_{(2)}$	…	$X_{(\frac{n+1}{2})}$	…	$X_{(n)}$
الترتيب	1	2	…	$\frac{n+1}{2}$	…	n.

الوسيط = $X_{(\frac{n+1}{2})}$ = القيمة في المنتصف

الوسيط للبيانات المفردة (غير مبوبة)

- **ثانيًا: إذا كان حجم العينة n عددًا زوجيًا:**

الوسيط = متوسط القيمتين في منتصف البيانات بعد ترتيبها وهما القيمتان المرتبتان ذاتا الترتيب $\frac{n}{2}$ و $\frac{n}{2}+1$.

البيانات مرتبة	$X_{(1)}$	$X_{(2)}$	…	$X_{(\frac{n}{2})}$	$X_{(\frac{n}{2}+1)}$	…	$X_{(n)}$
الترتيب	1	2	…	$\frac{n}{2}$	$\frac{n}{2}+1$	…	n.

القيمتان في المنتصف هما: $X_{(\frac{n}{2})}$ و $X_{(\frac{n}{2}+1)}$ لذلك فإن

$$\text{الوسيط} = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$$

الوسيط للبيانات المفردة (غير مبوبة)

مثال:

أوجد الوسيط لمجموعة الأوزان (بالكيلوجرام) التالية: 7.1, 2.5, 2.5, 5.4, 8.3.

الحل:

بما أن $n=5$ عدد فردي فإن الوسيط هو القيمة التي في المنتصف بعد ترتيب البيانات وهي القيمة ذات الترتيب $\frac{n+1}{2}=\frac{5+1}{2}=3$.

البيانات مرتبة	2.5	2.5	5.4	7.1	8.3
الترتيب	1	2	3	4	5

الوسيط هو القيمة ذات الترتيب 3 لذلك فإن: الوسيط = 5.4 كيلوجراما

الوسيط للبيانات المفردة (غير مبوبة)

مثال:

أوجد الوسيط لمجموعة الأوزان (بالكيلوجرام) التالية: 8.3, 9.2, 5.4, 2.5, 2.5, 7.1.

الحل:

بما أن n=6 عدد زوجي فإن الوسيط هو متوسط القيمتين اللتين في المنتصف بعد ترتيب البيانات

وهما القيمتان ذاتا الترتيب $\frac{n}{2}=\frac{6}{2}=3$ و $\frac{n}{2}+1=4$.

البيانات مرتبة	2.5	2.5	5.4	7.1	8.3	9.2
الترتيب	1	2	3	4	5	6

القيمتان في المنتصف هما 5.4 و 7.1 ولذلك فإن: الوسيط = $\frac{5.4+7.1}{2}$ = 6.25

كيلوجراما.

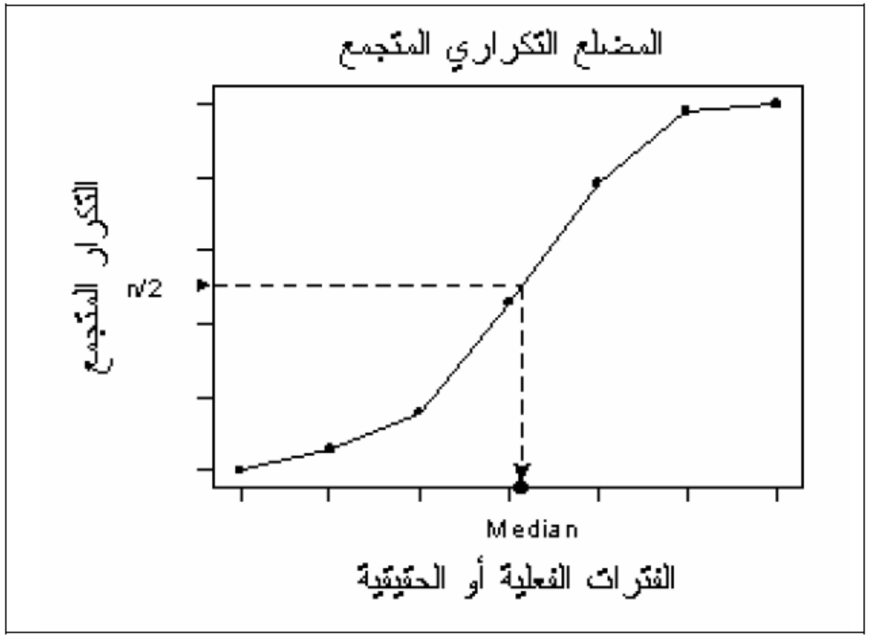
الوسيط للبيانات المبوبة

يمكن حساب الوسيط للبيانات الملخصة في جدول تكراري بطرقتين هما: طريقة حسابية وطريقة بيانية. ويستخدم الجدول التكراري المتجمع الصاعد لإيجاد الوسيط حسابيًا بينما يستخدم المـــضلع التكراري المتجمع الصاعد لإيجاد الوسيط بيانيًا. وفي حالة البيانات المبوبة فإننا نعرف ما يلي:

- رتبة (أو ترتيب) الوسيط = $\frac{n}{2}$ (سواءً كان عدد البيانات n زوجيًا أم فرديًا).
- الفترة الوسيطية = الفترة التي يقع فيها الوسيط

= أول فترة يزيد تكرارها المتجمع الصاعد عن $\frac{n}{2}$ أو يساويه

الوسيط للبيانات المبوبة



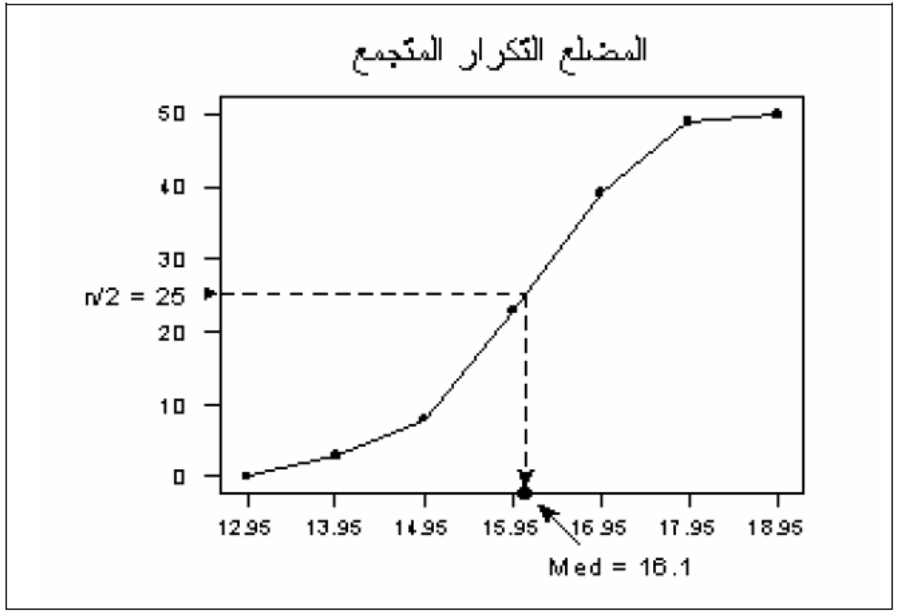
لإيجاد الوسيط بيانيًا نقوم بالخطوات التالية:

١. نرسم المضلع التكراري المتجمع الصاعد

٢. نحدد رتبة (أو ترتيب) الوسيط = $\frac{n}{2}$ (سواءً كان عدد البيانات فرديًا أو زوجيًا)

٣. في المضلع التكراري المتجمع الصاعد نحدد موقع رتبة الوسيط $\frac{n}{2}$ على المحور الرأسي (محور التكرار المتجمع) ومن ذلك الموقع نرسم خطًا أفقيًا يلتقي مع المضلع في نقطة. وعند نقطة الالتقاء نرسم عمودًا يتقاطع مع محور الفترات (المحور الأفقي) في نقطة. هذه النقطة هي قيمة الوسيط التقريبية. والشكل التالي يبين طريقة حساب الوسيط.

الوسيط للبيانات المبوبة



مثال

أوجد قيمة الوسيط بيانيًا لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

الحل:

نرسم المضلع التكراري المتجمع الصاعد كما مر معنا سابقًا ثم نطبق طريقة حساب الوسيط بيانيًا. مع ملاحظة أن رتبة الوسيط = $\frac{n}{2}=\frac{50}{2}=25$. باستخدام الشكل أدناه نجد أن القيمة التقريبية للوسيط هي:

الوسيط = 16.1.

مميزات وعيوب الوسيط

مميزات الوسيط:

إن الوسيط يعتبر من مقاييس النزعة المركزية الشائعة وذلك لما يتمتع به من بعض الصفات الجيدة. ومن مميزات الوسيط نذكر ما يلي:

١ـ الوسيط سهل التعريف والحساب.

٢ـ الوسيط وحيد لمجموعة البيانات الواحدة.

٣ـ الوسيط أقل تأثرا من المتوسط بالقيم الشاذة أو المتطرفة.

عيوب الوسيط :

بالرغم من أن الوسيط يعتبر من مقاييس النزعة المركزية الجيدة إلا أن له بعض العيوب نذكر منها ما يلي:

١ـ لا يأخذ الوسيط في الاعتبار جميع البيانات إذا أنه يعتمد فقط على القيم التي في المنتصف وعلى ترتيب البيانات بغض النظر عن قيمها.

٢ـ لا يمكن بشكل عام حساب الوسيط للبيانات الوصفية (النوعية).

ملاحظة:

وحدة الوسيط هي نفس وحدة البيانات الأصلية ـ فإذا كانت وحدة البيانات هي الكيلوجرام فإن وحدة الوسيط هي الكيلوجرام.

المنوال Mode

المنوال هو أحد مقاييس النزعة المركزية شائعة الاستخدام ولاسيما في حالة البيانات الوصفية (النوعية). ويعرف المنوال لمجموعة من البيانات على أنه تلك القيمة التي تتكرر أكثر من غيرها أي أنها القيمة ذات التكرار الأكبر (إن وجدت). يرمز للمنوال بالرمز (Mod) ومن تعريف المنوال تتضح لنا عدة أنواع من البيانات:

١- بيانات ليس لها منوال وتسمى عديمة المنوال.

٢- بيانات لها منوال واحد وتسمى وحيدة المنوال.

٣- بيانات لها أكثر من منوال وتسمى متعددة المنوال.

المنوال للبيانات المفردة (غير المبوبة)

المنوال = المشاهدة الأكثر تكراراً (إن وجدت)

مثال:

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس العمر (بالسنة) والوزن (بالكيلوجرام) والطول (بالسم) وفصيلة الدم. أوجد منوال للبيانات المختلفة.

رقم الشخص	1	2	3	4	5
العمر	25	20	25	30	35
الوزن	70	55	65	70	65
الطول	164	162	155	165	158
فصيلة الدم	O	A	B	A	AB

البيانات	المنوال	نوع البيانات بالنسبة للمنوال
العمر	25	وحيدة المنوال
الوزن	المنوال الأول = 65 المنوال الثاني = 70	متعددة المنوال (ثنائية المنوال)
الطول	لا يوجد	عديمة المنوال
فصيلة الدم	A	وحيدة المنوال

المنوال للبيانات المبوبة

- الفترة المنوالية هي الفترة ذات التكرار الأكبر وهي الفترة التي يقع فيها منوال .
- وفي الجدول التكراري قد يكون هناك فترة منوالية واحدة أو عدة فترات منوالية أو قد لا يوجد فترة منوالية.
- ويمكن حساب المنوال للبيانات الملخصة في جدول تكراري باستخدام المدرج التكراري

المنوال للبيانات المبوبة (الطريقة الحسابية)

لإيجاد المنوال حسابياً: **المنوال = مركز الفترة المنوالية**

مثال:

أوجد قيمة المنوال حسابياً لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم.

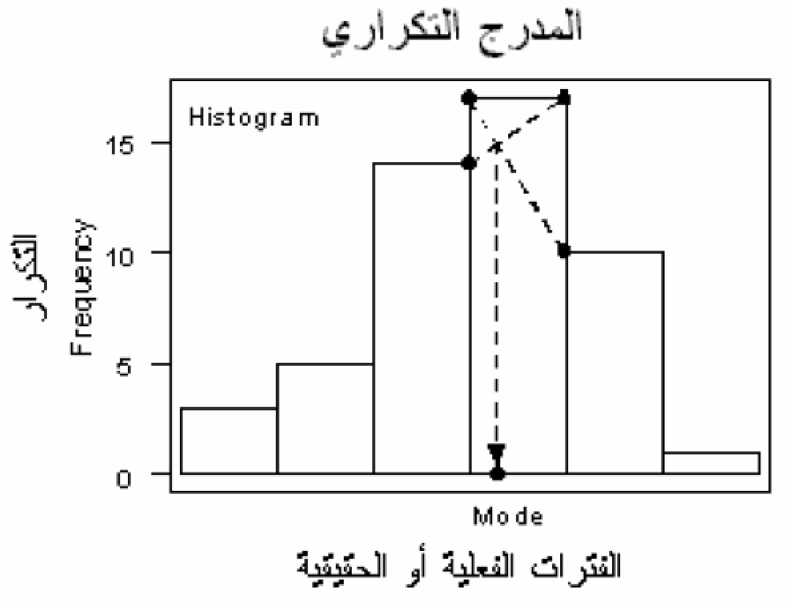
مستوى الهيموجلوبين	مركز الفترة	التكرار
12.95 – 13.95	13.45	3
13.95– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	**16.45**	**16**
16. 95– 17.95	17.45	10
17. 95– 18.95	18.45	1

الحل:

أكبر تكرار = ١٦ ← الفترة المنوالية هي: ١٥,٩٥ – ١٦,٩٥ ← المنوال = مركز الفترة المنوالية = ١٦,٤٥

المنوال للبيانات المبوبة (الطريقة البيانية)

- نستخدم المدرج التكراري لحساب المنوال .
- في المدرج التكراري، نحدد الفترة المنوالية وهي الفترة ذات التكرار الأكبر (المستطيل الأطول). بعد تحديد الفترة المنوالية نحدد الفترتين السابقة واللاحقة للفترة المنوالية . بعد ذلك نرسم خط مستقيم يصل القمة اليمنى لمستطيل الفترة المنوالية بالقمة اليمنى لمستطيل الفترة السابقة ونرسم خط مستقيم يصل القمة اليسرى لمستطيل الفترة المنوالية بالقمة اليسرى لمستطيل الفترة اللاحقة. وعند نقطة تقاطع الخطين نرسم عمود. القيمة على المحور السيني تمثل المنوال.

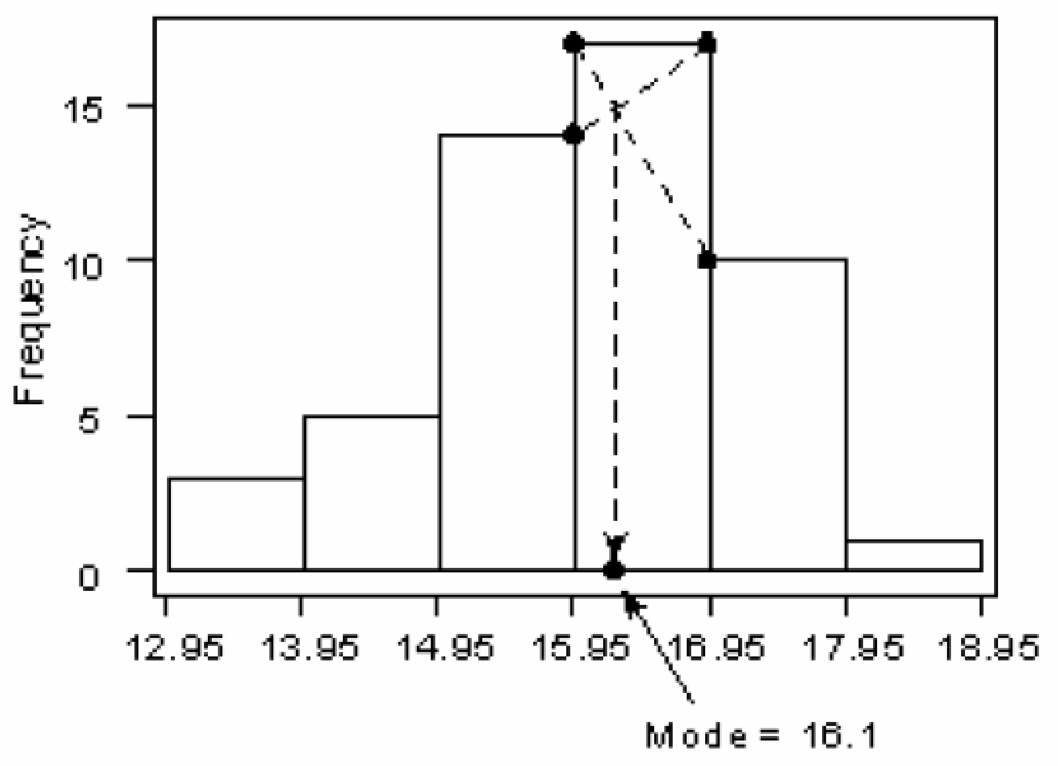


المنوال للبيانات المبوبة (الطريقة البيانية)

مثال:

أوجد قيمة ا لمنوال بيانياً لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم.

مستوى الهيموجلوبين	مركز الفترة	التكرار
12.95 – 13.95	13.45	3
13.95– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	**16.45**	**16**
16. 95– 17.95	17.45	10
17. 95– 18.95	18.45	1



الحل:

القيمة التقريبية للمنوال هي:

المنوال = ١٦,١

مميزات وعيوب المنوال

مميزات المنوال:

يعتبر المنوال من مقاييس النزعة المركزية الشائعة ومن مميزاته نذكر ما يلي:

١- المنوال سهل التعريف والحساب.

٢- المنوال أقل تأثراً من المتوسط بالقيم الشاذة أو المتطرفة.

٣- يمكن حساب المنوال للبيانات الكمية والوصفية (النوعية).

عيوب المنوال:

بالرغم من أن المنوال يعتبر من مقاييس النزعة المركزية الشائعة إلا أن له بعض العيوب نذكر منها ما يلي:

١- لا يأخذ المنوال في الاعتبار جميع البيانات إذا أنه يعتمد فقط على البيانات ذات التكرار الأكثر.

٢- قد لا يوجد منوال لمجموعة من البيانات أو قد يكون هناك أكثر من منوال.

ملاحظة:

وحدة المنوال هي نفس وحدة البيانات الأصلية . فإذا كانت وحدة البيانات هي الكيلوجرام فإن وحدة المنوال هي الكيلوجرام.