بعض المؤشرات الإحصائية :

إن تلخيص البيانات ســـواء عن المجتمع أو العينة وذلك بتبويبها في توزيعات تكرارية ومن ثم تمثيلها برسومات بيانية لإعطاء فكرة عن شكل الظاهرة لا يفيد كثيراً في التنبؤ لذلك يجب العمل على إيجاد بعض المقاييس (المؤشـــرات) الرقمية والتي من شـــانها إعطاء فكرة واضـــحة عن هذه البيانات كمقاييس النزعة المركزية أو التشتت ... الخ . والآتي بعض أهم هذه المقاييس :

1- الوسط الحسابي (المعدل) : Mathematical Mean (Average)

يعد الوســـط الحســـابي من أهم مقاييس النزعة المركزية لكثرة تطبيقاته ، فإذا كان لدينا مجتمع محدود مكون من (**N**) مشاهدة متمثلة بالمتغير العشوائي (**Y**) أي أن :

$$\mathbf{Y_i = Y_1, Y_2, Y_3, \ldots, Y_N}$$

وان :

$$\mathbf{y_i = y_1, y_2, y_3, \ldots, y_n}$$

عبارة عن مشـاهدات مأخوذة من نفس المجتمع وتمثل العينة بحجم (**n**) مشـاهدة حيث أن : ($\mathbf{n \le N}$) ، عندئذ يعرف الوسط الحسابي بالصيغة التالية :

$$\mu = \overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} \qquad \text{للمجتمع}$$

$$\hat{\mu} = \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{للعينة}$$

أي أن الوسط الحسابي لمجموعة مشاهدات ماهو إلا عملية إيجاد معدل هذه المشاهدات وهو مكافئ لعملية إيجاد توقع الدالة الذي تم شرحه سابقاً أي أن :

$$E[Y] = \sum_{\forall i} Y_i\, P(Y_i) = \mu = \overline{Y}$$

علماً بان وحدة قياس الوسط الحسابي هي نفس وحدة قياس المتغير العشوائي.

مثال :

لتكن $\mathbf{Y = 2,5,4,3,3,7}$ تمثل عدد الاخطاء الموجودة في ســـتة صـــفحات من كتاب معين ، المطلوب احسب معدل عدد الاخطاء .

الحل :

$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{2+5+4+3+3+7}{6} = 4$$

وباستخدام صيغة التوقع نجد أن:

$$E[Y]=\sum_{i=1} Y_i\,P(Y_i)=2(\frac{1}{6})+5(\frac{1}{6})+4(\frac{1}{6})+3(\frac{2}{6})+7(\frac{1}{6})=4$$

2- التباين: Variance

مثلما اشرنا سابقاً ، يعد هذا المقياس من أهم مقاييس التشتت ويستخدم هذا المقياس أيضاً للمقارنة بين مجموعتين أو أكثر (تحت شروط معينة) من حيث درجة التجانس، فكلما كان تباين المجوعة صغير دل هذا على كبر تجانس مشاهدات تلك المجموعة والعكس صحيح . إن الصيغة الرياضية الخاصة بالتباين هي :

للمجتمع :

$$V(Y)=\sigma^2_{(y)}=E(Y^2)-[E(Y)]^2=E[Y-\bar{Y}]^2=\frac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N}\quad \text{when } N\geq 30$$

$$\text{and}\quad V(Y)=\sigma^2_{(y)}=\frac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N-1}\qquad \text{when } N<30$$

للعينة :

$$\hat{V}(Y)=S^2_{(y)}=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}\qquad \text{when } n\geq 30$$

$$\text{and}\quad \hat{V}(Y)=S^2_{(y)}=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}\qquad \text{when } n<30$$

إن وحدة قياس التباين هي مربع وحدة قياس المتغير العشوائي أي انه إذا كانت وحدة قياس المتغير العشوائي هي (سم ، كغم ، م ،....الخ) فان وحدة قياس التباين هي (سم^2 ، كغم^2 ، م^2 ،....الخ) ولغرض توحيد وحدة القياس بين المتغير العشـــوائي ومقياس التباين يتم اخذ الجذر التربيعي للتباين عند ئذ يدعى هذا المقياس بالانحراف المعياري (القياسي) في حالة قياسات المجتمع وبالخطأ المعياري في حالة قياسات العينة . أي أن :

$$\sigma_{(y)}=\sqrt{\frac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N\ \text{or}\ N-1}}$$ للمجتمع

$$S_{(y)}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n\ \text{or}\ n-1}}$$ للعينة

مثال : لبيانات المثال السابق جد التباين والانحراف المعياري :

$$Y=2,5,4,3,3,7$$

الحل :

$$\sigma^2_{(y)} = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N} = \frac{(2-4)^2 + (5-4)^2 + (4-4)^2 + (3-4)^2 + (3-4)^2 + (7-4)^2}{6} = 2.667$$

وبطريقة أخرى :

$$V(Y) = \sigma^2_{(y)} = E(Y^2) - [E(Y)]^2$$

$$\because E(Y) = \overline{Y} = 4$$

$$\because E(Y^2) = \sum_i Y_i^2 P(Y_i) = (2)^2 \frac{1}{6} + (5)^2 \frac{1}{6} + (4)^2 \frac{1}{6} + (3)^2 \frac{2}{6} + (7)^2 \frac{1}{6} = \frac{112}{6} = 18.667$$

$$\Rightarrow V(Y) = 18.667 - (4)^2 = 2.667$$

وان الانحراف المعياري:

$$\sigma_{(y)} = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N}} = \sqrt{2.667} = 1.633$$

3- التباين المشترك : Covariance

وهو تباين واختلاف القيم عن بعضها البعض في مجموعتين (متغير مع متغير آخر) ، ويستخدم كمقياس نوعي لوجود العلاقة بين متغيرين . إن الصيغة الرياضة الخاصة بهذا المقياس تكون كالآتي :

$$Cov(X, Y) = E(XY) - E(X)E(Y) = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{N \text{ or } (N-1)}$$

ويمكن الاستنتاج من خلال العلاقة الأخيرة بان :

$$V(Y) = Cov(Y, Y) = E(YY) - E(Y)E(Y) = E(Y^2) - [E(Y)]^2$$

$$\Rightarrow Cov(Y, Y) = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})(Y_i - \overline{Y})}{N \text{ or } (N-1)} = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N \text{ or } (N-1)} = V(Y)$$

ملاحظة : يمكننا الاستنتاج بان التباين هو حالة خاصة من التباين المشترك ، والأمثلة الآتية توضح العلاقة ما بين التباين والتباين المشترك :

1- لتكن (a) كمية ثابتة فان:

- $V(Y+a)=V(Y)$

Sol : $V(Y+a)=Cov(Y+a,Y+a)=Cov(Y,Y+a)+Cov(a,Y+a)$

$=Cov(Y,Y)+Cov(Y,a)+Cov(a,Y)+Cov(a,a)$

$\because Cov(Y,a)=Cov(a,Y)$

$\therefore V(Y+a)=Cov(Y,Y)+2Cov(Y,a)+Cov(a,a)$

$\because Cov(Y,Y)=V(Y)$

$\because Cov(Y,a)=E(Ya)-E(Y)E(a)=aE(Y)-aE(Y)=0$

$\because Cov(a,a)=V(a)=E(aa)-E(a)E(a)=a^2-a^2=0$

$\Rightarrow V(Y+a)=V(Y)$

2- لتكن (a)و(b) كميات ثابتة فان :

$Cov(aX+bY,Z)=aCov(X,Z)+bCov(Y,Z)$

Sol : $Cov(aX+bY,Z)=Cov(aX,Z)+Cov(bY,Z)$

$\because Cov(aX,Z)=E(aXZ)-E(aX)E(Z)=aE(XZ)-aE(X)E(Z)$

$=a[E(XZ)-E(X)E(Z)]=aCov(X,Z)$

and $Cov(bY,Z)=bCov(Y,Z)$

$\Rightarrow Cov(aX+bY,Z)=aCov(X,Z)+bCov(Y,Z)$

3-اثبت ان :

$V(X-Y)=V(X)+V(Y)-2Cov(X,Y)$

$V(X-Y)=Cov(X-Y,X-Y)=Cov(X,X-Y)+Cov(-Y,X-Y)$

$\because Cov(X,X-Y)=Cov(X,X)+Cov(X,-Y)=V(X)+Cov(X,-Y)$

$\because Cov(X,-Y)=E[(X)(-Y)]-E(X)E(-Y)$

$=-E(XY)+E(X)E(Y)=-[E(XY)-E(X)E(Y)]=-Cov(X,Y)$

$\therefore Cov(X,X-Y)=V(X)-Cov(X,Y)$

and $Cov(-Y,X-Y)=Cov(-Y,X)+Cov(-Y,-Y)=-Cov(Y,X)+V(Y)$

$\Rightarrow V(X-Y)=V(X)-Cov(X,Y)-Cov(Y,X)+V(Y)$

$=V(X)+V(Y)-2Cov(X,Y)$

واجب : $V(X-2Y)=?$, $V[(X+Z)-Y]=?$