INSTITUT SUPERIEUR
DU MANAGEMENT

MAE

# Master Administration d’Entreprise

# Techniques quantitatives de gestion

**Intervenant**

DIA Nicolas



Abidjan , Cocody,Bvd F. Mitterand, Riviera Bonoumin, Immeuble La Paix
22 BP 876 Abidjan 22 Tél 20 01 05 75 Fax : 22 47 07 04
E-mail :ismconseil.ci@gmail.com Web: http:sites.google.com/site/ismcons

## SOMMAIRE

## Chapitre I. TRAITEMENT DE DONNEES STATISTIQUES

Présentation des données en tableau

- Caractère qualitative
  Il s'agit des caractères dont les modalités ne sont pas mesurables. Exemple : le caractère sexe dont les modalités sont : masculin, féminin

| Modalités du caractère étudié $x_i$ | Effectif associé $n_i$ | Fréquences relatives $f_i$ |
|---|---|---|
| $x_i$ | $n_1$ | $f_1$ |
| ……………………………….. | …………………………. | …………………………….. |
| $x_n$ | $n_n$ | $f_n$ |

L'effectif $n_i$ est le nombre de fois ou la modalité $x_i$ est observée

La fréquence relative $f_i$ est le pourcentage d'apparition de la modalité$x_i$. Soit N l'effectif total, $f_i = \frac{n_i}{N}$. On remarquera également que $\sum f_i = 1$
La production d'aluminium en 1988, dans cinq pays industrialisés, était la suivante (en milliers de tonnes)

| Pays | Production |
|---|---|
| U.S.A | 5 262 |
| Russie | 2 415 |
| Canada | 1 347 |
| Japon | 1 099 |
| France | 536 |

1 – Calculer la production totale
2- Calculer la proportion produite par chacun des pays

Correction

1 – Il suffit de sommer la colonne pour obtenir la production totale c'est-à-dire :10659000 tonnes

2- La proportion produite par chaque pays se calcule en faisant le quotient de la production d'un pays par la production totale.

| | $n_i$ | $f_i$ (en %) |
|---|---|---|
| USA | 5262 | 49,4 |
| URSS | 2415 | 22 ,7 |
| CANADA | 1347 | 12,6 |
| JAPON | 1099 | 10,3 |
| FRANCE | 536 | 5,0 |
| Total | 10 659 | 100,0 |

- Caractère quantitatif

Ce sont les caractères dont les modalités sont mesurables. On distingue à ce niveau deux types de caractère quantitatif, à savoir le quantitatif continu et le quantitatif discret.
– Le caractère quantitatif continu est celui dont les modalités sont regroupées dans des classes de type [a ; b [.
-Le caractère quantitatif discret est celui dont les modalités prennent des valeurs isolées
On appelle fréquences ou effectifs cumulés croissantes, le cumul des fréquences ou des effectifs associés aux valeurs du caractère strictement inférieures à $x_i$. Le cumul s'effectue du haut vers le bas du tableau et correspond à la notion « moins de »
On appelle fréquences ou effectifs cumulés décroissantes, le cumul des fréquences ou des effectifs associés aux valeurs du caractère strictement supérieures à $x_i$. Le cumul s'effectue du bas vers le haut du tableau et correspond à la notion « plus de »
Exemple : Lors d'une étude sur la population d'une ville, on a recensé le nombre de personnes par ménage. On a obtenu le tableau suivant :

| Nombre de personnes | 1 | 2 | 3 | 4 | 5 | 6 | 7 et plus |
|---|---|---|---|---|---|---|---|
| Nombre de ménages | 2 327 | 4 533 | 8 918 | 10 405 | 6 210 | 2 134 | 1 123 |

Calculer les fréquences et les fréquences cumulées croissantes.

Correction

Pour calculer les fréquences, on fait le quotient $n_i$ /n, ou ni est l'effectif associé à la modalité xi et n l'effectif total. Pour calculer les fréquences cumulées croissantes, on calcule pour chaque xi, la proportion de ménages ayant strictement moins de xi personnes. Le tableau suivant doit être interprété de la façon suivante : par exemple la valeur 0,192 est attachée à la valeur 3 du caractère. Donc 19 ,2% des ménages sont composés de moins de 3 personnes c'est-à-dire sont composés de 1 ou 2 personnes :

| xi | $n_i$ | $f_i$ | $F_i$ ↗ | $F_i$ ↘ |
|---|---|---|---|---|
| 1 | 2327 | 0,065 | 0,065 | 1,000 |
| 2 | 4533 | 0,127 | 0,192 | 0,935 |
| 3 | 8918 | 0,250 | 0,442 | 0,808 |
| 4 | 10405 | 0,292 | 0,734 | 0,558 |
| 5 | 6210 | 0,174 | 0,908 | 0,266 |
| 6 | 2134 | 0,060 | 0,963 | 0,092 |
| 7 | 1123 | 0,032 | 1,000 | 0,032 |
| Total | 35650 | 1,000 | | |

## Chapitre II. LES PARAMETERES DE POSITION ET DE DISPERSION

Les paramètres de position ou de tendance centrale (moyenne, médiane, mode) sont des valeurs qui résument une distribution statistique. Ils s'expriment dans la même unité que les observations. Les paramètres de dispersion (variance, écart type, coefficient de variation…) indiquent la disposition des observations autour des valeurs centrales

### 1- Les paramètres de position

- La moyenne

Soit une distribution $\{x_i\}$ i variant de 1 à n, d'effectif$n_i$. La moyenne arithmétique est le rapport : $\overline{\boldsymbol{X}} = \frac{\sum \boldsymbol{nixi}}{\sum \boldsymbol{ni}}$ Dans le cas d'une distribution continue, $x_i$ représente le centre des classes

- Le médiane

La médiane est la valeur du caractère qui partage la série en deux sous série de même taille (50% - 50%). Ce qui signifie que 50% des valeurs observées est inférieure à la médiane $M_e$

- Le mode

Le mode ou valeur probable est l'observation ayant l'effectif le plus élevé

2- Les paramètres de dispersion

- La variance

La variance est la moyenne arithmétique du carré des écarts à la moyenne arithmétique. Var(x)= $\frac{1}{n}\sum_{1}^{n}(xi - \bar{x})^2$

- L'écart-type

L'écart type est la racine carrée de la variance. Il détermine la dispersion des observations autour de la moyenne. σ(x) = $\sqrt{\mathrm{Var(x)}}$ = $\sqrt{\frac{1}{n}\sum_{1}^{n}(xi - \bar{x})^2}$

Exemples

La répartition des employés d'une entreprise suivant la prime de fin d'année à permis de dresser le tableau suivant

| PRIME EN F | NOMBRE D'EMPLOYES |
|---|---|
| [1250 ; 1750[ | 130 |
| [1750 ; 2250[ | 350 |
| [2250 ; 2750[ | 210 |
| [2750 ; 3250[ | 130 |
| [3250 ; 3750[ | 90 |
| [3750 ; 4250[ | 50 |
| [4250 ; 4750[ | 30 |
| [4750 ; 5250[ | 10 |

1- Quelle est l'étendue de cette série ?
2- Calculer la prime moyenne
3- Calculer les quartiles

4- Calculer l’écart interquartile
5- Calculer l’écart inter décile
6- Calculer l’écart-type

Correction

| PRIME EN F | ni | $x_i$ | $f_i$ | $F_i$ | $f_ix_i$ | $x_i^2 /10^4$ | $f_ix_i^2 /10^2$ |
|---|---|---|---|---|---|---|---|
| [1250 ; 1750[ | 130 | 1500 | 0,13 | 0,13 | 195 | 225 | 2925 |
| [1750 ; 2250[ | 350 | 2000 | 0,35 | 0,48 | 700 | 400 | 14000 |
| [2250 ; 2750[ | 210 | 2500 | 0,21 | 0,69 | 525 | 625 | 13125 |
| [2750 ; 3250[ | 130 | 3000 | 0,13 | 0,82 | 390 | 900 | 11700 |
| [3250 ; 3750[ | 90 | 3500 | 0,09 | 0,91 | 315 | 1225 | 11025 |
| [3750 ; 4250[ | 50 | 4000 | 0,05 | 0,96 | 200 | 1600 | 8000 |
| [4250 ; 4750[ | 30 | 4500 | 0,03 | 0,99 | 135 | 2025 | 6075 |
| [4750 ; 5250[ | 10 | 5000 | 0,01 | 1,00 | 50 | 2500 | 2500 |
| Total | 1000 | | 1,00 | | 2500 | | 69350 |

1- L’étendue de la série correspond à la différence entre la limite supérieure de la dernière classe et la limite inférieure de la première classe c'est-à-dire e = 5250 – 1250 = 4000 f
2- La prime moyenne est la moyenne arithmétique : $\bar{x} = \frac{\sum \text{nixi}}{\sum \text{ni}} = \sum \text{fixi} = 2500$ f
3- Il y a trois quartiles notés $Q_{25}$ , $Q_{50}$ et $Q_{75}$

$Q_{25}$= $1750+500.\frac{12}{35} \approx$ 1921,43 f ; $Q_{50}$= $2250 + 500.\frac{0{,}50-0{,}48}{0{,}69-0{,}48}$ ≈2297,62 f
$Q_{75}$= $2750 +5100.\frac{0{,}75-0{,}69}{0{,}82-0{,}69} \approx$ 2980,77 f

- $Q_{25}$ = 1921,43 f signifie que 25% des salariés touchent une prime inférieure à 1921,43 f
- $Q_{50}$ = 2297,62 f signifie que 50% des salariés touchent une prime inférieure à 2297,62 f
- $Q_{75}$ = 2980,77 f signifie que 75% des salariés touchent une prime inférieure à 2980,77 f

4- Les bornes de l’intervalle interquartile sont $Q_{25}$ et $Q_{75}$. L’écart interquartile est égal à $Q_{75} – Q_{25}$= 2980,77 – 1921,43= 1059,34 f
L’intervalle interquartile [1921,43 ; 2980,77] contient 50% des valeurs observées

5- Pour calculer l'écart interdécile, il faut calculer deux déciles qui correspondent à : $Q_{10}$ et $Q_{90}$. $Q_{10}$ est dans la classe [1250 ; 1750[ : $\frac{Q10-1250}{1750-1250} = \frac{0,10}{0,13}$ => $Q_{10}$≈ 1634,62 f

Q90 est dans la classe [3250 ; 3750[ : $\frac{Q90-3250}{3750-3250} = \frac{0,90}{0,91}$ => $Q_{90}$≈ 3694,44 f. L'écart interdécile est donc : $Q_{90} - Q_{10}$ = 3694,44 – 1634,62= 2059,83f

6- La variance est égale à : Var(x) = $\frac{\sum ni(xi-\overline{x})^2}{\sum ni}$ = $\sum fixi^2$-( $\sum fixi$)²= 6935000 – 2510²= 634900 L'écart type est égal à σ≈ 796,81 f. Ce qui signifie que la dispersion autour de la moyenne est de 796,81 f.

Exercice

Les bénéfices d'une entreprise ont augmenté :
de 5%, par an, pendant les 2 premières années
de 9%, par an, pendant les 5 années suivantes
de 12%, par an, pendant les 3 années suivantes

Quelle est l'augmentation moyenne sur les 10 ans ?

Correction

Soit Bo le bénéfice réalisé par l'entreprise l'année précédant la période écoulée.(Bo ≠0). Le bénéfice était :

1 an plus tard : $B_1$ = Bo x 1,05
2 ans plus tard : $B_2$= $B_1$ x 1,05 = Bo x (1,05)²
3ans plus tard : $B_3$ =$B_2$ x 1,09 = Bo x (1,05)²x 1,09

……………………………………………………………………………………………
……………………………..

10 ans plus tard : $B_{10}$ = Bo x $(1{,}05)^2$x $(1{,}09)^5$ x $(1{,}12)^3$

Soit a l'augmentation annuelle moyenne sur les 10 ans. On doit avoir : $B_{10}$ = $\text{Bo}(1+a)^{10}$ = Bo x $(1{,}05)^2$x $(1{,}09)^5$ x $(1{,}12)^3$ => $(1+a)^{10}$ = $(1{,}05)^2$x $(1{,}09)^5$x $(1{,}12)^3$ => 1+a = $[\,(1{,}05)^2 \text{x}\,(1{,}09)^5 \text{x}\,(1{,}12)^3\,]^{\frac{1}{10}}$ = 1,090728

**a = 9%**

## Chapitre III. DISTRIBUTION STATISTIQUE A DEUX DIMENSIONS

Soit deux caractères X et Y définis sur une même population d'effectif total n. La distribution statistique à deux dimensions relative au couple () est définie par la donnée :

- des P valeurs possibles $x_{1,}$ $x_{2,}$ $x_{3,}\ldots\ldots\ldots\ldots,$ $x_p$ de X
- des q valeurs possibles $y_{1,}$ $y_{2,}$ $y_{3,}\ldots\ldots\ldots\ldots,$ $y_q$ de Y
- des P x q effectifs correspondants aux observations (X = $x_{i,}$ et $Y = y_{j,}$) notée $n_{ij}$

### I- ETUDE D'UNE DISTRIBUTION STATISTIQUE DOUBLE

1. Présentation en tableau

- Tableau de données ponctuelles

| Observation numéro | Valeur de X | Valeur de Y |
|---|---|---|
| 1 | $x_1$ | $y_1$ |
| $x_i$ | $x_i$ | $y_i$ |
| n | $x_n$ | $y_n$ |

- Tableau à double entrée ou de contingence

| valeurs De X / valeurs de Y | $y_1$ | … | $y_j$ | … | $y_q$ |
|---|---|---|---|---|---|
| $x_1$ | $n_{11}$ | | $n_{1j}$ | | $n_{1q}$ |
| … | … | … | … | … | … |
| $x_i$ | $n_{i1}$ | | $n_{ij}$ | | $n_{iq}$ |
| … | … | … | … | … | … |
| $x_p$ | $n_{p1}$ | | $n_{pj}$ | | $n_{pq}$ |

A l'intersection de la $i^{ème}$ ligne et de la $j^{ème}$ colonne, on reporte l'effectif $n_{ij}$ correspondant à l'observation conjointe de X = $x_i$ et $Y = y_j$

2- Distributions marginales

De la distribution statistique du couple (X, Y), on peut déduire les distributions de X seul et Y seul. En indiquant par un point une sommation effectuée suivant l'indice i ou l'indice j, on note :

- $n_{i\bullet}$ le total des effectifs de la ligne i : $n_{i\bullet}= n_{i1}+ n_{i2} +\ldots\ldots + n_{iq} = \sum_{j=1}^{q} n_{ij}$
- $n_{\bullet j}$ le total des effectifs de la colonne j : $n_{\bullet j}= n_{1j}+ n_{2j} +\ldots\ldots + n_{pj} = \sum_{i=1}^{p} n_{ij}$
- $n_{\bullet\bullet}$ l'effectif total : des effectifs de la ligne i : $n_{\bullet\bullet} = \sum_{i=1}^{p} n_{i\bullet} = \sum_{j=1}^{q} n_{\bullet j} = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}$ = n

$(x_i, n_{i\bullet})$ est la distribution marginale de X
$(y_j, n_{\bullet j})$ est la distribution marginale de Y
La fréquence de l'observation $(X = x_i \text{et}\, Y = y_j)$est notée $f_{ij}$ : $f_{ij} = \frac{n_{ij}}{n}$.
La fréquence marginale de l'observation $(X = x_i)$ est notée $f_{i\bullet}$ : $f_{i\bullet} = \frac{n_{i\bullet}}{n} = \sum_{j=1}^{q} f_{ij}$
La fréquence marginale de l'observation $(Y = y_j)$ est notée $f_{\bullet j}$ : $f_{\bullet j} = \frac{n_{\bullet j}}{n} = \sum_{i=1}^{p} f_{ij}$
NB : Deux caractères X et Y sont dits indépendants ssi $\forall$ i $\epsilon$ {1,…., p} ; $\forall$ j $\epsilon$ {1,…., q} on a : $f_{ij} = f_{i\bullet}$ x $f_{\bullet j}$ ou $n_{ij} = \frac{n_{i\bullet} \text{ x } n_{\bullet j}}{n}$

3- Représentation graphique : le nuage de points.
On représente dans un repère orthogonal, l'ensemble des données ; chaque individu est alors représenté par un point de coordonnées $x_i$ et$y_i$. L'ensemble des points ainsi obtenus forme une sorte de nuage plus ou moins compact, plus ou moins étalé. Si les points sont groupés, il existe une relation étroite entre les deux variables ; si au contraire, ils sont très dispersés, les deux phénomènes sont indépendants. Cette représentation par un nuage de points est prolongée par l'étude de la corrélation entre les deux caractères observés

Exercice 1

On considère la statistique double définie par le tableau de contingence suivant

| X \ Y | Y1 | Y2 | Y3 |
|---|---|---|---|
| X1 | 128 | 81 | 17 |
| X2 | 64 | 22 | 88 |

Déterminer n , $n_{1.2}$, $n_{2.1}$, $n_{2.}$, $n_{2.}$, $f_{2.3}$, $f_{3.2}$

Correction

n = 128+81+17+64+22+88 = 400 ; $n_{1.2}$ = 81 ; $n_{2.1}$= 64 ; $n_{2.}$ = 64+22+88 = 174 ; $n_{.2}$ = 81+22 = 103 ; $f_{2.3} = \frac{n2.3}{n..} = \frac{88}{400}$ = 0,22 et $f_{.1} = \frac{n2.3}{n..} = \frac{192}{400}$ = 0,48

Exercice 2

Soit la distribution statistique double suivante

| X \ Y | 15 | 25 | 35 | 45 |
|---|---|---|---|---|
| -1 | 40 | 168 | 320 | 68 |
| 0 | 84 | 28 | 48 | 16 |
| 1 | 128 | 32 | 24 | 44 |

Déterminer les distributions marginales ; calculer leurs moyennes et variances

Correction

- Distribution marginale de X

| Xi | -1 | 0 | 1 | ∑ | ∑/n |
|---|---|---|---|---|---|
| ni. | 596 | 176 | 228 | 1000 | 1 |
| ni. xi | -596 | 0 | 228 | -368 | -0,368 |
| ni. xi² | 596 | 0 | 228 | 824 | 0,824 |

$\bar{x}$ = -0,368 et Var(x) = 0,824 - $(-0,368)^2 \approx$ 0,689

- Distribution marginale de Y

| yj | 15 | 25 | 35 | 45 | ∑ | ∑/n |
|---|---|---|---|---|---|---|
| n.j | 252 | 228 | 392 | 128 | 1000 | 1 |
| n.jyj | 3780 | 5700 | 13720 | 5760 | 28960 | 28,96 |
| n.j yj² | 56700 | 142500 | 480200 | 259200 | 938600 | 938,6 |

$\bar{Y}$= 28,96 et Var(Y) = 938,96 – $(28,96)^2 \approx$ 99,9

Exercice 3

Dans une université une enquête sur le tabagisme donne les résultats consignés dans le tableau ci –dessous

| X \ Y | Hommes | Femmes |
|---|---|---|
| Fumeurs | 420 | 75 |
| Non-fumeurs | 280 | 225 |

Déterminer

1- L'effectif de la population interrogée
2- Le nombre d'hommes qui fument
3- Le nombre de femmes
4- La proportion de fumeurs
5- La proportion de fumeurs parmi les hommes
6- La proportion de fumeurs parmi les femmes
7- X et Y sont ils indépendants ?

Correction

1- Effectif total : n = 420+75+280+225 = 1000
2- Le nombre d'hommes qui fument : $n_{1.1}$ = 420
3- Le nombre de femmes : $n_{.2}$ = 75+225 = 300
4- La proportion de femmes qui ne fument pas : $f_{2.2}$ = 225/1000 = 0,225 soit 22,5%
5- La proportion de fumeurs : $f_{1.}$= (420+75)/1000 = 0,495 soit 49,5%

6- Proportion de fumeurs parmi les hommes : $f_{x1/y1}$ = 420/ (420+280)= 0,60 soit 60%
7- Proportion de femmes parmi les fumeurs : f = 75/ (420+75) ≈ 0,15 soit 15%
8- On dit que X et Y sont indépendant si $n_{ij} = \frac{ni.\times n.j}{n..}$ => n..x $n_{ij}$ = $n_i$. x $n._j$

Vérifions cette égalité : n1.x n.1 = 495x700= 346500 et n..xn11 = 1000x420 = 420000. Ces deux résultats étant différents, X et Y ne sont pas indépendants

## II- REGRESSION ET PREVISION

- La corrélation renseigne sur l'intensité de la liaison entre X et Y. Cette liaison peut être : - nulle : X et Y sont non corrélés - intermédiaire : cas le plus fréquent - fonctionnelle : il existe une fonction f et/ou g telle que : Y = f (X) ; X = g (Y). Pour déterminer cette liaison, on calcul le coefficient de détermination suivant : r² (X ; Y) = $\frac{\mathrm{Cov}^2(X;Y)}{\mathrm{Var}(X)\mathrm{Var}(Y)}$ avec $Cov\ (x;$ y) = $\frac{1}{n}\sum_1^n(xi - \bar{x})(yi - \bar{y})$ = ( $\frac{1}{n}\sum_1^n xiyi$) - $\bar{x}\bar{y}$

r = $\frac{\mathrm{Cov}(X;Y)}{\sigma(X)\sigma(Y)}$ est appelé coefficient de corrélation. - 1 ≤r≤ 1
– Si r = 1 ou -1 alors il existe une liaison fonctionnelle affine entre X et Y.
– Si r> 0,5 alors il ya corrélation entre X et Y
– Si r< 0,5 alors X et Y ne sont pas corrélés

- La régression fournit une expression de la liaison entre X et Y sous forme d'une fonction mathématique. Lorsqu'on cherche à expliquer Y par X, on effectue une régression de Y en X ( si l'on cherche aussi à expliquer X par Y, on effectue une régression de X en Y) c'est-à-dire qu'on cherche à déterminer une fonction f des valeurs $\mathrm{x_i}$ de X dont la représentation graphique soit la plus proche possible du nuage. En général, les courbes de régression représentent des fonctions complexes. On préfère donc choisir un ajustement affine par la méthode des moindres carrés
- L'ajustement affine par la méthode des moindres carrés est approprié dans le cas où le nuage présente une forme allongée et linéaire. On choisit les fonctions affines définies par f(x) = ax + b où a et b sont deux réels à déterminer

**Droite d'ajustement de Y en X :** (D)

L'équation de cette droite s'écrit : y = ax + b avec a = $\frac{Cov(\mathrm{x;y})}{Var(\mathrm{x})}$ et b = $\bar{y}$ - a$\bar{x}$

**Droite d'ajustement de X en Y :** (D')

L'équation de cette s'écrit : x = a'y + b' avec a' = $\frac{Cov(\mathrm{x;y})}{Var(\mathrm{y})}$ et b = $\bar{x}$ – a'$\bar{y}$ ou y = $\frac{1}{a'}$ x – $\frac{b'}{a}$

**Exercice**

Dans une banque, on considère un échantillon de 12 clients choisis au hasard. On note X le nombre de chèques émis et Y le nombre de visites à l'agence, de chaque client durant un trimestre. On obtient

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 34 | 42 | 53 | 30 | 50 | 60 | 46 | 57 | 32 | 24 | 36 | 28 |
| Y | 12 | 14 | 15 | 10 | 15 | 17 | 12 | 14 | 10 | 9 | 11 | 10 |

1- Calculer Cov (X, Y)
2- Chercher la droite d'ajustement de X en Y, puis la droite d'ajustement de Y en X
3- Tracer sur un graphique les points (x, y) et les droites d'ajustement

Correction

1- Calcul de la covariance

| $x_i$ | $y_i$ | $x_iy_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|
| 34 | 12 | 408 | 1156 | 144 |
| 42 | 14 | 588 | 1764 | 196 |
| 53 | 15 | 795 | 2809 | 225 |
| 30 | 10 | 300 | 900 | 100 |
| 50 | 15 | 750 | 2500 | 225 |
| 60 | 17 | 1020 | 3600 | 289 |
| 46 | 12 | 552 | 2116 | 144 |
| 57 | 14 | 798 | 3249 | 196 |
| 32 | 10 | 320 | 1024 | 100 |
| 24 | 9 | 216 | 576 | 81 |
| 36 | 11 | 396 | 1296 | 121 |
| 28 | 10 | 280 | 784 | 100 |
| 492 | 149 | 6423 | 21774 | 1921 |

Pour calculer la covariance, il faut calculer la moyenne de chaque variable. On a :

$\bar{x} = \frac{\sum \mathrm{xi}}{\mathrm{ni}} = \frac{492}{12} = 41$ et $\bar{y} = \frac{\sum \mathrm{yi}}{\mathrm{ni}} = \frac{149}{12} = 12{,}42$

$Cov\ (x; \mathrm{y}) = \frac{1}{n}\sum_1^n (xi - \bar{x})(yi - \bar{y})$

$= \left(\frac{1}{n}\sum_1^n xiyi\right) - \bar{x}\bar{y} = \frac{6423}{12} - 41 \times 12{,}42$ Cov (x ; y) = 26,03

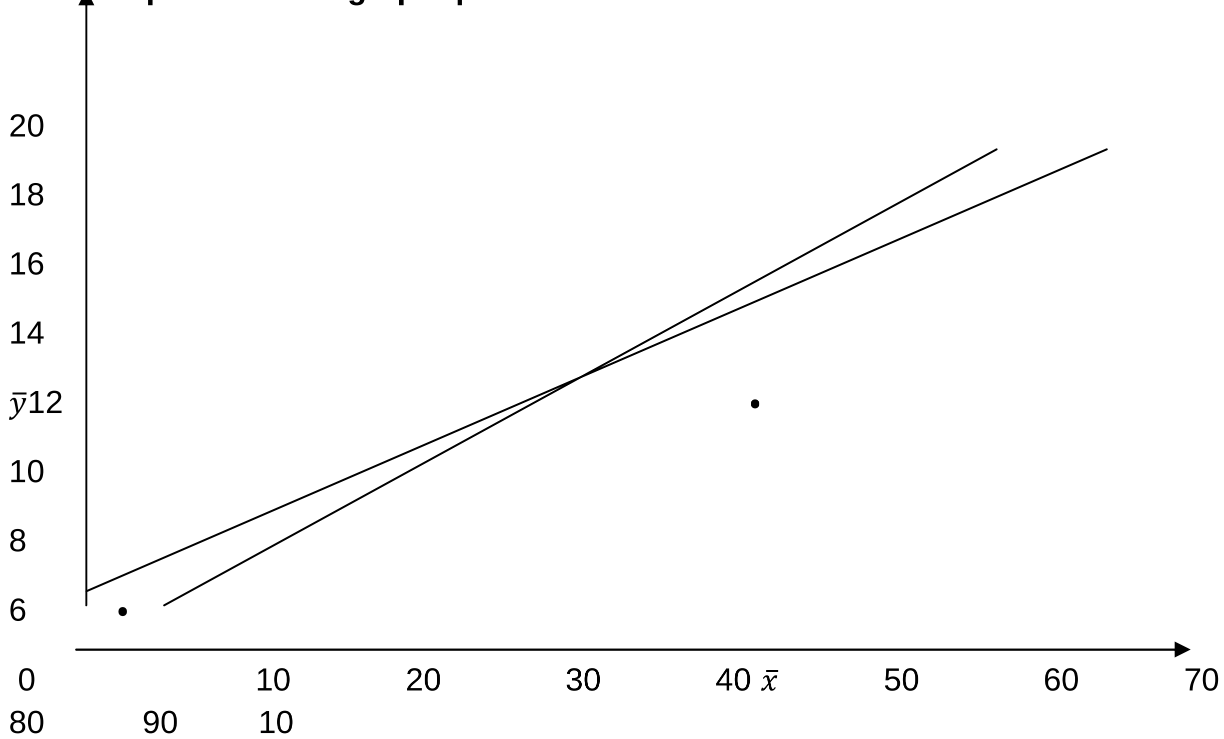
**2- Droite d'ajustement de Y en X**

L'équation de cette droite s'écrit : y - $\bar{y}$ = a(x - $\bar{x}$) avec a = $\frac{Cov(x;y)}{Var(x)}$ = $\frac{26,03}{133,5}$ car Var (x) = $\frac{21774}{12}$ - $41^2$ = 133,5. D'où l'équation de la droite est y = 0,195x + 4,426

**Droite d'ajustement de X en Y**

L'équation de cette droite s'écrit : y - $\bar{y}$ = $\frac{1}{a'}$(x - $\bar{x}$) avec $\frac{1}{a'}$ = $\frac{Var(y)}{Cov(x;y)}$ = $\frac{5,83}{26,03}$ car Var (y) = $\frac{1921}{12}$ – $12,42^2$ = 5,83. D'où l'équation de la droite est y = 0,224x + 3,237

**3- Représentation graphique**



# Chapitre IV. SERIES CHRONOLOGIQUES : PREVISIONS A COURT TREME

On appelle séries chronologiques ou chronique, des séries d'observations échelonnées dans le temps. Ce sont des séries à deux dimensions dont l'une est le temps. Les périodicités sont quelconques. L'année, le trimestre, le mois, le jour sont les périodicités les plus courantes

## I- COMPOSITION D'UNE SERIE CHRONOLOGIQUE

Une série chronologique se décompose en plusieurs composantes qui se combinent :
- la tendance (trend) ou mouvement conjoncturel noté $t_t$
- les variations saisonnières, notées $s_t$
- le cycle, noté $c_t$
- les variations accidentelles ou résiduelles, notées $r_t$

- la tendance ou trend : elle correspond à l'évolution générale du phénomène étudié dans un sens déterminé sur une longue durée. La méthode la plus employée pour estimer la tendance consiste à faire une régression sur les valeurs x. Sa forme est $x_t = t_t + r_t$

- les variations saisonnières : c'est-à-dire un ensemble de variations se produisant annuellement à des périodes bien déterminées sous l'influence de facteurs généralement connus (saisons, vacance, fêtes de fin d'année…). Si la série comporte des effets
saisonniers, la décomposition est alors du type suivant : $x_t = t_t + r_t + s_t$
- le cycle : le mouvement cyclique correspond à des périodes successives d'expansion et de récession, se traduisant par des fluctuations de part et d'autre du trend.

- les variations résiduelles dues à des facteurs exceptionnels, généralement imprévisibles (grève, conflit, etc.) et dont l'influence peut être plus ou moins importante.

## II- ETUDE DU MOUVEMENT SAISONNIERS

Deux structures peuvent être envisagées pour la variation saisonnière :
- la structure additive : $x_t = t_t + r_t + s_t + c_t$
- la structure multiplicative : $x_t = t_t . r_t . s_t . c_t$

**Comment caractériser un mouvement saisonnier ? Est-il additif ou multiplicatif** ?

Plusieurs méthodes permettent de choisir le modèle :
1- Méthode graphique
On relie par une courbe d'une part les maxima de la série, d'autre part, les minima. Si les deux courbes ainsi obtenues sont parallèles, le modèle est de type additif si les deux courbes ne sont pas parallèles, le modèle est de type multiplicatif

2-Méthodes analytiques
– On calcule pour chaque année de la série, la moyenne et l'écart type des observations. Si l'on peut considérer les écarts-types comme constants sur la période, le modèle est de type additif, dans le cas contraire il est de type multiplicatif
– On peut également, à partir des moyennes et des écarts types de chaque année, déterminer par la méthode des moindres carrés, les paramètres de l'ajustement : $\sigma_t = a\bar{x}_t + b$. Si a = 0, le modèle est additif ; dans le cas contraire il est multiplicatif

## III- PREVISION A COURT TERME

- Si la série ( $x_t$) ne comporte pas de composante saisonnière, la prévision peut être faite directement comme au chapitre précédent : $x_t$ = at + b
- Si la série ($x_t$) comporte une composante saisonnière, on l'élimine avec les moyennes mobiles et on obtient une nouvelle série ($y_t$) qui est une approximation de la tendance.
  – dans le cas d'un modèle additif : $y_t = x_t - s_t$.
  – dans le cas d'un modèle multiplicatif, $y_t = \frac{x_t}{s_t}$

**Comment désaisonnaliser une série ?**

Soit ($x_t$) une série de valeurs observées pendant n périodes. On appelle moyennes mobiles sur k périodes, (k < n), la série de valeurs ($y_t$) calculées comme suit :
– si k est impair, alors k = 2p + 1 : $y_t = \frac{x_{t-p} + \cdots + x_{t-1} + x_t + x_{t+1} + \ldots + x_{t+p}}{2p+1}$.
– si k est pair, alors k = 2p : $y_t = \frac{\frac{x_{t-p}}{2} + x_{t-P+1} + \ldots + x_t + \cdots + x_{t+P-1} + \frac{x_{t+p}}{2}}{2p}$

La série $y_t$ obtenue ne comporte plus la composante saisonnière. Cette série comporte n – 2p termes. Une régression, souvent un ajustement affine permet de

faire des estimations qui sont des prévisions ne prenant pas en compte l'effet saisonnier. Pour réintroduire la composante saisonnière, on utilise les coefficients saisonniers. On appelle **coefficient saisonnier** pour chaque période t :
- la valeur $c_t = x_t - y_t$ dans le cas d'un schéma additif
- la valeur $c_t = \frac{x_t}{y_t}$ dans le cas d'un schéma multiplicatif. Pour chacune des k saisons (mois, trimestres…), on obtient une série de coefficients saisonniers, qui devraient être égaux par définition, mais qui en fait diffèrent dans la pratique. On calcule alors un coefficient saisonnier $c_t$ qui est par exemple la moyenne des coefficients (on peut choisir aussi la médiane)
– dans le cas d'un modèle additif : $\sum_1^k c_t = 0$.
– dans le cas d'un modèle multiplicatif : $\sum_1^k c_t = k$

**NB** : on obtient une estimation $\hat{x}$ d'une valeur future de la série, à la saison i, en multipliant l'estimation $\hat{y}$ par le coefficient saisonnier $s_t$

## Chapitre V. TECHNIQUE QUANTITATIVE RELATIVE A L'ETUDE DE MARCHE

Ce sont des études qui permettent d'évaluer le potentiel tout en tenant compte de son évolution. Les études quantitatives consistent à chiffrer les données à partir d'un échantillon représentatif. Elle permet donc de quantifier le volume du produit, de déterminer l'âge, le sexe, les revenus, les lieux d'achat, la fréquence d'achat de l'échantillon interrogé. Les principales études quantitatives sont le recensement, les enquêtes répétitives (les baromètres, les panels) et le sondage

### I- LE RESENCEMENT

Cette méthode consiste à interroger tous les individus d'une population donnée et à dénombrer leurs caractéristiques et opinions. Le recensement est peu utilisé en marketing puisqu'il fait appel à une étude de population importante. Mais par contre, il est intéressant de disposer des chiffres de recensements nationaux fournis par les services publics, cela peut permettre de faire des statistiques ou d'autres applications.

### II LES ENQUËTES REPETITIVES

1- Les baromètres

Ce sont des enquêtes répétitives par lesquelles les échantillons sont renouvelés pour chaque vague d'enquête.

Par exemple : le baromètre Figaro Magazine Sofres est une enquête politique mensuelle comportant une douzaine des questions standards posées dans les mêmes termes à des échantillons successifs et distincts du corps électoral français.

2- Les panels

Ce sont des enquêtes répétitives menées sur un échantillon constant (ou permanent) du public étudié. Ces échantillons sont interrogés à intervalles de temps régulier. On distingue les panels de consommateurs, les panels de distributeurs…..

### III LE SONDAGE

Le sondage est une enquête effectuée sur une population dont on veut connaître les caractéristiques sur un sujet donné en interrogeant qu'un nombre limité de ses membres ou échantillon de la population. Statistiquement, la validité des résultats d'un sondage dépend essentiellement de la représentativité de l'échantillon et de la précision de l'estimation

1- Les méthodes d’enquête

Il existe deux grandes méthodes d’enquête par sondage :

- *les méthodes probabilistes* qui consistent à tirer au sort l’échantillon dans la population à étudier donnant à chaque élément de l’échantillon une probabilité connue, non nulle, d’être tiré ou sélectionné
- *les méthodes empiriques* qui reposent sur un choix raisonné d’individus de la population en respectant des règles concernant, soit les caractéristiques des individus (méthode des quotas) soit les lieux et les moments d’enquête (méthode des itinéraires)

a- Les méthodes probabilistes

- **Le sondage aléatoire simple** : il n’est réalisable que lorsqu’on dispose d’une base de sondage (annuaire téléphonique, liste professionnelle…). A partir de la base de sondage, un tirage au sort est organisé grâce à deux procédures :

- la table des nombres au hasard : il s’agit de liste de nombre dont le tirage a déjà été effectué aléatoirement. Il s’agit de la table de TIPPETT et de celle de KENDALL et BABINGTON SMITH
- le tirage systématique : il consiste à retenir chaque n/N ème individu de la base de sondage ; N étant la taille de la population à étudier et n celle de l’échantillon

**NB** : Lors d’un tirage systématique, les numéros tirés suivent une progression arithmétique de raison r = N/n et de premier terme $U_1$. Son terme général est : $Un = U_1 + (n – 1)r$. Le rapport n/N est appelé taux de sondage.

Exemple : l’on veut tirer 10 individus systématiquement sur un fichier de 220 noms.

Taux de sondage = 10/220 = 1/22. Ce qui signifie que le 1$^{er}$ nom doit être tiré entre la première place et la 22$^{ème}$ comprise. Imaginons que le premier nom tiré est le numéro 3 c'est-à-dire $U_1 = 3$. Les prochains seront :

U2 = 3+22= 25 ; U3= 25+22= 47 ; U4= 47+22 = 69 ; U5= 69+22= 91 ……..

- **le sondage en grappe** : cette méthode d’échantillonnage probabiliste consiste à tirer au sort un certain nombre de groupe d’unités statistiques appelées grappes puis à interroger tous les individus appartenant aux grappes retenues.
  Exemple : un ménage est une grappe de personnes physiques ; une entreprise est une grappe de salariés.

- **Le sondage aréolaire** : cette méthode consiste à découper en zones géographiques un territoire déterminé, à tirer au sort un certain nombre de ces zones et à interroger tous les individus statistiques y résidant
- **Le sondage à plusieurs degrés** : Cette méthode consiste à effectuer à différents niveaux successifs un tirage aléatoire. En effet on tire au sort dans une base de sondage des unités primaires, dans celles-ci, on tire des unités secondaires et ainsi de suite jusqu'au tirage au sort des personnes physiques ou morales à interviewer.
  Exemple : dans une ville, le tirage peut se faire par quartier (10 zones aréolaires), puis par pâté de maisons (5), puis dans les cinq maisons, par appartements (2) puis dans chaque appartement, interroger 2 personnes. L'échantillon sera dans ce cas de :

  n = 10 x 5 x 2 x 2 = 200 individus

- **Le sondage stratifié** : préalablement au sondage, la population à étudier est décomposée en strates homogènes à l'aide des critères liés au thème de l'étude
  Exemple : pour une étude de marché sur les habitudes alimentaires des ménages, on pourra utiliser la taille de la composition de la famille, l'origine ethnique, le niveau de revenu. Ensuite on effectue un sondage un sondage sur chaque strate ainsi crée.

b- Les méthodes non probabilistes ou empiriques

Ces techniques permettent une approximation de la représentativité en assurant une diversité contrôlée de l'échantillon. Elles n'ont donc pas la rigueur des techniques probabilistes, par contre, elles n'exigent pas de base de sondage ni de tirage au sort et sont plus économique pour des résultats satisfaisants.

- La méthode des quotas : elle consiste à reproduire, dans l'échantillon, les distributeurs statistiques de certaines variables, dites de contrôle, choisies pour leur liaison avec les caractéristiques à étudier. Elle repose sur le principe selon lequel un échantillon qui aurait la même répartition que la population mère suivant des critères déjà connus (le sexe, l'âge, la catégorie socio professionnelle…) a de fortes chances d'être représentatif de cette population.
- **La méthode des itinéraires** : cette méthode consiste à imposer aux enquêteurs un itinéraire précis dans une zone géographique donnée et les points d'enquêtes sur le trajet où ils procéderont à une interview.
- **La méthode par convenance** : l'interviewer choisit les répondants sans précaution, pour limiter les efforts et les coûts
- **L'échantillonnage en boule de neige** : elle repose sur la connaissance qu'ont les premiers répondants et leur bonne volonté à donner les noms d'autres personnes à interroger. Ici l'enquêteur,

après avoir choisit un interviewé, il lui demande de lui indiquer d'autres personnes à interroger.

2- Les méthodes de fixation de la taille de l'échantillon

Dans le domaine probabiliste, la taille de la population mère N n'a aucune incidence sur la taille de l'échantillon n. Cette conséquence de la loi des grands nombres n'est contrariée que quand le taux de sondage (n/N) est supérieur ou égale à 1/7.

La détermination de la taille de l'échantillon sera différente selon que nous sommes :

- dans un sondage non exhaustif ($n/N < 1/7$)
- dans un sondage exhaustif ($n/N > 1/7$)

a- Taille de l'échantillon d'un sondage aléatoire non exhaustif

Pour $n \geq 30$ ou $n/N < 1/7$, la taille minimale de l'échantillon peut être calculée par la formule suivante :

$\boxed{n = t^2pq/e^2}$ ou $\boxed{n = t^2\sigma^2/e^2}$ avec :

- t : la variable de la loi normale centrée réduite et dont la valeur dépend du seuil ou niveau de confiance

- e : la marge d'erreur que l'on peut tolérer dans l'estimation

- p : la fréquence observée dans l'échantillon

- $q = 1-p$

Deux cas sont à envisager selon qu'une enquête préalable est effectuée ou non :

**1$^{er}$ cas** : nous avons déjà réalisé une enquête sur un échantillon de au moins égale à 30 pour estimer la valeur de p.

Exemple : le directeur d'une grande surface s'intéresse au pourcentage de clients satisfaits de son magasin. Il a réalisé une pré enquête sur 40 clients et obtenu un taux de satisfaction voisin de 70%. Il veut l'estimer avec une marge d'erreur de 5% et un niveau de confiance de 95%

Déterminer la taille de l'échantillon

……………………………………………………………

Dans ce cas, p = 70%= 0,7 ; q= 1-0,7= 0,3 ; e= 0,05 ; pour un niveau de confiance de 95%, t=1,96

Donc $n = 1{,}96^2 \times 0{,}7 \times 0{,}3/0{,}05^2 = 323$ individus

**2$^{ème}$ cas** : aucune enquête préalable n'est réalisée (cas défavorable à l'entreprise)

Dans la pratique, il est assez rare qu'une enquête préalable soit réalisée. Il faut donc se placer dans une situation ¨à la normande¨du type 50% - 50% c'est-à-dire p = q = 50%

En reprenant l'exemple précédent, on obtient :

$n = 1{,}96^2 \times 0{,}5 \times 0{,}5/0{,}05^2 = 385$ individus

Quelques valeurs de t :

| Seuil ou niveau de confiance | Valeur de t |
|---|---|
| 68% | 1 |
| 90% | 1,65 |
| 95% | 1,96 souvent arrondi à 2 |
| 99% | 2,58 |

b- Taille de l'échantillon d'un sondage aléatoire exhaustif (n/N> 1/7)

Dans ce cas, la taille n' de l'échantillon se calcule par : $\boxed{n' = (n \times N)/(n+N)}$

avec n : taille de l'échantillon comme dans le cas d'un sondage non exhaustif et N la taille de la population mère.

c- La méthode budgétaire

Le budget de l'étude est composé des charges fixes globales (frais de transport, rémunération des encadreurs, autres charges,…..) et des charges variables en fonction d'un questionnaire. Ainsi, la taille de l'échantillon se détermine de la façon suivante :

$$n = \frac{\text{budget global – charges fixes globales}}{\text{Coût variable du questionnaire}}$$

Exemple : une société vous demande de déterminer la taille de l'échantillon d'une enquête sachant qu'elle dispose d'un budget de 1 800 000 FCFA pour la réalisation de l'étude. Ce budget est réparti entre les charges globales représentant 25% et les charges variables par questionnaire s'élevant à 4 000 FCFA

……………………………………………………………………..

- Charges fixes globales : 1 800 000 x 0,25 = 450 000
- n = $\frac{1\ 800\ 000 - 450\ 0000}{4\ 000}$

= 337,5 soit 338 individus

NB : La taille de l'échantillon se confond toujours avec le nombre de questionnaires à administrer

## IV Estimation par intervalle de confiance

De nombreux problèmes sont évalués à partir de jugements sur échantillon en faisant intervenir le calcul de probabilités. Aussi les entreprises, collectivités, ont souvent recours à l'étude d'échantillon pour déterminer les paramètres de la population statistiques étudiée. Au lieu de chercher une valeur approchée, on détermine les limites d'un intervalle à l'intérieur duquel on peut affirmer que la valeur du paramètre inconnu φ est comprise avec un risque d'erreur que l'on s'est fixé. Réaliser une estimation d'un paramètre φ par intervalle de confiance, c'est déterminer deux bornes A et B d'un intervalle tel que P (A<φ<B) = α. L'intervalle [A ;B] est un intervalle de confiance ; les extrémités A et B sont les limites de confiance et α le coefficient de confiance.

### 1- Estimation de la proportion f de la population mère

On démontre que $\boxed{f \in [p - e\ ;\ p + e]}$ avec $e = t.\sqrt{\frac{pq}{n}}$ ou $e = \frac{t\sigma}{\sqrt{n}}$

Exemple : Dans une étude réalisée sur un échantillon de 950 salariés, la proportion de salariés satisfaits de la politique de gestion des ressources humaines est de 60%. Au seuil de 95%, déterminer la proportion de salariés satisfaits de cette politique pour l'ensemble des salariés.

……………………………………………………………………………..

$e = \frac{t\sigma}{\sqrt{n}}$= 1,96√(0,6 x 0,4)/950 = 0,03 soit 3%.

La proportion f ε [60% – 3% ; 60% + 3%] = [57% ; 63%]

Ce qui signifie que la proportion de salariés satisfaits de l'ensemble est comprise entre 57% et 63%

**2- Estimation de la moyenne m de la population mère**

De même que pour la proportion, la moyenne de la population mère que $\boxed{m \in [\bar{x} - e\ ;\ \bar{x} + e]}$

Exemple : A partir d'une enquête auprès de 100 personnes, à la sortir des caisses d'un supermarché, il a été déterminé que le panier moyen des personnes interrogées est de 22 000 f avec un écart type de 9 500 f. Au seuil de confiance de 95%, déterminer l'estimation du panier moyen pour l'ensemble des clients du supermarché.

…………………………………………………………………………………………………….

e $=\frac{t\sigma}{\sqrt{n}}$ avec t = 1,96 ; σ = 9500 f ; n = 100 On en déduit que e = 1,96 x $\frac{9500}{\sqrt{100}}$ = 1862

Conclusion : $m \in$ [22 000-1862 ; 22 000+1862] = [20 138 ; 23 862]

Ce qui signifie que le panier moyen de l'ensemble des clients du super marché est compris entre 20 138 f et 23 862 f

Exercice :

Avant de prendre la décision de lancer un nouveau produit sur le marché, une rumeur vous apprend que la proportion de consommateurs favorables à cette décision est deux fois plus importante que la proportion de celles qui la condamnent.

1- Quelle sera la taille de l'échantillon de sondage au seuil de confiance de 95% afin que les rumeurs se justifient à 2,5% d'erreur près.
2- Dans la situation la plus défavorable, déterminez la taille de l'échantillon à interroger sur votre projet si e = 5% au seuil de confiance de 95%

.Exercice

La direction des ressources humaines d'une entreprise employant 3500 personnes a fait une enquête auprès d'un échantillon de 108 salariés. 35 d'entre eux ont déclaré être des fumeurs réguliers

1- Donnez une estimation ponctuelle du pourcentage p de fumeurs dans cette entreprise.

2- Déterminez et interprétez l'intervalle de confiance du pourcentage de fumeurs dans l'entreprise, avec un seuil de confiance de 95%
3- Des enquêtes réalisées au niveau national font apparaître que 27% des ivoiriens sont des fumeurs réguliers. Est-il possible d'affirmer que les salariés de l'entreprise étudiée fument plus que la moyenne nationale ?
4- Afin de déterminer les raisons qui poussent les salariés à fumer, on réalise une étude quantitative.
    - Définissez une étude quantitative
    - Citez et expliquez les différentes études qualitatives

## Chapitre VI. LA LOI NORMALE

Une variable aléatoire est une grandeur dont les valeurs dépendent du hasard. On dit qu'une variable aléatoire X suit une loi normale ou de Laplace-Gauce, si :

– ses valeurs sont dans IR

– et si sa densité de probabilité est définie par $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ où ses paramètres sont m = moyenne et σ = écart-type. Pour traduire que la variable aléatoire X suit une loi normale de moyenne m et d'écart-type σ, on note : X $\rightarrow$ N (m ; σ)

NB : il n'existe pas de table de la loi N (m ; σ) mais plutôt celle de N (0 ; 1) appelée loi normale centrée réduite dont la variable aléatoire est $t = \frac{X-m}{\sigma}$. Cette variable a pour densité de probabilité $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

Exemple :

Lors d'une étude de marché relative au lancement d'un nouveau produit, la SA KZ a demandé à 80 clients potentiels, le prix qu'ils seraient prêts à payer pour cet article. Elle a obtenu les résultats suivants :

| Prix cité | Nombre de clients |
|---|---|
| de 65 à moins de 75 F | 1 |
| de 75 à moins de 85 F | 3 |
| de 85 à moins de 95 F | 8 |
| de 95 à moins de 105 F | 18 |
| de 105 à moins de 115 F | 20 |
| de 115 à moins de 125 F | 16 |
| de 125 à moins de 135 F | 9 |
| de 135 à moins de 145 F | 4 |
| de 145 à moins de 155 F | 1 |

1- Déterminer :
   - m = prix moyen cité
   - $m_e$ = prix tel que, la moitié des clients ait proposé un prix inférieur
   - $m_o$ = prix cité le plus grand nombre de fois
2- Déterminer σ = écart-type de la distribution (nombre de décimales : 2)
3- Soit : A = m+σ et A'= m – σ. Déterminer, en pourcentage, le nombre d'observations comprises entre A et A'
4- L'entreprise décide de fixer le prix de vente à un niveau tel que seulement 20% des clients aient fixé un prix supérieur. En considérant que les prix cités sont distribués selon une loi normale de paramètres m = 110,37 et σ = 15,68, déterminer le prix de vente retenu

| Prix cité | Nombre de clients ni | Ni ↗ | $x_i$ | $x_i.ni$ | $ni.xi^2$ |
|---|---|---|---|---|---|
| [65 ; 75[ | 1 | 1 | 70 | 70 | 70 |
| [75 ; 85[ | 3 | 4 | 80 | 240 | |
| [85 ; 95[ | 8 | 12 | 90 | 720 | |
| [95 ; 105[ | 18 | 30 | 100 | | |
| [105 ; 115[ | 20 | 50 | 110 | | |
| [115 ; 125[ | 16 | 66 | 120 | | |
| [125 ; 135[ | 9 | 75 | 130 | | |
| [135 ; 145[ | 4 | 79 | 140 | | |
| [145 ; 155[ | 1 | 80 | 150 | | |
| | 80 | | | | |