Université Jinan
Faculté de Gestion
Tripoli - Liban

Statistiques

Examen Préparatoire

Version 1

Statistiques

Université de Jinan
Faculté de Gestion

Table des matières

Chapitre 1

Analyse statistique d'une variable

1.1 Données Statistiques

Les données statistiques sont présenté comme suit :

- sous la forme $\{(x_i, n_i) : 1 \leq i \leq p\}$ dans le cas d'une variable discrète avec $x_1 < x_2 < \cdots < x_p$;
- sous la forme $\{([a_i, a_{i+1}[, n_i) : 1 \leq i \leq p\}$ dans le cas d'une variable continue.

n_i est **l'effectif** de la valeur x_i (ou de la **classe** $[a_i, a_{i+1}[$).

On a évidemment $\displaystyle\sum_{i=1}^{p} n_i = n$.

1.2 Représentations graphiques

1.2.1 Variable discrète

DIAGRAMME EN BÂTONS

Le diagramme en bâtons des effectifs (resp. des fréquences) de la distridution statistique $\{(x_i, n_i) : 1 \leq i \leq p\}$ s'obtient en traçant les "bâtons" A_iB_i, c'est-à-dire les segments joignant les points $A_i(x_i, 0)$ $B_i(x_i, n_i)$ (resp. $B_i(x_i, f_i)$) pour $1 \leq i \leq p$.

POLYGONE DES EFFECTIFS OU DES FRÉQUENCES

Le polygone des effectifs (resp. des fréquences) de la distridution statistique $\{(x_i, n_i) : 1 \leq i \leq p\}$ s'obtient en joignant les points $B_i(x_i, n_i)$ (resp. $B_i(x_i, f_i)$) pour $1 \leq i \leq p$.

COURBES CUMULATIVE

La courbe cumulative (C) de la distridution statistique $\{(x_i, n_i) : 1 \leq i \leq p\}$ est la courbe représentative de la fonction F définie comme suit :
Si $x < x_1$, $F(x) = 0$.
Si $x_1 \leq x < x_2$, $F(x) = p_1$.
Si $x_2 \leq x < x_3$, $F(x) = p_1 + p_2$.
...
Si $x_n \leq x$, $F(x) = p_1 + ... + p_n = 1$.
F est donc une fonction en escalier, effectuant un "saut" à chaque point x_i.

1.2.2 Variable continue

HISTOGRAMME

L'histogramme des effictifs (resp. des fréquences) de la distibution statistique $\{([a_i, a_{i+1}[, n_i) : 1 \leq i \leq p\}$ s'obtient en traçant pour tout $1 \leq i \leq p$ le rectangle de base A_iA_{i+1} (A_i étant le point $(x_i, 0)$) et d'aire proportionnelle à n_i (resp. f_i), et donc de hauteur proportionnelle à $\dfrac{n_i}{a_{i+1} - a_i}$.
Remarque : si les classes ont toutes même amplitude, les hauteurs des rectangles sont proportionnelle aux n_i.

POLYGONE DES EFFECTIFS OU DES FRÉQUENCES

Le polygone des effictifs (resp. des fréquenes) de la distribution statistique $\{([a_i, a_{i+1}[, n_i) : 1 \leq i \leq p\}$ s'obtient en joignant les points $B_i(c_i = \dfrac{a_i + a_{i+1}}{2}, n_i)$ (resp. $B_i(c_i, f_i)$) pour $1 \leq i \leq p$.

COURBES CUMULATIVE

La courbe cumulative des fréquences de la distribution statistique $\{([a_i, a_{i+1}[, n_i) : 1 \leq i \leq p\}$ s'obtient en joignant les points $B_i(a_{i+1}, \sum_{j \leq i} f_j)$ pour i variant de 1 à p.

On obtient de même la courbe cumulative des effectifs.

1.3 Valeurs caractéristiques

1.3.1 Paramètres de position (ou de tendance centrale)

MODE

- Si X est une variable discrète, on appelle **mode** toute valeur x_i dont l'effectif (ou la fréquence) est maximum.
- Si X est une variable continue, on appelle **classe modale** toute classe $[a_i, a_{i+1}[$ pour la quelle $\dfrac{n_i}{a_{i+1} - a_i}$ (ou $\dfrac{f_i}{a_{i+1} - a_i}$) est maximum (sur l'histogramme, la hauteur du rectangle correspondant est maximum).
  Remarque : Il peut y avoir plusieurs modes (ou plusieurs classes modales)

MÉDIANE

- Si X est une variable discrète prenant N valeurs : $v_1 \leq v_2 \leq ... \leq v_N$ (les valeurs prises par X ne sont pas groupées), on appelle **médiane** un nombre réel m_e tel qu'il y ait autant des valeurs v_j inférieurs ou égales à m_e que des valeurs v_j inférieurs ou égales à m_e.
  Si $N = 2k + 1$, $m_e = v_{k+1}$
  Si $N = 2k$, $m_e = \dfrac{v_k + v_{k+1}}{2}$.

- Si X est une variable continue, on appelle **médiane** le nombre réel m_e abscisse du point d'ordonnée $\dfrac{1}{2}$ de la courbe cumulative des fréquences, c'est-à-dire le nombre réel solution de l'équation $F(x) = \dfrac{1}{2}$.
  m_e appartient à la première classe $[a_i, a_{i+1}[$ dont la fréquence cumulée p_i est supérieure ou égale à 0,5 (on a donc : $p_{i-1} < 0,5$ et $p_i \geq 0,5$).
  La valeur de m_e s'obtient en résolvant : $\dfrac{m_e - a_i}{a_{i+1} - a_i} = \dfrac{0,5 - p_{i-1}}{p_i - p_{i-1}}$, donc

$$m_e = a_i + \frac{0,5 - p_{i-1}}{p_i - p_{i-1}}(a_{i+1} - a_i)$$

MOYENNE

- Si X est une variable discrète de distridution $\{(x_i, n_i) : 1 \leq i \leq p\}$, on appelle **moyenne** le nombre réel : $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i$.
- Si X est une variable continue de distibution, $\{([a_i, a_{i+1}[, n_i) : 1 \leq i \leq p\}$. La moyenne se calcule comme pécédemment, en remplaçant x_i par le centre c_i de la classe $[a_i, a_{i+1}[$ ($c_i = \dfrac{a_i + a_{i+1}}{2}$).

1.3.2 Paramètres de dispersion

VARIANCE ET ÉCART TYPE

La **variance** est le nombre réel positif : $V = \overline{x^2} - \overline{x}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{p} n_i x_i^2 - \overline{x}^2 = \sum_{i=1}^{p} f_i x_i^2 - \overline{x}^2$ ou encore $V = \dfrac{1}{n}\displaystyle\sum_{i=1}^{p} n_i (x_i - \overline{x})^2 = \sum_{i=1}^{p} (f_i x_i - \overline{x})^2$. L'écart type est $\sigma = \sqrt{V}$.

1.4 Exercises

Soit le tableau statistique suivant

x_i	[155 ;160[	[160 ;165[	[165 ;170[	[170 ;175[	[175 ;180[
n_i	4	6	12	5	3

Trouver la moyenne, la médiane, la classe modale, la variance et l'ecart-type.

Answers :

la moyenne est 167, la médiane est 167,08, la classe modale est [165 ;170[.

Chapitre 2

Analyse statistique de deux variables

2.1 Présentation des données

Soient X et Y deux variables quantitatives définies sur une population P. L'étude d'une échantillon de taille n donne n couples de valeurs prises par le couple (X, Y) qui constituent la distribution statistique du couple (X, Y). Cette distribution est présenté comme suit :

• Sous forme de **données non groupées :**

individu	1	$\cdots$	i	$\cdots$	n
X	x_1	$\cdots$	x_i	$\cdots$	x_n
Y	y_1	$\cdots$	y_i	$\cdots$	y_n

(La première ligne de ce tableau est parfois omise).

2.2 Valeurs caractéristiques

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$
$$V(X) = \sigma_X^2 = \overline{X^2} - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{X}^2$$
$$V(Y) = \sigma_Y^2 = \overline{Y^2} - \overline{Y}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \overline{Y}^2$$
$$Cov(X,Y) = \overline{XY} - \overline{X}.\overline{Y}$$
$$r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

2.3 Représentation graphique

Dans un repère orthogonal, à chaque individu de l'échantillon on associe le point $M(x,y)$, x et y étant les valeurs prises par X et Y pour cet individus. Si (X,Y) prend la valeur (x,y) pour m individus, on dessine m points autour du point du coordonnées (x,y), ou un disque d'aire proportionnelle à m. On obtient ainsi le **nuage de points** représentant la distribution statistique.
Le point $G(\overline{X},\overline{Y})$ est appelé **point moyen** du nuage.

2.4 Ajustement linéaire

On appelle **droite de régression** de y en x ou droite des moindres carrés de y en x La droite suivante : (Δ) : $y = ax + b$ qui "ajuste" le nuage. On dit qu'on effectue un **ajustement linéaire.**
Où $a = \dfrac{Cov(X,Y)}{V(X)}$ et $b = \overline{Y} - a\overline{X}$.

2.5 Exercices

Soit le tableau statistique suivant :

x_i	480	450	480	540	570	420	390	520	470	480
y_i	22	18	20	24	24	22	14	22	18	16

1. Calculer $\overline{X}$, $\overline{Y}$, $V(X)$, $V(Y)$, σ_X et σ_Y.
2. Calculer $cov(X.Y)$ et r.

Réponse :

$\overline{X} = 480$, $\overline{Y} = 20$, $V(X) = 2600$, $V(Y) = 10,4$, $cov(X.Y) = 118$ et $r = 0,72$.