

مقاييس النزعة المركزية والتشتت

Measures of central tendency and dispersion

- ١ - مفهوم التجميع
- ٢ - مقاييس النزعة المركزية
- ٣ - مقاييس التشتت
- ٤ - تباين المتوسطات
- ٥ - معامل الاختلاف
- ٦ - العلاقة بين المدى والانحراف القياسي
- ٧ - استخدام برنامج SAS للحصول على الإحصاءات
التوصيفية
- ٨ - التشفير
- ٩ - الجداول التكرارية

obeikandi.com

يتطلب عرض البيانات في كثير من الأحيان ضرورة الحصول على أرقام توضح وتصف العشييرة أو العينة محل الدراسة بطريقة واضحة وليست مبهمه وتجعل منها شيئاً منفرداً يختلف عن العينات أو العشائر الأخرى. وكما ذكر سابقاً فإنه لا يمكن عمليا وصف كل الأفراد في عينة ما أو عشييرة ما، ولكن يتم الحصول من هذه البيانات على الأرقام التي توصف بها العينة (الإحصاءات statistics) والعشييرة (المعالم parameters). وغالبا ما يحتاج حساب هذه الأرقام إلى عمليات رياضية تتعامل مع البيانات الموجودة في العينة موضع الاعتبار والتي قد تشملها جميعاً أو قد تشمل، في بعض الأحيان، بعضاً منها فقط. وحتى يمكن التعامل مع البيانات بواسطة العمليات الحسابية فإنه يلزم الإلمام بمفهوم التجميع والذي غالباً ما يستعمل في الحسابات الإحصائية.

١-٣ مفهوم التجميع Summation notation

في علم الإحصاء كثيراً ما يتم جمع قيم البيانات أو مربعاتها ... الخ. فعند الرغبة في قياس متوسط أوزان مجموعة من العجول عند بدء تسمينها مثلاً فإنه بدلا من الإشارة إلى عملية التجميع عن طريق ذكرها حرفياً، فإن هناك طريقة متبعة لتوضيح ذلك باستعمال إشارات معينة لتدل على عملية التجميع وهي حرف \sum اليوناني وينطق سيجما (جدول ١٩ ملحق أ). وهذه الإشارة تعني التجميع، فإذا أريد مثلاً جمع أوزان 10 عجول فبدلاً من وضع ذلك في صورة $Y_1 + Y_2 + \dots + Y_{10}$ حيث Y_1 تمثل قيمة الوزن للمشاهدة الأولى و Y_2 تمثل قيمة الوزن للمشاهدة الثانية ... وهكذا، فإنه يمكن الإشارة إلى ذلك باستعمال إشارة التجميع $\sum_{i=1}^{10} Y_i$ حيث i تعتبر في هذه الحالة تمييزاً يشير إلى رقم أو ترتيب المشاهدات ويوضع تحت وفوق علامة التجميع المدى الذي يتم عليه ذلك بقيمتيه الدنيا والعليا، على الترتيب. ففي الحالة السابقة يتم التجميع من المشاهدة رقم $i=1$ إلى المشاهدة الأخيرة $i=10$.

مثال ١-٣

إذا كانت أوزان مجموعة عجول الفريزيان بالكيلوجرام عند بدء تسمينها هي كالتالي:

210, 180, 165, 215, 240, 195, 195, 158, 200, 160, 205 كج

فإن قيمة $\sum_{i=1}^{11} Y_i = 2123$ ، وهذا يمثل $210 + 180 + \dots + 160 = 2123$

وإذا افترض أن المطلوب هو تجميع القيم الثلاث الأولى فقط أى $i=1,2,3$ فإنه يشار إلى ذلك $\sum_{i=1}^3 Y_i$ وتصبح القيمة = 555 كج، أى $210+180+165 = 555$.

وعند افتراض أن عدد الأفراد أو المشاهدات فى عينة يساوى n وأريد التجميع

لكل أفراد العينة فإنه يشار إلى ذلك $\sum_{i=1}^n Y_i$ حيث تعنى i فى هذه الحالة $1,2,\dots,n$

ويمكن استعمال علامة التجميع فى تجميع دالات للمتغير Y حيث يمكن الحصول

على قيمة $\sum_{i=1}^n Y_i^2$ وهذه تعنى الحصول على مربع كل قيمة من قيم المتغير Y_i ثم يتم تجميع هذه القيم المربعة على المشاهدات من $i=1$ حتى $i=n$. ففى المثال السابق

$\sum_{i=1}^{i=11} Y_i^2 = 416189 = (160)^2 + \dots + (180)^2 + (210)^2$. ويمكن استعمال أى من الحروف كتمييز ولكن هناك شبه عرف بين الإحصائيين على استعمال الحروف اللاتينية البائدة من i ثم j ثم k وهكذا.

وفى كثير من الأحيان إذا كان التجميع على كل القيم الموجودة فى العينة فإنه يمكن التغاضى عن وضع المدى الذى يتم التجميع عليه ويكتفى بالإشارة إلى $\sum Y_i$ أو حتى يمكن إهمال التمييز كأن تستعمل $\sum Y$ فقط. وهو ما قد يشاهد فى خلال هذا الكتاب عند توضيح المعادلات أو إثباتها.

مثال ٢-٣

إذا كانت قيمة $Y_1 = 7$ ، $Y_2 = 3$ ، $Y_3 = 5$ فأوجد ما يلى:

$$\sum Y_i \quad (أ) \quad \sum_{i=1}^3 2Y_i^2 \quad (ب) \quad \sum_{i=2}^3 (Y_i - i) \quad (ج)$$

الحل:

$$\sum Y_i = Y_1 + Y_2 + Y_3 = 7 + 3 + 5 = 15 \quad (أ)$$

$$\sum_{i=1}^3 2Y_i^2 = 2(7)^2 + 2(3)^2 + 2(5)^2 = 98 + 18 + 50 = 166 \quad (ب)$$

$$\sum_{i=2}^3 (Y_i - i) = (3-2) + (5-3) = 3 \quad (ج)$$

ويمكن استعمال مفهوم التجميع أيضا للتعامل مع متغيرين بدلا من متغير واحد.

مثال ٣-٣

$$\text{إذا كانت قيم } Y_3 = 5, Y_2 = 3, Y_1 = 2$$

$$\text{وكانت قيم } X_3 = 5, X_2 = 2, X_1 = 4 \text{ فأوجد قيمة:}$$

$$\left(\sum_{i=2}^3 X_i \right) \left(\sum_{i=1}^2 Y_i^2 \right) \quad (\text{ب}) \quad , \quad \sum_{i=1}^3 X_i Y_i \quad (\text{أ})$$

الحل:

$$\sum_{i=1}^3 X_i Y_i = (4)(2) + (2)(3) + (5)(5) = 8 + 6 + 25 = 39 \quad (\text{أ})$$

$$\begin{aligned} \left(\sum_{i=2}^3 X_i \right) \left(\sum_{i=1}^2 Y_i^2 \right) &= (X_2 + X_3)(Y_1^2 + Y_2^2) = (2 + 5)(2^2 + 3^2) \\ &= (7)(4 + 9) = (7)(13) = 91 \quad (\text{ب}) \end{aligned}$$

وهناك بعض القواعد الأساسية للتعامل مع مفهوم التجميع وهي:

قاعدة ٣-١

تجميع حاصل جمع متغيرين أو أكثر يساوى حاصل جمع تجميعاتهما وعلى ذلك:

$$\sum_{i=1}^n (X_i + Y_i + Z_i) = \sum_{i=1}^n X_i + \sum_{i=1}^n Y_i + \sum_{i=1}^n Z_i$$

قاعدة ٣-٢

إذا كان ث (C) ثابتا فإن مجموع حاصل ضرب المتغير فى الثابت يساوى قيمة الثابت فى مجموع قيم المتغير أى:

$$\sum_{i=1}^n C Y_i = C \sum_{i=1}^n Y_i$$

قاعدة ٣-٣

إذا كان ث (C) ثابتاً فإن تجميع الثابت من 1 إلى n يساوي حاصل ضرب n في قيمة الثابت أي:

$$\sum_{i=1}^n C = nC$$

مثال ٣-٤

إذا كانت قيمة $X_2 = 4$ ، $X_1 = 2$ ، $Y_2 = -1$ ، $Y_1 = 3$

$$\sum_{i=1}^2 (3X_i - Y_i + 4)$$

فأوجد قيمة

الحل:

$$\begin{aligned} \sum_{i=1}^2 (3X_i - Y_i + 4) &= \sum_{i=1}^2 3X_i - \sum_{i=1}^2 Y_i + \sum_{i=1}^2 4 \\ &= 3 \sum_{i=1}^2 X_i - \sum_{i=1}^2 Y_i + (2)(4) \\ &= (3)(2+4) - (3-1) + 8 = 24 \end{aligned}$$

٢-٣ مقاييس النزعة المركزية "التمركز" Measures of central tendency

الجدول أو الرسوم البيانية يمكن أن توضح شكلاً عاماً لتوزيع متغير ما وتركيز قيم بيانات هذا المتغير. إلا أنه يوجد مقاييس عددية تدل على توزيع المتغير بطريقة كمية دون الحاجة إلى الرجوع إلى الرسوم البيانية.

تعريف ٣-١

أي مقياس يوضح مركز center أو موقع location مجموعة من البيانات مرتبة حسب قيمها يسمى بمقياس للمركزية. وأهم هذه المقاييس وأكثرها استخداماً وشيوعاً هو المتوسط الحسابي average أو arithmetic mean.

تعريف ٣-٢

إذا كانت مجموعة البيانات Y_1, Y_2, \dots, Y_N تمثل عشيرة حجمها N مشاهدة فإن متوسط العشيرة (population mean) (μ) يحسب من المعادلة:

$$\mu = \frac{\sum_{i=1}^N Y_i}{N} \quad (1-3)$$

ولكن في أغلب الأحيان فإنه من غير الممكن أو العملي قياس كل البيانات المتعلقة بعشيرة ما حجمها N وبالتالي يتم الحصول على عينة محددة العدد، ومن هذه العينة يتم حساب متوسط العينة.

تعريف 3-3

إذا كانت البيانات Y_1, Y_2, \dots, Y_n تمثل عينة محدودة حجمها n فإن متوسط العينة \bar{Y} sample mean يمكن حسابه من المعادلة:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \quad (2-3)$$

فإذا كانت العينة ممثلة تمثيلاً جيداً للعشيرة المأخوذة منها، فإن متوسط العينة يعتبر ممثلاً (تقديراً لنقطة point estimate) لمتوسط العشيرة.

مثال 3-5

في البيانات المذكورة في مثال 3-1 احسب متوسط العينة والتي حجمها $n = 11$

الحل:

$$\bar{Y} = \frac{\sum_{i=1}^{11} \bar{Y}}{n} = (210 + 180 + \dots + 160) / 11 = 193 \text{ kg}$$

ومن الملاحظ هنا أن جميع القيم الموجودة في العينة قد استخدمت في حساب المتوسط ومعنى هذا أن المتوسط تقدير شامل ويسمى هذا التقدير بأنه تقدير كاف sufficient من الناحية الرياضية.

وخاصيتان هامتان للمتوسط هما:

• مجموع انحرافات قيم المشاهدات عن متوسطها يساوى صفرا أى

$$\sum (Y - \bar{Y}) = 0$$

• مجموع مربعات فروق المشاهدات عن متوسطها هو أدنى ما يمكن، أى أن القيمة $\sum (Y - \bar{Y})^2$ تكون أدنى ما يمكن عما لو استخدمت أى قيمة أخرى غير \bar{Y} . وتعرف هذه الخاصية بـ "المربعات الصغرى least squares".

وهناك مقاييس أخرى للنزعة المركزية يكون استعمالها أقل من المتوسط خاصة عند معالجة البيانات البيولوجية أو الاقتصادية أو الاجتماعية، ومع ذلك ففي حالات كثيرة فإن تلك المقاييس قد تصف العشيرة أو العينة بطريقة أكثر تفهماً أو أكثر واقعية.

تعريف ٣-٤

الوسيط (الوسط) median لمجموعة من البيانات المرتبة حسب قيمها هو القيمة الوسطية (فى حالة وجود عدد فردى odd للملاحظات) أو المتوسط الحسابى للقيمتين الوسيطتين (فى حالة وجود عدد زوجى even للملاحظات).

وعلى ذلك فإن الوسط هو القيمة التى تقسم توزيع البيانات المرتبة تنازليا أو تصاعدياً إلى قسمين، فهو يقسم التوزيع بحيث إن نصف القيم يقع قبله والنصف الآخر يقع بعده.

تعريف ٣-٥

المنوال mode لمجموعة من البيانات هو تلك القيمة التى تظهر أكثر عدداً من المرات أو لها أعلى تكرار. وقد يوجد مجموعة من البيانات ليس لها منوال بمعنى عدم وجود أى قيمة تتكرر أكثر من مرة واحدة أو قد يكون هناك أكثر من قيمة لها نفس التكرار وفى هذه الحالة يعتبر التوزيع ذا منوالين bimodal أو أكثر multimodal.

مثال ٣-٦

احسب كل من الوسيط والمنوال للبيانات المذكورة فى المثال ٣-١

الحل:

إذا رتبنا البيانات فى العينة ترتيباً تنازلياً أو تصاعدياً فإن العينة تبدو كالتالى:

240, 215, 210, 205, 200, 195, 195, 180, 165, 160, 158

فإن قيمة الوسيط 195 kg وهي التي تقسم البيانات إلى قسمين متساويين أن هناك 5 قيم أكبر منها وخمس قيم أخرى أقل منها.

أما قيمة المنوال فهي أيضاً 195 kg حيث إنها هي القيمة الوحيدة التي تكررت في العينة مرتين.

ومن الملاحظ في المثال السابق أن قيمتي الوسيط والمنوال متساويتان وتقتربان من قيمة المتوسط الحسابي. وتعتبر هذه الظاهرة من مميزات البيانات التي تتبع في توزيعها التوزيع الطبيعي normal distribution.

وتلخيصاً لما سبق فإنه يمكن المقارنة بين المقاييس الثلاثة وهي المتوسط والوسيط والمنوال، فالمتوسط هو أكثر المقاييس استخداماً في الإحصاء خاصة في معالجة البيانات البيولوجية. ويرجع ذلك إلى أن توزيع متوسطات العينات عادة ما يكون معروفاً من الناحية الرياضية وبالتالي فإن الطرق المستخدمة في الاستدلال الإحصائي تكون المبنية على أساس المتوسط الحسابي وبالتالي يمكن معرفة مقدار الخطأ في التقديرات وتوزيعه. وكما سبق القول فإن المتوسط يعتبر تقديراً لمتوسط العشرة الأكثر تمثيلاً لها وذلك لأنه يستعمل كل البيانات في العينة بالإضافة إلى أن المتوسط يتمتع بأقل قدر من الخطأ (التباين) minimum variance. أما ما يعاب على المتوسط فهو أنه يتأثر تأثيراً كبيراً بالقيم المتطرفة (أى العالية جداً أو المنخفضة جداً outliers)، وذلك لدخول مثل هذه القيم في حسابه. أما الوسيط فهو سهل في حسابه ولا يتأثر بالقيم غير الطبيعية وبالتالي يعطى فكرة أوضح عن غالبية القيم في العينة عن المتوسط، ولكن يعاب عليه طبعاً أنه لا يستعمل كل بيانات العينة فهو غير كاف، وأيضاً فإن تباين قيمة العينات المختلفة يكون أكبر من حالة المتوسط. وعلى ذلك فإن المتوسط يكون أكثر ثباتاً من الوسيط وبالتالي فالمتوسط يمثل متوسط العشرة بقدر أكبر من الدقة أما المنوال فهو أقل المقاييس كفاءة وهو يكون غير ذي نفع في حالة العينات صغيرة الحجم. وفي بعض الحالات فإن المنوال قد لا يوجد أو يكون متعدد. إلا أنه في حالة البيانات العديدة فيكون له نفع خاص في البيانات العديدة مثل طول دورة الشبق وعدد الخلفه في البطن في بعض أنواع الحيوانات مثله مثل الوسيط في ذلك.

وهناك مقاييس أخرى للمركزية يمكن تعريفها هنا في هذا المجال وهي أقل استخداماً من سابقتها، مثال ذلك متوسطات الأرباع quartiles، الأعشار deciles والمئويات percentiles وهي تلك النقاط التي تقسم التوزيع إلى أرباع أو أعشار أو مئويات على التوالي. وعلى ذلك فيمكن القول بأن الوسيط يمثل الربع الثاني والعشر الخامس والمئوية الخمسين.

تعريف ٦-٣

المتوسط الهندسي لعدد (n) من القيم هو الجذر النوني لحاصل ضرب هذه القيم
أى:

$$G = \sqrt[n]{Y_1 \cdot Y_2 \cdots Y_n} = (Y_1 \cdot Y_2 \cdots Y_n)^{1/n} \quad (٣-٣)$$

(ملاحظة: بدلاً من بيان الضرب تفصيلاً يمكن التعبير عن هذه العملية باستخدام πY_i حيث π هي من الحروف اليونانية وينطق باي).

وبالنسبة للبيانات العديّة count data فإنه يمكن حساب ما يعرف بالمتوسط التوافقي harmonic mean وهو يستخدم في حالة إيجاد متوسط عدد المشاهدات في الفئات المختلفة أو التقسيمات المختلفة subclasses.

تعريف ٧-٣

المتوسط التوافقي لعدد (n) من القيم هو مقلوب متوسط مقلوب القيم في كل فئة (يرمز له بالرمز H) أى:

$$\frac{1}{H} = \frac{1}{n} \sum_i \left(\frac{1}{Y_i} \right) \quad (٤-٣)$$

وتستعمل هذه المتوسطات أيضاً في حساب متوسطات النسب ratios والمعدلات rates.

مثال ٧-٣

احسب المتوسط الهندسي للقيم التالية والتي تمثل معدل الزيادة العددية في بعض القطعان.

$$1.61, 1.4, 1.5, 1.27, 1.22, 1.25, 1.3, 1.2$$

الحل:

$$G = \sqrt[8]{(1.2)(1.3)(1.25)(1.22)(1.27)(1.5)(1.4)(1.61)} \\ = 1.337$$

مقارنة بالمتوسط الحسابي وهو 1.344

مثال ٣-٨

احسب متوسط عدد العجول في عمر سنة الموجودة في ثمانية قطعان من البيانات التالية:

81, 70, 75, 64, 61, 63, 65, 60

الحل:

$$H = 1 \div \frac{1}{H} = 1 \div \frac{1}{8} \left(\frac{1}{60} + \frac{1}{65} + \frac{1}{63} + \frac{1}{61} + \frac{1}{64} + \frac{1}{75} + \frac{1}{70} + \frac{1}{81} \right)$$

$$= 1 \div \frac{1}{8} (0.1199) = 66.72 \quad \text{عجل}$$

ويعتبر المتوسط التوافقي في حالة الأعداد أكثر تمثيلاً مقارنةً بالمتوسط الحسابي لنفس البيانات والذي قيمته 67.38 عجل.

صندوق ٣-١

مقاييس النزعة المركزية:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \quad \text{المتوسط :}$$

الوسيط (الوسط): القيمة الوسطية في العينة

المنوال: القيمة الأكثر تكراراً في العينة

$$G = \sqrt[n]{Y_1 \cdot Y_2 \cdots Y_n} = (Y_1 \cdot Y_2 \cdots Y_n)^{1/n} \quad \text{المتوسط الهندسي :}$$

$$H = 1 \div \frac{1}{n} \sum_i \left(\frac{1}{Y_i} \right) \quad \text{المتوسط التوافقي :}$$

٣-٣ مقاييس التشتت (التباين) Measures of dispersion or variation

إن مقاييس المركزية لا توضح مدى توزيع البيانات حول القيم المركزية للتوزيع. فقد تتماثل مجموعة من القيم في قيمة متوسطها أو منوالها ومع ذلك تختلف اختلافاً بيناً في توزيع القيم حول المتوسط.

مثال ٣-٩

إذا أخذت ثلاث مجموعات من الأرقام التالية:

(أ) 30, 31, 32, 33, 34

(ب) 34, 32, 28, 29, 37

(ج) 48, 45, 32, 20, 15

لاحظ أن كل من المجموعات الثلاث لها متوسط يساوي 32 في حين أن القيم في المجموعة (أ) كلها تقترب من القيمة المركزية (المتوسط) للمجموعة حيث أكبر شذوذ أو انحراف عن هذه القيمة هو وحدتان بالموجب أو بالسالب في حين أنه في المجموعة (ب) هناك انحرافات أكبر حيث تبعد إحدى القيم عن القيمة المتوسطة بخمس وحدات. أما بالنسبة للمجموعة الأخيرة (ج) فإنها أكثرها تشتتاً حيث بلغ انحراف إحدى القيم عن المتوسط 16 وحدة. وهكذا فإنه لا يمكن تحديد أى شكل للتوزيع لقيم المتغير محل الدراسة بمعرفة المتوسط فقط.

وكما أن هناك عدداً من المقاييس استخدمت لقياس مركزية (أو موقع) التوزيع فإنه يوجد أيضاً أكثر من مقياس لتشتت قيم المتغير أو تباينها حول مركزها.

تعريف ٣-٨

المدى range عبارة عن الفرق بين أعلى قيمة وأقل قيمة يأخذها المتغير. وهو يدل على مدى ابتعاد القيم عن المتوسط، وطبيعي أنه كلما شذت القيم عن مركزها كلما اتسع المدى. وفي المثال (٣-٩) كان المدى للمجموعة (أ) $4 = 34 - 30$ وحدات في حين أنه في المجموعة (ج) كان المدى فيها $33 = 48 - 15$ وحدة.

والمدى يعتبر مقياساً سريعاً يمكن منه أخذ فكرة مبدئية عن التشتت في العينة المدروسة ولكن طبعاً يعاب عليه أنه غير دقيق وأنه لا يستخدم في حسابه إلا على أعلى وأقل القيم في التوزيع. وبذا فهو غير ذى فائدة تقريباً في حالة العينات أو العنصر الكبيرة حيث يتأثر بالقيم الشاذة تأثراً كبيراً.

تعريف ٣-٩

متوسط الانحرافات average deviation لمجموعة من البيانات عددها (N) هو مجموع الانحرافات المطلقة عن المتوسط مقسوما على العدد (N) أى:

$$\frac{\sum_{i=1}^N |Y_i - \mu|}{N} \quad (٥-٣)$$

مثال: ٣-١٠

احسب متوسط الانحرافات لكل من العينتين أ ، ج فى المثال (٣-٩)

الحل:

المتوسط فى كل من العينتين يساوى 32 وحدة وعلى ذلك يكون:

متوسط الانحرافات فى (أ)

$$\frac{|-2| + |-1| + \text{صفر} + |1| + |2|}{5} = \frac{6}{5} = 1.2$$

متوسط الانحرافات فى (ج)

$$\frac{|16| + |13| + \text{صفر} + |-12| + |-17|}{5} = \frac{58}{5} = 11.6$$

وعلى ذلك فإنه كلما كانت القيم فى البيانات أكثر انحرافاً عن القيمة الوسطية فإن متوسط الانحرافات يزداد. يعاب على هذا القياس أن معالجته الرياضية تصبح غير سهلة ولكنه لا يضخم من قيمة الانحرافات الشاذة التى تحدث فى القيم.

تعريف ٣-١٠

تباين قيم متغير عشوائى هو متوسط مربع انحرافات قيم المتغير عن متوسط العشيرة ويرمز له بالرمز σ^2

$$\sigma_Y^2 = \frac{\sum_{i=1}^N (Y_i - \mu)^2}{N} \quad (6-3)$$

حيث N هي عدد القيم في العشرة للمتغير Y و μ هي متوسط هذه العشرة.

تعريف ١١-٣

الانحراف القياسي (المعياري) standard deviation لقيم متغير عشوائي هو الجذر التربيعي لتباين قيم المتغير.

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^N (Y_i - \mu)^2}{N}} \quad (7-3)$$

حيث N هي عدد القيم في العشرة للمتغير Y و μ هي متوسط هذه العشرة.

ويعتبر الانحراف القياسي من أكفأ مقاييس التشتت وهو الجذر التربيعي لتباين القيم. والتباين يعطى قدراً أكبر من الأهمية لتلك القيم الشاذة عن طريق تربيع الانحرافات عن المتوسط.

مثال ١١-٣

إذا كانت القيم المذكورة في مثال ٣-٩ تمثل 3 عشائر أ، ب، ج والمطلوب حساب التباين والانحراف القياسي لكل منها.

الحل:

بما أن المتوسط في كل عشيرة يساوي 32 فإنه بالتالي يكون:

التباين في العشرة (أ):

$$\sigma_Y^2 = \frac{(34-32)^2 + (33-32)^2 + (32-32)^2 + (31-32)^2 + (30-32)^2}{5} = \frac{10}{5} = 2$$

$$\sigma_Y = \sqrt{2} = 1.414 \quad \text{والانحراف القياسي:}$$

$$\sigma_Y^2 = \frac{54}{5} = 10.8 \quad \text{التباين في العشرة (ب):}$$

والانحراف القياسى: $\sigma_Y = 3.286$

التباين فى العشرة (ج): $\sigma_Y^2 = \frac{858}{5} = 171.6$

والانحراف القياسى: $\sigma_Y = 13.1$

ويتضح من المثال السابق أن هناك دلالة على أن العشائر الثلاث مختلفة عن بعضها حيث إن قيمة التباين فيها مختلفة خاصة العشرة (ج).

وبما أنه فى أغلب الأحيان لا يكون التعامل مع العشائر إنما مع عينات مأخوذة من هذه العشائر فإن المقياس للتباين والانحراف القياسى المحسوب من العينة سوف يصبح تقديراً للمعلم فى العشرة. ولكن يلاحظ هنا أنه يجب فى هذه الحالة حساب المتوسط

من العينة \bar{Y} حيث إنه تقديراً لمتوسط العشرة غير المعروف μ . وعلى هذا الأساس يمكن القول بأن التقدير للتباين المحسوب من العينة المسحوبة من العشرة يصبح تقديراً غير منحاز لقيمتة فى العشرة ولكنه نظراً لأن المتوسط يحسب من العينة فإنه للحصول على متوسط مربع الانحرافات تكون القسمة على عدد البيانات ناقص واحد أى $n-1$ حيث n هى عدد البيانات فى العينة ويرمز للتقدير المتحصل عليه وهو تباين العينة بالرمز S^2 والانحراف القياسى S . وبالتالي فإن:

$$S_Y^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1} \quad (٨-٣)$$

$$S_Y = \sqrt{S_Y^2} \quad (٩-٣)$$

وتعرف الكمية $(n-1)$ باسم درجات الحرية degrees of freedom، وهذا مفهوم يوضح أنه عند حساب المتوسط من العينة تصبح هناك قيمة واحدة فى العينة تحدد من معرفة بقية القيم الأخرى والتى عددها $n-1$ بالإضافة إلى المتوسط المحسوب وذلك من المعادلة (٢-٣) حيث:

$$\sum Y_i = n \bar{Y} \quad (١٠-٣)$$

وفى هذا المجال فإن استخدام المعادلة (٨-٣) لحساب S^2 من العينة قد تكون غير ملائمة للحساب خاصة فى حالة وجود كسور فى قيمة المتوسط نتيجة للقسمة على

عدد المشاهدات وبالتالي فإنه عند تربيع الانحرافات تصبح صعوبة التعامل وقد يكون هناك خطأ حسابي نتيجة للتقريب في حساب مربع الانحرافات. وهنا يمكن استخدام المعادلة الحسابية التي تساوى المعادلة التعريفية حيث يمكن إثبات أن:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - \frac{(\sum Y_i)^2}{n} \quad (11-3)$$

الإثبات:

$$\begin{aligned} \sum (Y_i - \bar{Y})^2 &= \sum Y_i^2 - 2\bar{Y}\sum Y_i + \sum \bar{Y}^2 \\ &= \sum Y_i^2 - 2n\bar{Y}^2 + n\bar{Y}^2 \\ &= \sum Y_i^2 - n\bar{Y}^2 \\ &= \sum Y_i^2 - \frac{n(\sum Y_i)^2}{n^2} \\ &= \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \end{aligned}$$

وهنا تجدر الإشارة إلى أن الكمية $\frac{(\sum Y_i)^2}{n}$ والتي تساوى $n\bar{Y}^2$ تعرف بمعامل التصحيح (CF) correction factor وهو يصحح مجموع المربعات غير المصحح $\sum Y_i^2$ لقيمة المتوسط حيث يمكن اعتباره أنه في كل مرة في عملية التجميع كان يجب طرح الكمية $n\bar{Y}^2$ وتكرر n من المرات. وعموماً فإن استعمال المعادلة (11-3) لحساب مجموع المربعات المصحح $\sum y_i^2 = \sum (Y - \bar{Y})^2$ يكون ممكناً في حالة البيانات قليلة العدد والتي تكون قيمتها صغيرة وقيمة المتوسط فيها غير كسرية في حين أن المعادلة (11-3)، وهي الأكثر استعمالاً وشيوعاً، هي أوفى أيضاً عند الحساب واستخدام الآلات الحاسبة. ويلاحظ أن بعض الآلات الحاسبة الصغيرة الإلكترونية حالياً تجرى حساباتها عن هذا الطريق وهذه الآلات تمت برمجتها بحيث تعطى كل من قيم $\sum Y_i$ ، $\sum Y_i^2$ ، n ومن خلال هذه القيم يمكن حساب المتوسط \bar{Y} ومجموع المربعات المصحح $\sum y_i^2$ ثم حساب التباين والانحراف القياسى للعينة والعشيرة.

المطلوب حساب تقدير لقيمة المتوسط والانحراف القياسي بالكيلوجرام لوزن الميلاد في العينة التالية والتي عددها 16 عجل من عجول الفريزيان:

كج 26, 26, 21, 24, 25, 24, 22, 26, 28, 22, 35, 24, 23, 22, 28, 26

الحل:

$$\sum Y_i = 26 + 28 + \dots + 26 = 402$$

متوسط العينة:

$$\bar{Y} = \frac{\sum Y}{n} = \frac{402}{16} = 25.125 \text{ kg}$$

مجموع المربعات غير المصحح: $\sum Y_i^2 = 26^2 + 28^2 + \dots + 26^2 = 10272$
معامل التصحيح:

$$CF = \frac{(\sum Y_i)^2}{n} = \frac{(402)^2}{16} = \frac{161604}{16} = 10100.25$$

مجموع المربعات المصحح:

$$\begin{aligned} \sum y^2 &= \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \\ &= \sum Y^2 - CF = 10272 - 10100.25 = 171.75 \end{aligned}$$

تباين العينة:

$$S^2 = \frac{\sum y^2}{n-1} = \frac{171.75}{16-1} = 11.45 \text{ kg}^2$$

$$S = \sqrt{S^2} = \sqrt{11.45} = 3.38 \text{ kg}$$

الانحراف القياسي:

لاحظ أن مجموع المربعات المصحح والذي أخذ فيه متوسط العينة كأساس لحساب الانحرافات يعتبر أقل مجموع مربعات وهو ما يعرف بالـ least squares ومعنى ذلك

أنه إذا استبدلت قيمة المتوسط بأى قيمة أخرى فإن قيمة مجموع مربعات الانحرافات عن هذه القيمة يكون أكبر من تلك المحسوبة على أساس المتوسط كما أنه واضح أن مجموع الانحرافات عن المتوسط $\sum (Y - \bar{Y})$ يساوى الصفر. ويلاحظ أن كلاً من المتوسط والانحراف القياسى أخذ نفس وحدة قياس مفردات العينة، وهذه إحدى مميزات الانحراف القياسى، بينما التباين يأخذ مربع وحدة القياس.

٣-٤ تباين المتوسطات Variance of means

كما سبق إيضاحه فإن متوسط العشييرة μ عادة ما يكون مجهولاً ويتحصل على تقدير له أو عدة تقديرات من العينات المأخوذة من العشييرة وبالتالي يكون هناك متوسط \bar{Y} من كل عينة تدرس. ويلاحظ أن هذه المتوسطات قد تختلف فيما بينها، وهى عادة ما تختلف، ولكن متوسطها يتوقع أن يساوى متوسط العشييرة وتكون المتوسطات للعينات المختلفة والتي كل منها حجمه n من الأفراد موزعة توزيعاً طبيعياً حول متوسط العشييرة. وهناك علاقة تربط بين تباين القيم فى العشييرة وتباين المتوسطات من العينات. ويرمز لتباين متوسطات العينات بالرمز $\sigma_{\bar{Y}}^2$ وهو يساوى $\frac{\sigma^2}{n}$ والقاعدة التالية تستخدم لإثبات هذه العلاقة.

قاعدة ٣-٤

$$V(aY) = a^2 V(Y)$$

حيث يشير الرمز V للتباين variance.

وحيث إن المتوسط عبارة عن $\bar{Y} = \frac{\sum Y}{n}$ فإن:

$$\begin{aligned} V(\bar{Y}) &= V\left(\frac{\sum Y_i}{n}\right) \\ &= V\left(\frac{1}{n} \sum Y_i\right) = \frac{1}{n^2} V(Y_1 + Y_2 + \dots + Y_n) \end{aligned}$$

وبما أن تباين كل القيم المفردة فى العشييرة يتوقع أن تتساوى فإن:

$$V\left(\bar{Y}\right) = \frac{1}{n^2} [nV(Y_i) + n(n-1)\text{Cov}(Y_i, Y_i')] \quad (12-3)$$

وحيث إن قيم Y_i أخذت كلها عشوائياً فإن

$$V\left(\bar{Y}\right) = \frac{1}{n^2} [nV(Y_i)] \quad , \text{Cov}(Y_i, Y_i') = 0$$

وبما أن قيمة $\sigma^2 = V(Y_i)$ هو تباين القيم في العشرة فتصبح المعادلة:

$$V(\bar{Y}) = \frac{\sigma^2}{n} \quad (13-3)$$

وبالتالى فإن الانحراف القياسى للمتوسط standard deviation of mean والذي يعرف بالخطأ القياسى standard error، ويختصر SE، عبارة عن:

$$\frac{\sigma}{\sqrt{n}} \text{ أو } \sqrt{\frac{\sigma^2}{n}} \quad (14-3)$$

لاحظ أن الخطأ القياسى يتناسب عكسياً مع حجم العينة أو العينات المقدر منها المتوسط، فكلما ازداد عدد أفراد العينة كلما انخفضت قيمة الخطأ القياسى. ففي المثال 12-3 كان حجم العينة 16 مشاهدة والتباين $S^2 = 11.45 \text{ kg}^2$ فيكون تباين المتوسط

$$\sigma_{\bar{Y}} = \sqrt{\frac{11.45}{16}} = 0.846 \text{ kg} \quad \text{وقيمة الخطأ القياسى } \sigma_{\bar{Y}}^2 = \frac{11.45}{16} = 0.716 \text{ kg}^2$$

ولا يستوجب حساب تباين المتوسط أو الخطأ القياسى وجود أكثر من متوسط واحد للعينة. والخطأ القياسى له مفهوم واضح حيث يوضح مدى الخطأ الذى قد يحدث كنتيجة لاستخدام المتوسط المحسوب من العينة كتقدير لمتوسط العشرة. ومن الطبيعى أنه بزيادة حجم العينة تزداد الثقة فى التقدير. وسوف نشير فى مواضع لاحقة إلى استخدامات الخطأ القياسى خاصة فى اختبارات الفروض الإحصائية لدراسة الفروق بين متوسطات المعاملات وجميعها يستخدم فيها الخطأ القياسى أو قيم أخرى بمعلوماته.

٣-٥ معامل الاختلاف أو معامل التباين Coefficient of variation

وهو مقياس آخر للتشتت ويستعمل كثيراً عند مقارنة نتائج تجارب مختلفة على نفس الصفة أو لمقارنة صفات مختلفة في مقدار تباينها ومتوسطاتها أو مختلفة في وحدات القياس المستخدمة ويرمز له بالرمز C.V.

معامل الاختلاف لصفة ما أو لتقدير ما هو النسبة المئوية للانحراف القياسي مقسوماً على متوسط العينة.

$$C.V. = \frac{S}{\bar{Y}} \times 100 \quad (١٥-٣)$$

ومن الاستعمالات المهمة لمعامل التباين هو توضيح فيما إذا كانت النتائج المتحصل عليها من عينة ما تعطي نتائج تتماشى مع ما هو معروف عن تباين تلك الصفة في العشيرة حيث إن هذا المقياس مستقل عن وحدات القياس. وبالتالي يمكن أيضاً من مقارنة صفة ما مفاصة على فترات زمنية مختلفة وتوضيح مدى تباينها. فمثلاً قد يراد معرفة ما إذا كانت الاختلافات في الوزن عند الميلاد تتماشى مع تلك التي في الوزن عند الفطام حيث من المعروف أن النسبة من الأوزان سواء الخفيفة جداً أو الثقيلة جداً يحدث فيها نفوق متزايد وبالتالي فقد يكون معامل الاختلاف أعلى عند الميلاد عنه عند الفطام وقد تدخل أسباب أخرى بيئية في إحداث زيادة أو نقصاً في التباين منسوباً إلى المتوسط.

ومعامل الاختلاف لوزن العجول عند الميلاد في مثال ٣-١٢

$$C.V. = \frac{3.38}{25.125} \times 100 = 13.5\%$$

وفي بعض الحالات يفيد معامل الاختلاف في دراسة العلاقة بين الانحراف القياسي والمتوسط عندما يلاحظ أن زيادة التباين تكون مواكبة لزيادة المتوسط في حالة دراسة الصفة على مراحل زمنية متعاقبة. ويفيد معامل الاختلاف في فهم طبيعة البيانات التي يدرسها المجرّب حيث إنه إذا كان معامل الاختلاف أعلى من معدله الطبيعي فإن ذلك يدل على أنه قد يوجد شيء غير طبيعي بالنسبة لمجموعة البيانات تحت الدراسة. ويجدر الملاحظة أن معامل الاختلاف ولو أنه مفيد إلا أنه لا يغني عن معرفة كل من المتوسط والانحراف القياسي حيث إن تغيير معامل الاختلاف، كأى نسبة، لا يحدد سبب هذا التغيير فيما إذا كان تغييراً في بسط النسبة أو في مقامها.

صندوق ٢-٣

مقاييس التشتت:

المدى: أعلى قيمة - أقل قيمة

متوسط الانحرافات المطلقة: $\frac{1}{n} \sum |Y - \bar{Y}|$ الانحراف المعياري للعشيرة: $\sigma_Y = \sqrt{\frac{\sum (Y - \mu)^2}{N}}$ الانحراف المعياري للعينة: $S_Y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n-1}}$ الخطأ القياسي (الانحراف المعياري للمتوسط): $S_{\bar{Y}} = \sqrt{\frac{S^2}{n}}$ معامل الاختلاف: $C.V. = \frac{S}{\bar{Y}} \times 100$

٢-٣ العلاقة بين المدى والانحراف القياسي

تجدر الإشارة إلى أن هناك مقاييس سريعة يمكن عن طريقها الحصول على فكرة سريعة عن التباين في العينات خاصة الصغيرة منها. فهناك علاقة بين المدى وبين الانحراف القياسي في العينة. وهذه العلاقة تكون أعلى ما يمكن عندما يكون حجم العينة يساوي 2، ففي هذه الحالة فإن المدى يساوي الانحراف القياسي فعلاً وكلما زاد حجم العينة كلما قلت كفاءة المدى كمقياس للانحراف القياسي وأيضاً كلما قلت العلاقة بين المدى والانحراف القياسي. وهناك جداول توضح هذه العلاقة العكسية كنسبة بين المدى والانحراف القياسي عند الأحجام المختلفة للعينة (جدول ٢ ملحق أ).

مثال ٣-١٣

في مثال ٣-١٢ كانت $n=16$ ، ومن جدول ٢ ملحق أ فإن النسبة بين الانحراف القياسي والمدى، $\sigma/R = 0.283$ ، وهذا يعنى أنه يمكن معرفة حجم σ بصورة مبدئية باستخدام هذه العلاقة. والمدى في هذه الحالة كان من 21 إلى 35 أى 14. وبالتالي فإن $\sigma = 14 \times 0.283 = 3.96 \text{ kg}$. وهى لا تختلف كثيراً عن قيمة σ المحسوبة من العينة S والتي تساوى 3.38 كج. وميزة هذه العلاقة هى فى سرعة الحصول على تقدير لحجم التباين فى العينة حيث إن المدى يسهل حسابه فى العينة وبسرعة دون اللجوء إلى إجراء الحسابات المطلوبة فى قياس الانحراف القياسى.

مثال ٣-١٤

بالنسبة للبيانات الخاصة بأوزان العجول الفريزيان المذكورة فى مثال ٣-١ كان أعلى الأوزان 240 كج فى حين كان أقلها وزناً هو 158 كج وبالتالي فإن المدى فى هذه الحالة $240 - 158 = 82 \text{ kg}$ ومن خلال علاقة المدى بالانحراف القياسى والموجودة فى جدول ٢ ملحق أ وعند حجم عينة 11، القيمة هى $\sigma/R = 0.315$. بناء على ذلك يصبح تقدير الانحراف القياسى $82(0.315) = 25.83 \text{ kg}$ وباستخدام المعادلة الأساسية للانحراف القياسى لهذه المجموعة معادلة (٣-٩) تكون قيمته 25.40 كج وهو اختلاف بسيط مقارنة بحجم الانحراف القياسى. ولكن الحصول عليه أسهل بكثير ويعطى فكرة واضحة وسريعة عن البيانات التى سوف تستخدم فى إجراء التجارب. وحيث إن معامل الاختلاف هنا حوالى 13% فإن هذه البيانات توضح أن العينة ملائمة لغرض التجارب حيث إنها لا تشذ عما هو متوقع بالنسبة لحجم التباين بالنسبة للقيم ومتوسطاتها. ولكن يجب افتراض أن المتغير يتبع التوزيع الطبيعى حتى يمكن تطبيق هذه العلاقة بين المدى والانحراف القياسى.

مثال ٣-١٥

البيانات التالية هى نتائج متحصل عليها من قياس إدرار اللبن بالكيلوجرام من 106 بقرة فريزيان فى أعمار مختلفة، ويمثل الإدرار الناتج فى خلال فترة 13 أسبوعاً الأولى من بداية الحليب.

$$\text{العدد: } n = 106, \text{ المجموع الكلى: } \sum Y_i = 157081$$

$$\text{مجموع المربعات غير المصحح: } \sum Y_i^2 = 247111137$$

احسب الانحراف المعياري والخطأ القياسى ومعامل الاختلاف.

الحل:

المتوسط:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{157081}{106} = 1481.9 \text{ kg}$$

مجموع المربعات المصحح:

$$\begin{aligned} \sum y^2 &= \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \\ &= 247111137 - \frac{(157081)^2}{106} = 14333395.86 \end{aligned}$$

$$S^2 = \frac{\sum y_i^2}{(n-1)} = \frac{14333395.86}{105} = 136508.5 \text{ kg}^2$$

التباين:

$$S = \sqrt{136508.5} = 369.5 \text{ kg} \quad \text{الانحراف القياسي:}$$

$$S_{\bar{Y}}^2 = \frac{S^2}{n} = \frac{136508.5}{106} = 1287.8 \text{ kg}^2 \quad \text{تباين المتوسط:}$$

$$S_{\bar{Y}} = \sqrt{S_{\bar{Y}}^2} = \sqrt{1287.8} = 35.9 \text{ kg} \quad \text{الخطأ القياسي:}$$

$$C.V. = \frac{S}{\bar{Y}} \times 100 = \frac{369.5}{1481.9} \times 100 = 24.93\% \quad \text{معامل الاختلاف:}$$

ويتضح من النتائج أن معامل الاختلاف في هذه العينة كان إلى حد ما مرتفعاً وقد يكون السبب في ذلك هو بعض الاختلافات في النواحي البيئية التي تتعرض لها الأبقار ومن أهمها: اختلاف الأبقار الداخلة في العينة في مواسم إدارها وأعمارها، بالإضافة إلى أنه قد يكون للموسم الذي بدأ فيه الحليب تأثير على الإنتاج. هذه الأسباب وتعليلها ومحاولة فهم كيفية تأثيرها وإن كان محسوساً إلا أنه أمر لا يدخل ضمن أساسيات علم الإحصاء.

٣-٧ استخدام برنامج SAS للحصول على الإحصاءات التوصيفية

الإحصاءات التوصيفية descriptive statistics هي عبارة عن الإحصاءات التي تستخدم لتوصيف عينة ما مثل: الوسيط، المنوال، المتوسط، أقل قيمة، أقصى قيمة، المدى، التباين، الانحراف القياسي، الخطأ القياسي، معامل الاختلاف. وتفيد معرفة هذه المعلومات في فحص وتفتيح أى بيانات قبل إجراء أى تحليلات إحصائية عليها. ويمكن باستخدام برنامج SAS الحصول على هذه الإحصاءات بطريقة سهلة.

ويوجد ببرنامج SAS عدة طرق يمكن استخدامها في التوصيف الإحصائي لمجموعة ما من البيانات وهي كالتالي:

الطريقة Procedure			الإحصاء Statistic
Summary	Univariate	Means	
√	√	√	N عدد المشاهدات
√	√	√	MEAN المتوسط الحسابي
√	√	√	SUM إجمالي قيم المشاهدات
√	√	√	MIN أقل قيمة في المشاهدات
√	√	√	MAX أعلى قيمة في المشاهدات
√	√	√	RANGE المدى
√	√	√	VAR تباين العينة
√	√	√	STD الانحراف القياسي للعينة
√	√	√	STDERR الخطأ القياسي للمتوسط
√	√	√	CV معامل الاختلاف
	√		MEDIAN الوسيط
	√		MODE المنوال

من الجدول السابق يتضح أنه عند الرغبة في حساب كل من الوسيط والمنوال فإنه لا بد من استخدام PROC UNIVARIATE كما سيتضح من الأمثلة التالية.

مثال ٣-١٦

استخدم برنامج SAS لحساب كل من: الوسيط، المنوال، المتوسط، أقل قيمة، أقصى قيمة، المدى، التباين، الانحراف القياسي، الخطأ القياسي، معامل الاختلاف لبيانات المثال رقم ٣-١٢.

الحل:

يتم كتابة برنامج SAS فى نافذة تحرير البرنامج program editor window كالتالى:

```
*----- Descriptive Statistics-----;
DATA CALVES;
INPUT BWEIGHT @@;
CARDS;
*----- Start of Data -----;
26 26 22 23 24 35 22 28 26 22 24 25 24 21 28 26
*----- End of Data -----;
*----- Start of procedures and options -----;
PROC MEANS MAXDEC=2 N MEAN MIN MAX RANGE VAR
STD STDERR CV;
TITLE OUTPUT FROM MEANS PROCEDURE;
RUN;
```

لاحظ:

- ١- جميع سطور البرنامج يجب أن تنتهى بعلامة الفاصلة المنقوطة (;)
- ٢- إذا بدء السطر بعلامة * معنى ذلك أن ما سوف يأتى بعدها عبارة عن معلومات توضيحية وليس لها أى تأثير فى عمل البرنامج.
- ٣- الكلمات الموضوع تحتها خط تشير إلى أنها كلمات أساسية ولن يعمل البرنامج بطريقة صحيحة إذا لم يتضمنها البرنامج (لاحظ أنه عند التنفيذ لا تضع خطوط تحت هذه الكلمات).
- ٤- جزء البيانات يمكن أن يوضع بحيث كل مشاهدة فى سطر منفصل أو فى سطور متصلة، ولكن فى هذه الحالة يلزم وضع العلامتين @@ فى سطر المدخلات بعد اسم المتغير.
- ٥- استخدام كلمة MAXDEC=2 تعنى إظهار النتائج باستخدام رقمين فقط بعد العلامة العشرية، ويمكن تغيير الرقم 2 حسب القيمة التى يأخذها المتغير.

بعد الانتهاء من كتابة البرنامج يضغط على الاختيار submit أو run لبدء التحليل بواسطة البرنامج. افتح نافذة المستكشف log window للتأكد من عدم وجود أخطاء.

النتائج:

في نافذة المخرجات output window نحصل على النتائج كالتالي:

OUTPUT FROM MEANS PROCEDURE					
Analysis Variable: BWEIGHT					
N	Mean	Minimum	Maximum	Range	Variance
16	25.13	21.00	35.00	14.00	11.45

	Std Dev	Std Error	CV		
	3.38	0.85	13.47		

لحساب الوسيط والمنوال يلزم استخدام proc univariate بدلاً من proc means فنظهر النتائج أن الوسيط = 24.5 و المنوال = 26

٣-٨ التشفير Coding

قد يكون من الملائم في كثير من الحالات عند إجراء الحسابات أن تجري عملية تشفير للبيانات بغرض تسهيل العمليات الحسابية وذلك بمعاملتها بطريقة رياضية (جبرية) كأن يتم مثلاً طرح قيمة معينة من جميع البيانات الواردة في العينة أو إضافة قيمة لها ... الخ. وقد يكون ملائماً في بعض الأحيان أيضاً إجراء عمليات أخرى مثل الضرب أو القسمة على ثابت أو حتى استخراج الجذور أو الحصول على لوغاريتمات الأعداد. كل ذلك لا يحدث تغيراً في طريقة حساب الإحصاءات مادامت البيانات جميعها عوملت بنفس المعاملة وأخذ هذا التشفير في الحسبان عند نهاية عملية الحسابات وذلك بغرض إرجاع القيم المحسوبة إلى حالتها (أو قيمتها) الأصلية وهو ما يعرف بعملية عكس التشفير decoding. وتختلف عملية عكس التشفير حسب نوع المعاملة الجبرية أو الحسابية التي أجريت على البيانات.

في حالة الجمع أو الطرح لقيمه ثابتة adding or subtracting فإن هذه العملية تؤثر على قيمة المتوسط فقط دون التباين. ولإرجاع قيمة المتوسط المحسوبة إلى قياسه الأساسي يتم عكس ما تم إحداثه من عمليات حسابية وذلك للحصول على القيم الحقيقية actual values، فإذا طرح ثابت من جميع القيم فإن عملية عكس التشفير تتم عقب حساب المتوسط بإضافة نفس القيمة المطروحة إلى التقدير الناتج، والعكس صحيح في حالة الجمع. أما بالنسبة لقيمة التباين وبالتالي الانحراف القياسي ... الخ، فإن قيمتها لا تتأثر بتلك العملية نظراً لأن تلك القيم تحسب على أساس انحراف من قيمة المتوسط.

وهنا تجدر الإشارة إلى بعض القواعد الجبرية التي تحكم العلاقات بين المتغير والثابت وحساب القيم منهما.

قاعدة ٣-٥

$$\mu (Y-C) = \mu y - C$$

الإثبات:

بما أن تقدير قيمة μ يتم عن طريق تقدير المتوسط \bar{Y} فيكون كالتالي:

$$\begin{aligned} \mu (Y-C) &= \sum (Y_i - C)/n \\ &= (\sum Y_i - \sum C)/n \\ &= (\sum Y_i - nC)/n \\ &= \frac{\sum Y_i}{n} - \frac{nC}{n} = \bar{Y} - C \end{aligned}$$

وبالتالي لعكس التشفير يتم إضافة قيمة الثابت C إلى التقدير الناتج من حساب متوسط القيم المشفرة coded values.

قاعدة ٣-٦

$$V(Y_i - C) = V(Y_i)$$

حيث إن تباين الثابت يساوي صفرًا لعدم تغير قيمه، أي أن $V(C) = 0$ ، أي أنه عند عملية التشفير لا يؤثر ذلك على التباين وبالتالي الانحراف القياسي وذلك لأن الحساب يتم من المتوسط فعند ذلك يمكن وضع قيمة $(Y_i - C) = Y_i$ ويصبح حساب التباين للقيم المشفرة للمتغير الجديد Y_i' وبالتالي فإن المتوسط سوف يحسب من مراكز القيم الجديدة، وعلى ذلك فلا يؤثر ذلك على التباين والانحراف القياسي.

مثال ٣-١٧

من مجموعة البيانات التالية اطرح ثابت قيمته 50 ثم احسب كل من المتوسط والانحراف القياسي.

$$51, 59, 57, 55, 53$$

الحل:

$$1, 9, 7, 5, 3$$

بعد الطرح تصبح القيم

$$\bar{Y} = \frac{25}{5} = 5 \text{ المتوسط}$$

بإضافة الثابت فإن المتوسط الفعلى يكون $50 + 5 = 55$

التباين

$$\frac{3^2 + 5^2 + \dots + 1^2 - \frac{(25)^2}{5}}{5-1} = \frac{165-125}{4} = 10$$

الانحراف القياسى $\sqrt{10} = 3.162$ و هذه القيمة لا تتأثر بالتشفير.

فى حالة القيم التى تشفر عن طريق قسمتها (أو ضربها) فى رقم ثابت فإنه تبعاً لذلك يتم حساب المتوسط على القيم المشفرة وبعد ذلك يتم الحصول على قيمة المتوسط الحقيقى بعكس التشفير وذلك بالضرب فى الثابت (أو القسمة عليه).

أما بالنسبة للتباين فإنه حسب قاعدة ٣-٤ فإن تباين الثابت مضروباً فى المتغير (أو مقسوماً عليه) يساوى تباين المتغير مقسوماً على مربع الثابت (أو مضروباً فيه). ويلزم فى هذه الحالة عكس عملية التشفير عند الرغبة فى الحصول على القيم الحقيقية. وعلى ذلك ففى التباين يضرب فى (أو يقسم على) مربع الثابت، أما فى حالة الانحراف القياسى فإنه فقط لعكس عملية التشفير يضرب فى (أو يقسم على) الثابت.

مثال ٣-١٨

القيم التالية تمثل تركيز عنصر نادر فى تحليلات العينات وهى:

$$0.0049, 0.0049, 0.0048, 0.0045, 0.0044, 0.0047$$

المطلوب حساب التباين والانحراف القياسى ومعامل الاختلاف باستخدام التشفير ثم حساب القيم الحقيقية.

الحل:

(١) يتم ضرب القيم فى ثابت قيمته 10000 للتخلص من العلامة العشرية وتصبح الأرقام صحيحة.

(٢) يتم طرح ثابت آخر قيمته 40 من جميع القيم وتصبح القيم 9, 9, 8, 5, 4, 7

(٣) يتم حساب المتوسط والانحراف القياسى باستخدام القيم المشفرة.

(٤) للحصول على القيم الحقيقية يتم عكس التشفير.

$$\text{المتوسط المشفر } = 7 = \frac{42}{6}, \text{ المتوسط الحقيقي } = 0.0047 = \frac{1}{10000} \times (7 + 40)$$

$$\text{التباين المشفر} = 4.4 = \frac{316 - 6(7)^2}{6 - 1} = \frac{22}{5}$$

$$\text{التباين الحقيقي} = 4.4 \times \frac{1}{(10000)^2} = 4.4 \times 10^{-8}$$

$$\text{الانحراف القياسي الحقيقي} = 2.0976 \times 10^{-4} = \sqrt{4.4 \times 10^{-8}}$$

$$\text{معامل الاختلاف المشفر} = 29.97\% = \frac{2.0976}{7} \times 100$$

$$\text{معامل الاختلاف الحقيقي} = 4.46\% = \frac{2.0976 \times 10^{-4}}{0.0047} \times 100$$

٣-٩ الجداول التكرارية Frequency table

يكون من الأوفق في كثير من الأحيان أن تلخص البيانات المتحصل عليها بحيث توضع في جداول تحتوي على تكرار كل قيمة تحدث في العينة أو توضع كل مجموعة قيم مع بعضها في فئة معينة ثم يذكر عدد المرات التي حدثت بها هذه الفئة. ويكون ذلك مفيداً في حالة زيادة عدد المشاهدات في العينة. فإذا كانت القيم المشاهدة تقع في عدد قليل من الأقسام (الفئات) وكل هذه الأقسام ممثلة كما في عدد الخلفة في البطن الواحدة في المعز مثلاً حيث تتراوح حصرياً ما بين 1 أو 5 فإن كل فئة تمثل أحد القيم ويوضع أمامها عدد الحالات التي شوهدت من هذه القيم، وهذا يمثل جدولاً تكرارياً. أما إذا كانت قيم المتغير لها صفة الاستمرارية إلى حد ما مثل عدد الخلفة في البطن في الأرانب (التي تتراوح ما بين 1 و 12) أو نسب الإصابة بمرض معين في مجاميع الحيوانات، فإن القيم توضع في فئات تشمل كل منها أكثر من قيمة. ووضع البيانات بهذه الطريقة مع بيان لعدد المشاهدات التي تدخل ضمن كل فئة أو داخل حدود كل تقسيم هو ما يعرف بالجدول التكرارية. وغالباً ما يمكن تمثيل هذه الجداول التكرارية أيضاً، وكما سبق الإشارة إليه، في صورة أشكال بيانية (هستوجرامات). وفي الجداول التكرارية توضع البيانات في عدد معقول من الأقسام أو الفئات cells or classes وذلك يضمن عدم فقدان الأساس في تلخيص البيانات، وهذه عادة ما يتراوح عددها بين

20-10 فئة. وتعرف الفئة على أنها تصنيف توضع فيه كل القيم التي تقترب من بعضها. وبناء على المدى الذي تأخذه البيانات عامة، فإنه يحدد مراحل الفئات أو حدود الفئات أو الخلايا cell intervals or class boundaries التي تحدد الحد الأدنى والأعلى لكل فئة، وهذه تصمم بحيث تكون هذه الحدود أعلى من قيم المشاهدات حتى لا يكون هناك أى تداخل بين الأقسام.

وتحدد نقاط تمثل مراكز الفئات أو الخلايا cell midpoint وهي عبارة عن متوسط مدى الفئات. وعند إجراء الحسابات فإن كل المشاهدات التي تقع في فئة معينة تعطى كلها قيمة مركز الفئة، فإذا كان المدى المستخدم في الفئات صغيراً فإنه بالتالي يمكن استخدام الجداول التكرارية في تمثيل المتغيرات ذات الطابع الاستمراري مثل الأوزان أو الإنتاج وخلافه. وقد يصاحب ذلك بعض الفقد في التفاصيل من حيث معاملة التوزيع المستمر كما لو كان توزيعاً متقطعاً، ولكن هناك فائدة في تلخيص النتائج وأن الفقد يكون صغيراً ما دامت الفئات حدودها ضيقة. ومن المستحسن في حالة الجداول التكرارية ألا تكون هناك فئات مفتوحة النهاية كأن توضع القيم الشاذة في فئة نهائية، فمثلاً توضع عدد الخلفة في البطن في الأرناب 20 فأكثر، حيث إن عدد الأمهات التي تعطى أكثر من 20 خلفة في البطن يكون قليلاً وقد تشمل 22 أو 23 أو أكثر، في حين أن تكون الفئات الأخرى الوسطية لها مدى حوالى 2 مثلاً من 10-8، 12-10... وهكذا.

وعموماً لتجميع البيانات في جداول تكرارية فإنه يجب:

أولاً: تحديد عدد الفئات التي سوف تبوب فيها البيانات، ويفضل أن يتراوح عددها بين 10، 20 فئة في حالة ما إذا كان المتغير مستمراً أو في عدد ملائم من الفئات في حالة المتغيرات المتقطعة؛ وذلك حتى لا تفقد قدرًا كبيراً من المعلومات، وليس عدداً كبيراً جداً حتى لا تفقد ميزة التجميع، ويلاحظ أنه كلما قل عدد البيانات قل عدد الفئات، ويقل بذلك احتمال دخول القيم الشاذة في العينة.

ثانياً: يتم تقسيم المدى المحسوب من البيانات، وهو الفرق بين أعلى قيمة وأقل قيمة في العينة، على عدد الفئات وهو ما تم إيضاحه سابقاً أنه يحسن أن يتراوح بين 10، 20 فئة. وهذا يعطى مدى الفئة وهو ما يجب أن يتسع لكل القيم.

ثالثاً: يتم تحديد المدى الخاص بالفئة الدنيا من بين كل الفئات. ومن الملائم والمعمول به في بعض الحالات أن تبدأ الفئة الدنيا برقم يمثل مضاعفات لمدى الفئة ولو أن هذا لا يعتبر قاعدة لبدء الفئات. ومن المهم هنا أنه لتحديد حدود الفئات أن تبدأ برقم عشري واحد أقل من ذلك الذي يتواجد في المشاهدات، فمثلاً إذا كانت الفئة أرقاماً صحيحة فإن المدى للفئات يبدأ برقم عشري واحد. فإذا كان المدى الكلي

مثلاً 20 وحدة وحدد عدد الفئات المستخدمة بعشرة فئات فإن مدى الفئة يصبح $20/10 = 2$ وحدة. وبالتالي فإن الفئات سوف تبدأ من 9.5-11.5 بدلاً من 10-12 وبذلك فإن القيم 10 بالضبط سوف توضع في فئة معينة حيث إن الفئة السابقة قد تكون 8-10 وهذا لا يوضح موضع الرقم 10 .

رابعاً: بعد تحديد الفئة الدنيا يضاف إلى حدود الفئات اتساع (أو مدى) الفئات فتعطي حدود الفئة التي تليها ... وهكذا.

خامساً: تحسب مراكز الفئات وذلك بقسمة حاصل جمع مدى الفئة الدنيا على 2 حيث يعطى مركز الفئة الدنيا ثم يجمع مدى الفئات ليعطى مراكز الفئات المتتالية.

سادساً: يوضع أمام مركز كل فئة في الجداول عدد الأفراد والمشاهدات التي تنتمي إلى هذه الفئة وهو ما يعرف بتكرار الفئة class frequency ويرمز له بالرمز f ويعطى حاصل جمع تكرارات الفئات مجموع المشاهدات.

وفي هذا المجال من الممكن أن تقسم التكرارات الخاصة بكل فئة على مجموع التكرارات معطية بذلك التكرار النسبي relative frequency distribution لكل فئة ومن هذه التكرارات النسبية ينشأ التوزيع التكراري النسبي. وقد يمكن التعبير عن التكرارات النسبية في صورة مئوية percentage distribution بالضرب في 100 وهذه طرق مختلفة للتعبير عن التكرارات للفئات المختلفة.

وهناك طرق أخرى أيضاً للتعبير عن التوزيع التكراري للفئات المختلفة كما سبق ذكره في الباب الثاني فهناك التوزيع التكراري المتجمع cumulative frequency distribution. وفي هذا المجال يكون من المرغوب فيه معرفة عدد الأفراد التي تقل أو تزيد عن قيمة معينة. وعلى ذلك يتم تجميع تكرار كل فئة مع تكرارات جميع الفئات التي تسبقها أو تليها.

مثال ٣-١٩

في دراسة عن العمر عند أول ولادة في أبقار الفريزيان وجد أن العمر يتراوح بين 19 إلى 44 شهراً. ولقد رُئي أن تقسم الأعمار إلى فترات تحتوى كل منها على 3 أشهر. على هذا الأساس وحتى لا يحدث هناك لبس في وضع القيم بدأ بعمر 18 شهراً إلى 45 شهراً لمراحل الفئات. وبما أن الأعمار للأبقار كانت مقيدة إلى أقرب شهر كامل فإنه بذلك تصبح حدود الفئات أقل من المراحل برقم عشري واحد وبذلك بدء في حدود الفئات من 17.5 وحتى 44.5 شهراً.

يتم بعد ذلك تحديد مراكز الفئات وذلك بقسمة حدى الفئة على 2 وفى خانة أخرى من الجدول يوضع فيها تكرار كل فئة من الفئات أى عدد المشاهدات التى تتبع تلك الفئة. وقد يضاف إلى ذلك خانات أخرى توضح توزيعاً تكرارياً تكاملياً أو توزيعاً تكرارياً نسبياً ... الخ.

ويوضح جدول ٣-١ تكوين جدول تكرارى لعدد 962 عجلة موضحاً فيه أعمار الأبقار عند أول ولادة لها.

جدول ٣-١ حساب المتوسط والتباين من جداول التوزيع التكرارى

(٧) مربع الانحراف X التكرار	(٦) الانحراف X التكرار	(٥) الانحراف	(٤) التكرار	(٣) مراكز الفئات	(٢) حدود الفئات	(١) مراحل الفئات
fd ²	fd	d	f	Y		
54	-18	-3	6	19	20.5-17.5	21-18
80	-40	-2	20	22	23.5-20.5	24-21
187	-187	-1	187	25	26.5-23.5	27-24
0	0	0	340	28	29.5-26.5	30-27
181	181	1	181	31	32.5-29.5	33-30
516	258	2	129	34	35.5-32.5	36-33
585	195	3	65	37	38.5-35.5	39-36
432	108	4	27	40	41.5-38.5	42-39
175	35	5	7	43	44.5-41.5	45-42
2210	532		962			المجموع

حيث $d_i = \frac{Y_i - 28}{3}$ وتعتبر 28 وسطاً فرضياً، أما قيمة 3 فتعبر عن مدى الفئة.

$$\text{فمثلاً } d_1 = \frac{19 - 28}{3} = \frac{-9}{3} = -3$$

حساب المتوسط والانحراف القياسى من الجداول التكرارية.

تتلخص الطريقة فيما يلى:

١- تقدر الفئة التى تحتوى على المتوسط وفى كثير من الأحيان، خاصة فى حالة التوزيع الذى يقترب من الطبيعي، فإنه غالباً ما تستخدم تلك الفئة التى تمثل أعلى تكرار على أنها الفئة التى قد تحتوى على المتوسط (ولا يعتبر ذلك أمراً ضرورياً حيث إن طريقة الحساب سوف تصحح لذلك).

٢- يكون عمود (٥) فى جدول ٣-١ للانحرافات deviations ويرمز لها بالرمز d. وفى هذه الخانة يوضع قيمة الصفر فى تلك الفئة التى اعتبرت أنها تحتوى على المتوسط. ويعنى هذا أنه قد طرح من كل قيمة من الفئات المذكورة فى الجدول قيمة تساوى قيمة مركز الفئة التى افترض أنها تحوى المتوسط.

٣- توضع قيمة موجبة أو سالبة تمثل الانحرافات التى تسبق أو تلى الفئة التى تحتوى المتوسط (عمود ٥)، فإذا كانت الفئة أكبر توضع قيمة موجبة وفى حالة الأقل توضع قيمة سالبة. وهذا يعنى أنه قد قسمت كل مركز فئة على مراحل الفئات المستخدمة أى أن وحدات القياس تصبح ممثلة بوحدات لمرحل الفئات، فمثلاً تصبح القيم السالبة ١,٠٠٠، -٢,٠٠٠، -١، فى حين تصبح القيم الموجبة ١,٠٠٠، ٢,٠٠٠. وتعبّر d عن انحراف القيمة عن المتوسط الفرضى مقسوماً على مدى الفئة.

٤- يتم الحصول على القيمة الجبرية لحاصل ضرب تكرار كل فئة فى قيمة انحرافها أى fd وتوضع فى عمود مستقل. بعد ذلك يتم الجمع الجبرى لكل القيم أى $\sum fd$ ، ويجدر الإشارة هنا إلى أنه لو كانت قيمة هذا المجموع صفر فإن ذلك يعنى أن المتوسط الحقيقى كان يساوى قيمة مركز الفئة التى اعتبرت أنها تحتوى المتوسط.

٥- يتم حساب المتوسط الحقيقى للعينة وذلك من المعادلة التالية:

$$\bar{Y} = M + \frac{i \sum fd}{\sum f} \quad (١٦-٣)$$

حيث

M : هى مركز الفئة التى افترض أنها تحتوى المتوسط.

i : هى مرحلة الفئة.

$\sum f$: هى مجموع عدد المشاهدات فى العينة.

٦- يتم تربيع قيم الانحرافات في كل فئة d^2 ثم ضربها في تكرار كل فئة للحصول على قيمة fd^2 وتوضع هذه في عمود آخر أو يتم ذلك بضرب قيمة (fd) في كل فئة (عمود ٧). ثم بالجمع (وهنا كل القيم تكون موجبة نظراً لأنه قد رُبعت قيم d) للحصول على قيمة $\sum fd^2$ أي مجموع حاصل ضرب مربعات الانحرافات في التكرارات ويتم حساب الانحراف القياسي للعينة من المعادلة التالية:

$$s^2 = \frac{i^2[\sum fd^2 - (\sum fd)^2 / \sum f]}{\sum f - 1} \quad (١٧-٣)$$

ومنها فإن الانحراف القياسي S عبارة عن الجذر التربيعي للتباين.

لاحظ أنه قد تختلف قيمة المتوسط المحسوب بهذه الطريقة اختلافاً بسيطاً عن قيمة المتوسط إذا تم حسابه من المعادلة (٣-٢) وذلك لأنه في المعادلة (٣-١٦) أعطيت كل القيم التي تنتمي إلى فئة معينة نفس القيمة والتي تمثل مركز الفئة الواقعة فيها وهذا قد يختلف عن استخدام كل قيمة بمفردها. ولكن في حالة البيانات الكثيرة العدد فإن الاختلافات في الحساب تصبح صغيرة ويمكن تجاهلها ولا توازى الجهد الزائد المبذول في عمليات الحساب، خاصة إذا كانت أعداد الفئات كبيرة وبالتالي مراحلها ليست متسعة جداً. ونفس الشيء طبعاً سوف يحدث بالنسبة لقيمة الانحراف القياسي.

مثال ٣-٢١

استكمالاً للمعلومات الموجودة في المثال ٣-١٩ فإنه يتم حساب المتوسط \bar{Y} من ضرب القيمة الموجودة في عمود (٤) وتمثل التكرارات في قيم العمود (٥) وتمثل الانحرافات وهذه مذكورة في العمود (٦) والتي يتم تجميعها في نهاية هذا العمود ويرمز لها بالرمز $\sum fd$. وتطبيق المعادلة (٣-١٦) يمكن الحصول على قيمة المتوسط كالتالي:

بما أن قيمة $M = 28$ ، $i = 3$

$$\bar{Y} = 28 + \frac{(3)(532)}{962} = 29.66 \quad \text{إذا}$$

أي أن متوسط العمر عند الولادة الأولى كان 29.66 شهراً

ومن استخدام العمود (٧) يمكن تطبيق المعادلة (٣-١٧) لحساب التباين كالتالي:

$$S^2 = 3^2 \left[\frac{2210 - (532)^2 / 962}{961} \right] = 17.942$$

$$S = \sqrt{17.942} = 4.236 \quad \text{ويكون الانحراف القياسي:}$$

أي أن الانحراف القياسي لهذه الصفة وهي العمر عند أول ولادة يساوي 4.236 شهراً وهنا تجدر الإشارة إلى أن هذه الطريقة قد وفرت وقتاً كثيراً كان سوف يستخدم في معالجة البيانات فردياً. ويمكن حساب معامل الاختلاف كالتالي:

$$C.V. = \frac{4.236}{29.66} \times 100 = 14.3\%$$

تمارين الباب الثالث

٣-١ لديك بيانات تمرين ٢-١ والمطلوب:

أ- حساب مقاييس النزعة المركزية لمجموعة البيانات وهي كل من المتوسط الحسابي، الوسيط، المنوال، المتوسط الهندسي.

ب- احسب المدى من خلال جدول ٢ ملحق أ، ما هو تقديرك لقيمة الانحراف القياسي.

ج- احسب التباين والانحراف القياسي.

د- احسب قيمة معامل الاختلاف وكل من تباين المتوسط والخطأ القياسي.

٣-٢ إذا كان القطاع المأخوذ في عظمة الفخذ دائرياً وأن مساحته تقدر بواسطة المعادلة $m = \pi r^2$ حيث $\pi = 3.14$ فأوجد المتوسط والتباين ومعامل الاختلاف لمساحة المقطع إذا كانت القيم التالية تمثل نصف القطر بالسنتيمتر:

2.1, 2.9, 2.8, 2.4, 3.6, 2.8, 2.9, 3.5, 3.1

٣-٣ إذا كانت قيمة $\bar{Y} = 40$ ومعامل الاختلاف 20.7% وحجم العينة $n = 16$ فرداً. احسب مجموع المربعات الكلي غير المصحح والخطأ القياسي.

٣-٤ البيانات التالية توضح عدد بويضات لطفيل ما والتي لوحظت في مجموعة من العينات والمطلوب حساب كل من المتوسط والانحراف والخطأ القياسي لها.

حدود الفئة	التكرار	حدود الفئة	التكرار
1 - 5	3	36 - 40	25
6 - 10	5	41 - 45	22
11 - 15	7	46 - 50	19
16 - 20	18	51 - 55	6
21 - 25	32	56 - 60	6
26 - 30	45	61 - 65	3
31 - 35	38	66 - 70	1

٣-٥ إذا كان متوسط الفرق في ضغط الدم قبل بدء العلاج بدواء معين وبعده يساوى 2 مم في تجربة ما قيس فيها ضغط الدم على 12 فرداً من مرضى ضغط الدم وشوهت 11 قيمة فقط من هذه الفروق فوجدت كالتالى:

$$3, 2, 8, 2, 1, -3, -7, 10, 1, 7, 8$$

احسب القيمة الغائبة والانحراف القياسى ومعامل التباين وتباين المتوسط والخطأ القياسى.

٣-٦ اكتب القيم التالية بصورة كاملة (مفكوك الكمية):

$$\sum_{m=1}^5 5(X_m - 3) \quad (\text{ج}) \quad \sum_{i=2}^4 (C_i - 1) \quad (\text{ب}) \quad \sum_{i=5}^{10} b_i^2 \quad (\text{أ})$$

$$X_4 = -1, X_3 = 6, X_2 = 3, X_1 = 4 \quad \text{إذا كانت قيمة}$$

فأوجد قيمة ما يلى:

$$\sum_{i=1}^4 (X_i - 3) \quad (\text{ب}) \quad \sum_{i=1}^4 X_i^2 - 6 \quad (\text{أ})$$

$$\sum_{i=1}^4 X_i^2 - 4 \left(\frac{\sum_{i=1}^4 X_i}{4} \right)^2 \quad (\text{د}) \quad \sum_{i=1}^4 (X_i + 1)^2 \quad (\text{ج})$$