٣ 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 القياسى

٧ – استخدام برنامج SAS للحصول على الإحصاءات التوصيفية

٨ – التشفير

٩ – الجداول التكرارية

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

٣-١ مفهوم التجميع Summation notation

فى علم الإحصاء كثيراً ما يتم جمع قيم البيانات أو مربعاتها ... الخ. فعند الرغبة فى قياس متوسط أوزان مجموعة من العجول عند بدء تسمينها مثلا فإنه بدلا من الإشارة إلى عملية التجميع عن طريق ذكرها حرفياً، فإن هناك طريقة متبعة لتوضيح ذلك باستعمال إشارات معينة لتدل على عملية التجميع وهى حرف $\sum$ اليونانى وينطق سيجما (جدول ١٩ ملحق أ). وهذه الإشارة تعنى التجميع، فإذا أريد مثلاً جمع أوزان 10 عجول فبدلاً من وضع ذلك فى صورة $Y_1 + Y_2 + \cdots + Y_{10}$ 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 فى هذه الحالة تمييزاً يشير إلى رقم أو ترتيب المشاهدات ويوضع تحت وفوق علامة التجميع المدى الذى يتم عليه ذلك بقيمتيه الدنيا والعليا، على الترتيب. ففى الحالة السابقة يتم التجميع من المشاهدة رقم $i = 1$ إلى المشاهدة الأخيرة $i = 10$.

مثال ٣-١

إذا كانت أوزان مجموعة عجول الفريزيان بالكيلوجرام عند بدء تسمينها هى كالتالى:

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

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

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 $555 = 165 + 180 + 210$.

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 $1,2,\cdots,n$

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 $\sum_{i=1}^{11} Y_i^2$ تعنى $416189 = (210)^2 + (180)^2 + \cdots + (160)^2$. ويمكن استعمال أى من الحروف كتمييز ولكن هناك شبه عرف بين الإحصائيين على استعمال الحروف اللاتينية البادئة 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 ٣-٢

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

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

الحل:

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

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

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

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

مثال ٣-٣

إذا كانت قيم $Y_1 = 2$ ، $Y_2 = 3$ ، $Y_3 = 5$

وكانت قيم $X_1 = 4$ ، $X_2 = 2$ ، $X_3 = 5$ فأوجد قيمة:

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

الحل:

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

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

وهناك بعض القواعد الأساسية للتعامل مع مفهوم التجميع وهى:

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 يساوى حاصل ضرب n فى قيمة الثابت أى:

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

مثال ٣-٤

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

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

الحل:

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

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

الجداول 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

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

تعريف ٣-٢

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

$$\mu = \frac{\sum_{i=1}^{N} Y_i}{N} \qquad (١-٣)$$

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

تعريف ٣-٣

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

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \qquad (٢-٣)$$

فإذا كانت العينة ممثلة تمثيلا جيداً للعشيرة المأخوذة منها، فإن متوسط العينة يعتبر ممثلا (تقديرا لنقطة point estimate) لمتوسط العشيرة.

مثال ٣-٥

فى البيانات المذكورة فى مثال ٣-١ احسب متوسط العينة والتى حجمها $n = 11$

الحل:

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

ومن الملاحظ هنا أن جميع القيم الموجودة فى العينة قد استخدمت فى 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 $\sum(Y-\overline{Y})=0$
- مجموع مربعات فروق المشاهدات عن متوسطها هو أدنى ما يمكن، أى أن القيمة $\sum(Y-\overline{Y})^2$ تكون أدنى ما يمكن عما لو استخدمت أى قيمة أخرى غير $\overline{Y}$. وتعرف هذه الخاصية بـــ "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 ففى حالات كثيرة فإن تلك المقاييس قد تصف العشيرة أو العينة بطريقة أكثر تفهما أو أكثر واقعية.

تعريف ٣-٤

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

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 عددا 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 رتبت البيانات فى العينة ترتيبا تنازليا أو تصاعديا فإن العينة تبدو كالتالى:

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

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

أما قيمة المنوال فهى أيضاً 195 kg حيث إنها هى القيمة الوحيدة التى تكررت فى العينة مرتين.

ومن الملاحظ فى المثال السابق أن قيمتى الوسيط والمنوال متساويتان وتقتربان من قيمة المتوسط الحسابى. وتعتبر هذه الظاهرة من مميزات البيانات التى تتبع فى توزيعها التوزيع الطبيعى normal distribution.

وتلخيصاً لما سبق فإنه يمكن المقارنة بين المقاييس الثلاثة وهى المتوسط والوسيط والمنوال، فالمتوسط هو أكثر المقاييس استخداماً فى الإحصاء خاصة فى معالجة البيانات البيولوجية. ويرجع ذلك إلى أن توزيع متوسطات العينات عادة ما يكون معروفاً من الناحية الرياضية وبالتالى فإن الطرق المستخدمة فى الاستدلال الإحصائى تكون المبنية على أساس المتوسط الحسابى وبالتالى يمكن معرفة مقدار الخطأ فى التقديرات وتوزيعه. وكما سبق القول فإن المتوسط يعتبر تقديراً لمتوسط العشيرة الأكثر تمثيلا لها وذلك لأنه يستعمل كل البيانات فى العينة بالإضافة إلى أن المتوسط يتمتع بأقل قدر من الخطأ (التباين) minimum variance. أما ما يعاب على المتوسط فهو أنه يتأثر تأثراً كبيراً بالقيم المتطرفة (أى العالية جداً أو المنخفضة جداً outliers)، وذلك لدخول مثل هذه القيم فى حسابه. أما الوسيط فهو سهل فى حسابه ولا يتأثر بالقيم غير الطبيعية وبالتالى يعطى فكرة أوضح عن غالبية القيم فى 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 ثباتا من الوسيط وبالتالى فالمتوسط يمثل متوسط العشيرة بقدر أكبر من الدقة أما المنوال فهو أقل المقاييس كفاءة وهو يكون غير ذى نفع فى حالة العينات صغيرة الحجم. وفى بعض الحالات فإن المنوال قد لا يوجد أو يكون متعددا. إلا أنه فى حالة البيانات العديدة فيكون له نفع خاص فى البيانات العدية مثل طول دورة الشبق وعدد الخلفه فى البطن فى بعض أنواع الحيوانات مثله مثل الوسيط فى ذلك.

وهناك مقاييس أخرى للمركزية يمكن تعريفها هنا فى هذا المجال وهى أقل استخداماً من سابقتها، مثال ذلك متوسطات الأرباع quartiles، الأعشار deciles والمئويات percentiles وهى تلك النقاط التى تقسم التوزيع إلى أرباع أو أعشار أو مئويات على التوالى. وعل ذلك فيمكن القول بأن الوسط يمثل الربع الثانى والعشر الخامس والمئوية الخمسين.

تعريف ٣-٦

المتوسط الهندسى لعدد (n) من القيم هو الجذر النونى لحاصل ضرب هذه القيم أى:

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

(ملاحظة: بدلاً من بيان الضرب تفصيلياً يمكن التعبير عن هذه العملية باستخدام πY_i حيث π هى من الحروف اليونانية وينطق باى).

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

تعريف ٣-٧

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

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

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

مثال ٣-٧

احسب المتوسط الهندسى للقيم التالية والتى تمثل معدل الزيادة العددية فى 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 الحسابى وهو 1.344

مثال ٣-٨

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

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

الحل:

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

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

ويعتبر المتوسط التوافقى فى حالة الأعداد أكثر تمثيلا مقارنة بالمتوسط الحسابى لنفس البيانات والذى قيمته 67.38 عجل.

صندوق ٣-١

مقاييس النزعة المركزية:

المتوسط : $\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$

الوسيط (الوسط): القيمة الوسطية فى العينة

المنوال: القيمة الأكثر تكرارا فى العينة

المتوسط الهندسى: $G = \sqrt[n]{Y_1.Y_2...Y_n} = (Y_1.Y_2 \cdots Y_n)^{1/n}$

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

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 فى قيمة متوسطها أو منوالها ومع ذلك تختلف اختلافاً بيناً فى 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 يساوى 32 فى حين أن القيم فى المجموعة (أ) كلها تقترب من القيمة المركزية (المتوسط) للمجموعة حيث أكبر شذوذ أو انحراف عن هذه القيمة هو وحدتان بالموجب أو بالسالب فى حين أنه فى المجموعة (ب) هناك انحرافات أكبر حيث تبتعد 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، وطبيعى أنه كلما شذت القيم عن مركزها كلما اتسع المدى. وفى المثال (٣-٩) كان المدى للمجموعة (أ) $34-30=4$ وحدات فى حين أنه فى المجموعة (ج) كان المدى فيها $48-15=33$ وحدة.

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

تعريف ٣-٩

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

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

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

$$\sigma_Y^2 = \frac{\sum_{i=1}^{N}(Y_i - \mu)^2}{N} \qquad (٣-٦)$$

حيث N هى عدد القيم فى العشيرة للمتغير Y و μ هى متوسط هذه العشيرة.

تعريف ٣-١١

الانحراف القياسى (المعيارى) standard deviation لقيم متغير عشوائى هو الجذر التربيعى لتباين قيم المتغير.

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \mu)^2}{N}} \qquad (٣-٧)$$

حيث N هى عدد القيم فى العشيرة للمتغير Y و μ هى متوسط هذه العشيرة.

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

مثال ٣-١١

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

الحل:

بما أن المتوسط فى كل عشيرة يساوى 32 فإنه بالتالى يكون:

التباين فى العشيرة (أ):

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

والانحراف القياسى: $\sigma_Y = \sqrt{2} = 1.414$

التباين فى العشيرة (ب): $\sigma_Y^2 = \frac{54}{5} = 10.8$

والانحراف القياسى: $\sigma_Y = 3.286$

التباين فى العشيرة (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 العشيرة (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 العشيرة. ولكن يلاحظ هنا أنه يجب فى هذه الحالة حساب المتوسط من العينة $\overline{Y}$ حيث إنه تقديراً لمتوسط العشيرة غير المعروف μ. وعلى هذا الأساس يمكن القول بأن التقدير للتباين المحسوب من العينة المسحوبة من العشيرة يصبح تقديراً غير منحاز لقيمته فى العشيرة 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. وبالتالى فإن:

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

$$S_Y = \sqrt{S_Y^2} \qquad (٣-٩)$$

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 (٣-٢) حيث:

$$\sum Y_i = n\,\overline{Y} \qquad (٣-١٠)$$

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

عدد المشاهدات وبالتالى فإنه عند تربيع الانحرافات تصبح صعبة التعامل وقد يكون هناك خطأ حسابى نتيجة للتقريب فى حساب مربع الانحرافات. وهنا يمكن استخدام المعادلة الحسابية التى تساوى المعادلة التعريفية حيث يمكن إثبات أن:

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - \frac{(\sum Y_i)^2}{n} \qquad (٣-١١)$$

الإثبات:

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

وهنا تجدر الإشارة إلى أن الكمية $\frac{(\sum Y_i)^2}{n}$ والتى تساوى $n\overline{Y}^2$ تعرف بمعامل التصحيح correction factor (CF) وهو يصحح مجموع المربعات غير المصحح $\sum Y_i^2$ لقيمة المتوسط حيث يمكن اعتبار أنه فى كل مرة فى عملية التجميع كان يجب طرح الكمية $n\overline{Y}^2$ وتتكرر n من المرات. وعموما فإن استعمال المعادلة (٣-٨) لحساب مجموع المربعات المصحح $\sum y_i^2 = \sum(Y - \overline{Y})^2$ يكون ممكناً فى حالة البيانات قليلة العدد والتى تكون قيمتها صغيرة وقيمة المتوسط فيها غير كسرية فى حين أن المعادلة (٣-١١)، وهى الأكثر استعمالاً وشيوعاً، هى أوفق أيضاً عند الحساب واستخدام الآلات الحاسبة. ويلاحظ أن بعض الآلات الحاسبة الصغيرة الإلكترونية حاليا 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 $\overline{Y}$ ومجموع المربعات المصحح $\sum y_i^2$ ثم حساب التباين والانحراف القياسى للعينة والعشيرة.

مثال ٣-١٢

المطلوب حساب تقدير لقيمة المتوسط والانحراف القياسى بالكيلوجرام لوزن الميلاد فى العينة التالية والتى عددها 16 عجل من عجول الفريزيان:

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

الحل:

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

متوسط العينة:

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

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

معامل التصحيح:

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

مجموع المربعات المصحح:

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

تباين العينة:

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

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

لاحظ أن مجموع المربعات المصحح والذى أخذ فيه متوسط العينة كأساس لحساب الانحرافات يعتبر أقل مجموع مربعات وهو ما يعرف بالـــ 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 $\sum(Y-\overline{Y})$ 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 العشيرة μ عادة ما يكون مجهولاً ويتحصل على تقدير له أو عدة تقديرات من العينات المأخوذة من العشيرة وبالتالى يكون هناك متوسط $\overline{Y}$ من كل عينة تدرس. ويلاحظ أن هذه المتوسطات قد تختلف فيما بينها، وهى عادة ما تختلف، ولكن متوسطها يتوقع أن يساوى متوسط العشيرة وتكون المتوسطات للعينات المختلفة والتى كل منها حجمه n من الأفراد موزعة توزيعاً طبيعياً حول متوسط العشيرة. وهناك علاقة تربط بين تباين القيم فى العشيرة وتباين المتوسطات من العينات. ويرمز لتباين متوسطات العينات بالرمز $\sigma^2_{\overline{Y}}$ وهو يساوى $\frac{\sigma^2}{n}$ و القاعدة التالية تستخدم لإثبات هذه العلاقة.

قاعدة ٣-٤

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

حيث يشير الرمز V للتباين variance.

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

$$V(\overline{Y}) = V\left(\frac{\sum Y_i}{n}\right)$$
$$= V\left(\frac{1}{n}\sum Y_i\right) = \frac{1}{n^2}V(Y_1 + Y_2 + \ldots + Y_n)$$

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

$$V\left(\overline{Y}\right)=\frac{1}{n^2}\left[nV(Y_i)+n(n-1)Cov(Y_i,Y_i^{'})\right] \qquad (١٢-٣)$$

وحيث إن قيم Y_i أخذت كلها عشوائياً فإن

$$V\left(\overline{Y}\right)=\frac{1}{n^2}\left[nV\left(Y_i\right)\right] \qquad ،Cov\left(Y_i,Y_{i'}\right)=0$$

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

$$V(\overline{Y})=\frac{\sigma^2}{n} \qquad (١٣-٣)$$

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

$$\frac{\sigma}{\sqrt{n}} \text{ أو } \sqrt{\frac{\sigma^2}{n}} \qquad (١٤-٣)$$

لاحظ أن الخطأ القياسى يتناسب عكسياً مع حجم العينة أو العينات المقدر منها المتوسط، فكلما ازداد عدد أفراد العينة كلما انخفضت قيمة الخطأ القياسى. ففى المثال ٣-١٢ كان حجم العينة 16 مشاهدة والتباين $S^2=11.45\ kg^2$ فيكون تباين المتوسط $\sigma^2_{\overline{Y}}=\frac{11.45}{16}=0.716\ kg^2$ وقيمة الخطأ القياسى $\sigma_{\overline{Y}}=\sqrt{\frac{11.45}{16}}=0.846\ kg$.

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 العشيرة. 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 بمعلوميته.

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 فى مقدار تباينها ومتوسطاتها أو مختلفة فى وحدات القياس المستخدمة ويرمز له بالرمز C.V.

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

$$C.V. = \frac{S}{\bar{\bar{Y}}} X100 \qquad (٣-١٥)$$

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

ومعامل الاختلاف لوزن العجول عند الميلاد فى مثال ٣-١٢

$$C.V. = \frac{3.38}{25.125} X100 = 13.5\%$$

وفى بعض الحالات يفيد معامل الاختلاف فى دراسة العلاقة بين الانحراف القياسى والمتوسط عندما يلاحظ أن زيادة التباين تكون مواكبة لزيادة المتوسط فى حالة دراسة الصفة على مراحل زمنية متعاقبة. ويفيد معامل الاختلاف فى فهم طبيعة البيانات التى يدرسها المجرب حيث إنه إذا كان معامل الاختلاف أعلى من معدله الطبيعى فإن ذلك يدل على أنه قد يوجد شيء غير طبيعى بالنسبة لمجموعة البيانات تحت الدراسة. ويجدر الملاحظة أن معامل الاختلاف ولو أنه مفيد إلا أنه لا يغنى عن معرفة كل من المتوسط والانحراف القياسى حيث إن تغير معامل الاختلاف، كأى نسبة، لا يحدد سبب هذا التغير فيما إذا كان تغيراً فى بسط النسبة أو فى مقامها.

صندوق ٣-٢

مقاييس التشتت:

المدى: أعلى قيمة – أقل قيمة

متوسط الانحرافات المطلقة: $\frac{1}{n}\sum\left|Y-\overline{Y}\right|$

الانحراف المعيارى للعشيرة: $\sigma_Y=\sqrt{\frac{\sum(Y-\mu)^2}{N}}$

الانحراف المعيارى للعينة: $S_Y=\sqrt{\frac{\sum(Y-\overline{Y})^2}{n-1}}$

الخطأ القياسى (الانحراف المعيارى للمتوسط): $S_{\overline{Y}}=\sqrt{\frac{S^2}{n}}$

معامل الاختلاف: $C.V.=\frac{S}{\overline{Y}}X100$

٣-٦ العلاقة بين المدى والانحراف القياسى

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

مثال ٣-١٣

فى مثال ٣-١٢ كانت $n = 16$، ومن جدول ٢ ملحق أ فإن النسبة بين الانحراف القياسى والمدى، $\sigma / R = 0.283$، وهذا يعنى أنه يمكن معرفة حجم σ بصورة مبدئية باستخدام هذه العلاقة. والمدى فى هذه الحالة كان من 21 إلى 35 أى 14. وبالتالى فإن $\sigma = 14 X 0.283 = 3.96\ kg$. وهى لا تختلف كثيراً عن قيمة σ المحسوبة من العينة S والتى 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 وبالتالى فإن المدى فى هذه الحالة $240 - 158 = 82\ kg$ 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 $(82)(0.315) = 25.83\ kg$ 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.

العدد: $n = 106$ ، المجموع الكلى: $\sum Y_i = 157081$

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

احسب الانحراف المعيارى والخطأ القياسى ومعامل الاختلاف.

الحل:

المتوسط:

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

مجموع المربعات المصحح:

$$\sum y^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

$$= 247111137 - \frac{(157081)^2}{106} = 14333395.86$$

التباين:

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

الانحراف القياسى:

$$S = \sqrt{136508.5} = 369.5 \text{ kg}$$

تباين المتوسط:

$$S_{\overline{Y}}^2 = \frac{S^2}{n} = \frac{136508.5}{106} = 1287.8 \text{ kg}^2$$

الخطأ القياسى:

$$S_{\overline{Y}} = \sqrt{S_{\overline{Y}}^2} = \sqrt{1287.8} = 35.9 \text{ kg}$$

معامل الاختلاف:

$$C.V. = \frac{S}{\overline{Y}} X100 = \frac{369.5}{1481.9} X100 = 24.93\%$$

ويتضح من النتائج أن معامل الاختلاف فى هذه العينة كان إلى حد ما مرتفعا وقد يكون السبب فى ذلك هو بعض الاختلافات فى النواحى البيئية التى تتعرض لها الأبقار ومن أهمها: اختلاف الأبقار الداخلة فى العينة فى مواسم إدرارها وأعمارها، بالإضافة إلى أنه قد يكون للموسم الذى 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 هى عبارة عن الإحصاءات التى تستخدم لتوصيف عينة ما مثل: الوسيط، المنوال، المتوسط، أقل قيمة، أقصى قيمة، المدى، التباين، الانحراف القياسى، الخطأ القياسى، معامل الاختلاف. وتفيد معرفة هذه المعلومات فى فحص وتنقيح 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 فى التوصيف الإحصائى لمجموعة ما من البيانات وهى كالتالى:

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

من الجدول السابق يتضح أنه عند الرغبة فى حساب كل من الوسيط والمنوال فإنه لابد 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 القياسى، الخطأ القياسى، 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:

فى نافذة المخرجات output window نحصل على النتائج كالتالى:

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 بدلا من proc means فتظهر النتائج أن الوسيط = 24.5 و المنوال = 26

٣-٨ التشفير Coding

قد يكون من الملائم فى كثير من الحالات عند إجراء الحسابات أن تجرى عملية تشفير للبيانات بغرض تسهيل العمليات الحسابية وذلك بمعاملتها بطريقة رياضية (جبرية) كأن يتم مثلا طرح قيمة معينة من جميع البيانات الواردة فى العينة أو إضافة قيمة لها ... الخ. وقد يكون ملائما فى 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 فى طريقة حساب الإحصاءات مادامت البيانات جميعها عوملت بنفس المعاملة وأخذ هذا التشفير فى 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 التى أجريت على البيانات.

فى 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 الأساسى 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 فى حالة الجمع. أما بالنسبة لقيمة التباين وبالتالى الانحراف القياسى ... 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 التى 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 $\overline{Y}$ فيكون كالتالى:

$$\begin{aligned}
\mu_{(Y-C)} &= \sum(Y_i - C)/n \\
&= \left(\sum Y_i - \sum C\right)/n \\
&= \left(\sum Y_i - nC\right)/n \\
&= \frac{\sum Y_i}{n} - \frac{nC}{n} = \overline{Y} - C
\end{aligned}$$

وبالتالى 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 يساوى صفراً لعدم تغير قيمه، أى أن $V(C) = 0$ ، أى أنه عند عملية التشفير لا يؤثر ذلك على التباين وبالتالى الانحراف القياسى 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' وبالتالى فإن المتوسط سوف يحسب من مراكز القيم الجديدة، وعلى ذلك فلا يؤثر ذلك على التباين والانحراف القياسى.

مثال ٣-١٧

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

51, 59, 57, 55, 53

الحل:

بعد الطرح تصبح القيم 1, 9, 7, 5, 3

المتوسط $\overline{Y} = \frac{25}{5} = 5$

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

التباين

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

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 مضروبا 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.

المتوسط المشفر $7 = \frac{42}{6}$، المتوسط الحقيقى $(7+40)X\frac{1}{10000} = 0.0047$

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

التباين الحقيقى $4.4X\frac{1}{(10000)^2} = 4.4X10^{-8}$

الانحراف القياسى الحقيقى $\sqrt{4.4X10^8} = 2.0976X10^{-4}$

معامل الاختلاف المشفر $\frac{2.0976}{7}X100 = 29.97\%$

معامل الاختلاف الحقيقى $\frac{2.0976X10^{-4}}{0.0047}X100 = 4.46\%$

٣-٩ الجداول التكرارية Frequency table

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

20 – 10 فئة. وتعرف الفئة على أنها تصنيف توضع فيه كل القيم التى تقترب من بعضها. وبناء على المدى الذى تأخذه البيانات عامة، فإنه يحدد مراحل الفئات أو حدود الفئات أو الخلايا cell intervals or class boundaries التى 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 وهى عبارة عن متوسط مدى الفئات. وعند إجراء الحسابات فإن كل المشاهدات التى تقع فى فئة معينة تعطى كلها قيمة مركز الفئة، فإذا كان المدى المستخدم فى الفئات صغيراً فإنه بالتالى يمكن استخدام الجداول التكرارية فى تمثيل المتغيرات ذات الطابع الاستمرارى مثل الأوزان أو الإنتاج وخلافه. وقد يصاحب ذلك بعض الفقد فى التفاصيل من حيث معاملة التوزيع المستمر كما لو كان توزيعاً متقطعاً، ولكن هناك فائدة فى تلخيص النتائج وأن الفقد يكون صغيرا ما دامت الفئات حدودها ضيقة. ومن المستحسن فى حالة الجداول التكرارية ألا تكون هناك فئات مفتوحة النهاية كأن توضع القيم الشاذة فى فئة نهائية، فمثلاً توضع عدد الخلفة فى البطن فى الأرانب 20 فأكثر، حيث إن عدد الأمهات التى تعطى أكثر من 20 خلفة فى البطن يكون قليلاً وقد تشمل 22 أو 23 أو أكثر، فى حين أن تكون الفئات الأخرى الوسطية لها مدى حوالى 2 مثلاً من 10 – 8، 12 – 10... وهكذا.

وعموماً لتجميع البيانات فى جداول تكرارية فإنه يجب:

أولاً: تحديد عدد الفئات التى سوف تبوب فيها البيانات، ويفضل أن يتراوح عددها بين 10، 20 فئة فى حالة ما إذا كان المتغير مستمرا أو فى عدد ملائم من الفئات فى حالة المتغيرات المتقطعة؛ وذلك حتى لا تفقد قدراً 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 فى العينة.

ثانياً: يتم تقسيم المدى المحسوب من البيانات، وهو الفرق بين أعلى قيمة وأقل قيمة فى 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 فى 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 عشرى واحد أقل من ذلك الذى يتواجد فى المشاهدات، فمثلاً إذا كانت الفئة أرقاماً صحيحة فإن المدى للفئات يبدأ برقم عشرى واحد. فإذا كان المدى الكلى

مثلا 20 وحدة وحدد عدد الفئات المستخدمة بعشرة فئات فإن مدى الفئة يصبح 2 = 20/10 وحدة. وبالتالى فإن الفئات سوف تبدأ من 11.5-9.5 بدلاً من 12-10 وبذلك فإن القيم 10 بالضبط سوف توضع فى 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 فتعطى حدود الفئة التى 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 فى الجداول عدد الأفراد والمشاهدات التى تنتمى إلى هذه الفئة وهو ما يعرف بتكرار الفئة class frequency ويرمز له بالرمز f ويعطى حاصل جمع تكرارات الفئات مجموع المشاهدات.

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

وهناك طرق أخرى أيضا للتعبير عن التوزيع التكرارى للفئات المختلفة كما سبق ذكره فى الباب الثانى فهناك التوزيع التكرارى المتجمع cumulative frequency distribution. وفى هذا المجال يكون من المرغوب فيه معرفة عدد الأفراد التى تقل أو تزيد عن قيمة معينة. وعلى ذلك يتم تجميع تكرار كل فئة مع تكرارات جميع الفئات التى تسبقها أو تليها.

مثال ٣-١٩

فى دراسة عن العمر عند أول ولادة فى أبقار الفريزيان وجد أن العمر يتراوح بين 19 إلى 44 شهراً. ولقد رئى أن تقسم الأعمار إلى فترات تحتوى كل منها على 3 أشهر. على هذا الأساس وحتى لا يحدث هناك لبس فى 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 عشرى واحد وبذلك بدء فى 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 تكراريا نسبيا ... 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

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

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

فمثلاً $-3 = \frac{-9}{3} = \frac{19-28}{3} = d_1$ مركز انحراف الفئة الأولى.

مثال ٣-٢٠

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 الطبيعى، 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 قيماً 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 لمراحل الفئات، فمثلاً تصبح القيم السالبة $-1,-2,\cdots$ فى حين تصبح القيم الموجبة $1,2,\cdots$. وتعبر 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 اعتبر أنها تحتوى المتوسط.

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

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

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

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

ومنها فإن الانحراف القياسى S عبارة عن الجذر التربيعى للتباين.

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

مثال ٣-٢١

استكمالا للمعلومات الموجودة فى المثال ٣-١٩ فإنه يتم حساب المتوسط $\overline{Y}$ من ضرب القيمة الموجودة فى عمود (٤) وتمثل التكرارات فى قيم العمود (٥) وتمثل الانحرافات وهذه مذكورة فى العمود (٦) والتى يتم تجميعها فى نهاية هذا العمود ويرمز لها بالرمز $\sum fd$. وبتطبيق المعادلة (٣-١٦) يمكن الحصول على قيمة المتوسط كالتالى:

بما أن قيمة $M = 28$ ، $i = 3$

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

أى أن متوسط العمر عند الولادة الأولى كان 29.66 شهراً

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

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

ويكون الانحراف القياسى: $S = \sqrt{17.942} = 4.236$

أى أن الانحراف القياسى لهذه الصفة وهى العمر عند أول ولادة يساوى 4.236 شهراً وهنا تجدر الإشارة إلى أن هذه الطريقة قد وفرت وقتاً كثيراً كان سوف يستخدم فى معالجة البيانات فردياً. ويمكن حساب معامل الاختلاف كالتالى:

$$C.V. = \frac{4.236}{29.66} X100 = 14.3\%$$

تمارين الباب الثالث

٣-١ لديك بيانات تمرين ٢-١ والمطلوب:

أ- حساب مقاييس النزعة المركزية لمجموعة البيانات وهى كل من المتوسط الحسابى، الوسيط ، المنوال، المتوسط الهندسى.

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

ج- احسب التباين والانحراف القياسى.

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

٣-٢ إذا كان القطاع المأخوذ فى عظمة الفخذ دائرياً وأن مساحته تقدر بواسطة المعادلة م = ط نق^{٢} حيث ط = 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 $\overline{Y} = 40$ ومعامل الاختلاف 20.7% وحجم العينة $n = 16$ فرداً. احسب مجموع المربعات الكلى غير المصحح والخطأ القياسى.

٣-٤ البيانات التالية توضح عدد بويضات لطفيل ما والتى لوحظت فى مجموعة من العينات والمطلوب حساب كل من المتوسط والانحراف والخطأ القياسى لها.

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

٣-٥ إذا كان متوسط الفرق فى ضغط الدم قبل بدء العلاج بدواء معين وبعده يساوى 2 مم فى تجربة ما قيس فيها ضغط الدم على 12 فرداً من مرضى ضغط الدم وشوهدت 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):

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

٣-٧ إذا كانت قيمة $X_1=4$ ، $X_2=3$ ، $X_3=6$ ، $X_4=-1$

فأوجد قيمة ما يلى:

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

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