

An aerial photograph of a dense forest of evergreen trees covered in snow, viewed from a high angle. The trees are scattered across a light-colored, snow-covered ground. The overall scene is monochromatic, with various shades of white, grey, and black.

Statistiques Déductives

Biostatistiques

A black and white photograph of a snowy mountain range. The foreground is a vast, flat expanse of snow. In the middle ground, there are several jagged, snow-covered mountain peaks. The sky is a uniform, light gray. A white rectangular text box is overlaid on the left side of the image, containing the text "Tests d'Hypothèse" in a bold, black, sans-serif font.

Tests d'Hypothèse

Tests de comparaison

Objectif : tirer des conclusions à partir d'observations/comparaisons de paramètres différents entre 2 populations différents par constitution de 2 échantillons et recherche de différence significative du paramètre

Deux hypothèses

$H_0 =$ *Hypothèse nulle* : Pas de différence entre deux populations pour le paramètre étudié (différence = hasard)

$H_1 =$ *Hypothèse alternative* : différence significative pour le paramètre étudié (différence \neq hasard)

Le test permet de savoir si on accepte ou rejette H_0 en fonction également du risque α (fixé en amont)

Les Etapes du Test

- **Définir H0 et H1**
- **Définir le bon test** (*f° de quantitatif ou qualitatif avec Z le résultat*)
- **Choisir le risque α** (*le + souvent 5%*)
- **Recueil des données**
- **Calcul de Z**
- **Utilisation de la règle** (*accepte ou rejette f° de H0 et α selon la position de Z sur le modèle théorique de distribution normale = gaussienne*)
- **Détermination du risque réel à postériori** (*=degré de signification de p*)
- **Interpréter les résultats** (*aussi niveau de la population*)

Le Risque

α = risque de 1^{er} espèce (5%) : **rejeter H0** alors que H0 est **vrai**

β = risque de 2^{ème} espèce (20%) : **accepter H0** alors que H0 est **faux**

$1-\beta$ = puissance du test (avoir juste) : **rejeter H0** si H0 est **faux**

On privilégie de maîtriser α quitte à ignorer β

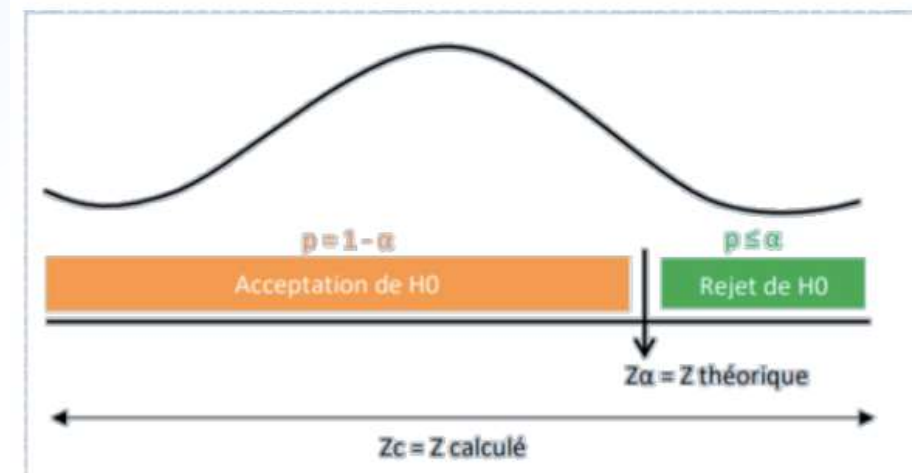
La règle de rejet du test est définie uniquement à partir de α et **H0**.

Entre 2 alternatives, on choisira pour **H0**
l'hypothèse qu'il serait **le plus grave de rejeter**
à tort.

		Décision du statisticien	
		Rejet H0	Non rejet H0
Réalité	H0 vraie	α	$1-\alpha$
	H1 vraie	$1-\beta$	β

Interprétation graphique de α

On cherche Z_t qui est donné par l'écart réduit et on examine la position de Z_c sur le modèle théorique probabiliste (*normal=gaussien*)

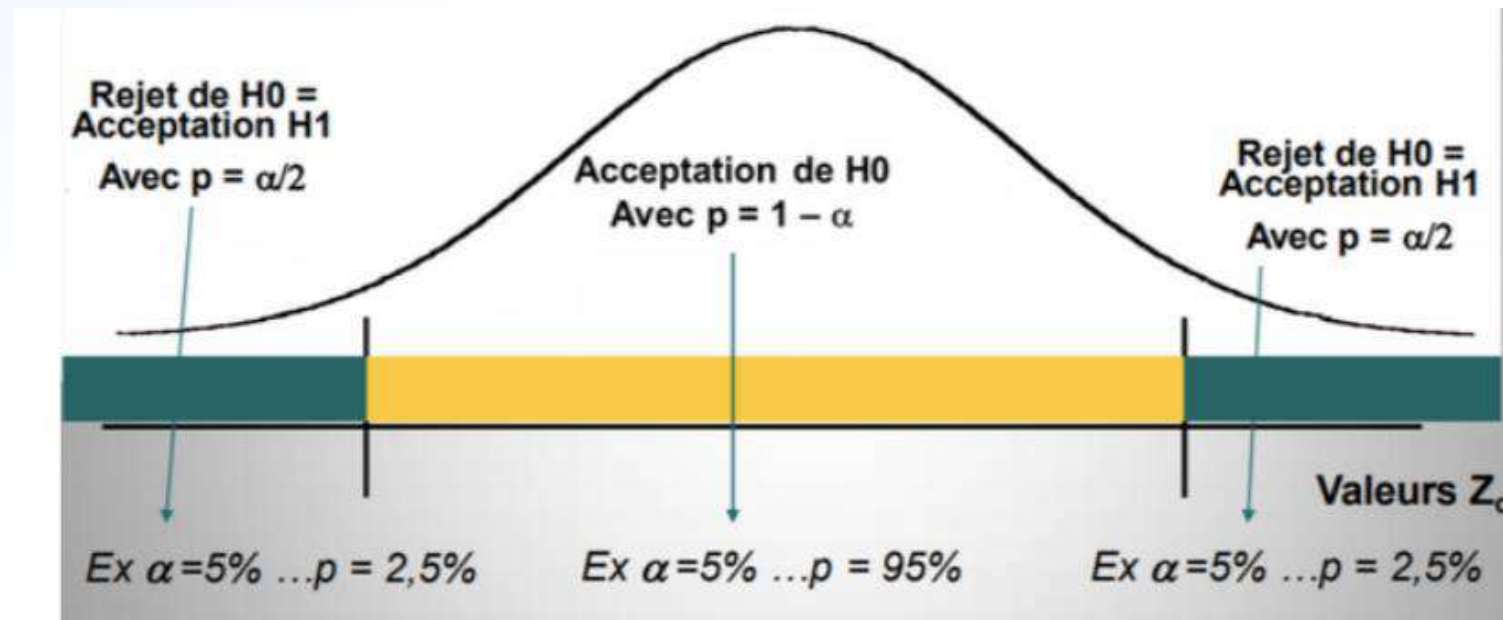


Acceptation de H0	Rejet de H0
$Z_c < Z_t$	$Z_c > Z_t$
$p = 1 - \alpha$	$p \leq \alpha$

p = le risque réel à postériori = le degré de signification du test = quantifie la conviction avec laquelle on rejette H0

2 Situation si H1

- Situation Unilatérale : la conclusion est absence ou présence d'une différence
- Situation Bilatérale : (dans le cas d'une différence) la conclusion permet de dire qu'une condition est meilleur que l'autre





Liaison entre 2 variables QUALITATIVES

Test de Comparaison de Pourcentages

On cherche Z qui est donné par l'écart réduit ε .

Avec: H_0 = Pas de différences significatives

H_1 = Différences significatives

On compare alors:

- ε_t = théorique : donné par la table d'écart réduit en fonction de α

- ε_c = calculé :
$$\frac{p_A - p_B}{\sqrt{\frac{p_A \cdot q_A}{n_A} + \frac{p_B \cdot q_B}{n_B}}}$$
 avec $q = 1 - p$

SI $\varepsilon_c > \varepsilon_t$ alors rejet de H_0

Si on cherche ε_t avec $\alpha = 5\%$

Table de l'écart réduit

		α								
		0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
α	∞	2,576	2,326	2,17	2,054	1,96	1,881	1,812	1,751	1,695
	0,1	1,645	1,598	1,555	1,514	1,44	1,405	1,372	1,341	1,311
	0,2	1,282	1,254	1,227	1,2	1,15	1,126	1,103	1,08	1,058
	0,3	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,86
	0,4	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,69
	0,5	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,539
	0,6	0,524	0,51	0,496	0,482	0,468	0,454	0,44	0,426	0,399
	0,7	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,266
	0,8	0,253	0,24	0,228	0,215	0,202	0,189	0,176	0,164	0,138
	0,9	0,126	0,113	0,1	0,088	0,075	0,063	0,05	0,038	0,013

Teste du χ^2

On cherche Z qui est donné par χ^2

Avec: H_0 = pas de différence dans la répartition

H_1 = différences dans la répartition

On compare alors:

- χ^2_t = théorique : donné la table du (Chi) χ^2 en fonction de α et du nombre de ddl

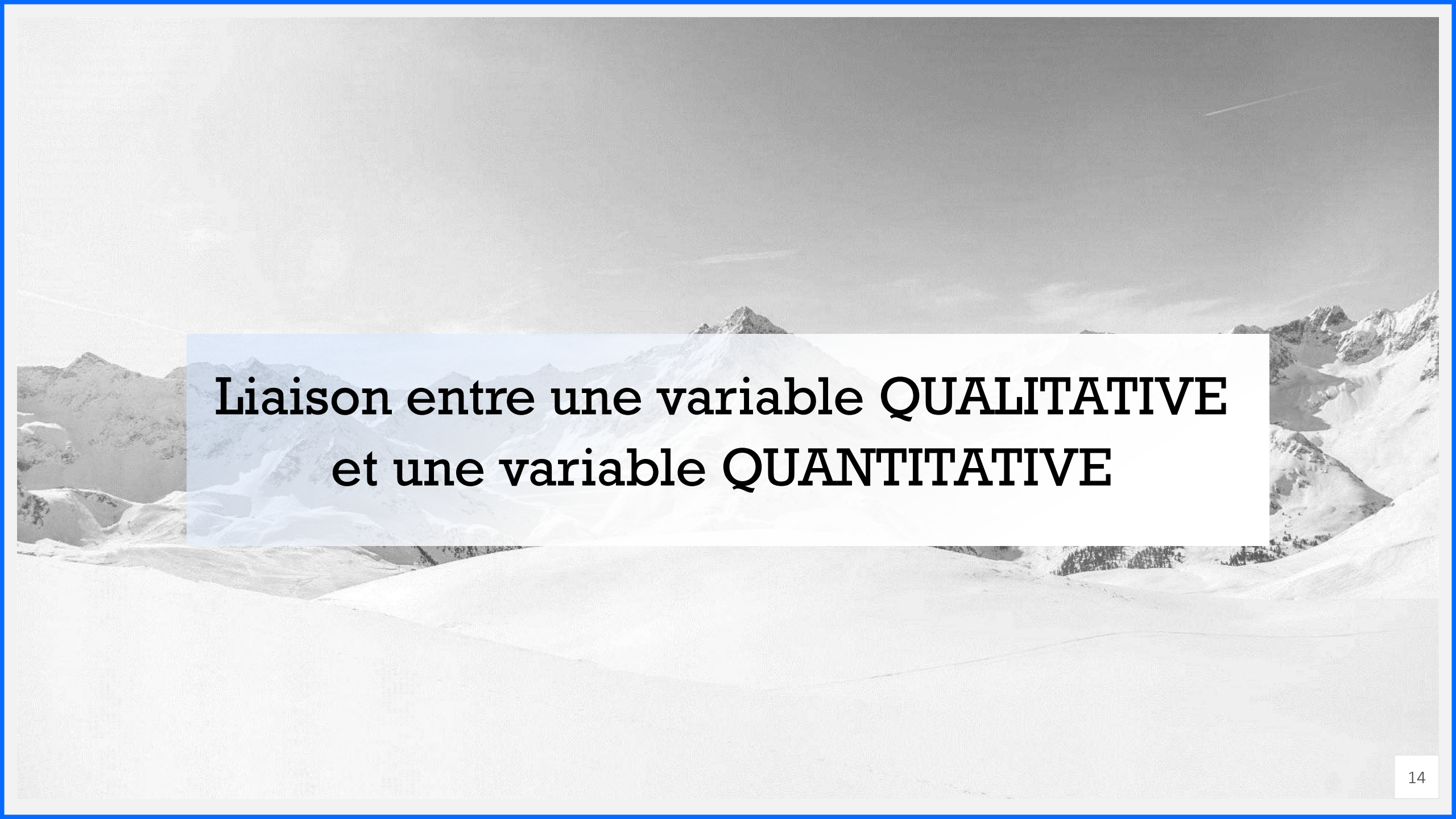
$$DDL = (\text{nombre de lignes} - 1) * (\text{nombre de colonnes} - 1)$$

- χ^2_c = calculé : $\frac{\sum(o_i - c_i)^2}{c_i}$ avec o_i = données observées et c_i = données calculées

SI $\chi^2_c > \chi^2_t$ alors rejet de H_0

Si on cherche X^2 dans sa table avec $\alpha = 5\%$ et $ddl = 1$

ddl	α								
	0,9	0,5	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,016	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,21	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
5	1,61	4,351	6,064	7,289	9,236	11,07	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,49	7,344	9,524	11,03	13,362	15,507	18,168	20,09	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	6,304	11,34	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	7,042	12,34	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,79	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32	39,252
17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,79

A grayscale photograph of a snowy mountain landscape. The foreground is a vast, flat expanse of snow. In the middle ground, there are rolling hills and valleys covered in snow. In the background, several sharp mountain peaks are visible, some with patches of snow. The sky is a uniform, light gray. A semi-transparent white rectangular box is centered over the image, containing the text.

**Liaison entre une variable QUALITATIVE
et une variable QUANTITATIVE**

Comparaison de moyennes (*grands échantillons = n_1 et $n_2 > 30$*)

On utilise **la table d'écart réduit** ($\varepsilon > 1,96$ et $\alpha < 5\%$)

Avec: H_0 = les moyennes ne sont pas différentes

H_1 = les moyennes son différentes

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

SI $\varepsilon_c > \varepsilon_t$ alors rejet de H_0

Test de Student (petits échantillons = n_1 et $n_2 < 30$)

On utilise la **table t de Student** avec **$(n_1-1) + (n_2-1)$** ddl

Avec: H_0 = les moyennes ne sont pas différentes

H_1 = les moyennes son différentes

$$\text{Avec } t = \frac{m_1 - m_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \quad \text{où } s = \sqrt{\frac{\sum(x_i - m_1)^2 + \sum(x_j - m_2)^2}{(n_1 - 1) + (n_2 - 1)}}$$

SI $t_c > t_t$ alors **rejet de H_0**

Table de Student

α d.d.l.	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	1,000	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,816	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,765	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,741	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,727	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,718	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,711	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,706	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,703	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,700	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,697	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,695	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,694	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,692	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,691	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,690	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,689	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,688	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,688	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,687	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,686	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,686	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,685	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,685	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,684	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,684	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,684	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,683	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,683	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,683	1,055	1,310	1,697	2,042	2,457	2,750	3,646
∞	0,126	0,674	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Séries appariées = méthode des couples

On l'utilise si les deux échantillons sont **non indépendants**

- Si **n > 30** : on utilise la **comparaison de moyennes**

$$\text{Avec } \varepsilon = \frac{m_d}{\sqrt{\frac{s^2}{n}}}$$

- Si **n < 30** : on utilise le **test de Student**

$$\text{Avec } t = \frac{m_d}{\sqrt{\frac{s^2}{n}}}$$

Calcul différent mais l'application est identique aux séries non appariées



Liaison entre 2 variables QUANTITATIVES

Corrélation et régression

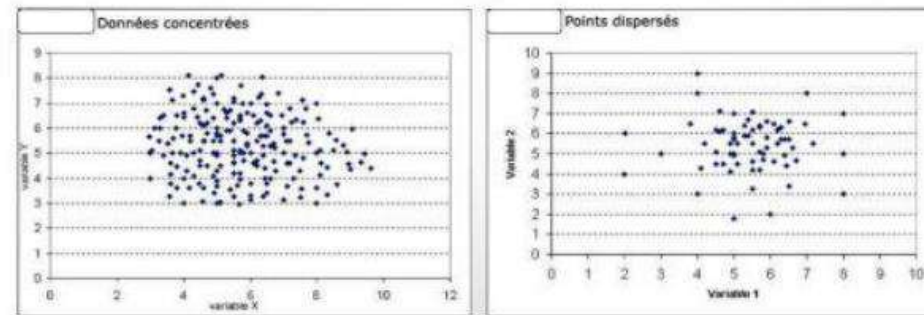
Corrélation : évaluation de la liaison entre 2 variables quantitatives

Régression : méthode mathématique expliquant les relations entre variables observées

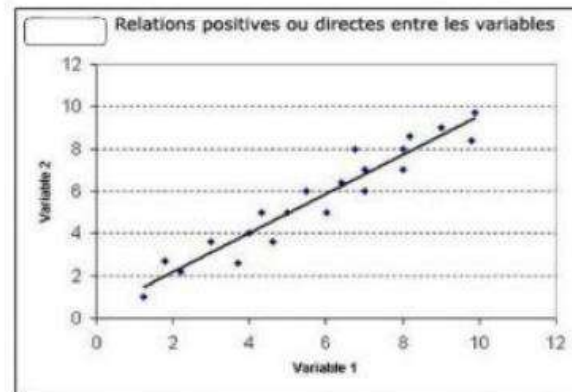
Représentation des données

Droite de régression = Droite des moindres carrés : passe au plus près de chaque point du nuage
permet la prédiction de la valeur d'une variable en connaissant l'autre

Nuage de points :



Droite de régression : permet de visualiser si une des 2 variables est **dépendante** de l'autre



Test du coefficient de corrélation = pente de la courbe

On utilise le **tableau du coefficient de corrélation** avec **ddl = n-2**

Avec: H0 = pas de lien

H1 = il y a un lien

$$\text{Avec } r = \frac{\sum(x_i - m_x)(y_i - m_y)}{\sqrt{\sum(x_i - m_x)^2 \sum(y_i - m_y)^2}} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n})(\sum y^2 - \frac{(\sum y)^2}{n})}}$$

SI $r > 0$ alors **liaison positive** (x et y varient dans le **même sens**)

SI $r < 0$ alors **liaison négative** (x et y varient dans le **sens inverse**)

Avec r TOUJOURS inférieur à 1 !!

Soit $r_c > r_t$ alors **rejet de H0**

Table des Coefficients de Corrélation

d.d.l. \ α	0,10	0,05	0,02	0,01
1	0,9877	0,9969	0,9995	0,9999
2	0,9000	0,9500	0,9800	0,9900
3	0,8054	0,8783	0,9343	0,9587
4	0,7293	0,8114	0,8822	0,9172
5	0,6694	0,7545	0,8329	0,8745
6	0,6215	0,7067	0,7887	0,8343
7	0,5822	0,6664	0,7498	0,7977
8	0,5494	0,6319	0,7155	0,7646
9	0,5214	0,6021	0,6851	0,7348
10	0,4973	0,5760	0,6581	0,7079
11	0,4762	0,5529	0,6339	0,6835
12	0,4575	0,5324	0,6120	0,6614
13	0,4409	0,5139	0,5923	0,6411
14	0,4259	0,4973	0,5742	0,6226
15	0,4124	0,4821	0,5577	0,6055
16	0,4000	0,4683	0,5425	0,5897
17	0,3887	0,4555	0,5285	0,5751
18	0,3783	0,4438	0,5155	0,5614
19	0,3687	0,4329	0,5034	0,5487
20	0,3598	0,4227	0,4921	0,5368
25	0,3233	0,3809	0,4451	0,4869
30	0,2960	0,3494	0,4093	0,4487
35	0,2746	0,3246	0,3810	0,4182
40	0,2573	0,3044	0,3578	0,3932
45	0,2428	0,2875	0,3384	0,3721

Tests Non Paramétrique

Utilisation

Utilisé en cas d'effectif trop faible soit $4 < n < 12$ avec un caractère quantitatif

Test avec forte robustesse car transforme données quantitatives en mesures ordinaires (=rangs)

Test de U-MANN & WHITNEY

On l'utilise la table U de MANN & WHITNEY

Avec: H_0 = les moyennes des deux groupes sont proches ($\mu_1 = \mu_2$)

H_1 = les moyennes ne sont pas proches

- On range les valeurs de E1 et E2 par ordre croissant
- Somme les valeurs du membre opposé

SI $U_c > U_t$ alors accepte de H_0

Exemple

G1: 1,5,10,23,34,67

G2: 7,13,56,134,235

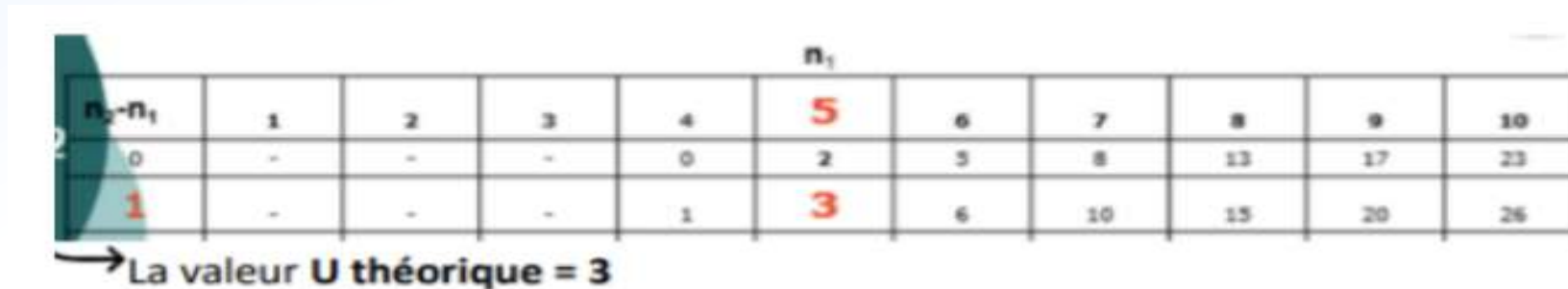
Nbr	1	5	7	10	13	23	34	56	67	134	235
Grp	A	A	B	A	B	A	A	B	A	B	B
U	0	0	2	1	3	2	2	5	3	6	6

$$U_{ab} = 2+3+5+6+6 = 22$$

$$U_{ba} = 0+0+1+2+2+3 = 8$$

RQ: $U_{BA} + U_{AB} = n_A \times n_B = 6 \times 5$ ($5 = n_1$ car le plus petit des deux n) $= 30$

On prend U le plus petit et le compare à U_t



Soit $U_c > U_t$ donc H_0 accepter

Test du coefficient r' de SPEARMAN

On l'utilise le tableau du r de SPERMAN

Avec: H_0 = pas de lien entre les variables

H_1 = lien entre les variable

$$\text{Avec } r' = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

SI $r'_c > r'_t$ alors **rejet de H_0**

Test du coefficient r' de SPEARMAN

On cherche r' fonction de α et n

$n \backslash \alpha$	0.2	0.1	0.05	0.02	0.01	0.002
4	1.000	1.000	—	—	—	—
5	0.800	0.900	1.000	1.000	—	—
6	0.657	0.829	0.886	0.943	1.000	—
7	0.571	0.714	0.786	0.893	0.929	1.000
8	0.524	0.643	0.738	0.833	0.881	0.952
9	0.483	0.600	0.700	0.783	0.833	0.917
10	0.455	0.564	0.648	0.745	0.794	0.879
11	0.427	0.536	0.618	0.709	0.755	0.845
12	0.406	0.503	0.587	0.678	0.727	0.818
13	0.385	0.484	0.560	0.648	0.703	0.791
14	0.367	0.464	0.538	0.626	0.679	0.771
15	0.354	0.446	0.521	0.604	0.654	0.750
16	0.341	0.429	0.503	0.582	0.635	0.729
17	0.328	0.414	0.488	0.566	0.618	0.711

Récapitulatif:

	Test	
	Paramétriques	Non paramétrique
Comparaison de 2 échantillons indépendants	<ul style="list-style-type: none">♣ Test t de Student♣ Test de comparaison de moyennes	<ul style="list-style-type: none">♣ Test de Mann-Whitney
Comparaison de 2 échantillons appariés	<ul style="list-style-type: none">♣ Test t de Student pour séries appariées♣ Test de comparaison de moyennes pour séries appariées	<ul style="list-style-type: none">♣ Test de Wilcoxon
Test de corrélation	<ul style="list-style-type: none">♣ Test du coefficient r	<ul style="list-style-type: none">♣ Test du coefficient r' de Spearman

Effectif	Données quantitatives	Données qualitatives	Données qualitatives et quantitatives
$n \geq 30$	<ul style="list-style-type: none"> • Coefficient de corrélation r • r' de Spearman 	<ul style="list-style-type: none"> • Comparaison de pourcentages • Chi-2 	<ul style="list-style-type: none"> • Comparaison de moyennes • T de Student • U Mann et Whitney
$30 > n \geq 12$	<ul style="list-style-type: none"> • Coefficient de corrélation r • r' de Spearman 	<ul style="list-style-type: none"> • Comparaison de pourcentages • Chi-2 	<ul style="list-style-type: none"> • T de Student • U Mann et Whitney
$12 > n > 4$	<ul style="list-style-type: none"> • r' de Spearman 	<ul style="list-style-type: none"> • Comparaison de pourcentages • Chi-2 	<ul style="list-style-type: none"> • U Mann et Whitney



**Je vous
remercie pour
votre attention**