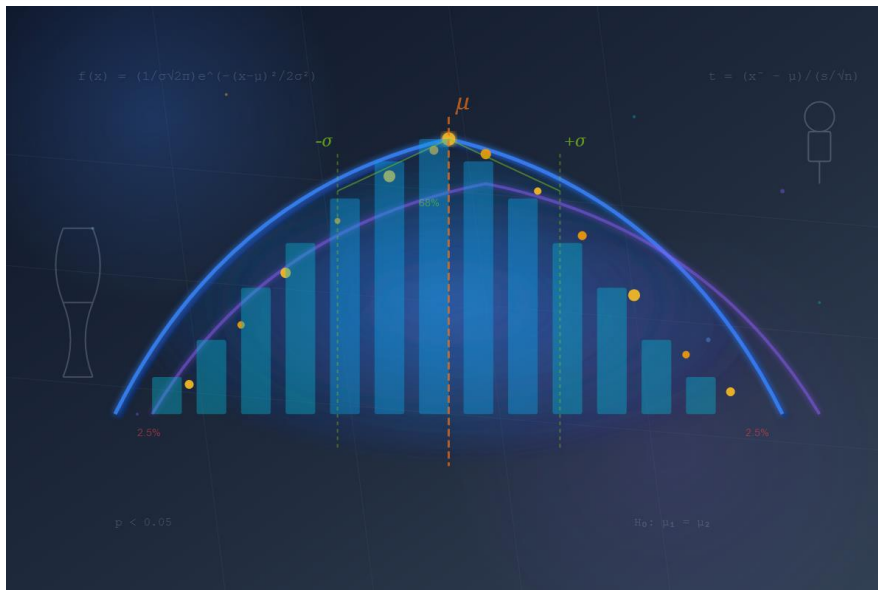


PASS : UE8



Support de Cours

Introduction aux Biostatistiques

Partie 2

Plan

2/99

Je me permets d'attirer votre attention
sur le fait que ce diaporama ne
constitue en aucun cas un polycopié du
cours dispensé en direct et ne saurait,
en aucune manière, s'y substituer



Plan

3/99

- 👉 Concepts de Base en Statistiques
- 👉 Description des Données
- 👉 La Distribution Normale
- 👉 Les Hypothèses
- 👉 Le Test d'Hypothèse (Appliqué aux Moyennes)

Plan

4/99

👉 Concepts de Base en Statistiques

👉 Description des Données

👉 La Distribution Normale

👉 Les Hypothèses

👉 Le Test d'Hypothèse (Appliqué aux Moyennes)

Les Termes Importants

5/99

- **Échantillon**

Pour étudier un phénomène, comme la dénomination, il est impossible de l'analyser chez tous les sujets d'intérêt (ex : tous les patients atteints d'Alzheimer). Cela nécessiterait trop de ressources. On prélève donc un **échantillon** d'observations de la **population parente**.

- **Échantillon aléatoire**

L'échantillon doit être tiré au hasard pour être représentatif. Ainsi, chaque patient de la population a une chance égale d'être inclus. Cet échantillon est dit **aléatoire**.

- **Population parente**

C'est l'ensemble des événements ou sujets d'intérêt (ex : scores de dénomination de tous les patients atteints d'Alzheimer).

Les Termes Importants

6/99

- **Validité externe**

Il n'est pas toujours possible d'obtenir un échantillon aléatoire. Par conséquent, les estimations de la population peuvent ne pas représenter fidèlement l'ensemble de la population, ce qui entraîne un manque de **validité externe**.

Exemple : Étudier les processus de lecture chez 5 patients ayant une lésion visuelle peut ne pas être généralisable, ce qui réduit la validité externe.

- **Variables discrètes versus continues**

- Les **variables discrètes** prennent un nombre limité de valeurs, comme le sexe (homme ou femme).
- Les **variables continues** peuvent prendre une infinité de valeurs dans une échelle définie, comme l'âge.

Les Termes Importants

7/99

- **Données de mesure vs données quantitatives**

La distinction entre variables continues et discrètes est liée à celle entre **données de mesure** (quantitatives) et **données catégorielles** (qualitatives).

- **Données de mesure** : Se réfèrent aux résultats de mesures, comme un score à un test ou le poids des personnes.
- **Données catégorielles** : Concernent les classifications, par exemple : « Dans notre étude, 10 sujets ont des performances normales et 12 des performances déficitaires ». On regroupe les sujets par catégories.

Les Termes Importants

8/99

- **Variable indépendante (VID) vs variable dépendante (VD) – IMPORTANT**
 - Les **variables indépendantes (VID)** sont les variables **manipulées** par l'expérimentateur.
 - Les **variables dépendantes (VD)** représentent les **données mesurées**.

Exemple : Comparer les scores de dénomination entre des patients ayant une lésion à l'hémisphère gauche et des participants neurotypiques.

- **Question :** Quelle est la variable dépendante ? Quelle est la variable indépendante ?

Remarque : Les VID peuvent être quantitatives ou qualitatives, tandis que les VD sont généralement quantitatives.

- VID = groupe (lésion gauche vs neurotypique)
- VD = score de dénomination

La variable indépendante est la cause testée, la variable dépendante est l'effet observé.

Les Termes Importants

9/99

STATISTICA - [Data: Spreadsheet2* (3v by 20c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

	1 Identification Sujets	2 Groupe	3 Score de lecture (/20)
1	SN1	Sujets normaux	20
2	SN2	Sujets normaux	19
3	SN3	Sujets normaux	18
4	SN4	Sujets normaux	16
5	SN5	Sujets normaux	18
6	SN6	Sujets normaux	19
7	SN7	Sujets normaux	20
8	SN8	Sujets normaux	14
9	SN9	Sujets normaux	20
10	SN10	Sujets normaux	17
11	P1	Patients	18
12	P2	Patients	13
13	P3	Patients	12
14	P4	Patients	15
15	P5	Patients	9
16	P6	Patients	12
17	P7	Patients	5
18	P8	Patients	8
19	P9	Patients	12
20	P10	Patients	12

For Help, press F1

Spreadsheet2 C11.V1 P1 Sel.OFF Weight.OFF CAP NUM REC

Quelle est la variable dépendante ? Quelle est la variable indépendante ?

Les Termes Importants

10/99

STATISTICA - [Data: Spreadsheet2* (5v by 20c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

	1 Identification Sujets	2 Groupe	3 Sexe	4 Age	5 Score de lecture (/20)
1	SN1	Sujets normaux	F	25	20
2	SN2	Sujets normaux	M	32	19
3	SN3	Sujets normaux	M	45	18
4	SN4	Sujets normaux	M	36	16
5	SN5	Sujets normaux	F	18	18
6	SN6	Sujets normaux	F	65	19
7	SN7	Sujets normaux	F	53	20
8	SN8	Sujets normaux	M	30	14
9	SN9	Sujets normaux	F	35	20
10	SN10	Sujets normaux	M	39	17
11	P1	Patients	M	40	18
12	P2	Patients	M	36	13
13	P3	Patients	M	29	12
14	P4	Patients	F	50	15
15	P5	Patients	F	66	9
16	P6	Patients	M	22	12
17	P7	Patients	M	35	5
18	P8	Patients	F	45	8
19	P9	Patients	M	25	12
20	P10	Patients	f	30	12

For Help, press F1

Sel.OFF Weight.OFF CAP NUM REC

Quelle est la variable dépendante ? Quelles sont les variables indépendantes ?

Les Termes Importants

11/99

STATISTICA - [Data: Spreadsheet2* (7v by 20c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

	1 Identification Sujets	2 Groupe	3 Sexe	4 Age	5 Score de lecture (/20) Avant l'opération	6 Score de lecture (/20) Juste après opération	7 Score de lecture (/20) Trois mois après l'opération													
1	SN1	Sujets normaux	F	25	20	18	20													
2	SN2	Sujets normaux	M	32	19	20	20													
3	SN3	Sujets normaux	M	45	18	18	18													
4	SN4	Sujets normaux	M	36	16	17	17													
5	SN5	Sujets normaux	F	18	18	18	18													
6	SN6	Sujets normaux	F	65	19	19	20													
7	SN7	Sujets normaux	F	53	20	19	20													
8	SN8	Sujets normaux	M	30	14	15	17													
9	SN9	Sujets normaux	F	35	20	19	20													
10	SN10	Sujets normaux	M	39	17	18	18													
11	P1	Patients	M	40	18	12	14													
12	P2	Patients	M	36	18	10	16													
13	P3	Patients	M	29	17	10	16													
14	P4	Patients	F	50	15	15	17													
15	P5	Patients	F	66	20	15	16													
16	P6	Patients	M	22	12	12	12													
17	P7	Patients	M	35	10	5	11													
18	P8	Patients	F	45	15	7	8													
19	P9	Patients	M	25	12	9	12													
20	P10	Patients	F	30	12	8	10													

For Help, press F1

Spreadsheet2 C8.V7 17 Sel:OFF Weight:OFF CAP NUM REC

Quelle est la variable dépendante ? Quelles sont les variables indépendantes ?

Statistiques Descriptives

12/99

⇒ 2 concepts importants

- **Statistique descriptive**

Permet de décrire les données pour une vision globale avant toute analyse plus complexe. C'est une **analyse exploratoire**.

- Créer des histogrammes pour représenter les données
- Calculer la moyenne, la médiane et le mode
- Mesurer la dispersion avec des écarts types

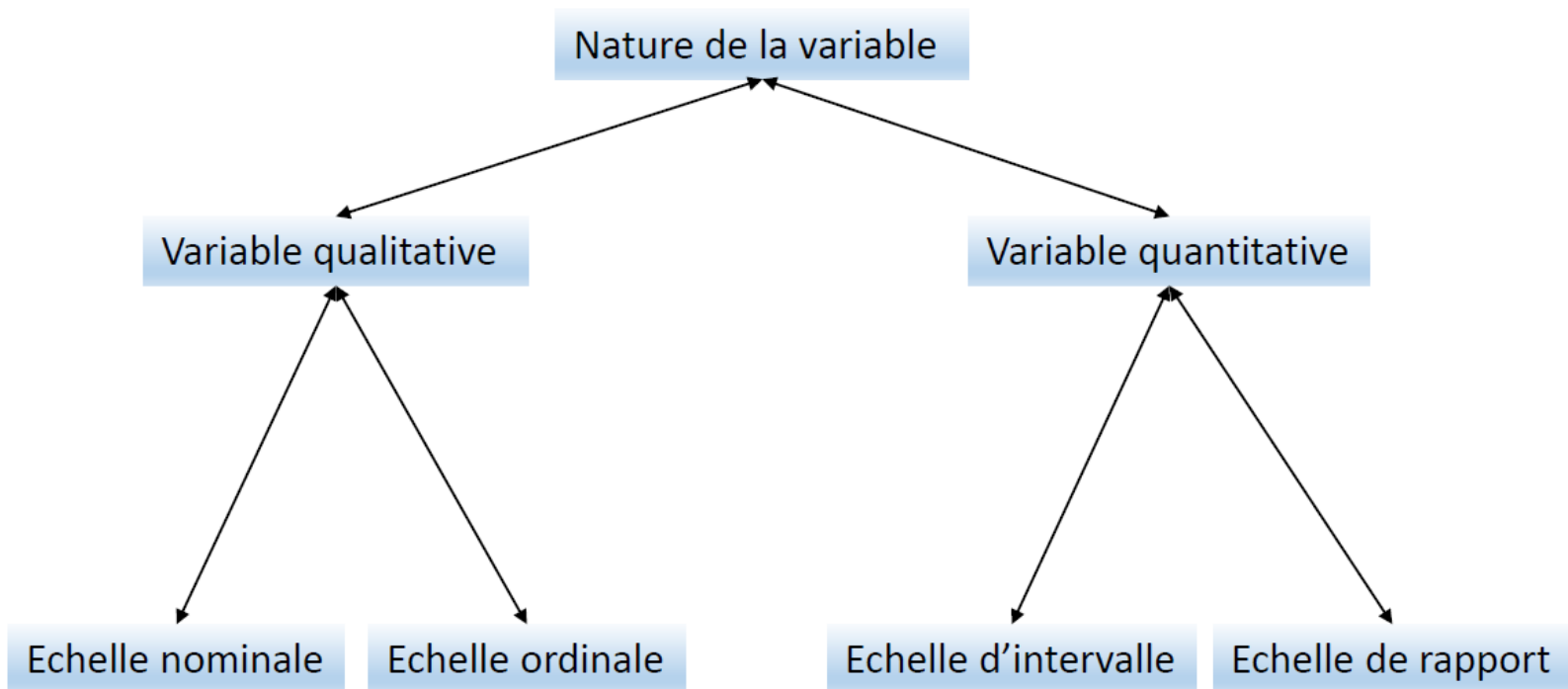
- **Statistique inférentielle**

Regroupe les méthodes pour vérifier des hypothèses et tirer des conclusions. Exemple : Tester si les patients avec une maladie neurodégénérative ont plus de difficultés de dénomination que les sujets contrôles.

Les Echelles de Mesure

13/99

⇒ Il existe 4 échelles de mesure différentes



Les Echelles de Mesure

14/99

⇒ Il existe 4 échelles de mesure différentes

- **Échelle nominale**

Les échelles nominales ne présentent pas de hiérarchie. Elles classent les données en catégories sans ordre spécifique. Par exemple, le sexe ou la latéralité manuelle sont des variables nominales. Ce sont des étiquettes.

- **Échelle ordinale**

Les échelles ordinales classent les données selon un ordre ou un continuum prédéfini, sans indiquer les écarts entre les niveaux. Par exemple, les grades militaires : sous-officier < officier < lieutenant < colonel.

Exemple : Lors d'un questionnaire, les sujets évaluent leur accord avec des affirmations selon cette échelle :

Fort désaccord — Désaccord — Accord — Parfaitement d'accord

Les Echelles de Mesure

15/99

⇒ Il existe 4 échelles de mesure différentes

- Échelle d'intervalles

Les échelles d'intervalles permettent de mesurer les différences entre les points de l'échelle, où chaque unité a la même signification partout. L'échelle de Celsius en est un exemple : une différence de 5° est identique quel que soit l'endroit sur l'échelle.

Cependant, on ne peut pas faire de rapports (dire que 10° est deux fois plus chaud que 5°) car le zéro est arbitraire.

Exemple :

La pointure = (taille du pied en cm + 1) $\times$ 1,5

Si la taille du pied est de 25 cm, la pointure sera de 39.

Toutefois, on ne peut pas dire qu'une personne avec une pointure de 42 (27 cm) a un pied deux fois plus grand qu'une personne avec une pointure de 21 (13 cm).

Les Echelles de Mesure

16/99

⇒ Il existe 4 échelles de mesure différentes

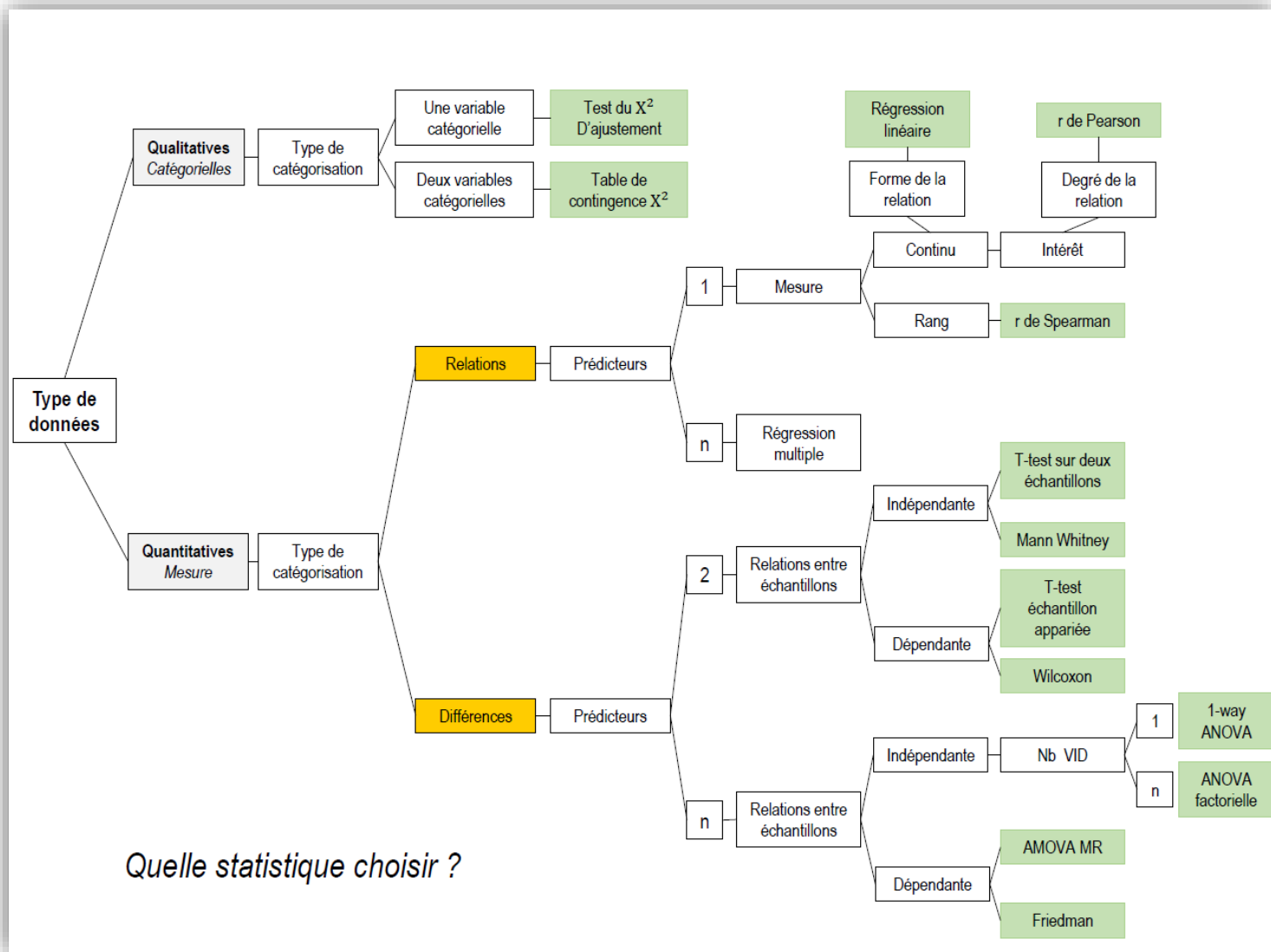
- Échelle de rapport

Les échelles de rapport sont des échelles où le point 0 n'est pas arbitraire (0 signifie l'absence de la chose mesurée).

Exemples classiques : Ces échelles s'appliquent aux grandeurs physiques comme la longueur, le poids ou les temps de réaction.

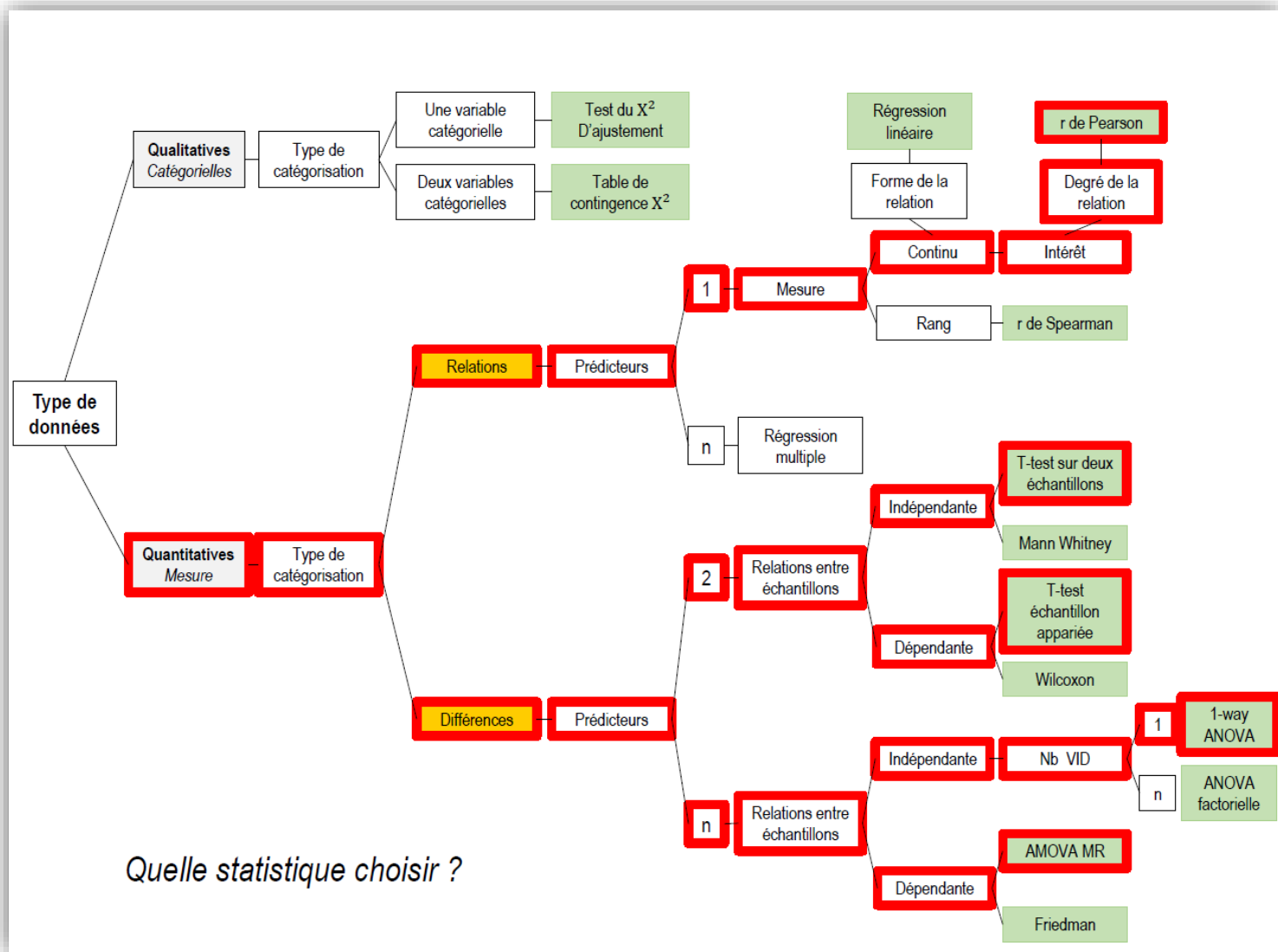
Quelles Statistiques Utiliser?

17/99



Quelles Statistiques Utiliser?

18/99



Plan

19/99

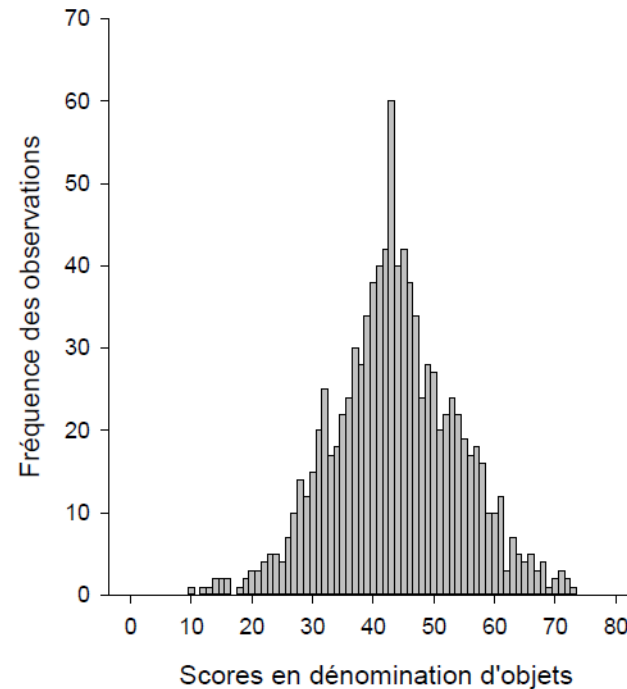
- 👉 Concepts de Base en Statistiques
- 👉 Description des Données
- 👉 La Distribution Normale
- 👉 Les Hypothèses
- 👉 Le Test d'Hypothèse (Appliqué aux Moyennes)

Représentation Graphique

20/99

DO Score	Fréquence	DO Score	Fréquence
1	0	41	40
2	0	42	42
3	0	43	60
4	0	44	40
5	0	45	42
6	0	46	38
7	0	47	34
8	0	48	24
9	0	49	28
10	1	50	27
11	0	51	20
12	1	52	22
13	1	53	24
14	2	54	22
15	2	55	19
16	2	56	17
17	0	57	18
18	1	58	16
19	2	59	10
20	3	60	10
21	3	61	12
22	4	62	3
23	5	63	7
24	5	64	5
25	4	65	4
26	7	66	5
27	10	67	3
28	14	68	4
29	12	69	1
30	15	70	2
31	20	71	3
32	25	72	2
33	17	73	1
34	18	74	0
35	22	75	0
36	24	76	0
37	30	77	0
38	28	78	0
39	34	79	0
40	38	80	0

Graphique des scores en dénomination par rapport à la fréquence



Performances en dénomination de 955 sujets normaux

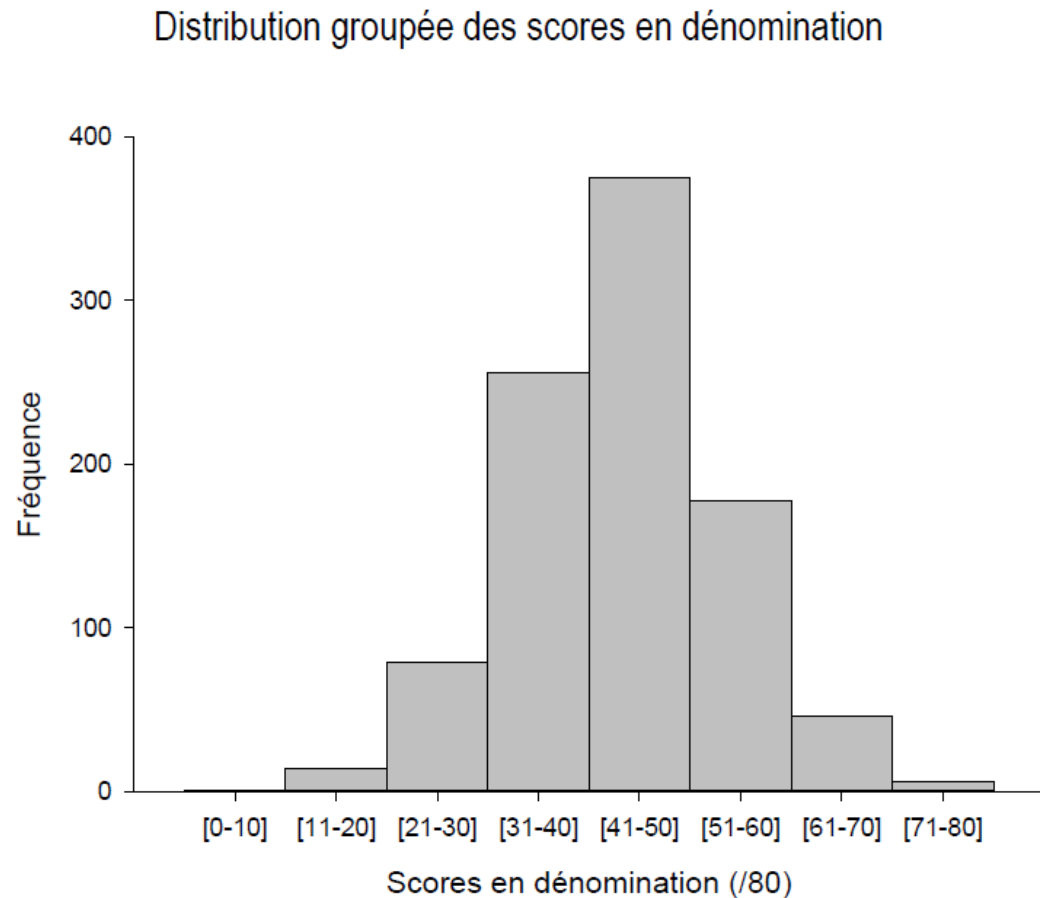
Histogramme de fréquence des observations

Représentation Graphique

21/99

Intervalle	Fréquence
[0-10]	1
[11-20]	14
[21-30]	79
[31-40]	256
[41-50]	375
[51-60]	178
[61-70]	46
[71-80]	6

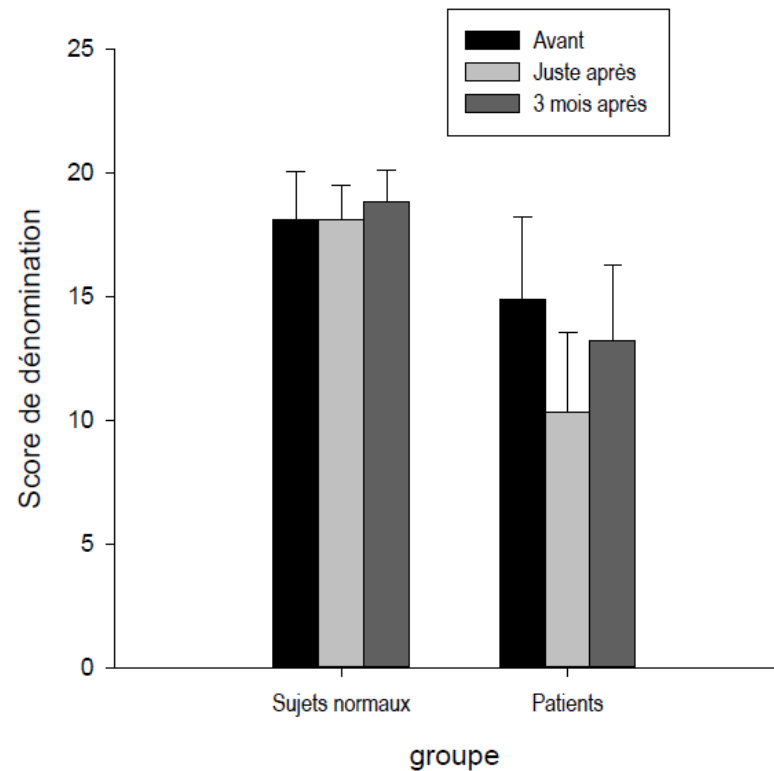
Données triées en intervalles
de score
(même données que la slide
précédente)



Histogramme groupé

Représentation Graphique

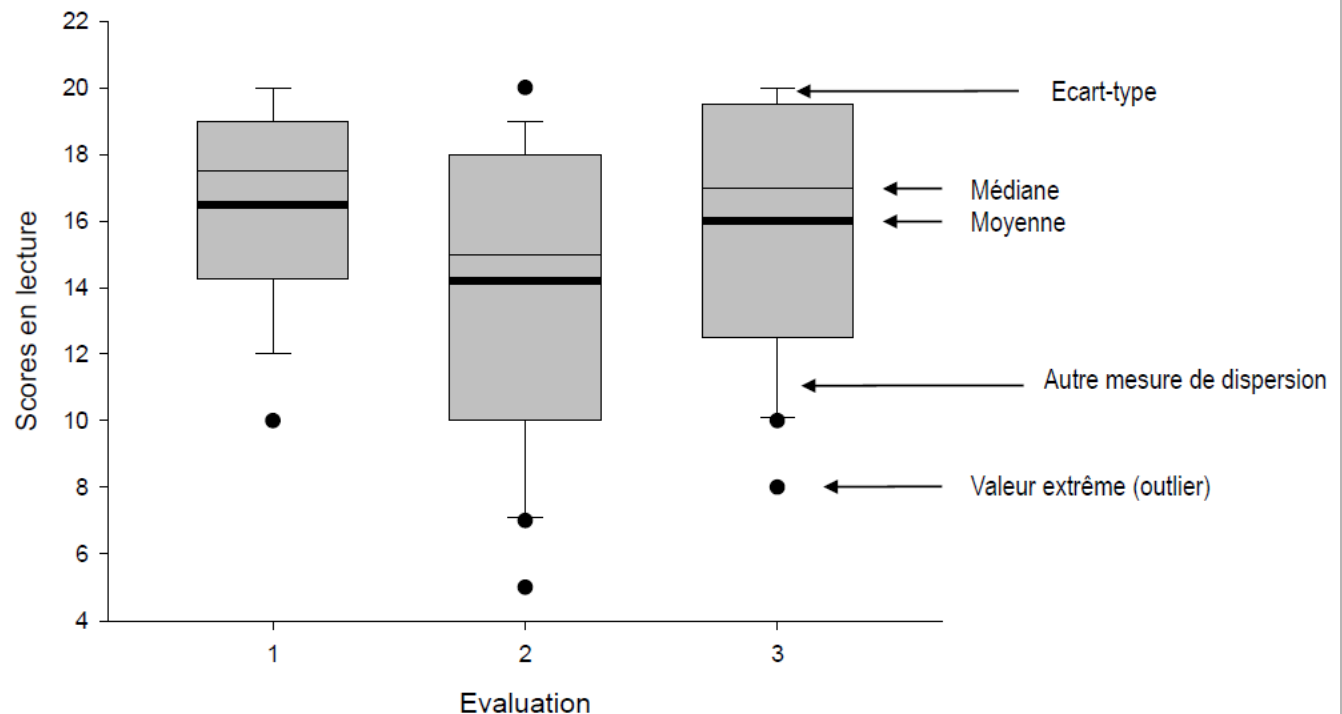
22/99



Moyennes et écarts-type en fonction du groupe et de l'évaluation

Représentation Graphique

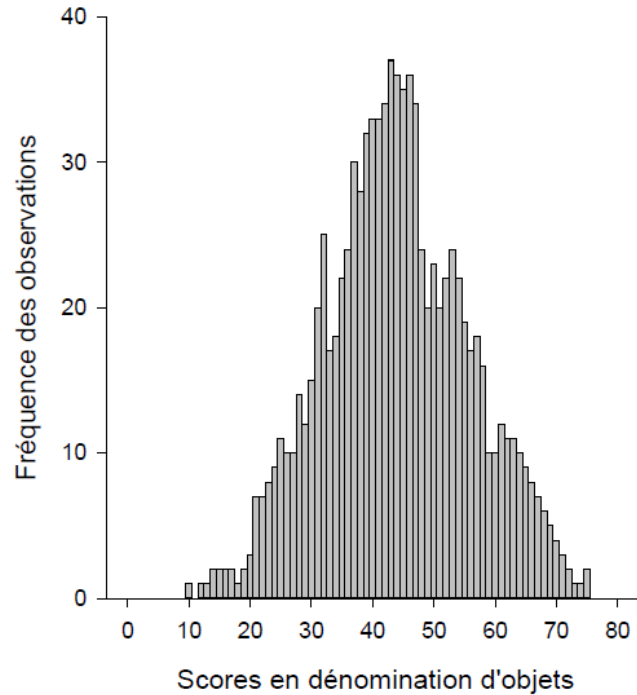
23/99



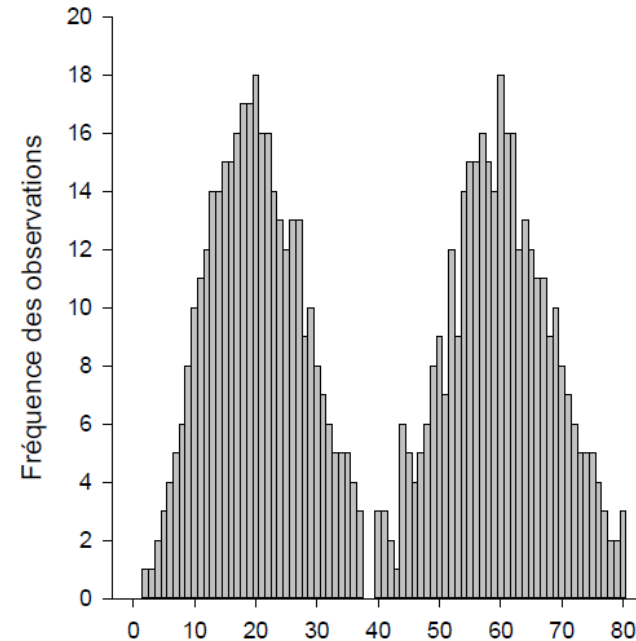
Boite à moustache (ou boxplot) des scores en fonction de l'évaluation

Caractérisation des Données

24/99



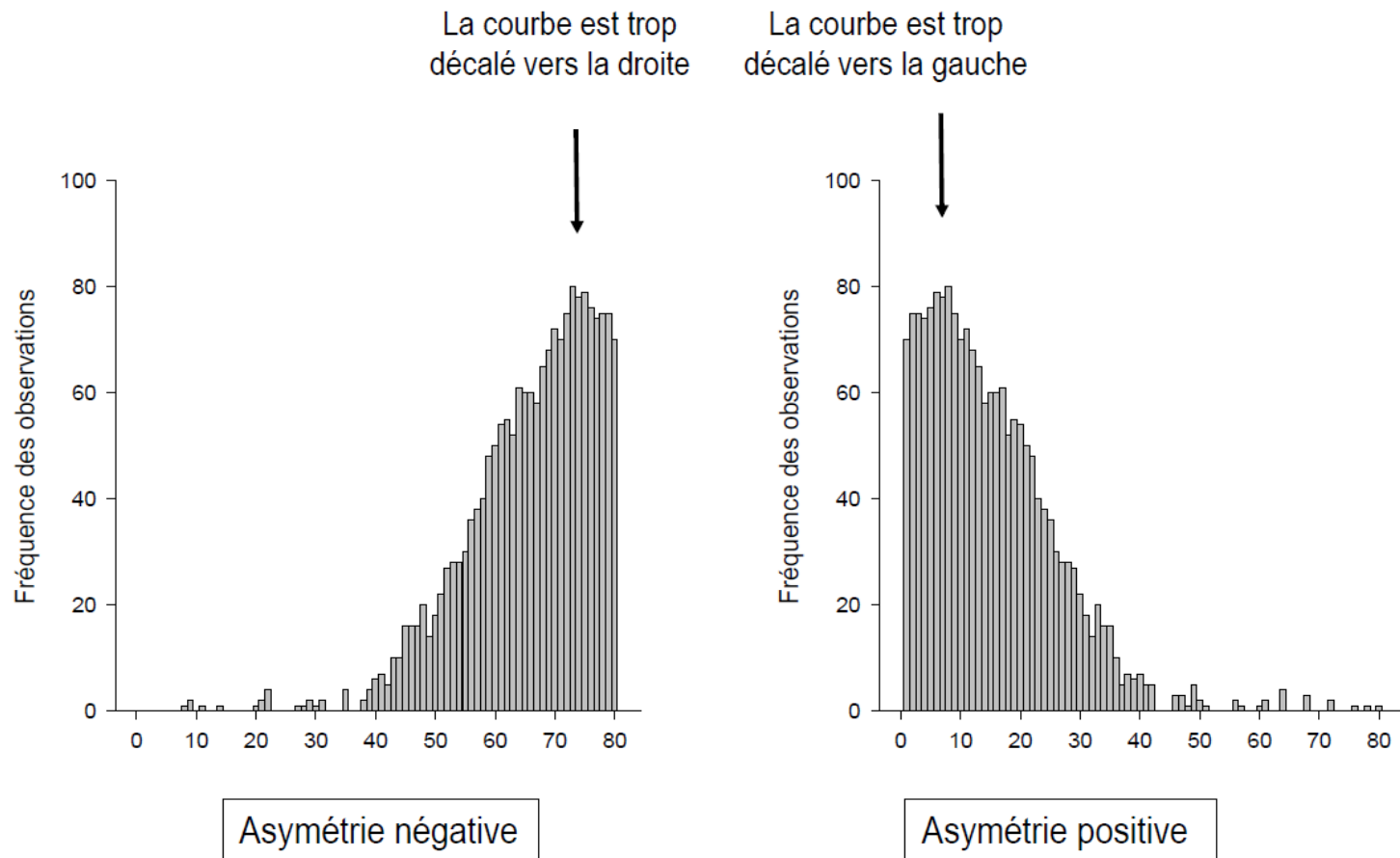
Courbe normale
Ou mésocurtique



Distribution bimodale

Caractérisation des Données

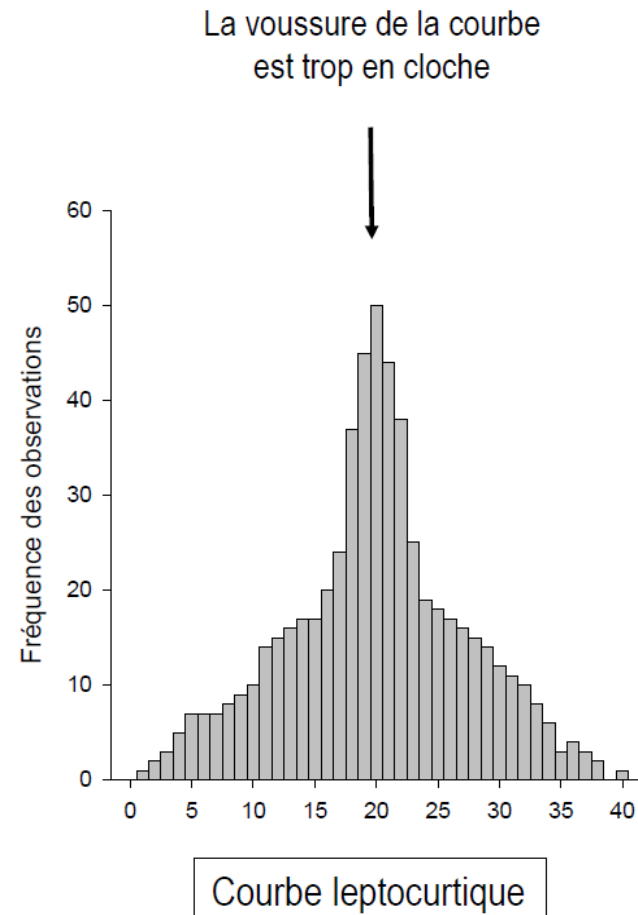
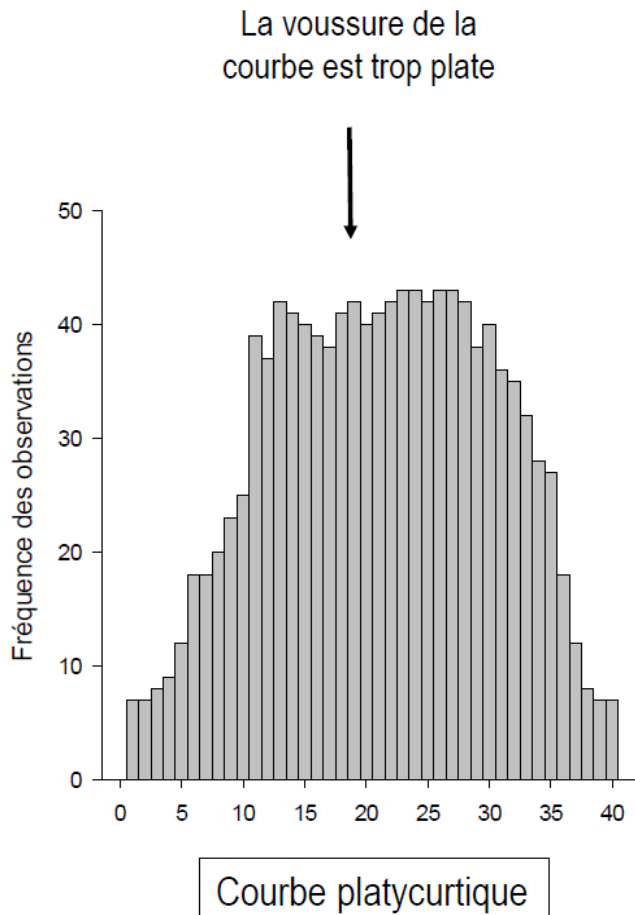
25/99



Remarque : il existe un indice pour mesurer l'asymétrie de la courbe : le Skewness

Caractérisation des Données

26/99



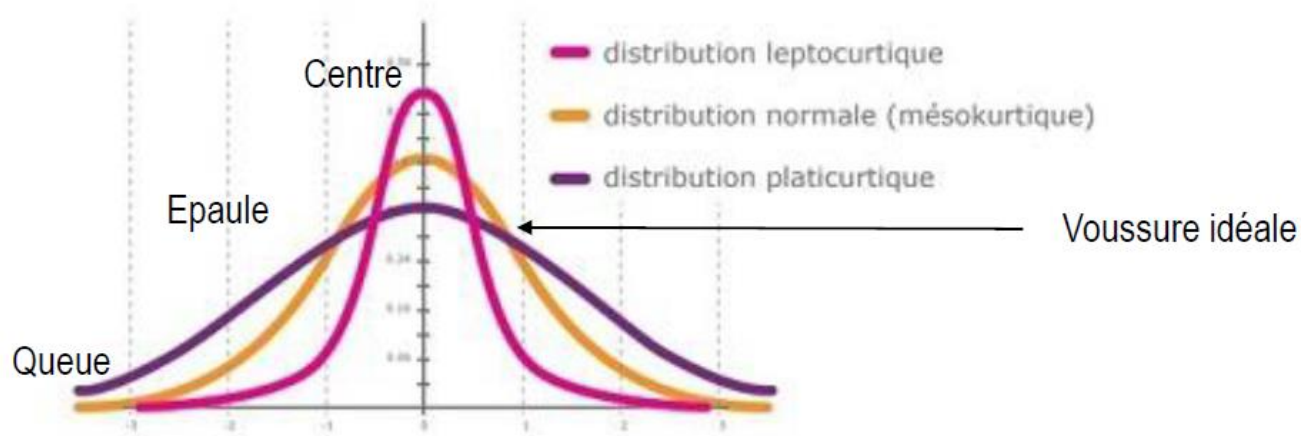
Remarque : il existe un indice pour mesurer l'aplatissement de la courbe : le Kurtosis

Caractérisation des Données

27/99

❖ Voissure d'une distribution

La voissure renvoie d'un point de vue mathématique à la concentration relative des scores au centre, dans les extrémités supérieure et inférieure (les queues) et au niveau des épaules entre les centres et les queues.



Notations Importantes

28/99

- **Les notations de variables**

- Soit une variable (par exemple, le score de lecture de 6 patients), notée X ou Y . Si les scores sont $[18, 20, 15, 18, 19, 14]$, X représente l'ensemble de ces valeurs.
- X_1 fait référence à 18, X_3 à 15, etc.
- X_i peut prendre n'importe quelle valeur entre 14 et 20, avec $i = 1$ à $i = 6$.

- **Notation de la somme (Sigma)**

- $\sum X_i$: « additionner les X_i »
- $\sum_{i=1}^N X_i$: « additionner tous les X_i de $i = 1$ à $i = N$ »

Notations Importantes

29/99

Score de lecture (10)	Score de compréhension (5)				
(X)	(Y)	(X ²)	(Y ²)	X-Y	XY
8	3	64	9	5	24
6	5	36	25	1	30
10	2	100	4	8	20
9	4	81	16	5	36
5	4	25	16	1	20

$$\sum X_i =$$

$$\sum Y_i =$$

$$\sum X_i^2 =$$

$$\sum Y_i^2 =$$

$$\sum (X_i - Y_i)$$

$$\sum X_i Y_i =$$

$$(\sum X_i)^2 =$$



Indices de Tendance Centrale

30/99

❖ Double indice

	Score en lecture (10)		Total
	Evaluation 1	Evaluation 2	
1	8	7	15
2	5	5	10
3	10	9	19
4	5	6	11
5	7	8	15
Total	35	35	

Dans ce type de tableau, i va représenter la ligne et j la colonne. Si vous voulez faire référence à une case spécifique de tableau, vous allez utiliser la notation X_{ij}

i allant de 1 à 5 et j allant de 1 à 2.

$X_{5,2} =$ 

La somme de toutes les données du tableau va s'écrire :

$$\sum_{i=1}^5 \sum_{j=1}^2 X_{ij}$$

Indices de Tendance Centrale

31/99

- **Définition formelle**

La mesure de tendance centrale (ou mesure de position) fait référence aux mesures indiquant le centre d'une distribution. Elle montre où la distribution se concentre sur l'échelle.

Ces mesures dépendent de l'échelle utilisée et peuvent être sensibles aux valeurs extrêmes (outliers).

Les trois indices principaux de tendance centrale sont : le **mode**, la **médiane** et la **moyenne**.

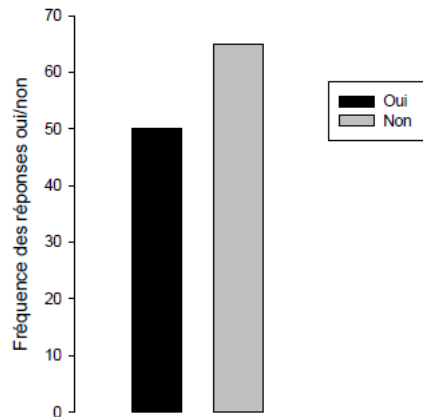
- **Le mode**

Le **mode** est la valeur la plus fréquente dans une distribution, c'est-à-dire celle qui revient le plus souvent. Par exemple, le mode est la valeur de X ayant le plus grand nombre d'occurrences.

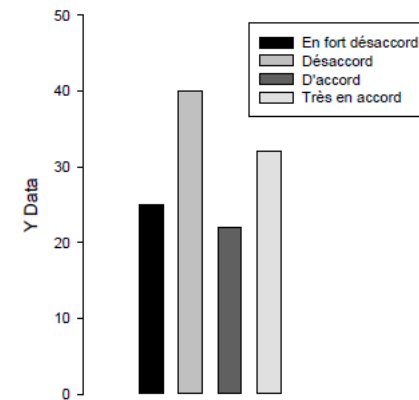
Le mode peut être utilisé avec n'importe quelle échelle et est l'unique indice pertinent pour les échelles catégorielles.

Indices de Tendance Centrale

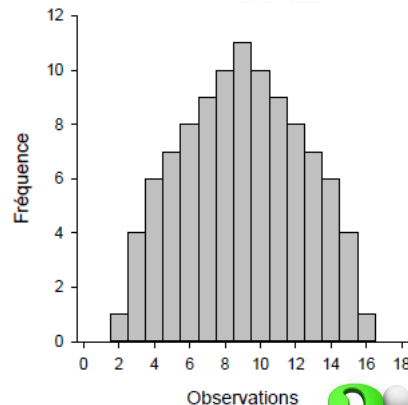
32/99



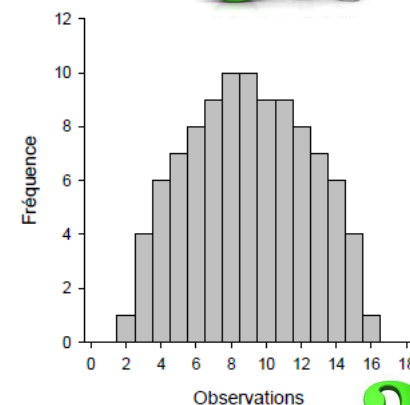
Le mode est ? (65)



Le mode est ? (40)



Le mode est ?



Le mode est ?

Indices de Tendance Centrale

33/99

❖ La médiane

D'un point de vue formel, c'est le score qui correspond au point *central* lorsque les données sont disposées en ordre numérique.

La position médiane dans une distribution d'écrit donc :

Patient	N°	Score (10)
Patient 1	1	0
Patient 2	2	2
Patient 3	3	2
Patient 4	4	2
Patient 5	5	3
Patient 6	6	3
Patient 7	7	4
Patient 8	8	4
Patient 9	9	5
Patient 10	10	5
Patient 11	11	6
Patient 12	12	7
Patient 13	13	7
Patient 14	14	8
Patient 15	15	8
Patient 16	16	8
Patient 17	17	9
Patient 18	18	9
Patient 19	19	10



$$\frac{N + 1}{2}$$

$$\text{Position Médiane} = (19+1)/2 = 10$$

$$\text{Médiane} = 5$$

Indices de Tendance Centrale

34/99

❖ La médiane

D'un point de vue formel, c'est le score qui correspond au point auquel sous lequel des scores sont inclus lorsque les données sont disposées en ordre numérique. La position médiane dans une distribution d'écrit donc :



Patient	N°	Score (10)
Patient 1	1	0
Patient 2	2	2
Patient 3	3	2
Patient 4	4	2
Patient 5	5	3
Patient 6	6	3
Patient 7	7	4
Patient 8	8	4
Patient 9	9	5
Patient 10	10	5
Patient 11	11	6
Patient 12	12	7
Patient 13	13	7
Patient 14	14	8
Patient 15	15	8
Patient 16	16	8
Patient 17	17	9
Patient 18	18	9
Patient 19	19	10
Patient 20	20	10

$$\frac{N + 1}{2}$$

$$\text{Position Médiane} = (20+1)/2 = 10,5$$

$$\text{Médiane} = 5,5$$

Indices de Tendance Centrale

35/99

- La moyenne

La moyenne, un indice de tendance centrale très utilisé, est calculée en divisant la somme des scores par le nombre de scores.

La notation est $\bar{X}$.

$$\bar{X} = \frac{\sum X}{N}$$

Exemple : Pour l'ensemble de valeurs [12, 6, 5, 4, 2], la moyenne est égale à :



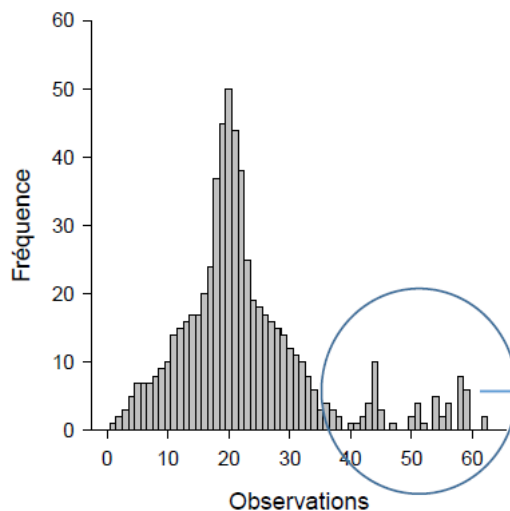
$$\frac{12+6+5+4+2}{5} = 5.8$$

Indices de Tendance Centrale

36/99

❖ Quelques remarques

- R1: La moyenne et la médiane ne seront égales que si la distribution est symétrique, et les trois indices de tendance que si la distribution est symétrique et unimodale.
- R2: La médiane n'est pas influencé par les scores extrêmes contrairement à la moyenne.



Données extrêmes qui vont influencer la moyenne (la décaler vers la droite)

Indices de Tendance Centrale

37/99

Patient	temps de réaction (ms)
Patient 1	540
Patient 2	623
Patient 3	750
Patient 4	422
Patient 5	453
Patient 6	472
Patient 7	650
Patient 8	524
Patient 9	583
Patient 10	701
Patient 11	423
Patient 12	499
Patient 13	523
Patient 14	522
Patient 15	605
Patient 16	702
Patient 17	621
Patient 18	530
Patient 19	589

Patient	temps de réaction (ms)
Patient 1	540
Patient 2	623
Patient 3	750
Patient 4	422
Patient 5	453
Patient 6	472
Patient 7	650
Patient 8	524
Patient 9	583
Patient 10	701
Patient 11	423
Patient 12	499
Patient 13	523
Patient 14	522
Patient 15	605
Patient 16	702
Patient 17	621
Patient 18	530
Patient 19	3523

Médiane =



Moyenne =

Médiane =



Moyenne =

Indices de Tendence Centrale

38/99

Si Bill Gates entre dans un bar, est-ce que tout le monde devient millionnaire ?

Données:

- 100 personnes dans le bar
- Richesse moyenne ordinaire par personne: 30 000 €
- Fortune de Bill Gates (2025): environ 140 milliards d'euros

Conclusion

La moyenne de richesse par personne dépasse 1 million d'euros. Cela signifie-t-il que tout le monde est millionnaire ? Non !

Ce calcul illustre le **piège de la moyenne** : un seul ultra-riche peut fausser complètement la vision statistique d'un groupe.

Pour décrire une réalité plus fidèle, il faut privilégier la **médiane**, qui ne serait ici qu'à 30 000 €, reflétant la richesse réelle du groupe sauf pour l'ultra-riche

Hypothèse

Calculons la richesse moyenne après l'arrivée de Bill Gates:

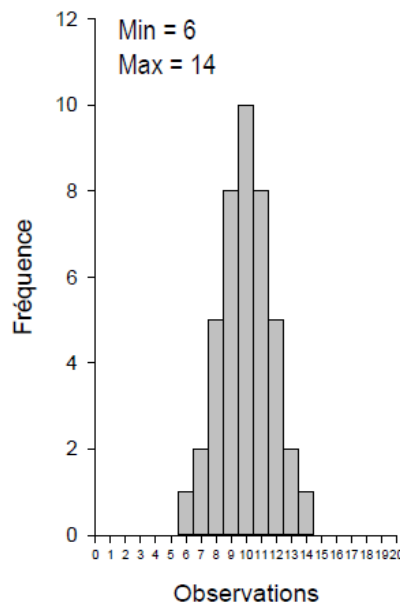
$$\text{Moyenne} = \frac{(100 \times 30\,000) + 140\,000\,000\,000}{101} \approx 1,386\,138\,613\text{€}$$

Mesures de Variabilité

39/99

❖ Dispersion

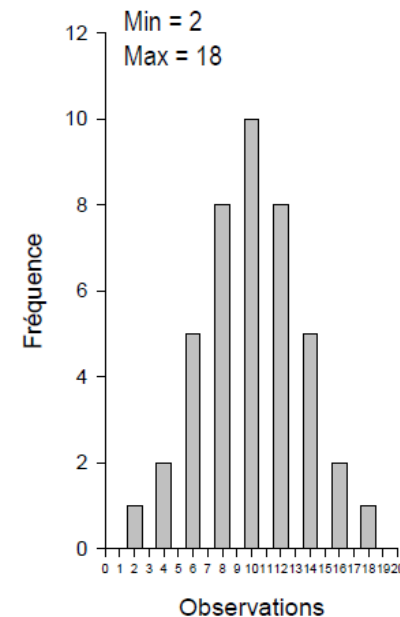
Les indices de tendance centrale permettent de caractériser les données, de les décrire, mais ils ne sont pas suffisamment en eux-mêmes, pour appréhender, **nous renseigner de façon complète**, la distribution des données.



Mode = 10

Moyenne = 10

Médiane = 10



Pourtant la variabilité autour des indices de tendance est bien plus grande sur la courbe de droite
(deux fois plus grandes)

LA DISPERSION N'EST PAS LA MÊME

Mesures de Variabilité

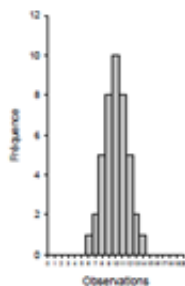
40/99

❖ Dispersion

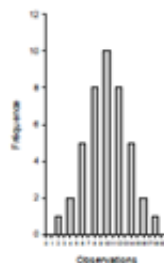
Plusieurs indices de dispersion/variabilité peuvent être utilisées, comme :

- L'étendue
- L'espace interquartile
- L'écart-moyen
- L'écart absolu moyen
- La variance
- L'écart-type

Par exemple: l'étendue est la distance entre la valeur la moins élevée et la valeur la plus élevée de la distribution.



Etendue = $14 - 6 = 8$



Etendue = $18 - 2 = 16$

Mesures de Variabilité

41/99

❖ Ecart moyen (écart à la moyenne)

Pour mesurer le degré de dispersion des scores autour de la moyenne (c'est-à-dire leur écart à la moyenne), on pourrait calculer tous les écarts à la moyenne et en faire la moyenne. Mais la moyenne des écarts à la moyenne (leur somme a priori) est toujours égal à 0. Nous ne pouvons utiliser cet écart moyen comme indice de dispersion.

Comme la moyenne des écarts bruts = 0, on utilise soit les valeurs absolues, soit les carrés des écarts → variance et écart-type.

	x	$x - \bar{x}$
	5	-5
	6	-4
	7	-3
	8	-2
	9	-1
	10	0
	11	1
	12	2
	13	3
	14	4
	15	5
Moyenne	10	0

Mesures de Variabilité

42/99

❖ La variance de l'échantillon

La variance est un concept clé et une statistique régulièrement utilisée, mais elle n'est pas facile à interpréter. Elle s'écrit :

$$s_X^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$



	x	x-<x>	(x-<x>)^2
	5	-5	25
	6	-4	16
	7	-3	9
	8	-2	4
	9	-1	1
	10	0	0
	11	1	1
	12	2	4
	13	3	9
	14	4	16
	15	5	25
Moyenne	10	0	10

$$s_X^2 = \frac{\sum (X - \bar{X})^2}{N - 1} = 11$$

Mesures de Variabilité

43/99

❖ Ecart-type

L'écart type (s) se définit comme la *racine carrée positive de la variance*. Elle est parfois appelée **déviati**on **standard**. C'est la mesure de dispersion la plus utilisée et la plus facilement interprétable. Formellement :

$$s_X = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$$

Si nous reprenons l'exemple d'avant, l'écart type est donc de :



Remarque : le calcul de la variance et de l'écart-type se base sur les données de l'échantillon. Si vous voulez faire une estimation sur la population, vous retirez le 1 à la formule. s_X^2 devient alors σ_X^2 ; s_X devient σ_X

Mesures de Variabilité

44/99

❖ Ecart-type

Pour caractériser une distribution dans un mémoire ou dans un article de recherche, on note généralement de la façon suivante :

Moyenne $\pm$ Ecart-type [min-max]

Par exemple,



	x	$x - \langle x \rangle$	$(x - \langle x \rangle)^2$
	5	-5	25
	6	-4	16
	7	-3	9
	8	-2	4
	9	-1	1
	10	0	0
	11	1	1
	12	2	4
	13	3	9
	14	4	16
	15	5	25
Moyenne	10	0	10

Mesures de Variabilité

45/99

❖ Ecart-type

L'écart-type est particulièrement sensible aux valeurs extrêmes, raison pour laquelle il représente un bon indice de dispersion mais seulement quand les données suivent une distribution normale.

Patient	temps de réaction (ms)
Patient 1	540
Patient 2	623
Patient 3	750
Patient 4	422
Patient 5	453
Patient 6	472
Patient 7	650
Patient 8	524
Patient 9	583
Patient 10	701
Patient 11	423
Patient 12	499
Patient 13	523
Patient 14	522
Patient 15	605
Patient 16	702
Patient 17	621
Patient 18	530
Patient 19	589

Médiane

Moyenn

ET

Patient	temps de réaction (ms)
Patient 1	540
Patient 2	623
Patient 3	750
Patient 4	422
Patient 5	453
Patient 6	472
Patient 7	650
Patient 8	524
Patient 9	583
Patient 10	701
Patient 11	423
Patient 12	499
Patient 13	523
Patient 14	522
Patient 15	605
Patient 16	702
Patient 17	621
Patient 18	530
Patient 19	3523

Médiane =

Moyenne

ET

Moyenne =

Moyenne =

Synthèse

46/99

On considère l'échantillon de notes suivantes (population de petite taille) :

$$X = \{4, 6, 6, 8, 10, 12, 12, 14\}$$

Questions :

1. Calculer la **moyenne**.
2. Calculer la **médiane**.
3. Calculer les **écarts à la moyenne** et leur somme.
4. Calculer la **variance** et l'**écart-type** (population et échantillon).

1. Moyenne :

$$\bar{X} = \frac{4 + 6 + 6 + 8 + 10 + 12 + 12 + 14}{8} = \frac{72}{8} = 9$$

2. Médiane (effectif pair → moyenne des 2 valeurs centrales = 8 et 10) :

$$Me = \frac{8 + 10}{2} = 9$$

3. Écarts à la moyenne :

$$(-5, -3, -3, -1, 1, 3, 3, 5) \rightarrow \text{somme} = 0 \quad \checkmark$$

4. Variance (population) :

Paramètres

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{N} = \frac{88}{8} = 11$$

$$\sigma = \sqrt{11} \approx 3,32$$

Variance (échantillon) :

Statistiques

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{N - 1} = \frac{88}{7} \approx 12,57$$

$$s = \sqrt{12,57} \approx 3,55$$

Plan

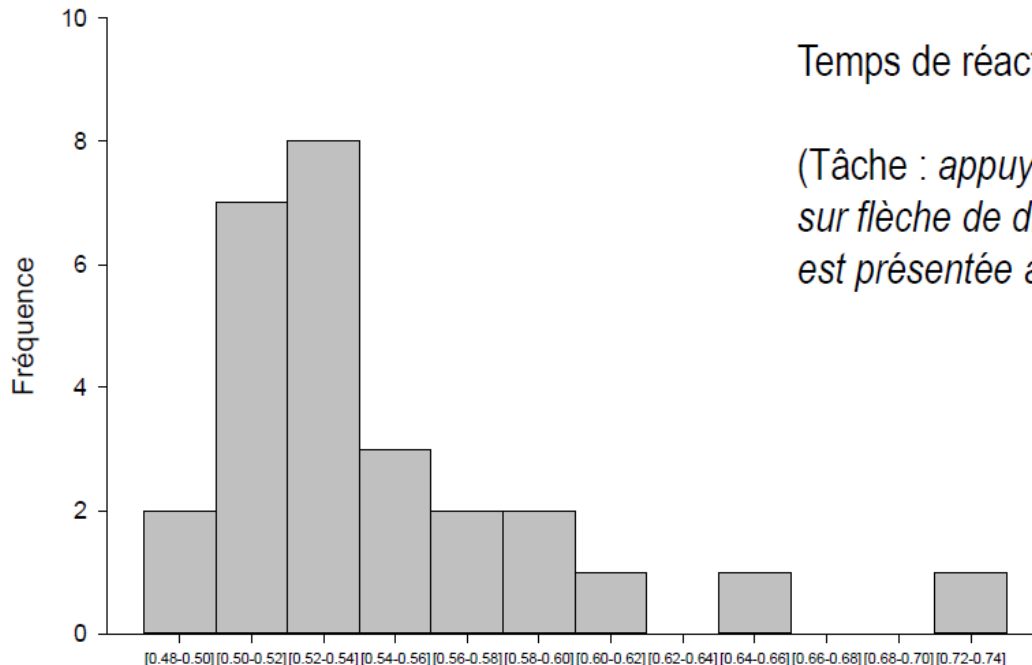
47/99

- 👉 Concepts de Base en Statistiques
- 👉 Description des Données
- 👉 La Distribution Normale
- 👉 Les Hypothèses
- 👉 Le Test d'Hypothèse (Appliqué aux Moyennes)

Description

48/99

- ❖ La distribution normale est l'une des distributions que l'on rencontre le plus (dans la nature), et ce d'autant plus quand on étudie des processus psychologiques.



Temps de réaction d'un patient

(Tâche : appuyer le plus rapidement possible sur flèche de droite quand une flèche droite est présentée à l'écran et vice-versa).

Distribution quasi normale, légèrement asymétrique avec deux outliers

Formulation Mathématique $\mathcal{N}(\mu, \sigma)$

49/99

Densité gaussienne

μ : moyenne ; σ : écart-type.

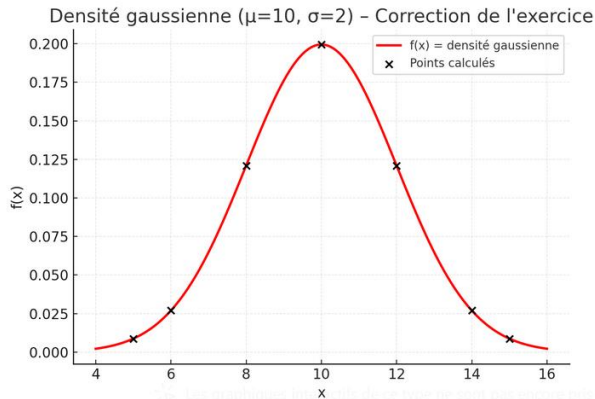
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

On fixe $\mu = 10$ et $\sigma = 2$.

Calculer $f(x)$ pour $x = [5, 6, 8, 10, 12, 14, 15]$.

(utiliser la formule directement ou Excel)

$$\frac{1}{\sigma\sqrt{2\pi}} = \frac{1}{2\sqrt{2\pi}} \approx 0,1995$$



x	$(x - \mu)^2 / (2\sigma^2)$	$\exp(-\dots)$	$f(x)$
5	3.1250	0.0439	0.0088
6	2.0000	0.1353	0.0270
8	0.5000	0.6065	0.1210
10	0.0000	1.0000	0.1995
12	0.5000	0.6065	0.1210
14	2.0000	0.1353	0.0270
15	3.1250	0.0439	0.0088

Moments d'ordre 3 & 4

50/99

Skewness S et Kurtosis (excès) C

Formules (versions "simples")

$$S = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3, \quad C = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 - 3,$$

$$\text{où } \bar{x} = \frac{1}{n} \sum x_i \text{ et } s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}.$$

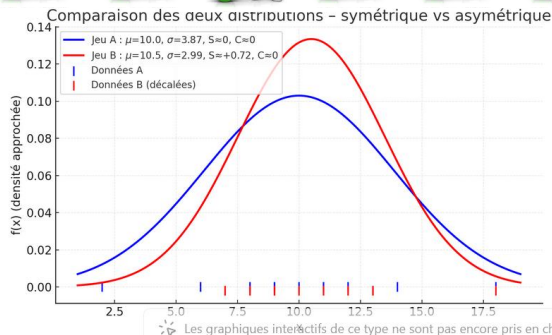
Jeu A — mésocurtique & symétrique

$x_A = [2, 6, 8, 8, 9, 9, 10, 11, 11, 12, 12, 14, 18]$

Donnés pour l'exercice : $\bar{x}_A = 10.0$, $s_A = 3.873$.

À faire : calculer S_A et C_A avec les formules ci-dessus.

Correction



À retenir

- Mésocurtique $\Leftrightarrow C \approx 0$ (cas normal en référence).
- Symétrie $\Leftrightarrow S = 0$.
- Queue droite $\Leftrightarrow S > 0$, queue gauche $\Leftrightarrow S < 0$.

— mésocurtique & asymétrique (queue droite)

$x_B = [7, 7, 7, 8, 8, 9, 9, 10, 11, 11, 12, 12, 13, 13, 13, 18]$

Donnés pour l'exercice : $\bar{x}_B = 10.5$, $s_B = 2.989$.

À faire : calculer S_B et C_B .

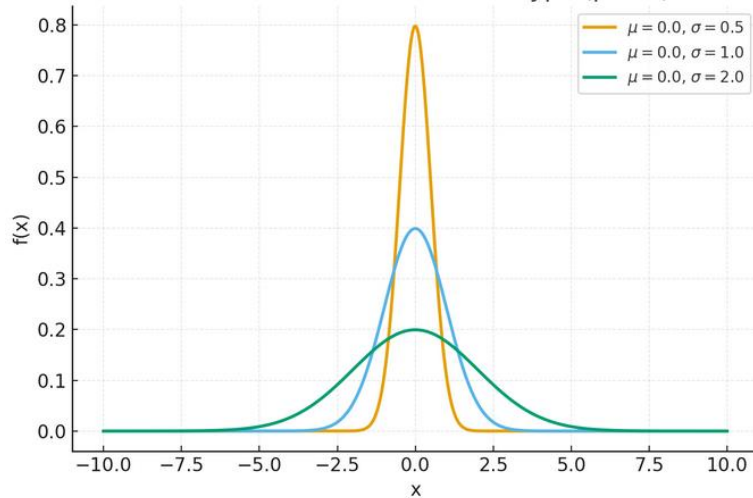
Correction



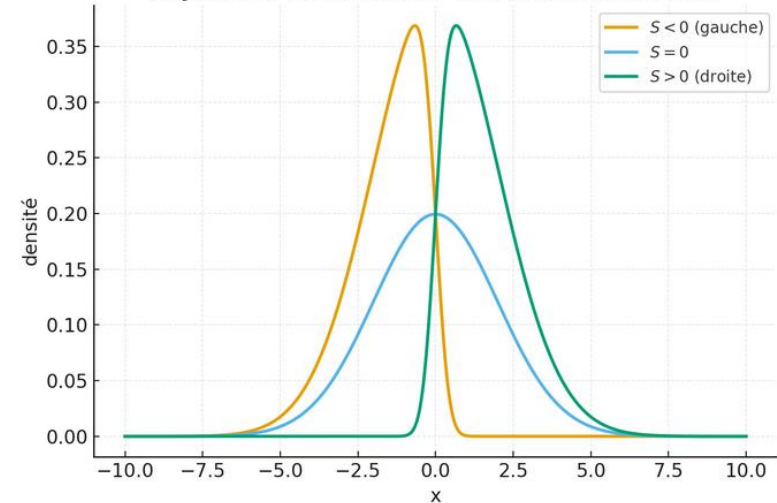
Différentes Formes

51/99

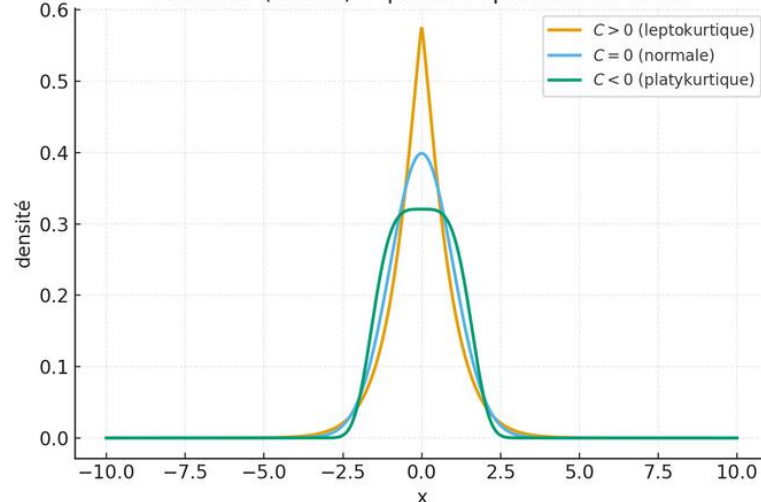
Loi normale : effet de l'écart-type (μ fixe)



Asymétrie (Skewness) : courbes skew-normal



Kurtosis (excès) : queues épaisses vs fines



La moyenne μ définit la position,
l'écart-type σ contrôle la dispersion,
l'asymétrie S oriente la distribution,
et la kurtosis C en façonne le relief.

Test de Normalité : Assistance Numérique

52/99

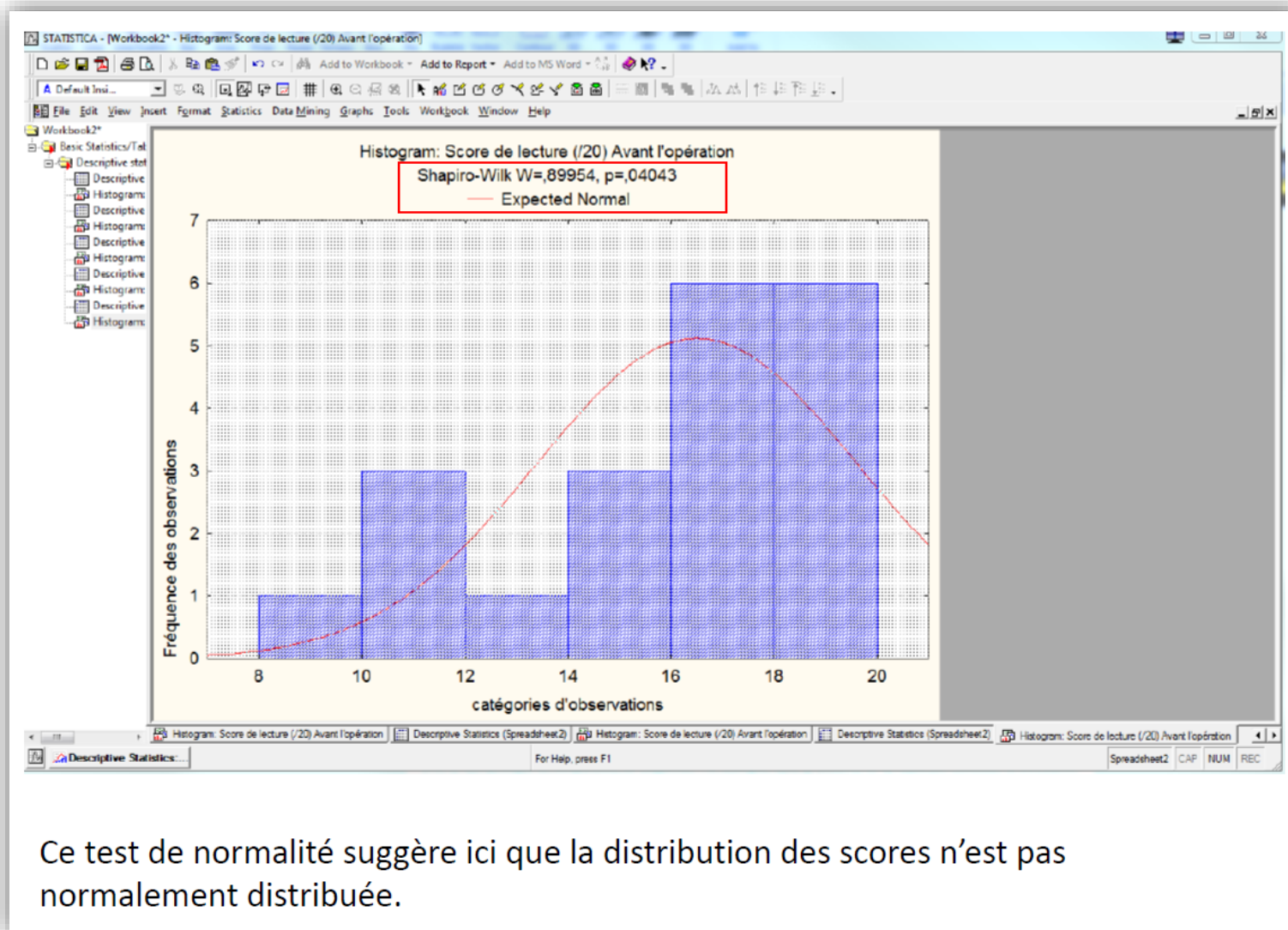
The screenshot displays the Minitab software interface. In the background, a spreadsheet titled 'STATISTICA - [Data: Spreadsheet2* (7v by 20c)]' is visible, containing data for 'Score de lecture (/20)' at three time points: 'Juste après opération' and 'Trois mois après l'opération'. The data is as follows:

5	6	7
Score de lecture (/20) Juste après opération	Score de lecture (/20) Juste après opération	Score de lecture (/20) Trois mois après l'opération
20	18	20
19	20	20
18	18	18
16	17	17
18	18	18
19	19	20
20	19	20
14	15	17
20	19	20

In the foreground, the 'Basic Statistics and Tables: Spreadsheet2' dialog box is open, with 'Descriptive statistics' selected. Below it, the 'Descriptive Statistics: Spreadsheet2' dialog box is also open, showing the 'Normality' tab. The 'Kolmogorov-Smirnov & Lilliefors test for normality' is selected under the 'Number of intervals' section. The 'Stem and leaf' section is also visible, with 'Stem & leaf plot' selected.

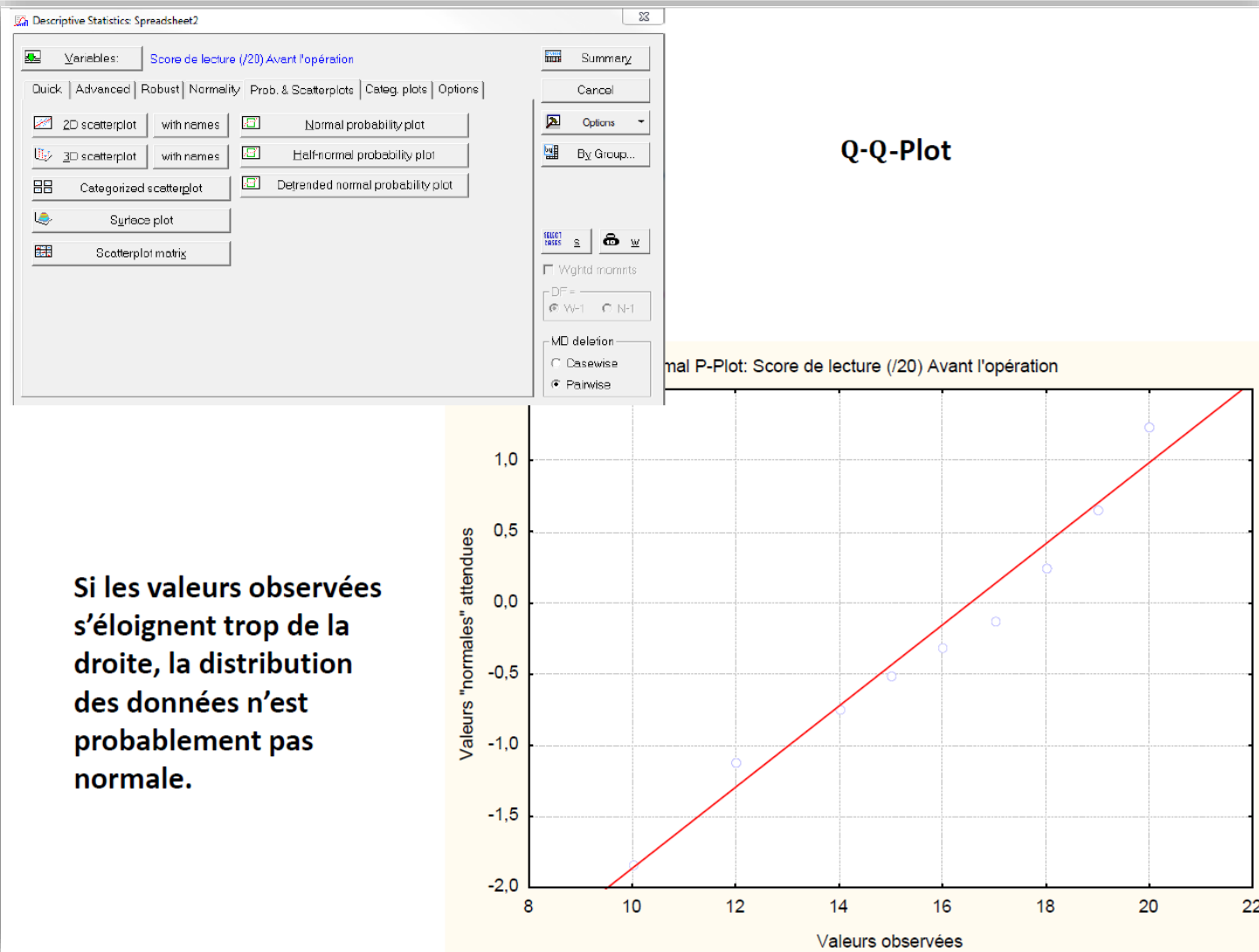
Test de Normalité : Tests Sophistiqués

53/99



Test de Normalité : Interprétation Graphique

54/99



Test de Normalité : Sous Excel

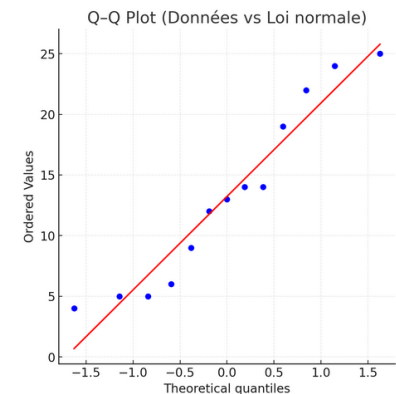
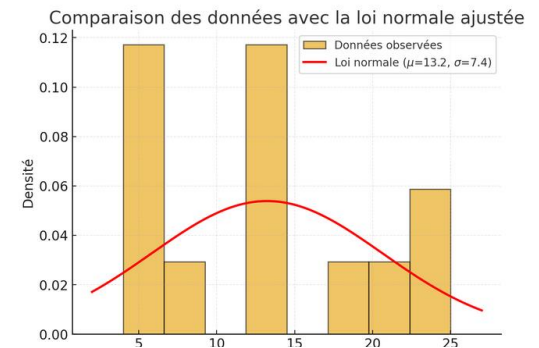
55/99

Le Test de Jarque-Bera est basé sur l'analyse du Kurtosis et du Skewness (par rapport à une distribution normale)

$$JB = (n/6) * (S^2 + (C^2/4))$$

Si $JB < 5,99$ alors la distribution est normale (χ^2 avec ddl=2 $n>20$)

Données			
4	observations	n	
5	asym	S	
5	kurt	C	
6	JB Stat	JB	
9			
12	p-Value	p	
13			
14			
14			
19			
22			
24			
25			



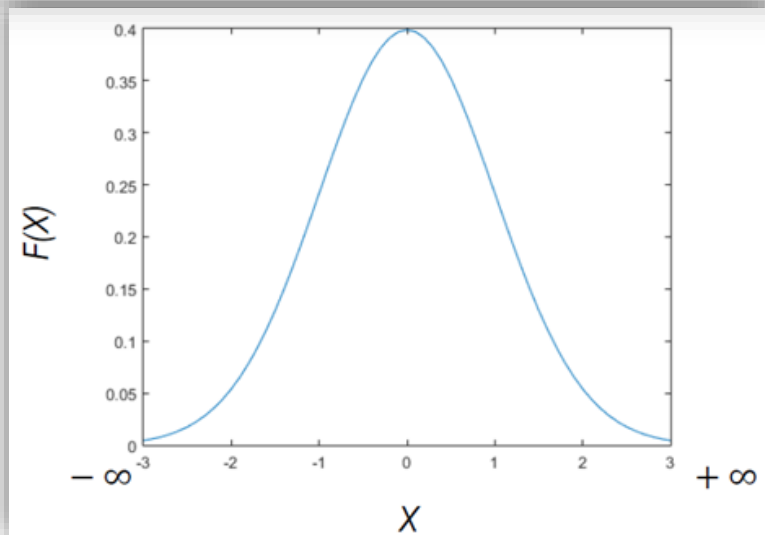
Description Normale Centrée Réduite

56/99

- Distribution centrée réduite**

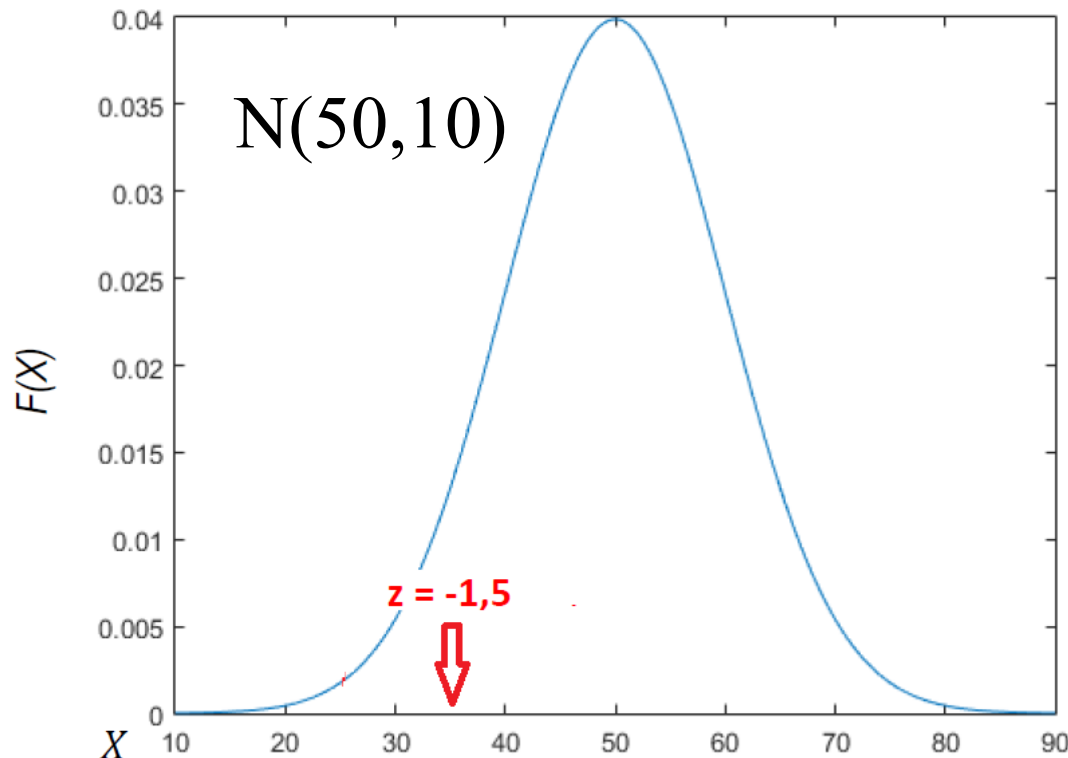
La distribution centrée réduite est une distribution normale de moyenne 0 et d'écart-type 1, notée $N(0, 1)$. Elle est utile pour effectuer des inférences, car toute distribution normale peut être transformée en distribution normale centrée réduite.

- La distribution $N(0, 1)$ est couramment utilisée pour situer un individu (ou score) par rapport à la distribution et pour déterminer la probabilité d'obtenir cette valeur dans la distribution.



Changement de Variables

57/99



Pour situer une valeur de distribution dans $N(0, 1)$, il suffit de procéder au calcul suivant :

$$z = \frac{X - \mu}{\sigma}$$

Par exemple, $X = 35$

$$z = \frac{35 - 50}{10}$$

$$z = \frac{-15}{10}$$

$$z = -1,5$$

$X - \mu$

z



$\mathcal{N}(0,1)$: Les Valeurs « Table »

58/99

La variable $Z = \frac{X - \mu}{\sigma}$ suit $N(0, 1)$.

L'aire sous la courbe représente une probabilité cumulée :

$$P(Z < z) = \text{surface à gauche de } z.$$

À $z = 0$, la moyenne divise la surface totale en deux moitiés égales (50 %).

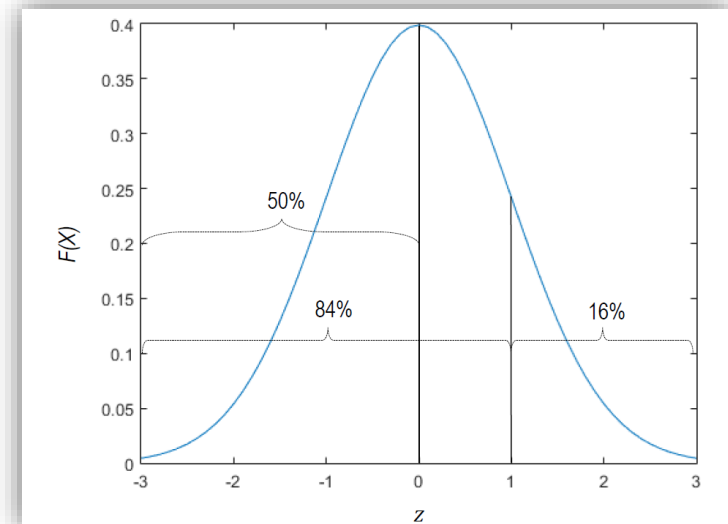
Entre 0 et $z = 1$, on trouve environ 34 % supplémentaires, soit 84 % au total à gauche de $+1\sigma$.

Au-delà de $z = 1$, il reste 16 % des observations.

Les valeurs de $P(Z < z)$ se lisent dans les **tables de la loi normale standard** ou via la fonction **NORM.S.DIST(z; TRUE)** d'Excel.

Les aires sous la courbe normale représentent des probabilités cumulées.

En pratique, les valeurs de $P(Z < z)$ sont lues dans les **tables de la loi normale standard**.



$\mathcal{N}(0,1)$: Les Valeurs de la Table ($z < 0$)

59/99

Number in the table represents $P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Exemple:

$z = -0.37$

$Z = -1.96$

Chaque case donne la probabilité cumulée

$P(Z \leq z)$

- Ligne → premières décimales de z
- Colonne → 2^e décimale

💡 Valeurs caractéristiques :

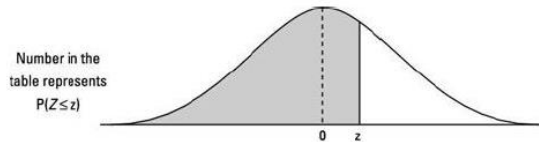
$P(Z < 0) = 0.5000, \quad P(Z < -1) = 0.1587$

⇒ donc 68 % des valeurs se trouvent entre -1σ et $+1 \sigma$.

Par symétrie : $P(Z < -a) = 1 - P(Z < a)$

$\mathcal{N}(0,1)$: Les Valeurs de la Table ($z > 0$)

60/99



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

Exemple:

$z = 0.37$



$Z = 1.96$



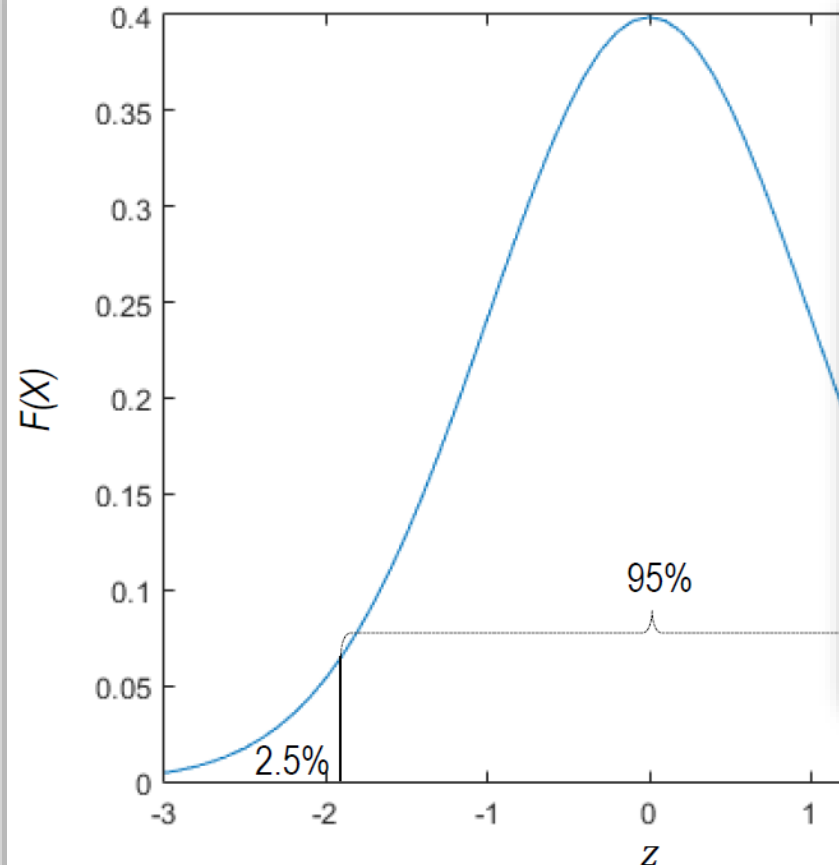
À retenir

- 68 % des valeurs sont dans $[\mu - 1\sigma, \mu + 1\sigma]$
- 95 % dans $[\mu - 1.96\sigma, \mu + 1.96\sigma]$
- 99 % dans $[\mu - 2.58\sigma, \mu + 2.58\sigma]$

Ces intervalles sont les **fondements des intervalles de confiance** et des **tests statistiques** (zones de 5 % ou 1% à rejeter).

$\mathcal{N}(0,1)$: Les Valeurs de la Table

61/99



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183

- Environ 98 % des sujets ont un score supérieur à 39.
- Le score de 39 est donc très faible, situé à 2 écarts-types sous la moyenne, donc hors de l'intervalle de confiance à 95 % $[-1.96\sigma, +1.96\sigma]$.

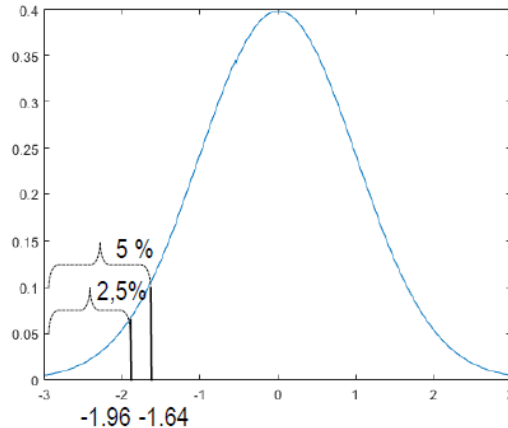
Exemple : vous obtenez les résultats en dénomination de 200 sujets (moyenne = 60,4; ET = 10,3). Vous souhaitez savoir où se situe le score d'un sujet individuel qui a obtenu un score de

39.



$\mathcal{N}(0,1)$: La Signification Probabiliste

62/99



Supposez que vous avez les données normatives en lecture de 100 sujets normaux (moyenne 25, ET = 3,5).

Vous obtenez les scores de deux patients (19 et 15), et souhaitez savoir où ils se situent par rapport aux sujet normaux.

Vous allez calculer les scores.



$$Z = \frac{X - \mu}{\sigma}; Z = \frac{19 - 25}{3,5}; Z = -1,71 \text{ (Valeur du tableau P = 0,0436)}$$

$$Z = \frac{X - \mu}{\sigma}; Z = \frac{15 - 25}{3,5}; Z = -2,85 \text{ (Valeur du tableau P = 0,0022)}$$

Ces probabilités (4,4 % et 0,2 %) ne représentent **pas** le risque d'erreur, mais la **rareté** du score sous l'hypothèse d'une population normale ($\mu = 25$, $\sigma = 3,5$).

Autrement dit :

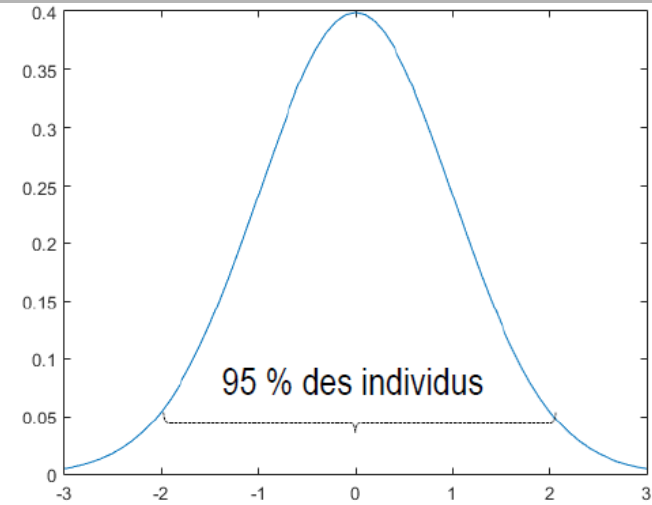
- "le patient 1 se situe dans les **4 % les plus faibles** de la population normale,"
- "le patient 2 dans les **0,2 % les plus faibles**."

Ces valeurs sont donc **statistiquement atypiques**, surtout pour le second.

Les Bornes d'Encadrement: Population Connue

63/99

- Vous connaissez la distribution des scores à un test de dénomination d'objets, avec :
 $\mu = 75$
 $\sigma = 5$
- Vous souhaitez calculer les bornes (les limites) entre lesquelles 95 % des scores sont distribués.

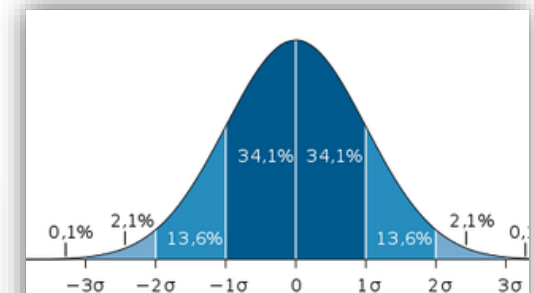


IC95% pour la Moyenne: Sampling

64/99

L'usage le plus simple des intervalles de confiance concerne les populations à distribution normale (en forme de cloche) dont on cherche à estimer la moyenne $\bar{X}$. Si on connaît l'écart type $\sigma(X)$ (ou si on en connaît une estimation assez fiable) de cette distribution, et si on mesure la moyenne $\bar{x}$ sur un échantillon de taille n pris au hasard, alors

- l'intervalle $\left[\bar{x} - \frac{\sigma(X)}{\sqrt{n}}; \bar{x} + \frac{\sigma(X)}{\sqrt{n}} \right]$ est un intervalle de confiance de $\bar{X}$ à environ 68 %
- l'intervalle $\left[\bar{x} - 2 \frac{\sigma(X)}{\sqrt{n}}; \bar{x} + 2 \frac{\sigma(X)}{\sqrt{n}} \right]$ est un intervalle de confiance de $\bar{X}$ à environ 95 %
- l'intervalle $\left[\bar{x} - 3 \frac{\sigma(X)}{\sqrt{n}}; \bar{x} + 3 \frac{\sigma(X)}{\sqrt{n}} \right]$ est un intervalle de confiance de $\bar{X}$ à environ 99,7 %



Le **théorème central limite** établit la convergence en loi de la somme d'une suite de variables aléatoires vers la loi normale. Intuitivement, ce résultat affirme qu'une somme de variables aléatoires indépendantes et identiquement distribuées tend vers une variable aléatoire **normale**.

Ces formules sont valables pour des échantillons supposés infinis ($n > 100$). Dans le cas d'échantillon plus petit, la consultation d'une table de distribution de la loi de Student est nécessaire.

Encore faut-il connaître ou avoir une estimation de l'écart type $\sigma(X)$. En pratique, on prend comme estimation de $\sigma(X)$ la valeur s , l'écart-type de la série de mesures issues de l'échantillon.

Ainsi l'on voit que pour augmenter la confiance, il faut élargir l'intervalle et pour obtenir un intervalle plus fin avec même degré de confiance, il faut augmenter la taille de l'échantillon.

IC95% pour la Moyenne : Exemple

65/99

Le staff médical d'une grande entreprise fait ses petites statistiques sur le taux de cholestérol de ses employés ; les observations sur 100 employés tirés au sort sont les suivantes.

Taux Cholesterol (TC)	Effectif (E)	TC x E	(TC-me)^2	E x (TC-me)^2
120	9	1080	8760,96	78848,64
160	22	3520	2872,96	63205,12
200	25	5000	184,96	4624
240	21	5040	696,96	14636,16
280	16	4480	4408,96	70543,36
320	7	2240	11320,96	79246,72
Somme	1320	100	28245,76	311104
Somme / E	13,2	1	213,6	3111,04

Calculer la moyenne m_e et l'écart-type σ_e sur l'échantillon.
Déterminer un intervalle de confiance pour la moyenne.



1,96

On obtient, sur l'échantillon, la moyenne $m_e = 214$, l'écart-type $\sigma_e = 55.77$.

On en déduit, au seuil 95%, un intervalle de confiance pour la moyenne : $[m_e - y_\alpha \frac{\hat{\sigma}_e}{\sqrt{n}}; m_e + y_\alpha \frac{\hat{\sigma}_e}{\sqrt{n}}]$

Ainsi le taux moyen de cholestérol est, à un seuil de confiance 95%, située entre 203 et 225 cg.

Plan

66/99

- 👉 Concepts de Base en Statistiques
- 👉 Description des Données
- 👉 La Distribution Normale
- 👉 Les Hypothèses
- 👉 Le Test d'Hypothèse (Appliqué aux Moyennes)

Distribution d'Echantillon

67/99

- Variabilité due au hasard

Lorsque vous tirez un échantillon au hasard pour établir des normes (ex : pour un test de compréhension orale), la distribution des scores peut varier d'un échantillon à un autre. Cela signifie que la moyenne ou l'écart-type peut légèrement différer entre les échantillons.

Cette variabilité naturelle, attendue en statistique, est appelée **variabilité due au hasard** ou parfois **erreur d'échantillonnage**.

Exemple : Dans l'illustration, on observe différentes distributions avec des moyennes (μ) et écarts-types (σ) légèrement différents selon l'échantillon choisi.

Distri 1: $\mu = 70, \sigma = 12$

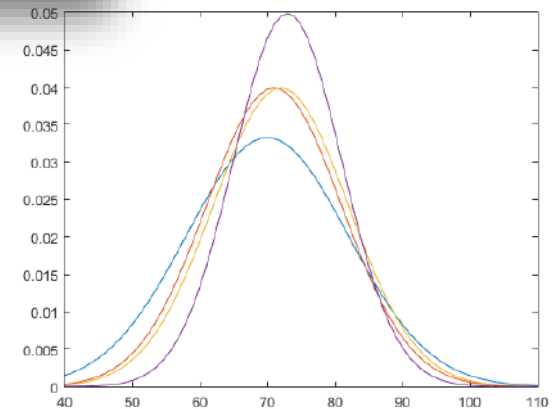
Distri 2: $\mu = 61, \sigma = 10$

Distri 3: $\mu = 72, \sigma = 10$

Distri 4: $\mu = 73, \sigma = 8$

► En d'autres termes, chaque échantillon donne une estimation légèrement différente des paramètres de la population

cette fluctuation aléatoire constitue l'essence de l'inférence statistique.



Variation de la distribution en fonction de l'échantillon

Distribution d'Echantillon

68/99

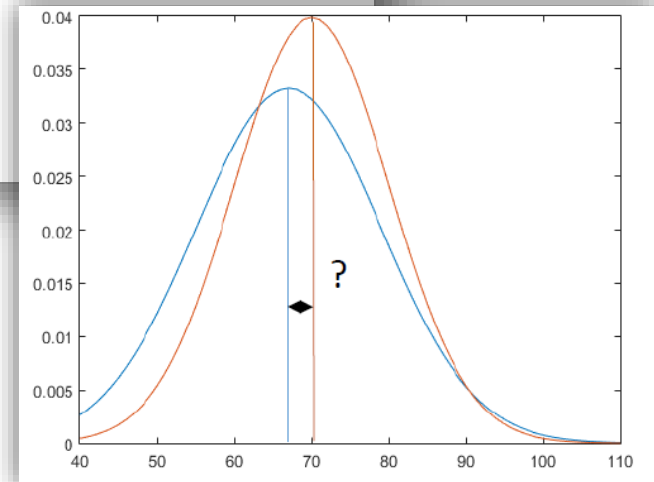
- **Les variations sont-elles dues au hasard ?**

Imaginons que vous disposiez des scores d'un échantillon de patients et des scores d'un échantillon de sujets normaux. La question est de savoir si la différence entre ces deux distributions est due au hasard.

- **Hypothèse 1** : La différence entre les moyennes est due au hasard ou à une erreur d'échantillonnage.
- **Hypothèse 2** : La différence entre les moyennes n'est pas due au hasard, mais au fait que les patients présentent une pathologie.

Exemple : Dans le graphique, les distributions ont les paramètres suivants :

- Distribution des patients : $\mu = 64, \sigma = 11$
- Distribution des sujets normaux : $\mu = 70, \sigma = 10$



Distribution d'Echantillon

69/99

- **Distribution d'échantillonnage**

Reprenons l'exemple précédent : « La différence entre les moyennes est-elle trop grande pour être attribuée au hasard ? » Pour répondre, nous utilisons des **distributions d'échantillonnage**.

Le concept de distribution d'échantillonnage est central dans les tests statistiques. Elle permet de déterminer les valeurs possibles d'une statistique (comme une moyenne) en fonction d'échantillons tirés d'une population donnée.

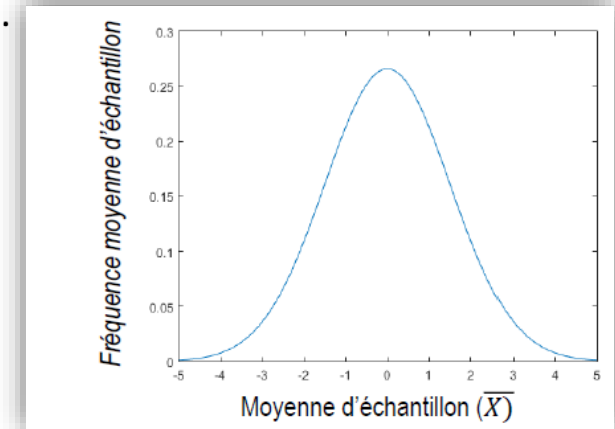
Ces distributions indiquent les valeurs que l'on pourrait observer pour une statistique, sous certaines conditions (par exemple, la différence maximale attendue entre les moyennes de deux échantillons tirés de populations identiques).

En résumé, la **distribution d'échantillonnage** d'une statistique est la distribution des valeurs obtenues pour cette statistique en répétant l'échantillonnage un grand nombre de fois.

Distribution d'échantillonnage des moyennes

70/99

- La distribution d'échantillonnage des moyennes représente la variabilité de la moyenne d'un échantillon ($\bar{X}$) lorsqu'on répète de très nombreux tirages aléatoires de taille n dans une même population.
- Sous certaines conditions (population de moyenne μ et variance σ^2), la distribution des moyennes d'échantillons est centrée sur μ et sa variance vaut σ^2/n .
- Lorsque n est grand, cette distribution tend vers une loi normale, quelle que soit la distribution initiale de la population (théorème central limite).
- Procédure théorique : on tire un grand nombre d'échantillons de taille n , on calcule leur moyenne, puis on observe la distribution des valeurs de $\bar{X}$.



Tests d'Hypothèses

71/99

- **Hypothèse de recherche**

- Par exemple, nous souhaitons tester l'hypothèse selon laquelle les patients atteints d'une lésion dans l'aire de Broca présentent des difficultés dans les aspects expressifs du langage, mesurées par un test de fluence phonologique (énoncer le maximum de mots commençant par la lettre "P" en deux minutes).
- Pour cela, nous avons recueilli les données de deux échantillons avec des caractéristiques sociodémographiques similaires (âge, niveau d'éducation, sexe, etc.), se distinguant uniquement par le fait que les sujets d'un échantillon ont une lésion dans l'aire de Broca.

- **Hypothèse nulle**

L'hypothèse nulle, notée H_0 , s'énonce comme suit :

Il n'y a pas de différence entre les moyennes des deux échantillons.

Ce qui signifie :

$$\mu_{\text{Sujets normaux}} = \mu_{\text{Patients}}$$

ou, de manière équivalente :

$$\mu_{\text{Sujets normaux}} - \mu_{\text{Patients}} = 0$$

L'Hypothèse Nulle

72/99

- **Hypothèse nulle**

La logique derrière l'utilisation de l'hypothèse nulle :

1. **Principe de réfutabilité (Fisher)** : On ne peut jamais prouver une hypothèse avec certitude, mais on peut démontrer son inexactitude.

Exemple : Observer que 100 personnes ont deux mains ne prouve pas que toutes les personnes en ont deux. Cependant, trouver une personne avec une main suffit pour rejeter l'hypothèse que "toutes les personnes ont deux mains".

2. **Aspect pratique** : L'hypothèse nulle est souvent utilisée pour tester la signification des résultats.

Exemple : Vous souhaitez démontrer que la moyenne des QI des orthophonistes est supérieure à 100. Pour cela, vous testez $H_0 : \mu = 100$ et pouvez éventuellement la rejeter si les résultats montrent une moyenne significativement supérieure (par exemple, $\mu = 102, 104$, ou 125).

L'Hypothèse Alternative H_1

73/99

- **L'hypothèse nulle et son opposée : l'hypothèse alternative**

Lorsqu'on souhaite démontrer que la moyenne des QI des orthophonistes est supérieure à 100, on doit formuler une hypothèse alternative.

- **Hypothèse nulle** : $H_0 : \mu = 100$
- **Hypothèse alternative** : Une première version pourrait être $H_1 : \mu \neq 100$.

Cependant, cela ne correspond pas exactement à ce que l'on souhaite tester (une valeur supérieure à 100). Il faut orienter l'hypothèse :

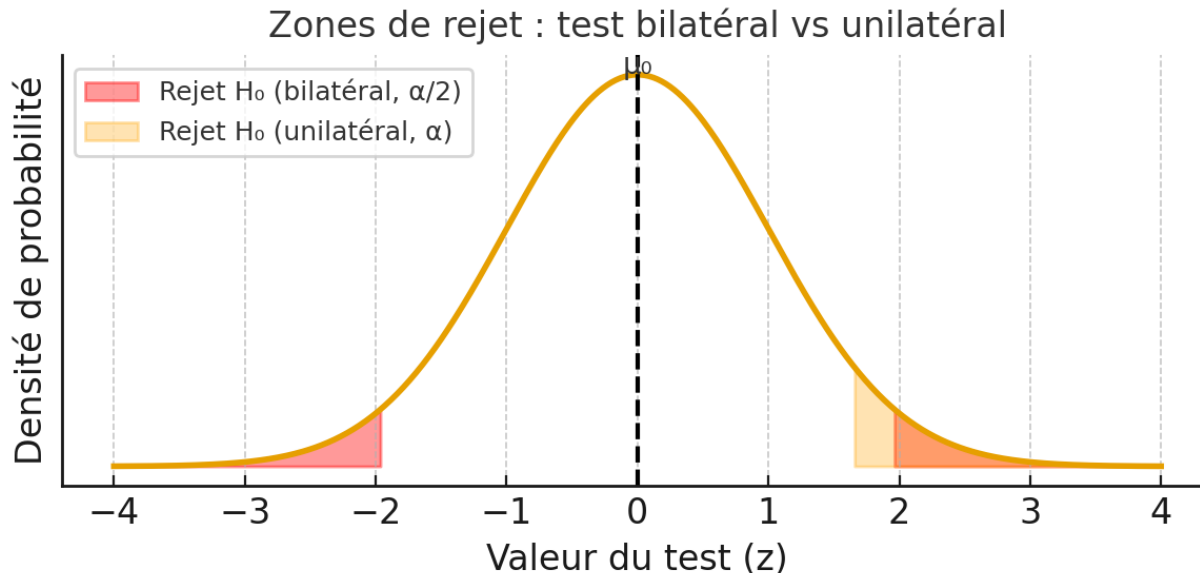
- **Hypothèse alternative orientée** : $H_1 : \mu > 100$

Test Bi- ou Uni-Latéral

74/99

- Un test bilatéral ($H_1 : \mu \neq \mu_0$) vérifie s'il existe une différence dans n'importe quel sens.
→ La zone de rejet est répartie dans les deux queues de la distribution ($\alpha/2$ de chaque côté).
- Un test unilatéral ($H_1 : \mu > \mu_0$ ou $\mu < \mu_0$) vérifie si l'effet va dans un sens précis.
→ Toute la probabilité α est placée d'un seul côté.

En pratique : si la direction n'est pas certaine avant l'expérience, on utilise le test bilatéral.



Statistiques de Test

75/99

- Les statistiques précédemment mentionnées (moyenne, médiane, écart-type, etc.) sont des **statistiques d'échantillon**. Il existe aussi des **statistiques de test**, qui sont associées à des procédures spécifiques et possèdent leurs propres distributions d'échantillonnage.
- Exemples : le **t de Student**, le **F de Fisher**.
- Le **t-test** (ou t de Student) est utilisé pour tester si la moyenne de deux échantillons est la même.
Hypothèse nulle : $H_0 : \mu_1 = \mu_2$.
- La distribution d'échantillonnage théorique du t-test est calculée en prenant un grand nombre de paires d'échantillons issus de deux populations identiques et en calculant le t pour chaque paire.

Loi Normale & Test d'Hypothèses

76/99

- **Hypothèse sur une observation spécifique**

- Imaginons que vous souhaitiez comparer le score de lecture d'un patient atteint d'une lésion dans l'aire visuelle des mots (aire occipito-temporale basale) avec les performances d'un échantillon de sujets normaux.

- **Hypothèse nulle :**

H_0 : *Le score du patient provient de la population normale* (c'est-à-dire que le score du patient n'est pas différent de celui de la population normale).

- En utilisant la moyenne et l'écart-type de la distribution de l'échantillon, vous pouvez calculer la probabilité d'obtenir un score aussi extrême.
 - Si cette probabilité est très faible, vous pouvez rejeter H_0 .
 - Si la probabilité n'est pas faible, vous ne pouvez pas remettre en question la validité de H_0 .

Loi Normale & Test d'Hypothèses

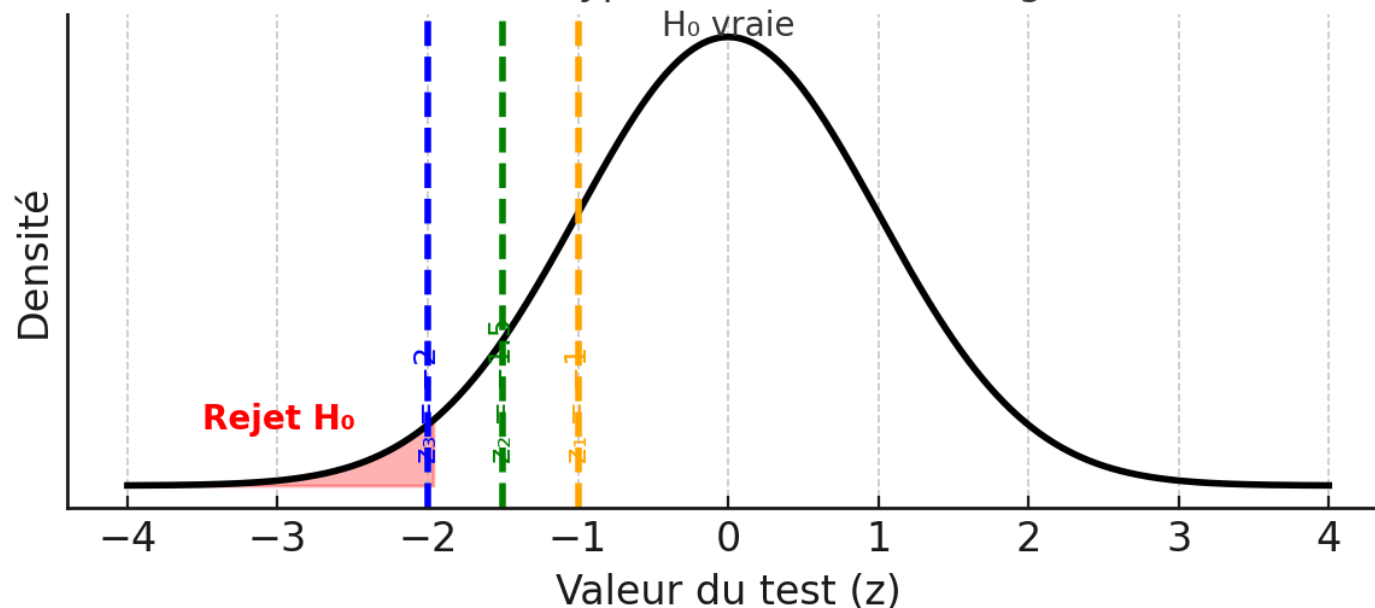
77/99

On souhaite tester si les scores individuels de lecture de 3 patients ($X_1=40$, $X_2=35$, $X_3=30$) diffèrent de la moyenne de référence de la population ($\mu_0=50$, $\sigma=10$).

→ Hypothèses :

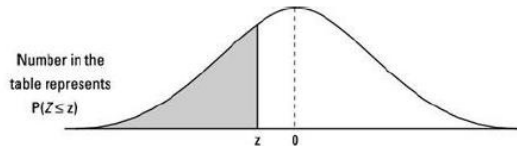
$H_0 : \mu = 50$ | $H_1 : \mu < 50$ (test unilatéral à gauche)

Illustration du test d'hypothèse unilatéral à gauche (z-test)



Loi Normale & Test d'Hypothèses

78/99



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Interprétation du test (unilatéral à gauche)

$$H_0 : \mu = 50 \quad ; \quad H_1 : \mu < 50$$

Chaque score observé (X_1, X_2, X_3) est converti en valeur z pour être comparé à la loi normale standardisée



X_3

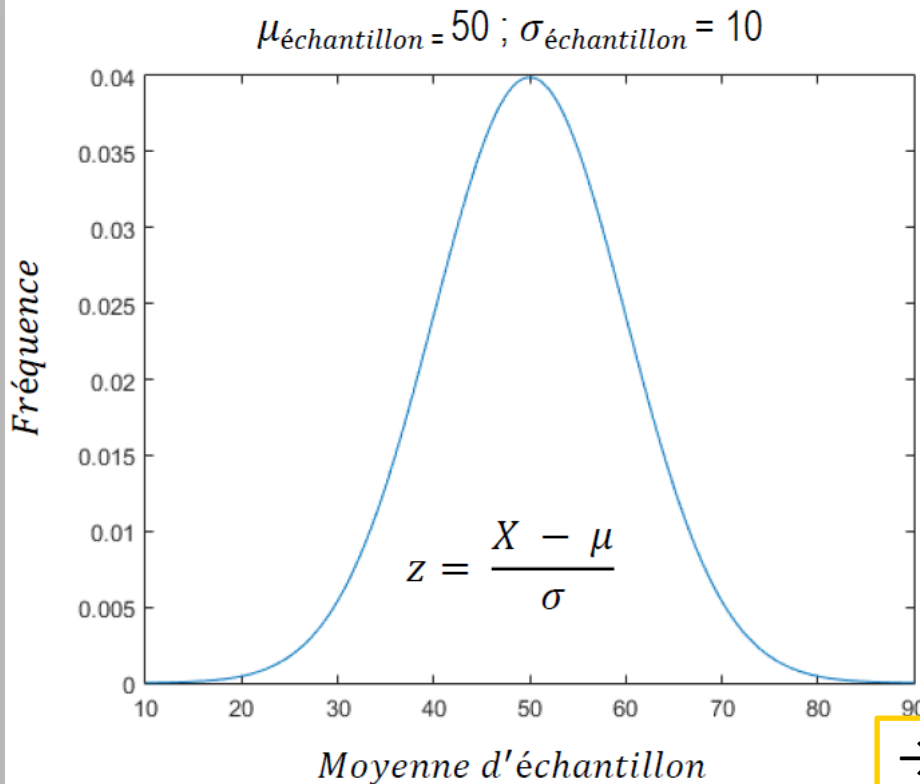
X_2

X_1

$$\begin{cases} H_0 : X_i \text{ appartient à la population normale } (\mu = 50) \\ H_1 : X_i \text{ a un score inférieur à la normale } (\mu < 50) \end{cases}$$

Loi Normale & Test d'Hypothèses

79/99



$X_1 = 40; z =$



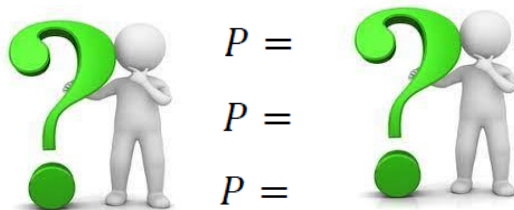
$X_2 = 35; z =$



$X_3 = 30 \quad z =$



Vous
regardez sur
les tables



$P =$
 $P =$
 $P =$

→ Interprétation : seul le troisième patient présente un score significativement inférieur ($p < 0,05$).

Les deux premiers restent compatibles avec la variabilité normale.

Loi Normale & Test d'Hypothèses

80/99

- En fonction de ces probabilités, vous devez décider de rejeter ou non votre hypothèse nulle.
- **Seuil de rejet (ou seuil de signification)** : Par convention, on rejette H_0 lorsque la probabilité sous H_0 est inférieure ou égale à 0,05 ($P \leq 0,05$).
- Le choix du seuil de rejet est arbitraire. En général, 0,05 est considéré comme suffisant, mais on peut aussi utiliser un seuil plus strict, comme 0,01.
- **Exemple** : Avec un seuil de rejet de 0,05, on peut rejeter l'hypothèse nulle pour X_3 . Avec un seuil de 0,01, on ne pourrait pas rejeter l'hypothèse pour les valeurs X_1 , X_2 , et X_3 .

Plan

81/99

- 👉 Concepts de Base en Statistiques
- 👉 Description des Données
- 👉 La Distribution Normale
- 👉 Les Hypothèses
- 👉 Le Test d'Hypothèses (Appliqué aux Moyennes)

Théorème Central Limite

82/99

• Rappel

La **distribution d'échantillonnage d'une statistique** représente les valeurs que l'on obtiendrait pour cette statistique en prélevant un nombre infini d'échantillons de la population et en calculant cette statistique pour chaque échantillon.

La distribution d'échantillonnage de la moyenne suit ce principe et repose sur le **théorème central limite** :

« Pour une population avec une moyenne μ et une variance σ^2 , la distribution d'échantillonnage de la moyenne aura :

- une moyenne égale à μ (c'est-à-dire $\mu_X = \mu$),
- une variance de $\frac{\sigma^2}{n}$ et
- un écart-type $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

La distribution tendra vers une distribution normale à mesure que n , la taille de l'échantillon, augmente. »

Comparaison des Moyennes : z-Test

83/99

Objectif

- H_0 (hypothèse nulle) : $\rightarrow H_0 : \mu = 100$
- H_1 (hypothèse alternative) : $\rightarrow H_1 : \mu \neq 100$ (test bilatéral)

Vérifier si la **pression artérielle moyenne (PAM)** d'un nouveau traitement antihypertenseur diffère de la valeur de référence normale de **100 mmHg**.

Données

- Taille de l'échantillon : $n = 36$ patients
- Moyenne observée : $\bar{X} = 92$ mmHg
- Écart-type de la population : $\sigma = 15$ mmHg
- Seuil de signification : $\alpha = 0,05$ (test bilatéral)

Calcul du test

Formule : $z = (\bar{X} - \mu_0) / (\sigma / \sqrt{n})$

$$\Rightarrow z = (92 - 100) / (15 / 6) = (-8) / 2,5 = -3,2$$

Décision

Valeur critique bilatérale : $z\alpha = \pm 1,96$

$|z_{\text{calculé}}| = 3,2 > 1,96 \Rightarrow \text{Rejet de } H_0$

Interprétation

La pression artérielle moyenne des patients traités est **significativement plus basse** que la valeur normale ($p < 0,01$).

→ Le traitement est **efficace** pour réduire la pression artérielle.

Comparaison des Moyennes : *t*-Test (Student)

84/99

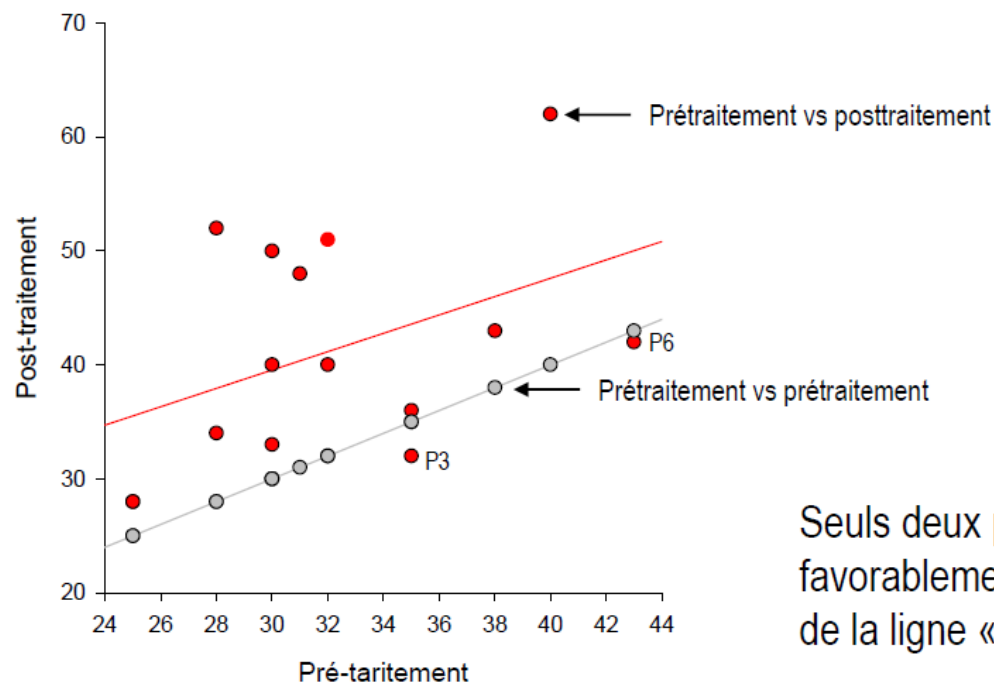
- Nous souhaitons comparer les moyennes de deux échantillons dits **pairés** (ce qui correspond à des **mesures répétées**). On peut aussi dire :
 - Échantillons liés
 - Échantillons corrélés
 - Échantillons appariés
 - Échantillons dépendants
- Dans ce cas, nous effectuerons un **t-test pour échantillons dépendants** (ou mesures répétées).
- **Exemple expérimental** : Vous voulez évaluer l'efficacité d'un programme de rééducation de l'accès au lexique. Pour cela, vous utilisez un test de dénomination passé par les patients avant (prétraitement) et après la rééducation (post-traitement).

Comparaison des Moyennes : *t*-Statistique

85/99

Tableaux des données

Identification	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	Moy.	ET.
Prétraitement	38	40	35	32	30	43	32	28	30	25	28	25	31	30	35	32,13	5,22
Posttraitement	43	62	32	40	40	42	51	52	33	28	34	28	48	50	36	41,27	9,84
Diff.	5	22	-3	8	10	-1	19	24	3	3	6	3	17	20	1	9,13	8,95



Seuls deux patients n'évaluent pas favorablement, ceux qui se situent en dessous de la ligne « grise »

Comparaison des Moyennes : t-Statistique

86/99

Tableaux des données

Identification	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	Moy.	ET.
Prétraitement	38	40	35	32	30	43	32	28	30	25	28	25	31	30	35	32,13	5,22
Posttraitement	43	62	32	40	40	42	51	52	33	28	34	28	48	50	36	41,27	9,84
Diff.	5	22	-3	8	10	-1	19	24	3	3	6	3	17	20	1	9,13	8,95

- Cette différence de moyenne est-elle assez grande pour que l'on considère que la réhabilitation a eu un effet favorable sur la dénomination ?

$$H_0: \mu_{\text{Prétraitement}} = \mu_{\text{Posttraitement}}$$

- Toutefois, il est plus intéressant de s'intéresser directement aux scores de différences et ce pour chaque patient. Si la rééducation n'a aucun effet, la moyenne des scores des différences devrait être égale ou proche de 0. L'hypothèse nulle va donc s'écrire de la façon suivante:

$$H_0: \mu_D = \mu_{\text{Posttraitement}} - \mu_{\text{Prétraitement}} = 0$$

Comparaison des Moyennes : t-Statistique

87/99

Tableaux des données

Identification	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	Moy.	ET.
Prétraitement	38	40	35	32	30	43	32	28	30	25	28	25	31	30	35	32,13	5,22
Posttraitement	43	62	32	40	40	42	51	52	33	28	34	28	48	50	36	41,27	9,84
Diff. (D)	5	22	-3	8	10	-1	19	24	3	3	6	3	17	20	1	9,13	8,95

❖ Calcul de t

$$t = \frac{\bar{D} - 0}{s_{\bar{D}}} = \frac{\bar{D} - 0}{\frac{s_D}{\sqrt{N}}}$$

Avec

s_D : écart-type des différences

N : Nombre de différences

$\bar{D}$: Moyenne des différences

Dans notre exemple:

$$t = \frac{\bar{D} - 0}{s_{\bar{D}}} = \frac{\bar{D} - 0}{\frac{s_D}{\sqrt{N}}} = \frac{9,13 - 0}{\frac{8,95}{\sqrt{15}}} = 3,95$$

Le degré de liberté est toujours le
nombre de paires moins 1
Donc $15 - 1 = 14$

Moyennes : t-Table Vs t-Statistique

88/99

Degrees of freedom	Significance level					
	20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.043	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.158	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

$$t = \frac{\bar{D} - 0}{S_{\bar{D}}} = \frac{\bar{D} - 0}{\frac{S_D}{\sqrt{N}}} = \frac{9,13 - 0}{\frac{8,95}{\sqrt{15}}} = 3,95$$

- Si nous prenons un seuil de signification de $\alpha = .05$, $t_{.05(14)} = \pm 2.145$.
- Etant donné que $t = 3,95$ dans notre exemple, nous pouvons rejeter l'hypothèse nulle et dire que les patients ont progressé de façon significative.
- Nous pouvons même rejeter l'hypothèse à un seuil inférieur à .01.
- Remarquez que si vous avions eu 4 sujets supplémentaires, nous aurions pu rejeter l'hypothèse à $p < .001$

Test de Student : Outils Numériques

89/99

STATISTICA - [Data: Spreadsheet2* (3v by 20c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

	1 Identification Sujets	2 Prétraitement	3 Posttraitement
1	P1	38	43
2	P2	40	62
3	P3	35	32
4	P4	32	40
5	P5	30	40
6	P6	43	42
7	P7	32	61
8	P8	28	52
9	P9	30	33
10	P10	25	28
11	P11	28	34
12	P12	25	28
13	P13	31	48
14	P14	30	60
15	P15	35	36
16			
17			
18			
19			
20			

Basic Statistics and Tables: Spreadsheet2

Quick

- Descriptive statistics
- Correlation matrices
- t-test, independent, by groups
- t-test, independent, by variables
- t-test, dependent samples
- t-test, single sample
- Breakdown & one-way ANOVA
- Breakdown; non-factorial tables
- Frequency tables
- Tables and banners
- Multiple response tables
- Difference tests: t, %, means
- Probability calculator

OK

Cancel

Options

Open Data

SELECT CASES S W

Test de Student : *p*-Value

90/99

T-Test for Dependent Samples: Spreadsheet2

Variables:

First list: none
Second list: none

Quick | Advanced

Summary: T-tests
Box & whisker plots

Summary
Cancel
Options
By Group...
SELECT CHSES
Wghtd momnts

Select one or two variable lists

1 - Identification Sujets
2 - Prètraitement
3 - Postraitement

OK
Cancel
[Bundles]...
Use the "Show appropriate variables"

Select All | Spread | Zoom

First variable list:

☐ Show appropriate variables only

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
Prètraitement	32.13333	5.221749						
Postraitement	41.26667	9.837731	15	-9.13333	8.951190	-3.95179	14	0.001447

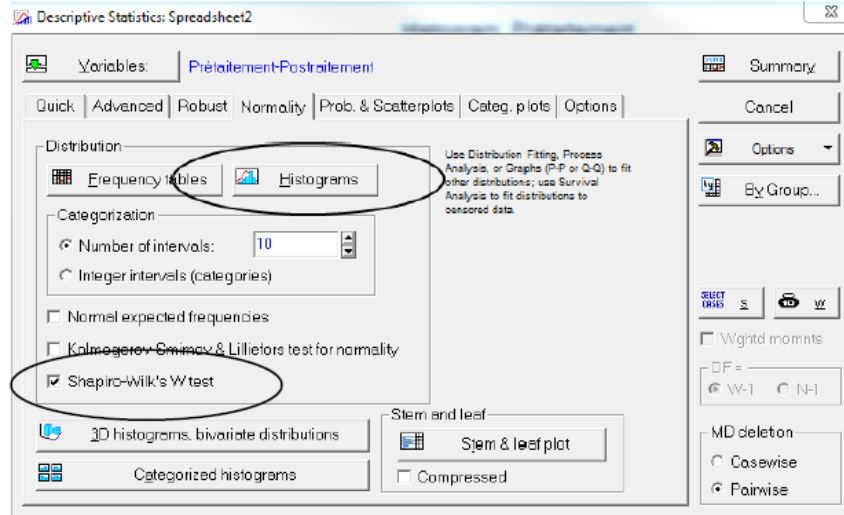
Dans les articles scientifiques, les résultats doivent être écrits de la manière suivante:

$$t_{(14)} = 3,95, p < 0.01 \text{ (ou } < 0.005)$$

Remarque : étant donné que le t-test est basé sur les paramètres de la loi normale, nous aurions dû vérifier la normalité des données. Faisons maintenant, *a posteriori*.

Test de Student : Faisabilité

91/99



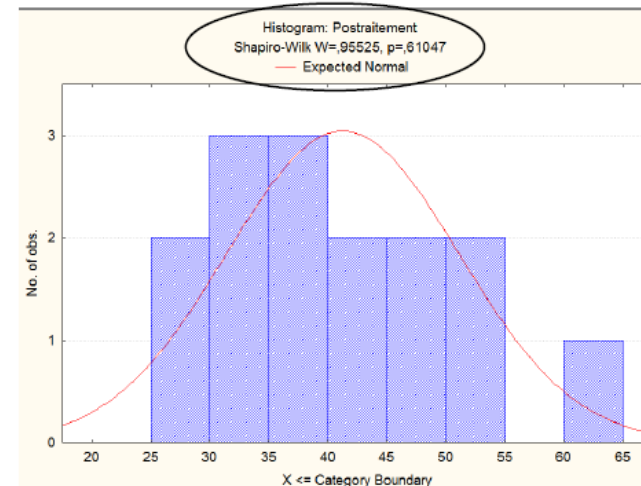
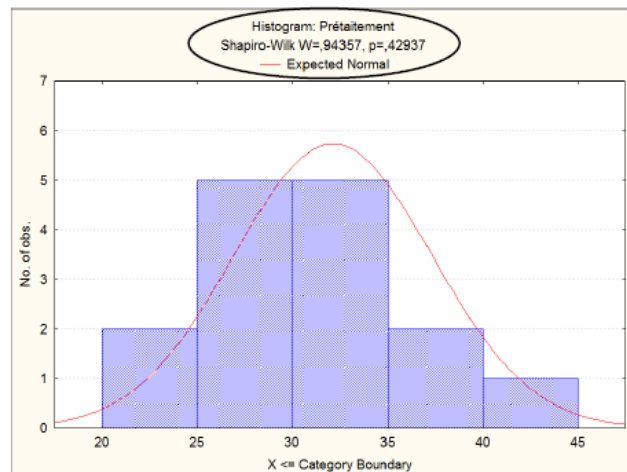
Vous allez utiliser le test de Shapiro-Wilk's

Interprétation du test de Shapiro-Wilk :

Hypothèses du test :

H_0 : la distribution est normale
 H_1 : la distribution n'est pas normale

- Si $p > 0.05$ → on ne rejette pas H_0 → les données sont compatibles avec une loi normale.
- Si $p \leq 0.05$ → on rejette H_0 → les données ne suivent pas une loi normale.



Test de Student Non-Apparié

92/99

- ❖ Nous souhaitons ici comparer les moyennes de deux échantillons dits indépendants ce qui veut dire que les données proviennent de **deux groupes indépendants** et pas *des mêmes sujets*.
- ❖ Dans ce cas précis, nous allons utiliser un *t*-test pour échantillon indépendant dont la formulation mathématique est la suivante :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Si les Variances des 2 groupes sont différentes (hétéroscédasticité)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad s_p = \sqrt{\frac{(n_1 - 1) s_{X_1}^2 + (n_2 - 1) s_{X_2}^2}{n_1 + n_2 - 2}}$$

Si les Variances des 2 groupes sont équivalentes (homoscédasticité)

Le *dl* va s'obtenir en calculant :

$$n_1 + n_2 - 2$$

Remarque importante: la variance rentre en considération dans ces équations. Une des conditions d'applications de ce test est l'homogénéité des variances. Celle-ci doit donc être vérifiée.

Test de Student : Synthèse

93/99

Cas	Quand l'utiliser	Formule du test t	Remarques / Hypothèses
t de Student apparié	Quand les deux mesures proviennent des mêmes sujets (ou de paires naturellement associées : avant/après traitement, œil gauche/droit, etc.)	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$ avec $\bar{d}$ la moyenne des différences et s_d leur écart-type.	Test sur les différences intra-sujets . Hypothèse : $H_0 : \bar{d} = 0$. Très puissant quand les observations sont corrélées.
t de Student non apparié	Quand les deux groupes sont indépendants (ex. hommes vs femmes, groupe contrôle vs traitement).	$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	Cas général (hétéroscédasticité) : variances inégales, test de Welch . On ne suppose pas d'égalité des variances. Si variances $\approx$ égales, on peut utiliser la version classique : $ddl = n_1 + n_2 - 2$.

⚠ En pratique

- On **n'utilise presque jamais** le cas d'**homoscédasticité stricte**.
- Les logiciels (R, Python, SPSS, etc.) appliquent automatiquement la **correction de Welch**.
- Dans les cours ou calculs manuels, on garde souvent $ddl = n_1 + n_2 - 2 \rightarrow$ **ce n'est pas une erreur**, c'est l'approche canonique du test de Student.

Les moyennes diffèrent significativement si $|t_{\text{calculé}}| > t_{\text{table}}$ ($ddl = n_1 + n_2 - 2$, $\alpha = 0.05$).

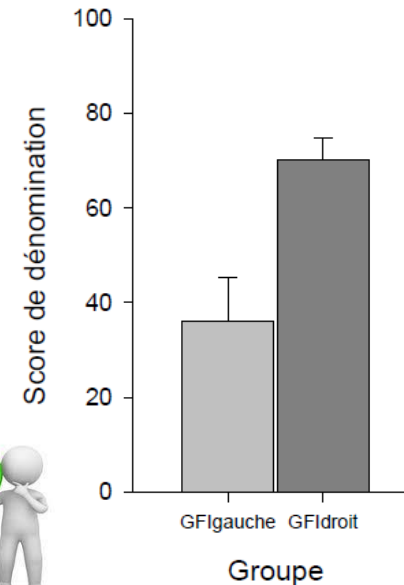
Test de Student Non-Apparié (hétéroSC)

94/99

Identification	Groupe	Dénomination
P1	GFIgauche	24
P2	GFIgauche	32
P3	GFIgauche	36
P4	GFIgauche	40
P5	GFIgauche	24
P6	GFIgauche	38
P7	GFIgauche	22
P8	GFIgauche	44
P9	GFIgauche	42
P10	GFIgauche	50
P11	GFIgauche	48
P12	GFIgauche	32
P13	GFI droit	70
P14	GFI droit	74
P15	GFI droit	72
P16	GFI droit	68
P17	GFI droit	69
P18	GFI droit	78
P19	GFI droit	77
P20	GFI droit	62
P21	GFI droit	64
P22	GFI droit	70
P23	GFI droit	71
P24	GFI droit	68

❖ Supposons que nous voulions comparer les performances en dénomination de patients avec une lésion du GFI gauche versus des patients atteints d'une lésion du GFI droit.

GFI = aire de Broca



Que peut-on suspecter en regardant cet histogramme ?

Test de Student Non-Apparié (hétéroSC)

95/99

The screenshot displays the STATISTICA software interface. The main window shows a spreadsheet with the following data:

	1 Identification Sujets	2 Groupe	3 Dénomination
1	P1	GFIgauche	24
2	P2	GFIgauche	32
3	P3	GFIgauche	36
4	P4	GFIgauche	40
5	P5	GFIgauche	24
6	P6	GFIgauche	38
7	P7	GFIgauche	22
8	P8	GFIgauche	44
9	P9	GFIgauche	42
10	P10	GFIgauche	50
11	P11	GFIgauche	48
12	P12	GFIgauche	32
13	P13	GFIdroit	70
14	P14	GFIdroit	74
15	P15	GFIdroit	72
16	P16	GFIdroit	68
17	P17	GFIdroit	69
18	P18	GFIdroit	78
19	P19	GFIdroit	77
20	P20	GFIdroit	62
21	P21	GFIdroit	64
22	P22	GFIdroit	70
23	P23	GFIdroit	71
24	P24	GFIdroit	68

The 'Basic Statistics and Tables: Spreadsheet2' dialog box is open, showing the 'Quick' tab. The 't-test, independent by groups' option is selected. Other options include Descriptive statistics, Correlation matrices, t-test, independent by variables, t-test, dependent samples, t-test, single sample, Breakdown & one-way ANOVA, Breakdown, non-factorial tables, Frequency tables, Tables and banners, Multiple response tables, Difference tests: r, %, means, and Probability calculator.

Test de Student Non-Apparié (hétéroSC)

96/99

STATISTICA - [Data: Spreadsheet2* (3v by 24c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

1 Identification Sujets 2 Groupe 3 Dénomination

1 P1 GFgauche 24

2 P2 GFgauche 32

3 P3 GFgauche 36

4 P4 GFgauche 36

5 P5 GFgauche 36

6 P6 GFgauche 36

7 P7 GFgauche 36

8 P8 GFgauche 36

9 P9 GFgauche 36

10 P10 GFgauche 36

11 P11 GFgauche 36

12 P12 GFgauche 36

13 P13 GFgauche 36

14 P14 GFgauche 36

15 P15 GFgauche 36

16 P16 GFgauche 36

17 P17 GFgauche 36

18 P18 GFgauche 36

19 P19 GFgauche 36

20 P20 GFgauche 36

21 P21 GFgauche 36

22 P22 GFgauche 36

23 P23 GFgauche 36

24 P24 GFgauche 36

25 P25 GFgauche 36

26 P26 GFgauche 36

27 P27 GFgauche 36

28 P28 GFgauche 36

29 P29 GFgauche 36

30 P30 GFgauche 36

31 P31 GFgauche 36

32 P32 GFgauche 36

33 P33 GFgauche 36

34 P34 GFgauche 36

35 P35 GFgauche 36

36 P36 GFgauche 36

37 P37 GFgauche 36

38 P38 GFgauche 36

39 P39 GFgauche 36

40 P40 GFgauche 36

41 P41 GFgauche 36

42 P42 GFgauche 36

43 P43 GFgauche 36

44 P44 GFgauche 36

45 P45 GFgauche 36

46 P46 GFgauche 36

47 P47 GFgauche 36

48 P48 GFgauche 36

49 P49 GFgauche 36

50 P50 GFgauche 36

51 P51 GFgauche 36

52 P52 GFgauche 36

53 P53 GFgauche 36

54 P54 GFgauche 36

55 P55 GFgauche 36

56 P56 GFgauche 36

57 P57 GFgauche 36

58 P58 GFgauche 36

59 P59 GFgauche 36

60 P60 GFgauche 36

61 P61 GFgauche 36

62 P62 GFgauche 36

63 P63 GFgauche 36

64 P64 GFgauche 36

65 P65 GFgauche 36

66 P66 GFgauche 36

67 P67 GFgauche 36

68 P68 GFgauche 36

69 P69 GFgauche 36

70 P70 GFgauche 36

71 P71 GFgauche 36

72 P72 GFgauche 36

73 P73 GFgauche 36

74 P74 GFgauche 36

75 P75 GFgauche 36

76 P76 GFgauche 36

77 P77 GFgauche 36

78 P78 GFgauche 36

79 P79 GFgauche 36

80 P80 GFgauche 36

81 P81 GFgauche 36

82 P82 GFgauche 36

83 P83 GFgauche 36

84 P84 GFgauche 36

85 P85 GFgauche 36

86 P86 GFgauche 36

87 P87 GFgauche 36

88 P88 GFgauche 36

89 P89 GFgauche 36

90 P90 GFgauche 36

91 P91 GFgauche 36

92 P92 GFgauche 36

93 P93 GFgauche 36

94 P94 GFgauche 36

95 P95 GFgauche 36

96 P96 GFgauche 36

97 P97 GFgauche 36

98 P98 GFgauche 36

99 P99 GFgauche 36

100 P100 GFgauche 36

T-Test for Independent Samples by Groups: Spreadsheet2

Variables: Dependent: none Grouping: none

Code for Group 1: Code for Group 2:

Quick Advanced Options

Summary: T-tests

Box & whisker plot

Summary Cancel Options By Group... SELECT CASES Weighted moments DF =

Select the dependent variables and one grouping variable

1 - Identification Sujets 2 - Groupe 3 - Dénomination

1 - Identification Sujets 2 - Groupe 3 - Dénomination

OK Cancel [Bundles]...

Use the "Show appropriate variables only" option to pre-screen variable lists and show categorical and continuous variables. Press F1 for more information.

Select All Spread Zoom Select All Spread Zoom

Dependent variables: 3

Grouping variable: 2

Show appropriate variables only

T-Test for Independent Samples by Groups: Spreadsheet2

Variables: Dependent: Dénomination Grouping: Groupe

Code for Group 1: GFgauche Code for Group 2: GFdroit

Quick Advanced Options

Summary: T-tests

Box & whisker plot

Summary Cancel Options By Group... SELECT CASES Weighted moments DF = MD deletion Casewise Pairwise

Test de Student Non-Apparié (hétéroSC)

97/99

Pour les plus courageux
Faire un calcul manuel

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Les déviations
standards sont
éloignées

Le test du F-ratio
démontre que les
variances sont
significativement
différentes entre els
deux groupes

itics/Tables (Spreadsheet2)	T-tests; Grouping: Groupe (Spreadsheet2)					
rtive statistics dialog	Group 1: GFgauche					
rogram: Dénomination	Group 2: GFdroit					
Groups	Variable	Mean	Mean	t-value	df	p
Descriptive Statistics (Spre	GFgauche	36,00000	70,25000	-11,2740	22	0,000000
Histogram: Dénomination	Dénomination					
upe=GFgauche						
Descriptive Statistics (Spre						
Histonram: Dénomination						

	GFgauche	GFIdroit	
Card	12	12	
moyenne	36	70,25	
variance	88,7272727	22,0227273	
ecart type	9,41951553	4,69283787	
t-test	-11,2740338		
Lecture table (0,05)	2,074		ddl=22
Comparaison	abs(-11,274)>2,074		
H0	rejeté		
Significativité	"+++"		

$$t_{(22)} = -11,27, p < 0.0000001$$

Panorama des Tests Paramétriques

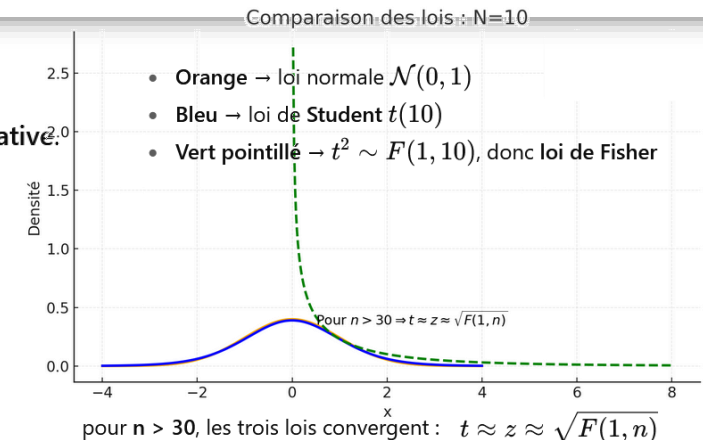
98/99

Les trois tests paramétriques fondamentaux

Test	Objectif	Formule	Exemple	Décision / Interprétation
Test de Student (t-test)	Compare les moyennes de deux groupes (indépendants ou appariés).	$t = (\bar{X}_1 - \bar{X}_2) / \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$	Comparer les scores moyens avant/après traitement.	On rejette $H_0 : \bar{X}_1 = \bar{X}_2$ si la valeur absolue de t calculé est supérieure à la valeur t table (ddl = $n_1 + n_2 - 2$). → Différence significative si $t_{\text{calculé}} > t_{\text{critique}} (\alpha = 0,05)$.
Test de Fisher (F-test)	Compare les variances de deux groupes (homogénéité des dispersions).	$F = s_1^2 / s_2^2$	Vérifier si deux groupes présentent la même variabilité.	On rejette $H_0 : s_1^2 = s_2^2$ si $F_{\text{calculé}} > F_{\text{table}} (\text{ddl}_1, \text{ddl}_2)$. → Variances différentes = hétéroscédasticité.
Test Z (normal)	Compare une moyenne observée à une valeur théorique (σ connue ou $n > 30$).	$z = (\bar{X} - \mu_0) / (\sigma / \sqrt{n})$	Tester si la moyenne d'un groupe diffère d'une norme de référence.	On rejette $H_0 : \bar{X} = \mu_0$ si la valeur absolue de z_calculé est supérieure à z_table. → Différence significative si $z_{\text{calculé}} > 1,96 (\alpha = 0,05 \text{ bilatéral})$.

À retenir :

- Si la statistique calculée dépasse la valeur critique, la différence est **statistiquement significative**.
- Ces tests supposent une **distribution normale** des données.
- Si ce n'est pas le cas, on utilise les **tests non paramétriques** :
 - Mann-Whitney (équivalent du t-test non apparié)
 - Wilcoxon (équivalent du t-test apparié)
 - Kruskal-Wallis (équivalent de l'ANOVA non paramétrique).



MERCI POUR VOTRE ATTENTION

99/99

