2- Les paramètres de dispersion

2-1 L'étendue

L'étendu, noté e, représente la différence entre les valeurs extrêmes de la distribution :

$e = x_n - x_1$.

Exemple : Soit la série suivante : 15 14 18 17 19 12 56 48 47 59

$e = xn – x1 = 59 – 12 = 47$.

2-2 La variance

C'est la moyenne arithmétique des carrés des écarts par rapport à la moyenne. Symbolisé par le signe (s^2) ou dans la littérature par (σ^2), et elle est donnée par la relation suivante :

$Sx^2 = 1/n\text{-}1 \sum(x_i - m)^2$

Exemple : Avec la série précédente

m = 30,5

$Sx^2 = 1/9 \sum (15\text{-}30{,}5)^2 + (14\text{-}30{,}5)^2 + \ldots + (59\text{-}30{,}5)^2$

$Sx^2 = 374{,}05$

2-3 L'écart-type

Aussi appelé déviation standard, c'est la racine carrée de la variance, symbolisé par le signe (S) ou (σ) et donné par : $Sx = \sqrt{\text{variance}}$

Avec l'exemple précédent on obtient

$Sx = \sqrt{374{,}05} = 19{,}34$

2-4 Le coefficient de variation

Il permet d'apprécier l'homogénéité des observations dans un échantillon. Symbolisé par le signe *V* ou C.V. et donné par la relation suivante :

C.V. ou $V = Sx / m \times 100$

Si le $V < 5\%$ il s'agit d'observations très homogènes ;

Si 5% < *V* < 10% il s'agit d'observations homogènes ;

Si 10% < *V* < 15% il s'agit d'observations moyennement homogènes ;

Si 15% < *V* < 30% il s'agit d'observations hétérogènes ;

Si le *V* > 30% il s'agit d'observations très hétérogènes.

Exemple

Avec la série précédente, *quel est le degré d'homogénéité entre les observations ?*

La solution

$V = Sx / m \times 100$ implique C.V. = 19,34 / 30,5 × 100 = 63,41%

Cette dernière valeur indique que les observations de la série sont très hétérogènes

Chapitre III. Statistique descriptive à deux dimensions

L'objectif de cette partie est d'étudier sur une même population de n individus, deux caractères différents X et Y et de rechercher s'il existe un lien entre ces deux variables.

Covariance

C'est la moyenne des produits des écarts pour chaque série d'observation.

$Cov(x,y) = S_{xy} = 1/n \sum(x_i - m)(y_i - \bar{y})$

La corrélation

La corrélation est la netteté ou l'intensité de la relation existante entre deux séries de données.

Le coefficient de corrélation (r)

Le coefficient de corrélation mesure la dépendance linéaire entre les variables X et Y. Le coefficient de corrélation, est précisément le rapport de la covariance sur le produit des écarts-types de deux variables X et Y.

$$r = Cov(x, y) / Sx . Sy$$

On a $-1 < r < 1$. Si r est proche de 1 ou -1, les variables X et Y sont dits : fortement corrélées.

Propriétés

Si le coefficient de corrélation est positif, les points du nuage sont alignés le long d'une droite croissante. Dans ce cas X et Y évoluent dans le même sens (figure 1).

Si le coefficient de corrélation est négatif, les points sont alignés le long d'une droite décroissante. Dans ce cas X et Y évoluent dans des sens opposés (figure 1).

Si le coefficient de corrélation est nul ou proche de zéro, il n'y a pas de dépendance linéaire (figure 1).

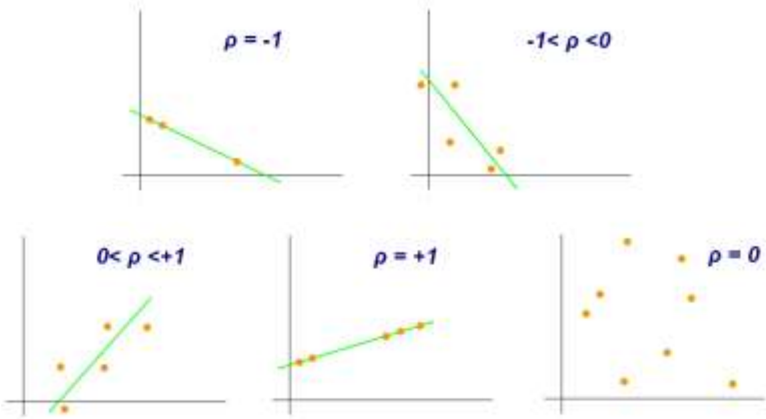

Figure 1: Exemples de diagrammes de dispersion avec différentes valeurs de coefficient de corrélation.

Exemple

L'étude de deux variables (X) le poids de 1000 grains en gramme et le rendement (Y) en quintaux par hectare chez le blé a donné les résultats suivants :

X (g)	43	46	48	50	52	55	56	58	60	62
Y (qx/ha)	30	33	35	36	37	39	39	42	43	45

1. Calculer la covariance Cov(xy) qui associe X et Y.
2. Calculer le coefficient de corrélation r_{xy} et quelle est votre conclusion ?

La solution

m= 53 Sx = 6,25

ȳ = 37,5 Sy= 4,65

Cov(xy) = 28,88.

r = 0,99.

Il existe une forte corrélation positive.