

L2 Géographie
UE Fondamentale 2



Statistiques – COURS 1

Salle 125

Intervenants : Nadège Martiny & Joseph Boyard-Micheau

UFR Sciences Humaines (Département de Géographie)
UMR Centre de Recherches de Climatologie (CRC)

Introduction générale

- 12 séances
- Epreuve Terminale : **12^{ème} séance, 16/12/2011** (durée **1h30**)
- + Contrôle continu : **2 mini-épreuves surprise** (notées sur 10)
- Emploi du temps :
Tous les vendredis 10h-12h G1 / 13h-15h G2
- Organisation des séances : Cours/TD
- Outil **Excel**
- Fonctionnement de la salle informatique

Bibliographie

- **Boursin J.L.** : Comprendre les statistiques descriptives. Armand-Colin, 1998
- **Chemla G.** : Statistiques appliquées à la géographie. Nathan Université, 1995
- **Groupe Chadule** : Initiation aux pratiques statistiques en géographie. Masson, 1997
- **Lahousse P. & Piedanna Y.** : L'outil statistique en géographie. Tome 2 : analyse bivariée. Armand-Colin, coll. Synthèse, 1999
- **Saporta G.** : Probabilités, analyse des données et statistiques. Technip, 1990

Introduction générale

- La compréhension des faits **géographiques** passe par la mise en évidence d'éventuelles **interactions** entre les phénomènes.

Exemple :

Y a-t-il des interactions entre l'intermodalité d'un réseau et la taille de l'agglomération?



Vision futuriste de la Gare de Dijon à l'horizon 2013



Gare d'Auxerre

Introduction générale

- Le géographe se pose donc souvent la question de l'existence d'une **relation de dépendance** entre deux **variables**.

*** Définition d'une variable :**

Caractère mesuré sur un certain nombre d'**observations**. En géographie les observations décrivent souvent l'espace ou le temps.

Introduction générale

Exemple :

Mesures de la radioactivité en France dans les années 1988-1991, à la suite de Tchernobyl (avril 1986)

Localité	Département	Césium 137 (Bq/m ²)
Saales		8675
Brumath		
Preuschedorf	67	6505
Neuve-Eglise	67	5407
Rothau	67	3249
Saverne	67	1078
Linthal	68	20200
Kruth	68	18780
Durmenach	68	17140

9 OBSERVATIONS
décrivant l'espace

2 VARIABLES

Extrait de base de données CRII-RAD, organisme drômois (Dumolard, Dubus et Charleux, « Les statistiques en Géographie »)

Plan du cours

CHAPITRE I.

Relation entre deux caractères ou plus

Introduction générale

- Le géographe est également souvent amené à étudier la **distribution (=répartition des valeurs) de variables** dans des **populations** très nombreuses.

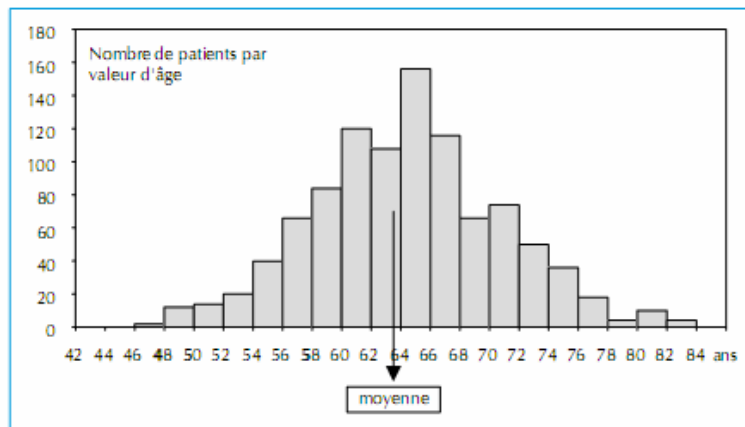


Figure 1, extraite de Laporte et Quenet, 2007

Exemple :

Distribution de l'âge sur une cohorte de 1000 patients au CHU de St Etienne

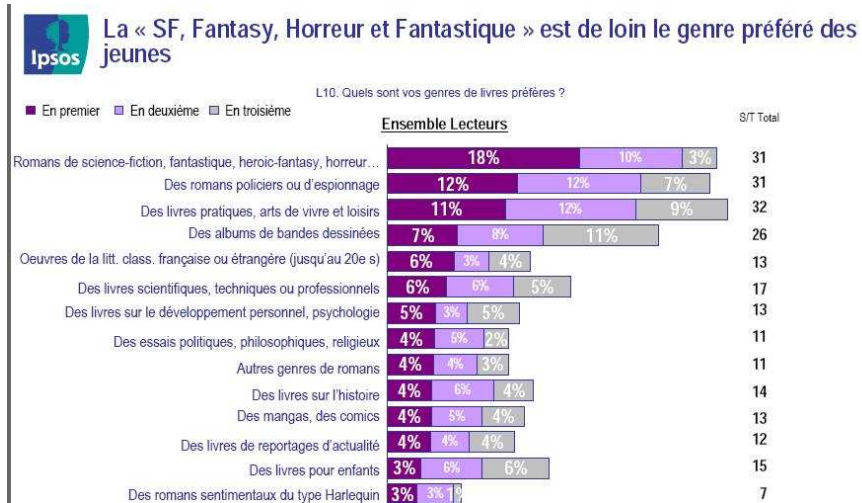
Variable = âge des patients

* Définition d'une population :

Ensemble analysé par les statistiques, qui constitue un **univers statistique**.

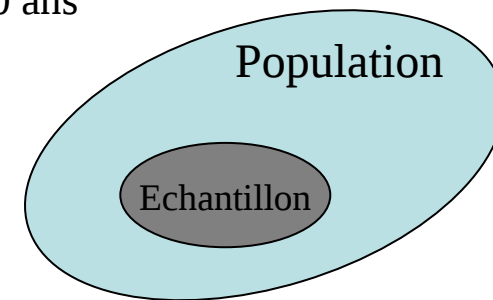
Introduction générale

- Les statistiques permettent aussi d'étudier des **échantillons** à partir desquels les **paramètres de l'ensemble** de la population peuvent être estimés.



Exemple :

Sondage IPSOS sur les genres de livres préférés des 18-30 ans



* Définition d'un échantillon :

Sous-ensemble des éléments d'une population choisis **au hasard** parmi celle-ci.

Plan du cours

CHAPITRE I.

Relation entre deux caractères ou plus

CHAPITRE II.

Echantillonnage

Introduction générale

- Les variables géographiques sont par ailleurs en constante **évolution dans le temps**, il est donc intéressant de mener des analyses temporelles **variable par variable**.

*** Définition d'une série chronologique :**

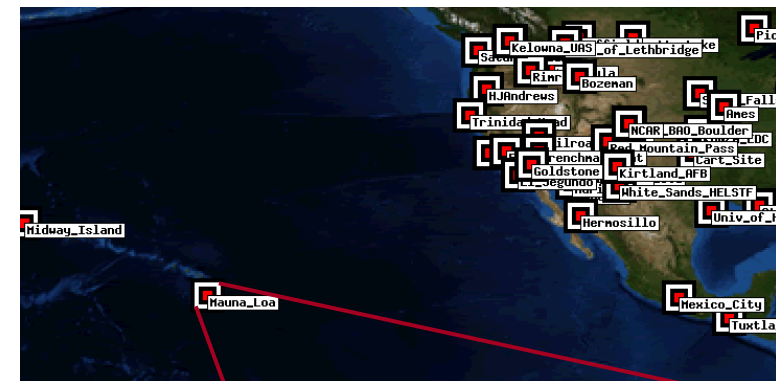
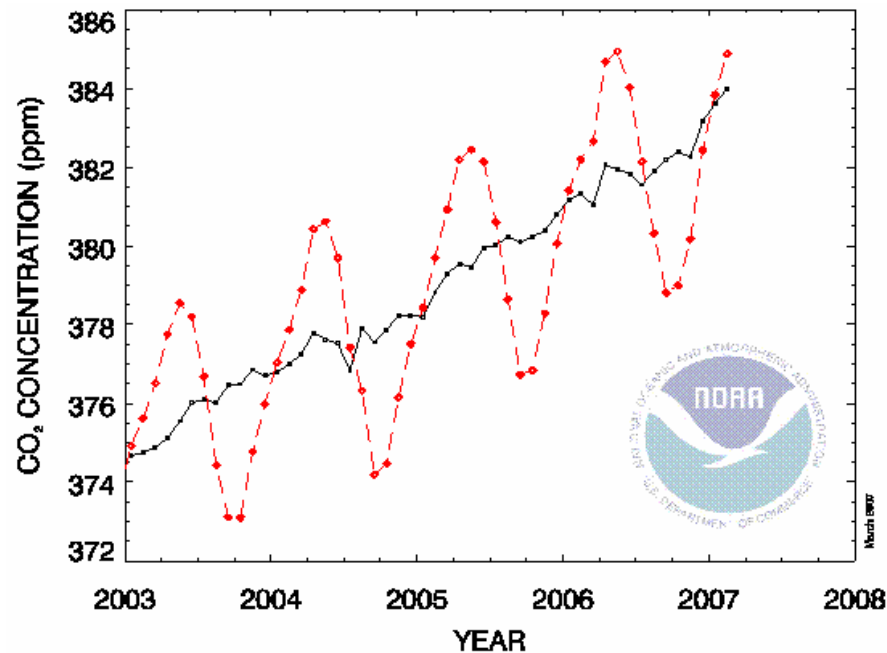
Suite d'observations chiffrées portant sur une **variable** (=caractère) **unique**, et ordonnées dans le **temps**.

On l'appelle également **chronique** ou **série temporelle**.

Introduction générale

Example :

Evolution de la concentration atmosphérique en CO₂ à **Mauna Loa**
en moyennes mensuelles (rouge) et corrigées de la saison (noir)



Réseau AERONET
NASA



Plan du cours

CHAPITRE I.

Relation entre deux caractères ou plus

CHAPITRE II.

Echantillonnage

CHAPITRE III.

Analyse de séries chronologiques

Chapitre I

I.1 Démarche générale

I.2 Relation entre deux caractères qualitatifs :
Tableau de contingence & Test du khi-deux

I.3 Relation entre un caractère qualitatif et un caractère quantitatif :
Analyse de variance

I.4 Relation entre deux caractères quantitatifs :
Corrélations et régressions linéaires simples

I.5 Relation entre plus de deux caractères quantitatifs :
Corrélations et régressions linéaires multiples

Démarche générale

Hypothèse de l'existence d'une relation entre 2 caractères



Constitution d'un **tableau** de données (**outil informatique**)



Quantification globale de la relation (**intensité**)



Test de significativité de la relation

*Effectué **si et seulement** si les données sont entachées d'erreur **aléatoire**
(données d'un échantillon et non d'une population, erreurs de mesures non systématiques)*



Si la relation est significative, **a-t-elle un sens?**

*Cette étape fait appel à la **culture** du géographe*

Exemple

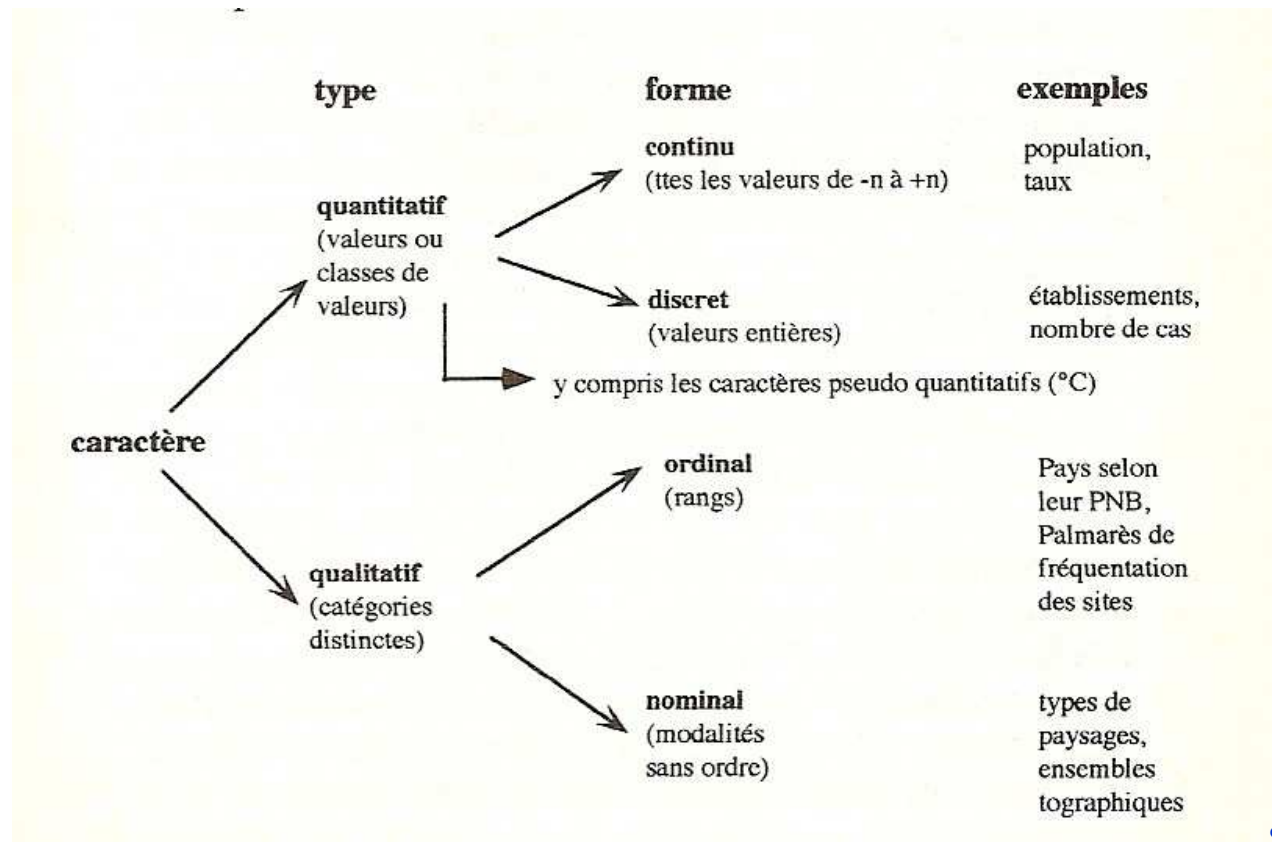
A propos de la **relation** entre :

- la **vente annuelle de chaussures pour 1000 habitants (Y)**
- le **taux annuel de maladies cardio-vasculaires pour 1000 habitants (X)** en Europe

- **Bonne relation statistique mais on ne peut en déduire** pour autant que la relation est **logique** :
 - (i) *Acheter beaucoup de chaussures provoque-t-il des maladies cardio-vasculaires?*
 - (ii) *Les malades cardio-vasculaires achètent-ils plus de chaussures que les autres?*
- **En réalité, X et Y sont toutes deux liées à **une 3^{ème} variable**, assez explicative pour X et Y : **le niveau de vie moyen des différents pays (Z)****
- **Imaginons maintenant qu'on s'intéresse à la relation entre X et Z**
- **Si la relation statistique entre ces 2 variables est **significative**, on peut supposer qu'il existe une relation physique ou de cause à effet entre elles... cependant, elle ne peut en aucune manière **le prouver** !!!**

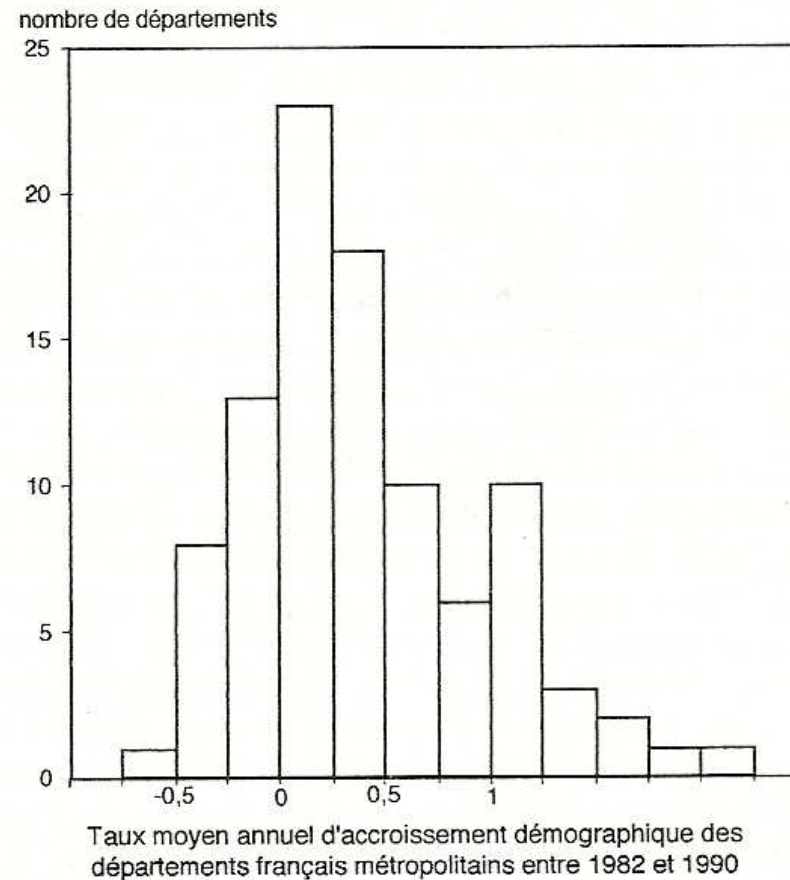
Classification des variables

- Avant d'effectuer le **calcul de la relation statistique** entre des variables, il faut identifier leur **type**.



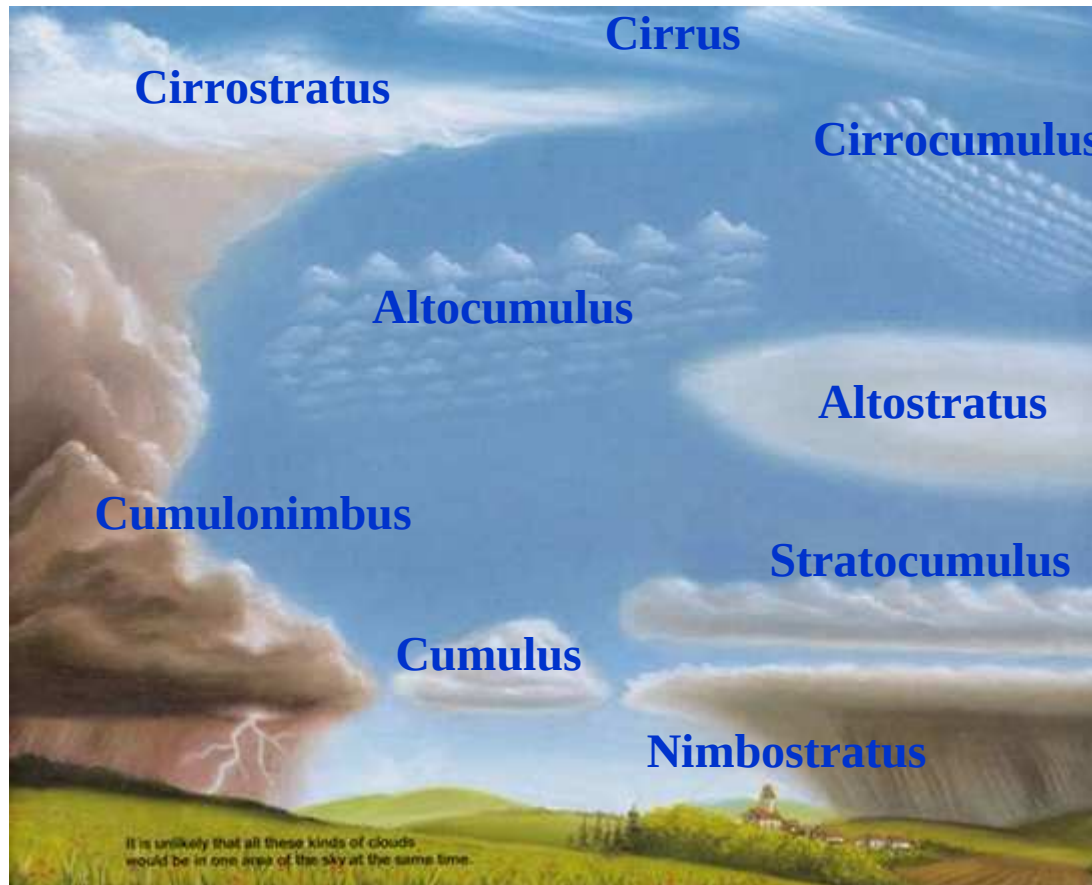
Source. Vigneron, 1997

Identification du type de variable



VARIABLE DE TYPE QUANTITATIF ET DE FORME CONTINUE

Identification du type de variable



Source. <http://la.climatologie.free.fr>

VARIABLE DE TYPE QUALITATIF ET DE FORME NOMINALE

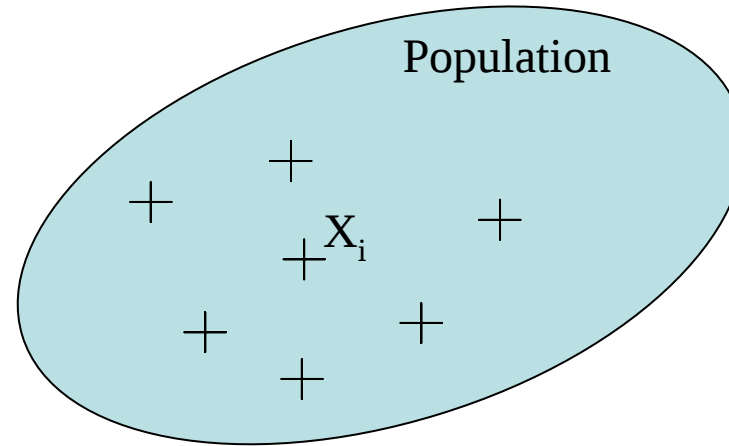
Types d'analyses (2 variables)

TYPE DE VARIABLES	QUALITATIVE	QUANTITATIVE
QUALITATIVE	TABLEAU DE CONTINGENCE & TEST DU KHI-2	ANALYSE DE VARIANCE
QUANTITATIVE	ANALYSE DE VARIANCE	REGRESSION & CORRELATION LINEAIRE SIMPLE

Notions de base

- Moyenne arithmétique :

$$M_X = \frac{1}{N} \sum_{i=1}^N X_i$$



X_i : individus appartenant à la population

N : nombre total d'individus

Propriété : elle est le centre de gravité d'une distribution

$$\sum_{i=1}^N (X_i - M_X) = 0$$

La moyenne est sensible aux valeurs extrêmes, elle décrit donc mal la tendance centrale d'une distribution dissymétrique, **au contraire de la médiane**

Notions de base

- Médiane :

Valeur telle que la moitié des valeurs lui est inférieure et l'autre moitié supérieure

Soient N individus :

- Classer d'abord les N valeurs des individus par ordre croissant

- 2 cas de figures :

- * N impair : la médiane est la valeur de la variable de rang $(N+1)/2$

- * N pair : la médiane correspond à la moyenne des valeurs de rangs $N/2$ et $(N+2)/2$

Exemple :

Série 1, 1, 4, 5, 12

N=5 (impair) donc médiane = 3^{ème} valeur de la série : **4**

Série 1, 1, 4, 5

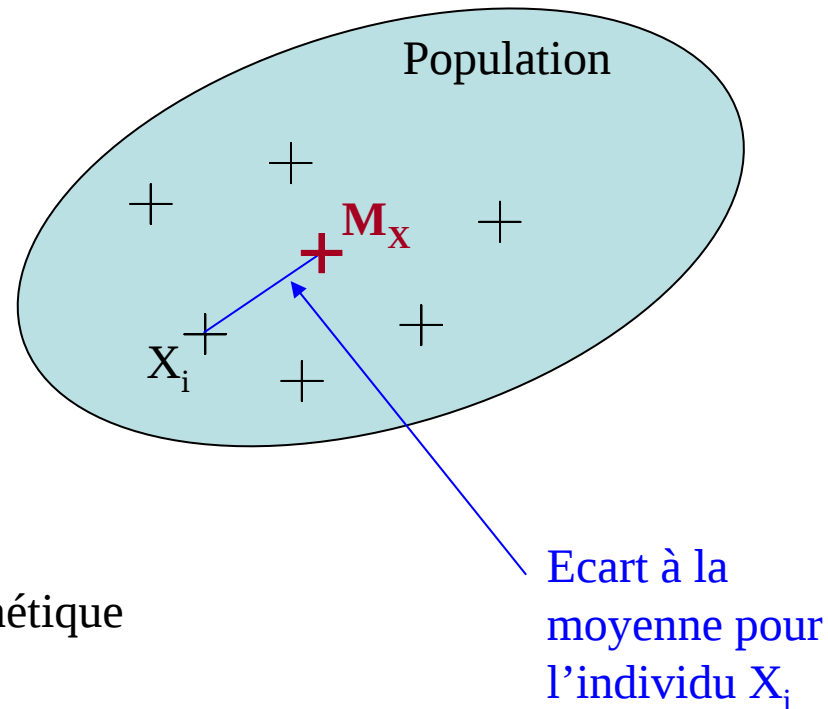
N=4 (pair) donc médiane = entre la 2^{ème} et la 3^{ème} valeur : **2.5**

Notions de base

- Variance :

$$V = \frac{1}{N} \sum_{i=1}^N (X_i - M_X)^2$$

Il s'agit du carré de l'écart-type, qui représente l'écart moyen des valeurs à leur moyenne arithmétique



L'écart-type est noté σ :

$$\sigma = \sqrt{V}$$

UNE SEULE VARIABLE

Notions de base

- Covariance :

$$COV = \frac{1}{N} \sum_{i=1}^N (X_i - M_X)(Y_i - M_Y)$$



Permet d'étudier les variations simultanées de **DEUX VARIABLES** quantitatives par rapport à leur moyenne respective

X_i : individus de la variable X

Y_i : individus de la variable Y

M_X : moyenne des individus de la variable X

M_Y : moyenne des individus de la variable Y

N : nombre d'individus

COV >>0 : les 2 variables varient dans le même sens

COV = 0 : pas de relation linéaire entre les 2 variables

COV <<0 : les 2 variables varient en sens inverse

Chapitre I

I.1 Démarche générale

I.2 Relation entre deux caractères qualitatifs :
Tableau de contingence & Test du khi-deux

I.3 Relation entre un caractère qualitatif et un caractère quantitatif :
Analyse de variance

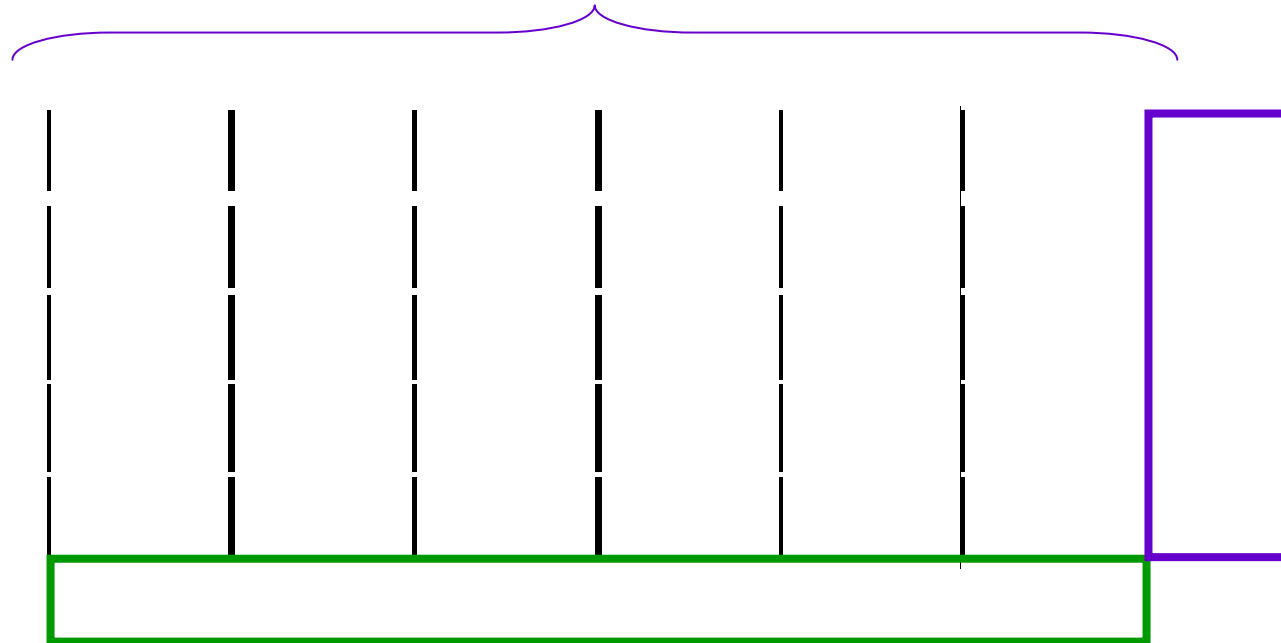
I.4 Relation entre deux caractères quantitatifs :
Corrélations et régressions linéaires simples

I.5 Relation entre plus de deux caractères quantitatifs :
Corrélations et régressions linéaires multiples

Tableau de contingence

j colonnes = **j** modalités de la variable B

i lignes
= ***i*** modalités de la variable A



La dernière **ligne** correspond aux **effectifs marginaux** de chaque colonne =>

La dernière **colonne** correspond aux effectifs marginaux de chaque ligne =>

Répartition des effectifs pour 1 caractère, sans tenir compte du second

Propriétés d'un tableau de contingence

(1) **Symétrie**

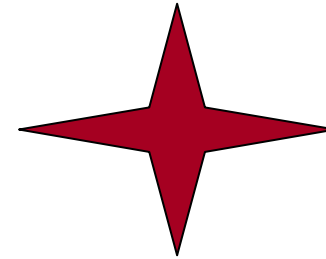
(2) Ne contient **pas de nbres < 0**

(3) Les sommes en ligne et en colonne ont une **signification**

Ex. Les sommes en ligne composent une dernière colonne qui représente la distribution du caractère Y

3 cas de figures

(1) Cas d'une liaison **parfaite**



(2) Cas d'une **absence** de liaison

(3) Cas ni d'une liaison parfaite ni d'une
absence de liaison :

Cas réel

(1) Cas d'une liaison parfaite

Exemple

On étudie la relation entre les **précipitations totales annuelles réparties (variable B)** en **3 modalités** (400-600, 600-800, 800-1000 mm/an) et observées dans **80 stations** d'une région.

Ces stations ont **une exposition (variable A)** selon **4 modalités** (N, S, E ou W).

3 colonnes = 3 modalités de la variable B

		400-600	600-800	800-1000	
4 lignes = 4 modalités de la variable A	N	20	0	0	
	E	0	18	0	
	S	0	0	23	
	O	0	0	19	

(1) Cas d'une liaison parfaite

On **regroupe** les **stations et les modalités** ayant des caractéristiques **analogues** (S et O)

3 colonnes = 3 modalités de la variable B

		400-600	600-800	800-1000	
3 lignes = 3 modalités de la variable A	N	20	0	0	20
	E	0	18	0	18
	S et O	0	0	42	42
		20	18	42	80

(1) Cas d'une liaison parfaite

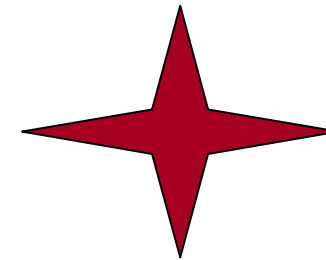
- La liaison est parfaite car on obtient (en **permutant** ou en **regroupant** certaines modalités) un tableau de contingence contenant des **nombre positifs sur la diagonale** et des **zéros ailleurs**.
- En outre, **chaque modalité** de la **variable A** correspond à **une unique modalité** de la **variable B** et réciiproquement.

Cls : dans notre cas d'études, la pluviométrie annuelle dépend **complètement** de l'exposition de la station météorologique

3 cas de figures

(1) Cas d'une liaison **parfaite**

(2) Cas d'une **absence** de liaison



(3) Cas ni d'une liaison parfaite ni d'une absence de liaison :

Cas réel

(2) Cas d'une absence de liaison

**Pas de liaison
apparente ici ...**

Dans ce cas, il faut
établir le tableau
des **profils-lignes** et
le tableau des
profils-colonnes

A B	400-600	600-800	800-1000	
N	8	20	12	40
E	4	10	6	20
S	2	5	3	10
O	2	5	3	10
	16	40	24	80

(2) Cas d'une absence de liaison

Tableau des profils-lignes : chaque élément d'une ligne est divisé par le total de la ligne, le tout multiplié par 100

	A B	400-600	600-800	800-1000	
40	N	$8/40 \cdot 100$	$20/40 \cdot 100$	$12/40 \cdot 100$	
20	E	$4/20 \cdot 100$	$10/20 \cdot 100$	$6/20 \cdot 100$	
10	S	$2/10 \cdot 100$	$5/10 \cdot 100$	$3/10 \cdot 100$	
10	O	$2/10 \cdot 100$	$5/10 \cdot 100$	$3/10 \cdot 100$	
80					

(2) Cas d'une absence de liaison

Résultat : chaque ligne est identique et le total est de 100%

A B	400-600	600-800	800-1000	
N	20%	50%	30%	100%
E	20%	50%	30%	100%
S	20%	50%	30%	100%
O	20%	50%	30%	100%

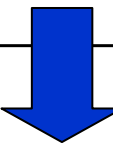
(2) Cas d'une absence de liaison

Il faut maintenant
établir le tableau
des **profils-**
colonnes

A B	400-600	600-800	800-1000	
N	8	20	12	40
E	4	10	6	20
S	2	5	3	10
O	2	5	3	10
	16	40	24	80

(2) Cas d'une absence de liaison

Il faut maintenant
établir le tableau
des **profils-
colonnes**

	16	40	24	80
				
A B	400-600	600-800	800-1000	
N	$8/16 * 100$	$20/40 * 100$	$12/24 * 100$	
E	$4/16 * 100$	$10/40 * 100$	$6/24 * 100$	
S	$2/16 * 100$	$5/40 * 100$	$3/24 * 100$	
O	$2/16 * 100$	$5/40 * 100$	$3/24 * 100$	

(2) Cas d'une absence de liaison

Résultat : chaque colonne est identique et le total est de 100%

A B	400-600	600-800	800-1000	
N	50%	50%	50%	
E	25%	25%	25%	
S	12,5%	12,5%	12,5%	
O	12,5%	12,5%	12,5%	
	100%	100%	100%	

(1) Cas d'une absence de liaison

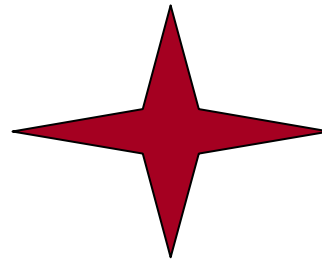
- Il n'existe aucune liaison entre les 2 variables qualitatives car on obtient un **tableau de profil-lignes** tel que toutes les lignes sont identiques et le total est égal à 100% ainsi qu'un **tableau de profil-colonnes** tel que toutes les colonnes sont identiques et le total est égal à 100%
- Ces calculs sont **indispensables** avant de se lancer dans des calculs plus complexes à réaliser dans le cas général...
- Le calcul de **l'un des deux tableaux** est suffisant :
 - Si toutes les lignes (ou colonnes) sont identiques et =100%, on conclut qu'il n'y a pas de liaison entre les 2 variables
 - Si ce n'est pas le cas, il faut se lancer dans des calculs plus complexes

Cls : dans notre cas d'études, la pluviométrie annuelle ne dépend **pas du tout de l'exposition de la station météorologique**

3 cas de figures

- (1) Cas d'une liaison **parfaite**
- (2) Cas d'une **absence** de liaison
- (3) Cas ni d'une liaison parfaite ni d'une absence de liaison :

Cas réel



(3) Cas réel

- Dans la réalité, les cas précédents existent **rarement**
- Pour savoir s'il **existe une liaison** et si celle-ci est **significative**,

On utilise le test du Khi-2

(3) Cas réel

DEMARCHE :

ETAPE 0. On pose l'hypothèse H0 d'indépendance entre les variables

ETAPE 1. Calcul d'un **tableau théorique**

A partir du tableau de contingence (=tableau réel)

- Ce tableau correspond à une indépendance entre les 2 caractères
- Les individus y sont répartis proportionnellement entre chaque modalité

ETAPE 2. Calcul d'un **tableau des écarts à l'indépendance**

A partir du tableau de contingence et du tableau théorique :

Tableau contingence – Tableau théorique

ETAPE 3. Test de **significativité** de la relation :

Calcul du **Khi-Deux** (χ^2) **réel**

Calcul du **Khi-Deux** (χ^2) **théorique**

Comparaison entre les 2 valeurs de Khi-Deux (χ^2)

(3) ETAPE 1

Soit le tableau suivant :

A B	400-600	600-800	800-1000	Total marginal colonne
N	7	10	3	20
E	8	8	3	19
S	6	9	9	24
W	0	5	13	18
Total marginal ligne	21	32	28	81

(3) ETAPE 1

Le tableau **théorique** se calcule comme suit :

$$n_{ij} = (n_i * n_j) / n$$

avec :

- n_{ij} : effectif théorique de la ligne i et de la colonne j
- n_i : effectif marginal de la ligne i
- n_j : effectif marginal de la colonne j
- n : effectif total

Exercice : Dresser le tableau **théorique** correspondant au tableau précédent

(3) ETAPE 1

Résultat :

A B	400-600	600-800	800-1000	Total marginal colonne
N	$(20 \times 21) / 81 = 5.2$	7.9	6.9	20
E	$(19 \times 21) / 81 = 4.9$	7.5	6.6	19
S	$(24 \times 21) / 81 = 6.2$	9.5	8.3	24
W	$(18 \times 21) / 81 = 4.7$	7.1	6.2	18
Total marginal ligne	21	32	28	81

(3) Cas réel

DEMARCHE :

ETAPE 0. On pose l'hypothèse H0 d'indépendance entre les variables

ETAPE 1. Calcul d'un **tableau théorique**

A partir du tableau de contingence (=tableau réel)

- Ce tableau correspond à une indépendance entre les 2 caractères
- Les individus y sont répartis proportionnellement entre chaque modalité

ETAPE 2. Calcul d'un **tableau des écarts à l'indépendance**

A partir du tableau de contingence et du tableau théorique :

Tableau contingence – Tableau théorique

ETAPE 3. Test de **significativité** de la relation :

Calcul du **Khi-Deux** (χ^2) **réel**

Calcul du **Khi-Deux** (χ^2) **théorique**

Comparaison entre les 2 valeurs de Khi-Deux (χ^2)

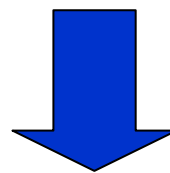
(3) ETAPE 2

Tableau de contingence (=tableau réel)

A B	400-600	600-800	800-1000	Total marginal colonne
N	7	10	3	20
E	8	8	3	19
S	6	9	9	24
W	0	5	13	18
Total marginal ligne	21	32	28	81

Tableau théorique

A B	400-600	600-800	800-1000	Total marginal colonne
N	5.2	7.9	6.9	20
E	4.9	7.5	6.6	19
S	6.2	9.5	8.3	24
W	4.7	7.1	6.2	18
Total marginal ligne	21	32	28	81



$7 - 5.2 = 1.8$	2.1	-3.9
$8 - 4.9 = 3.1$	0.5	-3.6
$6 - 6.2 = -0.2$	-0.5	0.7
$0 - 4.7 = -4.7$	-2.1	6.8

Il s'agit du **tableau des écarts à l'indépendance**

(3) ETAPE 2

	400-600	600-800	800-1000
N	$7 - 5.2 = 1.8$	2.1	-3.9
E	$8 - 4.9 = 3.1$	0.5	-3.6
S	$6 - 6.2 = -0.2$	-0.5	0.7
O	$0 - 4.7 = -4.7$	-2.1	6.8

Interprétation du tableau des écarts à l'indépendance :

- **Valeurs > 0**

Effectifs réels > Effectifs théoriques => **association positive**

Ex: le nombre de stations exposées Ouest appartenant à la classe des forts cumuls est > à celui qu'il devrait être en cas d'indépendance

- **Valeurs < 0**

Effectifs réels < Effectifs théoriques => **association négative**

Ex: le nombre de stations exposées Ouest appartenant à la classe des faibles cumuls est < à celui qu'il devrait être en cas d'indépendance

(3) Cas réel

DEMARCHE :

ETAPE 0. On pose l'hypothèse H0 d'indépendance entre les variables

ETAPE 1. Calcul d'un **tableau théorique**

A partir du tableau de contingence (=tableau réel)

- Ce tableau correspond à une indépendance entre les 2 caractères
- Les individus y sont répartis proportionnellement entre chaque modalité

ETAPE 2. Calcul d'un **tableau des écarts à l'indépendance**

A partir du tableau de contingence et du tableau théorique :

Tableau contingence – Tableau théorique

ETAPE 3. Test de **significativité** de la relation :

Calcul du **Khi-Deux** (χ^2) **réel**

Calcul du **Khi-Deux** (χ^2) **théorique**

Comparaison entre les 2 valeurs de Khi-Deux (χ^2)

(3) ETAPE 3

Pour tester la **significativité** de la relation entre 2 caractères, on calcule :

• χ^2 du tableau réel = χ^2 réel

Formule :

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^p \frac{(n_{ij} \text{obs} - n_{ij} \text{théo})^2}{n_{ij} \text{théo}}$$

Tableau des écarts à l'indépendance

Tableau théorique

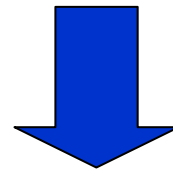
(3) ETAPE 3

Tableau des écarts à l'indépendance :

1.8	2.1	-3.9
3.1	0.5	-3.6
-0.2	-0.5	0.7
-4.7	-2.1	6.8

Tableau théorique

5.2	7.9	6.9
4.9	7.5	6.6
6.2	9.5	8.3
4.7	7.1	6.2



$(1.8)^2/5.2=0.62$	0.56	2.2
$(3.1)^2/4.9=1.96$	0.03	1.96
$(-0.2)^2/6.2=0.06$	0.03	0.06
$(-4.7)^2/4.7=4.7$	0.62	7.46

$$\chi^2_{\text{réel}} = 0.62 + 1.96 + 0.06 + \dots + 0.56 + 0.03 + 0.03 + \dots + 0.06 + 7.46 = 20.21$$

(3) ETAPE 3

Pour tester la **significativité** de la relation entre 2 caractères, on cherche également :

- χ^2 du tableau théorique : **χ^2 théorique**

Pour cela, il faut :

- **Calculer** le nombre de **degrés de liberté** →
- **Fixer** le niveau de **confiance** souhaité

Plus la **probabilité** est **faible**,
plus le niveau de confiance est élevé

P=0.1	→	S=90%
P=0.05	→	S=95%
P=0.01	→	S=99%

$$ddl = (k-1) * (p-1)$$

Modalités de la variables A

Modalités de la variables B

$k=4$

$p=3$

→ **ddl = 6**

1

Tableau 1 : Extrait de la table de la loi normale

Probabilité P d'avoir des valeurs comprises entre

-z et +z

	P (%)					
	68%	90%	95%	98%	99%	99.9%
z	1,0	1,65	1,96	2,33	2,57	3,3

2

Tableau 2 : Table en « t » de Student

Valeurs de t ayant la probabilité P d'être dépassées (ddl = nombre de degrés de liberté)

(test unilatéral)

P	0.05	0.01
ddl		
1	6.314	31.82
2	2.920	6.965
3	2.353	4.541
4	2.132	3.747
5	2.015	3.365
6	1.943	3.143
7	1.895	2.998
8	1.860	2.896
9	1.833	2.821
10	1.812	2.764
11	1.796	2.718
12	1.782	2.681
13	1.771	2.650
14	1.761	2.624
15	1.753	2.602
16	1.746	2.583
17	1.740	2.567
18	1.734	2.552
19	1.729	2.539
20	1.725	2.528
21	1.721	2.518
22	1.717	2.508
23	1.714	2.500
24	1.711	2.492
25	1.708	2.485
26	1.706	2.479
27	1.703	2.473
28	1.701	2.467
29	1.699	2.462
30	1.697	2.457
inf.	1.64	2.326

3

Tableau 3 : table du r de Bravais-Pearson

Probabilité P pour que le coefficient de corrélation égale ou dépasse, en valeur absolue, une valeur r, en fonction du nombre de degrés de liberté (d.d.l.).

d.d.l. P	0,10	0,05	0,01
1	0,9877	0,9969	0,9999
2	0,9000	0,9500	0,9900
3	0,8054	0,8783	0,9587
4	0,7293	0,8114	0,9172
5	0,6694	0,7545	0,8745
6	0,6215	0,7067	0,8343
7	0,5822	0,6664	0,7977
8	0,5494	0,6319	0,7646
9	0,5214	0,6021	0,7348
10	0,4973	0,5760	0,7079
11	0,4762	0,5529	0,6835
12	0,4575	0,5324	0,6614
13	0,4409	0,5139	0,6411
14	0,4259	0,4973	0,6226
15	0,4124	0,4821	0,6055
16	0,4000	0,4683	0,5897
17	0,3887	0,4555	0,5751
18	0,3783	0,4438	0,5614
19	0,3687	0,4329	0,5487
20	0,3598	0,4227	0,5368
25	0,3233	0,3809	0,4869
30	0,2960	0,3494	0,4487
35	0,2746	0,3246	0,4182
40	0,2573	0,3044	0,3932
45	0,2438	0,2875	0,3721
50	0,2306	0,2732	0,3541
60	0,2108	0,2500	0,3248
70	0,1954	0,2319	0,3017
80	0,1829	0,2172	0,2830
90	0,1726	0,2050	0,2673
100	0,1638	0,1946	0,2540

4

Tableau 4 : Table du Khi2

Valeurs de khi-deux ayant la probabilité P d'être dépassées, en fonction du nombre de degrés de liberté (d.d.l.).

d.d.l. P	0,10	0,05	0,01
1	2,706	3,841	12,101
2	4,605	5,991	18,475
3	6,251	7,879	23,685
4	7,779	9,488	28,755
5	9,236	11,070	33,567
6	10,645	12,592	38,566
7	12,017	14,067	43,773
8	13,362	15,507	49,002
9	14,691	16,919	54,156
10	15,987	18,307	59,342
11	17,275	19,675	64,558
12	18,549	21,026	69,658
13	19,812	22,362	74,778
14	21,064	23,685	79,901
15	22,307	24,996	85,029
16	23,542	26,296	90,179
17	24,769	27,587	95,354
18	25,989	28,869	100,541
19	27,204	30,144	105,756
20	28,412	31,410	110,991
21	29,615	32,671	116,256
22	30,813	33,924	121,541
23	32,007	35,172	126,845
24	33,196	36,415	132,160
25	34,382	37,652	137,484
26	35,563	38,885	142,816
27	36,741	40,113	148,156
28	37,916	41,337	153,503
29	39,087	42,557	158,857
30	40,256	43,773	164,226

Niveau de confiance

Nombre de degrés de liberté

P=0,05

ddl=6

=> χ^2 théorique =12,592

5

Tableau 5 : Table du F de Fisher

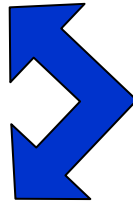
P=0.05

	ddl 1				
Ddl 2	1	2	3	4	5
1	161.44	199.5	215.7	224.58	230.16
2	18.51	19.00	19.16	19.25	19.29
3	10.13	10.24	10.28	10.31	10.33
4	7.7086	6.9443	6.5914	6.3882	6.2561
5	6.6079	5.7861	5.4095	5.1922	5.0503
6	5.9874	5.1433	4.7571	4.5337	4.3874
7	5.5914	4.7374	4.3468	4.1203	3.9715
8	5.3177	4.4590	4.0662	3.8379	3.6875
9	5.116	4.257	3.863	3.631	3.4817
10	4.966	4.107	3.713	3.480	3.3258
11	4.841	4.000	3.606	3.373	3.2039
12	4.7472	3.8853	3.4903	3.2592	3.1059
13	4.6672	3.8056	3.4105	3.1791	3.0254
14	4.6001	3.7389	3.3439	3.1122	2.9582
15	4.5431	3.6823	3.2874	3.0556	2.9013
16	4.4940	3.6337	3.2389	3.0069	2.8524
17	4.4513	3.5915	3.1968	2.9647	2.8100
18	4.4139	3.5546	3.1599	2.9277	2.7729
19	4.3807	3.5219	3.1274	2.8951	2.7401
20	4.3512	3.4928	3.0984	2.8661	2.7109
21	4.3248	3.4668	3.0725	2.8401	2.6848
22	4.3009	3.4434	3.0491	2.8167	2.6613
23	4.2793	3.4221	3.0280	2.7955	2.6400
24	4.2597	3.4028	3.0088	2.7763	2.6207
25	4.2417	3.3852	2.9912	2.7587	2.6030
26	4.2252	3.3690	2.9752	2.7426	2.5868
27	4.2100	3.3541	2.9604	2.7278	2.5719
28	4.1960	3.3404	2.9467	2.7141	2.5581
29	4.1830	3.3277	2.9340	2.7014	2.5454
30	4.1709	3.3158	2.9223	2.6896	2.5336
40	4.0847	3.2317	2.8387	2.6060	2.4495
60	4.0012	3.1504	2.7581	2.5252	2.3683
120	3.9201	3.0718	2.6802	2.4472	2.2899
inf	3.8415	2.9957	2.6049	2.3719	2.2141

(3) ETAPE 3

$$\chi^2_{\text{réel}} = 20.21$$

$$\chi^2_{\text{théorique}} = 12.59$$



Comparaison des deux valeurs pour tester la significativité de la relation statistique

Donc, pour une probabilité de 0,05, et dans ce cas précis, ce serait vérifié quelle que soit la probabilité choisie, nous avons $\chi^2_{\text{réel}} > \chi^2_{\text{théorique}}$

Interprétation :

- $\chi^2_{\text{réel}} > \chi^2_{\text{théorique}}$

La liaison statistique entre les 2 variables est **significative**

L'hypothèse H0 d'indépendance posée initialement est rejetée.

Les variables sont dépendantes l'une de l'autre.

- $\chi^2_{\text{réel}} < \chi^2_{\text{théorique}}$

La liaison statistique entre les 2 variables n'est **pas significative**

L'hypothèse H0 d'indépendance est validée.

Les variables ne sont pas dépendantes l'une de l'autre.

Interprétation

- (1) Dans notre cas la **cumul annuel** de précipitations dépend significativement (à un seuil de 99%) de l'**exposition** de la station météorologique

A B	400-600	600-800	800-1000
N	1,8	2,1	-3,9
E	3,1	0,5	-3,6
S	-0,2	-0,5	0,7
O	-4,7	-2,1	6,8

Tableau des écarts à l'indépendance

- (2) **Association positive** entre expositions N et E et faibles précipitations

- (4) Dans la région étudiée les stations **exposées à l'O**, et dans une moindre mesure au S, **sont plus arrosées** que celles qui font face au N et à l'E.

- (3) **Association négative** entre ces mêmes expositions et la classe de précipitations supérieure

Résumé

Tableau de contingence = Tableau réel

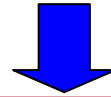


Tableau « théorique » $n_{ij} = (n_i * n_j) / n$

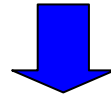
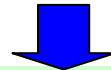
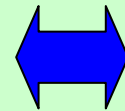


Tableau des écarts à l'indépendance $n_{ij} = (n_{ij}^{obs} - n_{ij}^{théo})$



χ^2 réel : $\chi^2 = \sum_{i=1}^k \sum_{j=1}^p \frac{(n_{ij}^{obs} - n_{ij}^{théo})^2}{n_{ij}^{théo}}$



χ^2 théorique : Lu dans le tableau des χ^2 à partir de ddl et P

Si χ^2 réel > χ^2 théorique,
Lien statistique
entre les 2 variables
H0 rejetée

Si χ^2 réel < χ^2 théorique,
Aucun lien statistique
entre les 2 variables
H0 validée

Interprétation plus fine