

M1 Management du sport 



ANALYSE STATISTIQUE

Présentation

- Objectifs de l'enseignement :
 - Connaître les bases de l'analyse statistiques
 - *Comprendre leur intérêt dans secteur du management du sport*
 - *Mettre en application les principaux tests statistiques*
- Déroulement de l'enseignement :
 - 18h TD – 9 séances
 - CT : séance 9

PLAN DU COURS

- Rappels
- Collecte de données
- Transformation et traitement des données
- Statistiques descriptives
- Statistiques inférentielles
- Tests d'hypothèses
- Mises en application diverses

INTRODUCTION ET RAPPELS

A quoi sert la statistique ?

Qu'est-ce que la statistique ?

« Méthode scientifique consistant à réunir des données sur des ensembles importants (population, échantillon) puis à décrire, analyser, commenter et critiquer ces données »

Quelle est son utilité ?

- Rendre lisible un flux important de données
- Quantifier le monde pour mieux le comprendre

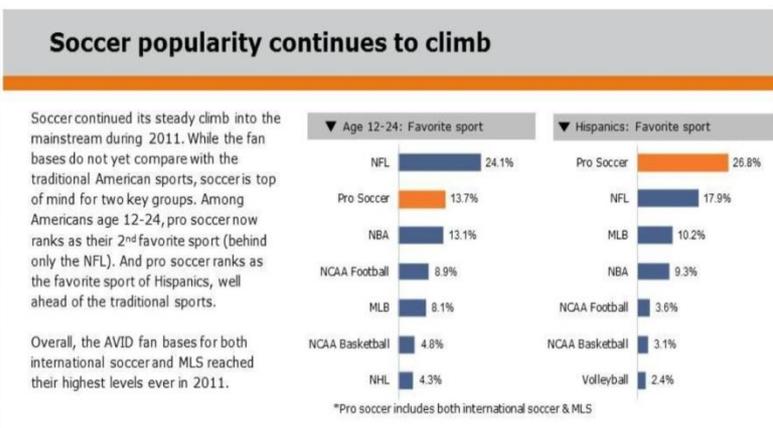
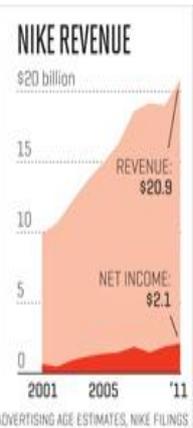
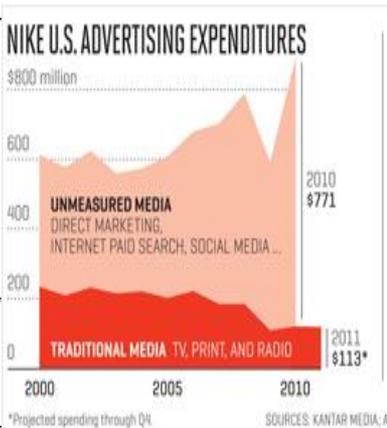
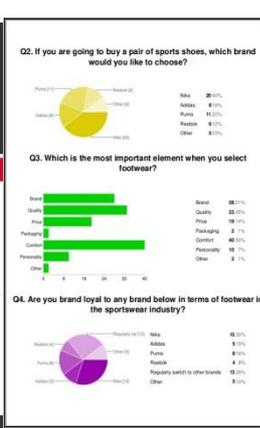
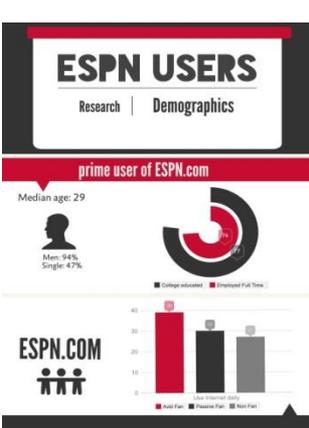
Est-ce une discipline mathématique (comprenez « pour les matheux » ?)

- Les outils statistiques ont été développés par des mathématiciens
- Mais leur utilisation ne nécessite pas forcément d'en comprendre les soubassements mathématiques
- Exemple : la variance = on passe par le carré des écarts à la moyennes.
-> On peut chercher à savoir pourquoi mais pas nécessaire

Intérêt de la statistique en management / marketing du sport

Quel intérêt de la statistique dans notre secteur ?

- ❑ Essentielle en marketing / stratégies privées
 - ✓ Etudes de marché
 - ✓ Comportement du consommateur
 - ✓ Etc...
- ❑ Essentielle en management / politiques publiques
 - ✓ Financement du sport
 - ✓ Etc...
- ❑ Importance renforcée en raison du big data
 - ✓ Profils recherchés : analyse de données (réseaux sociaux, objets connectés, etc...)



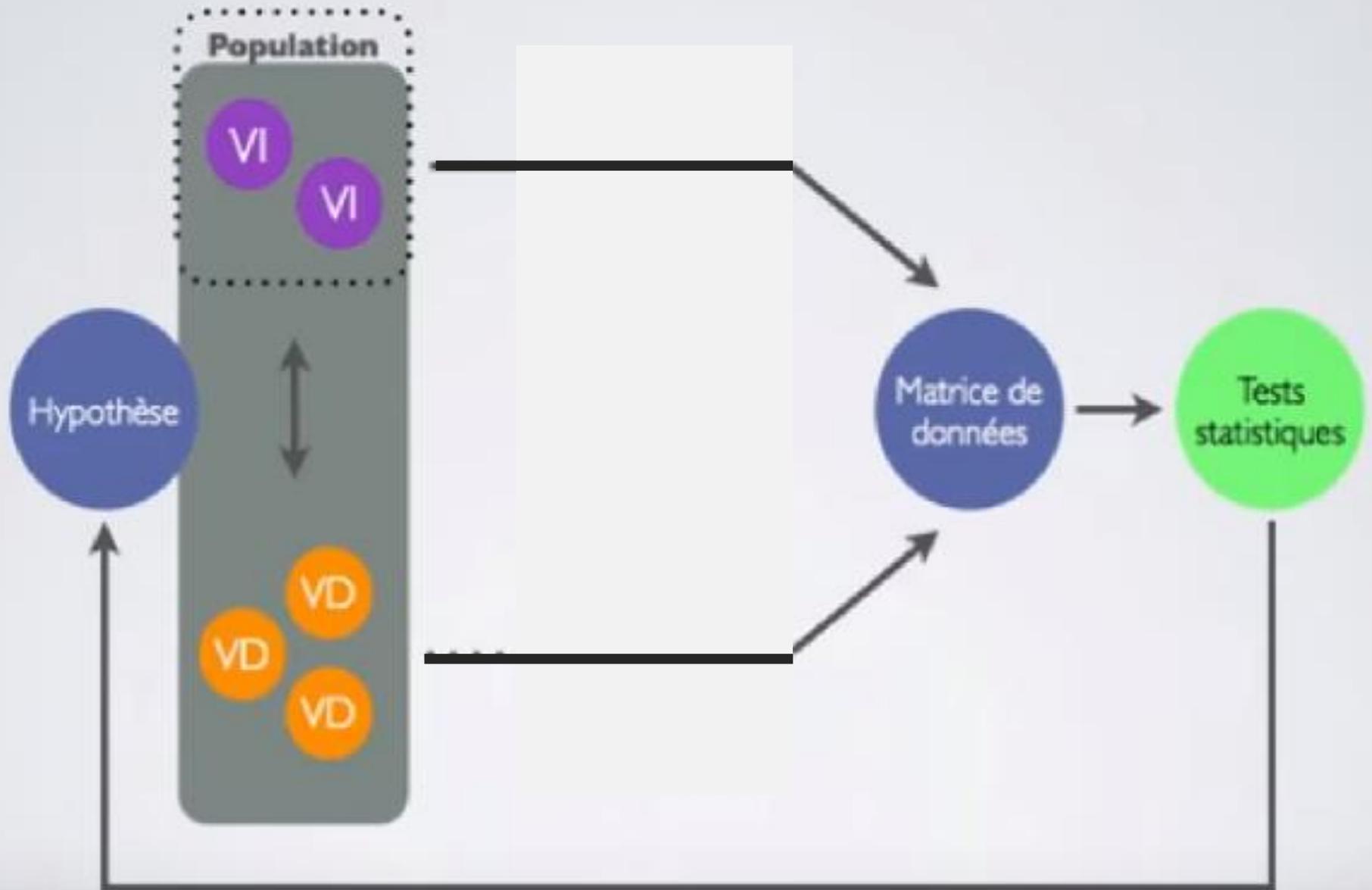
LES VARIABLES

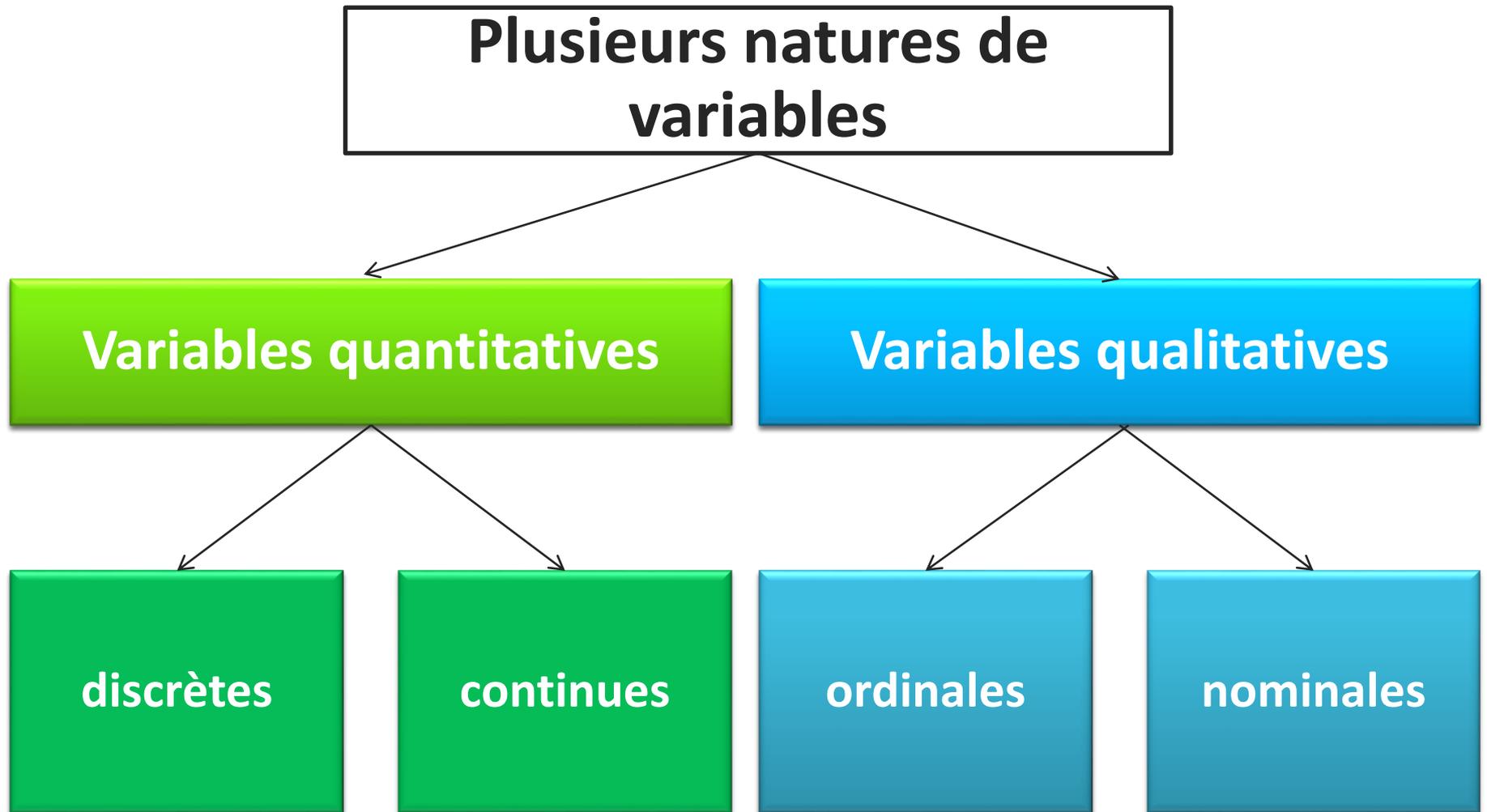
Qu'est-ce qu'une variable ?

« Propriété d'un individu, d'un objet ou d'un phénomène pouvant prendre différentes valeurs »

Faire de la statistique consiste à mesurer des variables (ou propriétés) au sein d'une population (d'individus, d'objets, de phénomènes), puis à les analyser, à les comparer, etc... afin d'en tirer des informations ou des connaissances (que l'on cherche souvent à généraliser à la population en question).

Les variables





Les variables

Variables quantitatives

= associées à un caractère mesurable

continues

= qui peuvent prendre toutes les valeurs dans un intervalle donné

discrètes

= qui prennent un nombre limité de valeurs entières

Variables qualitatives

= associées à un caractère non mesurable

ordinales

= dont les modalités peuvent être classées selon un ordre

nominales

= dont les modalités ne peuvent pas être classées selon un ordre

Les variables

Matrice de données = ensemble des variables mesurées

	Prix du billet	Âge	Satisfaction	Genre
Individu 1	30 €	23 ans	Plutôt satisfait-e	Femme
Individu 2	30 €	34 ans	Plutôt insatisfait-e	Homme
Individu 3	40 €	16 ans	Très insatisfait-e	Femme
Individu 4	40 €	25 ans	Satisfait-e	Homme
Individu 5	50 €	41 ans	Très satisfait-e	Homme

- Chaque **colonne** représente une **variable** de la matrice
- Chaque **ligne** représente une **unité de l'échantillon** (1 individu, une mesure relative à tel individu dans le cas de mesures répétées, etc...)

Les variables

Matrice de données = ensemble des variables mesurées

	Prix du billet	Âge	Satisfaction	Genre
Individu 1	30 €	23 ans	Plutôt satisfait-e	Femme
Individu 2	30 €	34 ans	Plutôt insatisfait-e	Homme
Individu 3	40 €	16 ans	Très insatisfait-e	Femme
Individu 4	40 €	25 ans	Satisfait-e	Homme
Individu 5	50 €	41 ans	Très satisfait-e	Homme

Quantitative discrète

Quantitative continue

Qualitative ordinale

Qualitative nominale

Les variables

Matrice de données = ensemble des variables mesurées

	Prix du billet	Âge	Satisfaction	Genre
Individu 1	score	score	score	score
Individu 2	score	score	score	score
Individu 3	score	score	score	score
Individu 4	score	score	score	score
Individu 5	score	score	score	score

- Chaque **observation** s'appelle un **score**
- Pour chaque variable, le **score** dépend de ses **modalités**

Les variables

Matrice de données = ensemble des variables mesurées

	Prix du billet	Âge	Satisfaction	Genre
Individu 1	score	score	score	score
Individu 2	score	score	score	score
Individu 3	score	score	score	score
Individu 4	score	score	score	score
Individu 5	score	score	score	score

- Les **tests statistiques** que l'on va pouvoir mener dépendent de la nature et des modalités de chaque variable

Modalité 1 = Homme
Modalité 2 = Femme
Variable à 2 modalités

Les variables

Matrice de données = ensemble des variables mesurées

	Prix du billet	Âge	Satisfaction	Genre
Individu 1	30	23	2	0
Individu 2	30	34	1	1
Individu 3	40	16	0	0
Individu 4	40	25	3	1
Individu 5	50	41	4	1

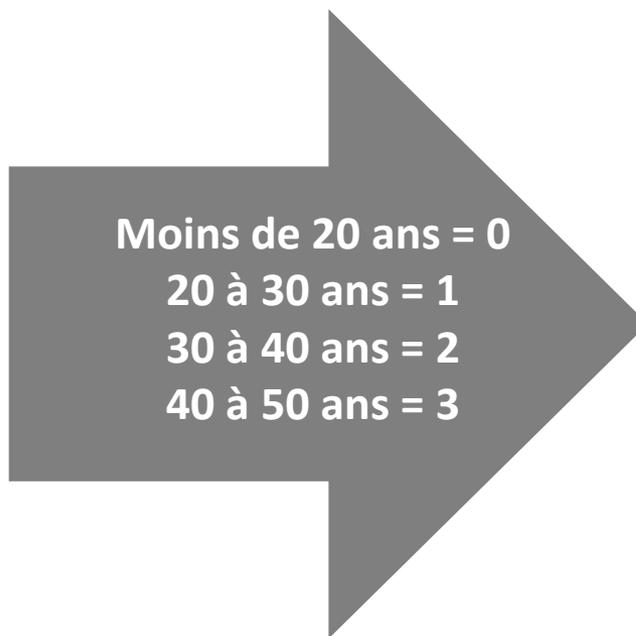
- Il faut y être très attentif étant donné qu'une vraie matrice ne comporte que des chiffres

Les variables

Matrice de données = ensemble des variables mesurées

Âge
23
34
16
25
41

Quantitative continue



- Il est également possible de changer la nature de la variable en fonction des besoins de l'analyse

Classe d'âge
1
2
0
1
3

Qualitative ordinale

LES STATISTIQUES DESCRIPTIVES

Qu'est-ce que la statistique descriptive ?

« résumer et présenter l'information contenue dans les données collectées »

- ❑ Il s'agit de décrire les variables collectées une par une
- ❑ Ce premier niveau d'analyse est indispensable (donc toujours présent)
- ❑ 4 niveaux de description :
 - ✓ Les tableaux et représentations graphiques
 - ✓ Les indices de tendance centrale
 - ✓ Les indices de dispersion
 - ✓ Les indices de forme

Les statistiques descriptives

Les tableaux de distribution :

Dépendent du type de variables

Présentent pour chaque modalité ou valeur prise par la variable :

- ❑ le nombre d'individus qui prennent cette modalité ou cette valeur, que l'on appelle **l'effectif**
- ❑ Parfois le pourcentage de cet effectif, que l'on appelle **la fréquence**

Les statistiques descriptives

Les tableaux de distribution : Effectifs (Nombre), Fréquence en %

	Nombre	%
Ensemble des équipements sportifs recensés (hors sites et espaces naturels relatifs aux sports de nature)	264 751	100 %
dont : Terrains de grands jeux (football, rugby, hockey,...)	44 217	16,7 %
Courts de tennis	41 410	15,6 %
Boulodromes	27 739	10,5 %
Plateaux d'Éducation Physique et Sportive	21 066	8,0 %
Salles multisports	17 657	6,7 %
Terrains extérieurs de petits jeux collectifs (basket-ball, beach volley, handball, volley-ball)	15 648	5,9 %
Salles non spécialisées	15 105	5,7 %
Salles ou terrains spécialisés	14 527	5,5 %
Équipements équestres	12 495	4,7 %
Équipements d'activités de forme et de santé	9 982	3,8 %
Équipements d'athlétisme	9 209	3,5 %
Salles de combat	6 989	2,6 %
Bassins de natation	6 308	2,4 %
Pas de tir	5 280	2,0 %
Skate park et vélo Freestyle	3 078	1,2 %
Murs et frontons	2 957	1,1 %
Structures artificielles d'escalade	2 665	1,0 %

Les statistiques descriptives

Les tableaux de distribution : Effectifs (Nombre), Fréquence en %

	Nombre	%
Ensemble des équipements sportifs recensés (hors sites et espaces naturels relatifs aux sports de nature)	264 751	100,0 %
dont : Terrains de grands jeux (football, rugby, hockey,...)	44 217	16,7 %
Courts de tennis	41 410	15,7 %
Boulodromes	27 739	10,5 %
Plateaux d'Éducation Physique et Sportive	21 066	8,0 %
Salles multisports	17 657	6,7 %
Terrains extérieurs de petits jeux collectifs (basket-ball, beach volley, handball, volley-ball)	15 648	5,9 %
Salles non spécialisées	15 105	5,7 %
Salles ou terrains spécialisés	14 527	5,5 %
Équipements équestres	12 495	4,7 %
Équipements d'activités de forme et de santé	9 982	3,8 %
Équipements d'athlétisme	9 209	3,5 %
Salles de combat	6 989	2,6 %
Bassins de natation	6 308	2,4 %
Pas de tir	5 280	2,0 %
Skate park et vélo Freestyle	3 078	1,2 %
Murs et frontons	2 957	1,1 %
Structures artificielles d'escalade	2 665	1,0 %

Parmi les **264 751** « équipements sportifs recensés », **44 217** sont des « terrains de grands jeux »

Les statistiques descriptives

Les tableaux de distribution : Effectifs (Nombre), Fréquence en %

	Nombre	%
Ensemble des équipements sportifs recensés (hors sites et espaces naturels relatifs aux sports de nature)	264 751	100 %
dont : Terrains de grands jeux (football, rugby, hockey,...)	44 217	16,7 %
Courts de tennis	41 410	15,6 %
Boulodromes	27 739	10,5 %
Plateaux d'Éducation Physique et Sportive	21 066	8 %
Salles multisports		
Terrains extérieurs de petits beach volley, handball, volley-		
Salles non spécialisées		
Salles ou terrains spécialisés		
Équipements équestres		
Équipements d'activités de fo		
Équipements d'athlétisme		
Salles de combat		
Bassins de natation	6 308	2,4 %
Pas de tir	5 280	2,0 %
Skate park et vélo Freestyle	3 078	1,2 %
Murs et frontons	2 957	1,1 %
Structures artificielles d'escalade	2 665	1,0 %

Les « terrains de grands jeux » représentent près de **17%** des « équipements sportifs recensés »

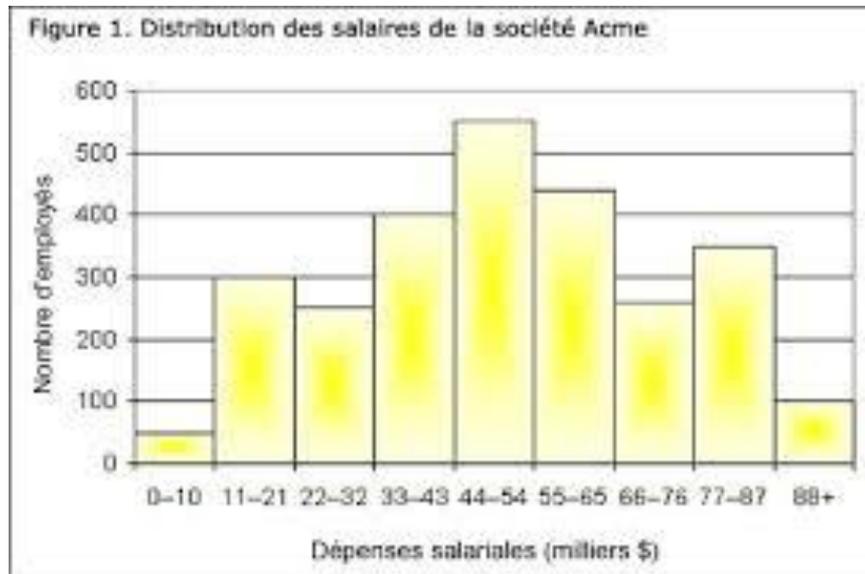
Les statistiques descriptives

Les représentations graphiques :

Dépendent du type de variables

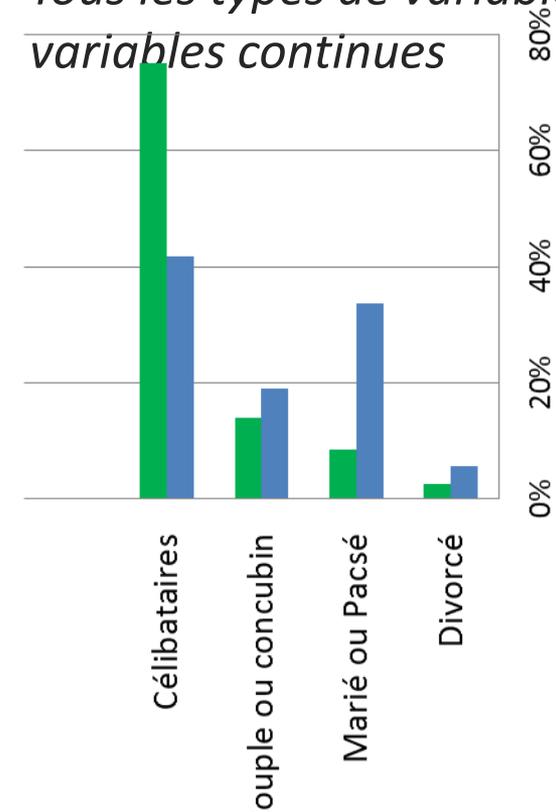
☐ L'histogramme

✓ Variables quantitatives continues
(classées)



☐ Le diagramme en bâtons

✓ Tous les types de variables hors
variables continues



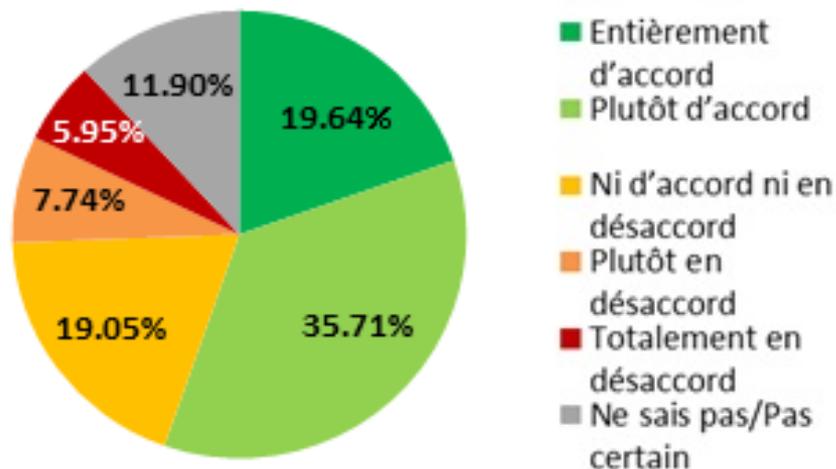
Les statistiques descriptives

Les représentations graphiques :

Dépendent du type de variables

☐ Le diagramme circulaire

✓ *Variables qualitatives avec relativement peu de modalités*



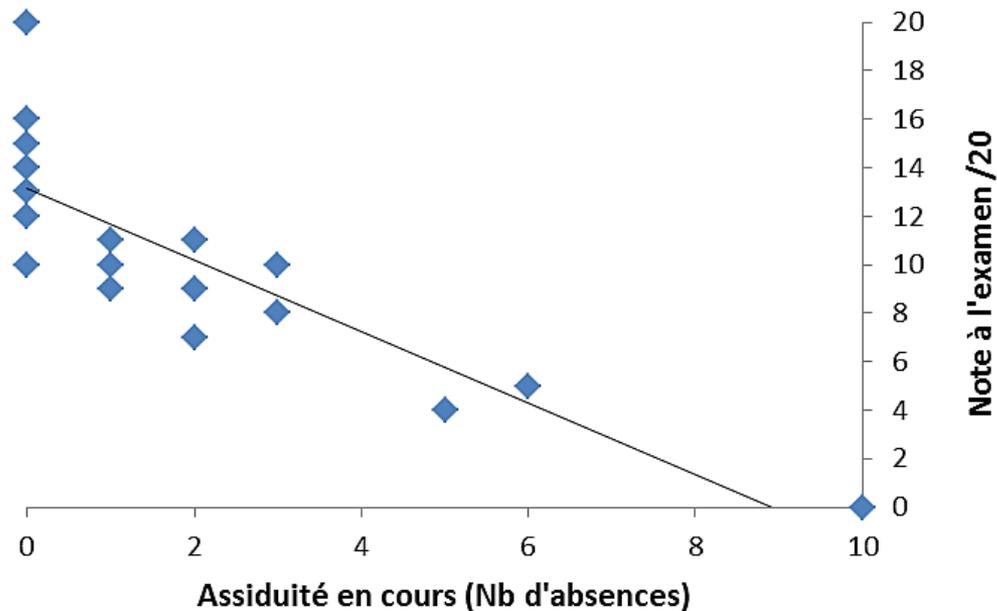
Les statistiques descriptives

Les représentations graphiques :

Dépendent du type de variables

☐ Le nuage de points

✓ *Variables quantitatives discrètes*



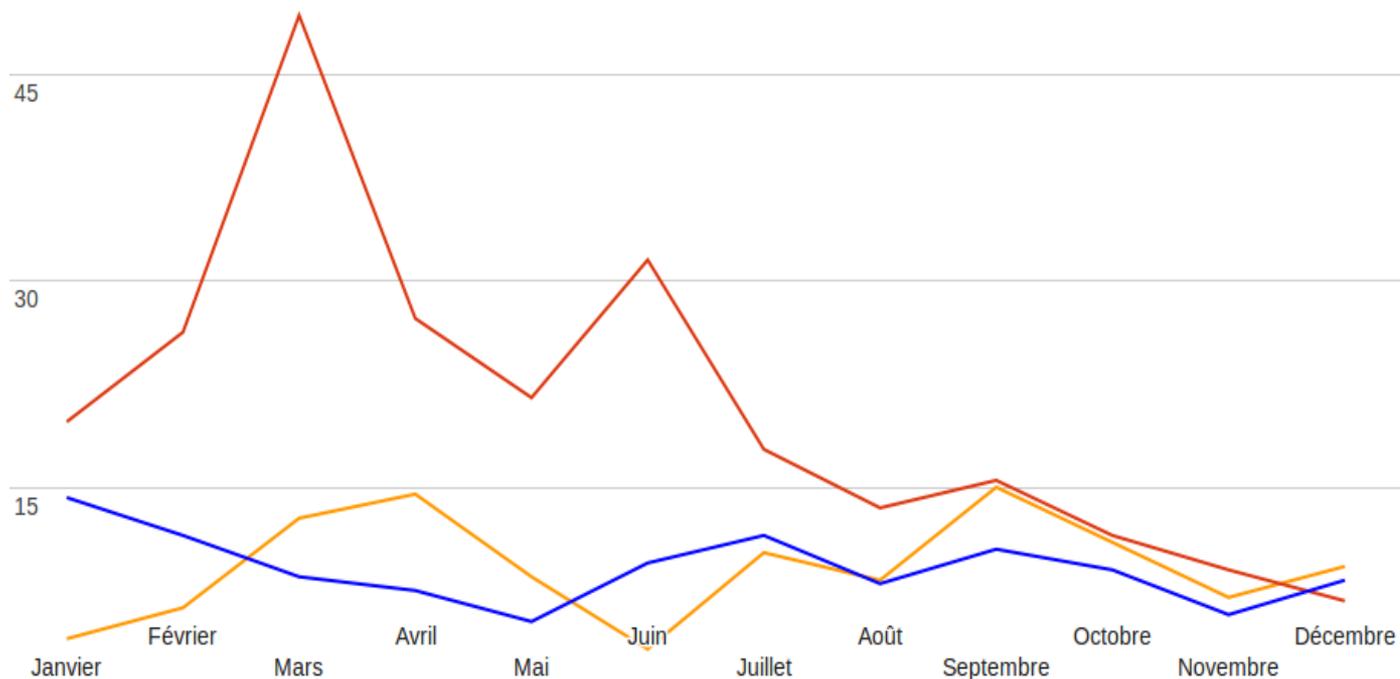
Les statistiques descriptives

Les représentations graphiques :

Dépendent du type de variables

☐ La courbe

✓ *Variables quantitatives continues*



Les statistiques descriptives

Les représentations graphiques :

Dépendent du type de variables

❑ Le diagramme figuratif

✓ *Surtout variables qualitatives nominales*



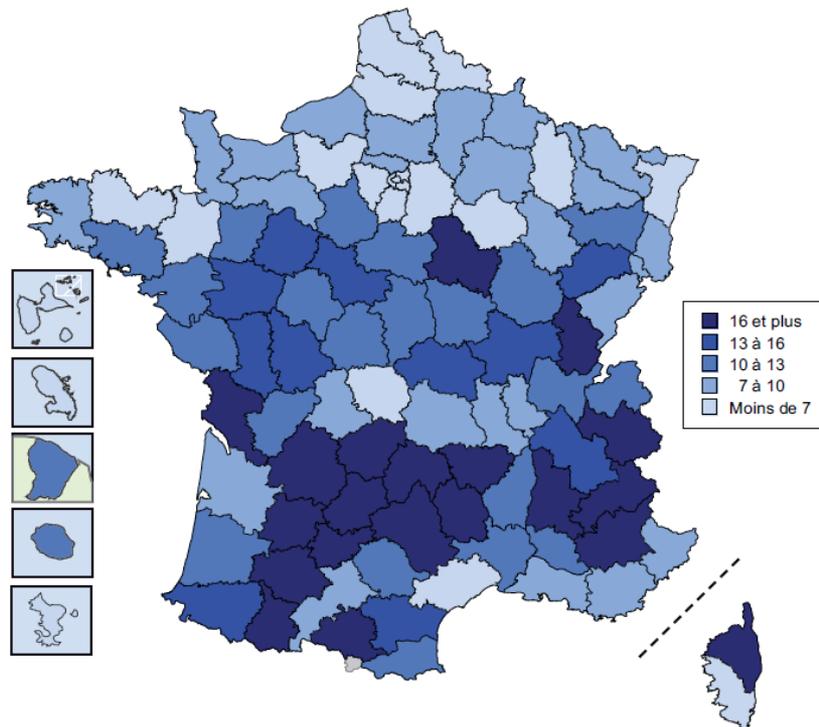
Les statistiques descriptives

Les représentations graphiques :

Dépendent du type de variables

☐ La représentation cartographique

✓ *Surtout variables qualitatives nominales ou ordinales*



Les statistiques descriptives

Les indicateurs de tendance centrale :

Décrivent « le cœur » de la distribution de différentes manières

□ La moyenne (arithmétique)

✓ *Somme des valeurs d'une série statistique divisée par le nombre de ses valeurs*

	Nb de billets achetés / saison
Individu 1	9
Individu 2	4
Individu 3	8
Individu 4	10
Individu 5	6
Individu 6	5
Individu 7	4
Individu 8	3
Individu 9	8

$$M = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots}{n}$$

x_i = valeurs que peut prendre la variable

n = effectif

$$M = \frac{9+4+8+10+5+6+4+3+8}{9} = 6,33$$

Les statistiques descriptives

Les indicateurs de tendance centrale :

Décrivent « le cœur » de la distribution de différentes manières

□ La médiane

- ✓ « Valeur du caractère qui partage une distribution en deux sous-ensembles de même effectif »

	Nb de billets achetés / saison
Individu 8	3
Individu 2	4
Individu 7	4
Individu 6	5
Individu 5	6
Individu 3	8
Individu 9	8
Individu 1	9
Individu 4	10
Individu 10	11

Suppose de reclasser les individus par score

Utilité : « revenu moyen » \neq « revenu médian »

50 % des
individus

50 % des
individus

Si dis
centr

Si distribution paire : milieu des deux
valeurs au centre de la distribution = 7

Les statistiques descriptives

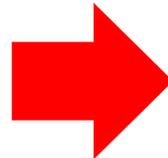
Les indicateurs de tendance centrale :

Décrivent « le cœur » de la distribution de différentes manières

☐ Le mode

✓ « Valeur qui correspond au plus grand effectif ».

	Nb de billets achetés / saison
Individu 1	9
Individu 2	4
Individu 3	8
Individu 4	10
Individu 5	6
Individu 6	5
Individu 7	4
Individu 8	3
Individu 9	8



Nb de billets achetés / saison	Effectifs
0 à 3 billets	1
4 à 6 billets	4
7 à 9 billets	3
+ de 9 billets	1

Les statistiques descriptives

Les indicateurs de dispersion :

Les indicateurs de tendance centrale ne suffisent pas et peuvent donner une fausse image des données

Ex : la moyenne

**Salaire moyen d'un pays =
1500 euros par mois**

```
graph TD; A[Salaire moyen d'un pays = 1500 euros par mois] --> B[Cas 1 : Tout le monde gagne entre 1400 et 1600 euros par mois = pays égalitaire]; A --> C[Cas 2 : Beaucoup de gens gagnent moins de 700 euros par mois. Quelques personnes gagnent + de 10 000 euros par mois = pays très inégalitaire (se voit via le salaire médian)]; A --> D[Cas 3 : Des effectifs relativement comparables gagnent entre 700 et 800 euros, entre 800 et 1200 euros, entre 1200 et 1700 euros, etc... (ne se voit pas via le salaire médian)];
```

❑ Cas 1 :

Tout le monde gagne entre 1400 et 1600 euros par mois = pays égalitaire

❑ Cas 2 :

Beaucoup de gens gagnent moins de 700 euros par mois. Quelques personnes gagnent + de 10 000 euros par mois = pays très inégalitaire (se voit via le salaire médian)

❑ Cas 3 :

Des effectifs relativement comparables gagnent entre 700 et 800 euros, entre 800 et 1200 euros, entre 1200 et 1700 euros, etc... (ne se voit pas via le salaire médian)

Les statistiques descriptives

Les indicateurs de dispersion :

Décrivent la manière dont les mesures et effectifs se répartissent

□ L'étendue et l'amplitude

- ✓ *Étendue = « Intervalle entre la plus grande et la plus petite des valeurs prises par la variable »*
- ✓ *Amplitude = « différence entre ces deux valeurs »*

	Nb de billets achetés / saison
Individu 1	9
Individu 2	4
Individu 3	8
Individu 4	10
Individu 5	6
Individu 6	5
Individu 7	4
Individu 8	3
Individu 9	8



Étendue = [3; 10]
Amplitude = 7

Les indicateurs de dispersion :

Décrivent la manière dont les mesures et effectifs se répartissent

□ La variance

- ✓ *« écart des scores par rapport à la moyenne »*
- ✓ *Pour des raisons mathématiques, on calcule la variance à partir du carré des écarts à la moyennes (« moyenne du carré des écarts à la moyenne »)*

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$$

Variance de l'échantillon

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

Variance de la population

Les indicateurs de dispersion :

Décrivent la manière dont les mesures et effectifs se répartissent

□ L'écart type

- ✓ « racine carrée de la variance »
- ✓ Permet essentiellement de lire la variance dans l'unité de valeur de départ (km, etc...)

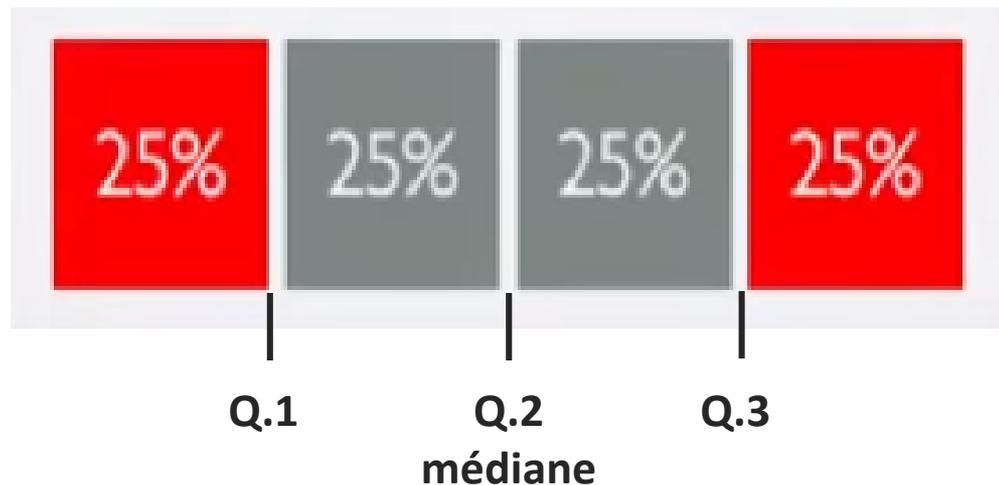
Les statistiques descriptives

Les indicateurs de dispersion :

Décrivent la manière dont les mesures et effectifs se répartissent

□ Les quartiles

- ✓ « Les 3 valeurs qui divisent les données en 4 parties égales, chacune représentant donc 1/4 de l'échantillon (ou 25%)
- ✓ Le 2^{ème} quartile équivaut ainsi à la médiane

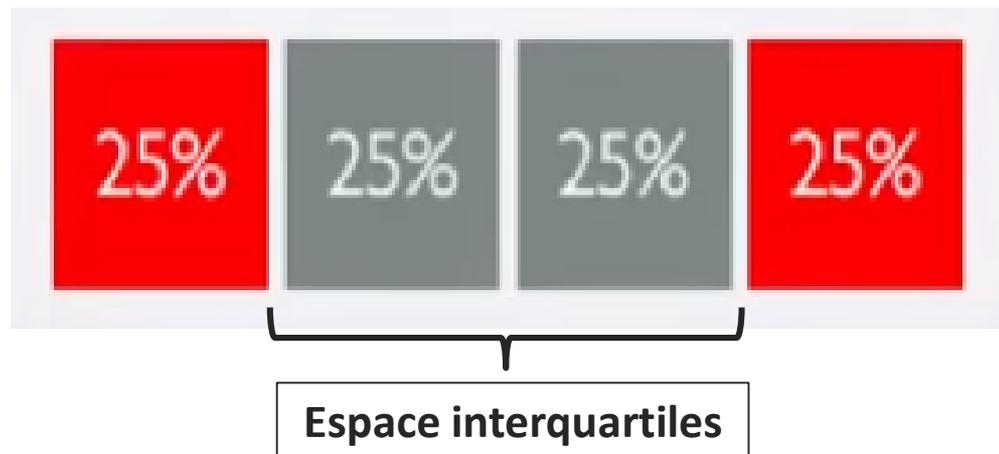


Les statistiques descriptives

Les indicateurs de dispersion :

Décrivent la manière dont les mesures et effectifs se répartissent

☐ Les quartiles

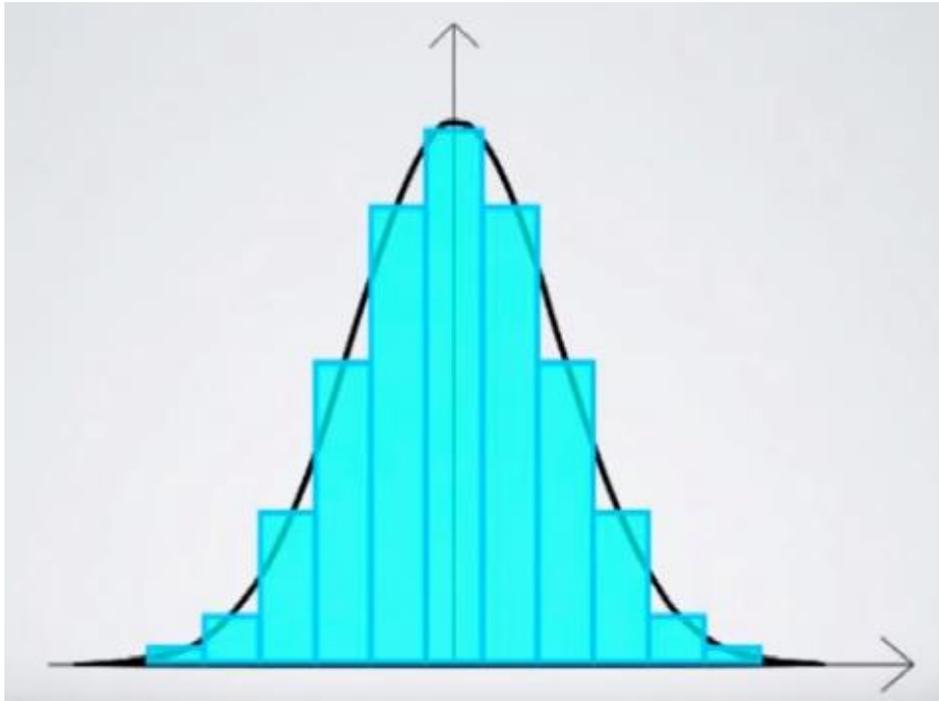


- ✓ *On utilise souvent l'espace interquartiles, à savoir l'espace qui contient les 50% de l'échantillon situés entre le 1^{er} et le 3^{ème} quartile (Q.3 – Q.1)*

Les statistiques descriptives

Les indicateurs de forme :

Revoient à la loi normale (que nous évoquerons plus en détail lors des tests statistiques)



- ✓ Unimodale (1 seul sommet)
- ✓ Symétrique
- ✓ Mesokurtique (ni trop plate, ni trop aigue)

Les statistiques descriptives

Les indicateurs de forme :

Renvoient à la loi normale (que nous évoquerons plus en détail lors des tests statistiques)



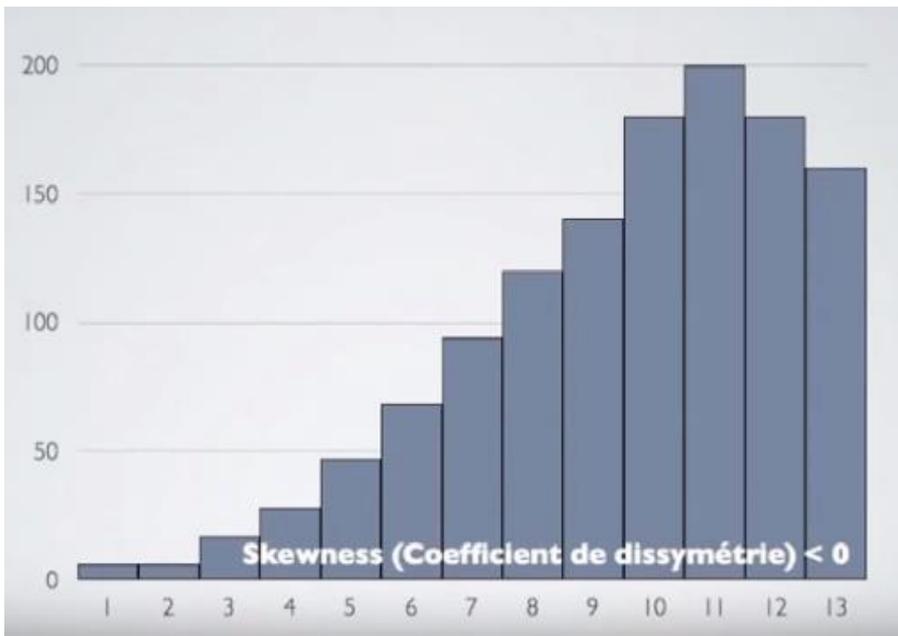
☐ Certaines distributions peuvent être :

- ✓ Bimodales ou multimodales (plusieurs sommets)
- ✓ Exemple : structure des revenus aux Etats-Unis (« suppression de la classe moyenne »)

Les statistiques descriptives

Les indicateurs de forme :

Revoient à la loi normale (que nous évoquerons plus en détail lors des tests statistiques)



☐ Certaines distributions peuvent être :

- ✓ Asymétriques négatives (negative skewness)
- ✓ Exemple : taux de mortalité

Les statistiques descriptives

Les indicateurs de forme :

Revoient à la loi normale (que nous évoquerons plus en détail lors des tests statistiques)



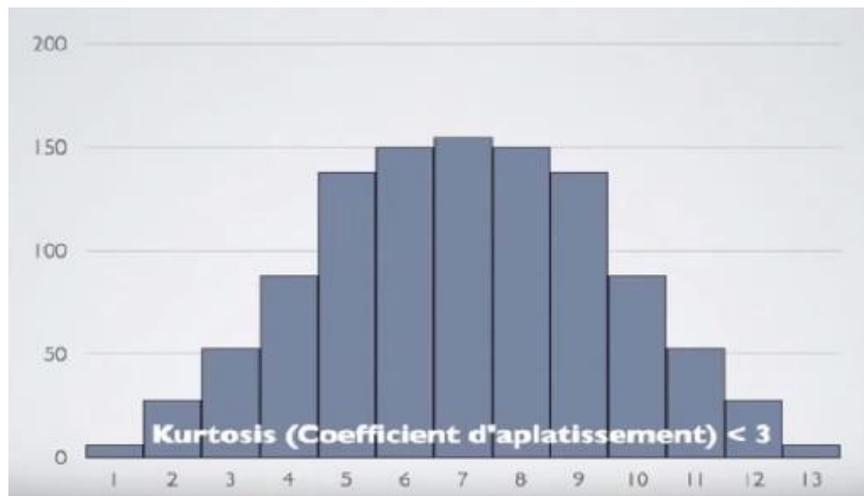
☐ Certaines distributions peuvent être :

- ✓ Asymétriques positives (positive skewness)
- ✓ Exemple : Âge de la maîtrise de la langue maternelle

Les statistiques descriptives

Les indicateurs de forme :

Revoient à la loi normale (que nous évoquerons plus en détail lors des tests statistiques)



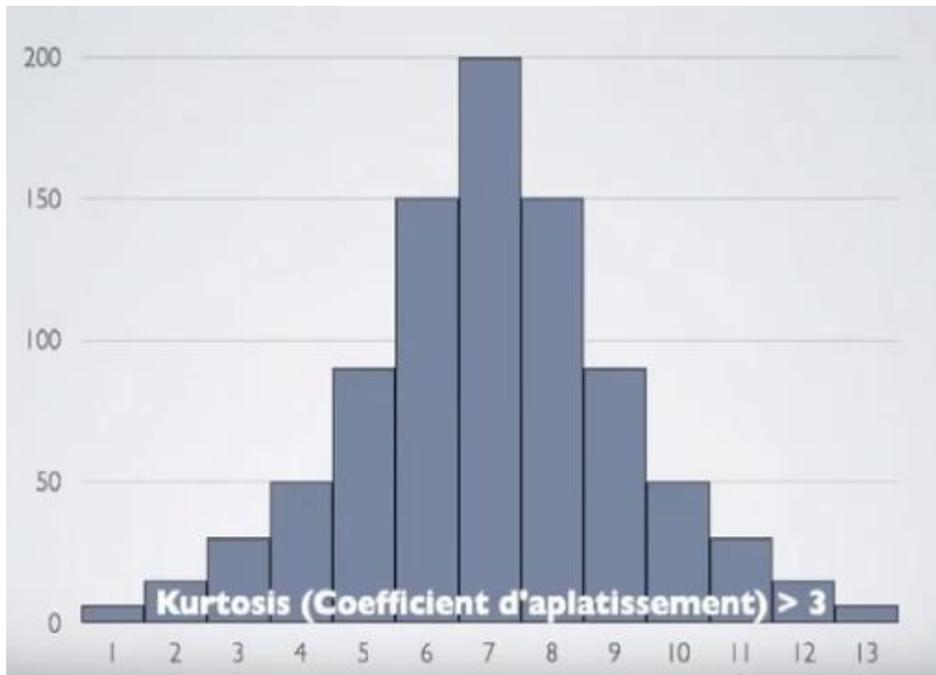
❑ Certaines distributions peuvent être :

- ✓ Plates ou platykurtiques (negative kurtosis)
- ✓ Exemple : chiffre d'affaire des PME (elle signifie qu'il y a une variabilité importante)

Les statistiques descriptives

Les indicateurs de forme :

Revoient à la loi normale (que nous évoquerons plus en détail lors des tests statistiques)



❑ Certaines distributions peuvent être :

- ✓ Pointues ou leptokurtiques (positive kurtosis)
- ✓ Exemple : Niveau d'instruction des pays développés (très peu d'analphabètes, peu d'individus de niveau école primaire, la plupart des gens ont un niveau bac, ensuite plus on monte dans le supérieur et moins il y a d'individus)

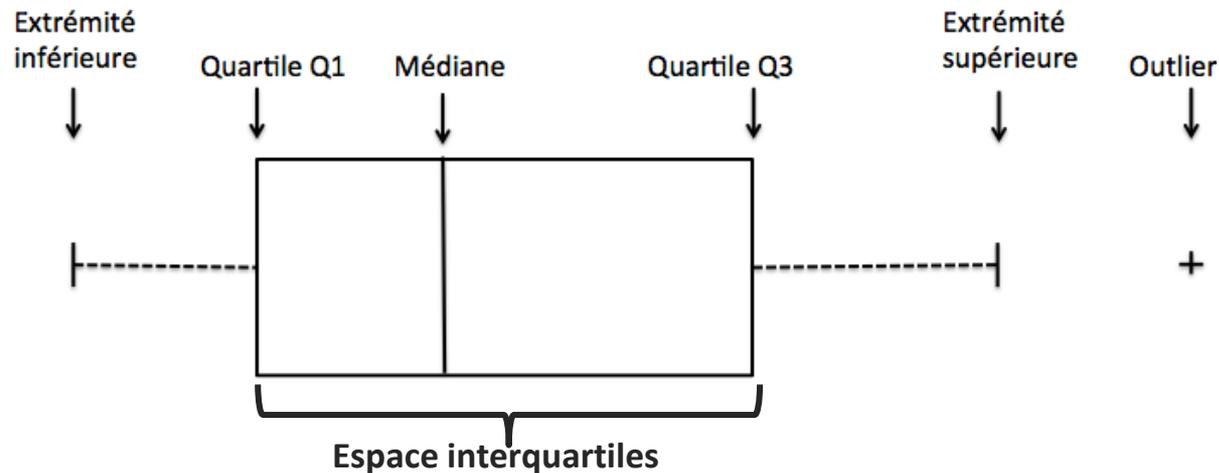
Les statistiques descriptives

Les indicateurs agrégés :

Contiennent plusieurs indicateurs en même temps

☐ Les **boxplots** ou **boîtes à moustaches**

- ✓ *Très utiles, car ils rassemblent la plupart des indices évoqués jusqu'à présent*

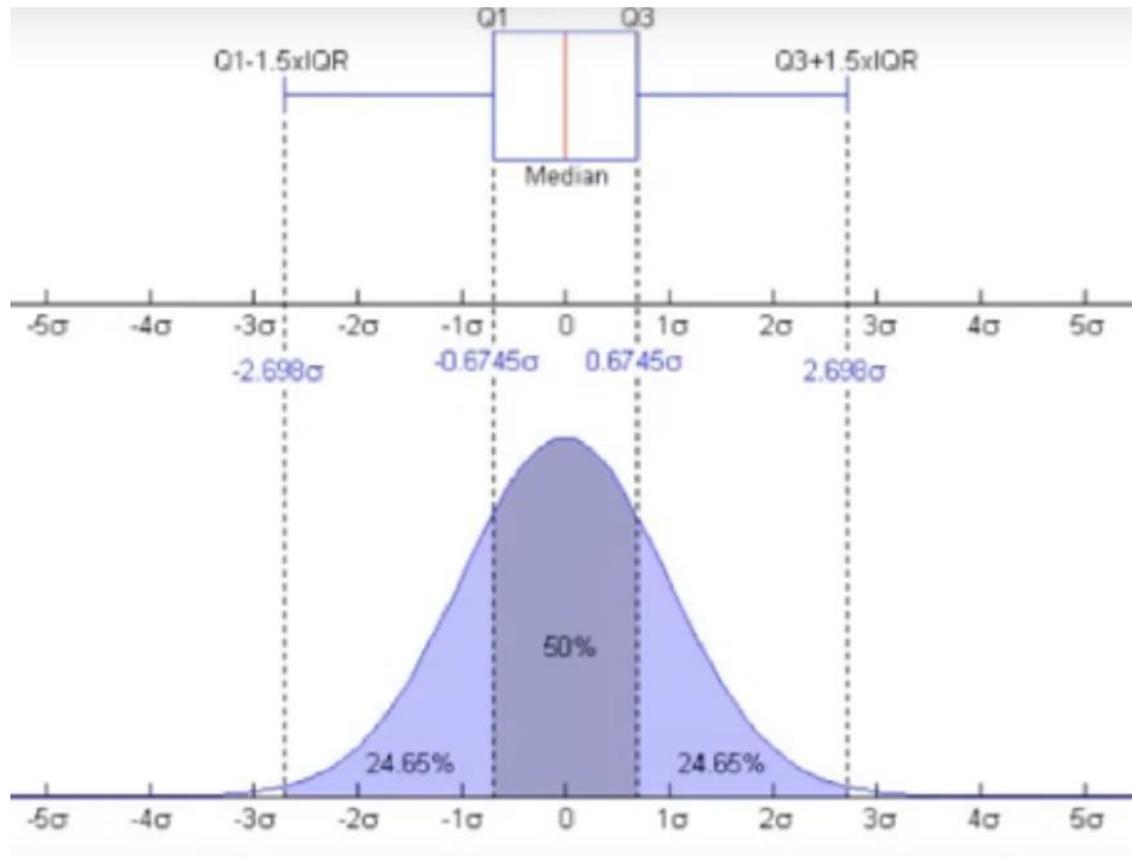


Les statistiques descriptives

Les indicateurs agrégés :

Contiennent plusieurs indicateurs en même temps

- ✓ *Boxplot d'une distribution normale*

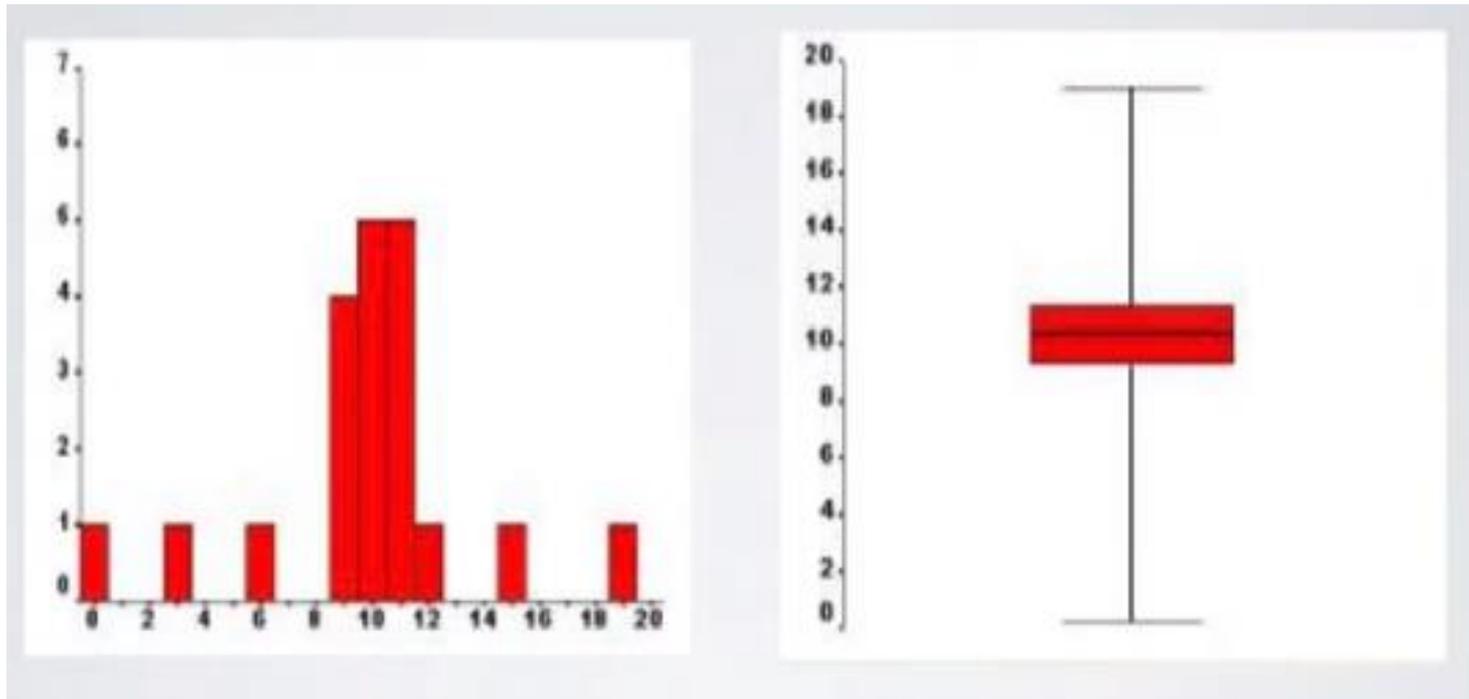


Les statistiques descriptives

Les indicateurs agrégés :

Contiennent plusieurs indicateurs en même temps

- ✓ *Boxplot d'une distribution leptokurtique (pointue)*

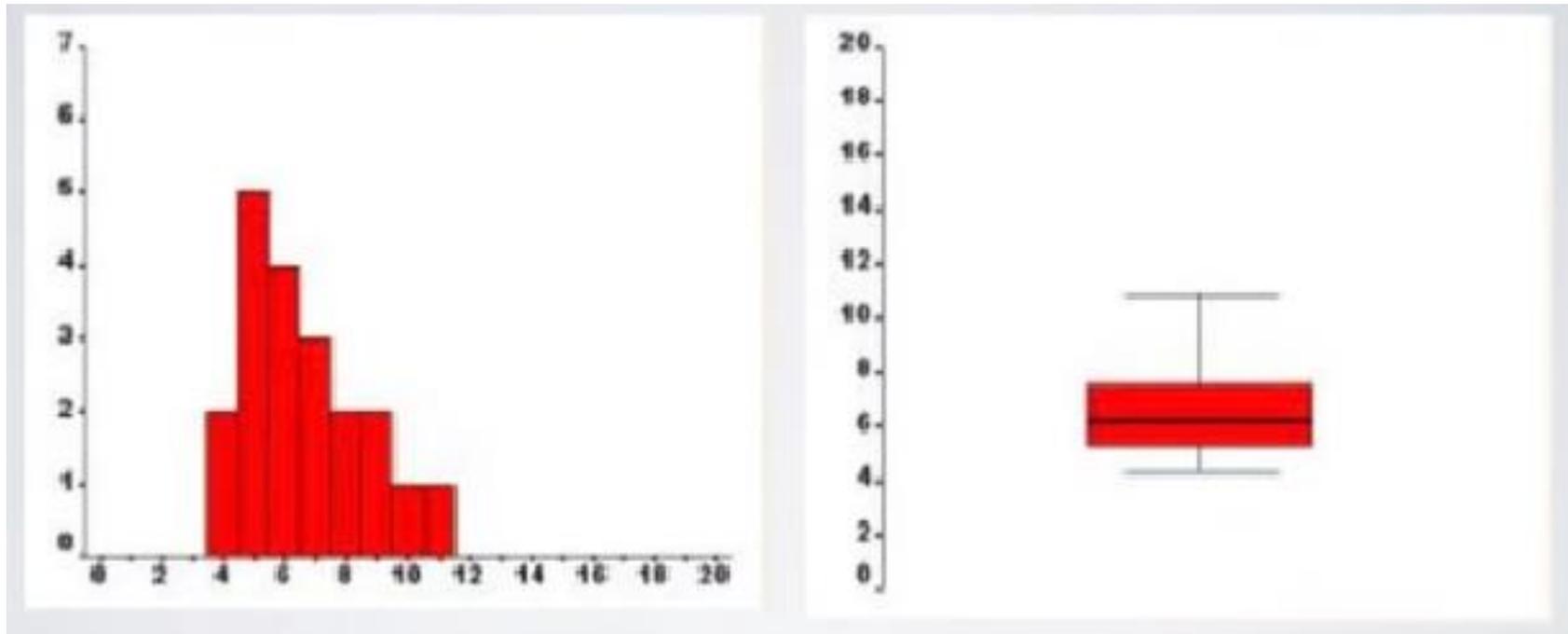


Les statistiques descriptives

Les indicateurs agrégés :

Contiennent plusieurs indicateurs en même temps

- ✓ *Boxplot d'une distribution asymétrique positive (positive skewness)*



MISE EN APPLICATION

Concevoir le journal de bord d'analyse « Code book »

- Le code book permet au manager de garder une trace écrite des transformations engendrées sur les données afin qu'elles puissent être traitables dans divers logiciels statistiques.

Exemple :

Var	Question (en fonction)	Modalités	Type
Prix	Quel prix avez-vous acheté le produit A?	1€=1, etc.	Continue
Client	Etes-vous client de StatFF?	Oui = 1 Non = 2	Nominale
Satisfaction	Si oui, quel est votre niveau de satisfaction (sur 10)?	1=1, etc.	Discrète

Document VariableCodebookNettoyage

A partir du fichier Excel retraçant la base de données « Exercice 1 Vendeur »

- Dans l'onglet Ex1 Variables, identifiez les variables :
 - qualitatives/quantitatives
 - Nominale/ ordinales / discrètes / continue
- Dans l'onglet Ex1 CodeBook, réalisez le code book:
 - **Code variable**
 - **Modalités**
 - **Type**

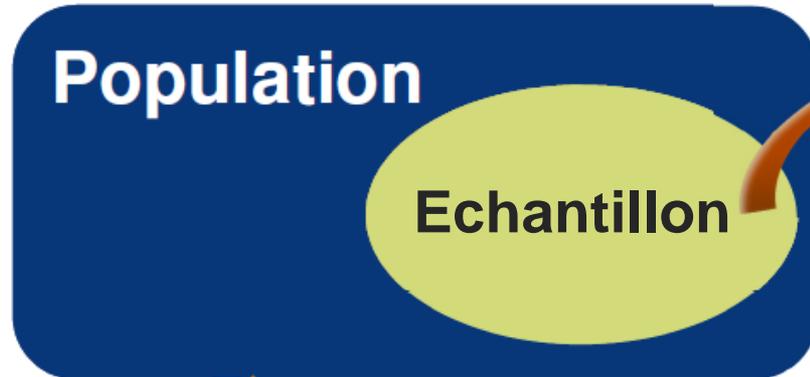
Document VariableCodebookNettoyage

A partir du fichier Excel retraçant la base de données « Exercice 2 Clients »

- Dans l'onglet Ex2 Statistiques descriptives, calculez :
 - Moyenne, Médiane, Mode, Etendue, Ecart type de la variable âge
 - Chez les hommes et chez les femmes
- Faites un graphique représentatif (moyennes)
 - Mettez-le en forme (couleurs, axes, titre)

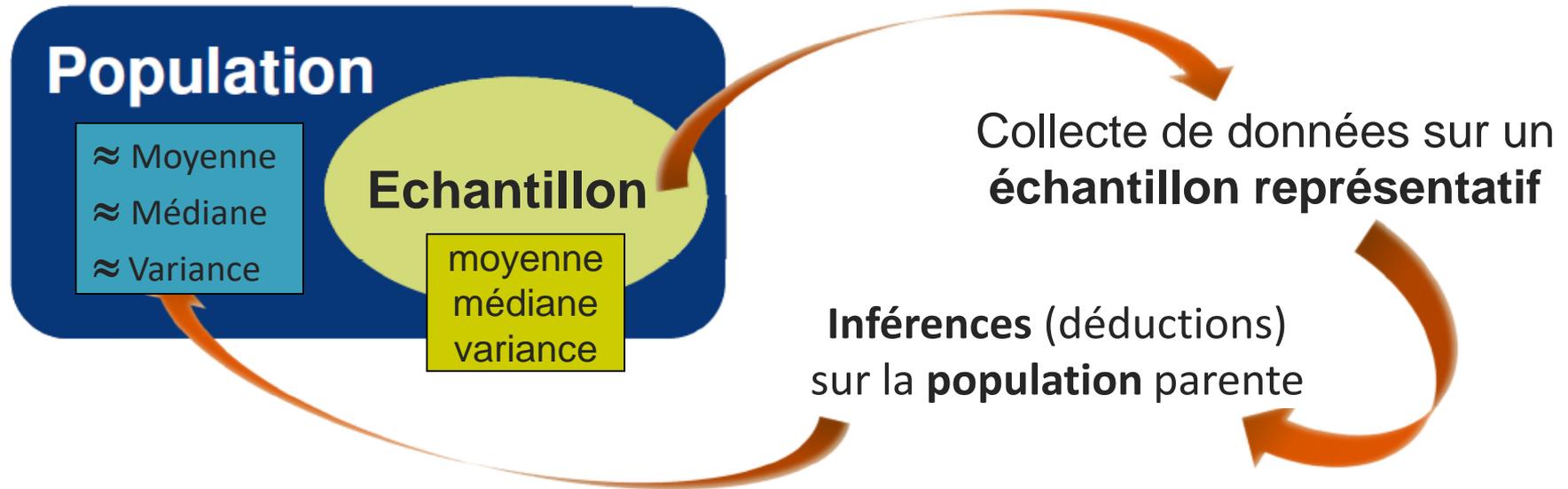
LES STATISTIQUES INFERENCELLES

Les statistiques inférentielles

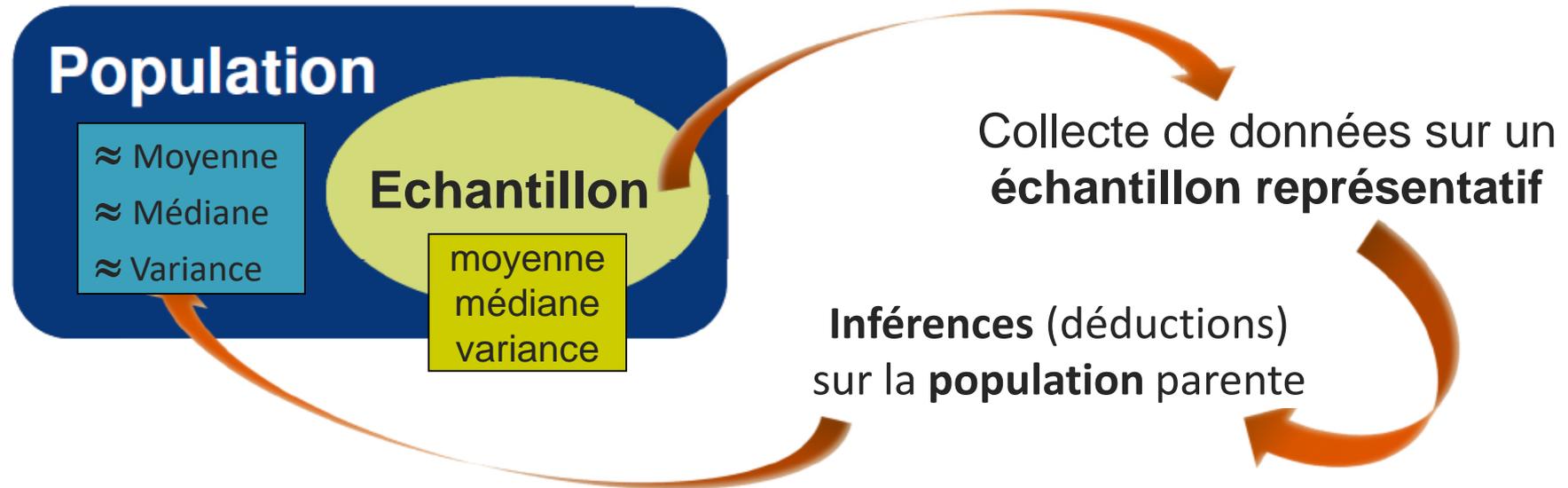


Collecte de données sur un
échantillon représentatif

Les statistiques inférentielles



Les statistiques inférentielles



- Objectif des statistiques inférentielles : **extrapoler** les paramètres d'un échantillon à la population globale
- Pour le faire, on s'appuie sur 2 grandes lois en fonction du type de variable que l'on étudie :

**La loi
normale**

**La loi
binomiale**

Les statistiques inférentielles

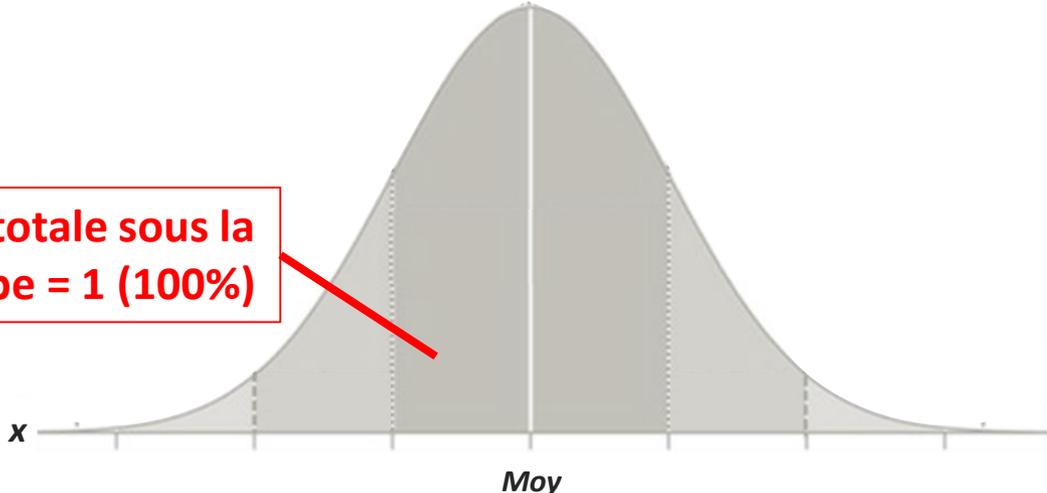
La Loi normale (ou courbe de Gauss)

- La loi normale est une loi de probabilité **qui s'applique aux variables quantitatives continues**
- Sa courbe se conçoit donc comme une **densité de probabilité**
- Cela veut dire que chaque valeur prise par une variable X (si celle-ci suit une loi normale) est associée à une probabilité (un effectif théorique)
- Le total de toutes les probabilités correspond à **l'aire sous la courbe** (d'où la notion de densité de probabilité). Par définition, il est égal à 1 (c'est-à-dire 100%)

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \times e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Fonction de la loi normale
Dans la pratique on s'en fiche !



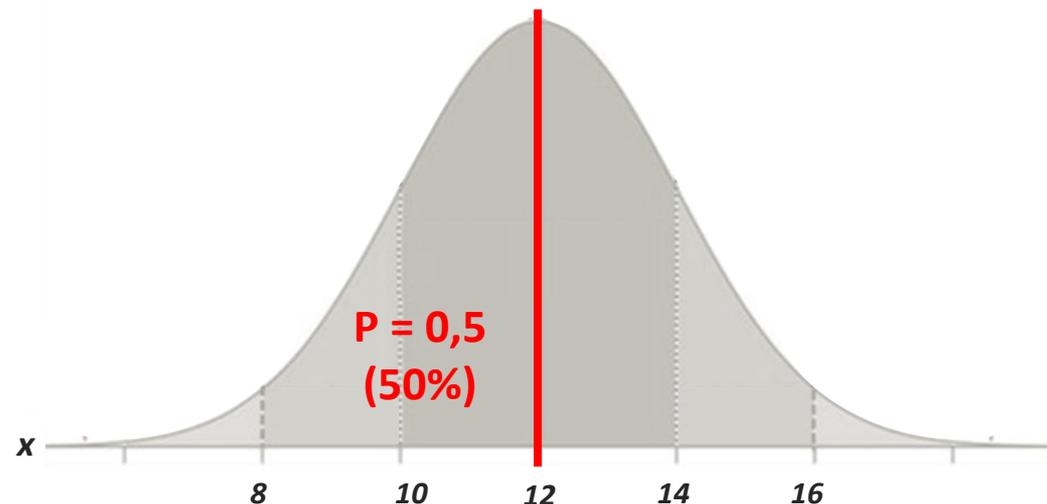
Aire totale sous la
courbe = 1 (100%)

x

Moy

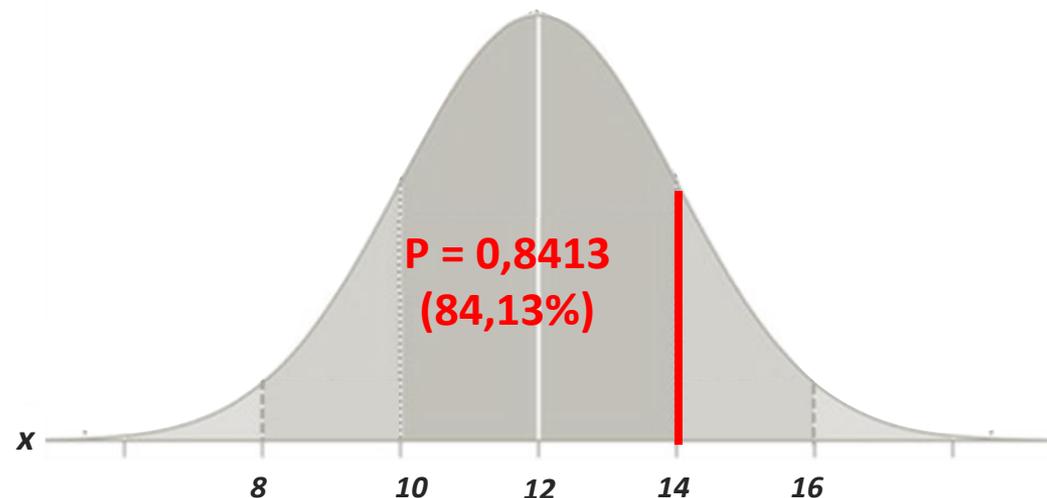
Loi normale (ou courbe de Gauss)

- ↪ Exemple : la moyenne générale d'une classe d'élèves est de 12.
- ↪ La variable « note » suivant une loi normale, on peut déterminer la probabilité d'élèves ayant moins de 12 (correspond à l'aire sous la courbe pour l'ensemble des valeurs de $X \leq 12$)
- ↪ Cette probabilité est de 0,5 (50%)



Loi normale (ou courbe de Gauss)

- ↪ Exemple : la moyenne générale d'une classe d'élèves est de 12.
- ↪ La variable « note » suivant une loi normale, on peut déterminer l'effectif théorique des élèves ayant moins de 12 (correspond à l'aire sous la courbe pour l'ensemble des valeurs de $X \leq 12$)
- ↪ Cette probabilité est de 0,5 (50%)
- ↪ De même, l'effectif théorique d'élèves ayant moins de 14 serait de 0,8413 (84,13%)
- ↪ **Comment calcule-t-on ces probabilités ? En utilisant la table Z**



Les statistiques inférentielles

- La table des Z fournit, pour chaque valeur d'une variable (suivant une loi normale), la probabilité de l'effectif qui lui est associé
- En d'autres termes, elle fournit **l'aire la plus grande sous la courbe (probabilité que tel élève ait moins de 14 par exemple)**
- Petite subtilité : on utilise généralement la table Z **bilatérale externe cumulée** -> fournit l'aire **la plus petite** sous la courbe (**probabilité que l'élève « n'ait pas moins de 14 »**) -> **pourquoi ?** Renvoie à la logique de l'hypothèse nulle, voir partie tests d'hypothèses).
- Pour l'utiliser, il faut **centrer et réduire** notre variable afin de la faire correspondre à la loi normale
 - Normaliser la moyenne (centrer) : « faire comme si » la moyenne de notre échantillon était égale à 0
 - Normaliser l'écart type (réduire) : pour qu'il corresponde à l'écart-type de la loi normale (appelé score Z)

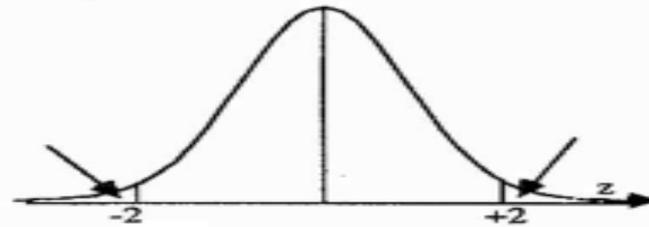
$$Z = \frac{X - \bar{X}}{\text{Écart-type}}$$

Table z, bilatérale externe cumulée



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	1.000	.993	.984	.976	.968	.960	.952	.944	.936	.928
0.1	.950	.941	.934	.926	.918	.910	.903	.895	.887	.879
0.2	.841	.834	.826	.818	.810	.803	.795	.787	.779	.772
0.3	.764	.757	.749	.741	.734	.726	.719	.711	.704	.697
0.4	.689	.682	.674	.667	.660	.653	.645	.638	.631	.624
0.5	.617	.610	.603	.596	.589	.582	.575	.568	.562	.555
0.6	.549	.542	.535	.529	.522	.516	.509	.503	.497	.490
0.7	.484	.478	.472	.465	.459	.453	.447	.441	.435	.430
0.8	.424	.418	.412	.407	.401	.395	.390	.384	.379	.373
0.9	.366	.361	.356	.352	.347	.342	.337	.332	.327	.322
1.0	.317	.313	.308	.303	.298	.294	.289	.285	.280	.276
1.1	.271	.267	.263	.258	.254	.250	.246	.242	.238	.234
1.2	.230	.226	.222	.219	.215	.211	.208	.204	.201	.197
1.3	.194	.190	.187	.184	.180	.177	.174	.171	.168	.165
1.4	.162	.159	.156	.153	.150	.147	.144	.142	.139	.136
1.5	.134	.131	.129	.126	.124	.121	.119	.116	.114	.112
1.6	.110	.107	.105	.103	.101	.099	.097	.095	.093	.091
1.7	.089	.087	.085	.084	.082	.080	.078	.077	.075	.073
1.8	.072	.070	.069	.067	.066	.064	.063	.062	.060	.059
1.9	.057	.056	.055	.054	.052	.051	.050	.049	.048	.047
2.0	.046	.044	.043	.042	.041	.040	.039	.038	.037	.036
2.1	.035	.033	.033	.032	.032	.031	.030	.029	.028	.028
2.2	.028	.027	.026	.026	.025	.024	.024	.023	.023	.022
2.3	.021	.021	.020	.020	.019	.019	.018	.018	.017	.017
2.4	.016	.016	.015	.015	.015	.014	.014	.014	.013	.013
2.5	.013	.013	.012	.012	.011	.011	.011	.010	.010	.010
2.6	.009	.009	.009	.009	.008	.008	.008	.008	.007	.007
2.7	.007	.007	.007	.007	.006	.006	.006	.006	.005	.005
2.8	.005	.005	.005	.005	.004	.004	.004	.004	.004	.004
2.9	.004	.004	.004	.004	.003	.003	.003	.003	.003	.003
3.0	.003	.002	.001							

Table z, bilatérale externe cumulée



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	1.00	.992	.984	.976	.968	.960	.952	.944	.936	.928
0.1	.920	.912	.904	.897	.887	.881	.873	.865	.857	.849
0.2	.841	.834	.826	.818	.810	.803	.795	.787	.779	.772
0.3	.764	.757	.749	.741	.734	.726	.719	.711	.704	.697
0.4	.689	.682	.674	.667	.660	.653	.645	.638	.631	.624
0.5	.617	.610	.603	.596	.589	.582	.575	.569	.562	.555
0.6	.549	.542	.535	.529	.522	.516	.509	.503	.497	.490
0.7	.484	.478	.472	.465	.459	.453	.447	.441	.435	.430
0.8	.424	.418	.412	.407	.401	.395	.390	.384	.379	.373
0.9	.368	.363	.358	.352	.347	.342	.337	.332	.327	.322
1.0	.317	.313	.308	.303	.298	.294	.289	.285	.280	.276
1.1	.271	.267	.263	.258	.254	.250	.246	.242	.238	.234
1.2	.230	.226	.222	.219	.215	.211	.208	.204	.201	.197
1.3	.194	.190	.187	.184	.180	.177	.174	.171	.168	.165
1.4	.162	.159	.156	.153	.150	.147	.144	.142	.139	.136
1.5	.134	.131	.129	.126	.124	.121	.119	.116	.114	.112
1.6	.110	.107	.105	.103	.101	.100	.097	.095	.093	.091
1.7	.089	.087	.085	.084	.082	.080	.078	.077	.075	.073
1.8	.072	.070	.069	.067	.066	.064	.063	.062	.060	.059
1.9	.057	.056	.055	.054	.052	.051	.050	.049	.048	.047
2.0	.046	.044	.043	.042	.041	.040	.039	.038	.038	.037
2.1	.035	.035	.034	.033	.032	.032	.031	.030	.029	.028
2.2	.028	.027	.026	.026	.025	.024	.024	.023	.023	.022
2.3	.021	.021	.020	.020	.019	.019	.018	.018	.017	.017
2.4	.016	.016	.016	.015	.015	.014	.014	.014	.013	.013
2.5	.012	.012	.012	.011	.011	.011	.010	.010	.010	.010
2.6	.009	.009	.009	.009	.008	.008	.008	.008	.007	.007
2.7	.007	.007	.007	.006	.006	.006	.006	.006	.005	.005
2.8	.005	.005	.005	.005	.004	.004	.004	.004	.004	.004
2.9	.004	.004	.004	.003	.003	.003	.003	.003	.003	.003
3.0	.003	.002	.001	-						

Ecart type centré réduit (Z)
D'abord ligne puis colonne

Probabilité (p) ou « plus petite aire sous la courbe »

Les statistiques inférentielles

En statistique inférentielle, l'utilisation de la loi normale est très importante

- Elle sert à estimer les paramètres d'une population par rapport à ceux de notre échantillon en déterminant des **intervalles de confiance**
- L'intervalle de confiance le plus utilisé est celui de la **moyenne à 95%** : il détermine un espace à l'intérieur duquel la moyenne de la population a 95% de se situer. -> Donc moins de 5% de chances qu'elle soit à l'extérieur
- A cet égard, on utilise le score **$z = 1,96$** (qui correspond à un intervalle allant de $- 1,96$ écart-type à $+ 1,96$ écart-type)

Probabilité p
⇒ Lecture de la **table des z**



$$x = M \pm 1.96 * ET$$

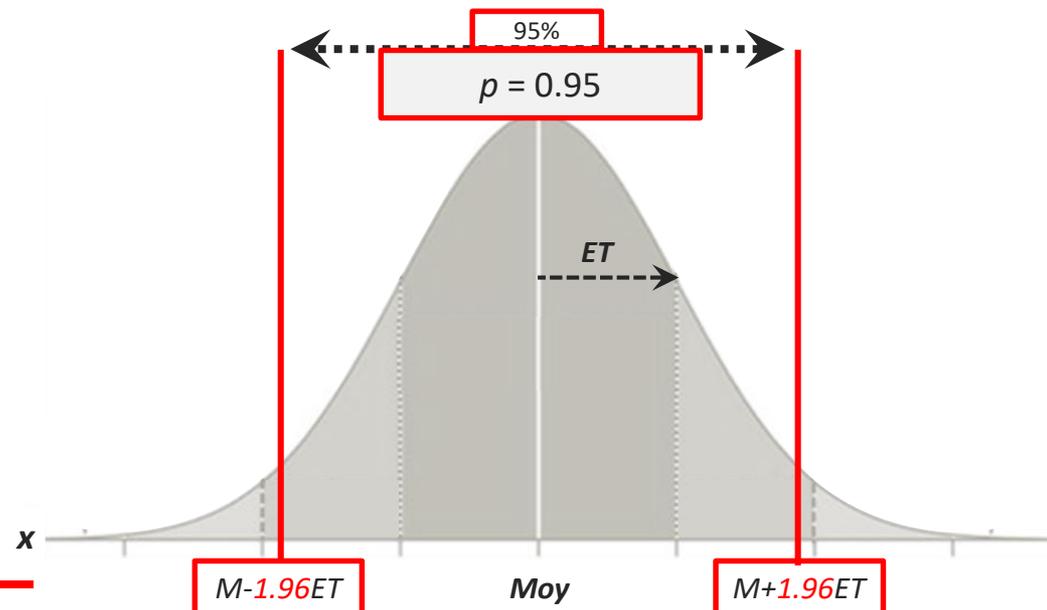
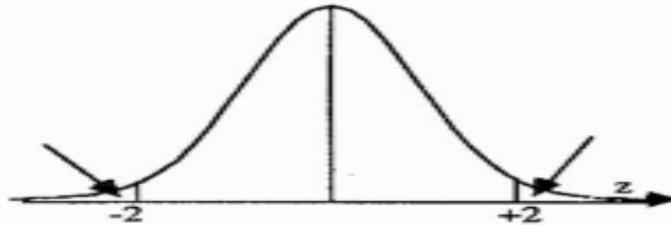


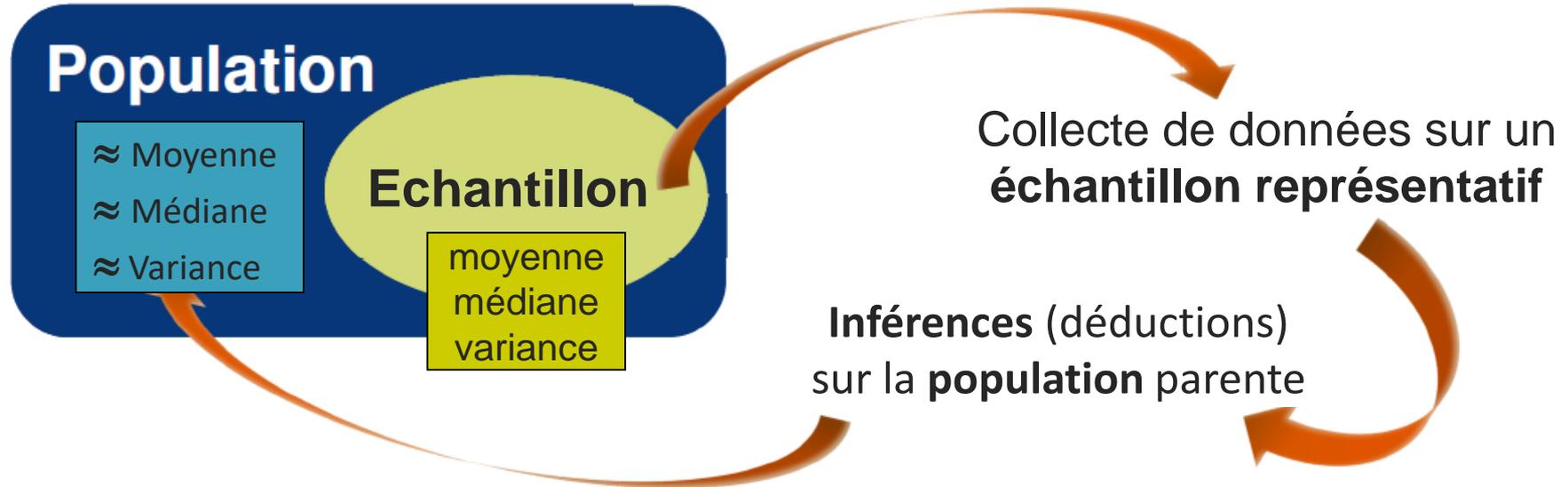
Table z, bilatérale externe cumulée



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	1.00	.992	.984	.976	.968	.960	.952	.944	.936	.928
0.1	.920	.912	.904	.897	.887	.881	.873	.865	.857	.849
0.2	.841	.834	.826	.818	.810	.803	.795	.787	.779	.772
0.3	.764	.757	.749	.741	.734	.726	.719	.711	.704	.697
0.4	.689	.682	.674	.667	.660	.653	.645	.638	.631	.624
0.5	.617	.610	.603	.596	.589	.582	.575	.569	.562	.555
0.6	.549	.542	.535	.529	.522	.516	.509	.503	.497	.490
0.7	.484	.478	.472	.465	.459	.453	.447	.441	.435	.430
0.8	.424	.418	.412	.407	.401	.395	.390	.384	.379	.373
0.9	.368	.363	.358	.352	.347	.342	.337	.332	.327	.322
1.0	.317	.313	.308	.303	.298	.294	.289	.285	.280	.276
1.1	.271	.267	.263	.258	.254	.250	.246	.242	.238	.234
1.2	.230	.226	.222	.219	.215	.211	.208	.204	.201	.197
1.3	.194	.190	.187	.184	.180	.177	.174	.171	.168	.165
1.4	.162	.159	.156	.153	.150	.147	.144	.142	.139	.136
1.5	.134	.131	.129	.126	.124	.121	.119	.116	.114	.112
1.6	.110	.107	.105	.103	.101	.100	.097	.095	.093	.091
1.7	.089	.087	.085	.084	.082	.080	.078	.077	.075	.073
1.8	.072	.070	.069	.067	.066	.064	.063	.062	.060	.059
1.9	.057	.056	.055	.054	.052	.051	.050	.049	.048	.047
2.0	.046	.044	.043	.042	.041	.040	.039	.038	.038	.037
2.1	.035	.035	.034	.033	.032	.032	.031	.030	.029	.028
2.2	.028	.027	.026	.026	.025	.024	.024	.023	.023	.022
2.3	.021	.021	.020	.020	.019	.019	.018	.018	.017	.017
2.4	.016	.016	.016	.015	.015	.014	.014	.014	.013	.013
2.5	.012	.012	.012	.011	.011	.011	.010	.010	.010	.010
2.6	.009	.009	.009	.009	.008	.008	.008	.008	.007	.007
2.7	.007	.007	.007	.006	.006	.006	.006	.006	.005	.005
2.8	.005	.005	.005	.005	.004	.004	.004	.004	.004	.004
2.9	.004	.004	.004	.003	.003	.003	.003	.003	.003	.003
3.0	.003	.002	.001	-						

Plus petite aire sous la courbe (ou « chances que la moyenne se situe à l'extérieur de l'intervalle ») = 0,05 soit 5%

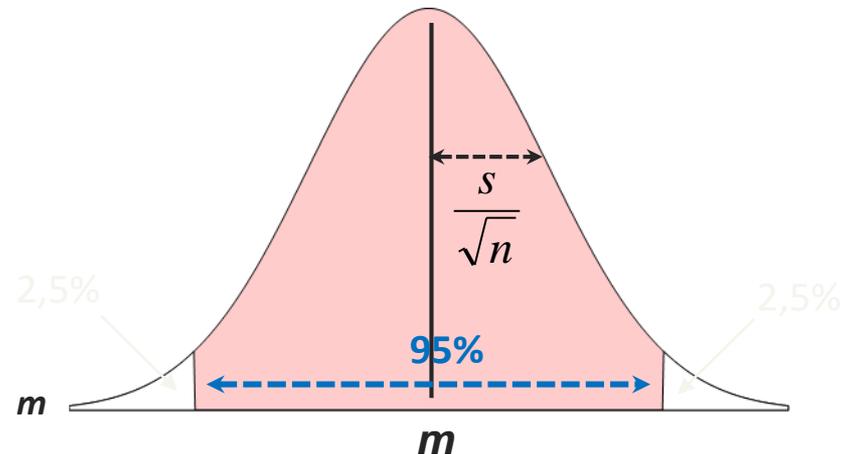
Les statistiques inférentielles



L'intervalle de confiance à 95% (IC_{95%})

$$Lim_{inf} = m - 1,96 \frac{s}{\sqrt{n}}$$

$$Lim_{sup} = m + 1,96 \frac{s}{\sqrt{n}}$$



Les statistiques inférentielles

Exercice d'application 1 :

Dépenses en **achats de matériels sportifs** chez **40 étudiants STAPS**

Moyenne observée : $m = 166,5 \text{ €}$

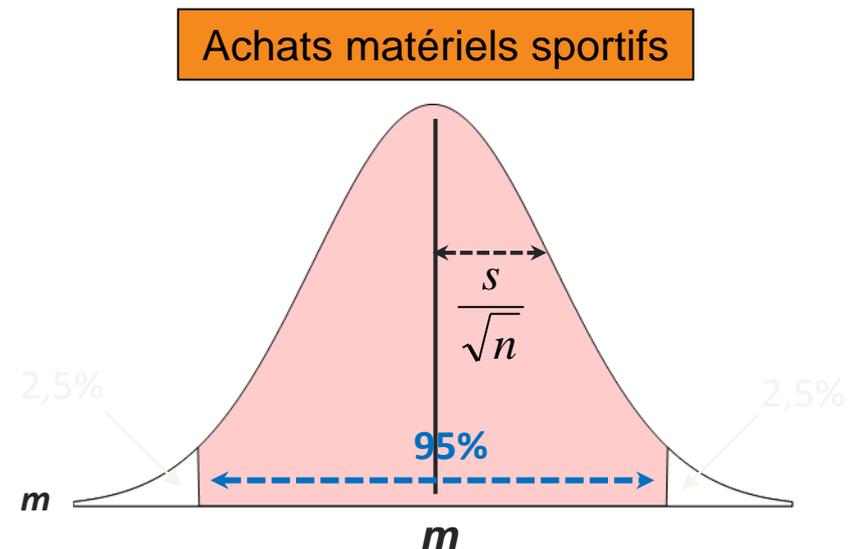
Écart type estimée : $s = 15,7 \text{ €}$

Quel est l'intervalle de confiance à 95% de la moyenne ?

L'intervalle de confiance à 95% (IC_{95%})

$$Lim_{\text{inf}} = m - 1,96 \frac{s}{\sqrt{n}}$$

$$Lim_{\text{sup}} = m + 1,96 \frac{s}{\sqrt{n}}$$



Les statistiques inférentielles

Exercice d'application 1 :

Dépenses en **achats de matériels sportifs** chez **40 étudiants STAPS**

Moyenne observée : $m = 166,5 \text{ €}$

Écart type estimée : $s = 15,7 \text{ €}$

Quel est l'intervalle de confiance à 95% de la moyenne ?

L'intervalle de confiance à 95% ($IC_{95\%}$)

$$Lim_{inf} = m - 1,96 \frac{s}{\sqrt{n}}$$

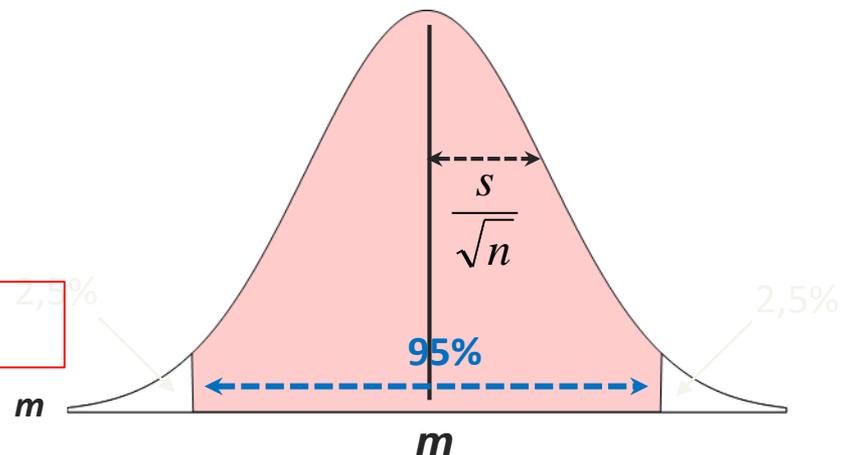
$$166,5 - 1,96 \times 2,482 \\ = 161,6 \text{ €}$$

$$Lim_{sup} = m + 1,96 \frac{s}{\sqrt{n}}$$

$$166,5 + 1,96 \times 2,482 \\ = 171,4 \text{ €}$$

$$IC_{95\%} = [161,6 - 171,4 \text{ €}] \Rightarrow \text{Amplitude : } 9,8 \text{ €}$$

Achats matériels sportifs



Les statistiques inférentielles

Exercice d'application 2 :

Dépenses en **achats de matériels sportifs** chez **228** étudiants STAPS

Moyenne observée : $m = 166,5 \text{ €}$

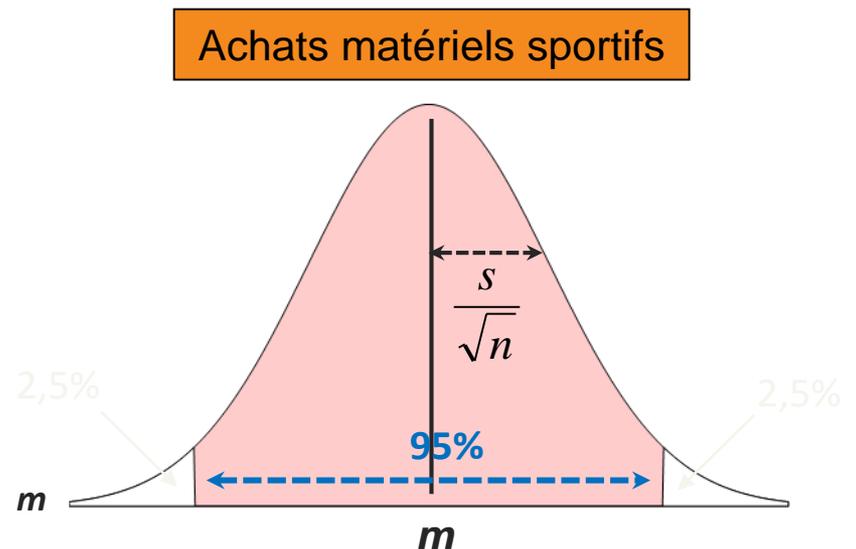
Écart type estimée : $s = 15,7 \text{ €}$

Quel est l'intervalle de confiance à 95% de la moyenne ?

L'intervalle de confiance à 95% (IC_{95%})

$$Lim_{\text{inf}} = m - 1,96 \frac{s}{\sqrt{n}}$$

$$Lim_{\text{sup}} = m + 1,96 \frac{s}{\sqrt{n}}$$



Les statistiques inférentielles

Exercice d'application 2 :

Dépenses en **achats de matériels sportifs** chez **228** étudiants STAPS

Moyenne observée : $m = 166,5 \text{ €}$

Écart type estimée : $s = 15,7 \text{ €}$

Quel est l'intervalle de confiance à 95% de la moyenne ?

L'intervalle de confiance à 95% ($IC_{95\%}$)

$$Lim_{inf} = m - 1,96 \frac{s}{\sqrt{n}}$$

$$166,5 - 1,96 \times 1,04 =$$

$$164,5 \text{ €}$$

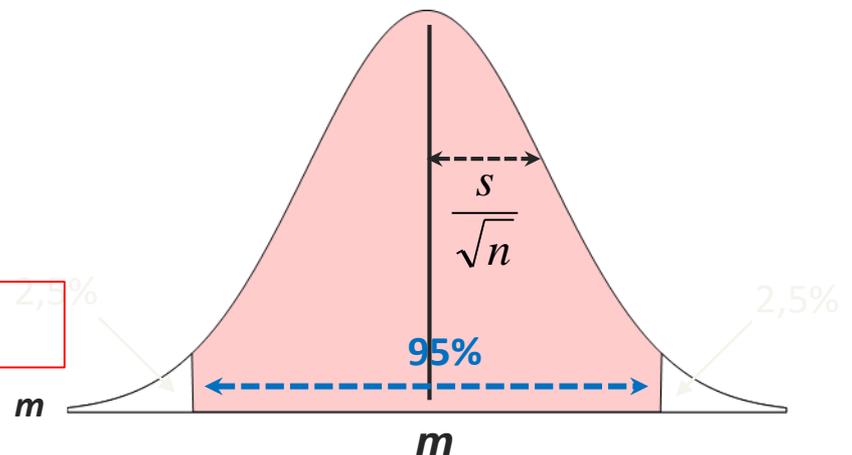
$$Lim_{sup} = m + 1,96 \frac{s}{\sqrt{n}}$$

$$166,5 + 1,96 \times 1,04$$

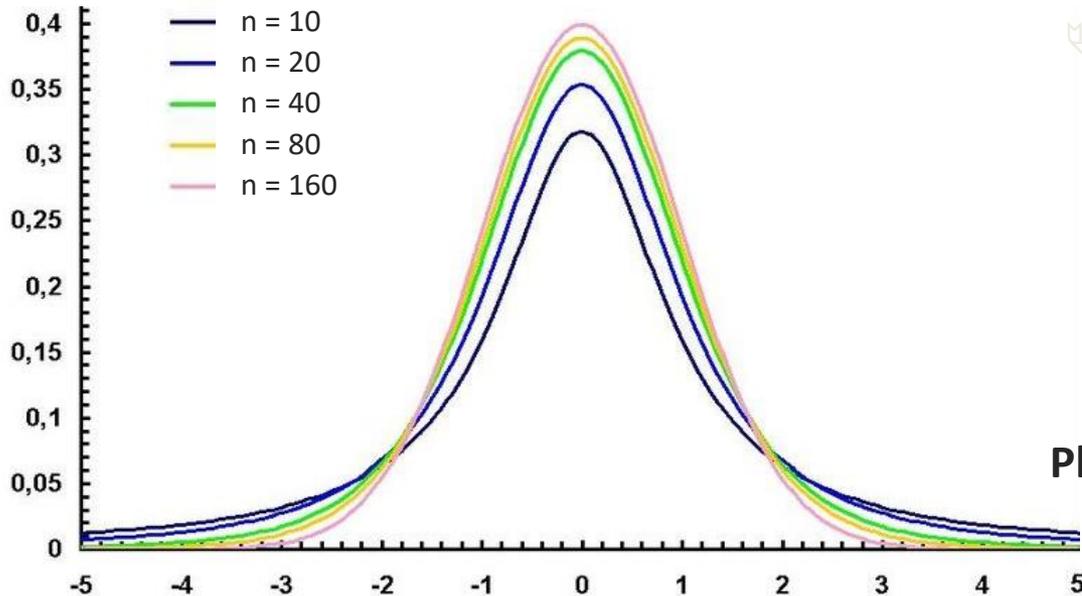
$$= 168,5 \text{ €}$$

$$IC_{95\%} = [164,5 - 168,5 \text{ €}] \Rightarrow \text{Amplitude} : 4 \text{ €}$$

Achats matériels sportifs



Les statistiques inférentielles



Plus l'échantillon est grand

- ↳ + mesure précise
- ↳ + intervalle de confiance étroit
- ↳ Attention : l'augmentation de la précision n'est pas proportionnelle à celle de l'échantillon mais à sa racine carrée (pour une précision 10 fois plus importante, il faut 100 individus supplémentaires)

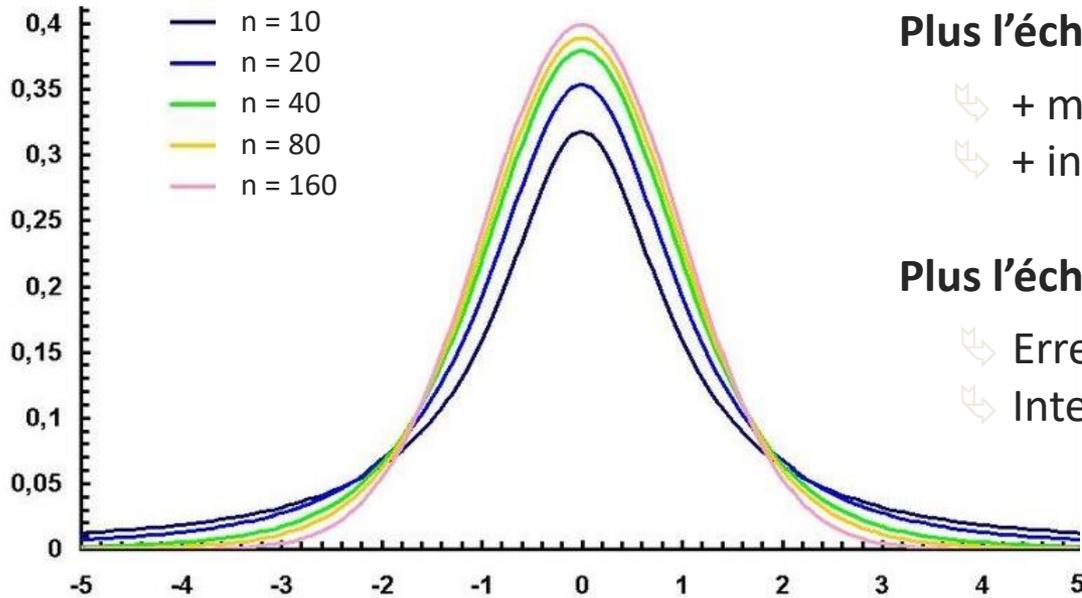
Plus l'échantillon est petit

- ↳ Erreur estimation variance importante
- ↳ Intervalle de confiance + large

Pour les petits échantillons ($N < 30$)

- ↳ On n'utilise pas la table des z mais le **t de Student**

Les statistiques inférentielles



Plus l'échantillon est grand

- + mesure précise
- + intervalle de confiance étroit

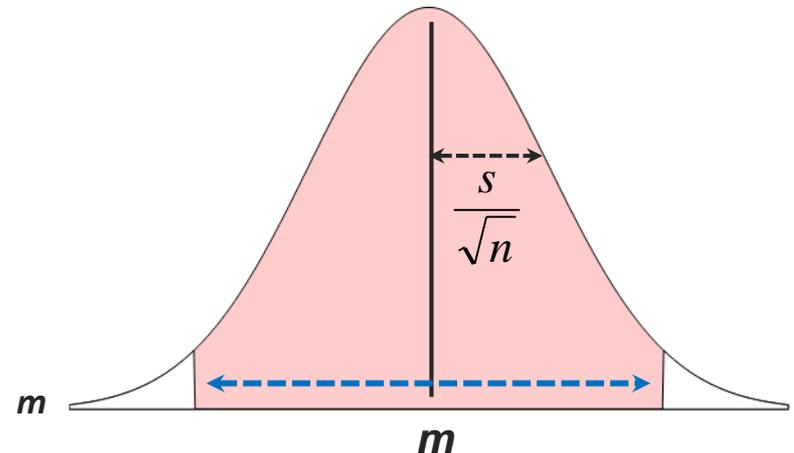
Plus l'échantillon est petit

- Erreur estimation variance importante
- Intervalle de confiance + large

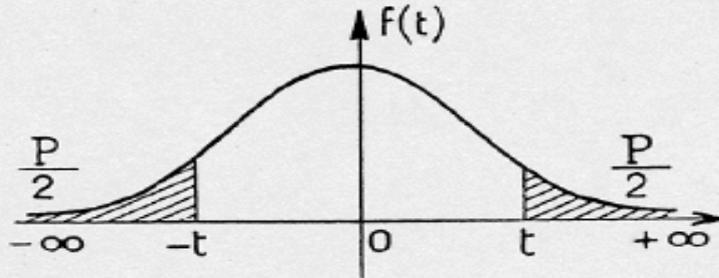
⇒ **Le cas de petits échantillons ($N < 30$)**

$$Lim_{\text{inf}} = m - t \frac{s}{\sqrt{n}}$$

$$Lim_{\text{sup}} = m + t \frac{s}{\sqrt{n}}$$



ddl = n-1 (exemple pour un échantillon de 20 individus : n-1 = 19)



P = risque d'erreur choisi (généralement 5% donc 0,05)

$\nu \backslash P$	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,141	0,285	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,785	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Les statistiques inférentielles

Exercice d'application 3 :

Dépenses en **achats de matériels sportifs** chez **10 étudiants STAPS**

Moyenne observée : $m = 166,5 \text{ €}$

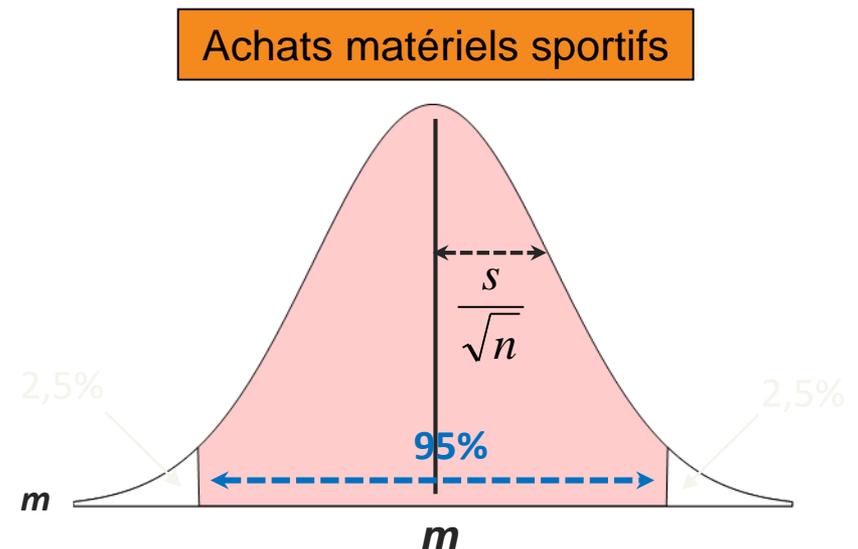
Écart type estimée : $s = 15,7 \text{ €}$

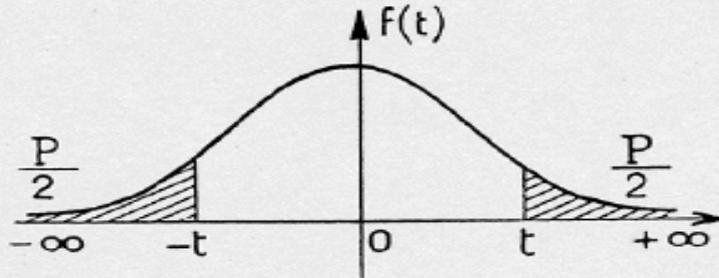
Quel est l'intervalle de confiance à 95% de la moyenne ?

L'intervalle de confiance à 95% ($IC_{95\%}$)

$$Lim_{\text{inf}} = m - t \frac{s}{\sqrt{n}}$$

$$Lim_{\text{sup}} = m + t \frac{s}{\sqrt{n}}$$





ddl = n-1

ν \ P	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,141	0,277	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,785	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,542	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Les statistiques inférentielles

Exercice d'application 3 :

Dépenses en **achats de matériels sportifs** chez **10 étudiants STAPS**

Moyenne observée : $m = 166,5 \text{ €}$

Écart type estimée : $s = 15,7 \text{ €}$

Quel est l'intervalle de confiance à 95% de la moyenne ?

L'intervalle de confiance à 95% ($IC_{95\%}$)

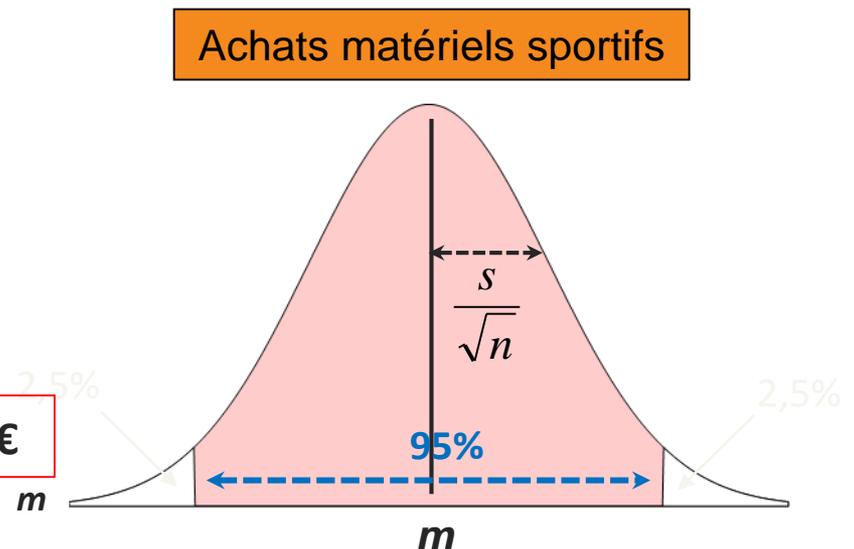
$$Lim_{\text{inf}} = m - t \frac{s}{\sqrt{n}}$$

$$166,5 - 2,262 \times 4,96 \\ = 155,3 \text{ €}$$

$$Lim_{\text{sup}} = m + t \frac{s}{\sqrt{n}}$$

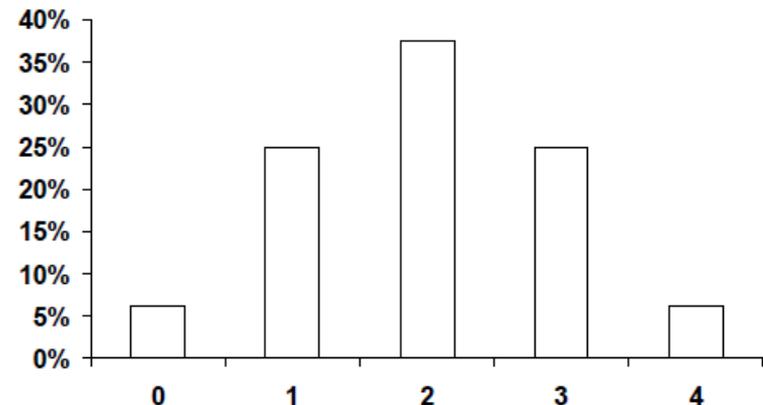
$$166,5 + 2,262 \times 4,96 \\ = 177,7 \text{ €}$$

IC95% : [155,3 — 177,7 €] => Amplitude : 22,4 €



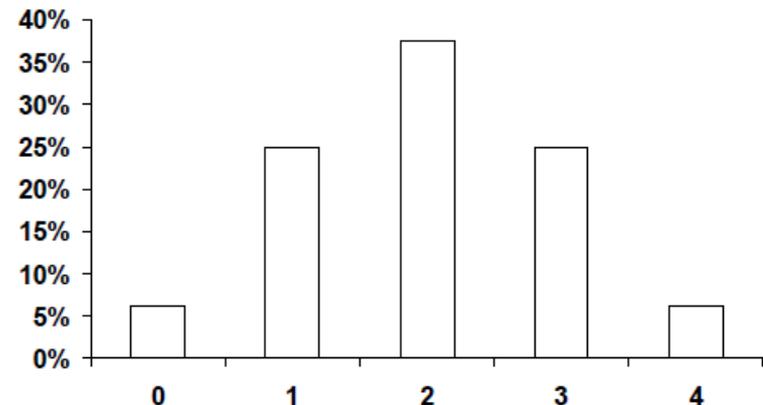
La Loi binomiale

- La loi binomiale est une loi de probabilité qui s'applique aux **variables quantitatives discrètes** ou aux **variables qualitatives**
- Il ne s'agit plus de définir une densité mais une pure probabilité qu'une variable X prenne telle ou telle valeur (« oui ou non », « homme ou femme »)
- On l'appelle aussi « loi des pourcentages » -> très importante en management/marketing



Loi binomiale

- Elle se note de la façon suivante : **B** ($n : p$), où n est la taille de l'échantillon et p la probabilité de l'événement qu'on cherche à étudier.
- Elle permet de visualiser la distribution des probabilités (notées en pourcentages)
- Exemple :
 - on part d'une population constituée de 50% d'hommes et de 50% de femmes ($p = 0,5$ ou 50%)
 - on tire 4 individus au hasard au sein d'une population
- Donc la loi se note B (4 : 0,5)



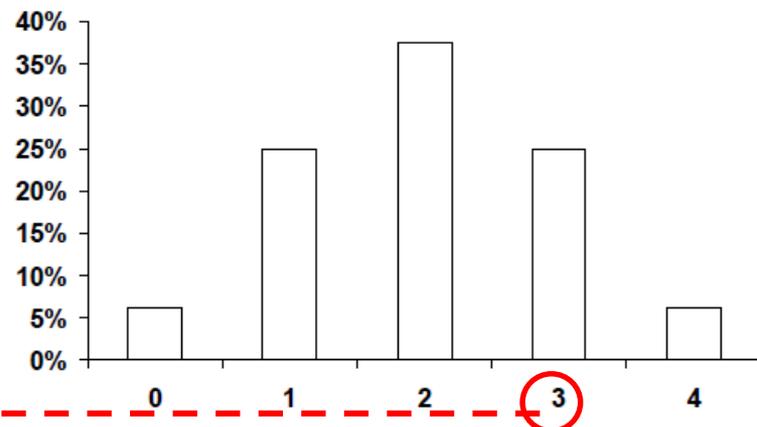
Loi binomiale

- L'objectif de base est de déterminer la probabilité « p » d'obtenir l'événement « k » qui nous intéresse (par exemple « quelle est la probabilité que 3 des 4 individus tirés au hasard au sein de ma population soient des hommes ? »)
- On notera « q » la probabilité inverse, qui est par définition égale à $1 - p$ (puisque une probabilité de 100% est égale à 1).
- A partir d'une formule **qu'on n'a pas besoin de comprendre** $P(X = k) = \binom{n}{k} p^k q^{n-k}$

La loi nous permet de déterminer qu'il y a :

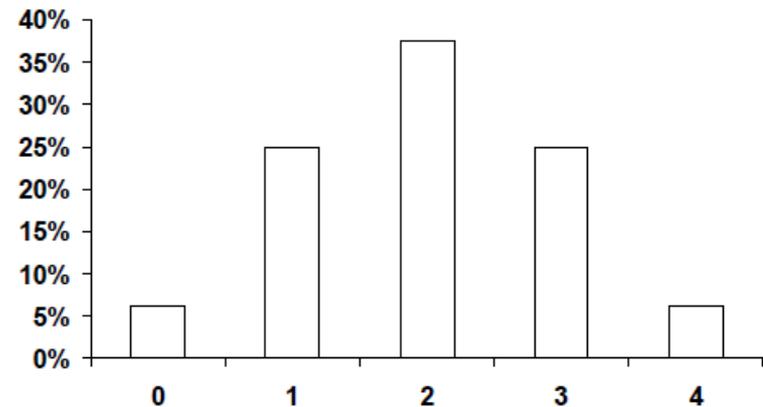
- 5% de chances de ne tirer aucune homme
- 25% de chances de tirer 1 homme
- 37,5% de chances de tirer 2 hommes
- 25% de chances de tirer 3 hommes
- 5% de chances de tirer 4 hommes

Ainsi $p(k=3)$ est égale à 25%, j'ai 25% de chances que 3 individus sur 4 soient des hommes



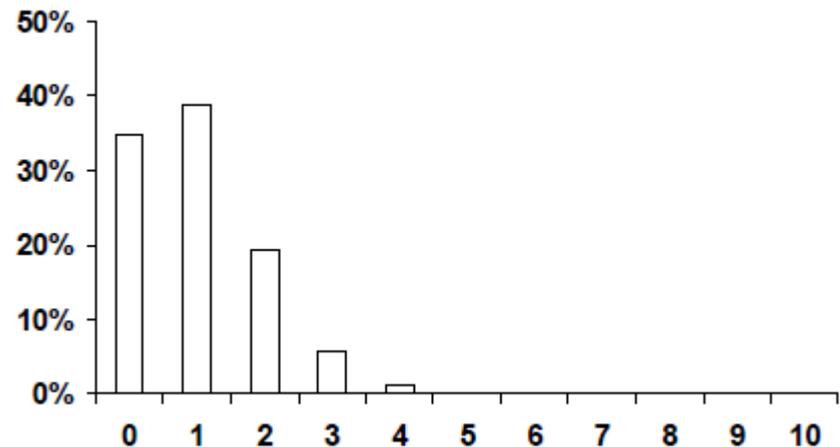
Loi binomiale

- Attention, la loi binomiale ne ressemble à la loi normale que lorsque p tend vers 0,5
- Les distributions binomiales peuvent s'éloigner de la loi normale



Loi binomiale

- Attention, la loi binomiale ne ressemble à la loi normale que lorsque p tend vers 0,5
- Les distributions binomiales peuvent s'éloigner de la loi normale

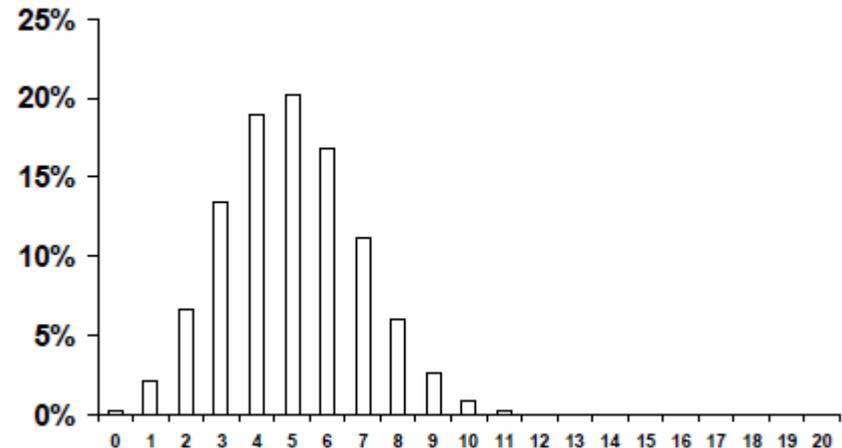


Distribution des nombres de gauchers attendus dans un échantillon aléatoire de 10 personnes prélevé au sein d'une population comportant exactement 10% de gauchers ($p=0,1$)

Les statistiques inférentielles

Loi binomiale

- Attention, la loi binomiale ne ressemble à la loi normale que lorsque p tend vers 0,5
- Les distributions binomiales peuvent s'éloigner de la loi normale



Distribution des notes attendues chez des étudiants répondant au hasard à un QCM sachant qu'il y a 4 réponses possibles pour chaque question, qu'une seule est bonne et que chaque bonne réponse vaut 1 point ($p=0,25$)

Loi binomiale

- La loi binomiale permet ainsi de calculer des intervalles de confiance pour les variables **qualitatives** et **quantitatives discrètes**, autrement dit pour **les pourcentages**

Intervalles de confiance d'un pourcentage :

- Grand échantillon (np et $nq > 5$)
- à 95%

$$Lim_{\text{inf}} = p(\text{obs}) - 1,96 \left(\sqrt{\frac{p(\text{obs})q(\text{obs})}{(n-1)}} \right)$$

$$Lim_{\text{sup}} = p(\text{obs}) + 1,96 \left(\sqrt{\frac{p(\text{obs})q(\text{obs})}{(n-1)}} \right)$$

Exercice d'application 4 :

Sur 146 étudiants de maîtrise BPE ayant fourni l'information, 20 étaient gauchers ou ambidextres (soit $p_{obs} = 13,698\%$ et $q_{obs} = 86,302\%$ droitiers).

Quel est l'intervalle de confiance du pourcentage de la catégorie [gauchers & ambidextres] ?

L'intervalle de confiance à 95% ($IC_{95\%}$)

$$Lim_{inf} = p(obs) - 1,96 \left(\sqrt{\frac{p(obs)q(obs)}{(n-1)}} \right)$$

$$Lim_{sup} = p(obs) + 1,96 \left(\sqrt{\frac{p(obs)q(obs)}{(n-1)}} \right)$$

Les statistiques inférentielles

Exercice d'application 4 :

Sur 146 étudiants de maîtrise BPE ayant fourni l'information, 20 étaient gauchers ou ambidextre (soit $p_{obs} = 13,698\%$ et $q_{obs} = 86,302\%$ droitiers).

Quel est l'intervalle de confiance du pourcentage de la catégorie [gauchers & ambidextres] ?

L'intervalle de confiance à 95% ($IC_{95\%}$)

vérification préliminaire : $np = 146 \times 0,13698 = 19,99 > 5$; $nq = 146 \times 0,86302 = 126 > 5$. On a bien $np > 5$ et $nq > 5$

$$Lim_{inf} = p(obs) - 1,96 \left(\sqrt{\frac{p(obs)q(obs)}{(n-1)}} \right)$$

$$Lim_{sup} = p(obs) + 1,96 \left(\sqrt{\frac{p(obs)q(obs)}{(n-1)}} \right)$$

$$\sqrt{[p(obs) q(obs)]/(n-1)} = \sqrt{[(0,13698 \times 0,86302)/(146 - 1)]} = 0,02845$$

$$\text{Borne inf : } 0,13698 - 1,96 \times 0,02845 = 0,081218 \\ = 8,1\%$$

$$\text{Borne sup : } 0,13698 + 1,96 \times 0,02845 = 0,192742 \\ = 19,3\%$$

$$IC_{95\%} : [8,1\% \text{ ----- } 19,3\%]$$

L'intervalle est très large même avec un échantillon conséquent (donc inutile pour les petits échantillons)

MISE EN APPLICATION

Mise en application statistiques inférentielles

Document Exercices TD 2 (Dokeos)

Vous êtes Mathieu Djaballah (quelle chance !). Suite à une enquête par questionnaire sur les Fan Zones de l'Euro 2016 en Ile de France, la ville de Paris aimerait connaître les caractéristiques des spectateurs.

*Elle doute que vous puissiez estimer correctement **l'âge** et la **proportion d'hommes et de femmes** adultes avec seulement quelques centaines de questionnaires pour une fréquentation totale cumulée de près d'un million d'individus.*

Etes-vous capable de lui donner tort ?

Rappels :

Intervalle de confiance à 95%, loi normale

$$Lim_{\text{inf}} = m - 1,96 \frac{s}{\sqrt{n}}$$

$$Lim_{\text{sup}} = m + 1,96 \frac{s}{\sqrt{n}}$$

Intervalle de confiance à 95%, loi binomiale

$$Lim_{\text{inf}} = p(\text{obs}) - 1,96 \left(\sqrt{\frac{p(\text{obs})q(\text{obs})}{(n-1)}} \right)$$

$$Lim_{\text{sup}} = p(\text{obs}) + 1,96 \left(\sqrt{\frac{p(\text{obs})q(\text{obs})}{(n-1)}} \right)$$

Mise en application statistiques inférentielles

Document Exercices TD 2 (Dokeos)

Attention, les calculs ne font pas tout !

Pour que les résultats soient fiables, il fallait s'assurer le plus possible du caractère aléatoire de la passation des questionnaires, ce qui a été fait en :

- ✓ *Multipliant le nombre de jours (pas qu'un seul jour/match)*
- ✓ *Couvrant tout le périmètre de la Fan Zone*
- ✓ *Délivrant des consignes spécifiques aux enquêteurs*
- ✓ *Multipliant le nombre d'enquêteurs (pour éviter qu'un même biais se répète)*
- ✓ *Réalisant un ensemble de tests et corrections*

LES TESTS D'HYPOTHESES

Test d'hypothèses - définition :

- Un intérêt majeur des statistiques est de tester des hypothèses sur nos données afin d'expliquer différents phénomènes.
- Une hypothèse se traduit (de manière très générale) comme une influence supposée d'une (ou plusieurs) variable(s) sur une (ou plusieurs) autre(s). On distingue donc :
 - les **variables dépendantes (VD)** qui subissent l'influence...
 - ... de **variables indépendantes ou explicatives (VI)**

Les tests d'hypothèses : principes généraux

Test d'hypothèses - définition :

- Identifiez les **VD** et les **VI** :
 - L'impact de la **masse salariale** des clubs pros sur leur **classement au championnat**.
 - **L'accueil d'un grand événement sportif** permet-il **d'accroître la pratique sportive** ?
 - **I Am Bolder: A Social Cognitive Examination of Road Race Participant Behavior**.
 - **Points of Attachment** and **Sponsorship Outcomes** in an Individual Sport.

Les tests d'hypothèses : principes généraux

Test d'hypothèses - définition :

- Les VD sont les variables mesurées
- Les VI sont les variables manipulées (parfois on les mesure aussi, notamment dans le cas des corrélations)

Test d'hypothèses - méthode :

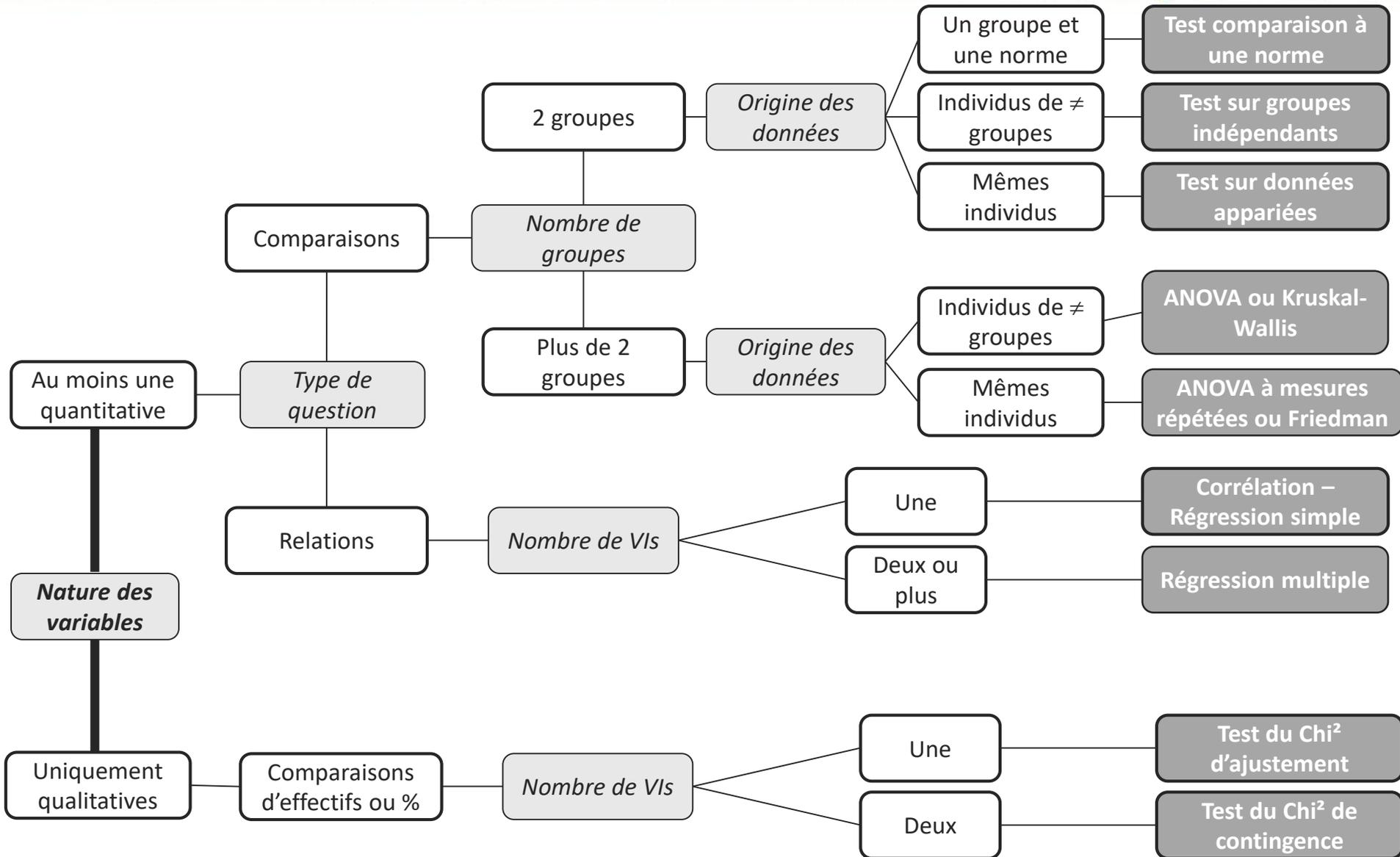
- En statistiques on raisonne à l'inverse de la manière dont on raisonnerait naturellement.
- On ne peut valider notre hypothèse qu'en rejetant l'hypothèse nulle (ou H_0) selon laquelle il n'y a aucune influence des VI sur les VD (par exemple aucune différence entre le chiffre d'affaire des magasins Décathlon d'Ile de France et ceux des autres régions)

Les tests d'hypothèses : principes généraux

Test d'hypothèses - méthode :

- Tout test d'hypothèse dépend :
 - ✓ Du type de question posée (différence entre 2 moyennes ou distributions, différence entre n moyennes ou distributions, liens entre des variables, etc...)
 - ✓ De la nature des variables à analyser (quantitatives, qualitatives ordinales ou nominales, échantillons indépendants ou appariés, etc...)
 - ✓ De leur distribution (loi normale ou pas)
- Il s'agit donc avant tout de **déterminer le test adéquat**

Arbre de décision statistique (simplifié)



Test d'hypothèses - méthode :

- A partir de là, le principe général est toujours le même :
 - On pose notre hypothèse (notée H_1)
 - On calcule le paramètre du test (**valeur calculée**)
 - Qu'on compare à une **valeur critique** sous H_0
 - Cette valeur critique dépend du **risque** qu'on choisit (généralement 5% soit 0,05)
 - Selon que notre **valeur calculée est $>$ ou $<$ à la valeur critique**, on rejette ou non H_0
 - Si **H_0 est rejetée**, notre hypothèse **H_1 est validée**

Risque et significativité statistique :

- Le risque de rejeter H_0 alors qu'elle est vraie (autrement dit que notre hypothèse est fautive) est appelé risque α (ou risque de type 1)
 - Comme énoncé ci-avant, ce risque est choisi à l'avance (généralement 5%)
 - comme il s'agit d'une probabilité (que H_0 soit vraie) on le note « $P = 0,05$ » ou « $.05$ »
 - Il signifie qu'on rejettera H_0 s'il n'y a que 5% de chances qu'elle soit vraie
 - En pratique, plus ce risque est petit, plus le rejet de H_0 est « sûr »
 - Les logiciels fournissent souvent directement la valeur exacte du P , qui est appelée significativité du test

Risque et significativité statistique :

- Mais le risque inverse (celui de ne pas rejeter H_0 alors qu'elle est fausse, autrement dit celui de rejeter notre hypothèse alors qu'elle est vraie bien que notre test dit le contraire) existe aussi. Il se note β (ou risque de type 2)
 - Malheureusement il ne peut pas être calculé la plupart du temps
 - Et cela peut poser problème : par exemple si l'on conclut qu'une substance chimique n'a pas d'effet négatif alors qu'elle en a un !
 - On peut le soupçonner lorsque notre valeur calculée est proche de notre valeur critique bien qu'elle ne la dépasse pas (autrement dit que notre P est proche de 0,05, par exemple 0,06744...)
 - La raison est alors souvent que notre échantillon n'est pas assez grand

Les tests d'hypothèses : principes généraux

Risque et significativité statistique :

	H0 vraie (en réalité)	H0 fausse (en réalité)
H0 rejetée	Risque α	Pas d'erreur
H0 non-rejetée	Pas d'erreur	Risque β

Comparaisons de moyennes

Objectif des comparaisons de moyennes : à quelle(s) question(s) veut-on répondre ?

*Qui dit « moyennes » dit variables quantitatives

Comparer un groupe (échantillon) à une norme :

- La moyenne d'âge des spectateurs de la Fan Zone de Paris est elle la même que celle de la population parisienne ?
- Les magasins Décathlon d'Ile de France génèrent-ils un chiffre d'affaire supérieur à la moyenne nationale ?
- Les recettes match-day des clubs de Ligue 1 sont-elles équivalentes à la moyenne de celles des clubs européens ?

Comparaisons de moyennes

Objectif des comparaisons de moyennes : à quelle(s) question(s) veut-on répondre ?

*Qui dit « moyennes » dit variables quantitatives

Comparer deux groupes (échantillons) indépendants* :

- La moyenne d'âge des spectateurs de la Fan Zone de Paris est elle la même que celle des spectateurs de la Fan Zone de Saint Denis ?
- Les magasins Décathlon d'Ile de France génèrent-ils un chiffre d'affaire supérieur à ceux de la région PACA ?
- Les recettes match-day des clubs de Ligue 1 sont-elles équivalentes à celles des clubs de Bundesliga ?

*groupes indépendants : individus différents / valeurs d'un échantillon n'apportent aucune information concernant celles de l'autre

Comparaisons de moyennes

Objectif des comparaisons de moyennes : à quelle(s) question(s) veut-on répondre ?

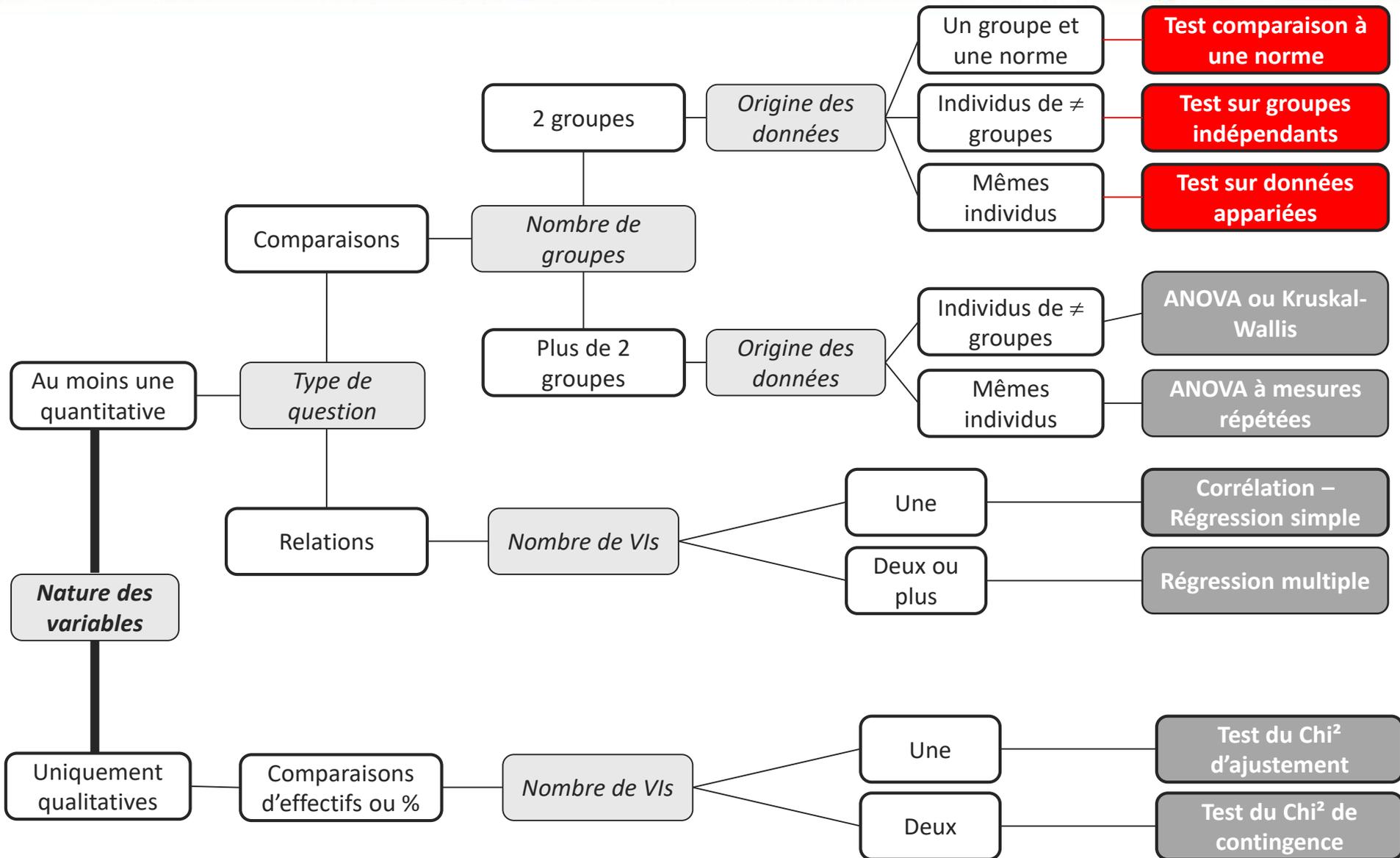
*Qui dit « moyennes » dit variables quantitatives

Comparer deux groupes (échantillons) dépendants ou appariés* :

- L'amélioration de l'expérience stade permet-elle d'accroître les recettes match day des clubs de Ligue 1 ?

***groupes appariés : mêmes individus mesurés plusieurs fois (avant-après)**

Comparaisons de moyennes



Comparaisons de moyennes

Comparaisons d'une moyenne à une norme

Une fois que l'on a choisi le test de comparaison à une norme, reste à déterminer :

- La taille de l'échantillon (grand ou petit)
- La distribution (normale ou non)

		Taille de l'échantillon	
		Grand	Petit
Distribution	Normale	Test Z	t de Student
	Non-normale	Test Z	Bootstrap

Comparaisons de moyennes

Comparaisons d'une moyenne à une norme

Grand échantillon (> 30) : test Z

$$|Z| = \frac{|m - M|}{\sqrt{\frac{s^2}{n}}}$$

- m = moyenne de l'échantillon
- M (parfois μ) = norme (moyenne attendue, par exemple moyenne de la population)
- s^2 = variance de l'échantillon (**rappel la variance est le carré de l'écart type et non pas l'écart type lui-même !**)
- n = taille de l'échantillon
- $| \quad |$ = signifie valeur absolue (on ne prend pas en compte le signe + ou - des valeurs)

Si $|Z| > 1,96$ on rejette H_0 (au risque $\alpha = 0,05$)

Si $|Z| < 1,96$ on ne rejette pas H_0 (au risque β inconnu)

Comparaisons de moyennes

Exercice d'application 1 :

Bénéfices des magasins décathlon en Ile de France (47 magasins)

Moyenne observée : $m = 351,46$ K€

Écart type estimée : $s = 112,51$ K€ (il faudra le mettre au carré pour le calcul !)

La moyenne de l'ensemble des magasins en France est $M = 317,84$ K€

Peut-on affirmer que les magasins d'Ile de France ont en moyenne une rentabilité différente des autres magasins de France ?

H0 : la rentabilité n'est pas différente

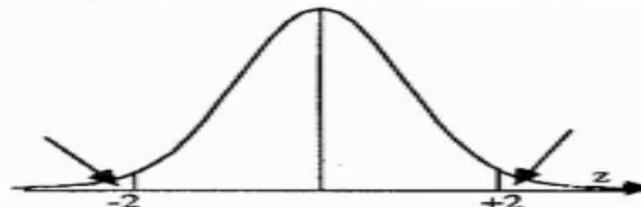
H1 : la rentabilité est significativement différente

$$|z| = \frac{|m - M|}{\sqrt{\frac{s^2}{n}}} \Rightarrow |z| = \frac{|351,46 - 317,84|}{\sqrt{\frac{112,51^2}{47}}} = 2,048$$

2,048 > 1,96 : on peut rejeter H0. La moyenne est significativement différente de la norme à P = 0,05

Comparaisons de moyennes

Table z, bilatérale externe cumulée



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	1.00	.992	.984	.976	.968	.960	.952	.944	.936	.928
0.1	.920	.912	.904	.897	.887	.881	.873	.865	.857	.849
0.2	.841	.834	.826	.818	.810	.803	.795	.787	.779	.772
0.3	.764	.757	.749	.741	.734	.726	.719	.711	.704	.697
0.4	.689	.682	.674	.667	.660	.653	.645	.638	.631	.624
0.5	.617	.610	.603	.596	.589	.582	.575	.569	.562	.555
0.6	.549	.542	.535	.529	.522	.516	.509	.503	.497	.490
0.7	.484	.478	.472	.465	.459	.453	.447	.441	.435	.430
0.8	.424	.418	.412	.407	.401	.395	.390	.384	.379	.373
0.9	.368	.363	.358	.352	.347	.342	.337	.332	.327	.322
1.0	.317	.313	.308	.303	.298	.294	.289	.285	.280	.276
1.1	.271	.267	.263	.258	.254	.250	.246	.242	.238	.234
1.2	.230	.226	.222	.219	.215	.211	.208	.204	.201	.197
1.3	.194	.190	.187	.184	.180	.177	.174	.171	.168	.165
1.4	.162	.159	.156	.153	.150	.147	.144	.142	.139	.136
1.5	.134	.131	.129	.126	.124	.121	.119	.116	.114	.112
1.6	.110	.107	.105	.103	.101	.100	.097	.095	.093	.091
1.7	.089	.087	.085	.084	.082	.080	.078	.077	.075	.073
1.8	.072	.070	.069	.067	.066	.064	.063	.062	.060	.059
1.9	.057	.056	.055	.054	.052	.051	.050	.049	.048	.047
2.0	.046	.044	.043	.042	.041	.040	.039	.038	.038	.037
2.1	.035	.035	.034	.033	.032	.032	.031	.030	.029	.028
2.2	.028	.027	.026	.026	.025	.024	.024	.023	.023	.022
2.3	.021	.021	.020	.020	.019	.019	.018	.018	.017	.017
2.4	.016	.016	.016	.015	.015	.014	.014	.014	.013	.013
2.5	.012	.012	.012	.011	.011	.011	.010	.010	.010	.010
2.6	.009	.009	.009	.009	.008	.008	.008	.008	.007	.007
2.7	.007	.007	.007	.006	.006	.006	.006	.006	.005	.005
2.8	.005	.005	.005	.005	.004	.004	.004	.004	.004	.004
2.9	.004	.004	.004	.003	.003	.003	.003	.003	.003	.003
3.0	.003	.002	.001	-						

Comparaisons de moyennes

Comparaisons d'une moyenne à une norme

Petit échantillon (< 30) : t de Student général

$$|t| = \frac{|m - M|}{\sqrt{\frac{s^2}{n}}}$$

- m = moyenne de l'échantillon
- M (parfois μ) = norme (moyenne attendue, par exemple moyenne de la population)
- s^2 = variance de l'échantillon (*rappel la variance est le carré de l'écart type et non pas l'écart type lui-même !*)
- n = taille de l'échantillon
- $| |$ = signifie valeur absolue (on ne prend pas en compte le signe + ou - des valeurs)

On compare le $|t|$ calculé au $|t|$ lu sur la table du t de Student (au ddl correspondant, c'est-à-dire $n-1$)

- Si $|t|$ calculé $>$ $|t|$ lu (valeur critique de la table t) on rejette H_0
- Si $|t|$ calculé $<$ $|t|$ lu on ne rejette pas H_0

Comparaisons de moyennes

Exercice d'application 2 :

Bénéfices des magasins Décathlon en Ile de France (26 magasins)

Moyenne observée : $m = 351,46$ K€

Écart type estimée : $s = 112,51$ K€ (il faudra le mettre au carré pour le calcul !)

La moyenne de l'ensemble des magasins en France est $M = 317,84$ K€

Peut-on affirmer que les magasins d'Idf ont en moyenne une rentabilité différente des autres magasins de France ?

H_0 : la rentabilité n'est pas différente

H_1 : la rentabilité est significativement différente

$$|t| = \frac{|m - M|}{\sqrt{\frac{s^2}{n}}} \quad \rightarrow \quad |t| = \frac{|351,46 - 317,84|}{\sqrt{\frac{112,51^2}{26}}} = 1,523$$

Ddl = 25 (26 - 1) \rightarrow $|t|_{lu} = 2,06$ (à $P = 0,05$)

$1,523 < 2,06$: on ne peut pas rejeter H_0 . La moyenne n'est pas significativement différente \rightarrow on dira « nos données ne nous permettent pas d'établir une différence significative »

Comparaisons de moyennes

- Je peux tout de même regarder si mon test, non-significatif à $P=0,05$, ne le serait pas avec un risque d'erreur plus grand ($P=0,10$ ou 10% - il s'agit quand-même d'un gros risque !).
- Ce n'est pas le cas : le $|t|$ calculé reste supérieur au $|t|$ lu (1,71)
- Mais je peux tout de même voir que je ne suis pas extrêmement loin, ce qui devrait au moins m'inciter à refaire le test avec un échantillon plus grand.

		<i>t de Student</i>				
		P	.10	.05	.02	.01
d.d.l.						
1		6,31	12,71	31,82	63,66	
2		2,92	4,30	6,97	9,93	
3		2,35	3,18	4,54	5,84	
4		2,13	2,78	3,75	4,60	
5		2,02	2,57	3,37	4,03	
6		1,94	2,45	3,14	3,71	
7		1,90	2,37	3,00	3,5	
8		1,86	2,31	2,90	3,36	
9		1,83	2,26	2,82	3,25	
10		1,81	2,23	2,76	3,17	
11		1,80	2,2	2,72	3,11	
12		1,78	2,18	2,68	3,06	
13		1,77	2,16	2,65	3,01	
14		1,76	2,14	2,62	2,98	
15		1,75	2,13	2,6	2,95	
16		1,75	2,12	2,58	2,92	
17		1,74	2,11	2,57	2,90	
18		1,73	2,10	2,55	2,88	
19		1,73	2,09	2,54	2,86	
20		1,73	2,09	2,53	2,85	
21		1,72	2,08	2,52	2,83	
22		1,72	2,07	2,51	2,82	
23		1,71	2,07	2,50	2,81	
24		1,71	2,06	2,49	2,80	
25		1,71	2,06	2,49	2,79	
26		1,71	2,06	2,48	2,78	
27		1,70	2,05	2,47	2,77	
28		1,70	2,05	2,47	2,76	
29		1,70	2,04	2,46	2,76	
30		1,70	2,04	2,46	2,75	
40		1,68	2,02	2,42	2,70	
60		1,67	2,00	2,39	2,66	
Infini		1,65	1,96	2,33	2,58	

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Une fois que l'on a choisi le test de comparaison sur groupes indépendants, reste à déterminer :

- La taille des échantillons (grands ou petits)
- Leurs distributions (normales ou non)

		Taille des échantillons	
		Les 2 sont grands	Au moins 1 est petit
Distribution	Normale	Test Z sur échantillons indépendants	t de Student sur échantillons indépendants (avec pondération des variances)
	Non-normale pour le petit échantillon		Test de Mann-Whitney

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

2 grands échantillons (> 30) : test Z sur échantillons indépendants

$$Z = \frac{m1 - m2}{\sqrt{\frac{s1^2}{n1} + \frac{s2^2}{n2}}}$$

- $m1$ = moyenne de l'échantillon 1 (idem $m2$)
- $s1^2$ = variance échantillon 1 (idem $s2$) *rappel la variance est le carré de l'écart type !*
- $n1$ = taille de l'échantillon 1 (idem $n2$)

Si $|Z| > 1,96$ on rejette H_0 (au risque $\alpha = 0,05$)

Si $|Z| < 1,96$ on ne rejette pas H_0 (au risque β inconnu)

Comparaisons de moyennes

Exercice d'application 3 :

Spectateurs de la Fan Zone de Paris ($n_1 = 497$ spectateurs interrogés)

Moyenne d'âge observée : $m_1 = 25,729$

Écart type estimé : $s_1 = 8,303$

Spectateurs de la Fan Zone de Saint-Denis ($n_2 = 452$ spectateurs interrogés)

Moyenne d'âge observée : $m_2 = 36,153$

Écart type estimé : $s_2 = 12,326$

Les moyennes d'âge des spectateurs de Paris et de Saint-Denis sont-elles significativement différentes ?

H0 : les moyennes ne sont pas différentes

H1 : les moyennes sont significativement différentes

$$z = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \Rightarrow \quad z = \frac{25,729 - 36,153}{\sqrt{\frac{8,303^2}{497} + \frac{12,326^2}{452}}} = |15,127|$$

15,127 > 1,96 : on peut rejeter H0. Les moyennes d'âge sont significativement différentes -> et même très très significativement (P= de l'ordre de 0,0000000001, On a une chance sur plus d'un milliard de se tromper !)

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable suit une loi normale : t de Student sur échantillons indépendants

Une étape préalable : pondérer les deux variances

$$S^2 = \frac{s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)}{n_1 + n_2 - 2}$$

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable suit une loi normale : t de Student sur échantillons indépendants

$$t = \frac{m1 - m2}{\sqrt{\frac{s^2}{n1} + \frac{s^2}{n2}}}$$

- $m1$ = moyenne de l'échantillon 1 (idem $m2$)
- s^2 = *variance pondérée*
- $n1$ = taille de l'échantillon 1 (idem $n2$)

- Si $|t|$ calculé > $|t|$ lu (valeur critique de la table t) on rejette H_0
- Si $|t|$ calculé < $|t|$ lu on ne rejette pas H_0

Comparaisons de moyennes

Exercice d'application 4 :

Spectateurs de la Fan Zone de Paris ($n_1 = 26$ spectateurs interrogés)

Moyenne d'âge observée : $m_1 = 25,729$

Écart type estimé : $s_1 = 8,303$

Spectateurs de la Fan Zone de Saint-Denis ($n_2 = 28$ spectateurs interrogés)

Moyenne d'âge observée : $m_2 = 36,153$

Écart type estimé : $s_2 = 12,326$

Les moyennes d'âge des spectateurs de Paris et de Saint-Denis sont-elles significativement différentes ?

H_0 : les moyennes ne sont pas différentes

H_1 : les moyennes sont significativement différentes

Pondération des variances

$$s^2 = \frac{s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)}{n_1 + n_2 - 2} \quad \Rightarrow \quad S^2 = \frac{8,303^2 (26 - 1) + 12,326^2 (28 - 1)}{26 + 28 - 2} = 112,031$$

Attention pour la suite : s^2 est la variance et non l'écart type, donc pas besoin de le mettre au carré dans la formule du t !

Comparaisons de moyennes

Exercice d'application :

Spectateurs de la Fan Zone de Paris ($n_1 = 26$ spectateurs interrogés)

Moyenne d'âge observée : $m_1 = 25,729$

Écart type estimé : $s_1 = 8,303$

Spectateurs de la Fan Zone de Saint-Denis ($n_2 = 28$ spectateurs interrogés)

Moyenne d'âge observée : $m_2 = 36,153$

Écart type estimé : $s_2 = 12,326$

Les moyennes d'âge des spectateurs de Paris et de Saint-Denis sont-elles significativement différentes ?

H_0 : les moyennes ne sont pas différentes

H_1 : les moyennes sont significativement différentes

$$t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$



$$t = \frac{25,729 - 36,153}{\sqrt{\frac{112,031}{26} + \frac{112,031}{28}}} = |3,616|$$

Pas besoin de mettre au carré, c'est déjà la variance !

Ddl = - 1 à chaque échantillon = $(26 - 1) + (28 - 1) = 52 \rightarrow |t|_{lu} = 2,02$ (à $P = 0,05$)

$3,616 > 2,02$: on peut rejeter H_0 . Les moyennes d'âge restent significativement différentes même sur ces petits échantillons.

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable ne suit pas une loi normale : test de Mann-Whitney

- Aussi appelé Wilcoxon-Mann-Whitney
- Ne pas confondre avec le test des rangs signés de Wilcoxon
- C'est un test non-paramétrique (n'utilise pas les paramètres que sont la moyenne et la variance)

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable ne suit pas une loi normale : test de Mann-Whitney

Prenons 2 échantillons A (d'effectif $X_1, X_2 \dots X_{nA}$) et B (d'effectif $Y_1, Y_2 \dots Y_{nB}$)

Principe : si les individus proviennent de la même population (hypothèse H_0) alors la probabilité qu'un X pris au hasard soit supérieur à un Y pris au hasard est de 0,5 (une chance sur deux).

Si en revanche la moyenne est plus élevée dans A que dans B, on aura plus souvent $x > y$ ou $y > x$ (et, si on réalise un classement on trouvera respectivement les x et les y plutôt en haut ou en bas du classement .

➤ **Il va donc s'agir de comparer les situations où $x > y$ et celles où les $y > x$**

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable ne suit pas une loi normale : test de Mann-Whitney

1^{ère} étape : on compte, pour chaque x , le nombre de y qui lui sont inférieurs et on somme les résultats obtenus. On fait la même chose avec les y .
Lorsqu'un x et un y sont égaux, on ne compte que 0,5 pour chacun

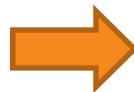
Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable **ne suit pas** une loi normale : test de Mann-Whitney

Situation de départ

A	B
x1 = 27	Y1 = 23
x2 = 25	Y2 = 27
x3 = 21	Y3 = 22
x4 = 32	
x5 = 27	
x6 = 24	



classement

A	B
x4 = 32	Y2 = 27
x1 = 27	Y1 = 23
x5 = 27	Y3 = 22
x2 = 25	
x6 = 24	
x3 = 21	

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable **ne suit pas** une loi normale : test de Mann-Whitney

classement

Points	Uxy	A	B	Uyx	Points
6	3 pts	x4 = 31	Y2 = 25	1,5 pts	3
5	3 pts	x1 = 28	Y1 = 23	1 pts	2
4	3 pts	x5 = 28	Y3 = 22	1 pts	1
3	3 pts	x2 = 26			
2	2,5 pts	x6 = 25			
1	0 pts	x3 = 21			
	14,5 pts			3,5 pts	

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable ne suit pas une loi normale : test de Mann-Whitney

2^{ème} étape : on sélectionne la plus petite des deux valeurs entre U_{xy} et U_{yx} (dans notre cas il s'agit de U_{yx} qui est égal à 3,5 (case en bleu ci-avant))

Et on se réfère à la table du U de Mann-Whitney (à 5%)

Attention (désolé, ce n'est pas de ma faute !) : cette table est présentée « à l'envers » par rapport à la logique des autres tables (loi normale, t de student, etc...).

- Le test est significatif si la valeur calculée est **inférieure** à la valeur de la table (et non supérieure)

Table 3 Critical values of U (5% significance).

$n_1 \backslash n_2$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2								0	0	0	0	1	1	1	1	1	2	2	2	2
3					0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4				0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5			0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6			1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7			1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8		0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9		0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10		0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11		0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12		1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13		1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14		1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15		1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16		1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17		2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18		2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19		2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20		2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Comparaisons de moyennes

Comparaisons de deux moyennes observées (indépendantes)

Au moins 1 petit échantillon (< 30) mais la variable ne suit pas une loi normale : test de Mann-Whitney

Dans notre cas $U(\text{calculé}) > U(\text{table})$ -> On ne peut pas rejeter H_0

« La différence est non significative : les populations dont ces échantillons sont tirés ont la même moyenne »

Comparaisons de moyennes

Exercice d'application 5 :

Les recettes match-day des clubs de ligue 1 et de Premiere League sont elles équivalentes ?

Ligue 1	
Paris SG	33
AS Monaco	20
Lille OSC	11
AS Saint-Etienne	11
Olympique Lyonnais	9
Olympique de Marseille	8
FC Girondins de Bordeaux	7
FC Lorient	6
Toulouse FC	6
SC Bastia	5
Stade de Reims	5
Stade Rennais FC	5
FC Nantes	4
Evian TG FC	4
Montpellier HSC	3
EA Guigamp	3
OGC Nice	3
FC Metz	2
RC Lens	2
SM Caen	2

Premiere League	
Chelsea	71
Man City	43
Arsenal	100
Manchester United	91
Tottenham	41
Liverpool	59
Southampton	18
Swansea	8
Stoke City	8
Crystal Palace	10
Everton	18
West Ham	20
WBA	8
Leicester	11
Newcastle	27
Sunderland	11
Aston Villa	14
Hull City	7
Burnley	6
QPR	8

Comparaisons de moyennes

Exercice d'application 5 :

Les recettes match-day des clubs de ligue 1 et de Premiere League sont elles comparables ?

	Uxy	Ligue 1	miere leag	Uyx	
20	14	33	100	20	20
19	12,5	20	91	20	19
18	8	11	71	20	18
17	8	11	59	20	17
16	6	9	43	20	16
15	4	8	41	20	15
14	1,5	7	27	19	14
13	0,5	6	20	18,5	13
12	0,5	6	18	18	12
11	0	5	18	18	11
10	0	5	14	18	10
9	0	5	11	17	9
8	0	4	11	17	8
7	0	4	10	16	7
6	0	3	8	14,5	6
5	0	3	8	14,5	5
4	0	3	8	14,5	4
3	0	2	8	14,5	3
2	0	2	7	13,5	2
1	0	2	6	12	1
	55			345	

$U(\text{calculé}) = 55$

$U(\text{table}) = 127$

- $U(\text{calculé}) < U(\text{table})$
- On peut rejeter H_0
- Les moyennes sont significativement différentes

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

Une fois que l'on a choisi le test de comparaison sur groupes appariés, reste à déterminer :

- La taille de l'échantillon (grand ou petit)
- La distribution (normale ou non)

		Taille de l'échantillon	
		Grand (>30)	Petit (<30)
Distribution	Normale	t de Student pour échantillons appariés	t de Student pour échantillon appariés
	Non-normale	En fonction de la possibilité de quantifier les différences : - Test des signes (si non) - Test des rangs signés de Wilcoxon (si oui)	En fonction de la possibilité de quantifier les différences : - Test des signes (si non) - Test des rangs signés de Wilcoxon (si oui)

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi normale : t de Student sur échantillons appariés

1^{ère} étape :

- Calculer la variable d = différence entre X_2 (« après ») et X_1 (« avant »)

- $x_{11} - x_{21} = ?$
- $x_{12} - x_{22} = ?$
- $x_{13} - x_{23} = ?$
-
- $x_{1n} - x_{2n} = ?$



X1	X2	d
avant	après	différence
3	6	3
8	14	6
4	8	4
6	4	-2
9	16	7
2	7	5
12	19	7

md = moyenne de la variable d

sd = écart type de la variable d

n = taille de l'échantillon

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

2 grands échantillons (> 30) + loi normale : t de Student sur échantillons appariés

1^{ère} étape :

- Calculer la variable d = différence entre X_2 (« après ») et X_1 (« avant »)

- $x_{11} - x_{21} = ?$
- $x_{12} - x_{22} = ?$
- $x_{13} - x_{23} = ?$
-
- $x_{1n} - x_{2n} = ?$



X1	X2	d
avant	après	différence
3	6	3
8	14	6
4	8	4
6	4	-2
9	16	7
2	7	5
12	19	7

$$md = 4,28$$

$$sd = 3,15$$

$$n = 7$$

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

2 grands échantillons (> 30) + loi normale : t de Student sur échantillons appariés

2^{ème} étape :

$$t = \frac{md}{\left(\frac{sd}{\sqrt{n}}\right)}$$

- md = moyenne de la variable d
- sd = écart type
- n = taille de l'échantillon

- Si $|t|$ calculé $>$ $|t|$ lu (valeur critique de la table t) on rejette H_0
- Si $|t|$ calculé $<$ $|t|$ lu on ne rejette pas H_0
- **Attention dans ce test ddl = n-1 (car les 2 échantillons sont composés des mêmes individus)**

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

2 grands échantillons (> 30) + loi normale : t de Student sur échantillons appariés

2^{ème} étape :

$$t = \frac{4,28}{\left(\frac{3,15}{\sqrt{7}}\right)}$$

- md = moyenne de la variable d
- sd = écart type
- n = taille de l'échantillon

- t calculé = 3,60
- t lu ($ddl = n-1$ donc 6) = 2,45 (à 0,05)
- Je peux rejeter H_0 et valider H_1
- Je peux même aller jusqu'à une significativité de 0,02 (t lu = 3,14)
- Mais pas jusqu'à 0,01 : t lu = 3,71 (devient supérieur à mon t calculé).

Comparaisons de moyennes

d.d.l.	t de Student			
	P	.10	.05	.02
1	6,31	12,71	31,82	63,66
2	2,92	4,30	6,97	9,93
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,02	2,57	3,37	4,03
6	1,94	2,45	3,14	3,71
7	1,90	2,37	3,00	3,5
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
11	1,80	2,2	2,72	3,11
12	1,78	2,18	2,68	3,06
13	1,77	2,16	2,65	3,01
14	1,76	2,14	2,62	2,98
15	1,75	2,13	2,6	2,95
16	1,75	2,12	2,58	2,92
17	1,74	2,11	2,57	2,90
18	1,73	2,10	2,55	2,88
19	1,73	2,09	2,54	2,86
20	1,73	2,09	2,53	2,85
21	1,72	2,08	2,52	2,83
22	1,72	2,07	2,51	2,82
23	1,71	2,07	2,50	2,81
24	1,71	2,06	2,49	2,80
25	1,71	2,06	2,49	2,79
26	1,71	2,06	2,48	2,78
27	1,70	2,05	2,47	2,77
28	1,70	2,05	2,47	2,76
29	1,70	2,04	2,46	2,76
30	1,70	2,04	2,46	2,75
40	1,68	2,02	2,42	2,70
60	1,67	2,00	2,39	2,66
Infini	1,65	1,96	2,33	2,58

Comparaisons de moyennes

Exercice d'application :



L'application Running Heroes incite-t-elle les runners à parcourir davantage de km ?

	Km parcourus par semaine	
	Avant utilisation	Après utilisation
indiv 1	11,1	14,6
indiv 2	16,2	14,3
indiv 3	5,1	10,5
indiv 4	9,2	10,4
indiv 5	12,4	11,6
indiv 6	3,5	7
indiv 7	5,1	4,3
indiv 8	14	16,7
indiv 9	21	19,7
indiv 10	9,8	10,2
indiv 11	19	20,2
indiv 12	8,5	12,3
indiv 13	10	13,3
indiv 14	15,2	14,7
indiv 15	20,5	18,2
indiv 16	4,5	6,3
indiv 17	11,2	13,1
indiv 18	13,2	16,1
indiv 19	18,1	16,7
indiv 20	2,2	4,5
indiv 21	9,2	10,3
indiv 22	5,6	3,8
indiv 23	21,1	21,1
indiv 24	7,2	9,3
indiv 25	11,4	8,2
indiv 26	5,8	9,1
indiv 27	11,3	12,7
indiv 28	16,2	14,2

Comparaisons de moyennes

Exercice d'application :



L'application Running Heroes incite-t-elle les runners à parcourir davantage de km ?

	Km parcourus par semaine		d
	Avant utilisation	Après utilisation	
indiv 1	11,1	14,6	3,5
indiv 2	16,2	14,3	-1,9
indiv 3	5,1	10,5	5,4
indiv 4	9,2	10,4	1,2
indiv 5	12,4	11,6	-0,8
indiv 6	3,5	7	3,5
indiv 7	5,1	4,3	-0,8
indiv 8	14	16,7	2,7
indiv 9	21	19,7	-1,3
indiv 10	9,8	10,2	0,4
indiv 11	19	20,2	1,2
indiv 12	8,5	12,3	3,8
indiv 13	10	13,3	3,3
indiv 14	15,2	14,7	-0,5
indiv 15	20,5	18,2	-2,3
indiv 16	4,5	6,3	1,8
indiv 17	11,2	13,1	1,9
indiv 18	13,2	16,1	2,9
indiv 19	18,1	16,7	-1,4
indiv 20	2,2	4,5	2,3
indiv 21	9,2	10,3	1,1
indiv 22	5,6	3,8	-1,8
indiv 23	21,1	21,1	0
indiv 24	7,2	9,3	2,1
indiv 25	11,4	8,2	-3,2
indiv 26	5,8	9,1	3,3
indiv 27	11,3	12,7	1,4
indiv 28	16,2	14,2	-2
moyennes	11,34285714	12,26428571	
		md	0,92142857
		sd	2,24538327
		n	28
		t	2,17145187

Comparaisons de moyennes

- t calculé = 2,17
- t (lu) à 0,05 = 2,05
- Je peux rejeter H_0 : l'application Running Heroes incite les runners à parcourir davantage de km
- Mais cet effet semble limité (à 0,02 mon t calculé passe en dessous du t lu)

		<i>t de Student</i>				
		P	.10	.05	.02	.01
d.d.l.						
1		6,31	12,71	31,82	63,66	
2		2,92	4,30	6,97	9,93	
3		2,35	3,18	4,54	5,84	
4		2,13	2,78	3,75	4,60	
5		2,02	2,57	3,37	4,03	
6		1,94	2,45	3,14	3,71	
7		1,90	2,37	3,00	3,5	
8		1,86	2,31	2,90	3,36	
9		1,83	2,26	2,82	3,25	
10		1,81	2,23	2,76	3,17	
11		1,80	2,2	2,72	3,11	
12		1,78	2,18	2,68	3,06	
13		1,77	2,16	2,65	3,01	
14		1,76	2,14	2,62	2,98	
15		1,75	2,13	2,6	2,95	
16		1,75	2,12	2,58	2,92	
17		1,74	2,11	2,57	2,90	
18		1,73	2,10	2,55	2,88	
19		1,73	2,09	2,54	2,86	
20		1,73	2,09	2,53	2,85	
21		1,72	2,08	2,52	2,83	
22		1,72	2,07	2,51	2,82	
23		1,71	2,07	2,50	2,81	
24		1,71	2,06	2,49	2,80	
25		1,71	2,06	2,49	2,79	
26		1,71	2,06	2,48	2,78	
27		1,70	2,05	2,47	2,77	
28		1,70	2,05	2,47	2,76	
29		1,70	2,04	2,46	2,76	
30		1,70	2,04	2,46	2,75	
40		1,68	2,02	2,42	2,70	
60		1,67	2,00	2,39	2,66	
Infini		1,65	1,96	2,33	2,58	

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + impossibilité de quantifier la différence :

test des signes

- Test non-paramétrique
- Tire son nom du fait qu'il utilise les signes + et -, au lieu de données quantitatives.
- Particulièrement utile dans les cas où il n'est pas possible de mesurer la différence précise entre « avant » et « après » (on sait juste que « avant » est supérieur ou inférieur à « après »)
- Sert donc notamment pour les variables qualitatives nominales

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + impossibilité de quantifier la différence :

test des signes

X1	X2
avant	après
.	Plus
.	Moins
.	Moins
.	Moins
.	Identique
.	Plus
.	Moins

Principe :

On a simplement besoin de :

- x = nombre de signes les moins fréquents
- n = taille de l'échantillon
- Les différences nulles (« identique ») ne sont pas prises en compte ($n - \text{différences nulles}$)

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + impossibilité de quantifier la différence :

test des signes **cas des petits échantillons**

Moins	4
Plus	16
Identique	25

Exemple :

- $x = 4$
- $n = 4 + 16$ (on retire les 25 identiques) = 20
- Je suis donc en présence d'un petit échantillon (alors qu'au départ il semblait grand ($4+16+25 = 45$))
- Je lis directement le résultat dans la table du test binomial

Comparaisons de moyennes

- Ici la significativité est de 0,006
- Donc bien au-delà de 0,05
- Je peux valider l'hypothèse selon laquelle le score de mes individus est plus élevé (puisque davantage de « plus ») après qu'avant.

Table 3 des valeurs critiques du test binomial

x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
N																	
5	031	188	500	812	969	*											
6	016	109	344	656	891	984	*										
7	008	062	227	500	773	938	992	*									
8	004	035	145	363	637	855	965	996	*								
9	002	020	090	254	500	746	910	980	998	*							
10	001	011	055	172	377	623	828	945	989	999	*						
11		006	033	113	274	500	726	887	967	994	*	*					
12		003	019	073	194	387	613	806	927	981	997	*	*				
13		002	011	046	133	291	500	709	867	954	989	998	*	*			
14		001	006	029	090	212	395	605	788	910	971	994	999	*	*		
15			004	018	069	151	304	500	696	849	941	982	996	*	*	*	
16			002	011	058	105	227	402	598	773	895	962	989	998	*	*	
17			001	006	045	072	166	315	500	685	834	928	975	994	999	*	
18			001	004	045	048	119	240	407	593	760	881	952	985	996	999	
19				002	040	032	084	180	324	500	676	820	916	968	990	998	
20					001	006	021	058	132	252	412	588	748	868	942	979	994
21				001	004	013	039	095	192	332	500	668	808	905	961	987	
22					002	008	026	067	143	262	416	584	738	857	933	974	
23					001	005	017	047	105	202	339	500	661	798	895	953	
24					001	003	011	032	076	154	271	419	581	729	846	924	
25						002	007	022	054	115	212	345	500	655	788	885	

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + impossibilité de quantifier la différence :

test des signes **cas des grands échantillons**

X1	X2
avant	après
.	Plus
.	Moins
.	Moins
.	Moins
.	Identique
.	Plus
.	Moins

Lorsque mon échantillon > 30, je peux réaliser une approximation normale :

$$z = \frac{(x + 0,5) - 1/2n}{1/2\sqrt{n}}$$

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + impossibilité de quantifier la différence :

test des signes **cas des grands échantillons**

Moins	9
Plus	43
Identique	7

Exemple :

- $x = 9$
- $n = 9 + 43$ (on retire les 7 identiques) = 52

$$z = \frac{(9 + 0,5) - 1/2(52)}{1/2\sqrt{52}} = -4,576$$

$|4,576| > 1,96$: on peut rejeter H_0 .

Le score « après » est significativement plus élevé (davantage de « plus ») que le score « avant »

Comparaisons de moyennes

Exercice d'application :

Suite à une étude sur l'image du Crédit Agricole passée auprès de 1000 clients (positive, neutre, négative), on les réinterroge dans les semaines qui suivent l'affaire Knysna lors de la Coupe du Monde 2010 de la FIFA. Quel a été l'impact de l'événement sur l'image du sponsor ?

	Avant Knysna	Après Knysna
Client 1	positive	négative
Client 2	neutre	positive
Client 3	négative	négative
Client 4	négative	positive
Client 5	positive	positive
Client 6	positive	neutre
Client 7	neutre	neutre
Client 7	négative	neutre
Client 8	neutre	négative
...
Client 1000	neutre	négative



Dégradation	287
Pas de changement	535
Amélioration	178

Campagne de contre-communication du Crédit Agricole sur les valeurs du football amateur

Comparaisons de moyennes

Exercice d'application :

Suite à une étude sur l'image du Crédit Agricole passée auprès de 1000 clients (positive, neutre, négative), on les réinterroge dans les semaines qui suivent l'affaire Knysna lors de la Coupe du Monde 2010 de la FIFA. Quel a été l'impact de l'événement sur l'image du sponsor ?

*Pourquoi peut-on dire « dégradé » ? Car il y a davantage de cas de dégradation que d'amélioration

Dégradation	287
Pas de changement	535
Amélioration	178

- $x = 178$
- $n = 178 + 287 = 465$

$$z = \frac{(178 + 0,5) - 1/2(465)}{1/2\sqrt{465}} = -5,008$$

$|5,008| > 1,96$: on peut rejeter H_0 .

L'affaire Knysna a significativement **dégradé l'image*** du Crédit Agricole auprès des clients, et ce malgré la campagne de contre-communication.

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

- A utiliser dans les cas où il est possible de mesurer la différence précise entre « avant » et « après »
- Sert donc pour les variables quantitatives (ne suivant pas une loi normale) ainsi que les qualitatives ordinales

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après
Indiv 1	82	63
Indiv 2	69	42
Indiv 3	73	74
Indiv 4	43	37
Indiv 5	58	51
Indiv 6	56	69
Indiv 7	76	80
Indiv 8	65	62

Principe :

- On calcule les différences d_i entre après/avant
- On range ces différences sans tenir compte du signe
 - Les différences nulles (0) sont abandonnées
 - Pour les différences égales entre plusieurs individus (ex : si plusieurs individus ont une différence de -2 entre avant et après) : On leur donne le même rang, à savoir la moyenne des rangs qu'auraient eu ces individus si leurs différences n'étaient pas égales
- On réaffecte les signes

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di
Indiv 1	82	63	-19
Indiv 2	69	42	-27
Indiv 3	73	74	1
Indiv 4	43	47	4
Indiv 5	58	62	4
Indiv 6	61	61	0
Indiv 7	56	69	13
Indiv 8	76	80	4
Indiv 9	65	62	-3

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di
Indiv 1	82	63	-19
Indiv 2	69	42	-27
Indiv 3	73	74	1
Indiv 4	43	47	4
Indiv 5	58	62	4
Indiv 6	61	61	0
Indiv 7	56	69	13
Indiv 8	76	80	4
Indiv 9	65	62	-3

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe
Indiv 1	82	63	-19	
Indiv 2	69	42	-27	
Indiv 3	73	74	1	1
Indiv 4	43	47	4	
Indiv 5	58	62	4	
Indiv 6	61	61	0	
Indiv 7	56	69	13	
Indiv 8	76	80	4	
Indiv 9	65	62	-3	

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe
Indiv 1	82	63	-19	
Indiv 2	69	42	-27	
Indiv 3	73	74	1	1
Indiv 4	43	47	4	
Indiv 5	58	62	4	
Indiv 6	61	61	0	
Indiv 7	56	69	13	
Indiv 8	76	80	4	
Indiv 9	65	62	-3	2

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe
Indiv 1	82	63	-19	
Indiv 2	69	42	-27	
Indiv 3	73	74	1	1
Indiv 4	43	47	4	3
Indiv 5	58	62	4	4
Indiv 6	61	61	0	
Indiv 7	56	69	13	
Indiv 8	76	80	4	5
Indiv 9	65	62	-3	2



Moyenne de
 $3+4+5 = 4$

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe
Indiv 1	82	63	-19	
Indiv 2	69	42	-27	
Indiv 3	73	74	1	1
Indiv 4	43	47	4	4
Indiv 5	58	62	4	4
Indiv 6	61	61	0	
Indiv 7	56	69	13	
Indiv 8	76	80	4	4
Indiv 9	65	62	-3	2

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe
Indiv 1	82	63	-19	
Indiv 2	69	42	-27	
Indiv 3	73	74	1	1
Indiv 4	43	47	4	4
Indiv 5	58	62	4	4
Indiv 6	61	61	0	
Indiv 7	56	69	13	6
Indiv 8	76	80	4	4
Indiv 9	65	62	-3	2



Je reprends à
partir de 5

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe
Indiv 1	82	63	-19	
Indiv 2	69	42	-27	
Indiv 3	73	74	1	1
Indiv 4	43	47	4	4
Indiv 5	58	62	4	4
Indiv 6	61	61	0	
Indiv 7	56	69	13	6
Indiv 8	76	80	4	4
Indiv 9	65	62	-3	2

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe
Indiv 1	82	63	-19	7
Indiv 2	69	42	-27	
Indiv 3	73	74	1	1
Indiv 4	43	47	4	4
Indiv 5	58	62	4	4
Indiv 6	61	61	0	
Indiv 7	56	69	13	6
Indiv 8	76	80	4	4
Indiv 9	65	62	-3	2

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe
Indiv 1	82	63	-19	7
Indiv 2	69	42	-27	8
Indiv 3	73	74	1	1
Indiv 4	43	47	4	4
Indiv 5	58	62	4	4
Indiv 6	61	61	0	
Indiv 7	56	69	13	6
Indiv 8	76	80	4	4
Indiv 9	65	62	-3	2

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe	Affect signes
Indiv 1	82	63	-19	7	-7
Indiv 2	69	42	-27	8	
Indiv 3	73	74	1	1	
Indiv 4	43	47	4	4	
Indiv 5	58	62	4	4	
Indiv 6	61	61	0		
Indiv 7	56	69	13	6	
Indiv 8	76	80	4	4	
Indiv 9	65	62	-3	2	

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe	Affect signes
Indiv 1	82	63	-19	7	-7
Indiv 2	69	42	-27	8	-8
Indiv 3	73	74	1	1	1
Indiv 4	43	47	4	4	4
Indiv 5	58	62	4	4	4
Indiv 6	61	61	0		
Indiv 7	56	69	13	6	6
Indiv 8	76	80	4	4	4
Indiv 9	65	62	-3	2	-2

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe	Affect signes	Signe moins fréquent
Indiv 1	82	63	-19	7	-7	-7
Indiv 2	69	42	-27	8	-8	-8
Indiv 3	73	74	1	1	1	
Indiv 4	43	47	4	4	4	
Indiv 5	58	62	4	4	4	
Indiv 6	61	61	0			
Indiv 7	56	69	13	6	6	
Indiv 8	76	80	4	4	4	
Indiv 9	65	62	-3	2	-2	-2

Comparaisons de moyennes

Comparaisons de deux moyennes observées (dépendantes ou appariées)

loi non-normale + possibilité de quantifier la différence :

test des rangs signés de Wilcoxon

	avant	après	di	Rang sans signe	Affect signes	Signe moins fréquent
Indiv 1	82	63	-19	7	-7	-7
Indiv 2	69	42	-27	8	-8	-8
Indiv 3	73	74	1	1	1	
Indiv 4	43	47	4	4	4	
Indiv 5	58	62	4	4	4	
Indiv 6	61	61	0			
Indiv 7	56	69	13	6	6	
Indiv 8	76	80	4	4	4	
Indiv 9	65	62	-3	2	-2	-2
						w = 17

$|w| = 17$
 $n = 8$ (9 - 1 identique)
-> Je regarde la table de Wilcoxon

Comparaisons de moyennes

- Attention, dans ce test je rejette H_0 si mon w calculé est $<$ à mon w_{lu}
- Ici ce n'est pas le cas : $17 > 3$
- Je ne peux pas rejeter H_0
- Il n'y a pas de différence significative entre « avant » et « après »

Test bilatéral		
n	risque 5%	risque 1%
6	0	
7	2	
8	3	0
9	5	1
10	8	3
11	10	5
12	13	9
13	17	9
14	21	12
15	25	15
16	29	19
17	34	23
18	40	27
19	46	32
20	52	37
21	59	43
22	66	49
23	73	55
24	81	61
25	89	68

Comparaisons de moyennes

Exercice d'application :

Les recettes match-day des clubs de ligue 1 ont-elles significativement augmentée d'une saison à l'autre ?

2013/2014	
Paris SG	32,31
AS Monaco	3,7336
Lille OSC	9,4776
AS Saint-Etienne	7,0364
OL	11,488
OM	16,0832
FC. Gir. Bordeaux	7,6108
FC Lorient	2,154
Toulouse FC	2,7284
SC Bastia	4,0208
Stade de Reims	5,026
Stade Rennais FC	4,1644
FC Nantes	7,7544
Evian TG FC	2,5848
Montpellier HSC	2,4412
EA Guigamp	3,3028
OGC Nice	4,4516
Valenciennes FC	
FC Sochaux	
AC Ajaccio	

2014/2015	
Paris SG	33,3802
AS Monaco	3,4804
Lille OSC	11,074
AS Saint-Etienne	6,74354
OL	11,865
OM	20,4078
FC.Gir. Bordeaux	8,0682
FC Lorient	2,8476
Toulouse FC	2,5312
SC Bastia	3,7968
Stade de Reims	5,3788
Stade Rennais FC	5,2206
FC Nantes	9,492
Evian TG FC	3,4804
Montpellier HSC	2,8476
EA Guigamp	4,1132
OGC Nice	4,2714
FC Metz	
RC Lens	
SM Caen	

Comparaisons de moyennes

	2013/2014	2014/2015	di	Valeur absolue	Rang sans signe	affect signes
Paris SG	32,31	33,3802	1,0702	1,0702	14	14
AS Monaco	3,7336	3,4804	-0,2532	0,2532	4	-4
Lille OSC	9,4776	11,074	1,5964	1,5964	15	15
AS Saint-Etienne	7,0364	6,74354	-0,29286	0,29286	5	-5
OL	11,488	11,865	0,377	0,377	7	7
OM	16,0832	20,4078	4,3246	4,3246	17	17
FC.Gir. Bordeaux	7,6108	8,0682	0,4574	0,4574	9	9
FC Lorient	2,154	2,8476	0,6936	0,6936	10	10
Toulouse FC	2,7284	2,5312	-0,1972	0,1972	2	-2
SC Bastia	4,0208	3,7968	-0,224	0,224	3	-3
Stade de Reims	5,026	5,3788	0,3528	0,3528	6	6
Stade Rennais FC	4,1644	5,2206	1,0562	1,0562	13	13
FC Nantes	7,7544	9,492	1,7376	1,7376	16	16
Evian TG FC	2,5848	3,4804	0,8956	0,8956	12	12
Montpellier HSC	2,4412	2,8476	0,4064	0,4064	8	8
EA Guigamp	3,3028	4,1132	0,8104	0,8104	11	11
OGC Nice	4,4516	4,2714	-0,1802	0,1802	1	-1
						-15

w_{lu} (pour $n = 17$, à $0,05$) = 34

15 < 34 : on peut rejeter H_0 . On peut même la rejeter à 0,01.

Les recettes match-day des clubs de Ligue 1 ont significativement augmenté entre la saison 2013/2014 et la saison 2014/2015

Comparaisons de moyennes

w calculé = 15

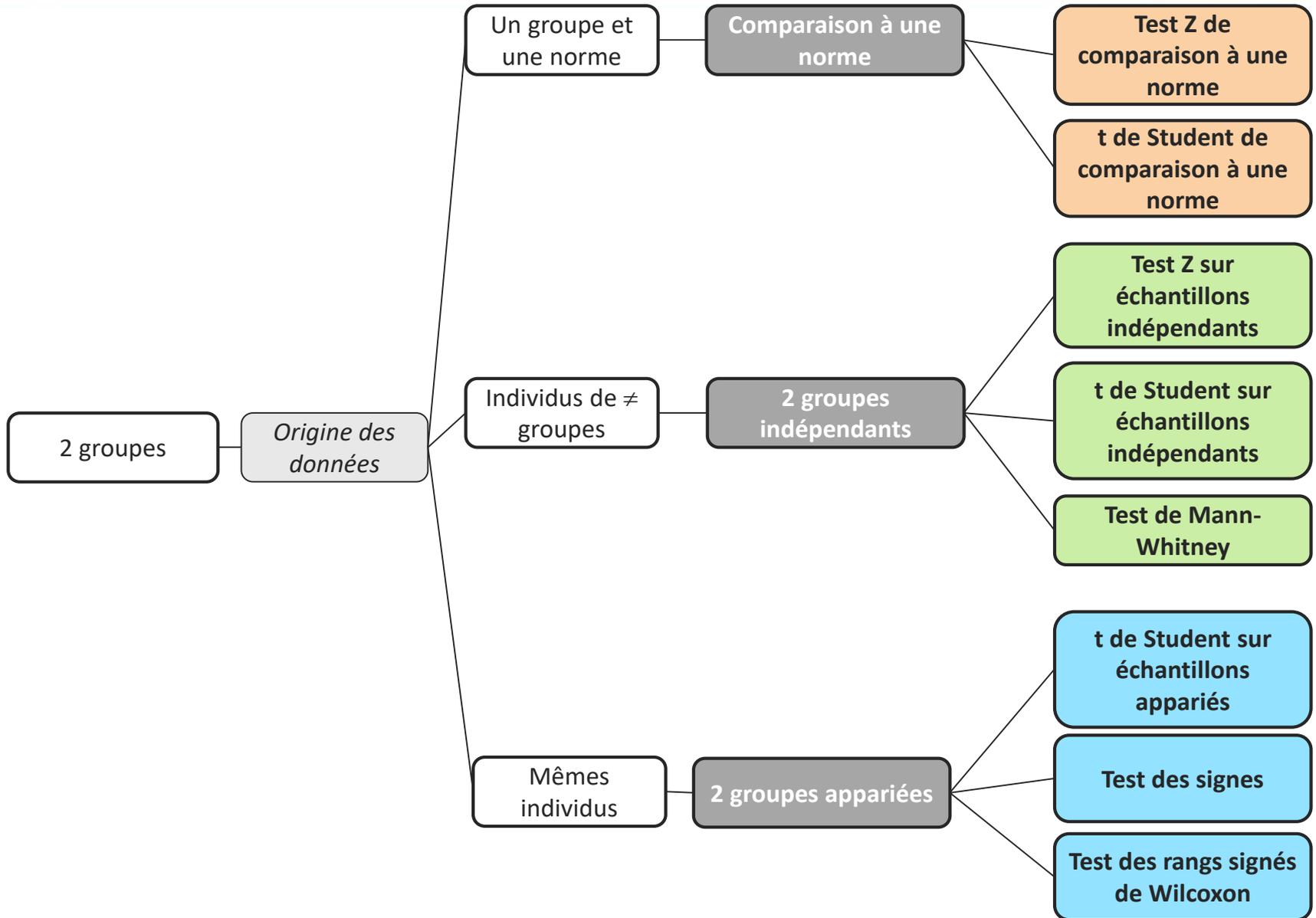
w lu à 0,05 = 34

..... à 0,01 = 23

- w calculé < w lu : je peux rejeter H0 dans les deux cas

Test bilatéral		
n	risque 5%	risque 1%
6	0	
7	2	
8	3	0
9	5	1
10	8	3
11	10	5
12	13	9
13	17	9
14	21	12
15	25	15
16	29	19
17	34	23
18	40	27
19	46	32
20	52	37
21	59	43
22	66	49
23	73	55
24	81	61
25	89	68

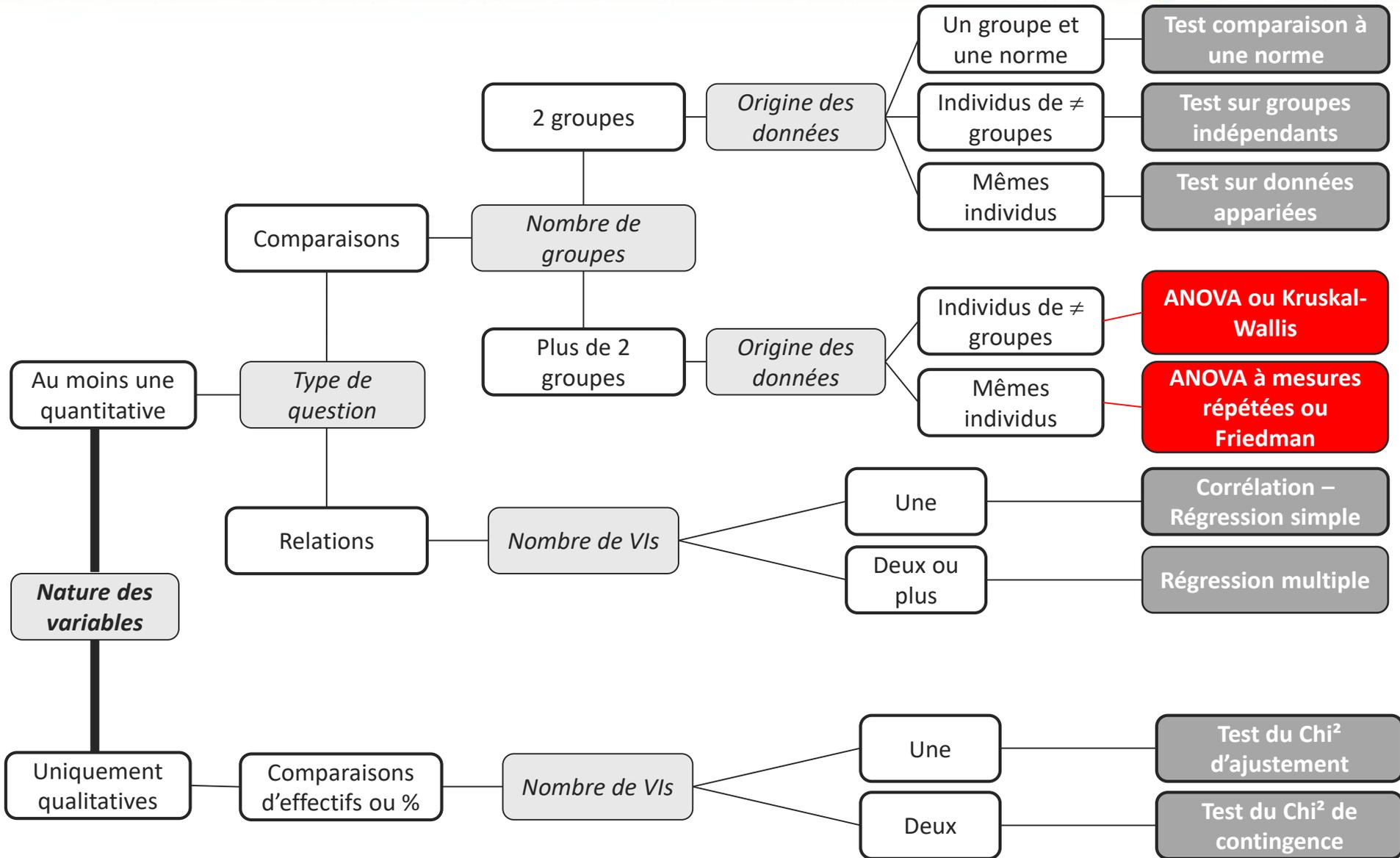
Résumé



Comparaisons de moyennes (plus de 2 groupes)



Comparaisons de moyennes (plus de 2 groupes)



Comparaisons de moyennes (plus de 2 groupes)

Objectif des comparaisons de plus de 2 groupes : à quelle(s) question(s) veut-on répondre ?

Comparer des groupes indépendants :

- Les fan zones des 10 villes d'accueil ont-elles accueilli des publics d'âges comparables ?
- Le chiffre d'affaire des magasins Décathlon dépend-il de leur région d'implantation ?
- Les clubs de football européens ont-ils des recettes match-day équivalentes ?

Comparaisons de moyennes (plus de 2 groupes)

Objectif des comparaisons de plus de 2 groupes : à quelle(s) question(s) veut-on répondre ?

Comparer des groupes appariés :

- Les recettes match-day des clubs de Ligue 1 connaissent-elles une évolution significative au fil des saisons ?

Comparaisons de moyennes (plus de 2 groupes)

Comparaisons (à 1 facteur) de plus de 2 groupes

Choix du test :

		Variables	
		Numériques normales	Numériques non-normales ou ordinales
Groupes	Indépendants	ANOVA	Test Kruskal-Wallis
	Appariés	ANOVA à mesures répétées	Test Friedman

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi normale : l'ANOVA ou analyse de la variance

- Test paramétrique
- Conditions :
 - *Normalité des variables*
 - *Homogénéité des variances*
- Si ces conditions ne sont pas remplies on peut toujours effectuer un test non-paramétrique (Kruskal-Wallis)

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi normale : l'ANOVA ou analyse de la variance

Principe

- Soit n le nombre d'individus par groupe
- Soit k le nombre de groupes
- Soit m la moyenne de chaque groupe (m_1, m_2, \dots, m_k)
- La variance totale (de l'ensemble des individus tous groupes compris) se compose :
 - De la variance inter-groupes
 - De la variance intra-groupes
- Par définition, si la variance inter-groupes est significativement supérieure à la variance intra-groupes, c'est que notre VI a une influence

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi normale : l'ANOVA ou analyse de la variance

Principe

- Ainsi on posera toujours les hypothèses suivantes :
 - $H_0 = m_1 = m_2 = \dots = m_k$ (autrement dit « les moyennes ne sont pas significativement différentes, la VI n'a pas d'influence »)
 - $H_1 =$ au moins une des moyennes est significativement différente (autrement dit « la VI a une influence »)
- L'ANOVA ne permet pas de dire quel(s) groupe(s) est (sont) différent(s). Il est alors nécessaire de faire des tests post-hoc si on en a besoin (dans la plupart des cas qui nous intéressent on n'en a pas besoin)

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi normale : l'ANOVA ou analyse de la variance

Mise en oeuvre

- Il faut calculer les SCE (sommes des carrés des écarts)
 - *SCE totale* : (sous excel) `SOMME.CARRES.ECARTS` sur tous les individus
 - *SCE inter-groupes* : $(\text{Moy } G1 - \text{Moy générale})^2 * n$
 $+ (\text{Moy } G2 - \text{Moy générale})^2 * n$
 $+ (\text{Moy } Gk - \text{Moy générale})^2 * n$
 - *SCE intra-groupes* : $\sum (\text{chaque indiv } G1 - \text{Moyenne } G1)^2$
 $+ \sum (\text{chaque indiv } G2 - \text{Moyenne } G2)^2$
 $+ \sum (\text{chaque indiv } Gk - \text{Moyenne } Gk)^2$
correspond à $\sum (\text{SOMME.CARRES.ECARTS de chaque groupe})$

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi normale : l'ANOVA ou analyse de la variance

Mise en oeuvre

- Il faut déterminer les ddl
 - *ddl total : $kn - 1$*
 - *ddl inter-groupes : $k-1$*
 - *ddl intra-groupes : $k(n-1)$*
- Calculer les variances :
 - *Variance inter-groupes = $SCE\ inter / ddl\ inter$*
 - *Variance intra-groupes = $SCE\ intra / ddl\ intra$*

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi normale : l'ANOVA ou analyse de la variance

Mise en oeuvre

- A partir de là, notre valeur calculée F :

$$F = \text{variance inter} / \text{variance intra}$$

- Enfin, on compare le F calculé au F lu sur la table de Fisher (qui se lit grâce aux ddl inter et intra)
 - Si $F(\text{calculé}) > F(\text{lu})$: on peut rejeter H_0
 - Si $F(\text{calculé}) < F(\text{lu})$: on ne peut pas rejeter H_0

Table F : valeurs critiques de la distribution F

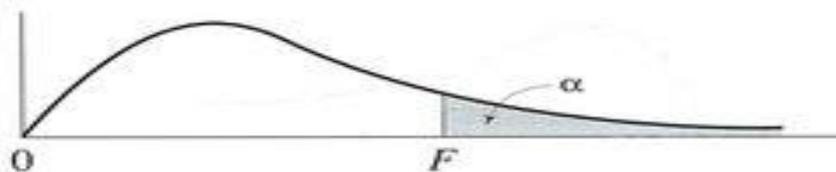


TABLE 1 $\alpha = .05$ → Une table par niveau de risque (ici 0,05)

	Degrés de liberté pour le numérateur															
	1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8	252.6
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.58
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.70
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.44
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.75
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.32
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.02
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.80
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.64
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.51
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.40
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.31
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.24
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.18
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15	2.12
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10	2.08
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06	2.04
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03	2.00
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99	1.97
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.97	1.94	1.89	1.86
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.94	1.90	1.85	1.82
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.91	1.87	1.82	1.79
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.88	1.84	1.79	1.76
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.78	1.74	1.69	1.66
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.73	1.69	1.63	1.60
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.69	1.65	1.59	1.56
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66	1.60	1.55	1.50	1.46
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.72	1.62	1.56	1.52	1.46	1.41
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.69	1.59	1.53	1.48	1.42	1.38
1000	3.85	3.01	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.68	1.58	1.52	1.47	1.41	1.36

Comparaisons de moyennes

Mise en application théorique :

G1	G2	G3
50	162	120
52	350	120
123	125	122
100	320	221
200	112	253
250	200	141
220	40	182
220	162	175
300	160	160
220	250	214

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3
	50	162	120
	52	350	120
	123	125	122
	100	320	221
	200	112	253
	250	200	141
	220	40	182
	220	162	175
	300	160	160
	220	250	214
moyennes	173,5	188,1	170,8

=MOY (indiv de chaque groupe)

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3		
	50	162	120		
	52	350	120		
	123	125	122		
	100	320	221		
	200	112	253		
	250	200	141		
	220	40	182		
	220	162	175		
	300	160	160		
	220	250	214		
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667

Moyenne générale = MOY (tous les individus)

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3		
	50	162	120		
	52	350	120		
	123	125	122		
	100	320	221		
	200	112	253		
	250	200	141		
	220	40	182		
	220	162	175		
	300	160	160		
	220	250	214		
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667
totale				SCE totale	170017,4667

SCE totale = SOMME.CARRES.ECARTS (de tous les individus)

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3		
	50	162	120		
	52	350	120		
	123	125	122		
	100	320	221		
	200	112	253		
	250	200	141		
	220	40	182		
	220	162	175		
	300	160	160		
	220	250	214		
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667
totale				SCE totale	170017,4667
Inter	157,344444	1130,67778	444,444444	SCE inter	

Inter = ((moy Groupe – moy générale)² * n) pour chaque groupe

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3		
	50	162	120		
	52	350	120		
	123	125	122		
	100	320	221		
	200	112	253		
	250	200	141		
	220	40	182		
	220	162	175		
	300	160	160		
	220	250	214		
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667
totale				SCE totale	170017,4667
Inter	157,344444	1130,67778	444,444444	SCE inter	1732,466667

SCE inter = SOMME. des inter de chaque groupe

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3		
	50	162	120		
	52	350	120		
	123	125	122		
	100	320	221		
	200	112	253		
	250	200	141		
	220	40	182		
	220	162	175		
	300	160	160		
	220	250	214		
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667
totale				SCE totale	170017,4667
Inter	157,344444	1130,67778	444,444444	SCE inter	1732,466667
Intra	67010,5	81440,9	19833,6	SCE intra	

Intra = SOMME.CARRE.ECARTS pour chaque groupe

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3		
	50	162	120		
	52	350	120		
	123	125	122		
	100	320	221		
	200	112	253		
	250	200	141		
	220	40	182		
	220	162	175		
	300	160	160		
	220	250	214		
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667
totale				SCE totale	170017,4667
Inter	157,344444	1130,67778	444,444444	SCE inter	1732,466667
Intra	67010,5	81440,9	19833,6	SCE intra	168285

Intra = SOMME des intras de chaque groupe

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3				
	50	162	120				
	52	350	120				
	123	125	122				
	100	320	221				
	200	112	253				
	250	200	141				
	220	40	182				
	220	162	175				
	300	160	160				
	220	250	214				
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667		
totale				SCE totale	170017,4667	ddl : kn-1	29
Inter	157,344444	1130,67778	444,444444	SCE inter	1732,466667	ddl : k-1	2
Intra	67010,5	81440,9	19833,6	SCE intra	168285	ddl : k(n-1)	27

ddl = appliquer les 3 formules

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3						
	50	162	120						
	52	350	120						
	123	125	122						
	100	320	221						
	200	112	253						
	250	200	141						
	220	40	182						
	220	162	175						
	300	160	160						
	220	250	214						
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667				
totale				SCE totale	170017,4667	ddl : kn-1	29		
Inter	157,344444	1130,67778	444,444444	SCE inter	1732,466667	ddl : k-1	2	Variance inter	866,233333
Intra	67010,5	81440,9	19833,6	SCE intra	168285	ddl : k(n-1)	27	Variance intra	6232,7778

Variance inter = SCE inter / ddl inter

Variance intra = SCE intra / ddl intra

Comparaisons de moyennes

Mise en application théorique :

	G1	G2	G3						
	50	162	120						
	52	350	120						
	123	125	122						
	100	320	221						
	200	112	253						
	250	200	141						
	220	40	182						
	220	162	175						
	300	160	160						
	220	250	214						
moyennes	173,5	188,1	170,8	Moyenne générale	177,4666667				
totale				SCE totale	170017,4667	ddl : kn-1	29		
Inter	157,344444	1130,67778	444,444444	SCE inter	1732,466667	ddl : k-1	2	Variance inter	866,233333
Intra	67010,5	81440,9	19833,6	SCE intra	168285	ddl : k(n-1)	27	Variance intra	6232,77778
								F	0,1389803

$$F = \text{variance inter} / \text{variance intra}$$

Table F : valeurs critiques de la distribution F

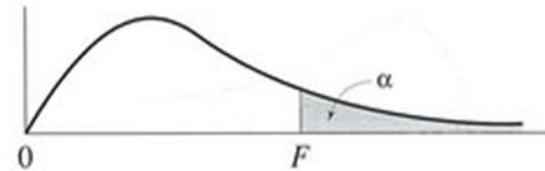


TABLE 1 α = .05

	Degrés de liberté pour le numérateur															
	1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8	252.6
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.58
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.70
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.44
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.75
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.32
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.02
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.80
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.64
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.51
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.40
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.31
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.24
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.18
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15	2.12
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10	2.08
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06	2.04
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03	2.00
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99	1.97
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.97	1.94	1.89	1.86
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.94	1.90	1.85	1.82
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.91	1.87	1.82	1.79
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.88	1.84	1.79	1.76
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.78	1.74	1.69	1.66
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.73	1.69	1.63	1.60
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.69	1.65	1.59	1.56
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66	1.60	1.55	1.50	1.46
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.72	1.62	1.56	1.52	1.46	1.41
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.69	1.59	1.53	1.48	1.42	1.38
1000	3.85	3.01	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.68	1.58	1.52	1.47	1.41	1.36

F calculé = 0,14

F lu à 0,05 = 3,355

- F calculé < F lu : je ne peux pas rejeter H0
- Ma VI n'a pas d'influence significative

Comparaisons de moyennes

Exercice d'application :

Les valeurs des franchises US sont-elles équivalentes ?



NFL		MLB		NBA	
Arizona Cardinals	1420	Arizona Diamondbacks	1110	Atlanta Hawks	990
Atlanta Falcons	1770	Atlanta Braves	1410	Boston Celtics	2900
Baltimore Ravens	1610	Baltimore Orioles	1200	Brooklyn Nets	1860
Carolina Panthers	1230	Boston Red Sox	2760	Charlotte Hornets	800
Chicago Bears	2100	Chicago Cubs	2640	Chicago Bulls	3150
Cincinnati Bengals	890	Chicago White Sox	1260	Cleveland Cavaliers	1120
Cleveland Browns	980	Cincinnati Reds	1086	Dallas Mavericks	1480
Dallas Cowboys	4750	Cleveland Indians	960	Denver Nuggets	1026
Denver Broncos	1620	Colorado Rockies	1032	Detroit Pistons	725
Green Bay Packers	1645	Detroit Tigers	1380	Golden State Warriors	2280
Houston Texans	2120	Houston Astros	1320	Houston Rockets	1510
Indianapolis Colts	1905	Kansas City Royals	1038	Indiana Pacers	1008
Jacksonville Jaguars	1365	Angeles Angels of Anah	1608	Los Angeles Clippers	2400
Kansas City Chiefs	870	Los Angeles Dodgers	3000	Los Angeles Lakers	3240
Los Angeles Rams	2060	Miami Marlins	810	Memphis Grizzlies	815
Miami Dolphins	1280	Milwaukee Brewers	1050	Miami Heat	1360
Minnesota Vikings	1350	Minnesota Twins	1092	Milwaukee Bucks	810
New England Patriots	3180	New York Mets	1980	Minnesota Timberwolves	864
New Orleans Saints	935	New York Yankees	4080	New Orleans Pelicans	780
New York Giants	2970	Oakland Athletics	870	New York Knicks	3950
New York Jets	2225	Philadelphia Phillies	1482	Oklahoma City Thunder	940
Oakland Raiders	1230	Pittsburgh Pirates	1170	Orlando Magic	970
Philadelphia Eagles	1680	San Diego Padres	1068	Philadelphia 76ers	840
Pittsburgh Steelers	1575	San Francisco Giants	2700	Phoenix Suns	980
San Diego Chargers	1760	Seattle Mariners	1440	Portland Trail Blazers	875
San Francisco 49ers	2400	St Louis Cardinals	1920	Sacramento Kings	960
Seattle Seahawks	1750	Tampa Bay Rays	780	San Antonio Spurs	1380
Tampa Bay Buccaneers	1810	Texas Rangers	1470	Toronto Raptors	995
Tennessee Titans	1420	Toronto Blue Jays	1080	Utah Jazz	790
Washington Redskins	2365	Washington Nationals	1560	Washington Wizards	930

Comparaisons de moyennes

	NFL	MLB	NBA					
	1420	1110	990					
	1770	1410	2900					
	1610	1200	1860					
	1230	2760	800					
	2100	2640	3150					
	890	1260	1120					
	980	1086	1480					
	4750	960	1026					
	1620	1032	725					
	1645	1380	2280					
	2120	1320	1510					
	1905	1038	1008					
	1365	1608	2400					
	870	3000	3240					
	2060	810	815					
	1280	1050	1360					
	1350	1092	810					
	3180	1980	864					
	935	4080	780					
	2970	870	3950					
	2225	1482	940					
	1230	1170	970					
	1680	1068	840					
	1575	2700	980					
	1760	1440	875					
	2400	1920	960					
	1750	780	1380					
	1810	1470	995					
	1420	1080	790					
	2365	1560	930					
moyennes	1808,833333	1545,2	1424,266667	moy gen	1592,76667			
totale				SCE totale	59402942,1	kn-1	89	
inter	1400544,133	67877,63333	851767,5	SCE inter	2320189,27	k-1	2	variance inte 1160094,63
intra	17981334,17	17081548,8	22019869,87	SCE intra	57082752,8	k(n-1)	87	variance intr 656123,596
								F 1,76810382
								F critique 3,15

Table F : valeurs critiques de la distribution F

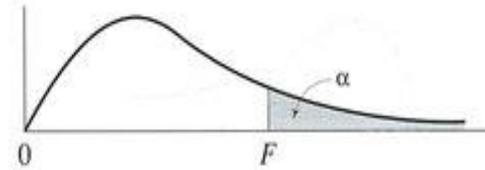


TABLE 1 α = .05

		Degrés de liberté pour le numérateur															
		1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
Degrés de liberté pour le dénominateur	1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8	252.6
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48	19.48
	3	10.13	10.55	10.72	10.81	10.86	10.89	10.91	10.92	10.93	10.94	10.95	10.95	10.96	10.96	10.96	10.96
	4	7.71	8.14	8.31	8.40	8.45	8.48	8.50	8.51	8.52	8.53	8.53	8.54	8.54	8.54	8.54	8.54
	5	6.61	7.04	7.21	7.30	7.35	7.38	7.40	7.41	7.42	7.43	7.43	7.44	7.44	7.44	7.44	7.44
	6	5.99	6.42	6.59	6.68	6.73	6.76	6.78	6.79	6.80	6.81	6.81	6.82	6.82	6.82	6.82	6.82
	7	5.59	6.02	6.19	6.28	6.33	6.36	6.38	6.39	6.40	6.41	6.41	6.42	6.42	6.42	6.42	6.42
	8	5.32	5.75	5.92	6.01	6.06	6.09	6.11	6.12	6.13	6.14	6.14	6.15	6.15	6.15	6.15	6.15
	9	5.12	5.55	5.72	5.81	5.86	5.89	5.91	5.92	5.93	5.94	5.94	5.95	5.95	5.95	5.95	5.95
	10	4.96	5.39	5.56	5.65	5.70	5.73	5.75	5.76	5.77	5.78	5.78	5.79	5.79	5.79	5.79	5.79
	11	4.84	5.27	5.44	5.53	5.58	5.61	5.63	5.64	5.65	5.66	5.66	5.67	5.67	5.67	5.67	5.67
	12	4.75	5.18	5.35	5.44	5.49	5.52	5.54	5.55	5.56	5.57	5.57	5.58	5.58	5.58	5.58	5.58
	13	4.67	5.10	5.27	5.36	5.41	5.44	5.46	5.47	5.48	5.49	5.49	5.50	5.50	5.50	5.50	5.50
	14	4.60	5.03	5.20	5.29	5.34	5.37	5.39	5.40	5.41	5.42	5.42	5.43	5.43	5.43	5.43	5.43
	15	4.54	4.97	5.14	5.23	5.28	5.31	5.33	5.34	5.35	5.36	5.36	5.37	5.37	5.37	5.37	5.37
16	4.49	4.92	5.09	5.18	5.23	5.26	5.28	5.29	5.30	5.31	5.31	5.32	5.32	5.32	5.32	5.32	
17	4.45	4.88	5.05	5.14	5.19	5.22	5.24	5.25	5.26	5.27	5.27	5.28	5.28	5.28	5.28	5.28	
18	4.41	4.84	5.01	5.10	5.15	5.18	5.20	5.21	5.22	5.23	5.23	5.24	5.24	5.24	5.24	5.24	
19	4.38	4.81	4.98	5.07	5.12	5.15	5.17	5.18	5.19	5.20	5.20	5.21	5.21	5.21	5.21	5.21	
20	4.35	4.78	4.95	5.04	5.09	5.12	5.14	5.15	5.16	5.17	5.17	5.18	5.18	5.18	5.18	5.18	
22	4.30	4.73	4.90	4.99	5.04	5.07	5.09	5.10	5.11	5.12	5.12	5.13	5.13	5.13	5.13	5.13	
24	4.26	4.69	4.86	4.95	5.00	5.03	5.05	5.06	5.07	5.08	5.08	5.09	5.09	5.09	5.09	5.09	
26	4.23	4.66	4.83	4.92	4.97	5.00	5.02	5.03	5.04	5.05	5.05	5.06	5.06	5.06	5.06	5.06	
28	4.20	4.63	4.80	4.89	4.94	4.97	4.99	5.00	5.01	5.02	5.02	5.03	5.03	5.03	5.03	5.03	
30	4.17	4.60	4.77	4.86	4.91	4.94	4.96	4.97	4.98	4.99	4.99	5.00	5.00	5.00	5.00	5.00	
40	4.08	4.51	4.68	4.77	4.82	4.85	4.87	4.88	4.89	4.90	4.90	4.91	4.91	4.91	4.91	4.91	
50	4.03	4.46	4.63	4.72	4.77	4.80	4.82	4.83	4.84	4.85	4.85	4.86	4.86	4.86	4.86	4.86	
60	4.00	4.43	4.60	4.69	4.74	4.77	4.79	4.80	4.81	4.82	4.82	4.83	4.83	4.83	4.83	4.83	
120	3.92	4.35	4.52	4.61	4.66	4.69	4.71	4.72	4.73	4.74	4.74	4.75	4.75	4.75	4.75	4.75	
200	3.89	4.32	4.49	4.58	4.63	4.66	4.68	4.69	4.70	4.71	4.71	4.72	4.72	4.72	4.72	4.72	
500	3.86	4.29	4.46	4.55	4.60	4.63	4.65	4.66	4.67	4.68	4.68	4.69	4.69	4.69	4.69	4.69	
1000	3.85	4.28	4.45	4.54	4.59	4.62	4.64	4.65	4.66	4.67	4.67	4.68	4.68	4.68	4.68	4.68	

F calculé = 1,768

F lu à 0,05 = entre 3,07 et 3,15

- F calculé < F lu : je ne peux pas rejeter H0
- Les valeurs des 3 franchises sont équivalentes
- Cela signifie que les réalités divergent davantage à l'intérieur de chaque ligue qu'entre les ligues...
- ... et montre bien à quel point les moyennes sont des indicateurs trompeurs !

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

- Test non-paramétrique
- Conditions :
 - Variables quantitatives (numériques) ou ordinales
- C'est une extension du test des rangs

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Principe

- On posera toujours les hypothèses suivantes :
 - H_0 = il n'y a pas de différence entre les groupes
 - H_1 = il y a une différence entre les groupes

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Exemple

Groupe 1	Groupe 2	Groupe 3
27	20	34
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 1 : attribuer un rang à toutes les mesures (**attention Fonction MOYENNE-RANG**)

Groupe 1	Groupe 2	Groupe 3
27	20	34
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 1 : attribuer un rang à toutes les mesures (**attention Fonction MOYENNE-RANG**)

Groupe 1	Groupe 2	Groupe 3
27	20	34
2 = rang 1	8	31
4	14	3
18	36	23
7	21	30
9	22	6

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 1 : attribuer un rang à toutes les mesures (**attention Fonction MOYENNE-RANG**)

Groupe 1	Groupe 2	Groupe 3
27	20	34
1	8	31
4	14	3 = rang 2
18	36	23
7	21	30
9	22	6

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 1 : attribuer un rang à toutes les mesures (**attention Fonction MOYENNE-RANG**)

Groupe 1	Groupe 2	Groupe 3
14	10	17
1	6	16
3	8	2
9	18	13
5	11	15
7	12	4

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 2 : calculer la somme des rangs pour chaque groupe (notées T_k)

	Groupe 1	Groupe 2	Groupe 3
	14	10	17
	1	6	16
	3	8	2
	9	18	13
	5	11	15
	7	12	4
T_k	39	65	67

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 3 : insérer les effectifs de chaque groupe (notés n_k)

	Groupe 1	Groupe 2	Groupe 3
	14	10	17
	1	6	16
	3	8	2
	9	18	13
	5	11	15
	7	12	4
Tk	39	65	67
n_k	6	6	6

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 4 : appliquer la formule du H de Kruskal-Wallis

$$H = \frac{12}{N(N+1)} \times \left(\sum \frac{T_k^2}{n_k} \right) - 3(N + 1)$$

N = effectif total

T_k = somme des rangs de chaque groupe

n_k = effectif de chaque groupe

	Groupe 1	Groupe 2	Groupe 3
	14	10	17
	1	6	16
	3	8	2
	9	18	13
	5	11	15
	7	12	4
T _k	39	65	67
n _k	6	6	6

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 4 : appliquer la formule du H de Kruskal-Wallis

$$H = \frac{12}{18(18+1)} \times \left(\frac{39^2}{6} + \frac{65^2}{6} + \frac{67^2}{6} \right) - 3(18 + 1)$$

$$H = 2,854$$

	Groupe 1	Groupe 2	Groupe 3
	14	10	17
	1	6	16
	3	8	2
	9	18	13
	5	11	15
	7	12	4
Tk	39	65	67
nk	6	6	6

Comparaisons de moyennes (plus de 2 groupes)

Groupes indépendants + loi non-normale : test de Kruskal-Wallis

Etape 5 : comparer la valeur calculé à la valeur lue sur la table du χ^2 (ou « Khi2 »)

Risque : 0,05

ddl = $K-1$ (nombre de groupes -1)

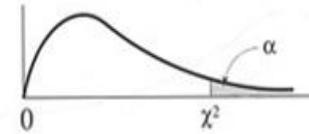
-> $3 - 1 = 2$

	Groupe 1	Groupe 2	Groupe 3
	14	10	17
	1	6	16
	3	8	2
	9	18	13
	5	11	15
	7	12	4
Tk	39	65	67
nk	6	6	6

Si H (calculé) > H (lu) : je peux rejeter H_0

Si H (calculé) < H (lu) je ne peux pas rejeter H_0

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2



2,854 < 5,99

- Je ne peux pas rejeter H0
- Il n'y a pas de différence significative entre les groupes

df	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

Comparaisons de moyennes

Exercice d'application :

Les spectateurs de la Fan Zone de Paris ont-ils le même niveau de satisfaction selon leur tranche d'âge ?



18-25	26-35	36-45	Plus de 45
1	4	3	3
4	4	4	4
4	5	4	5
5	3	5	5
2	4	3	4
4	5	2	2
5	5	3	3
2	5	4	4
5	5	5	1
5	5	5	3
3	5	1	2
4	5	3	4
5	5	5	5
4	5	4	4
5	4	5	4
2	4	5	5
4	5	3	5
5	5	1	3
4	5	4	4
5	5	4	2
5	4	4	5
3	5	5	5
5	5	3	2
1	5	4	4
5	5	3	
3	4		
	5		

Comparaisons de moyennes

Exercice d'application :

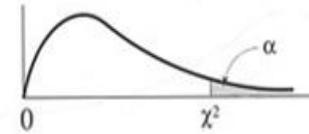
Les spectateurs de la Fan Zone de Paris ont-ils le même niveau de satisfaction selon leur tranche d'âge ?

- Variable ordinale
- nk différent d'un groupe à l'autre
- Le test s'applique de la même façon
- Appliquer la formule MOYENNE.RANG lorsque de nombreux scores sont identiques

Comparaisons de moyennes

Exercice d'application :

	18-25	26-35	36-45	Plus de 45
	3	43,5	21	21
	43,5	43,5	43,5	43,5
	43,5	80,5	43,5	80,5
	80,5	21	80,5	80,5
	9,5	43,5	21	43,5
	43,5	80,5	9,5	9,5
	80,5	80,5	21	21
	9,5	80,5	43,5	43,5
	80,5	80,5	80,5	3
	80,5	80,5	80,5	21
	21	80,5	3	9,5
	43,5	80,5	21	43,5
	80,5	80,5	80,5	80,5
	43,5	80,5	43,5	43,5
	80,5	43,5	80,5	43,5
	9,5	43,5	80,5	80,5
	43,5	80,5	21	80,5
	80,5	80,5	3	21
	43,5	80,5	43,5	43,5
	80,5	80,5	43,5	9,5
	80,5	43,5	43,5	80,5
	21	80,5	80,5	80,5
	80,5	80,5	21	9,5
	3	80,5	43,5	43,5
	80,5	80,5	21	
	21	43,5		
		80,5		
Tk	1287,5	1855	1074	1036,5
nk	26	27	25	24
N	102			
H	13,2207467			

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2 

H calculé = 13,22

H lu (ddl : k-1 soit 3) = 7,82

- H calculé > H lu
- Je peux rejeter H0
- La tranche d'âge a un impact sur le niveau de satisfaction des spectateurs fan zones
- Il faudrait faire des tests post-hoc pour identifier les tranches d'âge qui diffèrent significativement (a priori les 26-35 ans, peut-être les 18-25 ans)

dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi normale : ANOVA à mesures répétées

- Test paramétrique
- Conditions :
 - *Idem ANOVA*

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi normale : ANOVA à mesures répétées

Principe

- Idem ANOVA à une exception près :
- *A l'intérieur de la variance intra-groupes on distinguera :*
 - *La variance individus*
 - *La variance erreur*
- Cela ajoute simplement un calcul supplémentaire

Comparaisons de moyennes

Mise en application théorique :

Avant	Après	Encore après
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après
	9	7	4
	8	6	3
	7	6	2
	8	7	3
	8	8	4
	9	7	3
	8	6	2
moyennes	8,142857143	6,714285714	3

=MOY (indiv de chaque groupe)

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après		
	9	7	4		
	8	6	3		
	7	6	2		
	8	7	3		
	8	8	4		
	9	7	3		
	8	6	2		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952381

Moyenne générale = MOY (tous les individus)

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après		
	9	7	4		
	8	6	3		
	7	6	2		
	8	7	3		
	8	8	4		
	9	7	3		
	8	6	2		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952381
totales				SCE totale	108,9524

SCE totale = SOMME.CARRES.ECARTS (de tous les individus)

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après		
	9	7	4		
	8	6	3		
	7	6	2		
	8	7	3		
	8	8	4		
	9	7	3		
	8	6	2		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952381
totales				SCE totale	108,9524
Inter	33,58730159	4,063492063	61,01587302	SCE inter	

$$\text{Inter} = ((\text{moy Groupe} - \text{moy générale})^2 * n) \text{ par groupe}$$

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après		
	9	7	4		
	8	6	3		
	7	6	2		
	8	7	3		
	8	8	4		
	9	7	3		
	8	6	2		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952381
totales				SCE totale	108,9524
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66667

SCE inter = SOMME. des inter de chaque groupe

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après		
	9	7	4		
	8	6	3		
	7	6	2		
	8	7	3		
	8	8	4		
	9	7	3		
	8	6	2		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952381
totales				SCE totale	108,9524
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66667
Intra	2,857142857	3,428571429	4	SCE intra	

Intra = SOMME.CARRE.ECARTS par groupe

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après		
	9	7	4		
	8	6	3		
	7	6	2		
	8	7	3		
	8	8	4		
	9	7	3		
	8	6	2		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952381
totales				SCE totale	108,9524
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66667
Intra	2,857142857	3,428571429	4	SCE intra	10,28571

Intra = SOMME des intras de chaque groupe

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après		
	9	7	4		
	8	6	3		
	7	6	2		
	8	7	3		
	8	8	4		
	9	7	3		
	8	6	2		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952
totales				SCE totale	108,952381
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429
<i>Sujets</i>				SCE sujets	
<i>Erreur</i>				SCE erreur	

Il faut maintenant décomposer la SCE intra en:

- SCE sujets
- SCE erreur

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après	moyennes sujets	CE sujets*k
	9	7	4		
	8	6	3		
	7	6	2		
	8	7	3		
	8	8	4		
	9	7	3		
	8	6	2		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952
totales				SCE totale	108,952381
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429
<i>Sujets</i>				SCE sujets	
<i>Erreur</i>				SCE erreur	

Pour ça on a besoin :

- **Des moyennes de chaque sujet (donc MOY. en lignes)**

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après	moyennes sujets	CE sujets*k
	9	7	4	6,666666667	
	8	6	3	5,666666667	
	7	6	2	5	
	8	7	3	6	
	8	8	4	6,666666667	
	9	7	3	6,333333333	
	8	6	2	5,333333333	
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952
totales				SCE totale	108,952381
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429
<i>Sujets</i>				SCE sujets	
<i>Erreur</i>				SCE erreur	

Pour ça on a besoin :

- **Des moyennes de chaque sujet (donc MOY. en lignes)**

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après	moyennes sujets	CE sujets*k
	9	7	4	6,666666667	1,530612245
	8	6	3	5,666666667	0,244897959
	7	6	2	5	2,721088435
	8	7	3	6	0,006802721
	8	8	4	6,666666667	1,530612245
	9	7	3	6,333333333	0,43537415
	8	6	2	5,333333333	1,149659864
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952
totales				SCE totale	108,952381
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429
<i>Sujets</i>				SCE sujets	7,619047619
<i>Erreur</i>				SCE erreur	2,666666667

Pour ça on a besoin :

- Des carrés des écarts des moyennes sujets à la moyenne générale * le nb de mesures (k)

Donc $((\text{Moy sujet} - \text{Moy générale})^2 * k)$ (bloquer la moyenne générale avec des \$)

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après	moyennes sujets	CE sujets*k
	9	7	4	6,666666667	1,530612245
	8	6	3	5,666666667	0,244897959
	7	6	2	5	2,721088435
	8	7	3	6	0,006802721
	8	8	4	6,666666667	1,530612245
	9	7	3	6,333333333	0,43537415
	8	6	2	5,333333333	1,149659864
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952
totales				SCE totale	108,952381
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429
<i>Sujets</i>				SCE sujets	7,619047619
<i>Erreur</i>				SCE erreur	

SCE sujets = SOMME. de ces « CE sujets*k »

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après	moyennes sujets	CE sujets*k
	9	7	4	6,666666667	1,530612245
	8	6	3	5,666666667	0,244897959
	7	6	2	5	2,721088435
	8	7	3	6	0,006802721
	8	8	4	6,666666667	1,530612245
	9	7	3	6,333333333	0,43537415
	8	6	2	5,333333333	1,149659864
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952
totales				SCE totale	108,952381
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429
<i>Sujets</i>				SCE sujets	7,619047619
<i>Erreur</i>				SCE erreur	2,666666667

$$\text{SCE Erreur} = \text{SCE intra} - \text{SCE sujets}$$

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après	moyennes sujets	CE sujets*k		
	9	7	4	6,666666667	1,530612245		
	8	6	3	5,666666667	0,244897959		
	7	6	2	5	2,721088435		
	8	7	3	6	0,006802721		
	8	8	4	6,666666667	1,530612245		
	9	7	3	6,333333333	0,43537415		
	8	6	2	5,333333333	1,149659864		
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952		
totales				SCE totale	108,952381	ddl : kn - 1	20
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667	ddl : k - 1	2
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429	ddl : kn - k	18
Sujets				SCE sujets	7,619047619	ddl : n - 1	6
Erreur				SCE erreur	2,666666667	ddl : ddl intra - ddl sujets	12

ddl : on applique les formules

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après	moyennes sujets	CE sujets*k				
	9	7	4	6,666666667	1,530612245				
	8	6	3	5,666666667	0,244897959				
	7	6	2	5	2,721088435				
	8	7	3	6	0,006802721				
	8	8	4	6,666666667	1,530612245				
	9	7	3	6,333333333	0,43537415				
	8	6	2	5,333333333	1,149659864				
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952				
totales				SCE totale	108,952381	ddl : kn - 1	20		
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667	ddl : k - 1	2	variance inter	49,33333333
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429	ddl : kn - k	18		
Sujets				SCE sujets	7,619047619	ddl : n - 1	6		
Erreur				SCE erreur	2,666666667	ddl : ddl intra - ddl sujets	12	variance erreur	0,22222222

On a uniquement besoin des variances inter et erreur

Variance inter = SCE inter / ddl inter

Variance erreur = SCE erreur / ddl erreur

Comparaisons de moyennes

Mise en application théorique :

	Avant	Après	Encore après	moyennes sujets	CE sujets*k				
	9	7	4	6,666666667	1,530612245				
	8	6	3	5,666666667	0,244897959				
	7	6	2	5	2,721088435				
	8	7	3	6	0,006802721				
	8	8	4	6,666666667	1,530612245				
	9	7	3	6,333333333	0,43537415				
	8	6	2	5,333333333	1,149659864				
moyennes	8,142857143	6,714285714	3	moyenne générale	5,952380952				
totales				SCE totale	108,952381	ddl : kn - 1	20		
Inter	33,58730159	4,063492063	61,01587302	SCE inter	98,66666667	ddl : k - 1	2	variance inter	49,33333333
Intra	2,857142857	3,428571429	4	SCE intra	10,28571429	ddl : kn - k	18		
Sujets				SCE sujets	7,619047619	ddl : n - 1	6		
Erreur				SCE erreur	2,666666667	ddl : ddl intra - ddl sujets	12	variance erreur	0,222222222
								F	222

On peut calculer notre F

$$F = \text{Variance inter} / \text{Variance erreur}$$

Table F : valeurs critiques de la distribution F

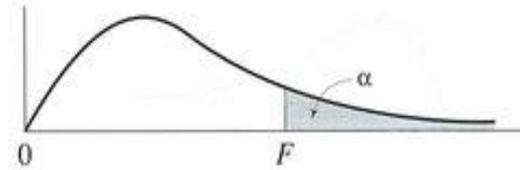


TABLE 1 $\alpha = .05$

	Degrés de liberté pour le numérateur															
	1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8	252.6
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48	19.48
3	10.13	10.55	10.28	10.12	10.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.58
4	7.71	8.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.70
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.44
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.75
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.32
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.02
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.80
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.64
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.51
12	4.73	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.40
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.31
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.24
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.18
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15	2.12
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10	2.08
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06	2.04
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03	2.00
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99	1.97
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.97	1.94	1.89	1.86
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.94	1.90	1.85	1.82
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.91	1.87	1.82	1.79
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.88	1.84	1.79	1.76
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.78	1.74	1.69	1.66
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.73	1.69	1.63	1.60
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.69	1.65	1.59	1.56
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66	1.60	1.55	1.50	1.46
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.72	1.62	1.56	1.52	1.46	1.41
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.69	1.59	1.53	1.48	1.42	1.38
1000	3.85	3.01	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.68	1.58	1.52	1.47	1.41	1.36

F calculé = 222 (ddl inter = 2; ddl erreur = 12)

F lu à 0,05 = 3,89

- F calculé > F lu : je peux rejeter H0
- Ma VI a une influence significative
- Et même très significative (222 est bcp plus grand que 3,89)

Comparaisons de moyennes

Exercice d'application :

A l'aide d'un eye tracker, on mesure la durée moyenne (en millisecondes) des saccades oculaires des spectateurs sur 3 cibles où sont affichés des sponsors :

- le centre de l'octogone
- les montants de la cage
- les shorts des combattants

Ces 3 types de saccades ont-elles une durée équivalente ?



	Centre	Montant	Short
Indiv 1	127,95	147,45	257,65
Indiv 2	251,06	231,98	167,44
Indiv 3	177,5	103,12	215,06
Indiv 4	230,08	183,61	124,42
Indiv 5	142,69	52,82	50,99
Indiv 6	244,81	161,87	150,45
Indiv 7	160,86	105,23	83,77
Indiv 8	100,4	164,88	109,17
Indiv 9	112,21	139,87	90,96
Indiv 10	133,53	242,92	134,25
Indiv 11	116,87	153,44	49,44
Indiv 12	195,97	115,88	61,19
Indiv 13	103,21	71,97	97,14
Indiv 14	212,82	172,2	79,46
Indiv 15	231,67	52,93	115,31
Indiv 16	221,21	93,53	146,86
Indiv 17	161,95	229,09	124
Indiv 18	161,36	118,63	114,86
Indiv 19	70,28	165,18	84,35
Indiv 20	140,09	179,82	59,06
Indiv 21	88,76	63,42	88,34
Indiv 22	194,65	140,29	217,09
Indiv 23	52,46	180,34	200,56
Indiv 24	189,88	69,44	257,54
Indiv 25	266,64	107,62	253,79

Comparaisons de moyennes

Exercice d'application :

	Centre	Montant	Short	moyennes sujets	CE sujets*k				
Indiv 1	127,95	147,45	257,65	177,6833333	3218,76179				
Indiv 2	251,06	231,98	167,44	216,8266667	15508,31232				
Indiv 3	177,5	103,12	215,06	165,2266667	1236,123844				
Indiv 4	230,08	183,61	124,42	179,37	3558,781646				
Indiv 5	142,69	52,82	50,99	82,16666667	11816,90468				
Indiv 6	244,81	161,87	150,45	185,71	4989,547198				
Indiv 7	160,86	105,23	83,77	116,62	2404,005946				
Indiv 8	100,4	164,88	109,17	124,8166667	1213,381096				
Indiv 9	112,21	139,87	90,96	114,3466667	2805,62938				
Indiv 10	133,53	242,92	134,25	170,2333333	1921,09993				
Indiv 11	116,87	153,44	49,44	106,5833333	4410,90971				
Indiv 12	195,97	115,88	61,19	124,3466667	1270,75738				
Indiv 13	103,21	71,97	97,14	90,77333333	8798,140442				
Indiv 14	212,82	172,2	79,46	154,8266667	293,9587243				
Indiv 15	231,67	52,93	115,31	133,3033333	405,3893257				
Indiv 16	221,21	93,53	146,86	153,8666667	239,7064363				
Indiv 17	161,95	229,09	124	171,68	2147,029914				
Indiv 18	161,36	118,63	114,86	131,6166667	531,5641363				
Indiv 19	70,28	165,18	84,35	106,6033333	4406,309566				
Indiv 20	140,09	179,82	59,06	126,3233333	1038,385982				
Indiv 21	88,76	63,42	88,34	80,17333333	12579,44876				
Indiv 22	194,65	140,29	217,09	184,01	4582,239438				
Indiv 23	52,46	180,34	200,56	144,4533333	0,675545653				
Indiv 24	189,88	69,44	257,54	172,2866667	2245,511812				
Indiv 25	266,64	107,62	253,79	209,35	12450,63379				
moyennes	163,5564	137,9012	133,326	moyenne générale	144,9278667				
totales				SCE totale	274691,9985				
inter	8675,55635	1234,35111	3365,08275	SCE inter	13274,99022	ddl : k-1	2	variance inter	6637,49511
intra	84859,8014	72467,2439	104089,963	SCE intra	261417,0082	ddl : kn-k	72		
sujets				SCE sujets	104073,2088	ddl : n-1	24		
erreur				SCE erreur	157343,7994	ddl : ddl intra - ddl sujets	48	variance erreur	3277,99582
								F	2,02486381

Table F : valeurs critiques de la distribution F

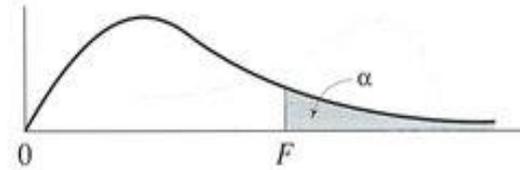


TABLE 1 α = .05

	Degrés de liberté pour le numérateur															
	1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8	252.6
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48	19.48
3	10.13	10.55	10.78	10.92	11.01	11.08	11.13	11.17	11.20	11.22	11.24	11.25	11.26	11.27	11.27	11.27
4	7.71	8.14	8.33	8.47	8.55	8.61	8.65	8.68	8.70	8.72	8.73	8.74	8.75	8.75	8.75	8.75
5	6.61	7.07	7.23	7.35	7.43	7.48	7.52	7.55	7.57	7.58	7.59	7.60	7.60	7.60	7.60	7.60
6	5.99	6.48	6.62	6.73	6.80	6.85	6.88	6.91	6.92	6.93	6.94	6.94	6.94	6.94	6.94	6.94
7	5.59	6.10	6.23	6.33	6.39	6.44	6.47	6.49	6.50	6.51	6.51	6.51	6.51	6.51	6.51	6.51
8	5.32	5.85	5.97	6.06	6.11	6.15	6.18	6.20	6.21	6.22	6.22	6.22	6.22	6.22	6.22	6.22
9	5.12	5.67	5.78	5.86	5.91	5.94	5.96	5.97	5.98	5.98	5.98	5.98	5.98	5.98	5.98	5.98
10	4.96	5.53	5.63	5.70	5.74	5.77	5.79	5.80	5.81	5.81	5.81	5.81	5.81	5.81	5.81	5.81
11	4.84	5.42	5.51	5.58	5.62	5.64	5.66	5.67	5.67	5.67	5.67	5.67	5.67	5.67	5.67	5.67
12	4.75	5.34	5.43	5.49	5.53	5.55	5.56	5.57	5.57	5.57	5.57	5.57	5.57	5.57	5.57	5.57
13	4.67	5.27	5.35	5.41	5.44	5.46	5.47	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.48
14	4.60	5.20	5.28	5.34	5.37	5.39	5.40	5.41	5.41	5.41	5.41	5.41	5.41	5.41	5.41	5.41
15	4.54	5.15	5.22	5.28	5.31	5.33	5.34	5.35	5.35	5.35	5.35	5.35	5.35	5.35	5.35	5.35
16	4.49	5.10	5.17	5.23	5.26	5.27	5.28	5.28	5.28	5.28	5.28	5.28	5.28	5.28	5.28	5.28
17	4.45	5.06	5.13	5.19	5.22	5.23	5.24	5.24	5.24	5.24	5.24	5.24	5.24	5.24	5.24	5.24
18	4.41	5.03	5.10	5.16	5.19	5.20	5.21	5.21	5.21	5.21	5.21	5.21	5.21	5.21	5.21	5.21
19	4.38	5.01	5.08	5.14	5.17	5.18	5.19	5.19	5.19	5.19	5.19	5.19	5.19	5.19	5.19	5.19
20	4.35	5.00	5.06	5.12	5.15	5.16	5.17	5.17	5.17	5.17	5.17	5.17	5.17	5.17	5.17	5.17
22	4.30	4.97	5.03	5.09	5.12	5.13	5.14	5.14	5.14	5.14	5.14	5.14	5.14	5.14	5.14	5.14
24	4.26	4.95	5.01	5.07	5.10	5.11	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12
26	4.23	4.93	4.99	5.05	5.08	5.09	5.10	5.10	5.10	5.10	5.10	5.10	5.10	5.10	5.10	5.10
28	4.20	4.92	4.97	5.03	5.06	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08
30	4.17	4.91	4.96	5.02	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07
40	4.08	4.83	4.88	4.94	4.97	4.98	4.99	4.99	4.99	4.99	4.99	4.99	4.99	4.99	4.99	4.99
50	4.03	4.78	4.83	4.89	4.92	4.93	4.94	4.94	4.94	4.94	4.94	4.94	4.94	4.94	4.94	4.94
60	4.00	4.76	4.81	4.87	4.90	4.91	4.92	4.92	4.92	4.92	4.92	4.92	4.92	4.92	4.92	4.92
120	3.92	4.70	4.75	4.81	4.84	4.85	4.86	4.86	4.86	4.86	4.86	4.86	4.86	4.86	4.86	4.86
200	3.89	4.68	4.73	4.79	4.82	4.83	4.84	4.84	4.84	4.84	4.84	4.84	4.84	4.84	4.84	4.84
500	3.86	4.66	4.71	4.77	4.80	4.81	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82
1000	3.85	4.65	4.70	4.76	4.79	4.80	4.81	4.81	4.81	4.81	4.81	4.81	4.81	4.81	4.81	4.81

F calculé = 2,02 (ddl inter = 2; ddl erreur = 48)

F lu à 0,05 = 3,18

- F calculé < F lu : je ne peux pas rejeter H0
- Les durées des 3 types de saccades sont équivalentes
- Il s'agit de données fictives !

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

- Test non-paramétrique
- Conditions :
 - Variables quantitatives (numériques) ou ordinales
- Principe semblable au test de Kruskal-Wallis

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Exemple

Avant	Après	Encore Après
27	20	34
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 1 : Attribuer un rang aux 3 mesures de chaque individu (donc en ligne)

Avant	Après	Encore Après
27	20	34
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

=MOYENNE.RANG(valeur que l'on compare; champ de valeurs auxquelles on la compare; 1)

Attention aux \$ (uniquement avant la lettre des valeurs du champ)

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 1 : Attribuer un rang aux 3 mesures de chaque individu (donc en ligne)

Avant	Après	Encore Après
Rang = 2	20	34
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

=MOYENNE.RANG(valeur que l'on compare; champ de valeurs auxquelles on la compare; 1)

Attention aux \$ (uniquement avant la lettre des valeurs du champ)

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 1 : Attribuer un rang aux 3 mesures de chaque individu (donc en ligne)

Avant	Après	Encore Après
2	Rang = 1	34
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

=MOYENNE.RANG(valeur que l'on compare; champ de valeurs auxquelles on la compare; 1)

Attention aux \$ (uniquement avant la lettre des valeurs du champ)

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 1 : Attribuer un rang aux 3 mesures de chaque individu (donc en ligne)

Avant	Après	Encore Après
2	1	Rang = 3
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

=MOYENNE.RANG(valeur que l'on compare; champ de valeurs auxquelles on la compare; 1)

Attention aux \$ (uniquement avant la lettre des valeurs du champ)

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 1 : Attribuer un rang aux 3 mesures de chaque individu (donc en ligne)

Avant	Après	Encore Après
2	1	3
1	2	3
2	3	1
1	3	2
1	2	3
2	3	1

Tirer la première ligne vers le bas

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 2 : Calculer le total des rangs de chaque groupe

	Avant	Après	Encore Après
	2	1	3
	1	2	3
	2	3	1
	1	3	2
	1	2	3
	2	3	1
Total rangs groupes	9	14	13

=SOMME (pour chaque colonne)

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 3 : appliquer la formule du test de Friedman (formule du χ^2 de rangs ou χ^2_r)

$$\chi^2_r = \frac{12}{nk(k+1)} \times \sum(Rangs^2) - 3n(k+1)$$

	Avant	Après	Encore Après
	2	1	3
	1	2	3
	2	3	1
	1	3	2
	1	2	3
	2	3	1
Total rangs groupes	9	14	13

n = effectif de chaque groupe

k = nb de groupes

Calculez les 3 termes de l'équation séparément pour éviter les erreurs

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 3 : appliquer la formule du test de Friedman (formule du χ^2 de rangs ou χ^2_r)

$$\chi^2_r = \frac{12}{(6*3(3+1))} \times (9^2+14^2+13^2) - 3*6*(3+1)$$

$$\chi^2_r = 2,33$$

	Avant	Après	Encore Après
	2	1	3
	1	2	3
	2	3	1
	1	3	2
	1	2	3
	2	3	1
Total rangs groupes	9	14	13

Calculez les 3 termes de l'équation séparément pour éviter les erreurs

Comparaisons de moyennes (plus de 2 groupes)

Groupes appariés + loi non-normale : test de Friedman

Etape 4 : comparer la valeur calculé à la valeur lue sur la table du χ^2 (ou « Khi2 »)

Risque : 0,05

ddl = $K-1$ (nombre de groupes -1)

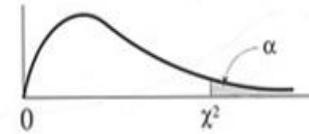
-> $3 - 1 = 2$

	Avant	Après	Encore Après
	2	1	3
	1	2	3
	2	3	1
	1	3	2
	1	2	3
	2	3	1
Total rangs groupes	9	14	13
χ^2	2,333333333		

Si χ^2_r (calculé) > χ^2 (lu) : je peux rejeter H0

Si χ^2_r (calculé) < χ^2 (lu) je ne peux pas rejeter H0

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2



2,33 < 5,99

- Je ne peux pas rejeter H0
- Il n'y a pas de différence significative entre les groupes

df	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

Comparaisons de moyennes

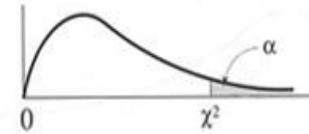
Exercice d'application :

Les spectateurs de la Fan Zone de Paris sont-ils davantage sensibles à certaines nuisances plutôt qu'à d'autres ?

Bruit	Pb transports	Déchets	Présence policière
2	3	4	4
1	4	1	1
3	4	5	3
2	3	3	1
2	2	2	1
1	1	1	1
2	2	2	3
4	2	1	1
2	3	3	2
1	4	1	1
1	1	1	1
2	2	2	1
2	3	1	1
1	2	4	1
3	1	1	1
1	2	1	1
3	1	2	1
2	3	1	1
1	1	3	2
1	1	2	1
5	5	5	1
4	5	5	1
1	1	3	3
1	1	3	1
1	3	3	2
2	3	1	1
2	1	1	1
2	2	1	1
4	5	4	3
2	1	1	2
4	4	5	1
1	1	1	1
3	2	2	3
3	3	3	1
2	1	1	4
3	4	2	1
1	1	1	1

Comparaisons de moyennes

	Bruit	Pb transports	Déchets	Présence policière
	1	2	3,5	3,5
	2	4	2	2
	1,5	3	4	1,5
	2	3,5	3,5	1
	3	3	3	1
	2,5	2,5	2,5	2,5
	2	2	2	4
	4	3	1,5	1,5
	1,5	3,5	3,5	1,5
	2	4	2	2
	2,5	2,5	2,5	2,5
	3	3	3	1
	3	4	1,5	1,5
	1,5	3	4	1,5
	4	2	2	2
	2	4	2	2
	4	1,5	3	1,5
	3	4	1,5	1,5
	1,5	1,5	4	3
	2	2	4	2
	3	3	3	1
	2	3,5	3,5	1
	1,5	1,5	3,5	3,5
	2	2	4	2
	1	3,5	3,5	2
	3	4	1,5	1,5
	4	2	2	2
	3,5	3,5	1,5	1,5
	2,5	4	2,5	1
	3,5	1,5	1,5	3,5
	2,5	2,5	4	1
	2,5	2,5	2,5	2,5
	3,5	1,5	1,5	3,5
	3	3	3	1
	3	1,5	1,5	4
	3	4	2	1
	2,5	2,5	2,5	2,5
Total rangs groupes	94	104	98,5	73,5
n	37			
k	4			
χ^2_r	8,618918919			

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2 

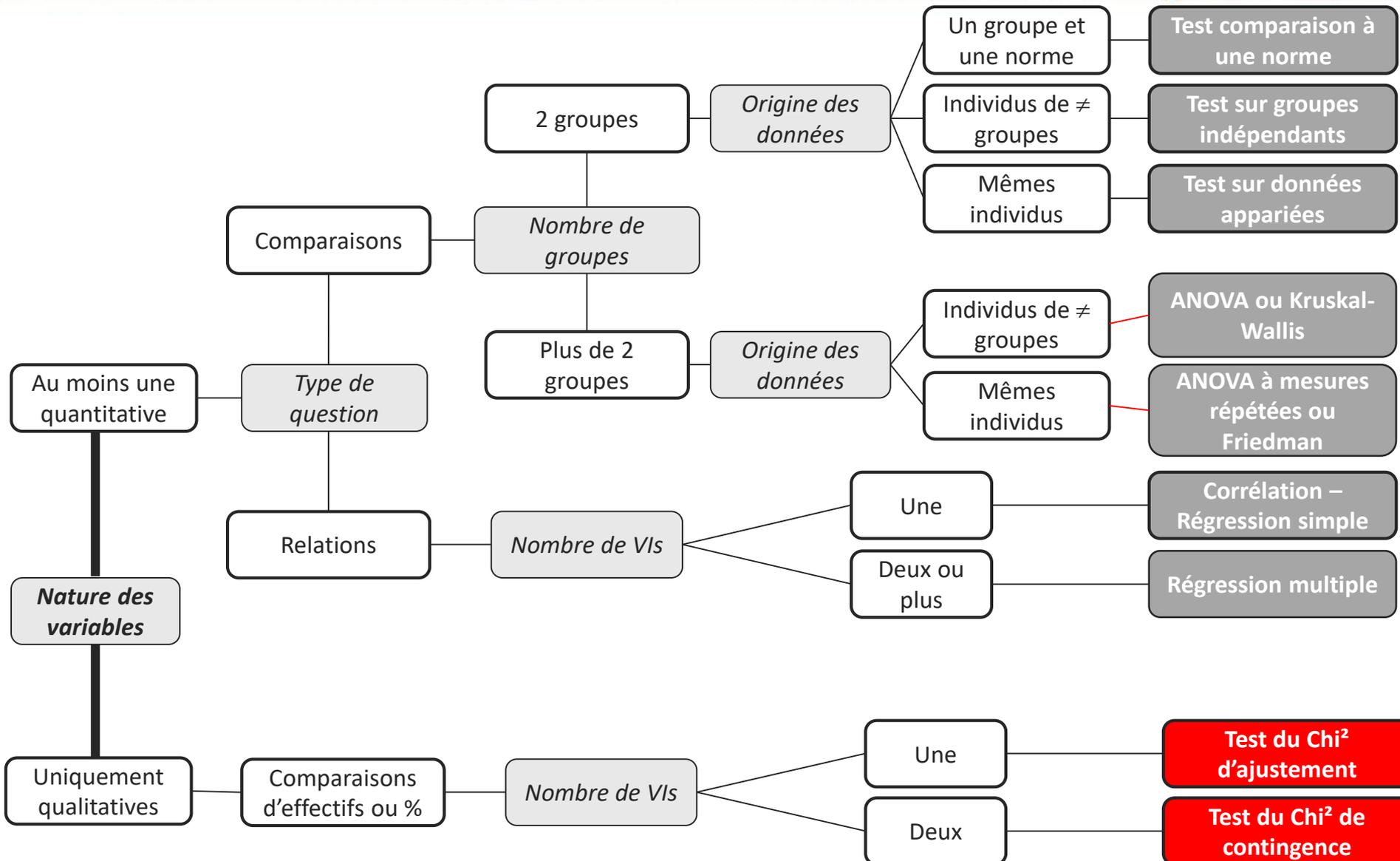
F calculé = 8,62

F lu (ddl = k - 1 soit 3) = 7,82

- F calculé > F lu
- Je peux rejeter H0
- Les spectateurs sont particulièrement sensibles à certaines nuisances
- Il faudrait faire des tests post-hoc pour identifier les nuisances significativement plus gênantes (a priori les problèmes de transports, peut-être aussi les déchets).
- Quant à la présence policière elle semble significativement moins gênante

dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

Comparaisons de pourcentages



Comparaisons de pourcentages

- De la même manière que l'on compare des variables quantitatives (et ordinales), il est possible de comparer des variables qualitatives (qui s'expriment généralement en pourcentages)
- Pour cela, on utilise les test du χ^2 (dont on a déjà vu la table)
- Condition : les modalités de nos variables doivent être mutuellement exclusives (soit « droitier », soit « gaucher », le cas échéant « soit ambidextre »)
- Il existe 2 types de test du χ^2 :
 - Test d'ajustement
 - Test de contingence

Le χ^2 d'ajustement

- Il s'utilise lorsqu'on n'a qu'un seul groupe (avec k modalités).
- Il permet :
 - De comparer une distribution observée à une distribution théorique (principe de la comparaison à une norme que l'on a déjà vu)
 - De comparer deux distributions observées

Comparaisons de pourcentages

Le χ^2 d'ajustement (1)

- Comparer une distribution observée à une distribution théorique.

Exemple – La fréquentation hebdomadaire du magasin Décathlon de Brétigny-sur-Orge se répartit de la manière suivante :

L	M	M	J	V	S
442	511	947	765	1041	1763

Cette fréquentation est-elle conforme à la fréquentation nationale ?

L	M	M	J	V	S
10%	10%	15%	15%	20%	30%

Comparaisons de pourcentages

Le χ^2 d'ajustement (1)

- Comparer une distribution observée à une distribution théorique.

Distribution observée (en effectifs)



L	M	M	J	V	S
442	511	947	765	1041	1763

Distribution théorique (en pourcentages)



L	M	M	J	V	S
10%	10%	15%	15%	20%	30%

Comparaisons de pourcentages

Le χ^2 d'ajustement (1)

- Comparer une distribution observée à une distribution théorique.

1^{ère} étape : je mets ma distribution théorique en effectifs

	L	M	M	J	V	S	TOTAL
observée	442	511	947	765	1041	1763	5469

Comparaisons de pourcentages

Le χ^2 d'ajustement (1)

- Comparer une distribution observée à une distribution théorique.

1^{ère} étape : je mets ma distribution théorique en effectifs

	L	M	M	J	V	S	TOTAL
observée	442	511	947	765	1041	1763	5469
théorique %	10%	10%	15%	15%	20%	30%	100%
théorique effectif							


5469*10%

Comparaisons de pourcentages

Le χ^2 d'ajustement (1)

- Comparer une distribution observée à une distribution théorique.

1^{ère} étape : je mets ma distribution théorique en effectifs

	L	M	M	J	V	S	TOTAL
observée	442	511	947	765	1041	1763	5469
théorique %	10%	10%	15%	15%	20%	30%	100%
théorique effectif	546,9						



$5469 * 10\%$

Comparaisons de pourcentages

Le χ^2 d'ajustement (1)

- Comparer une distribution observée à une distribution théorique.

1^{ère} étape : je mets ma distribution théorique en effectifs

	L	M	M	J	V	S	TOTAL
observée	442	511	947	765	1041	1763	5469
théorique %	10%	10%	15%	15%	20%	30%	100%
théorique effectif	546,9	546,9	820,35	820,35	1093,8	1640,7	5469

Comparaisons de pourcentages

Le χ^2 d'ajustement (1)

- Comparer une distribution observée à une distribution théorique.

2^{ème} étape : je calcule mon χ^2 via la formule suivante :

$$\chi^2 = \sum \frac{(\text{effectif observé} - \text{effectif théorique})^2}{\text{effectif théorique}}$$

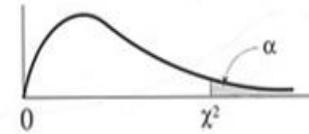
	L	M	M	J	V	S	TOTAL
observée	442	511	947	765	1041	1763	5469
théorique %	10%	10%	15%	15%	20%	30%	100%
théorique effectif	546,9	546,9	820,35	820,35	1093,8	1640,7	5469
χ^2	20,1207	2,356573	19,5529	3,734531	2,548766	9,116408	57,42988

Le χ^2 d'ajustement (1)

- Comparer une distribution observée à une distribution théorique.

3^{ème} étape : je compare mon χ^2 calculé avec mon χ^2 lu sur la table

- à 0;05
- ddl = (nb de modalités – 1) soit 5

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2 

dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

χ^2 calculé = 57,43

χ^2 lu (ddl = 5) = 11,07

- χ^2 calculé > χ^2 lu
- Je peux rejeter H0
- La fréquentation du magasin Décathlon de Brétigny-sur-Orge est différente de la fréquentation nationale

Comparaisons de pourcentages

Le χ^2 d'ajustement (2)

- Comparer deux distributions observées.

Exemple – La fréquentation hebdomadaire du magasin Décathlon de Brétigny-sur-Orge est-elle comparable à celle du magasin Décathlon de Toulouse-centre ?

Brétigny-sur-Orge

L	M	M	J	V	S
442	511	947	765	1041	1763

Toulouse-centre

L	M	M	J	V	S
321	416	1001	875	921	1813

Comparaisons de pourcentages

Le χ^2 d'ajustement (2)

- Comparer deux distributions observées.

1^{ère} étape : je réunis les deux échantillons pour calculer mes proportions théoriques...

	L	M	M	J	V	S	TOTAL Bretigny	
	442	511	947	765	1041	1763	5469	
+	L	M	M	J	V	S	TOTAL Toulouse	
	321	416	1001	875	921	1813	5347	
=	L	M	M	J	V	S	TOTAL Théorique	
	En effectifs	763	927	1948	1640	1962	3576	10816
	En pourcentages	7,05%	8,57%	18,01%	15,16%	18,14%	33,06%	100,00%

Pour la partie « en pourcentages » : il suffit de diviser l'effectif par le total et de mettre le résultat au format pourcentage

Comparaisons de pourcentages

Le χ^2 d'ajustement (2)

- Comparer deux distributions observées.

2^{ème} étape : ... que je réaffecte aux échantillons

	L	M	M	J	V	S	TOTAL Bretigny
	442	511	947	765	1041	1763	5469
=TOTAL Brétigny * 7,05%, etc...	385,80316	468,728	984,986	829,249	992,065	1808,17	5469
	L	M	M	J	V	S	TOTAL Toulouse
	321	416	1001	875	921	1813	5347
=TOTAL Toulouse * 7,05%, etc...	377,19684	458,272	963,014	810,751	969,935	1767,83	5347
	L	M	M	J	V	S	TOTAL Théorique
En effectifs	763	927	1948	1640	1962	3576	10816
En pourcentages	7,05%	8,57%	18,01%	15,16%	18,14%	33,06%	100,00%

Comparaisons de pourcentages

Le χ^2 d'ajustement (2)

- Comparer deux distributions observées.

3^{ème} étape : je calcule mon χ^2 via la formule suivante :

$$\chi^2 = \sum \frac{(\text{effectif observé} - \text{effectif théorique})^2}{\text{effectif théorique}}$$

L	M	M	J	V	S	TOTAL Bretagne
442	511	947	765	1041	1763	5469
385,803162	468,728	984,986	829,249	992,065	1808,17	5469
8,18574059	3,81226	1,46495	4,97796	2,41376	1,12829	
L	M	M	J	V	S	TOTAL Toulouse
321	416	1001	875	921	1813	5347
377,196838	458,272	963,014	810,751	969,935	1767,83	5347
8,37251081	3,89925	1,49838	5,09154	2,46883	1,15403	

Attention je dois faire la somme des valeurs calculées pour les deux échantillons

Comparaisons de pourcentages

Le χ^2 d'ajustement (2)

- Comparer deux distributions observées.

3^{ème} étape : je calcule mon χ^2 via la formule suivante :

$$\chi^2 = \sum \frac{(\text{effectif observé} - \text{effectif théorique})^2}{\text{effectif théorique}}$$

L	M	M	J	V	S	TOTAL Bretagne
442	511	947	765	1041	1763	5469
385,803162	468,728	984,986	829,249	992,065	1808,17	5469
8,18574059	3,81226	1,46495	4,97796	2,41376	1,12829	21,98296585
L	M	M	J	V	S	TOTAL Toulouse
321	416	1001	875	921	1813	5347
377,196838	458,272	963,014	810,751	969,935	1767,83	5347
8,37251081	3,89925	1,49838	5,09154	2,46883	1,15403	22,48454091
					χ^2	44,46750676

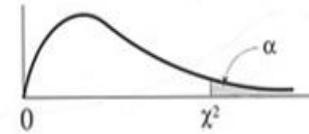
Attention je dois faire la somme des valeurs calculées pour les deux échantillons

Le χ^2 d'ajustement (2)

- Comparer deux distributions observées.

4^{ème} étape : je compare mon χ^2 calculé avec mon χ^2 lu sur la table

- à 0;05
- ddl = (nb de groupes – 1) x (nb de modalités – 1)
= (2-1) x (6-1) = 5

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2 

dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

χ^2 calculé = 44,47

χ^2 lu (ddl = 5) = 11,07

- χ^2 calculé > χ^2 lu
- Je peux rejeter H0
- Les fréquentations des deux magasins ne sont pas équivalentes

Le χ^2 de contingence

- Il s'utilise lorsqu'on a deux variables qualitatives différentes (qui ont chacune un certain nb de modalités).
- Il permet :
 - De déterminer si ces deux variables sont liées ou pas

Le χ^2 de contingence

Exemple – y a-t-il un lien entre l'attachement à un sport (mainstream) et l'attachement à une marque de sport (mainstream) ?

A partir d'un questionnaire :

- 1) Quel est votre sport favori ?
- 2) Quelle est votre marque de sport favorite ?

Le χ^2 de contingence

1^{ère} étape : faire un tableau de contingence (tableau croisé dynamique sur excel)

- *Insertion*
- *Tableau croisé dynamique*
- *On sélectionne l'ensemble des données*
- *Sport en ligne, marques en colonnes (ou l'inverse)*
- *Remplir le champ valeurs (avec l'un ou l'autre c'est pareil)*

Comparaisons de pourcentages

Le χ^2 de contingence

1^{ère} étape : faire un tableau de contingence (tableau croisé dynamique sur excel)

	Nombre de Sp	Étiquettes	Adidas	Nike	Puma	Total général
Athlétisme	15	26	71	112		
Basketball	41	60	13	114		
Football	32	26	46	104		
Rugby	54	42	9	105		
Total général	142	154	139	435		

Comparaisons de pourcentages

Le χ^2 de contingence

2^{ème} étape : copier-coller le tableau croisé en tableau excel pour pouvoir travailler dessus

	Adidas	Nike	Puma	Total général
Athlétisme	15	26	71	112
Basketball	41	60	13	114
Football	32	26	46	104
Rugby	54	42	9	105
Total général	142	154	139	435

Comparaisons de pourcentages

Le χ^2 de contingence

3^{ème} étape : générer le tableau des effectifs théoriques

	Adidas	Nike	Puma	Total général
Athlétisme	15	26	71	112
Basketball	41	60	13	114
Football	32	26	46	104
Rugby	54	42	9	105
Total général	142	154	139	435

Pour chaque case :
$$\frac{\text{total en ligne} * \text{total en colonne}}{\text{total général}}$$

	Adidas	Nike	Puma	Total général
Athlétisme	36,56091954	39,65057471	35,78850575	112
Basketball	37,2137931	40,35862069	36,42758621	114
Football	33,94942529	36,8183908	33,23218391	104
Rugby	34,27586207	37,17241379	33,55172414	105
Total général	142	154	139	435

Pour aller plus vite :
\$ colonne pour le total en ligne
\$ ligne pour le total en colonne
\$\$ pour le total général

Comparaisons de pourcentages

Le χ^2 de contingence

4^{ème} étape : générer le tableau des χ^2

$$\chi^2 = \sum \frac{(\text{effectif observé} - \text{effectif théorique})^2}{\text{effectif théorique}}$$

	Adidas	Nike	Puma	Total général
Athlétisme	15	26	71	112
Basketball	41	60	13	114
Football	32	26	46	104
Rugby	54	42	9	105
Total général	142	154	139	435

	Adidas	Nike	Puma	Total général
Athlétisme	36,56091954	39,65057471	35,78850575	112
Basketball	37,2137931	40,35862069	36,42758621	114
Football	33,94942529	36,8183908	33,23218391	104
Rugby	34,27586207	37,17241379	33,55172414	105
Total général	142	154	139	435

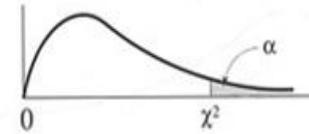
	Adidas	Nike	Puma	Total général
Athlétisme	12,71503171	4,699507922	34,64378581	52,05832544
Basketball	0,38521638	9,5588941	15,06692737	25,01103785
Football	0,11193883	3,178780415	4,905399182	8,196118427
Rugby	11,35030875	0,626959248	17,96590708	29,94317507
Total général	24,56249567	18,06414169	72,58201943	115,2086568

Comparaisons de pourcentages

Le χ^2 de contingence

5^{ème} étape : je compare mon χ^2 calculé avec mon χ^2 lu sur la table

- à 0;05
- ddl = (lignes - 1) x (colonnes - 1)
= (4-1) x (3-1) = 6

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2 

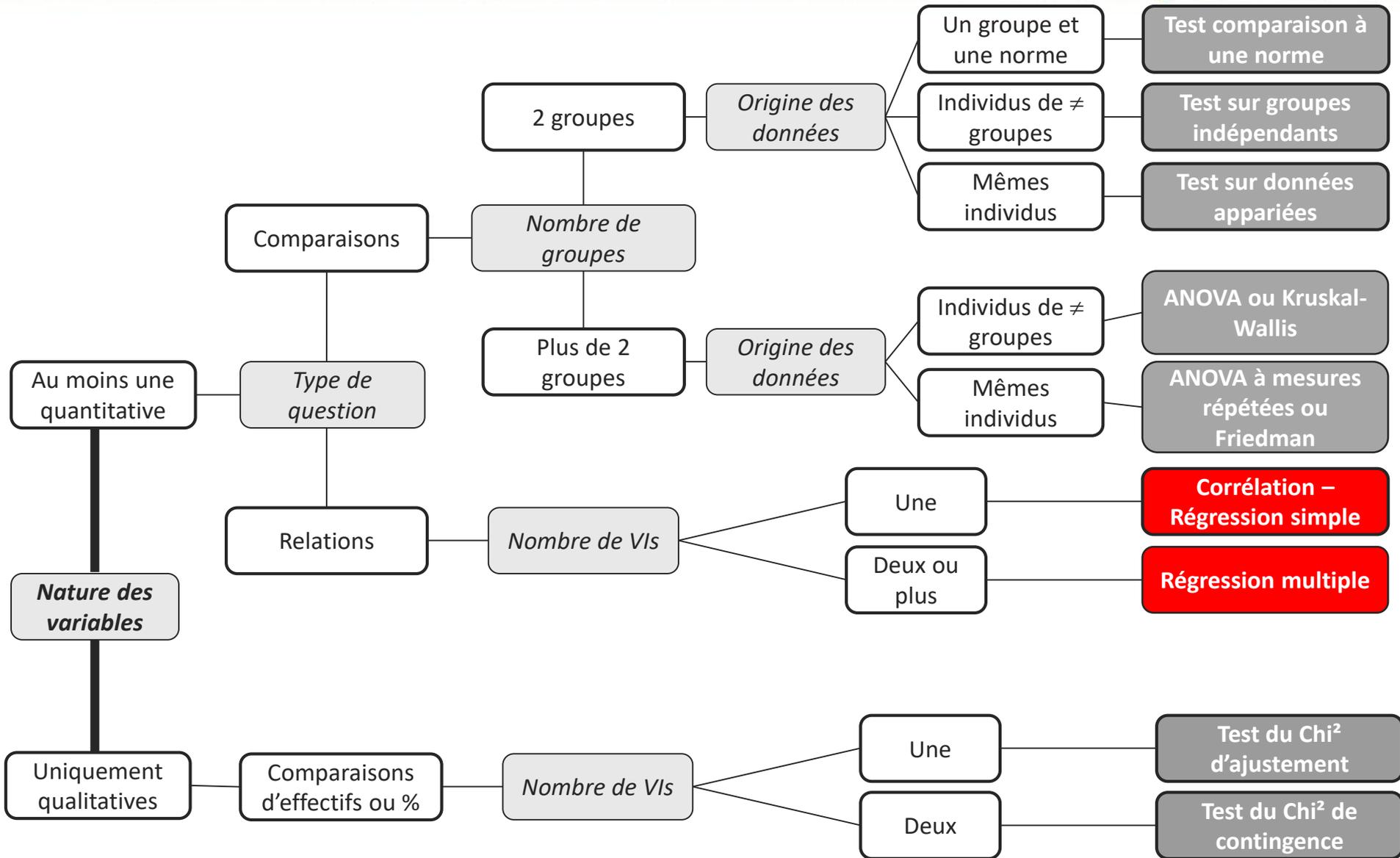
dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.20	7.88	10.59	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

χ^2 calculé = 115,21

χ^2 lu (ddl = 5) = 12,59

- χ^2 calculé > χ^2 lu
- Je peux rejeter H0
- Il y a un lien (fort) entre l'attachement à un sport et l'attachement à une marque sportive
- A partir de là je peux regarder les différences observées vs théoriques pour interpréter
- Exemple : les fan d'athlétisme apprécie davantage la marque Puma

Corrélations et régressions



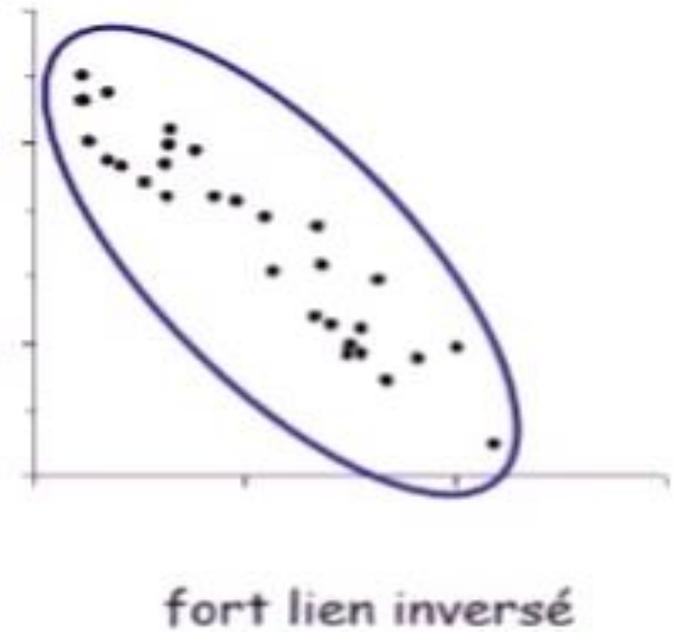
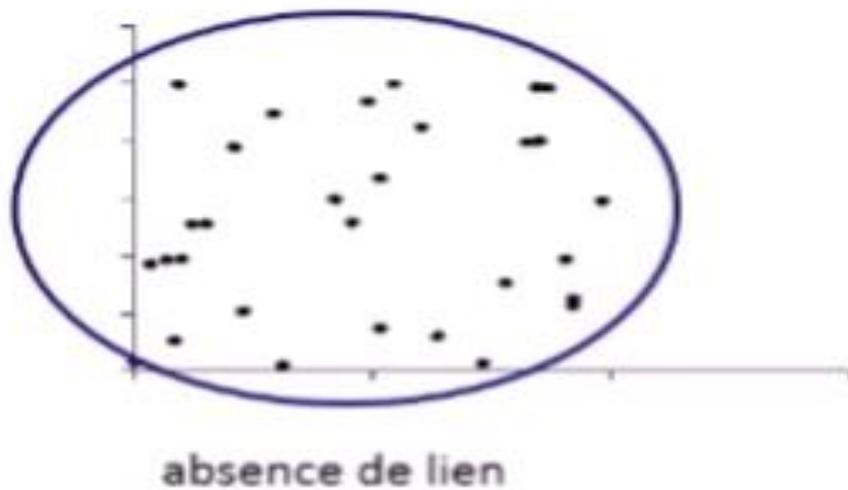
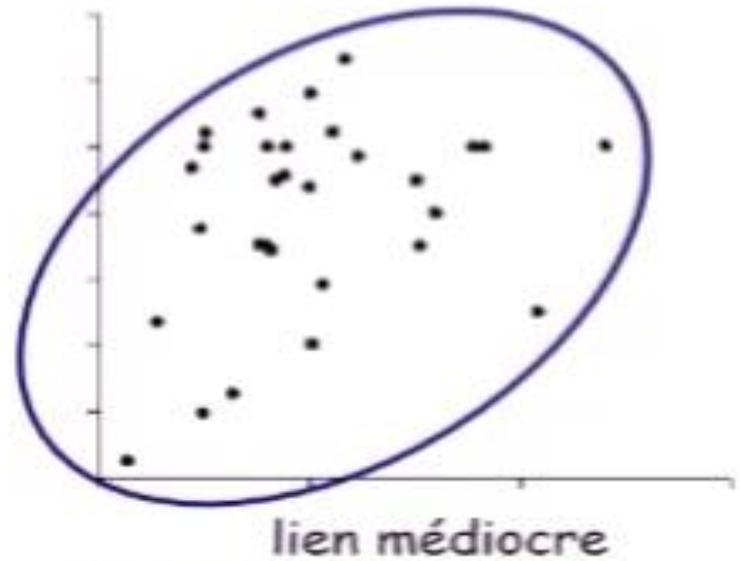
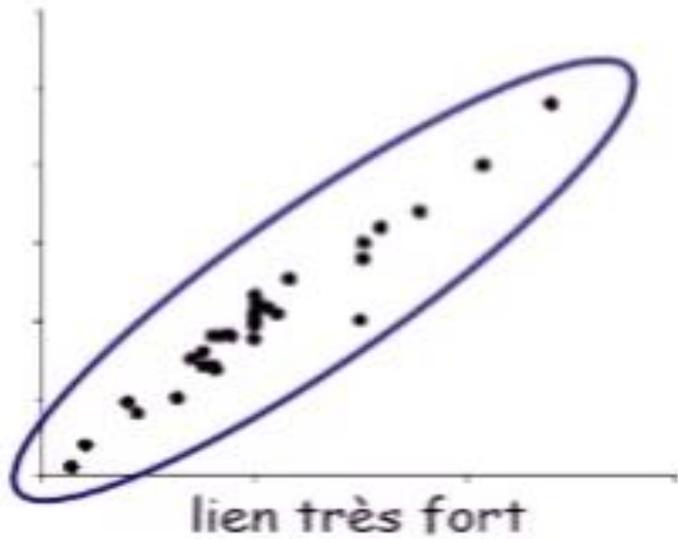
La corrélation

- corrélation et covariance : La **covariance** est le lien qui peut exister entre plusieurs variables, telle que les variations de leurs valeurs soient de même sens (positive) ou de sens opposé (négative).
- La **corrélation** est l'expression standardisée de la covariance. Elle permet d'exprimer de manière uniformisée le lien entre des variables mesurées dans des unités différentes (kg, cm, années, euros, etc...) -> un peu le même rôle que l'écart-type.
- Il existe différents types de corrélations. Nous ne nous intéressons qu'à la corrélation linéaire.

La corrélation

- On utilise la corrélation pour déterminer le lien entre deux variables n'ayant pas de « rôle » présumé (VI ou VD)
- Exemples :
 - Y a-t-il un lien entre les notes en maths et en physique ?

Corrélations et régressions



Corrélations et régressions

La corrélation

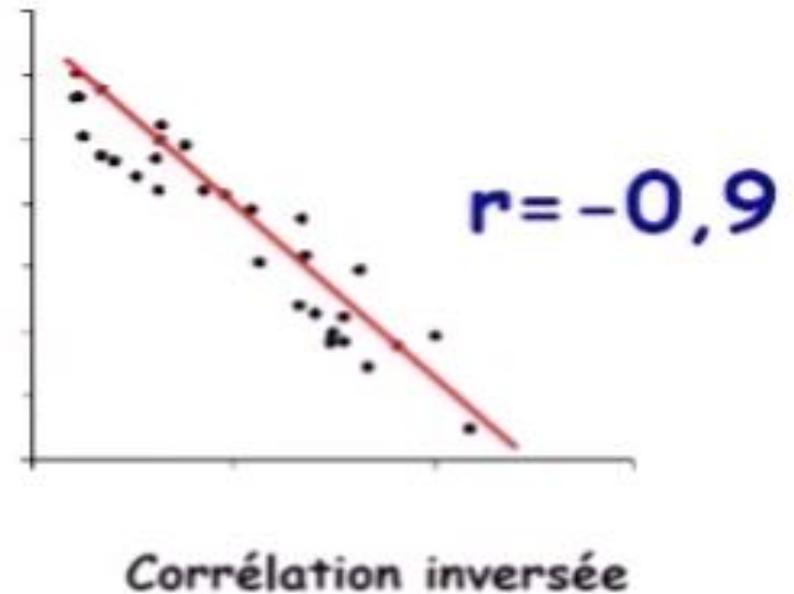
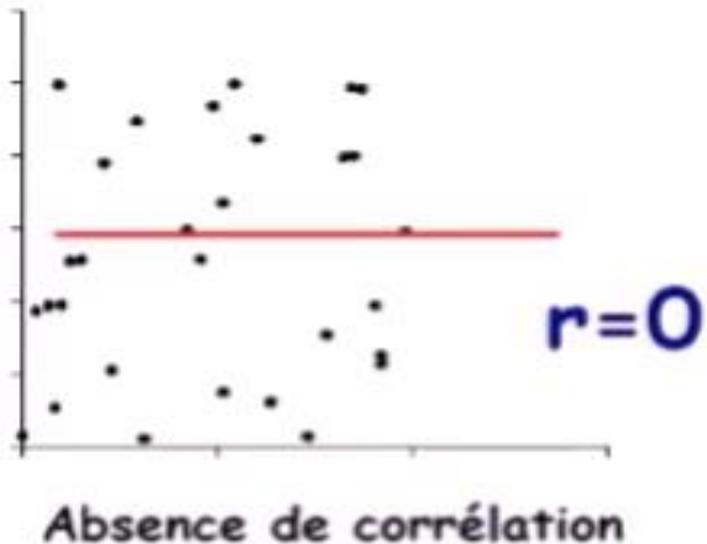
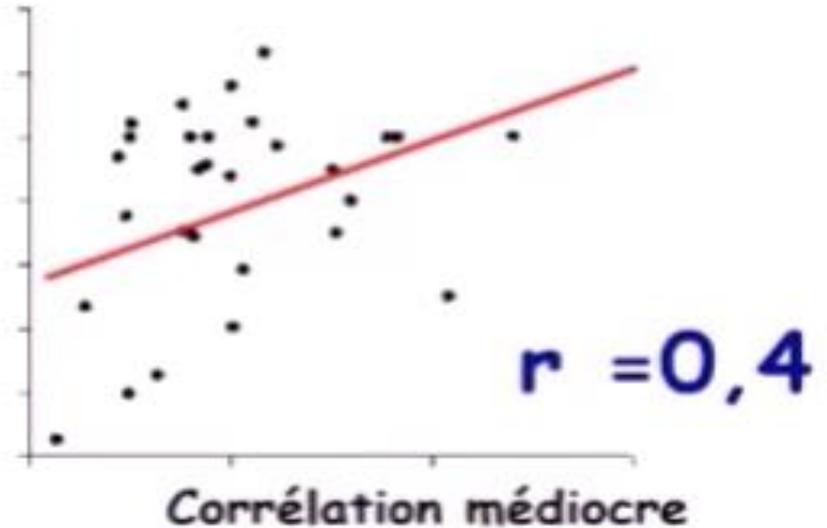
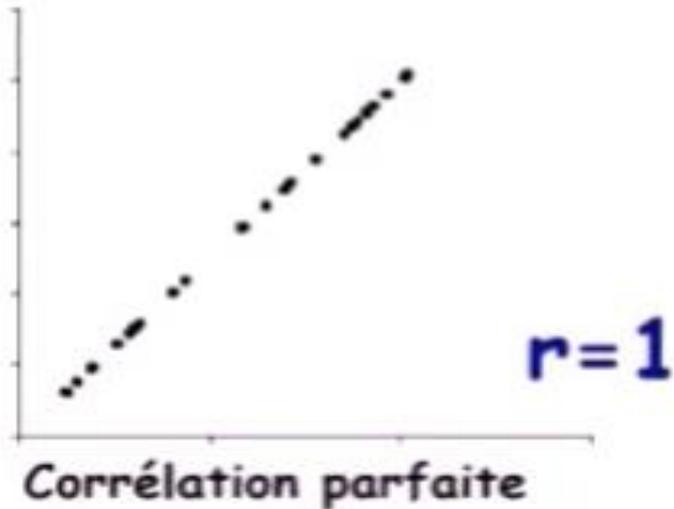
Le coefficient de corrélation « r » de Pearson

- Le coefficient de corrélation « r » exprime l'intensité du lien entre les deux variables
- Par construction, il est forcément compris entre -1 et 1.
- Si le coefficient se rapproche de 0, le lien est faible. Plus il se rapproche de -1 ou de 1, plus le lien est fort.

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}$$


$$r = \frac{\text{covariance } XY}{\sqrt{\text{var } X \text{ var } Y}}$$

Corrélations et régressions



La corrélation

Le test de significativité d'une corrélation.

- Le coefficient de corrélation donne l'intensité du lien entre les données de notre échantillon, mais il ne permet pas d'affirmer l'existence de la corrélation (statistique inférentielle).
- Pour cela, il faut appliquer un test de significativité (ce qu'on fait depuis le début !).



- H_0 : le coefficient de corrélation est nul : il n'y a pas de lien significatif entre les deux variables
- H_1 : le coefficient de corrélation n'est pas nul (positif ou négatif) : il y a un lien significatif entre les deux variables

Corrélations et régressions

La corrélation

Mise en œuvre sous excel

variable x	variable y
9,7	24,7
9	21,9
15,3	42,7
4,4	10
11,7	29,7
8,3	16
13,9	37,8
5,7	12,2
12,6	36,1
6,8	19,5
16,2	34,1
11,6	19,2
4,8	17,1
7,3	11,9
13,3	26,3
5,1	23,6
9,2	14,1
11,8	16,3
8,6	31,8
15,2	26,4

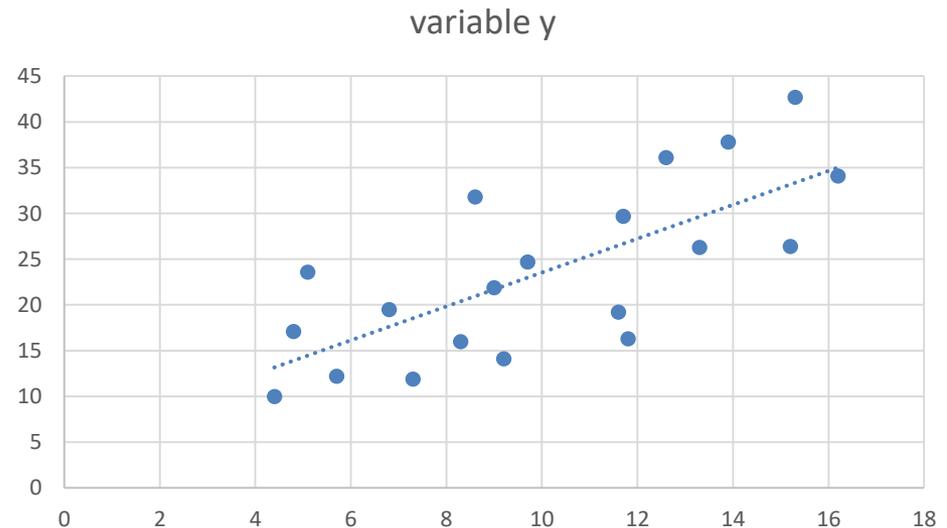
Corrélations et régressions

La corrélation

Mise en œuvre sous excel

Etape 1 : visualiser le nuage de point

- Sélectionner l'ensemble des données
- Insertion – nuage de points
- + courbe de tendance linéaire



Il semble y avoir une corrélation linéaire

Corrélations et régressions

La corrélation

Mise en œuvre sous excel

Etape 2 : déterminer le coefficient de corrélation

- = COEFFICIENT.CORRELATION (une colonne; l'autre)

variable x	variable y
9,7	24,7
9	21,9
15,3	42,7
4,4	10
11,7	29,7
8,3	16
13,9	37,8
5,7	12,2
12,6	36,1
6,8	19,5
16,2	34,1
11,6	19,2
4,8	17,1
7,3	11,9
13,3	26,3
5,1	23,6
9,2	14,1
11,8	16,3
8,6	31,8
15,2	26,4

coeff corrél 0,72735135

Le coefficient est assez élevé

Corrélations et régressions

La corrélation

Mise en œuvre sous excel

Etape 3 : tester la significativité de la corrélation

- Un peu compliqué ! Excel ne propose pas le test direct
- Données – utilitaire d'analyse (tout à droite)

test régression.xlsx - Excel

CONNEXION

Utilitaire d'analyse

Outils d'analyse de données

Outils pour l'analyse de données financières et scientifiques

FUNCRES.XLAM

En savoir plus

variable y

45

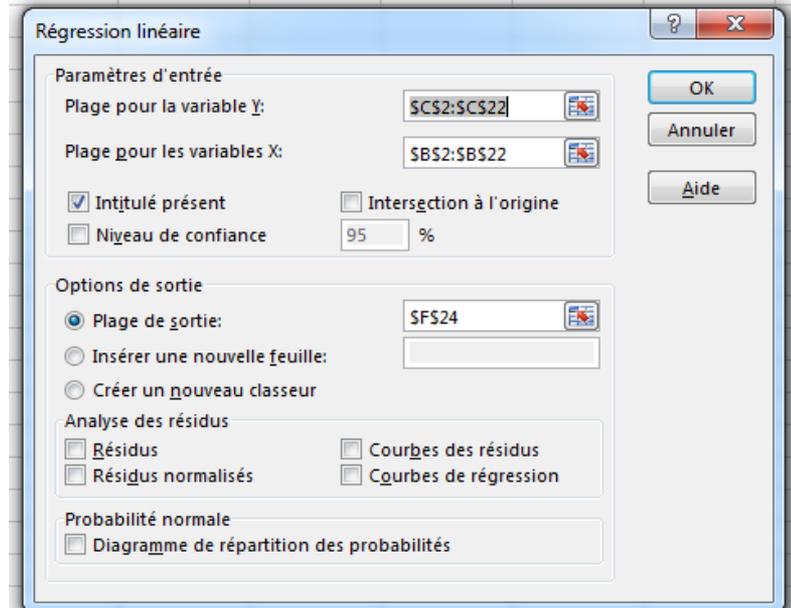
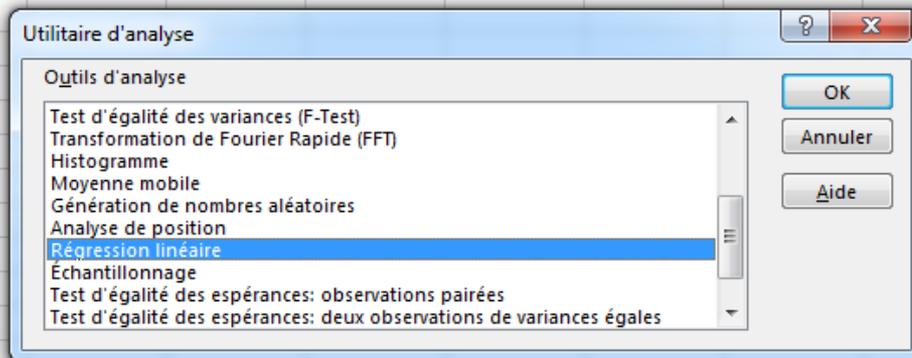
Corrélations et régressions

La corrélation

Mise en œuvre sous excel

Etape 3 : tester la significativité de la corrélation

- Aller à « régression linéaire » (et non à « corrélation »), entrez les champs nécessaires...



Corrélations et régressions

La corrélation

Mise en œuvre sous excel

Etape 3 : tester la significativité de la corrélation

- Ne regardez que la valeur critique de F du tableau d'analyse de la variance

RAPPORT DÉTAILLÉ									
<i>Statistiques de la régression</i>									
Coefficient c	0,72735135								
Coefficient c	0,52903999								
Coefficient c	0,50287554								
Erreur-type	6,64599871								
Observation	20								
<i>ANALYSE DE VARIANCE</i>									
	<i>Degré de liberté</i>	<i>mme des car</i>	<i>enne des ca</i>	<i>F</i>	<i>Valeur critique de F</i>				
Régression	1	893,094622	893,094622	20,2198053	0,000279058				
Résidus	18	795,047378	44,1692988						
Total	19	1688,142							
	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>	<i>seur pour seuil de conf</i>	<i>pour seuil de</i>	<i>our seuil de</i>	<i>our seuil de</i>	<i>confiance = 95,0%</i>
Constante	4,99416552	4,39021544	1,1375673	0,27021771	-4,22933486	14,2176659	-4,22933486	14,2176659	
variable x	1,85295107	0,41207424	4,49664378	0,00027906	0,987215216	2,71868692	0,98721522	2,71868692	

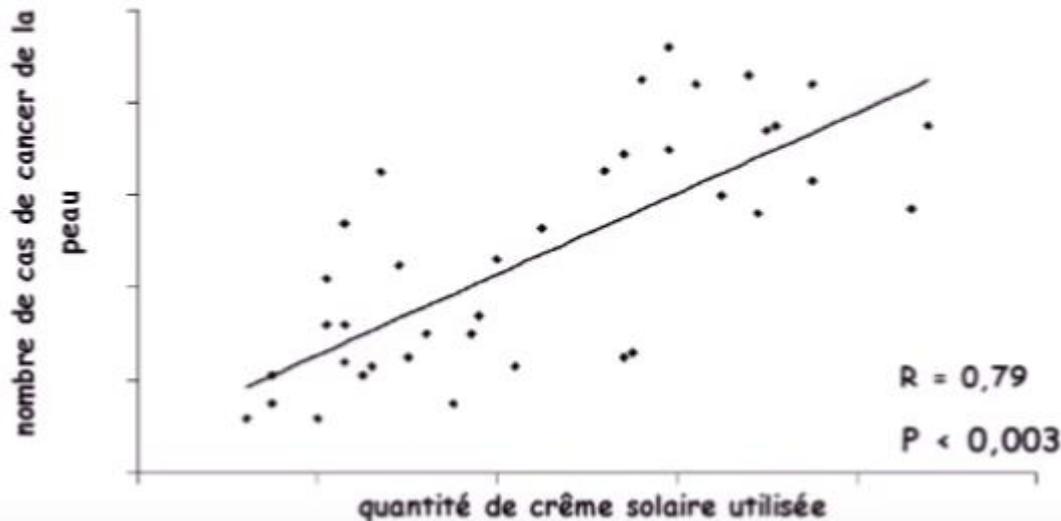
Il s'agit directement du p. Il est très inférieur à 0,05 : la corrélation est très significative et assez forte.

Corrélations et régressions

La corrélation

Attention ! Corrélation n'est pas causalité !

Utilisation de crème solaire et cancer de la peau (par pays)



Les crèmes solaires seraient cancérigènes ?
Non, ces deux variables ont une même cause : la quantité d'exposition au soleil (qui est à la fois une cause de cancer de la peau et d'utilisation de crème solaire)

La corrélation

Attention ! Corrélacion n'est pas causalité !

- Le fait de dormir avec des chaussures est corrélié au fait de se réveiller avec le mal de tête. Peut-on en conclure que dormir avec des chaussures fait mal à la tête ? Une explication plus vraisemblable est que ces deux événements font suite à des soirées trop arrosées...
- Imaginons que la consommation de cannabis soit corréliée avec des résultats scolaires inférieurs à la moyenne. Il se peut que fumer soit la cause de moins bons résultats. Mais il se peut aussi qu'avoir de moins bons résultats conduise à fumer. Ou encore que certaines personnes tendent à la fois à fumer du cannabis et à prendre leurs résultats moins au sérieux (cause psychologique ?)

Corrélations et régressions

Exercice d'application :

Y a-t-il un lien entre le temps passé à jouer à FIFA 18 et le temps consacré à suivre l'actualité du football ?

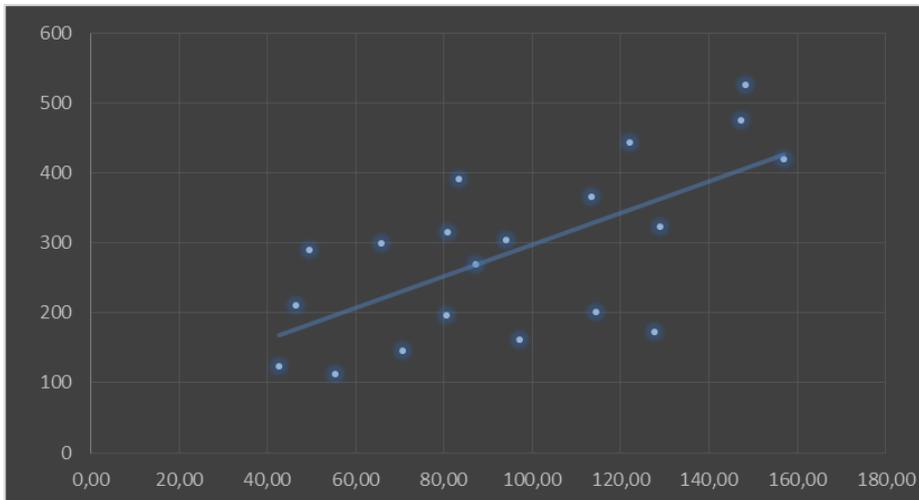
	Nb h / FIFA 18	Nb h /suivi actu foot
Indiv1	94,02	303,81
Indiv2	87,23	269,37
Indiv3	148,29	525,21
Indiv4	42,65	123
Indiv5	113,40	365,31
Indiv6	80,45	196,8
Indiv7	80,88	314,94
Indiv8	55,25	112,56
Indiv9	122,12	444,03
Indiv10	65,91	299,85
Indiv11	157,02	419,43
Indiv12	97,05	161,16
Indiv13	46,52	210,33
Indiv14	70,75	146,37
Indiv15	128,91	323,49
Indiv16	49,43	290,28
Indiv17	127,63	173,43
Indiv18	114,37	200,49
Indiv19	83,35	391,14
Indiv20	147,32	474,72



Corrélations et régressions

Exercice d'application :

Y a-t-il un lien entre le temps passé à jouer à FIFA 18 et le temps consacré à suivre l'actualité du football ?



coeff corré 0,658536246

ANALYSE DE VARIANCE					
	Degré de liberté	mmme des car	enne des cai	F	leur critique d
Régression	1	122021,983	122021,983	13,7835884	0,00159322
Résidus	18	159348,613	8852,70072		
Total	19	281370,596			

Il y a une corrélation significative ($p=0,001$) et modérée ($r = 0,66$) entre le temps passé à jouer à FIFA 18 et le temps consacré à suivre l'actualité du football

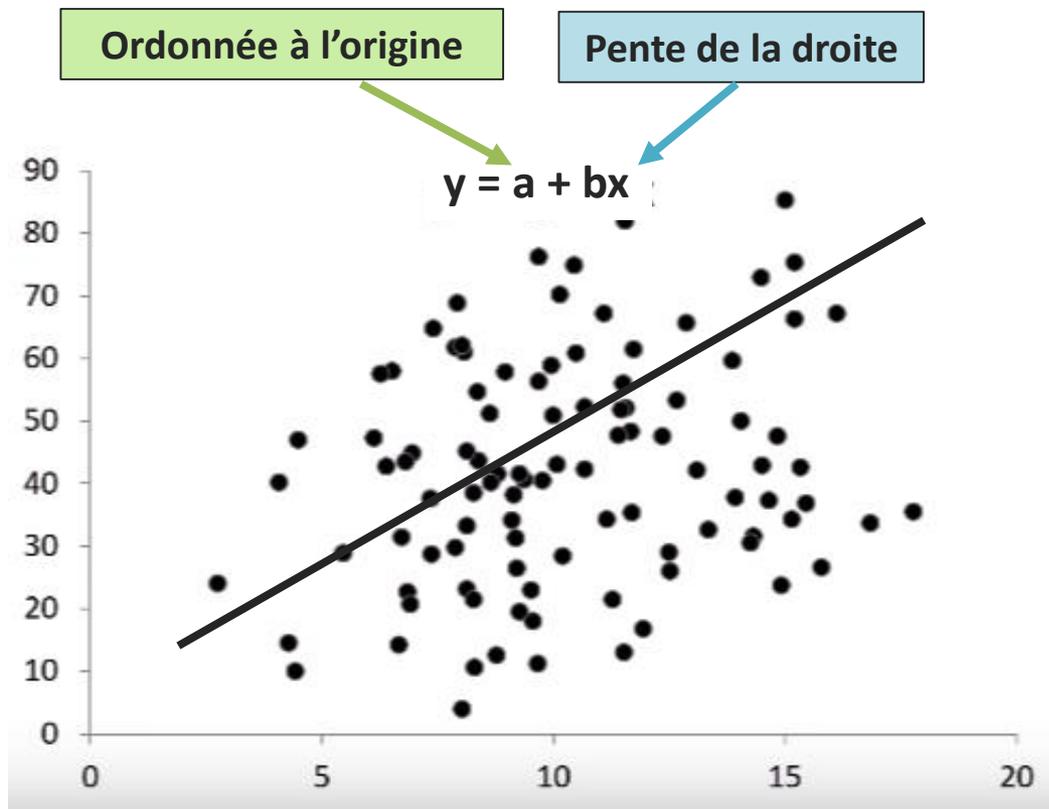
La régression linéaire simple

- Par plusieurs aspects, corrélation et régression se ressemblent
- A la différence de la corrélation, la régression a un sens déterminé : une variable dépendante (notée y) et une variable indépendante/explicative (notée x)
- Exemples :
 - **y a-t-il un lien entre le QI et la note en maths ?**
 - Autrement dit le QI (VI) détermine-t-il la note en maths (VD) ?
 - **y a-t-il un lien entre l'âge et le revenu ?**
 - Autrement dit, est-ce que plus on avance en âge (VI), plus notre revenu est important (VD) ?
 - Inverser le rôle des variables serait absurde (plus on gagne d'argent et plus on vieillit !)

Corrélations et régressions

La régression linéaire simple

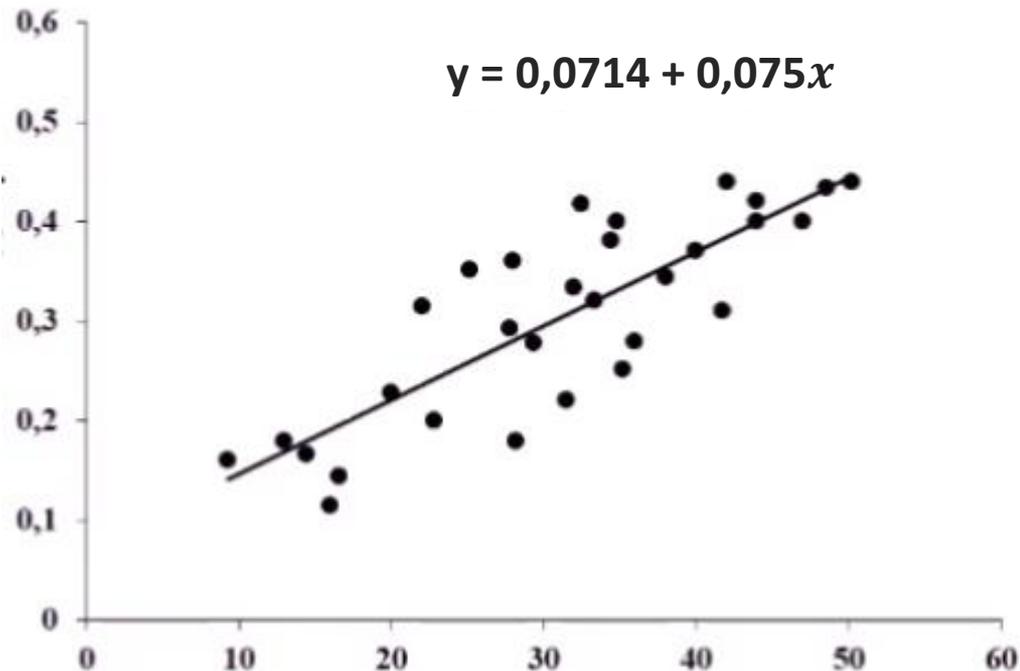
Droite de régression = droite des moindres carrés



Corrélations et régressions

La régression linéaire simple

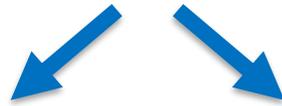
Droite de régression = droite des moindres carrés



*Dans le cadre
de la régression
on s'intéresse à
l'équation de
cette droite*

La régression linéaire simple

- Test de significativité d'une régression



Test de la pente

**Analyse de
variance**

- Ces deux tests sont équivalents mais ils présentent des informations complémentaires

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

variable x	variable y
9,7	24,7
9	21,9
15,3	42,7
4,4	10
11,7	29,7
8,3	16
13,9	37,8
5,7	12,2
12,6	36,1
6,8	19,5
16,2	34,1
11,6	19,2
4,8	17,1
7,3	11,9
13,3	26,3
5,1	23,6
9,2	14,1
11,8	16,3
8,6	31,8
15,2	26,4

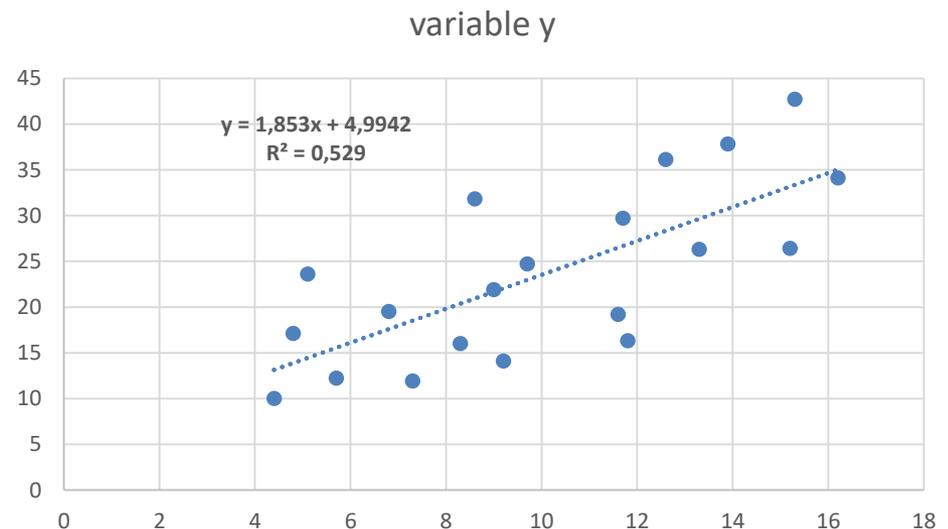
Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 1 : visualiser le nuage de point

- Sélectionner l'ensemble des données
- Insertion – nuage de points
- + courbe de tendance linéaire
- + courbe de tendance ->
autres options ->
afficher l'équation sur graphique
afficher coefficient détermination



Il semble y avoir un lien linéaire

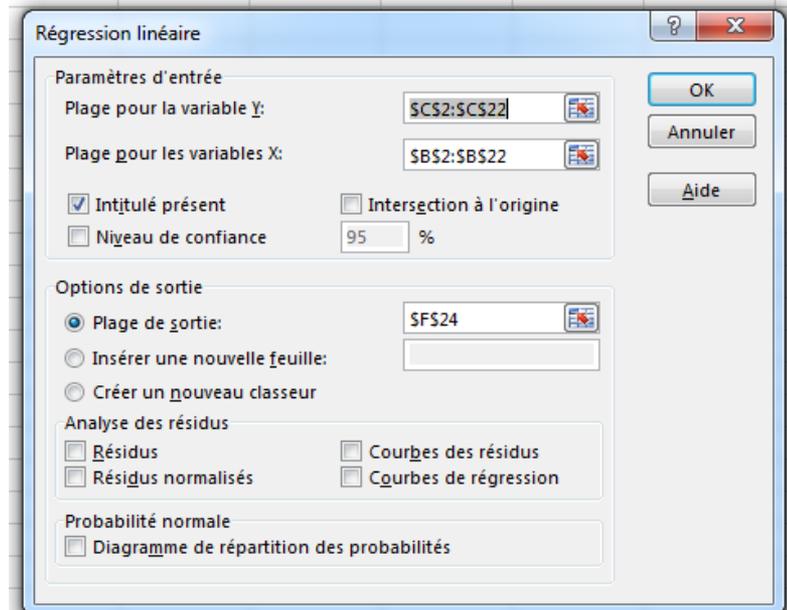
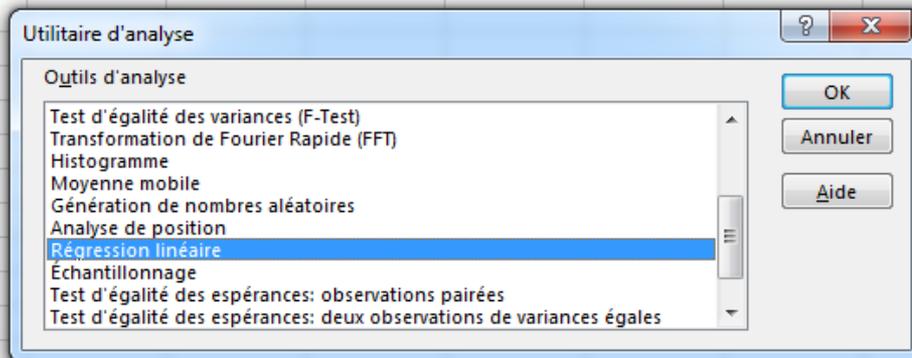
Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 2 : générer l'ensemble des résultats de la régression

- Régression linéaire – entrez les champs nécessaires



Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

- 3 tableaux différents :

RAPPORT DÉTAILLÉ					
<i>Statistiques de la régression</i>					
Coefficient de détermination multiple	0,72735135				
Coefficient de détermination R ²	0,529039987				
Coefficient de détermination R ²	0,502875542				
Erreur-type	6,645998706				
Observations	20				
ANALYSE DE VARIANCE					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	1	893,0946216	893,0946216	20,21980529	0,000279058
Résidus	18	795,0473784	44,1692988		
Total	19	1688,142			
	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>	
Constante	4,994165521	4,390215441	1,137567299	0,270217713	
variable x	1,85295107	0,41207424	4,496643781	0,000279058	

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Statistiques de la régression :

Le **coefficient de détermination R^2**

- C'est la part de la variabilité de y expliquée par x . Autrement dit c'est la force de l'influence de x sur y . Il s'agit en fait du carré du coefficient de corrélation
 - Comment lire ce tableau :

<i>Statistiques de la régression</i>				
Coefficient de détermination multiple	0,72735135	"mal dit", il s'agit en fait du coefficient de corrélation		
Coefficient de détermination R^2	0,529039987	Il s'agit du coefficient de détermination dans les deux cas (le 2ème est la version "ajustée" que l'on choisira préférentiellement)		
Coefficient de détermination R^2	0,502875542			
Erreur-type	6,645998706	C'est la différence moyenne attendue entre le modèle et la réalité		
Observations	20			

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Analyse de variance :

C'est le 1er **test de significativité** de la régression

- Il s'agit d'une ANOVA (on a déjà vu les principes) : ici il s'agit de comparer la variance « régression » due à notre modèle et la variance « résiduelle » due au hasard.
 - Si le rapport entre les deux est au dessus du F critique, on peut rejeter H0 et conclure à la significativité de la régression.
- Comment lire ce tableau :

ANALYSE DE VARIANCE

	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	1	893,0946216	893,0946216	20,21980529	0,000279058
Résidus	18	795,0473784	44,1692988		
Total	19	1688,142			

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Analyse de variance :

C'est le 1er **test de significativité** de la régression

- Il s'agit d'une ANOVA (on a déjà vu les principes) : ici il s'agit de comparer la variance « régression » due à notre modèle et la variance « résiduelle » due au hasard.
- Si le rapport entre les deux est au dessus du F critique, on peut rejeter H0 et conclure à la significativité de la régression.

➤ Comment lire ce tableau :

ANALYSE DE VARIANCE

	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	1	893,0946216	893,0946216	20,21980529	0,000279058
Résidus	18	795,0473784	44,1692988		
Total	19	1688,142			

Vous connaissez le principe des degrés de liberté, il n'y a rien à interpréter ici

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Analyse de variance :

C'est le 1er **test de significativité** de la régression

- Il s'agit d'une ANOVA (on a déjà vu les principes) : ici il s'agit de comparer la variance « régression » due à notre modèle et la variance « résiduelle » due au hasard.
 - Si le rapport entre les deux est au dessus du F critique, on peut rejeter H0 et conclure à la significativité de la régression.
- Comment lire ce tableau :

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	1	893,0946216	893,0946216	20,21980529	0,000279058
Résidus	18	795,0473784	44,1692988		
Total	19	1688,142			

Idem pour les sommes des carrés, il s'agit ici de carrés d'écart particuliers mais pas nécessaire de comprendre

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Analyse de variance :

C'est le 1er **test de significativité** de la régression

- Il s'agit d'une ANOVA (on a déjà vu les principes) : ici il s'agit de comparer la variance « régression » due à notre modèle et la variance « résiduelle » due au hasard.
- Si le rapport entre les deux est au dessus du F critique, on peut rejeter H0 et conclure à la significativité de la régression.

➤ Comment lire ce tableau :

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	1	893,0946216	893,0946216	20,21980529	0,000279058
Résidus	18	795,0473784	44,1692988		
Total	19	1688,142			

Les « moyennes des carrés » sont ce qu'on appelle les variances en ANOVA classique (sommés des carrés divisées par les degrés de liberté correspondants)

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Analyse de variance :

C'est le 1er **test de significativité** de la régression

- Il s'agit d'une ANOVA (on a déjà vu les principes) : ici il s'agit de comparer la variance « régression » due à notre modèle et la variance « résiduelle » due au hasard.
 - Si le rapport entre les deux est au dessus du F critique, on peut rejeter H0 et conclure à la significativité de la régression.
- Comment lire ce tableau :

ANALYSE DE VARIANCE

	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	1	893,0946216	893,0946216	20,21980529	0,000279058
Résidus	18	795,0473784	44,1692988		
Total	19	1688,142			

L'essentiel est ici : la statistique F calculée et son p exact (qu'il suffit de comparer à 0,05 pour conclure sur la significativité). Ici le test est très significatif, on peut rejeter H0

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Estimateurs :

Ce tableau contient :

- Les paramètres de l'équation de la droite de régression
- Le 2nd test de significativité : le test de la pente
- Les intervalles de confiance à 95% (dont nous n'avons pas besoin ici)
 - Comment lire ce tableau :

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>
Constante	4,994165521	4,390215441	1,137567299	0,270217713
variable x	1,85295107	0,41207424	4,496643781	0,000279058

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Estimateurs :

Ce tableau contient :

- Les paramètres de l'équation de la droite de régression
 - Le 2nd test de significativité : le test de la pente
 - Les intervalles de confiance à 95% (dont nous n'avons pas besoin ici)
- Comment lire ce tableau :

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>
Constante	4,994165521	4,390215441	1,137567299	0,270217713
variable x	1,85295107	0,41207424	4,496643781	0,000279058

Constante = ordonnée à l'origine
Variable x = pente de la droite



Il est dès lors possible de prédire la valeur de y pour n'importe quelle valeur de x -> dimension prédictive très utile !

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Estimateurs :

Ce tableau contient :

- Les paramètres de l'équation de la droite de régression
 - Le 2nd test de significativité : le test de la pente
 - Les intervalles de confiance à 95% (dont nous n'avons pas besoin ici)
- Comment lire ce tableau :

	Coefficients	Erreur-type	Statistique t	Probabilité
Constante	4,994165521	4,390215441	1,137567299	0,270217713
variable x	1,85295107	0,41207424	4,496643781	0,000279058

Exemple : pour $x = 15$

$$y = 1,85295101 * 15 + 4,994165521 = 32,79$$

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Estimateurs :

Ce tableau contient :

- Les paramètres de l'équation de la droite de régression
- Le 2nd test de significativité : le test de la pente
- Les intervalles de confiance à 95% (dont nous n'avons pas besoin ici)
 - Comment lire ce tableau :

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>
Constante	4,994165521	4,390215441	1,137567299	0,270217713
variable x	1,85295107	0,41207424	4,496643781	0,000279058

Les erreurs-types décrivent les différences entre la droite de régression trouvée à partir de l'échantillon et la « vraie » droite de régression dans la population totale (pas besoin ici)

Corrélations et régressions

La régression linéaire simple

Mise en œuvre sous excel

Etape 3 : interpréter les résultats de la régression

Estimateurs :

Ce tableau contient :

- Les paramètres de l'équation de la droite de régression
- Le 2nd test de significativité : le test de la pente
- Les intervalles de confiance à 95% (dont nous n'avons pas besoin ici)
 - Comment lire ce tableau :

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>
Constante	4,994165521	4,390215441	1,137567299	0,270217713
variable x	1,85295107	0,41207424	4,496643781	0,000279058

La statistique t et son p associé constituent le test de la pente -> il est forcément conforme à l'analyse de la variance vue dans le tableau précédent (donc ici significatif, pas besoin de tenir compte de la constante ou ordonnée à l'origine)

Corrélations et régressions

Exercice d'application :

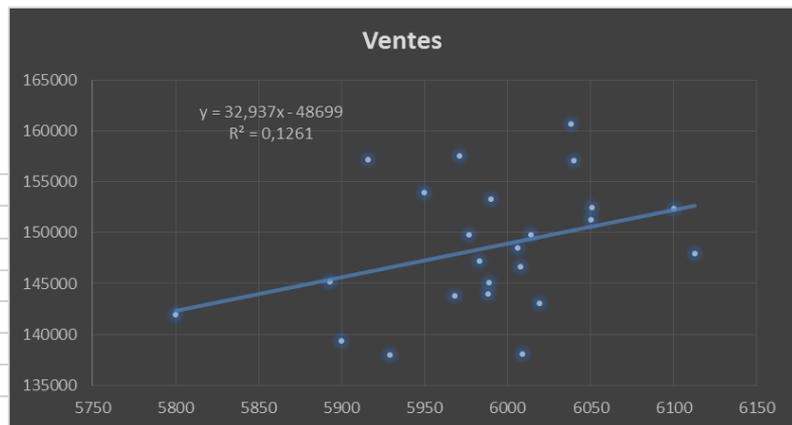
Y a-t-il un lien entre les dépenses publicitaires et les ventes mensuelles ?

Mois	Dépenses publicité	Ventes
janv_2016	6113	147918
fevr_2016	6019	143088
mars_2016	5916	157174
avri_2016	6038	160728
mai_2016	5971	157542
juin_2016	5988	143972
juil_2016	6014	149770
aout_2016	6008	146690
sept_2016	5990	153318
octo_2016	5968	143760
nove_2016	5983	147196
dece_2016	5977	149786
janv_2017	5929	138006
fevr_2017	6040	157084
mars_2017	5989	145086
avri_2017	6006	148494
mai_2017	6051	152506
juin_2017	5893	145164
juil_2017	6009	138044
aout_2017	6100	152400
sept_2017	5900	139402
octo_2017	5950	154010
nove_2017	5800	141974
dece_2017	6050	151286

Corrélations et régressions

Exercice d'application :

Y a-t-il un lien entre les dépenses publicitaires et les ventes mensuelles ?



RAPPORT DÉTAILLÉ	
<i>Statistiques de la régression</i>	
Coefficient de déter	0,35504995
Coefficient de déter	0,12606046
Coefficient de déter	0,08633594
Erreur-type	6049,4223
Observations	24

ANALYSE DE VARIANCE					
	Degré de liberté	mm des car	enne des ca	F	leur critique de F
Régression	1	116130956	116130956	3,17336623	0,08866006
Résidus	22	805101224	36595510,2		
Total	23	921232180			

	Coefficients	Erreur-type	Statistique t	Probabilité
Constante	-48698,7546	110715,29	-0,43985573	0,66433242
Dépenses publicité	32,9373851	18,489664	1,78139446	0,08866006

Le lien entre les dépenses publicitaires et les ventes mensuelles n'est pas significatif ($p=0,09$). A partir de là pas la peine de commenter les autres résultats