COURS. STATISTIQUE DESCRIPTIVE
Chapitre 04. Série Statistique Double

Département Mathématiques.

PRESENTATION
Dr. André Souleye Diabang

29 novembre 2021

Plan

Plan

Introduction

.

Les statistiques à une variable s'intéressaient, pour une population donnée, à un caractère donné : les notes à un devoir surveillé d'une classe, les salaires dans une entreprise, etc. . .

Introduction

.

Les statistiques à une variable s'intéressaient, pour une population donnée, à un caractère donné : les notes à un devoir surveillé d'une classe, les salaires dans une entreprise, etc...
Lorsque l'on s'intéresse à l'étude simultanée de deux caractères d'une même population, on fait ce que l'on appelle des statistiques à deux variables, en étudiant des séries statistiques doubles.

Introduction

.

On considère une population d'effectif n, si on étudie deux caractères X et Y de cette population,

Introduction

.

On considère une population d'effectif n, si on étudie deux caractères X et Y de cette population, on dit que l'on étudie une série statistique double.

Introduction

.

On considère une population d'effectif n, si on étudie deux caractères X et Y de cette population, on dit que l'on étudie une série statistique double.
Chaque individu de cette population est désigné par un nombre compris entre 1 et n.

Introduction

.

On considère une population d'effectif n, si on étudie deux caractères X et Y de cette population, on dit que l'on étudie une série statistique double.
Chaque individu de cette population est désigné par un nombre compris entre 1 et n. A chaque individu i $(1 \leq i \leq n)$ correspond un couple $(x_i; y_i)$, ou x_i est la modalité du caractère X et y_i est la modalité du caractère Y associé à l'individu i.

Introduction

.

On considère une population d'effectif n, si on étudie deux caractères X et Y de cette population, on dit que l'on étudie une série statistique double.
Chaque individu de cette population est désigné par un nombre compris entre 1 et n. A chaque individu i $(1 \leq i \leq n)$ correspond un couple $(x_i; y_i)$, ou x_i est la modalité du caractère X et y_i est la modalité du caractère Y associé à l'individu i. L'ensemble des couples $(x_i; y_i)$ définit une série statistique à deux variables.

Introduction

.

On considère une population d'effectif n, si on étudie deux caractères X et Y de cette population, on dit que l'on étudie une série statistique double.
Chaque individu de cette population est désigné par un nombre compris entre 1 et n. A chaque individu i $(1 \leq i \leq n)$ correspond un couple $(x_i; y_i)$, ou x_i est la modalité du caractère X et y_i est la modalité du caractère Y associé à l'individu i. L'ensemble des couples $(x_i; y_i)$ définit une série statistique à deux variables.
Lorsque l'un des deux caractères est une année, une date, on dit que la série statistique double est une série chronologique.

Application. B1

Exemple

Le tableau ci-dessous permet de suivre l'évolution de l'espérance de vie à la naissance (en années) d'un pays de 2012 à 2021 pour les femmes :

Année	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
rang (xi)	0	1	2	3	4	5	6	7	8	9
Espérance de vie (yi)	80,9	81,1	81,3	81,4	81,8	81,9	82,0	82,3	82,4	82,4

Plan

Nuage de points.

Définition.

Si l'on appelle $x1$, $x2$, . . ., xn les n valeurs du premier caractère (on notera cette série (xi)),

Nuage de points.

Définition.

Si l'on appelle $x1$, $x2$, . . ., xn les n valeurs du premier caractère (on notera cette série (xi)),et si l'on appelle $y1$, $y2$, . . ., yn les n valeurs du second caractère (on notera cette série (yi)),

Nuage de points.

Définition.

Si l'on appelle $x1$, $x2$, ..., xn les n valeurs du premier caractère (on notera cette série (xi)),et si l'on appelle $y1$, $y2$, ..., yn les n valeurs du second caractère (on notera cette série (yi)), alors on représente cette série statistique double par un nuage de points dans un repère du plan, constitué des points Mi de coordonnées (xi, yi)

Nuage de points.

Définition.

Si l'on appelle $x1$, $x2$, . . ., xn les n valeurs du premier caractère (on notera cette série (xi)),et si l'on appelle $y1$, $y2$, . . ., yn les n valeurs du second caractère (on notera cette série (yi)), alors on représente cette série statistique double par un nuage de points dans un repère du plan, constitué des points Mi de coordonnées (xi, yi)

Exemple.

Représenter le nuage de points de la série B1.

Plan

Le point moyen.

Définition.

Soit $\overline{x}$ la moyenne de la série (xi) et $\overline{y}$ la moyenne de la série (yi).

Le point moyen.

Définition.

Soit $\overline{x}$ la moyenne de la série (xi) et $\overline{y}$ la moyenne de la série (yi). Le point G de coordonnées $(\overline{x}, \overline{y})$ est appelé point moyen du nuage de points associé cette série statistique double. Ainsi

Le point moyen.

Définition.

Soit $\overline{x}$ la moyenne de la série (xi) et $\overline{y}$ la moyenne de la série (yi). Le point G de coordonnées $(\overline{x}, \overline{y})$ est appelé point moyen du nuage de points associé cette série statistique double. Ainsi

$$G(\overline{x}, \overline{y})$$

avec

Le point moyen.

Définition.

Soit $\overline{x}$ la moyenne de la série (xi) et $\overline{y}$ la moyenne de la série (yi). Le point G de coordonnées $(\overline{x}, \overline{y})$ est appelé point moyen du nuage de points associé cette série statistique double. Ainsi

$$G(\overline{x}, \overline{y})$$

avec

$$\overline{x} = \frac{\sum x_i}{N}$$

Le point moyen.

Définition.

Soit $\overline{x}$ la moyenne de la série (xi) et $\overline{y}$ la moyenne de la série (yi). Le point G de coordonnées $(\overline{x}, \overline{y})$ est appelé point moyen du nuage de points associé cette série statistique double. Ainsi

$$G(\overline{x}, \overline{y})$$

avec

$$\overline{x} = \frac{\sum x_i}{N} \quad et \quad \overline{y} = \frac{\sum y_i}{N}$$

Le point moyen.

Définition.

Soit $\overline{x}$ la moyenne de la série (xi) et $\overline{y}$ la moyenne de la série (yi). Le point G de coordonnées $(\overline{x}, \overline{y})$ est appelé point moyen du nuage de points associé cette série statistique double. Ainsi

$$G(\overline{x}, \overline{y})$$

avec

$$\overline{x} = \frac{\sum x_i}{N} \quad et \quad \overline{y} = \frac{\sum y_i}{N}$$

Exemple

Déterminer le point moyen de la série B1.

Plan

Définition

Covariance de (x,y) et les variances

On appelle covariance de la série statistique double de variables x et y le nombre réel :

$$cov(x,y) = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{N} = \frac{\sum x_i y_i}{N} - \overline{xy}$$

Définition

Covariance de (x,y) et les variances

On appelle covariance de la série statistique double de variables x et y le nombre réel :

$$cov(x,y) = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{N} = \frac{\sum x_i y_i}{N} - \overline{xy}$$

les variances de x et de y sont

$$V(x) = cov(x,x) \quad et \quad V(y) = Cov(y,y)$$

Remarque

La variance d'une variable est positive et le nombre $\sqrt{V(x)} = \sigma_x$ est l'écart-type de x.

Définition

Droites de régressions. Par la méthode des moindres carrés

- La droite de régression D_1 de y en x est l'équation

Définition

Droites de régressions. Par la méthode des moindres carrés

- La droite de régression D_1 de y en x est l'équation $y = ax + b$ avec $a = \frac{cov(x,y)}{V(x)}$ et $b = \overline{y} - a\overline{x}$.

Définition

Droites de régressions. Par la méthode des moindres carrés

- La droite de régression D_1 de y en x est l'équation $y = ax + b$ avec $a = \frac{cov(x,y)}{V(x)}$ et $b = \overline{y} - a\overline{x}$.
- La droite de régression D_2 de x en y est l'équation

Définition

Droites de régressions. Par la méthode des moindres carrés

- La droite de régression D_1 de y en x est l'équation $y = ax + b$ avec $a = \frac{cov(x,y)}{V(x)}$ et $b = \overline{y} - a\overline{x}$.
- La droite de régression D_2 de x en y est l'équation $x = a'x + b'$ avec $a' = \frac{cov(x,y)}{V(y)}$ et $b' = \overline{x} - a'\overline{y}$.

Définition

Droites de régressions. Par la méthode des moindres carrés

• La droite de régression D_1 de y en x est l'équation
$y = ax + b$ avec $a = \frac{cov(x,y)}{V(x)}$ et $b = \overline{y} - a\overline{x}$.
• La droite de régression D_2 de x en y est l'équation
$x = a'x + b'$ avec $a' = \frac{cov(x,y)}{V(y)}$ et $b' = \overline{x} - a'\overline{y}$.

Remarque

Le point moyen G du nuage appartient toujours aux deux droites de régressions

Définition

Le coefficient de corrélation.

Le coefficient de corrélation linéaire d'une série statistique de variables x et y est le nombre r défini par :

Définition

Le coefficient de corrélation.

Le coefficient de corrélation linéaire d'une série statistique de variables x et y est le nombre r défini par :

$$r = \frac{cov(x,y)}{\sigma(x).\sigma(y)}.$$

Coefficient de corrélation.

Interprétation.

• Ce coefficient sert à mesurer la qualité d'un ajustement affine.

• Plus le coefficient de corrélation linéaire est proche de 1 en valeur absolue, meilleur est l'ajustement linéaire.

Coefficient de corrélation.

Interprétation.

• Ce coefficient sert à mesurer la qualité d'un ajustement affine.
• Plus le coefficient de corrélation linéaire est proche de 1 en valeur absolue, meilleur est l'ajustement linéaire.
• Lorsque $r = \pm 1$, la droite de régression passe par tous les points du nuage, qui sont donc alignés.

Coefficient de corrélation.

Interprétation.

• Ce coefficient sert à mesurer la qualité d'un ajustement affine.
• Plus le coefficient de corrélation linéaire est proche de 1 en valeur absolue, meilleur est l'ajustement linéaire.
• Lorsque $r = \pm 1$, la droite de régression passe par tous les points du nuage, qui sont donc alignés.
• Le coefficient de corrélation linéaire r vérifie $-1 \leq r \leq 1$.