I.U.T. de REIMS

Département
TECHNIQUES DE COMMERCIALISATION

Finalité Tertiaire

Auteur : Jacky HUSSENET

SOMMAIRE

OBJECTIFS du cours de Statistiques Descriptives

Objectifs de £1 : Généralités :

Manipuler correctement le symbole de sommation
Prendre conscience du besoin de prévision ;
Distinguer les statistiques des probabilités, des sondages, des estimations .

Objectifs de £2 : Présentation

Réviser des méthodes élémentaires mais importantes pour présenter une distribution .
Tableaux .

Pour exploiter une série, vous devrez pouvoir interpréter et juger la situation .

Pour y parvenir vous devez être capable de présenter la série :
définir l'individu, la population, les caractères, les modalités, les effectifs;
présenter un tableau synthétique des éléments nécessaires à la demande ;

Objectifs de £3 : Représentation

Réviser des méthodes élémentaires mais importantes pour représenter une distribution .
Graphiques .

Pour exploiter une série, vous devrez pouvoir interpréter et juger la situation .

Pour y parvenir vous devez être capable, à partir d'un tableau synthétique des éléments nécessaires à la demande de :
définir l'individu, la population, les caractères, les modalités, les effectifs;
tracer des représentations des distributions en respectant les conventions :
en particulier pour l'histogramme, et les grandeurs cumulées
sans oublier la légende ;

Objectifs de £4 : Caractéristiques

Pour analyser les formes possibles d'une distribution et pouvoir en particulier les comparer à des distributions de référence, vous calculerez différents paramètres.

Vous devrez être capable dans le cas de séries discrètes ou continues :

de séparer les paramètres de position ou de tendance, de dispersion, de forme et d'aplatissement ;

d'obtenir, la moyenne arithmétique, géométrique ou harmonique d'une série en fonction du cas étudié ;

de calculer la variance, l'écart-type,

d'obtenir les autres paramètres de la série : mode, médiane, étendue, quartiles, coefficient de variation ;

de calculer, à partir des formules, les coefficients de Pearson et de Yule;

de réagir à un résultat manifestement erroné .

Objectifs de £5 : Ajustement :

Objectifs : Comprendre la notion de corrélation différente de la causalité ;
Connaître le principe de l'ajustement linéaire
Exploiter une courbe d'ajustement comme outil de prévision .

Vous devrez être capable dans le cas de séries discrètes ou continues :

de tracer le nuage de points ;

de calculer et interpréter le coefficient de corrélation;

de déterminer l'équation de la droite d'ajustement dans le cas des méthodes des moindres carrés, d'allométrie .

d'exploiter les résultats pour une estimation éventuelle .

Objectifs de £6 : Indices :

Objectifs : Cas particuliers de variables économiques dans des séries de caractère unique.

Vous devrez être capable :

calculer un indice simple ou synthétique,

enchaîner des indices,

différencier des types d'indices.

Objectifs de £7 : Séries Chronologiques :

Objectifs : Cas particuliers de variables économiques dans des séries doubles dont le temps.

Vous devrez être capable, dans le cadre de l'ajustement
de représenter la série pour apercevoir la tendance ou analyser le type de situation ;
décomposer les éléments constituant la série après avoir procédé à un ajustement.

Objectifs de £8 : Etude de Contingence :

Objectifs : Etude en autonomie de séries doubles pour introduire et préciser le vocabulaire relatif aux tableaux de contingence.

Objectifs de £9 : Etude de séries multiples :

Objectifs : Généraliser les définitions des caractéristiques,
Généraliser les méthodes d'ajustement
à des techniques se ramenant à l'ajustement linéaire,
ou à un ajustement multilinéaire par calcul matriciel .

£1 : GENERALITES

£1A : LE SYMBOLE SIGMA

I DEFINITION:

1) Problème:

Ecrire la somme des 10 premiers nombres entiers; puis celle des 100 premiers;

2) Notation:

$S=\Sigma_1^5 x_i= x_1+x_2+...+x_5$. Définir l'indice, le pas, le premier terme, le dernier.

3) Exercice:

On donne $a_1=1$,$a_2=2$,$a_3=4$,$a_4=8$,et $x_p=2p+1$. Calculer$\Sigma_1^4 a_i.x_i$.

II PROPRIETES:

1) On admettra : $\Sigma_1^n (x_i+y_i)=:\Sigma_1^n x_i+:\Sigma_1^n y_i$; $:\Sigma_1^n(ax_i)=a.\Sigma_1^n x_i$.

2) En déduire : $:\Sigma_1^n a = n.a$; $:\Sigma_1^n(x_i+y_i)^2 =:\Sigma_1^n x_i^2 +2:\Sigma_1^n x_i y_i +:\Sigma_1^n y_i^2$.

III EXERCICES :

1) Simplifier:

simplifier l'écriture de :
$S_1=0+1+2+3+4+5+6+7+8$; $S_2=0+1+4+9+16+25+36$;

$S=1+4+12+32=\Sigma k(2^{(k-1)})$, à calculer par dérivation de :$\Sigma_1^n(1+x)^k$ avec $x=1$.

2) Problème:

Le chiffre d'affaires journalier d'un commerçant est noté f(i) où i est le numéro du jour du mois.
Que représente f(3) ? Exprimer le chiffre d'affaires du mois de décembre.
Exprimer le bénéfice du mois sachant qu'il est de 25% du chiffre d'affaires .
Tenter d'exprimer le chiffre d'affaires de l'année

£1B : INTRODUCTION

I GENERALITES :

1)Objectifs :

La statistique descriptive a pour but de décrire , de synthétiser et de comparer des données recueillies en fonction du lieu ou du temps.
Elle correspond à un besoin constant de connaissances économiques, politiques et maintenant personnelles.
C'est une nécessité de prévision et d'aide à la décision pour l'état, l'entreprise .
Après observation de faits, elle résumera un ensemble de données obtenues par des tableaux, des graphiques, des indicateurs.

2) Définitions:

La statistique est l'ensemble de méthodes techniques et mathématiques qui ont pour but l'étude des résultats d'une expérience réalisée . Elle doit fournir une image de la situation et être une aide à la décision .
C'est l'étude numérique d'un grand nombre de faits.
Dans la Statistique, il faut distinguer :

statistique descriptive, étude d'un événement réalisé,
probabilités, étude d'un événement à réaliser,
les sondages, exploitation des deux techniques précédentes .

Le recensement est uniquement un stockage total .

II METHODE:

C'est le domaine d'application qui justifie le choix de la méthode ;
Il est essentiel de ne pas réduire la statistique à l'application mécanique de formules .

1) La collecte des renseignements. pour recueillir et rassembler les résultats.
Elle est exécutée au moyen d'un questionnaire .

2) Organiser la présentation des résultats dans des tableaux ou des graphiques.

3) L'analyse des données par des manipulations numériques

4) L'interprétation des résultats.

pour étendre à une population les résultats synthétiques obtenus ;
(ne pas oublier l'importance de la date).
(l'interprétation n'est pas d'ordre mathématique).

III VOCABULAIRE:

1) Population : ensemble étudié ;(personnes ou objets).
C'est l'ensemble de référence des observations ;
La taille de la population est son nombre d'éléments, effectif total, noté N .

2) Individu ou unité statistique :
C'est l'élément étudié, élément de la population .

3) L'observation
Elle est exhaustive (totale) ou partielle (sondage ou échantillonnage)

4) Caractère : propriété étudiée ;
il peut être qualitatif (signe , propriété non mesurable)
ou quantitatif (nombre associée à une grandeur) ;
un caractère quantitatif peut être discret (nombre entier)
ou continu (décimal sur un intervalle) .
Les résultats possibles d'un caractère ou valeurs prises par le caractère sont appelées **modalités** ; elles doivent être incompatibles ; elles sont notées par une lettre majuscule .

5) Classe : c'est l'intervalle choisi pour organiser la série à travers un regroupement;
sa largeur est appelée **amplitude**, (a_{i+1}-a_i)
son centre est le **centre de classe** .(a_{i+1}+a_i)/2

6) Effectif : (nommé parfois effectif absolu)
C'est le nombre d'individus ou d'observations qui ont une certaine valeur du caractère ou qui appartiennent à la classe considéré ; souvent noté n_i. $N=\Sigma n_i$ est l'effectif total .
L'effectif cumulé croissant de X en a est la somme des effectifs dont la variable associée est inférieure ou égale à a.

7) Fréquence : (nommé parfois effectif relatif)
Elle est définie comme le rapport de l'effectif à l'effectif total , notée $f_i=n_i/N$; elle correspond à l'idée intuitive de pourcentage mais est exprimée par un nombre à trois décimales ; elle sera prolongée par la notion de probabilité.

8) Série statistique :
C'est l'ensemble des données relevées dans une étude ;
C'est une suite de couples (ou plus généralement de p-uplets). $(x_i;n_i)$ ou $([a_i,a_{i+1}[;n_i)$
Elle est définie à une date donnée .

9) Exemples :
a) On étudie le nombre d'enfants de 1000 femmes ;
l'individu est la femme, la population l'ensemble des femmes, l'effectif total 1000, le caractère le nombre d'enfants ; c'est un caractère quantitatif discret .
b) Cas particulier d'un lancer de dé :
Le résultat peut être considéré comme qualitatif (comme une couleur) ou comme quantitatif en s'intéressant à la valeur .

IV APPLICATION :

Cas particulier :

On relève les résultats d'un lancer répété de dé :

Xi	1	2	3	4	5	6
ni	6	2	2	4	3	3

Considéré comme repère de face, le caractère est quantitatif,
quantité de jeu, le caractère est qualitatif .

1) Exercice 1 : Cas A

Questionnaire :

Numéro du logement		Date :	
Type de logement	foyer ☐	appartement ☐	maison ☐
Nombre de pièces	------------		
Surface habitable	------------		
Nombre de personnes	------------		

Vocabulaire :

préciser et associer le vocabulaire .

2) Exercice 2 : Cas B

En étudiant le nombre de personnes par ménage dans une ville, on a obtenu :

Nb de personnes	1	2	3	4	5	6	7 et +
Nb de ménages	2327	4533	8918	10405	6210	2134	1123

Présenter la série,
Présenter les résultats des effectifs, des fréquences, des fréquences cumulées .

Nb de personnes	1	2	3	4	5	6	7 et +	
Nb de ménages	2327	4533	8918	10405	6210	2134	1123	35650
fi	0,065	0,127	0,250	0,291	0,174	0,059	0,031	
Fi	0,065	0,192	0,442	0,734	0,908	0,968	0,999	

3) Exercice 3 : Cas C

Relever et présenter les séries obtenues en fonction de :ou du
Bac obtenu des étudiants du groupe ,
l'année , de l'âge ;

T.P.

Résultats des recensements en Champagne-Ardenne :

Population	Ardennes	Aube	Marne	Hte-Marne	Total
1982	302338	289300	543627	210670	1345935
1990	295400	289200	555800	203500	1343900

Ardennes en 90	Charl-Méz	Sedan	Revin	Rethel	
	58320	22242	9464	8605	
Aube	Troyes	Chapelle/L	Romilly/S	St-André	Ste-Savine
en 90	60309	15813	15711	11340	9751
Marne	Reims	Chalons/M	Epernay	Vitry-le-F	Tinqueux
en 90	183663	51117	26611	17283	10153
Hte-Marne	St-Dizier	Chaumont	Langres		
en 90	35275	28677	10942		

I 1 ° En utilisant une carte, donner les coordonnées de chaque chef-lieu .

2 ° En considérant la région par ses départements, ramenés en leur chef-lieu,

a) écrire une condition qui caractérise un point moyen ,

b) déterminer ce point moyen géographique .

3 ° En considérant la région par ses départements, ramenés en leur chef-lieu, déterminer ce point moyen pondéré par la population .

II Faire la même étude en ne considérant que les villes de plus de 10000 habitants .

sur l'ensemble de la région .

£2 : PRESENTATION DES RESULTATS construction de tableaux

COURS

I COLLECTE DES FAITS:

1)Remarque:

L'enquêteur doit avoir prévu des questions précises .
Préciser la population, l'individu, le caractère, nécessairement la date.
Le tableau présentera les modalités Xi, les effectifs ni, les fréquences fi, les masses ni.xi .

2)Données brutes:

On désire analyser les résultats de la classe du professeur YAKAFER dont on a relevé les notes : relevé direct des notes ,données brutes;

2	7	10	9	17	9	7	20	12	15	17	15	14
8	16	14	16	13	12	6	14	3	7	9	8	20
Ab	6	8	11	14	8	7	14	4	9	7	7	8
14	20	12	5	15	10	12	15	11	16	14	2	17
2	14	16	2	11	14	14	20	12	4	7	13	10

II MISE EN ORDRE:

1)Données ordonnées:

Pour exploiter ces résultats, on dépouille les résultats, on ordonne les valeurs du caractère, et on compte le nombre de notes de chaque valeur, c'est l'effectif .

notes dénbr	0	1	2	3	4	5	6	7	8	9	10
nb n_i	0	0	4	1	2	1	2	7	5	4	3

notes dénbr	11	12	13	14	15	16	17	18	19	20	:Σ
nb n_i	3	5	2	10	4	4	3	0	0	4	64

L'ensemble des couples, notes-nombres ou valeurs du caractère-effectif, est une série statistique.
$(x_i;n_i)$ ou $([a_i,a_{i+1}[;n_i)$

2)Données regroupées:

Le regroupement est parfois nécessaire. Il facilite la compréhension, mais perd de l'information.
Il est toujours effectué dans le cas de variables continues .
Exemple :
Les professeurs considèrent :

	Yves HALFER	Annie MATION
au niveau E ,	les notes de 0 à 6 ,	de 0 à 4 ;
au niveau D ,	les notes de 7 à 9 ,	de 5 à 7 ;
au niveau C ,	les notes de 10 à 12 ,	de 8 à 11 ;
au niveau B ,	les notes de 13 à 15 ,	de 12 à 14 ;
au niveau A ,	les notes de 16 à 20 ,	de 15 à 20 .

Compléter le tableau et comparer l'effet du regroupement;

Classe	*A*	*B*	*C*	*D*	*E*	*Total*
n_i *H.*	*11*	*16*	*11*	*16*	*10*	*64*
n_i *M.*	*15*	*17*	*15*	*10*	*7*	*64*

3) Récapitulation :

Compléter le tableau suivant :

caractère	0	1	2	3	4			18	19	20	total
effectif											
fréquence											
$\sum n_i$:_											
$\sum n_i \uparrow$											
$\sum f_i \uparrow$											
$\sum n_i x_i$											705
$\sum n_i x_i^2$											9213

On obtient une moyenne de 11,02 et un écart type de 4,75 .

III VOCABULAIRE : Rappels

1)Effectif, effectif cumulé:

n_i est l'effectif du caractère x_i; c'est le nombre d'éléments dont le caractère est x_i .
$N = :\sum n_i$ est l'effectif total .
On appelle effectif cumulé croissant, la somme des effectifs des valeurs du caractère inférieures ou égales à x_i .
On appelle effectif cumulé décroissant, la somme des effectifs des valeurs du caractère supérieures ou égales à x_i .
Combien de notes sont inférieures à 6? sont supérieures ou égales à 15?
Construire le tableau des effectifs cumulés croissants et décroissants.

2)Fréquence:

On appelle fréquence le rapport de l'effectif à l'effectif total ; il est calculé avec 3 chiffres après la virgule, arrondi en fonction du quatrième, mais souvent exprimé en pourcentage .
La fréquence est un nombre de l'intervalle [0;1] . On définit aussi les fréquences cumulées .

IV TABLEAU de CONTINGENCE Cas d'une série double :

Construction :

Salaire net Yj Age Xi	[5;6[	[6;7[	[7;8[	Total
[20;22[	1200	500	100	1800
[22;24[	2500	3500	600	6600
[24;26[	1800	5000	2300	9100
Total	5500	9000	3000	17500

ni,j est l'effectif
n2,1=2500
n.,j est l'effectif marginal de X
n.,2=9000
ni,. est l'effectif marginal de Y
n2,.=6600
n.,.=N=17500

Définitions :

Effectif partiel : ni,j effectif correspondant à la modalité Xi et Yj .
Fréquence partielle : fi,j = ni,j /N
Fréquence marginale : f.,j = n.,j / N
Fréquence conditionnelle : fi/j = ni,j / n.j ; c'est la proportion d'individus de la modalité Xi et Yj par rapport (parmi) la modalité Yj .

V APPLICATIONS :

1) Exercice 1 : Cas C

Relever et présenter les séries obtenues en fonction de :ou du
Bac obtenu des étudiants du groupe ,
l'année , de l'âge ;
l'année et l'âge

2) Exercice 2 : Cas D

Répartition des ouvriers d'une entreprise en fonction du salaire en milliers de francs ;

Classes	de 4,5 à 4,7	de 4,7 à 5,1	de 5,1 à 5,3	de 5,3 à 5,7	de 5,7 à 6,3	Total
effectifs	21	49	100	24	6	200

Préciser les centres de classes, les largeurs de classes .
Présenter les fréquences de répartition de salaires ;
Présenter dans un tableau, le nombre d'ouvriers dont le salaire est inférieur à t .
Présenter dans un tableau, le nombre d'ouvriers dont le salaire est supérieur à t .

3) Exercice 3 Cas E

Traduire les indications du tableau :

Nb d'enfants Revenus	0	1	2	3	>3	Total
de 3000 à 6000	1	2				3
de 6000 à 9000	2	4	3	2	1	12
de 9000 à 12000	1	2	3	1	2	9
de 12 à 15000	2	1	1	1	1	6
Total	6	9	7	4	4	30

Définir et préciser les variables marginales .
Comment peut-on représenter la situation ? cf TCC1S03

T.P.

cf TDSTAT1.XLS

On dispose des renseignements suivant relatifs à trente appartements :

N° i	Nb de Pièces	Surface	Catégorie
1	2	36	C
2	4	81	C
3	3	77	B
4	3	58	C
5	2	55	B
6	4	83	B
7	4	86	B
8	4	71	C
9	4	75	B
10	1	23	B
11	4	88	A
12	3	71	B
13	3	62	C
14	2	44	B
15	2	51	B

N° i	Nb de Pièces	Surface	Catégorie
16	5	138	A
17	4	101	B
18	1	17	B
19	3	53	B
20	2	32	B
21	3	51	C
22	4	80	B
23	1	21	C
24	2	28	A
25	2	39	C
26	3	70	B
27	3	72	B
28	5	117	C
29	4	104	B
30	5	106	B

I Présenter la série statistique .

II 1° Etudier le caractère " catégorie " :

Préciser la nature du caractère
Présenter les tableaux des effectifs, des fréquences de ce caractère .
Faire une représentation graphique adéquate (adaptée) de la série .

2° Etudier de même :

le nombre de pièces .
la surface .

III 1° Dans le cas du nombre de pièces , interpréter les effectifs cumulés ;

Représenter la série .

2 ° Faire la même étude pour la surface .

IV 1° Etudier les paramètres (caractéristiques) de la série de la catégorie .

2 ° Etudier de même le nombre de pièces , puis des surfaces .

N° i	Nb de Pièces	Surface	Catégorie
1	2	36	C
2	4	81	C
3	3	77	B
4	3	58	C
5	2	55	B
6	4	83	B
7	4	86	B
8	4	71	C
9	4	75	B
10	1	23	B
11	4	88	A
12	3	71	B
13	3	62	C
14	2	44	B
15	2	51	B
16	5	138	A
17	4	101	B
18	1	17	B
19	3	53	B
20	2	32	B
21	3	51	C
22	4	80	B
23	1	21	C
24	2	28	A
25	2	39	C
26	3	70	B
27	3	72	B
28	5	117	C
29	4	104	B
30	5	106	B
total	92	1990	
	3,07	66,33	

Catégorie	effectif	fréquence
A	3	0,10
B	18	0,60
C	9	0,30
total	30	1,00

Nb de Pièces	effectif	fréquence	calcul moy
1	3	0,10	3,00
2	7	0,23	14,00
3	8	0,27	24,00
4	9	0,30	36,00
5	3	0,10	15,00
	30	1,00	3,07

£3 : REPRESENTATIONS GRAPHIQUES
Cas d'un caractère

COURS

La forme de représentation d'une série change l'impression de l'image qu'elle en donne donc de son interprétation .

I VARIABLES QUALITATIVES :

Principe :

Modalité en abscisses, effectifs ou fréquences en ordonnées

1) Représentation en bâtons: des effectifs ou des fréquences

2) Barres ou tuyaux d'orgues : barres disjointes

Image significative si le caractère n'est pas ordonné, avec la barre la plus haute proche de l'axe des ordonnées pour lui donner plus d'importance .

3) Figuratif : fonction de la dimension, surface ou volume .

4) Exemple :

a) On a relevé les langues étudiées par des étudiants (LVI)

	GB	D	Autres	Total
Ni	934	351	215	1500
fi				

b) cf Exercice 3 du IV

II VARIABLES QUANTITATIVES DISCRETES :

1) Représentation cartésienne : points dans un repère orthonormé

des effectifs ou des fréquences . Le faire pour la série de notes .

2) Représentation en bâtons : des effectifs ou des fréquences

Avec les fréquences, en joignant les sommets, on obtient la courbe de distribution .

3) Barres ou tuyaux d'orgues : barres jointes

4) Secteurs ou rubans : adaptées à des fréquences .

$\theta i = fi*360$ $li = L*fi$

5) Cas d'une grandeur cumulée : effectif ou fréquence

C'est la courbe représentant la fonction F telle que F(t)=N(X<=t)
On obtient une fonction en escaliers qui, dans le cas particulier de fréquences, représente la **fonction de répartition** .

6) Exemple :

a) On a relevé le nombre d'enfants par famille ;

Xi	0	1	2	3	>3	Total
ni	6	9	7	4	4	30
Σni	6	15	22	26	30	

Représenter le polygone des effectifs,
le polygone des effectifs cumulés croissants ,
la fonction de répartition de la série .

b) cf Exercice 4 du IV Cas B

III VARIABLES QUANTITATIVES CONTINUES:

1) Polygone: des effectifs ou des fréquences:

On joint les points représentatifs placés au centre de chaque tranche.
Le polygone des fréquences corrigées est plus significatif ($f_i^*=f_i/(a_{i+1}-a_i)$) ;
Il représente la **fonction de densité** de la série par sa courbe dite de distribution .

2) Histogramme:

La représentation est faite par des rectangles
dont la surface du pavé est proportionnelle à l'effectif , ou à la fréquence .
La hauteur du pavé est obtenue par la mesure corrigée : $X_i^*=X_i/(a_{i+1}-a_i)$
Le tracer pour les notes regroupées par YAKAFER .
La courbe obtenue en joignant les milieux supérieurs des bandes de l'histogramme des fréquences est appelée **courbe de distribution**

3) Polygone d'une grandeur cumulée: effectif ou fréquence

Une grandeur ,(effectif ou fréquence),est représentée par une courbe polygonale qui représente la fonction de répartition de la série .F(t)=N(X<=t)
Les points sont placés en fin de tranche si elle est cumulée croissante,
en début de tranche si elle est cumulée décroissante.
l'abscisse due leur point d'intersection est **la médiane**, paramètre de la série .
Le polygone des fréquences cumulées croissantes est aussi appelée **courbe de répartition** .

4) Exemple :

a) Cas de tranches de même largeur
En étudiant les revenus par famille, on a en milliers de francs :

Classes	de 3 à 6	de 6 à 9	de 9 à 12	de 12 à 15	Total
effectifs	3	12	9	6	30
Cumul	3	15	24	30	

Représenter la série par un histogramme,
le polygone des effectifs cumulés croissants, décroissants .

b) Cas de tranches de largeurs inégales
On relève le nombre d'exploitations par tranche, en fonction de leur superficie ;

Classes	de 0 à 10	de 10 à 30	de 30 à 70	de 70 à 120	Total
effectifs	16	38	30	16	100
fi	0,16	0,38	0,30	0,16	1,00
fi*	0,16	0,19	7,5	3,2	

Représenter le polygone des fréquences, l'histogramme,
le polygone des fréquences cumulées croissantes, décroissantes, en déduire la médiane de la série .

III AUTRES REPRESENTATIONS:

Figuratifs ;

Polaire pour des séries chronologiques ;

Secteurs pour des fréquences ;

Pyramide des âges .

Semi-Logarithmique

IV APPLICATIONS :

1) Exercice 1 : Cas B

En étudiant le nombre de personnes par ménage dans une ville, on a obtenu :

Nb de personnes	1	2	3	4	5	6	7 et +
Nb de ménages	2327	4533	8918	10405	6210	2134	1123

Présenter la série, les résultats des effectifs, des fréquences, des fréquences cumulées .
Représenter la série

Nb de personnes	1	2	3	4	5	6	7 et +	
Nb de ménages	2327	4533	8918	10405	6210	2134	1123	35630
fi	0,065	0,127	0,250	0,291	0,174	0,059	0,031	
Fi	0,065	0,192	0,442	0,734	0,908	0,968	0,999	

2) Exercice 2 : Cas D

Répartition des ouvriers d'une entreprise en fonction du salaire en milliers de francs ;

Classes	de 4,5 à 4,7	de 4,7 à 5,1	de 5,1 à 5,3	de 5,3 à 5,7	de 5,7 à 6,3	Total
effectifs	21	49	100	24	6	200

Préciser les centres de classes, les largeurs de classes .
Présenter les fréquences de répartition de salaires ;
Présenter dans un tableau, le nombre d'ouvriers dont le salaire est inférieur à t .
Présenter dans un tableau, le nombre d'ouvriers dont le salaire est supérieur à t .
Représenter les résultats .

3) Exercice 3 : Cas C

Relever et *présenter* les séries obtenues en fonction de ou du :
Bac obtenu des étudiants du groupe ,
l'année , de l'âge ;
Représenter les résultats dans chaque situation ..

4) Exercice 4 Cas E

Traduire les indications du tableau :

Nb d'enfants Revenus	0	1	2	3	>3	Total
de 3000 à 6000	1	2				3
de 6000 à 9000	2	4	3	2	1	12
de 9000 à 12000	1	2	3	1	2	9
de 12 à 15000	2	1	1	1	1	6
Total	6	9	7	4	4	30

Définir et préciser les variables marginales .
Comment peut-on représenter la situation ?

Par une etude de covariance :

Moy(X) = 9300 Moy(Y) = 1,7 $\sigma(x)$ =2750 $\sigma(y)$ = 1,3

Cov(X,Y) = 640 r(X,Y) = 0,18

Par le χ^2 = : à vérifier :

Nb d'enfants Revenus	0	1	2	3	>3	Total
de 3000 à 6000	0,6	0,9	0 ,7	0,4	0,4	
de 6000 à 9000	2,4	3,6	2,8	1,6	1,6	
de 9000 à 12000	1,8	2,7	2,1	1,2	1,2	
de 12 à 15000	1,2	1,8	1,4	0,8	0,8	
Total						

ν = 12 α = 12 χ^2_{th} =18,5 χ^2_o = 6,028

T.P.

cf. TD STAT 2 .XLS

On étudie le caractère "diamètre (mm)" dans une population de pièces fabriquées.

Les données ont été regroupées dans le tableau ci - dessous :

Diamètres	de 208 à 210	de 210 à 212	de 212 à 214	de 214 à 216	de 216 à 218	plus de 218
Effectifs	78	199	523	122	48	30

1° Déterminer les fréquences et construire l'histogramme.

2° Construire l'histogramme et construire la courbe de distribution

cas d'une analyse plus fine (diamètres au 1/10 de mm)

Diamètres	2080-2100	2100-2110	2110-2120	2120-2125	2125-2130	2130-2135
Effectifs	78	71	128	93	140	183

Diamètres	2135-2140	2140-2150	2150-2160	2160-2180	2180 et plus	
Effectifs	107	87	35	48	30	

£4 : CARACTERISTIQUES D'UNE SERIE
cas d'un caractère

COURS

I PARAMETRES DE POSITION : ou paramètres de tendance centrale

1) Le mode: ou dominante

Définition : C'est la valeur du caractère dont l'effectif (ou la fréquence) corrigé est maximum ; il correspond à l'abscisse du sommet de la représentation polygonale. :
(Une distribution peut être bimodale : cas du lancer de poids d'une classe mixte).

Dans le cas discret, il est obtenu en la valeur correspondant au sommet du diagramme en bâtons.

Exemple E2 : cas discret de situations multiples :

Xi	10	15	20	25	30	35	40	Total
Ni	2	5	15	8	13	8	3	54

Le mode est de 20

Dans le cas d'une variable continue, série regroupée par classes, c'est le centre de la tranche modale c'est-à-dire la classe ou l'intervalle dont la grandeur corrigée est maximum
Il faut éventuellement considérer l'effectif ou la fréquence corrigé.

Exemple E3 : cas continu : **prime d'intéressement**

Xi	1000-1500	1500-2000	2000-2500	2500-3000	3000-3500	Total
Ni	6	12	25	17	5	65

La tranche modale est 2000-2500 ; le mode peut être 2250.

2) La médiane:

Définition : Valeur du caractère qui partage la population ordonnée en deux sous-ensembles de même effectif . Elle correspond à la valeur du caractère d'indice (N+1)/2 ;
C'est la solution de F(X)=1/2 avec F la fonction de répartition des fréquences ;.
il faut distinguer les cas de variables discrètes et continues .
C'est aussi l'abscisse du point d'intersection :
soit des polygones des effectifs cumulés croissant, décroissant, et de l'ordonnée (n+1)/2 .
soit des polygones des fréquences cumulées croissante, décroissante, et de l'ordonnée 0,5 .
La médiane est indépendante des valeurs extrêmes , mais dépend de leur nombre .

Détermination sur des exemples :

Exemple E1a : Cas discret de situations uniques N impair:

Xi	4	6	8	10	12	13	15	16	20

La médiane est la 5 ième donc 12.

Autre Exemple E1b : cas discret de situations uniques N pair:

Xi	4	6	8	10	11	12	13	15	16	20

La médiane est entre la 5 ième et la 6 ième donc 11,5.

Exemple E2 : cas discret de situations multiples :

Xi	10	15	20	25	30	35	40	Total
Ni	2	5	15	8	13	8	3	54
ΣNi	2	7	22	30	43	51	54	

La médiane est localisée en 27,5 donc parmi les 8 valeurs 25. La médiane est 25..

Cas continu : Classe puis interpolation linéaire ;
Graphiquement avec les courbes de fréquences cumulées .

Exemple E3 : cas continu : **prime d'intéressement**

Xi	1000-1500	1500-2000	2000-2500	2500-3000	3000-3500	Total
Ni	6	12	25	17	5	65
ΣNi	6	18	43	60	65	

La médiane est localisée en 33 donc dans la tranche 2000-2500.
Elle est calculée par interpolation en 2300 .

Exemple E4 **cas discret particulier** :

Xi	3	4	5	6	10	11	11	12	13	13	13	13	16	17	18
Yi	6	7	7	8	9	9	10	12	13	13	13	14	14	15	15

Etudier les paramètres des séries X et Y .
Modes : 13 ; médianes : 12 ; Moyennes : 11 .
Représenter les polygones des effectifs .

3) La moyenne:

Définition :C'est la moyenne arithmétique pondérée des valeurs de la série, notée $\overline{X}$,ou E(X) .
On a : $E(X)=(\sum n_i x_i)/(\sum n_i) = \sum f_i x_i$. $(\sum n_i)$= N effectif total .
On pourra faire la moyenne de la série de notes données, ou celles des notes regroupées.

Détermination sur des exemples

Exemple E2 : cas discret de situations multiples :

Xi	10	15	20	25	30	35	40	Total
Ni	2	5	15	8	13	8	3	54
ΣNi	2	7	22	30	43	51	54	
ΣNiXi	20	75	300	200	390	280	120	1385

La moyenne vaut 1385/54 .

Exemple E3 : Cas continu : **prime d'intéressement**

Xi	1000-1500	1500-2000	2000-2500	2500-3000	3000-3500	Total
Ni	6	12	25	17	5	65
ΣNi	6	18	43	60	65	
Ni*Xi	7500	21000	56250	46750	16250	147750

La moyenne obtenue est de 2273 .

Propriétés :

A la main, il peut être plus facile d'effectuer un changement de variables

: $y=ax+b$ ou $y=(x-x_0)/h$ alors $E(y)=a.E(x)+b$ ou $E(x)=x_0+h.E(y)$

. La somme des écarts à la moyenne est nulle .

La somme des carrés des écarts à la moyenne est minimum .

Elle est sensible aux valeurs extrêmes, à la différence de la médiane.

4) Autres moyennes :

Rappels mathématiques :

puissances, exposants, racines
logarithmes, exponentielle .

Définition :

Moyenne géométrique : $g=[\Pi x_i^{ni}]^{1/N} = \Pi x_i^{fi}$;
Moyenne harmonique : h telle que $1/h=(1/N).\sum n_i/x_i$;
Moyenne quadratique : $q=\sqrt{[(1/N).\sum n_i x_i^2]}$;

On démontre $h<g<E(x)<q$. (cf. thème sur les moyennes) .

Exemples

Remarque : Si f est une fonction monotone, la valeur moyenne de la série (xi,n_i) est le nombre M défini par $f(M)=(1/N)(\Sigma n_i\ f(xi))$; (avec f=Id , f= Ln , f=1/Id , $f=Id^2$)

5) Remarques :

Dans le cas d'une distribution parfaitement symétrique, mode, moyenne et médiane sont confondus ;
Sinon, la médiane est comprise entre le mode et la moyenne, toujours plus proche de la moyenne
Relation de Pearson : Me ≈ (Mode+2*Moyenne)/3 .

6) Quantiles :

Généralisation de la médiane,

les quartiles Q1, Q2, Q3 sont définis par les solutions de F(X) = 1/4 , 1/2 , 3/4
les déciles Dk sont définis par les solutions de F(X) = k/10

7) Médiale : (paramètre de concentration)

Médiale : Dans le cas d'une série groupée par classes, $[(a_i,a_{i+1}); n_i]$,

la médiale est la médiane de la série $[(a_i,a_{i+1}); n_i x_i]$
Elle partage la série en deux sous-ensembles de même masse ;

II PARAMETRES DE DISPERSION:

Remarque :

Exemple 2 : cas discret très particulier
:

Xi	3	4	5	6	10	11	11	12	13	13	13	13	16	17	18
Yi	6	7	7	8	9	9	10	12	13	13	13	14	14	15	15

Etudier les paramètres des séries X et Y .
Modes : 13 ; médianes : 12 ; Moyennes : 11 .
Représenter les polygones des effectifs .

1)L'étendue : ou intervalle de variation

C'est la différence des valeurs extrêmes du caractère . (caractéristique très imparfaite)
Elle peut être calculée à partir des valeurs extrêmes ou des centres de tranches extrêmes .

2)L'écart moyen : ou absolu moyen

C'est la moyenne arithmétique des écarts à la moyenne. (ou à la médiane)
$\mathbf{e=(\sum n_i|(x_i-\overline{X})|)/N}$. Difficile à calculer, il est peu utilisé .

3)La variance :

a) Définition :

C'est la moyenne arithmétique des carrés des écarts à la moyenne.
formules théoriques : $\mathbf{V(x)=(\sum n_i(x_i-X)^2)/N=\sum f_i(x_i-X)^2}$. (c'est (xi-X) qui est au carré
formules pratiques $\mathbf{V(x)=(\sum n_i x_i^2)/N-X^2=\sum f_i x_i^2-X^2}$. (conséquence de l'identité)

b) Propriétés :

Dans le cas d'un changement de variables
$\mathbf{Y=aX+b}$ **ou** $\mathbf{Y=(X-x_0)/h}$ **, on a** $\mathbf{V(Y)=a^2.V(X)}$ **,** $\mathbf{V(X)=h^2.V(Y)}$ **.**

Dans le cas d'un regroupement des données par classes,
la variance est modifiée par le regroupement ; elle peut-être corrigée par la relation :
$Vdc = Vdg - (h^2)/12$ avec h l'intervalle de classe, (correction de Sheppard)
Vdc variance des données corrigées, Vdg variance des données groupées.

Généralisation aux moments centrés d'ordre k **:**
$\mathbf{\mu_k=(1/N)\Sigma n_i(x_i-X)^k}$. $\mathbf{\mu_1=O}$, $\mathbf{\mu_2=V(X)}$,
Avec les moments d'ordre k , $\mathbf{X=m1, V(X)=m2-m1^2}$ **, et** $\mathbf{mk=(1/N)\Sigma n_i x_i^k}$

Dans le cas d'une variable associée à un caractère issu de deux populations distinctes,
Si E=E1UE2, N=N1+N2 et Xi et Vi(X) les paramètres dans Ei,
X=(1/N)*[N1.X1+N2.X2]
et $\mathbf{V(X)=(1/N)(N1.V1+N2.V2)+(1/N)[N1.(X1-X)^2+N2.(X2-X)^2]}$
La moyenne est la moyenne pondérée des moyennes,
La variance globale est la moyenne pondérée des variances augmentée de la variance pondérée des moyennes. **Vglobale = Vintra établissement + Vinter établissement** .

4)Ecart type :

Variable d'écart , c'est la racine carrée de la variance. **$\sigma=\sqrt{V(x)}$.**

Dans le cas d'un changement de variables Y=aX+b ou $Y=(X-x_0)/h$,
on a $\sigma(Y)=|a|.\sigma(X)$, $\sigma(X)=|h|.\sigma(Y)$.

L'écart type est toujours supérieur à l'écart moyen .

5)Remarques :

a) Un élève a une moyenne de 12 avec un écart type de 4 ; un autre a une moyenne de 12 avec un écart type de 2. Le second peut être plus confiant que le premier.

b) L'intervalle $[E(x)\pm2\sigma]$ contient toujours plus de 75% des valeurs (Tchebitcheff) .

c) L'expérience montre que pour une distribution uni modale symétrique régulière (Normale)
l'intervalle $[E(x)\pm\sigma]$ contient environ 68% des valeurs ,
l'intervalle $[E(x)\pm2\sigma]$ contient environ 95% des valeurs.

d) Preuve des trois écarts-types :
Si la distribution est sensiblement Normale et les paramètres calculés correctement, presque toutes les valeurs du caractère sont dans l'intervalle $[m-3\sigma ; m+3\sigma]$

6) Autres :

Ecart inter quartile : **Q3-Q1**

Rapport des déciles : **D9 / D1**

Coefficient de variation : **Cv= σx / X** souvent exprimé en pourcentage .

7) Exemples :

Classe	Ni	Xi	Ni*Xi	|Xi-X|	Ni*|Xi-X|	Ni*|Xi-X|^2	Ni*Xi^2
[3500;3700[	21	3600	75600	523,5	10993,5	5755097,3	
[3700;4100[	49	3900	191100	223,5	10951,5	2447660,3	
[4100;4300[	100	4200	420000	76,5	7650	585225	
[4300;4700[	24	4500	108000	376,5	9036	3402054	
[4700;5300[	6	5000	30000	876,5	5259	4609513,4	
Total	200		824700		43890	16799550,1	

Moyenne de X = 4123,5 ; Mode = 4200 ; Médiane = 200*(30,5)/100 + 4100 ;
Ecart absolu moyen = 219 ; Variance = 83997,75 ; Ecart-type de 289,8 ;

T.P.

EXERCICE 1 : TDXL5 -3 :

Soit la distribution
:

Classes	<260	260-280	280-300	300-320	320-340	340-360	360-380
Effectifs	21	122	309	321	156	55	16

1° Présenter la série dans un tableau qui précisera :
les centres de tranches, les fréquences, les fréquences cumulées.

2° Construire l'histogramme,
la courbe de distribution,
la courbe de répartition .

3° Calculer les paramètres de la série :

J'ai obtenu une moyenne de 304 et une médiane de 303 .

T.P.A

EXERCICE 1 : TDXL5 -4 :

Considérons la distribution des revenus imposables annuels en France en 1975.

Classes en KF	moins de 10	de 10 à 20	de 20 à 30	de 30 à 40	de 40 à 50
Effectifs	479000	4939000	3451000	1809000	854000

Classes en KF	de 50 à 100	de 100 à 200	de 200 à 400	plus de 400
Effectifs	975000	204000	44000	10000

Faire une étude complète de la distribution :

Calculs des fréquences (bâtons, histogramme, cumul)
Représentation des fonctions de distribution, de répartition ;
Calculer les paramètres (ou caractéristiques) de la distribution .

Travaux Dirigés de STATISTIQUES

EXERCICE 1 : TDXL5 -5 :

Considérons la distribution .

Classes	de 150 à 200	de 200 à 250	de 250 à 300	de 300 à 400	de 400 à 600	de 600à1000
Effectifs	131	548	497	211	43	15

Faire une étude complète de la distribution :

Représentation des fonctions de distribution, de répartition ;

Calculer les paramètres (ou caractéristiques) de la distribution .

Calcul éventuel de l'écart moyen

EXERCICE 1 : TDXL5 -6 :

On a relevé les poids d'une population fictive.
Les résultats sont donnés dans le tableau ci-dessous :

Classes	de 50 à 55	de 55 à 60	de 60 à 65	de 65 à 70
Effectifs	5	18	48	80

Classes	de 70 à 75	de 75 à 80	de 80 à 85	de 85 à 90
Effectifs	52	24	10	3

Faire une étude complète de la distribution :

Représentation des fonctions de distribution, de répartition ;

Calculer les paramètres (ou caractéristiques) de la distribution .

Calcul éventuel de l'écart moyen

£4 bis: CARACTERISTIQUES D'UNE SERIE cas d'un caractère

COURS

III CARACTERISTIQUES de FORME :

Préambule :

Rappels sur les moments :

$\mu_k=(1/N)\Sigma n_i(x_i-X)^k$. **$\mu_1=O$, $\mu_2= V(X)$,**

Avec les moments d'ordre k , $X=m1$, $V(X)=m2-m1^2$, et $(1/N)\Sigma n_i x_i^k$

1) Définition :

Dans le cas d'une distribution parfaitement symétrique, mode, moyenne et médiane sont confondus ;
Sinon, la médiane est comprise entre le mode et la moyenne, toujours plus proche de la moyenne

Relation de Pearson : Me≈(Mode+2*Moyenne)/3 .

La distribution est oblique à gauche si (Mode < Médiane < Moyenne) ou étalée à droite

(cas de lancers de poids dont l'effectif diminue si la performance approche un record).

La distribution est aplatie par comparaison à la courbe de Gauss .

2) Dissymétrie :

Coefficients empiriques :

Coefficient de YULE : **$s = [(Q3-Q2)-(Q2-Q1)]/(Q3-Q1)$**

Si le coefficient est nul, la distribution est symétrique,
s'il est positif, la distribution est oblique à gauche ; (Mode < Médiane < Moyenne)

Coefficient de PEARSON : **$s = (\text{Moyenne}- \text{Mode}) / \sigma$**

s'il est positif, la distribution est oblique à gauche ; (Mode < Médiane < Moyenne)

Coefficients calculés :

Coefficient de PEARSON

$$\beta_1= \mu^2_3 / \mu^3_2$$

Coefficient de FISHER : $\gamma_1= \mu_3/\sigma^3$ avec $\mu_3=(1/N)\Sigma n_i(x_i-X)^3$ donc $\gamma_1=\sqrt{\beta_1}$.

Si la distribution est symétrique, le coefficient est nul,
Si elle est étalée à droite, (oblique à gauche) le coefficient est positif .

3) Aplatissement :

Coefficient de PEARSON : $\boldsymbol{\beta_2 = \mu_4 / \mu^2_2}$

Plus il est proche de 1, plus la courbe est plate . Elle est normale si $\beta_2=3$.

Coefficient de FISHER : $\boldsymbol{\gamma_2 = \mu_4/\sigma^4 - 3}$ avec $\mu_4=(1/N)\Sigma n_i(x_i-X)^4$ et $\gamma_2=\beta_2-3$.

La référence est la courbe de Gauss,

si γ_2 est positif, la distribution est moins aplatie que la distribution Normale de mêmes moyenne et écart-type

Le coefficient est toujours supérieur à -2 .

IV CARACTERISTIQUES de CONCENTRATION :

1) Par le calcul :

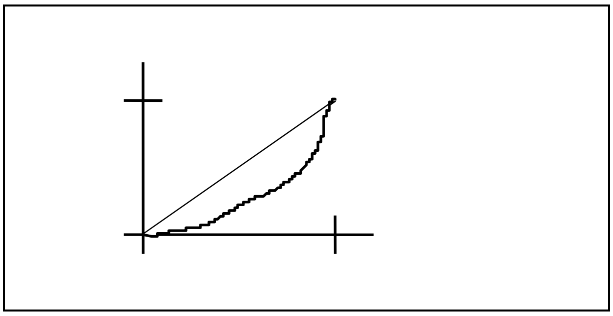
Médiale : Dans le cas d'une série groupée par classes, $[(a_i,a_{i+1}); n_i]$,
la médiale est la médiane de la série $[(a_i,a_{i+1}); n_ix_i]$
Elle partage la série en deux sous-ensembles de même masse ;

Indice de concentration :
C'est le rapport (Médiale-Médiane)/Etendue

2) par le graphique :

Courbe de LORENTZ : ou courbe de concentration :
La courbe de Lorentz est obtenue en plaçant les points de coordonnées (Σf_i ; ΣF_i) .
Elle permet de comparer la distribution à une distribution uniforme (égalitaire)
La courbe polygonale est construite en associant les fréquences cumulées croissantes de la série aux fréquences cumulées croissantes des masses
Plus la courbe est proche de la bissectrice, moins la concentration est forte .

Indice de GINI :
C'est le rapport entre l'aire de concentration et l'aire du demi-carré ,
c'est aussi le double de l'aire de concentration ; ceci étant l'aire du domaine compris entre la diagonale et la courbe de concentration ;
Compris entre 0 et 1, plus il est grand , plus la concentration est forte .



4) Exemple :

Classe	n_i	x_i	n_ix_i	f_i	Σf_i	F_i	ΣF_i
[3500;3700[	21	3600	75600	0,105		0,0917	0,0917
[3700;4100[	49	3900	191100	0,245	0,350	0,2317	0,3234
[4100;4300[	100	4200	420000	0,500	0,850	0,5093	0,8327
[4300;4700[	24	4500	108000	0,120	0,970	0,1309	0,9636
[4700;5300[	6	5000	30000	0,030	1,000	0,0364	1,0000
Total	200		824700	1,000		1,0000	

Rappel du II 7)
Moyenne de X = 4123,5 ; Mode = 4200 ; Médiane = 200*(100,5-70)/100 + 4100 = ;
Ecart absolu moyen = 219 ; Variance = 83997,75 ; Ecart-type de 289,8 ;
Médiale : = 200*(50-32,34)/50,93 + 4100 = 4169,43
donc la moitié de la masse salariale de l'entreprise est versée aux salariés dont le salaire est inférieur à 4170 F .

EX : N4 **Formule de Paréto** : On suppose que le nombre Y de personnes ayant un revenu inférieur à R francs est donné par la relation :

$$Y = N * \left(1 - \left(\frac{Ro}{R}\right)^{3/2}\right)$$

N représente l'effectif de la population active : 20 000 000 ;
Ro représente le revenu correspondant au SMIG : 5000 F (R ≥ Ro)
Déterminer et interpréter la limite de Y quand R devient infini (arbitrairement grand).
Etudier et représenter graphiquement la fonction f telle que Y=f(R)) .
Etudier la concentration .(Pb HEC)
Remarque : Il est rappelé que : $A^x = .e^{x.\ln A}$

Situations d'Exemples

Exemple 1 : cas discret :

Xi	10	15	20	25	30	35	40	Total
Ni	2	5	15	8	13	8	3	54
ΣNi	2	7	22	30	43	51	54	
ΣNiXi	20	75	300	200	390	280	120	1385
ΣNiXi^2								

Exemple 2 : cas discret particulier :

Xi	3	4	5	6	10	11	11	12	13	13	13	13	16	17	18
Yi	6	7	7	8	9	9	10	12	13	13	13	14	14	15	15

Etudier les paramètres des séries X et Y .
Modes : 13 ; médianes : 12 ; Moyennes : 11 .
Représenter les polygones des effectifs .

Exemple 3 : cas continu : prime d'intéressement

Xi	1000-1500	1500-2000	2000-2500	2500-3000	3000-3500	Total
Ni	6	12	25	17	5	65
ΣNi	6	18	43	60	65	
Ni*Xi	7500	21000	56250	46750	16250	147750

Travaux Dirigés de
STATISTIQUES

EXERCICE 1 : TDXL5 -7 :

Considérons la distribution des salaires dans une entreprise en 1975.

Classes en MF	moins de 4	de 4 à 4,5	de 4,5 à 5	de 5 à 5,5
Effectifs	13	228	421	312

Classes en KF	de 5,5 à 6	de 6 à 8	de 8 à 12	de 12 à 20
Effectifs	105	80	22	3

Faire une étude complète de la distribution :
On pourra étudier la dissymétrie
éventuellement la courbe de Lorentz .

£4 ter : CARACTERISTIQUES D'UNE SERIE cas d'un caractère

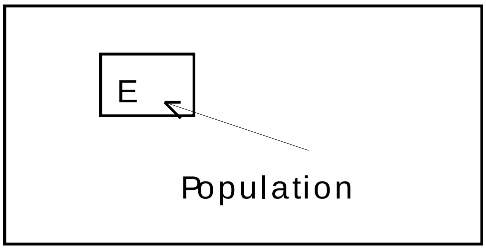
COURS

V NORMALITE D’UNE DISTRIBUTION :

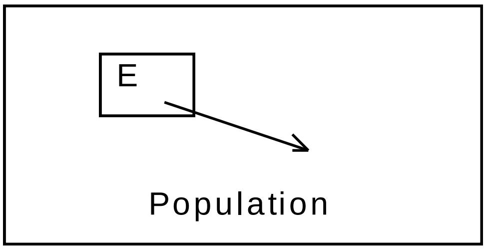
Problème :

L'échantillonnage : permet de passer de la population connue à l'échantillon ;
Il permet l'étude des liaisons entre une population et divers échantillons tirés dans cette population mais aussi de savoir si des différences observées entre deux échantillons sont significatives ou dues au hasard.

L'estimation : permet de passer de l'échantillon observé à une évaluation des paramètres de la population à étudier



Echantillonnage



Estimation

La méthode statistique comporte des phases d'observation, de description, de décision .

L’échantillonnage permet :
l’étude des liaisons entre la population et les échantillons tirés de cette population.
de savoir si les différences observées entre des échantillons sont significatives ou dues au hasard.
Les choix de tirages peuvent être :
exhaustifs (tous) , le résultat est une statistique avec ou sans remise ;
non exhaustifs (partiel), avec ou sans remise : c’est un échantillon.
Si la population est hétérogène, l’échantillonnage peut-être dirigé par la connaissance préalable de caractéristiques déjà connues (strates ou quotas) .

Rappel :

La somme des écarts à la moyenne est nulle : $\mathbf{0=(\sum n_i(x_i-X))/N}$.

La variance est la moyenne arithmétique des carrés des écarts à la moyenne.

formules théoriques : $\mathbf{V(x)=(\sum n_i(x_i-X)^2)/N =\sum f_i(x_i-X)^2}$.

formules pratiques $\mathbf{V(x)=(\sum n_i x_i^2)/N -X^2 =\sum f_i x_i^2-X^2}$. (conséquence de l'identité)

La variance est connue dans l'échantillon, on voudrait celle de la population.

Elle est estimée par la correction de Bessel : **Vp= Ve*N/(N-1)**

Preuve des trois ecarts-types :

Si la distribution est sensiblement Normale, si les caractéristiques sont correctement calculés, toutes les valeurs de la distribution sont sensiblement comprises dans l'intervalle de confiance de trois sigma centré sur la moyenne.

Vérification de la Normalité d'une distribution :

Méthode :

On compare la distribution observée à la distribution théorique Normale.
68 % de la population doit se situer dans l'intervalle de confiance à 1 sigma ;

	Nb dans Ik	Nb à l'extérieur	Total
Observé	140	60	200
Théorique	136	64	200

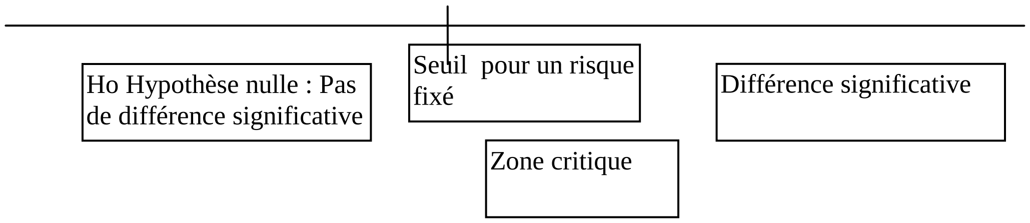
Khi 2 Observé : (140-136)² / 136 + (60-64)² / 64 = 0,36 .

Nombre de degrés de liberté : 1 car un seul choix possible ou (Nl-1)*(Nc-1) .

Au seuil de risque de 5%, le Khi 2 théorique lu dans la table est de 3,84 .

Il semble qu'il n'y ait pas de différence significative entre la distribution observée et une distribution Normale.

Interprétation :



Cas particulier :

Pour chaque nombre théorique inférieur à 10, on effectue la correction de Yates : on retranche 0,5 à l'écart observé.

Khi 2 Observé = (14-13,6)² / 136 + ((6-6,4)-0,5)² / 6,4 = 12,11 .

La distribution n'est pas Normale (au seuil de 5%)

De l'usage des statistiques :

Dans une ville nouvelle, on prévoit la construction de 20 000 logements avec un parking pour tous. On envisage de proposer un garage complémentaire au libre choix du propriétaire.
Une enquête a été effectuée auprès de 900 familles acheteur potentiel en ce lieu pour savoir si elles souhaiteraient acheter un garage supplémentaire pour leur logement.
630 réponses ont été affirmatives.

1) Déterminer, avec un coefficient de confiance de 0,95 , l'intervalle de confiance qui contient le nombre de garages à prévoir.

2) Combien aurait-il fallu interroger de familles pour que la précision de l'estimation de la fréquence de réponses favorables soit inférieure à 0,02 ?

3) a Déterminer le nombre de garages à construire pour satisfaire au moins 98% du maximum des demandes prévues dans les conditions du 1°).

b Déterminer le nombre de garages à construire pour satisfaire au moins 95% des demandes effectives toujours dans les mêmes conditions du 1°).

c Déterminer le nombre de garages à construire pour satisfaire au moins 95% du nombre maximum des demandes prévisibles encore dans les mêmes conditions du 1°).

Indication de solution :

1)
Fréquence de réponses favorables dans l'échantillon : 0,7
Fréquence estimée dans la population : 0,7
Variance de cette va , fréquence, dans la population : 0,7*0,3/900, soit 0,00023333
Ecart-type associé : 0,015.
Coefficient k, associé au coefficient de confiance de 0,95 : 1,96 soit 2.
D'où N, le nombre de garages à prévoir : N = 0,7*2 000
Ou plus précisément (0,7 – 2* 0,015)*20 000 < N < (0,7 + 2* 0,015)*20 000
Soit : 13 400 < N < 14 600 .

2)
Il faut que $k*\sigma$ soit < 0,02 , donc $2* \sqrt{(0,7*0,3/N)} < 0,02$; Il faut que N > 2100 .

3a) Pour satisfaire 98% des 14 600 demandes, il faut prévoir 14 310 garages.

3b) Avec un coefficient de confiance de 0,95 unilatéral, k vaut 1,645 ;
Il faut prévoir : (0,7 + 1,645*0,015)*20 000 soit 14 494 .

3c) Le nombre maximum prévisible est (0,7 + 3*0,015)*20 000 soit 14 900 ;
Pour en prévoir 95% , il faut en construire 14 155 .

£5 : ETUDE CLASSIQUE D'UNE SERIE

COURS

£5A : **Problèmes d'ajustement**
cas d'un caractère unique

I ETUDE des PARAMETRES :

Rappels : cf. cours
Exemples : cf. TP

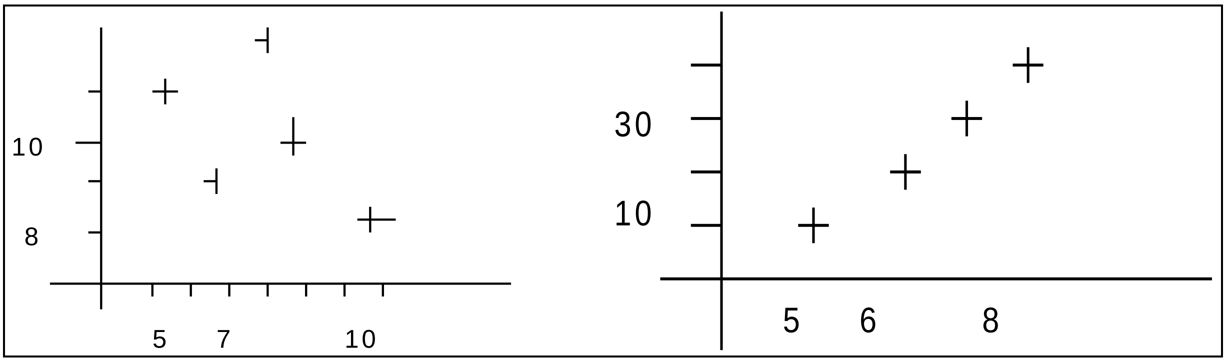
II TEST d'AJUSTEMENT d'une Série à une Loi :

1) Graphiquement : par comparaison des fonctions de densité

Exemple : On a relevé les rémunérations mensuelles de 50 salariés d'une entreprise.

						Total
Salaire en milliers	X<6	6<X<7	7<X<8	8<X<9	X>9	
Nombre	11	9	12	10	8	50
Cumul	11	20	32	42	50	

Représenter la fonction de densité puis la fonction de répartition , observer ;.



2) test du $\aleph^2$

Méthode :
Dans un tableau, on présente les effectifs de chaque classe, les effectifs théoriques,
on y calcule une distance $\chi^2 = \Sigma(n_{io}-n_{it})^2/n_{it}$;

On détermine **le nombre de degrés de liberté** qui précisera la position dans la table en fonction du coefficient de confiance choisi.

Le nombre de degrés de liberté est égal au nombre de classes moins le nombre de paramètres connus.

La valeur lue permet de justifier ou non l'ajustement (ou approximation).

Exemple :
On a relevé les rémunérations mensuelles de 50 salariés d'une entreprise.

Salaire en milliers	X<6	6<X<7	7<X<8	8<X<9	X>9	Total
Nombre	11	9	12	10	8	50

Les rémunérations sont-elles réparties uniformément au risque de 10% ?
Si oui, chaque classe aurait 10 salariés.

Salaire en milliers	X<6	6<X<7	7<X<8	8<X<9	X>9	Total
Nombre observé	11	9	12	10	8	50
Nombre théorique	10	10	10	10	10	
(nio-nit)^2	1	1	4	0	4	
(nio-nit)^2 / nit	0,1	0,1	0,4	0	0,4	1,0

Il y a 5 classes, l'effectif total connu, donc $\nu=4$ degrés de liberté.
Dans la table, avec $\alpha=0,10$ et $\nu=4$, $\chi^2_{th}=7,779$;or $\chi^2_c= 1 < \chi^2_{th}$ donc l'ajustement est accepté;

3) Droite de Henry : pour un ajustement à une loi Normale :

4) Applications :

cf. Math tome 2 Techniplus p 352

£5B : **Problèmes d'ajustement**
cas de plusieurs caractères
Corrélation Ajustement

I CAS d'une série de DEUX VARIABLES : NOTION de CORRELATION :

1) Situation :

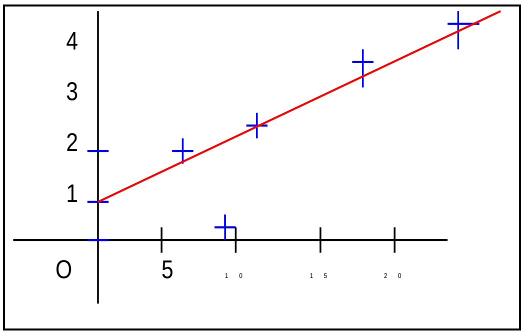
But : Une étude statistique étudie un couple de variables dans une enquête de publicité :
X désigne le coût variable d'une campagne publicitaire, Y les quantités vendues ;

i	1	2	3	4	5	6	7	8	9	10
X	13,2	0	18,6	8,8	12,3	28,3	0	23,5	5,6	0
Y	2	1	3,5	0,5	2,5	4	0	5	2	2

Chaque situation est unique, la série est de la forme : [(xi,yi);1]
La question est de savoir s'il existe un lien de cause à effet entre X et Y
Il n'est pas possible de répondre au problème, mais il est possible de chercher des liens théoriques à travers des formules mathématiques ; c'est la corrélation .

Représentation :

Construire le nuage de points associé à la série en plaçant dans un repère orthonormé les points Mi(xi,yi)
Le dessin suggère une relation affine entre X et Y : Y=aX+b ;
il reste à déterminer a et b .
Déterminer le(s) point(s) moyen(s)



Le lien mathématique traduit une notion de dépendance entre les variables appelée corrélation.
Cette dépendance est théorique et distincte de la notion de causalité .
Elle permettra d'estimer une valeur de y pour une valeur donnée de x .
cf. calculs p6

2) Covariance :

Définition : $\mathbf{Cov(X,Y) = (1/N)\Sigma n_i(x_i\text{-}X)(y_i\text{-}Y)}$ formule théorique

$\mathbf{Cov(X,Y) = (1/N)(\Sigma n_i x_i\ y_i)\text{-}XY}$ formule pratique

C'est la moyenne pondérée des produits moins le produit des moyennes .

Remarque : Cov(X,X) = V(X)

3) Coefficient de corrélation linéaire :

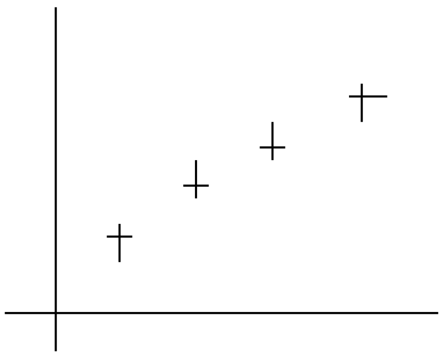
Définition : $\mathbf{r = [Cov(X,Y)] / (\sigma_X \sigma_Y)}$ On a toujours $-1 \leq r \leq 1$.

Si $0,8 \leq |r| \leq 1$, la corrélation est forte et l'ajustement peut-être justifié donc envisagé .

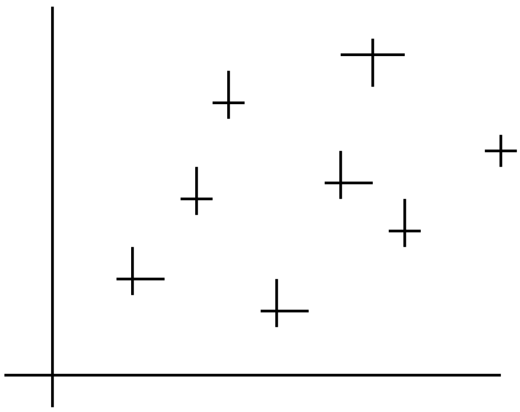
Il traduit une forte concentration des points au voisinage d'une droite ou une courbe , pas de notion de cause à effet ; par exemple, le nombre vente de postes de télévision et le nombre d'entrée en hospitalisation de 1955 à 1995 peuvent être fortement corrélées mais éventuellement liés à un phénomène de société.

Représentations :

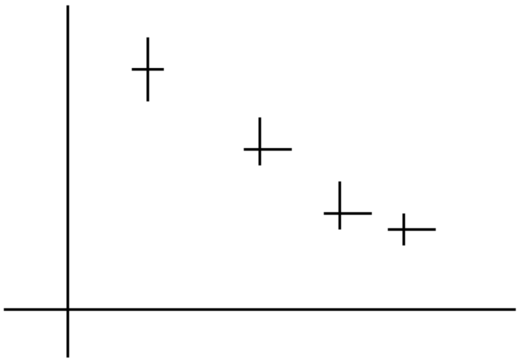
Cas de r positif voisin de 1



Cas de r positif mais faible



Cas de r voisin de -1



4) Exemple :

Cas a) On a relevé les réponses de personnes interrogées par deux questions,
Etes-vous une femme ?, Etes-vous veuve ?
On attribue la valeur 1 au caractère X si la personne est une femme,
1 au caractère Y si elle est veuve
Les résultats sont :

	H : X=0	F : X=1	Total
NV : Y=0	5	3	8
V : Y=1	0	2	2
Total	5	5	10

Calculer la covariance Cov(X,Y)

Moyenne de X = 5/10 Moyenne de Y = 2/10 Cov(X,Y) = 2/10-0,1

Cas b) On étudie la consommation de tabac sur une population de 200 personnes .

Tabac Sexe	ne fument pas	de 0 à 20 cigarettes	plus de 20 cigarettes	total
H	27 ou x	71 ou y	22 ou z	120
F	23 ou t	39 ou u	18 ou v	80
	50	110	40	200

La question est de savoir s'il y a corrélation entre le sexe et le tabac .
x, y, z, t, u, v sont les effectifs théoriques tels que x/50=120/200 et y/110=120/200 et ..
donc x=30, y=66, z=24, t=20, u=14, v=16 ;

On utilise le test du χ^2 :

							Total
n_i	27	71	22	23	39	18	
n_{iTh}	30	66	24	20	44	16	
$n_i - n_{iTh}$	-3	5	-2	3	-5	2	
$(n_i - n_{iTh})$^2	9	25	4	9	25	4	
( )^2/ n_{iTh}	0,3	0,3788	0,1667	0,450	0,5682	0,250	2,1137

On a 2 lignes et 3 colonnes donc (2-1)*(3-1) =2 degrés de liberté . 2,77 < 2,11 < 1,39
La valeur calculée est inférieure à la valeur critique au seuil de risque de 10% , même de 25%: elle est supérieure à celle au seuil de 0,50 . Ce qui veut dire que l'on a 0,50 risque de rejeter l'hypothèse de non corrélation alors qu'elle est vraie .

***règle :* Si le $\chi 2$ calculé est supérieur au $\chi 2$ théorique, on rejette l'hypothèse avec le seuil de risque correspondant de se tromper c'est à dire qu'elle soit vraie .**
Si le$\chi 2$ calculé est inférieur au $\chi 2$ théorique, on ne peut rejeter l'hypothèse .

II AJUSTEMENT LINEAIRE

1) Situation :

Dans le cas de régression linéaire, on cherche à remplacer le nuage de points par une droite ;
Pour minimiser à O la somme des écarts d'ordonnées entre les points représentatifs et la droite, la droite solution du problème doit passer par le barycentre du nuage de points, G(X;Y) appelé aussi point moyen du nuage ..

2) Méthode de Mayer :

Amélioration des méthodes purement graphiques ;
On partage le nuage de points en deux sous-ensembles de même effectif .
On cherche le point moyen de chaque nuage, G1, G2 ;
La droite d'ajustement sera la droite (G1,G2) qui passe par G(X;Y)
La méthode est ancienne ;
La droite dépend du choix des nuages, mais elle est valable si G1 et G2 sont éloignés ;

3) Méthode des moindres carrés : la plus répandue

La méthode consiste à minimiser la somme des carrés des écarts d'ordonnées entre les points représentatifs et la droite ; $S = \Sigma(y_i-ax_i-b)^2$;
la meilleure droite solution du problème vérifie l'équation :
Y-y = a.(X-x) avec a = Cov(X;Y) / V(X) et G(x,y)
Cette droite est appelée droite de régression de Y en X .
La droite de régression de X en Y ,
qui minimise la somme des carrés des écarts d'abscisses, a pour équation :
X-x = a'.(Y-y) ou X = a'Y+b' avec a' = Cov(X;Y) / V(Y) et G(x,y)
ou Y = (1/a') X-(b'/a')
Les deux droites passent par G, point moyen du nuage ;
Remarque : $a*a'=r^2$
Les droites D et D' sont confondues si et seulement si a=1/a' ou r*r=1 d'où l'importance du calcul de r qui validera l'étude préalablement .

4) Droite d'Allométrie : *ref Gacogne p 116*

Pour faire jouer un rôle symétrique à X et Y , on cherche la droite qui minimise le produit des écarts d'abscisses et d'ordonnées entre les points représentatifs et la droite à déterminer.
Elle vérifie l'équation :
Y-y = a.(X-x) avec $|a| = \sigma_x/\sigma_y$ et G(x,y)

5) Exemple :

Une entreprise étudie l'impact de la durée de publicité par semaine sur les quantités vendues la semaine suivante.
X désigne la durée et Y la quantité.
Représenter le nuage, calculer le coefficient de corrélation, déterminer la droite d'ajustement .

											Total
X	13,2	0	18,6	8,8	12,3	28,3	0	23,5	5,6	0	110,3
Y	2	1	3,5	0,5	2,5	4	0	5	2	2	22,5
X^2	174,24	0	345,96	77,44	151,29	800,89	0	552,25	31,36	0	2133,43
Y^2	4	1	12,25	0,25	6,25	16	0	25	4	4	72,75
X*Y	26,4	0	65,1	44	30,75	113,2	0	117,5	11,2	0	368,55

Mayer :

Représenter le nuage, déterminer une droite d'ajustement .

$X1=10,58$ $Y1=1,9$ $X2=11,48$ $Y2=2,6$

$Y-1,9 = 0,7 * (X - 10,58) / 0,9$ soit $Y = 0,78 * X - 6,33$

MMC :

Moyenne de X : 11,03 ; Moyenne de Y : 2,25 ;
$\Sigma Xi^2 = 2199,43$ $\Sigma Yi^2 = 72,75$
Variance de X : 91,6821 ; Variance de Y : 2,2125 ;
Covariance de X et Y : 12,0375 ; Coefficient de corrélation : 0,8452 ;
Droite d'ajustement linéaire $Y=0,1313*X+0,8018$;

Droite d'allométrie : $Y-2,25 = 6,43*(X-11,03)$

Estimer alors le montant des ventes pour 20 mn de publicité ; *(3,43)*
Peut-on encore l'estimer pour 50 mn ?

III AJUSTEMENTS SE RAMENANT A LA FORME LINEAIRE :

1) Situation :

Sur cet exemple, dessiner le nuage de points, calculer le coefficient de corrélation ;

						Total
X	2	5	3	1	4	15
Y	7	104	18	3	43	175
X^2	4	25	9	1	16	55
Y^2	49					13 047
X*Y	14					763
Z=LnY	1,9459	4,6444	2,8904	1,0986	3,7612	14,3045
Z^2	3,7865	21,5705	8,3544	1,2069	14,1466	49,0649
Z*X	3,8918	23,2220	8,6712	1,0986	15,0448	51,928

Moyennes de : X =3 Y = 35 Z = 2,86
Variances de V(X) = 2 V(Y) = 1 384,4 V(Z) = 180
Cov(X,Y) = 47,6 Cov(X,Z) = 1,78
r(X,Y) = 0,9046 r(X,Z) = 0,9998

2) Régression exponentielle :

Si Y = k. exp(aX) = k.r^X = exp(aX+b) Z=LnY=aX+b ;
Il suffit de faire un ajustement linéaire entre X et Z.

3) Régression logarithmique :

Si Y = b + a.Ln(X) , Il suffit de faire un ajustement linéaire entre LnX et Y

4) Régression monomiale :

Si Y = k.X^a ; LnY = a.LnX + b ; Il suffit de faire un ajustement linéaire entre LnX et LnY

5) Exemple :

								Total
	Xi	3,5	4,25	5,13	6,05	7,32	8,55	
	Yi	100	154	225	310	455	620	
R.L.S.	Xi^2							
	Yi^2							
	XiYi							
R.L.E	Z=LnY							
	Zi^2							
	ZiXi							
R.L.Ln	T=LnXi							
	Ti^2							
	TiYi							
R.L.M.	T=LnXi							
	Z=LnY							
	Ti^2							
	Zi^2							
	TiZi							

Pour chacun des modèles, on obtient comme coefficient de corrélation :

RLS 0,9934 RLE : 0,9895 RLL : 0,9709 RLM : 0,9997

Le meilleur ajustement sera sous forme de puissance, pourtant l'ajustement linéaire affine est acceptable et plus simple .

RLS Moyenne de X = 5,8; V(X) = 3,01
Moyenne de Y = 310,67 V(Y) = 32130
Cov(X;Y) = 309,11 r = 0,9934 a = 102,6

Dans l'ajustement linéaire en puissance, l'exposant de la variable donne l'ordre de la fonction.

Dans la pratique, on commence par tracer le nuage de points pour avoir une idée de la situation .

Le nuage de points peut-être dessiné sur du papier millimétré particulier : logarithmique ou semi-logarithmique .

IV APPLICATIONS :

1) Exercice 1 : Cas E

Traduire les indications du tableau relatif à l'étude de familles :

Nb d'enfants Revenus	0	1	2	3	>3	Total
de 3000 à 6000	1	2				3
de 6000 à 9000	2	4	3	2	1	12
de 9000 à 12000	1	2	3	1	2	9
de 12 à 15000	2	1	1	1	1	6
Total	6	9	7	4	4	30

Définir et préciser les variables marginales .
Comment peut-on représenter la situation ?

cf TCC1M042.DOC

£6 : ETUDE D'UNE SERIE cas particulier des indices

COURS

I NOTION d'INDICE :

Dans le domaine des Sciences Economiques, Sociales, les grandeurs varient dans le temps, l'espace ou même les deux.

Par exemple, la production industrielle varie selon les secteurs,
un bien A passe de 53 à 64 et B de 128 à 154 unités ;

Pour comparer des résultats statistiques et les interpréter facilement on considère les rapports des grandeurs de même nature à comparer.

I=G(t)/G(o)*100 Indice de prix, de production, taux de chômage, de change .

Exemple : Prix de la baguette de pain en 1970 0,57 F , en 1975 0,90 F
en 1994 3,70 F.

Quel est l'indice de prix en 1975 , en 1994 avec la base de 100 en 1970 ?
la base de 100 en 1975 ?

base 100	en 70	100*0,90/0,57=157,89 ;	en 94	100*3,70/0,57=649,12 .
base 100	en 75	100*0,90/0,90=100	en 94	100*3,70/0,90=411,11

On distingue les indices simples ou élémentaires calculés sur une seule grandeur,
des indices synthétiques calculés à partir de plusieurs grandeurs.

II INDICES ELEMENTAIRES

1) Définition :

I(G;t/o)=G(t)/G(o)*100 Soit i(G;t/o)=I/100, i est un coefficient multiplicateur
avec G la grandeur considérée, t l'époque étudiée, o la référence.

2) Propriétés :

a) Réversibilité :

i(G;o/t)=1/i(G;t/o) donc **I(G;o/t)=1 / I(G;t/o) *100**^**2**

b) Enchaînement :

i(G;c/a) = i(G;c/b) * i(G;b/a) propriété du coefficient multiplicateur .

i(;75/70) = i(;75/94) * i(; 94/70) = (1/411,11)*649,12=157,89 .

c) Changement de base :

i(G;c/b) = i(G;c/a) / i(G;b/a)

i(;94/75) = i(;94/70) / i(;75/70)

III INDICES SYNTHETIQUES

1) Exemple :

a) Situation : Données :

Articles	*Prix en francs*			*Quantité en Kg*			*Indice de prix I(P;1/0)*		
Périodes	0	1	2	0	1	2	0	1	2
Viande	20	30	38	10	12	14	100	150	
Pain	4	8	12	30	20	15	100	200	

Compléter le tableau .

b) ***Difficultés des moyennes*** : et de leur signification :

moyenne arithmétique des indices de prix I(P;1/0) = (150+200)/2 = 175 ;

moyenne géométrique des indices de prix I(P;1/0) = $\sqrt{(150*200)}$ = 173,20 ;

moyenne harmonique : 2/I = 1/I1 + 1/I2 ;

Les produits changent dans leurs prix et dans leurs quantités ;

les indices doivent être significatifs avant d'être utilisables ;

il faut introduire des pondérations dans les calculs .

c) Il semblerait naturel de prendre pour indice de prix, **l'indice global** de coût .

I(C;1/0) = (12*30+20*8)/(10*20+30*4)*100 = 162,5 ;

Cet indice nécessite la connaissance de nombreuses valeurs .

2) Indices Classiques :

a) ***de Laspeyres***

Objet :

On compare dans le temps, les valeurs d'un panier de consommation

Consommation fixée à l'époque de référence .

pour l'indice de prix, le volume est fixé ;

pour l'indice de quantité, le prix est fixé.

Le relevé des prix est simple, le relevé des quantités consommées est long et coûteux.

. *Exemple* :

IL(P;1/0) = (30*10+8*30)/(20*10+4*30)*100 = 168,75

IL(Q;1/0) = (12*20+20*4)/(20*10+4*30)*100 = 100

Formules :

Indice des prix de Laspeyres : **IL(P;1/0) = [ΣP(i;1)*Q(i;0)] / [ΣP(i;0)*Q(i;0)] *100**

C'est l'indice des coûts à volumes constants à l'époque 0.

C'est aussi : **IL(P;1/0) = [ΣP(i;0)*Q(i;0)*I(Pi;1/0)] / [ΣP(i;0)*Q(i;0)]**

Soit la moyenne arithmétique pondérée des indices élémentaires des prix ,
indices pondérés par les coûts élémentaires à la date de référence.

Indice des volumes de Laspeyres : **IL(V;1/0) = [ΣP(i;0)*Q(i;1)] / [ΣP(i;0)*Q(i;0)] *100**

b) ***de Paasche***

Objet :
On compare dans le temps, les valeurs d'un panier de consommation ;
Consommation fixée à l'époque considérée
pour l'indice de prix, le volume est fixé ;
pour l'indice de quantité, le prix est fixé.

Exemple :
IP(P;1/0) = (30*12+8*20)/(20*12+4*20)*100 = 162,50

Formules :
Indice des prix de Paasche : **IP(P;1/0) = [ΣP(i;1)*Q(i;1)] / [ΣP(i;0)*Q(i;1)] *100**
C'est l'indice des coûts à volumes constants à l'époque 1.
C'est aussi **1 / IP(P;1/0) = [1 / ΣP(i;1)*Q(i;1)] * [ΣP(i;1)*Q(i;1)/I(Pi;1/0)]**
C'est en fait la moyenne harmonique pondérée des indices élémentaires des prix , indices pondérés par les coûts élémentaires à la date considérée.

c) ***Remarque*** : Ces indices sont non enchainables .
Mais **IL(P;1/0) * IP(P;0/1) = 10000** et **IL(P;/)*IP(Q;/) = IL(Q;/)*IP(P;/)**

3)Exercice d'application :

Un responsable a relevé en novembre et décembre les prix et quantités de trois produits A, B, C ;

Mois	Articles	Quantités	Prix	Valeurs
	A	15	10	150
11	B	20	5	100
	C	25	8	200
	A	25	12	300
12	B	25	6	150
	C	35	15	525

Indice de quantités (Prix fixés)
formule de Laspeyres :IL(Q;12/11) = (25*10+25*5+35*8)/(150+100+200) = 145,55 ;
formule de Paasche : IP(Q;12/11) = (300+150+525)/(15*12+20*6+25*15) = 144,44
Indice de prix (Quantités fixées)
formule de Laspeyres :IL(P;12/11) = (15*12+20*6+25*15)/(150+100+200) = 150 ;
formule de Paasche : IP(Q;12/11) = (300+150+525)/(25*10+25*5+35*8) = 148,85

4) Indices Usuels :

La source officielle des données statistiques est l'INSEE ; elle publie annuaires, bulletins .
indice de hausse des prix : à qualité égale ;
indice de coût de la vie : prix minimum même avec une amélioration de qualité ;
indice de consommation : pondération en fonction des besoins .
Indices de la consommation de l'INSEE :
295 articles : base 100 en 1970 indice de Laspeyeres car consommation fixe .

Indice de Construction ;
Indices des prix des matières premières , Indices boursiers .
Indice de la production industrielle

indice de volume de production selon la formule de Laspeyres, base 100 en 1970, les pondérations sont proportionnelles aux valeurs ajoutées aux coûts des facteurs de chaque branche.

T.P.

EXEMPLE 1 :

	1981			1987		
	Prix Po	Production Qo	Valeur de production	Prix P1	Production Q1	Valeur de production
Pétrole	1460F/t	1676000 t	*2447*	1618F/t	1213000 t	*1963*
Charbon	515F/t	18588000 t	*9573*	635F/t	13455000 t	*8544*
Gaz	1,17F/m3	7084*10^6 m3	*8288*	2,31F/m3	3842*10^ 6m3	*8875*
Electricité	0,30F/KW	172830*10^6 kw	*51849*	0,52F/KW	325200*10^6 kw	*169104*
	Coût total production	72 157 MF		Coût total production	188 486 MF	

1) Calculer l'indice simple du coût total de production de 87 par rapport à 81
(indice global de valeur de production) .
2) Calculer les indices de Laspeyres :
a) des prix de 87 par rapport à 81 ;
b) des quantités de 87 par rapport à 81 ;
3) Calculer les indices de Paasche :
a) des prix de 87 par rapport à 81 ;
b) des quantités de 87 par rapport à 81 ;

£7 : ETUDE D'UNE SERIE Chronologique

I SITUATION :

1) Définition :

Une série chronologique ou chronique est une série double dont la première variable est le temps, souvent repéré à intervalles réguliers ;
le but de l'étude est de pouvoir faire des prévisions au-delà des périodes observées.

2) Exemple :

Evolution du chiffre d'affaires d'une entreprise :

						Total
Année	1990	1991	1992	1993	1994	
N° X	1	2	3	4	5	15
C.A. Y	111	115	118	119	123	
Xi^2						
Yi^2						
Xi*Yi						

Moyenne de X = 3 ; Moyenne de Y = 117,2 ;
V(X) = 2 ; V(Y) = 16,16 ; Cov(X,Y) = 5,6 ; r = 0,985 ;
Un ajustement linéaire donne :
$Y = 2,8 * X + 108,8$; On peut espérer 125,6 de C.A. en 1995 ;

3) Problème :

La périodicité des mesures dépend de la nature de la variable et des objectifs recherchés.
Si une production industrielle nationale doit être étudiée sur l'année, le montant des ventes publicitaires d'un magasin peut-être journalière.
Les antécédents relevés sont fonction de la projection envisagée, court , moyen ou long terme.
La résolution de chaque problème ne doit pas se réduire à des calculs ; c'est risquer de graves erreurs que d'ignorer des facteurs économiques, psychologiques sociologiques ou même politiques.

4) Décomposition de séries chronologiques :

Il est facile d'imaginer qu'une série chronologique n'aie pas la régularité d'une fonction ;
des variations peuvent être dues à l'évolution, à la saison ou cycle, ou encore à des accidents .
Les observations sont décomposées :

sur une longue période par la " tendance " ou trend ,
sur le moyen terme, par le mouvement cyclique ou saisonnier,
et en variations accidentelles ou résiduelles .

II REPRESENTATIONS :

Exemples :

Ex A :

Année		**91**				**92**				**93**				**94**		
trim.	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Xi	10	12	13	14	11	15	16	13	12	17	18	15	13	17	19	16

Ex B :

Cas d'un chiffre d'affaires trimestriel industriel

Trimestres Années	1	2	3	4
1991	8	10	7	30
1992	13	14	8	40
1993	16	18	11	50
1994	20	20	14	60

Ex C :

Année		**91**				**92**				**93**				**94**		
trim.	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Xi	100	130	192	313	298	290	303	347	311	259	237	234	160	99	72	80

Graphique arithmétique :

Faire la représentation graphique pour apercevoir l'impression de tendance ;

sur les quatre ans,
puis en quatre graphiques superposés sur un an .

Graphique polaire :

Faire la représentation graphique pour apercevoir l'impression de tendance ;

Anamorphoses semi-logarithmique ou logarithmique :

III ANALYSE de TENDANCE :

1) Méthode des moindres carrés :

La méthode d'ajustement des moindres carrés permet par des calculs classiques déjà étudiés d'obtenir une tendance linéaire ou s'y ramenant .
Dans le cas de l'exemple B , on obtient : Y = 1,67*X+7 avec le numéro du trimestre repéré.

2) Méthode des moyennes mobiles :

a) Principe : On remplace une mesure par une moyenne des mesures voisines
La période doit être choisie en fonction de la périodicité des résultats ; sur un an on considère les douze mois ou les quatre trimestres ou encore les trois qua trimestres.
Sur un an, les moyennes mensuelles peuvent être calculées sur 3, 4 et même 12 mois si les variations sont importantes.
b) Méthode :
Sur trois mois, (X1+X2+X3)/3 remplacera X2 des mesures (X1;X2;X3) ;
Sur quatre mois, (0,5*X1+X2+X3+X4+0,5*X5) remplacera X3 des mesures (X1;X2;X3;X4;X5)

Exemple A :

Année		**91**				**92**				**93**				**94**		
trim.	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Xi	10	12	13	14	11	15	16	13	12	17	18	15	13	17	19	16
Yi			12,4													

Exemple C :

Année		**91**				**92**				**93**				**94**		
trim.	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Xi	100	130	192	313	298	290	303	347	311	259	237	234	160	99	72	80
Yi			209	253	287											

c) Dessiner les graphiques correspondants associés au nuage de points

d) Remarque : Il est des cas où il faut corriger les données : par exemple dans le cas de productions en fonction du nombre de jours ouvrables .
On retrouve là, la notion de précaution et de jugement .

IV ANALYSE des VARIATIONS SAISONNIERES :

1) Situations : *(illustrer les deux situations par un graphique)*

Dans le cas de l'exemple B, on constate que une dilatation générale des variations saisonnières comme de la tendance ; mathématiquement, il s'agit d'une homothétie.
Dans le cas de l'exemple C, la courbe reste dans une bande " parallèle autour de la courbe de tendance ; il s'agit de translation .
Dans le cas B, on cherchera un modèle multiplicatif, dans le cas C un modèle additif .

2) Modèle multiplicatif :

a) Méthode :
rapport saisonnier : **Ri** c'est pour chacune des mesures le rapport de la valeur brute observée à la valeur de tendance correspondante .
estimation des coefficients saisonniers : ***Ci*** pour chacune des périodes, on calcule la moyenne des rapports saisonniers.(par exemple par trimestre)
détermination des coefficients saisonniers : **Ci** la somme des coefficients devrait être égale au nombre N de mesures par période (4 pour des trimestres) ;
on prendra **Ci = N * *Ci* / Σ*Ci***
b) Exemple :

3) Modèle additif :

a) Méthode :
différence saisonnière : **Di** c'est pour chacune des mesures la différence entre la valeur brute observée et la valeur de tendance correspondante .
estimation des coefficients saisonniers : ***Di*** pour chacune des périodes, on calcule la moyenne des différences saisonnières.(par exemple par trimestre)
détermination des coefficients saisonniers : **Di** la somme des coefficients devrait être égale à 0 ;
On prendra : **Di = *Di* - (Σ*Di*)/ N** avec N le nombre de mesures par période.
b) Exemple :

V ANALYSE des VARIATIONS RESIDUELLES :

1) Définition :

C'est la différence entre la mesure observée ou donnée brute et la valeur de tendance corrigée par le coefficient saisonnier correspondant .

2) Exemples :

£8 : ETUDE D'UNE SERIE
cas de plusieurs caractères
Problèmes de Contingence

TP CM

I PRESENTATION d'une SERIE DOUBLE (effectifs) :

1) Situation :

On étudie dans la classe, le nombre de filles ou garçons ainsi que le nombre d'enfants de la famille .

Sexe Nb d'enfants	Y1 = F	Y2 = G
X1 = 1	3	0
X2 = 2	10	4
X3 = 3	2	7
X4 > 3	0	2

La population est la classe, l'individu est l'étudiant, un caractère est quantitatif, l'autre est qualitatif.
Les modalités de la variable Y sont F et G ; celles de Xi sont i .

2) Notations :

Le tableau des effectifs est appelé tableau de contingence de la série double ;
Zi,j = Xi et Yj ; l'effectif associé à Zi,j est noté $n_{i,j}$. Déterminer $n_{1,2}$;

3) Exercice :

Etudier la série définie par le sexe après avoir présenté la série ;
Etudier de même la série définie par le nombre d'enfants ;
On peut définir une série qualitative avec Z en perdant la notion d'association .

4) Définition :

Reconstruire le tableau :

Sexe / Nb d'enfants	Y1 = F	Y2 = G	Total
X1 = 1	3	0	
X2 = 2	10	4	
X3 = 3	2	7	
X4 > 3	0	2	
Total			

Combien vaut $n_{2,1}$? Que représente $n_{2,1}$?
Combien y a-t-il d'enfants uniques ? On le note N1,.
Combien de filles dans la classe ? Comment le noter ?
Ecrire N3,. ; N.,2 ;
Où écrire N = N.,. l'effectif total ? Vérifier que N=ΣNi,.=ΣN.,j ;
N2,. = ΣN2,j. ; c'est l'effectif marginal de la valeur X2 .
(les effectifs marginaux s'écrivent dans la marge)
Déterminer l'effectif marginal de Y2 .
L'étude d'un seul caractère revient à l'étude de la série marginale des effectifs .

5) Exercice :

On donne le tableau des séries marginales, peut-on reconstruire le tableau de contingence ?

Y / X	Y1	Y2	Ni,.
X1			15
X2			5
N.,j	8	12	

Combien de tableaux différents peut-on reconstruire avec des effectifs entiers ? *(6)*

II PRESENTATION d'une SERIE DOUBLE (fréquences) :

1) Situation :

Reprendre la situation de la partie précédente et étudier les fréquences .

Sexe Nb d'enfants	Y1 = F	Y2 = G	fréquence
X1 = 1	3/28	0/28	
X2 = 2	10/28	4/28	
X3 = 3	2/28	7/28	
X4 > 3	0/28	2/28	
fréquence			

a) Calcul marginal

Calculer les fréquences marginales ;

Déterminer : $f_{i,j}$; F2,. ; F.,1 ;

b) Calcul conditionnel

Reprendre la situation en ne s'intéressant qu'aux familles de 2 enfants .

	Sexe Nb d'enfants	Y1 = F	Y2 = G	fréquence
effectifs	X2 = 2	10	4	
fréquences	X2 = 2			

Quelle est la fréquence de Y1 parmi les familles de 2 enfants ?

On étudie ainsi le caractère Y conditionné par la valeur X2 de X .

f(Y1/X2) = $n_{1,2}$ / N2,.= $f_{1,2}$ / F2,.

10/14 est la fréquence de filles conditionnées par X2 .

Que vaut f(X2/Y1) ?

III CAS d'une série double Quantitative de données regroupées :

1) Situation :

On a relevé les couples de résultats poids , taille dans la classe .

Tranche	Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8
	X1= 40	4	3	1	0
	X2= 50	2	5	2	0
	X3= 60	1	2	6	2
	X4= 70	0	1	2	3
	X5= 80	1	0	3	2

2) Calculs marginaux :

Tranche	Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8	Ni,.	Ni,.*Xi	Ni,.*Xi^2
	X1= 40	4	3	1	0			
	X2= 50	2	5	2	0			
	X3= 60	1	2	6	2			
	X4= 70	0	1	2	3			
	X5= 80	1	0	3	2			
	N.,j							
	N.,j*Yj							
	N.,j*Yj^2							

Compléter le tableau ;
Faire apparaître les effectifs marginaux ;
Calculer la moyenne X de la série marginale des effectifs du caractère X (moyenne marginale)
Calculer la moyenne Y de la série marginale des effectifs du caractère Y (moyenne marginale)
Calculer la variance V(X) de la série marginale du caractère X (variance marginale)
Calculer la variance marginale V(Y) de la série des effectifs du caractère Y

3) Calculs conditionnels :

a) La série obtenue en se limitant à X=60 est appelée ensemble de régression de Y en X ;

Tranche	Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8	$\Sigma n_{3,j}$	$\Sigma n_{3,j}$*Yj
	X3= 60	1	2	6	2		

Calculer la moyenne de cette série ; que représente-t-elle ? On la notera Y/X3

b) Calculer les moyennes de Y conditionnées par les autres valeurs de X .

Tranche	Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8	Ni,.	$\Sigma n_{i,j}$*Yj	Y/Xi
	X1= 40	4	3	1	0	8	12,5	1,56
	X2= 50	2	5	2	0		14,4	1,60
	X3= 60	1	2	6	2		18,5	1,68
	X4= 70	0	1	2	3		10,4	1,73
	X5= 80	1	0	3	2		10,2	1,70
	N.,j							
	$\Sigma n_{i,j}$ *Xi							
	X/Yj							

c) Calculer les moyennes de X conditionnées par les valeurs de Y .

d) On peut représenter ces deux courbes de régression avec deux couleurs et un bon choix du repère commun .

4) Calculs globaux :

a) en Ni*X : en Ni*Xi^2

Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8
X1= 40	4*40	3*40	1*40	0
X2= 50	2*50	5*50	2*50	0
X3= 60	1*60			
X4= 70	0*70			
X5= 80	1*80			
Σ				

Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8
X1= 40	4*1600			
X2= 50	2*2500			
X3= 60				
X4= 70				
X5= 80				
Σ				

b) en Nj*Yj

Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8
X1= 40	4*1,5	3*1,6	1*1,7	0
X2= 50	2*1,5	5*1,6	2*1,7	0
X3= 60	1*1,5			
X4= 70	0*1,5			
X5= 80	1*1,5			
Σ				

en Nj*Yj ^2

Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8
X1= 40	4*2,25			
X2= 50	2*2,25			
X3= 60				
X4= 70				
X5= 80				
Σ				

c) en X*Y

Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8	Σ
X1= 40	4*40*1,5	3*40*1,6	1*40*1,7	0	
X2= 50	2*50*1,5	5*50*1,6	2*50*1,7	0	
X3= 60	1*60*1,5				
X4= 70	0*70*1,5				
X5= 80	1*80*1,5				
Σ					

Calculer les moyennes de X, de Y puis les variances de X et Y .et enfin la covariance de XY .

5) Pratique :

Taille Y Poids X	Y1 1,5	Y2 1,6	Y3 1,7	Y4 1,8	Ni,.	Ni,.*Xi	Ni,.*Xi^2	$\Sigma n_{i,j}$*Yj	Xi* $\Sigma n_{i,j}$*Yj
X1= 40	4	3	1	0					
X2= 50	2	5	2	0					
X3= 60	1	2	6	2					
X4= 70	0	1	2	3					
X5= 80	1	0	3	2					
N.,j					N	X	V(X)		cov(X,Y)
N.,j*Yj					Y				
N.,j*Yj^2					V(Y)				
Σ $n_{i,j}$ *Xi									
Yj*Σ $n_{i,j}$ *Xi					cov(X,Y)				

£9 : ETUDE D'UNE SERIE cas de plusieurs caractères

I PRESENTATION d'une SERIE DOUBLE Regroupée :

1) Tableau de Contingence :

Construction : Tableau de données groupées :

Salaire net Yj Age Xi	[5;6[	[6;7[	[7;8[	Total
[20;22[	1200	500	100	1800
[22;24[	2500	3500	600	6600
[24;26[	1800	5000	2300	9100
Total	5500	9000	3000	17500

$n_{i,j}$ est l'effectif
$n_{2,1}$=2500
N.,2 est l'effectif marginal de Y2
N.,2=9000
N1,. est l'effectif marginal de X1
N2,.=6600
N.,.=N=17500

2) Définitions :

Effectif partiel : $n_{i,j}$ effectif correspondant à la modalité Xi et Yj notée Zij.
Fréquence partielle : $f_{i,j} = n_{i,j} /N$
Ni,. = $\Sigma n_{i,j}$ N.,j = $\Sigma n_{i,j}$
Fréquence marginale : f.,j = N.,j / N
Fréquence conditionnelle : f(Xi/Yj) = $n_{i,j}$ / N.,j ; c'est la proportion d'individus de la modalité Xi et Yj par rapport (parmi) la modalité Yj .

3) Exemple :

Pour étudier la variable X, on considère la série (Ni,. ; Xi)

Pour étudier la variable Y, on considère la série (N.,j ; Yj)

Pour étudier la variable Z, on considère la série ($n_{i,j}$; Zi,j) ou (n_k , Z_k)

II CARACTERISTIQUES d'une SERIE DOUBLE Regroupée:

Il faut distinguer la variable couplée, la variable marginale et la variable conditionnelle .

1) Variable couplée : Zi,j = (Xi , Yj) d'effectif nij

Peut-on faire des carrés sur la différence de couples ? !
Faut-il se limiter à Z=X ou Z=Y

Moyenne : $M(Z)= (1/N).\Sigma\Sigma n_{i,j}\, z_{i,j}$

Variance : $V(Z) = (1/N)\Sigma\Sigma n_{i,j}\, (z_{i,j} -Z)^2 = (1/N)(\Sigma\Sigma n_{i,j}\, z_{i,j}^2)-Z^2$

2) Variable marginale : Zi,. = Xi *(en étudiant uniquement Xi)*

Moyenne : $Mm(X)= (1/N)\Sigma N_{i,.}z_{i,.} = (1/N).\Sigma N_{i,.}x_i$

Variance : $Vm(X) = (1/N)\Sigma\, N_{i,.}\, (z_{i,.} -X)^2 = (1/N)\Sigma\, N_{i,.}z_{i,.}^2 -X^2$

3) Variable conditionnelle : Zi/j *(en étudiant Xi en se limitant à Yj)*

Moyenne : $Mc(X)= (1/N.,j).\Sigma n_{i,j}z_{i,j}$,

Variance : $Vc(X) = (1/N.,j).\Sigma n_{i,j}(z_{i,j} -Xj)^2 = (1/N.,j).\Sigma n_{i,j}.z_{i,j}^2 -Xj^2$

4) Relations entre caractéristiques marginales et conditionnelles :

Mm(X)=(1/N)ΣN.,jXj .

5) Exemple : reprendre exemple d'introduction

III CAS d'une série de DEUX VARIABLES :quantitatives situations non uniques

1) Situation : Cas E

Traduire les indications du tableau relatif à l'étude de familles :

Y Nb d'enfants X Revenus	0	1	2	3	>3	Total
de 3000 à 6000	1	2				3
de 6000 à 9000	2	4	3	2	1	12
de 9000 à 12000	1	2	3	1	2	9
de 12 à 15000	2	1	1	1	1	6

Total	6	9	7	4	4	30

Définir et préciser les variables marginales .
Comment peut-on représenter la situation ?

2) Covariance :

Définition : **$Cov(X,Y) = (1/N)*\Sigma\Sigma n_{i,j}.(x_i-X)(y_i-Y)$** formule théorique

$Cov(X,Y) = (1/N)*\Sigma(\Sigma n_{i,j}.x_i\ y_i)-XY)$ formule pratique

C'est la moyenne pondérée des produits moins le produit des moyennes .
Remarque : Cov(X,X) = V(X)

3) Coefficient de corrélation linéaire :

Définition : **$r = [Cov(X,Y)] / (\sigma_X.\sigma_y)$** On a toujours $-1 \leq r \leq 1$.
Si $0,8 \leq |r| \leq 1$, la corrélation est forte et l'ajustement peut-être justifié donc envisagé .
Il traduit une forte concentration des points au voisinage d'une droite ou une courbe , pas de notion de cause à effet ; par exemple, le nombre vente de postes de télévision et le nombre d'entrée en hospitalisation de 1955 à 1995 peuvent être fortement corrélées mais éventuellement liés à un phénomène de société.

4) Ajustement Linéaire *Rappels*

Dans le cas de régression linéaire, on cherche à remplacer le nuage de points par une droite ;
Pour minimiser à O la somme des écarts d'ordonnées entre les points représentatifs et la droite, la droite solution du problème doit passer par le barycentre du nuage de points, G(x;y)
Méthode des moindres carrés : la plus répandue
La méthode consiste à minimiser la somme des carrés des écarts d'ordonnées entre les points représentatifs et la droite, meilleure droite solution du problème ; elle vérifie l’équation :
$Y-y = a.(X-x)$ avec $a = Cov(X;Y) / V(X)$ et $G(x,y)$
Cette droite est appelée droite de régression de Y en X .
La droite de régression de X en Y a pour équation :
qui minimise la somme des carrés des écarts d'abscisses, a pour équation :
$X-x = a'.(Y-y)$ ou $X = a'Y+b'$ avec $a' = Cov(X;Y) / V(Y)$ et $G(x,y)$
ou $Y = (1/a') X-(b'/a')$
Les droites D et D' sont confondues si et seulement si a=1/a' ou r*r=1 d'où l'importance du calcul préalable de r qui validera l'étude ultérieure .
Qualité d'ajustement : elle est mesurée par **$\sigma_c=\sigma_y*\sqrt{(1-r^2)}$** *ref Gacogne p 122*

5) Droite d'Allométrie : *ref Gacogne p 116*

Pour faire jouer un rôle symétrique à X et Y , on cherche la droite qui minimise le produit des écarts d'abscisses et d'ordonnées entre les points représentatifs et la droite à déterminer.
Elle vérifie l'équation :
$Y-y = a.(X-x)$ avec $|a| = \sigma_X/\sigma_y$ et $G(x,y)$

6) Exemple :

Exemple 1 : celui du II 4
Une entreprise étudie l'impact de la durée de publicité par semaine sur les quantités vendues la semaine suivante.(cas non linéaire par décalage rectifié dans le tableau)
X désigne la durée et Y la quantité.
Représenter le nuage, calculer le coefficient de corrélation, déterminer la droite d'ajustement .

											Total
X	13,2	0	18,6	8,8	12,3	28,3	0	23,5	5,6	0	110,3
Y	2	1	3,5	0,5	2,5	4	0	5	2	2	22,5
X^2	174,24	0	345,96	77,44	151,29	800,89	0	552,25	31,36	0	2133,43
Y^2	4	1	12,25	0,25	6,25	16	0	25	4	4	72,75
X*Y	26,4	0	65,1	44	30,75	113,2	0	117,5	11,2	0	368,55

Moyenne de X : 11,03 ; Moyenne de Y : 2,25 ;
Variance de X : 91,6821 ; Variance de Y : 2,2125 ;
Covariance de X et Y : 12,0375 ; Coefficient de corrélation : 0,8452 ;
Droite d'ajustement linéaire Y=0,1313*X+0,8018 ;
Droite d'allométrie : Y-2,25 = 6,43*(X-11,03)
Estimer alors le montant des ventes pour 20 mn de publicité ;
Peut-on encore l'estimer pour 50 mn ?

Exemple 2

Y X	2	4	6	8	N'i	Ni*Xi	NiXi^2	Xi* ΣnijYj
20	4	5	0	0	9	180	3600	560
40	0	3	10	2	15	600	24000	3520
60	0	1	20	28	49	2940	176400	20880
N"j	4	9	30	30	73	3720	204000	24960
N"j*Yj	8	36	180	240	464			
N"j*Yj^2	16	144	1080	1920	3160			

M(X) = 3720/73 = 50,96 V(X) = 204000/73 - 50,96^2 = 197,71 .
M(Y) = 464/73 = 6,36 V(Y) = 3160/73 - 6,36^2 = 2,89 .
Cov(X,Y) = 24960/73 - 50,96*6,36 = 18,01 . r^2 = 18,01^2 / (V(X).V(Y)) = 0,75^2
a = 18,01 / 197,71 = 0,09 Y = 0,09*X +1,77

IV CAS d'une série double de données regroupées :

V CAS d'une série double de variables qualitatives :

VI GENERALISATION de l' AJUSTEMENT

1) Ajustements se ramenant à la forme linéaire :

Situation :
Sur cet exemple, dessiner le nuage de points, calculer le coefficient de corrélation ;

						Total
X	2	5	3	1	4	15
Y	7	104	18	3	43	
X^2	4	25	9	1	16	55
Y^2						
X*Y						
Z=LnY	1,9459	4,6444	2,8904	1,0986	3,7612	14,3045
Z^2	3,7865	21,5705	8,3544	1,2069	14,1466	49,0649
Z*X	3,8918	23,2220	8,6712	1,0986	15,0448	51,928

Régression exponentielle :
Si Y = k. exp(aX) = k.r^X = exp(aX+b) Z=LnY=aX+b ;
Il suffit de faire un ajustement linéaire etre X et Z.

Régression logarithmique :
Si Y = b + a.Ln(X) , Il suffit de faire un ajustement linéaire entre LnX et Y

Régression monomiale :
Si Y = k.X^a ; LnY = a.LnX + b ; Il suffit de faire un ajustement linéaire entre LnX et LnY

Exemple :

								Total	
	Xi	3,5	4,25	5,13	6,05	7,32	8,55		
	Yi	100	154	225	310	455	620		
R.L.S.	Xi^2								
	Yi^2								
	Xi*Yi								
R.L.E	Z=LnY								
	Zi^2								
	Zi*Xi								
R.L.Ln	T=LnXi								
	Ti^2								
	Ti*Yi								
R.L.M.	T=LnXi								
	Z=LnY								
	Ti^2								
	Zi^2								
	Ti*Zi								

Pour chacun des modèles, on obtient comme coefficient de corrélation :

RLS 0,9934 RLE : 0,9895 RLL : 0,9709 RLM : 0,9997

Le meilleur ajustement sera sous forme de puissance, pourtant l'ajustement linéaire affine est acceptable et plus simple .

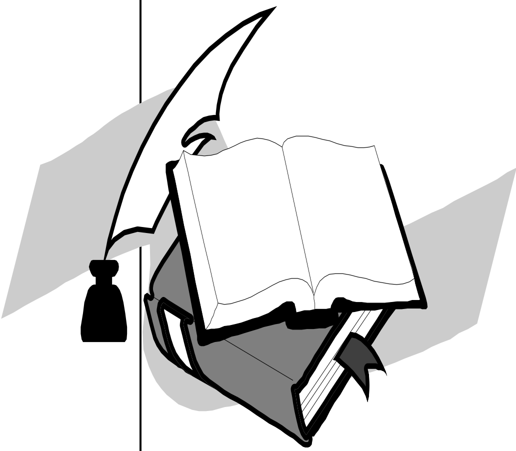
RLS	Moyenne de X = 5,8;	V(X) = 3,01	
	Moyenne de Y = 310,67	V(Y) = 32130	
	Cov(X;Y) = 309,11	r = 0,9934	a = 102,6

L'ajustement linéaire en puissance donne l'ordre de la fonction, l'exposant de la variable.

Dans la pratique, on commence par tracer le nuage de points pour avoir une idée de la situation .
Le nuage de points peut-être dessiné sur du papier millimétré particulier : logarithmique ou semi-logarithmique .

I.U.T. de REIMS

Département T.C.



RESUME du COURS de STATISTIQUES

Finalité Tertiaire

M. HUSSENET Jacky

SERIE à une Variable

Généralités

VOCABULAIRE :	
	Une série statistique est le résultat d'une expérience réalisée. Population **: c'est l'ensemble étudié.** Individu **: c'est l'élément étudié.** Caractère : X **c'est la propriété étudiée ;** **il peut être** qualitatif **ou** quantitatif **,** **s'il est quantitatif, il est** discret **s'il ne prend que des valeurs entières,** **ou** continu **s'il prend ses valeurs sur un intervalle.** Modalité **:**x_i **c'est le résultat possible du caractère.** Effectif **:** n_i **c'est le nombre d'individus associés à une modalité.**

DEFINITIONS :	
	Effectif total **:** $N = \Sigma n_i$ Série statistique **: ensemble des couples (xi ; ni)** Fréquence : $f_i = n_i / N$ Effectif cumulé **: croissant ou décroissant** Fréquence cumulée **: croissante ou décroissante.**

SERIE à une Variable

PRESENTATION

VOCABULAIRE :	
	Données brutes : Données ordonnées : Données regroupées :

PROCEDES :	
Tableaux :	Série simple : **modalité, effectif, fréquence, cumul.** Série double **: Tableau de contingence.**
Vocabulaire :	Série Simple : **(une variable)** **largeur de classe, centre de classe.** Série double **: (deux variables)** effectif marginal **:** N.,j = $\Sigma n_{i,j}$ fréquence partielle **:** $n_{i,j}$ / N fréquence marginale : N.,j / N **ou N**i,. / N fréquence conditionnelle **:** $n_{i,j}$ / N.,j ou $n_{i,j}$ / Ni,. **.**

SERIE à une Variable

REPRESENTATION

VOCABULAIRE :	
	Diagramme en bâtons Colonnes Bandes, Secteurs, Polygone des effectifs, des fréquences, (fonction de densité) **la représentation au centre de la bande ;** Histogramme : **la représentation de la surface est proportionnelle à l'effectif** ; Polygone des effectifs cumulés, des fréquences cumulées, (fonction de répartition) **la grandeur cumulée croissante est représentée fin de bande,** **la grandeur cumulée décroissante est représentée début de bande** Polaire, Pyramide des âges Figuratif Nuage de points.

PROCEDES :	
Série simple : **.** Série double **:**	discret **:** **Diagramme en bâtons, en Colonnes, en Bandes ou Secteurs** continu **Polygone ou Histogramme des effectifs ou des fréquences,** **Polygone des effectifs cumulés, des fréquences cumulées,** **Courbes de Distribution, de Répartition,** Série chronologique **:** **Polaire**, Série double **:** **Nuage de points.**

CARACTERISTIQUES

d'une série à une variable quantitative

RAPPELS :	
	Différencier : Effectif **: n_i, N** Caractères **X, Y** modalités : **x_i , y_j**

DEFINITIONS :	
Paramètres de position :	Mode ou dominante : **valeur du caractère dont l'effectif (éventuellement corrigé) est maximum.** Médiane : *cas discret* : **valeur du caractère d'abscisse (N+1)/2** *cas continu* : **valeur de l'abscisse du caractère tel que la fréquence cumulée vale 0,5.** Quartiles **: Qi tel que Σf_j = i.0,25. Q2=M .** Moyenne arithmétique : $M = (\Sigma n_i x_i)/N = \Sigma f_i x_i$
Paramètres de dispersion :	Etendue : $E = x_N - x_1$. Ecart interquartile : **Q3-Q1** Ecart moyen **: $\Sigma n_i \vert x_i - M \vert /N = \Sigma f_i \vert x_i - M \vert$** Variance : $V(X) = \Sigma n_i (x_i - M)^2/N = (\Sigma n_i x_i^2)/N - M^2$ $= \Sigma f_i (x_i - M)^2 = (\Sigma f_i x_i^2) - M^2$ Ecart-type : $\sigma = \sqrt{V(X)}$ Coefficient de variation : σ/M
Autres :	Coefficients de forme et de concentration : médiale : **Médiane de la série des masses.**

PROBLEMES D'AJUSTEMENT

SERIE à une Variable : Ajustement à une loi :	
Test du Khi-2 :	**La série statistique peut-elle être approchée par une loi de distribution classique : Uniforme, Poisson, Normale ?** Test du Khi-2 : **on calcule** $\chi^2_c = \Sigma (n_{io}-n_{iTh})^2/n_{iTh}$ **avec n_{io}, l'effectif observé, n_{iTh} l'effectif théorique ;** **on compare** χ^2_c **au** χ^2_α **, avec la table en fonction de α.** **si $\chi^2_c < \chi^2_\alpha$, l'ajustement est accepté au seuil α.**

SERIE à deux variables quantitatives : Ajustement fonctionnel :	
<u>Ajustement linéaire :</u> du type : Y=aX+b	Méthode des moindres carrés **à situations uniques : ni=1 ou à situations multiples.ni** **On calcule :** $Mx = \Sigma n_i x_i / N$ $My = \Sigma n_i y_i / N$et G(Mx ; My) $V(X) = (\Sigma n_i x_i^2 / N) - M^2x$ $V(Y) = (\Sigma n_i y_i^2 / N) - M^2y$ $Cov(X ; Y) = \Sigma n_i(x_i-Mx).(y_i-My) / N = (\Sigma n_i x_i.y_i)/ N - Mx.My$ $r = Cov(X ; Y) / (\sigma x. \sigma y)$ et si $\lvert r \rvert >0,8$, a = Cov(X ; Y) / V(X) **(ajustement de Y en X)** **on a :** Y-My = a(X-Mx) équation de la droite d'ajustement
	Allométrie : **(cas symétrique)** Y-My = a(X-Mx) avec $\lvert a \rvert = \sigma x / \sigma y$
Ajustement non linéaire : du type : Y=f(X)	*Exponentiel :* Y = exp(aX+b) *Logarithmique* : Y = aLn(X) + b *Monomial* : Y = kX^^{a}

SERIE à plus de deux variables quantitatives : :	
	Ajustement linéaire multiple ou matriciel :

SERIE CHRONOLOGIQUE

NOTION :	
	Série statistique d'une grandeur économique en fonction du temps.

REPESENTATIONS :	
	Graphiques **arithmétique polaire anamorphose** **Cas de dilatation (modèle multiplicatif)** **ou de glissement (modèle additif)**

CARACTERISTIQUES :	
Analyse de tendance	Y = aX + b **par la méthode des moindres carrés**
	Lissage **par la méthode des moyennes mobiles**

VARIATIONS SAISONNIERES :	
Modèle multiplicatif	**On calcule le rapport saisonnier :** Ri = Yio / Yit **puis une estimation du coefficient saisonnier :** *Ci* **la moyenne des rapports saisonniers.** **enfin le coefficient saisonnier** Ci = N.*Ci* / Σ*Ci*
Modèle additif	**On calcule la différence saisonnière :** Di = Yio - Yit **puis une estimation du coefficient saisonnier :** *Di* **la moyenne des rapports saisonniers.** **enfin le coefficient saisonnier** Di =(*Di* - (Σ*Di*)/N

VALEURS RESIDUELLES :	
	Cas multiplicatif Yi = (aXi +b) . Ci +Rsi Cas additif Yi = (aXi +b) - Di +Rsi