

TESTS NON PARAMÉTRIQUES

1

- Utilisation obligatoire si les effectifs sont trop faibles (< 5), avec des **caractères quantitatifs**. Dans ce cas, les populations ne se distribuent pas normalement.
- Présentent une excellente robustesse
- 1) LIAISON QUANTITATIFS / QUALITATIFS
 - **U de Mann & Whitney**
- 2) LIAISON ENTRE QUANTITATIFS
 - **Spearman (= Corrélation)**

2

Le **test de Wilcoxon-Mann-Whitney** (ou **test U de Mann-Whitney** ou encore **test de la somme des rangs de Wilcoxon**) est un test statistique non paramétrique qui permet de tester l'hypothèse selon laquelle les médianes de chacun de deux groupes de données sont proches.

2 échantillons indépendants E_1 et E_2 de taille n_1 et n_2

On souhaite tester l'hypothèse H_0 disant que les moyennes expérimentales dans les 2 échantillons sont égales $\mu_1 = \mu_2$

On trie les valeurs obtenues dans la réunion des 2 échantillons par ordre croissant

Pour chaque valeur x_i issue de E_1 , on compte le nombre de valeurs issues de E_2 situées après lui dans la liste ordonnée (celles qui sont égales à x_i ne comptent que pour $1/2$)

On note u_1 la somme des nombres ainsi associés aux différentes valeurs issues de E_1 .
On fait de même en échangeant les rôles des deux échantillons, ce qui donne la somme u_2 . Soit u la plus petite des deux sommes obtenues : $u = \min\{u_1 ; u_2\}$.



3

On note U la variable aléatoire associée. • Pour n_1 et n_2 quelconques, on lit dans les tables du test de Mann et Whitney le nombre m_α tel que, sous (H_0) , $P(U \leq m_\alpha) = \alpha$.

Pour n_1 et n_2 quelconques, on lit dans les tables du test de Mann et Whitney le nombre m_α tel que, sous (H_0) , $P(U \leq m_\alpha) = \alpha$.

On rejette (H_0) au risque d'erreur α si $u \leq m_\alpha$. Autrement on accepte (H_0) .

Si n_1 et n_2 sont assez grands (≥ 20 en général), sous (H_0) , U suit approximativement la loi normale $N(\mu, \sigma)$



4

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

On a recensé pour 6 étudiants les notes obtenues au concours PAES en Biostatistique, et le classement final à ce même concours.

On cherche à établir si il existe une relation entre cette note et le classement final.

X Biostat	Y Classement
12,4	210
4,9	555
18,1	6
5,4	445
19,4	5
16	14





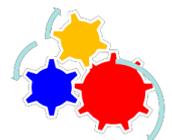
5

H0 : Il n'y a pas de lien entre ces 2 séries de valeurs numériques. Il s'agit de 2 séries indépendantes.

Données quantitatives : Coeff de corrélation, faibles effectifs → r' Spearman

X Biostat	Rg X	Y Classe ment	Rg Y	d _i rang	(d _i) ²
12,4	3	210	4	-1	1
4,9	1	555	6	-5	25
18,1	5	6	2	3	9
5,4	2	445	5	-3	9
19,4	6	5	1	5	25
16	4	14	3	1	1

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



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

$$r' = 1 - \frac{6 \times 70}{6 \times 35} = -1$$

Dans la table théorique, avec $n=6$: $r' = 0,89$ avec $\alpha = 5\%$
 $r' = 1$ avec $\alpha = 1\%$

Le r' calculé à $-1 \leq r'$ théorique lu dans la table.

On repousse donc H_0 ($p < 1\%$)

On met en évidence un lien très significatif entre ces 2 séries.

Il s'agit de 2 séries corrélées. Plus la note de Biostat est élevée, plus petit est le rang de classement (d'où le signe $-$ pour r').

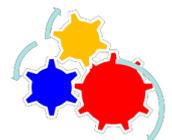


Table r' de Spearman

7

α

n	0.05	0.01
5	1.00	-
6	0.89	1.00
7	0.79	0.93
8	0.74	0.88
9	0.68	0.83
10	0.65	0.79

Effectif	Données Quantitatives	Données Qualitatives	Données Qualitatives - Quantitatives
>4 & <12	r' de Spearman	Comp % ou χ^2	U Mann & Withney
>=12 & < 30	Coeff de corrélation r	Comp % ou χ^2	t Student
>= 30	Coeff de corrélation r	Comp % ou χ^2	Comp moyennes

EXERCICES ET QCM ENTRAINEMENT

Remarques :

- 1) Certains exercices sont différents dans leur forme de ce que sera l'épreuve du concours
- 2) Les buts sont de vous faire réfléchir, de réviser le cours, et d'appliquer vos connaissances

Dans une ville de 10000 habitants, on décide de réaliser une étude. Pour cela on réalise un échantillon de cette population en choisissant, à partir des listes électorales, les 200 premières personnes par ordre alphabétique. Quelles sont les propositions exactes?

- A) Il s'agit d'un échantillon représentatif de la population.
- B) Il ne s'agit pas d'un échantillon représentatif de la population.
- C) Les noms de famille sont répartis aléatoirement dans la population
- D) Prendre une personne sur 50 est équivalent à réaliser un TAS.
- E) 200 Personnes est un trop petit échantillon pour être représentatif.

Réponse **B** : Le seul moyen d'obtenir un **échantillon représentatif est le TAS**. Tous les autres items sont faux : les noms de famille **ne sont pas répartis aléatoirement** (consonance étrangère), 1 pers/50 est certes mieux que 200 successifs, **mais ce n'est pas un TAS**, et il n'y a pas de taille mini pour un échantillon représentatif, **seule la méthode importe**.



En vue de comparer deux traitements T1 et T2, 250 patients sont répartis entre ces 2 traitements par T.A.S. Quel est le meilleur traitement ?

Etat du malade après 5 jours de traitement

Traitement	Stationnaire	Amélioré	Guéri	Total
T1	15	70	35	120
T2	25	85	20	130

- A) Le test à effectuer est un test du χ^2
- B) Le test à effectuer est un test de corrélation
- C) L'hypothèse nulle testée est que les nombres de patients stationnaires, améliorés, guéris sont identiques avec les 2 traitements
- D) L'hypothèse nulle testée est que les nombres de patients stationnaires, améliorés, guéris sont différents avec les 2 traitements
- E) Si le test statistique effectué est significatif, cela montre que les 2 traitements ont globalement des efficacités différentes



L'entraînement sportif intensif est réputé être à l'origine de perturbations hormonales, surtout chez la femme. On étudie ces perturbations sur le cycle menstruel.

Soit 3 groupes de femmes : Gr 1 = femmes non sportives, Gr 2 = femmes faisant du jogging loisir, Gr 3 = femmes courant le marathon

Troubles du cycle

Groupe	Oui	Non
1	14	40
2	9	14
3	46	42

Existe-t-il un retentissement hormonal en fonction de la pratique sportive?



L'entraînement sportif = variable qualitative nominale à 3 modalités

Troubles cycle = variable qualitative binaire à 2 modalités

H0 = La pratique sportive n'a pas d'influence sur le cycle menstruel :

les troubles du cycle sont les mêmes dans les 3 groupes.

H1 = La pratique sportive a une influence sur le cycle menstruel :

les troubles du cycle sont significativement différents dans les 3 groupes.

Test de χ^2 , avec $(3-1)(2-1) = 2$ ddl.

$$\chi^2 = \sum \frac{(O_i - C_i)^2}{C_i}$$



Troubles du cycle

Groupe	Oui	Non	Total
1	14 (22,58)	40 (31,42)	54
2	9 (9,62)	14 (13,38)	23
3	46 (36,80)	42 (51,20)	88
Total	69	96	165

Rappel :

$$69/165 \times 54 = 22,58$$

$$69/165 \times 23 = 9,62, \text{ etc..}$$

$$\chi^2 = \sum \frac{(O_i - C_i)^2}{C_i}$$

$$\chi^2 \text{ calculé} = 9,63 \quad \text{ddl} = 2$$

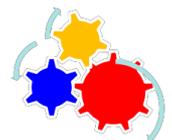
$$\chi^2 \text{ théorique} = 5,99 \quad \alpha = 5\%$$

χ^2 calculé > χ^2 théorique : **on rejette H0**

Il existe un retentissement hormonal statistiquement significatif en fonction de la pratique sportive ($\alpha < 5\%$).

Troubles cycle : Gr1 14/54=26% Gr2 9/23=39% Gr3 46/88=52%

Plus la pratique sportive est importante, plus les troubles apparaissent.



15

On teste un traitement T modifiant la glycémie. Deux groupes de 49 patients T.A.S : les patients du premier groupe (groupe T) sont traités par T ; ceux de l'autre groupe sont traités par placebo (groupe P). Au bout d'une semaine :

Les moyennes et écart type calculés dans les deux groupes sont : $m_T = 6,1$ mmol/ml, $m_P = 5,6$ mmol/ml, $s_T = 0,2$, $s_P = 0,3$. Le paramètre calculé avec le test adapté est 3,8
On donne un extrait de la table de l'écart réduit

α	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
ε	2,576	2,326	2,17	2,054	1,96	1,881	1,812	1,751	1,695

- 1.L'hypothèse nulle testée est que les moyennes calculées m_T et m_P sont identiques
- 2.Le test utilisé montre que m_T et m_P diffèrent significativement
- 3.Le degré de signification est $<1\%$
- 4.Le degré de signification est compris entre 5 % et 1 %
- 5.Le test est valide car les tailles des groupes sont suffisantes



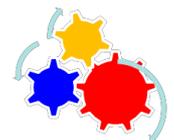
16

On veut savoir si une nouvelle molécule présente un effet anti dépresseur. Pour cela, on organise un essai portant sur 20 malades dépressifs, répartis en 2 groupes.

Les 20 malades sont répartis par TAS en 2 groupes de 10 sujets, l'un recevant la nouvelle molécule, l'autre recevant un placebo.

On évalue les patients à l'aide d'une échelle numérique de 0 (non déprimé) à 50 (très déprimé). Le groupe témoin reçoit le placebo.

Les patients des 2 groupes sont évalués avant traitement, puis après traitement au bout de 28 jours.



Témoins		Traités	
J0	J28	J0	J28
34	31	27	22
30	28	32	25
25	26	30	23
27	25	28	26
31	24	25	20
24	25	33	27
28	26	29	21
30	27	31	26
35	32	32	25
26	25	29	23

Y a-t-il un effet placebo ?

On compare les scores (J0 et J28) Témoins

On compare les scores (J0 et J28) Témoins / Traités

On compare les scores J0 Témoins / Traités

On compare les scores J28 Témoins / Traités

H0 = le placebo n'a aucun effet (les scores J0 ne diffèrent pas des scores J28)

Comparaison de moyennes (n=10) donc

Test t student séries appariées ou Mann et Whitney



Groupe Témoins

J0	J28	Diff
34	31	3
30	28	2
25	26	-1
27	25	2
31	24	7
24	25	-1
28	26	2
30	27	3
35	32	3
26	25	1

$$m_d = 21/10 = 2,1$$

$$S^2_d = 5,21$$

$$t = \frac{2,1}{\sqrt{\frac{5,21}{10}}} = 2,91$$

t théorique (avec $10-1 = 9$ ddl et $\alpha = 5\%$) = **2,26**

t calculé > t théorique Rejet de H0

Le placebo a un effet significatif



2. On évalue les patients à l'aide d'une échelle numérique de 0 (non déprimé) à 50 (très déprimé). Le groupe témoin reçoit le placebo.

Les patients des 2 groupes sont évalués avant traitement, puis après traitement au bout de 28 jours.

19

Témoins		Traités	
J0	J28	J0	J28
34	31	27	22
30	28	32	25
25	26	30	23
27	25	28	26
31	24	25	20
24	25	33	27
28	26	29	21
30	27	31	26
35	32	32	25
26	25	29	23

Le traitement est il efficace ?

Pour chaque patient :

On compare les scores (J0 et J28) Témoins

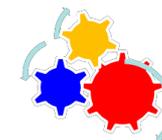
On compare les scores (J0 et J28) Témoins/Traités

On compare les scores J0 Témoins / Traités

On compare les scores J28 Témoins / Traités

Efficacité du traitement différent de Effet traitement !

Comparer les différences ($J_{28} - J_0$) de chaque patient, entre les 2 groupes



Le traitement est il efficace ?

Comparer les différences ($J_{28} - J_0$) de chaque patient, entre les 2 groupes.

Questions préliminaires :

Quelles sont les variables? Quel est le bon test ?

Variables qualitatives / variables quantitatives (Tt ou Placebo / Scores)

2 groupes indépendants de faible effectif ($n=10$).

Test t de Student ou Mann et Whitney.



Mann & Whitney :

21

Témoins $d=J_0-J_{28}$	3	2	-1	2	7	-1	2	3	3	1
Traités $d=J_0-J_{28}$	5	7	7	2	5	6	8	5	7	6

Valeurs de différences rangées par ordre croissant : ex aequo, classés ensemble.

Témoins
Traités

Ex aequo : rangs
1 et 2 donc ->
rang $3/2=1,5$

Ex aequo : rangs
4,5,6,7 donc ->
rang $22/4=5,5$

Différences	-1	-1	1	2	2	2	2	3	3	3
Rangs	1,5	1,5	3	5,5	5,5	5,5	5,5	9	9	9

Différences	5	5	5	6	6	7	7	7	7	8
Rangs	12	12	12	14,5	14,5	17,5	17,5	17,5	17,5	20



-1	-1	1	2	2	2	2	3	3	3
1,5	1,5	3	5,5	5,5	5,5	5,5	9	9	9

Témoins
Traités

5	5	5	6	6	7	7	7	7	8
12	12	12	14,5	14,5	17,5	17,5	17,5	17,5	20

Calcul des paramètres U

Soit U_1 : pour chaque témoin, cumul des **traités** classés AVANT

$$U_1 = 9$$

$$U_2 = 100 - 9 = 91 \text{ car on calcule } n_1 \times n_2 = 100.$$

On compare donc $U=9$ (le plus petit des 2) avec la table théorique

($\alpha=5\%$, $n_1=10$, $n_2=10$) U théorique = 23 : U calculé < U théorique :

Peu d'imbrication :

Rejet de H_0 : les différences sont significativement plus importantes pour le traitement que pour le placebo, avec $\alpha < 5\%$



Table U de Mann Whitney ($\alpha = 5\%$)

n_1 est le plus petit des 2 effectifs, U le plus petit des 2 U calculés

n_1

$n_2 - n_1$	1	2	3	4	5	6	7	8	9	10
0	-	-	-	0	2	5	8	13	17	23
1	-	-	-	1	3	6	10	15	20	26
2	-	-	0	2	5	8	12	17	23	29
3	-	-	0	3	6	10	14	19	26	33
4	-	-	1	4	7	11	16	22	28	36
5	-	-	2	4	8	13	18	24	31	39
6	-	0	2	5	9	14	20	26	34	42
7	-	0	3	6	11	16	22	29	37	45
8	-	0	3	7	12	17	24	31	39	48
9	-	0	4	8	13	19	26	34	42	52
10	-	1	4	9	14	21	28	36	45	55
11	-	1	5	10	15	22	30	38	48	
12	-	1	5	11	17	24	32	41	50	
13	-	1	6	11	18	25	34	43		
14	-	1	6	12	19	27	36	45		
...										
18	-	2	8	16	24	33				
19	-	3	9	17	25					
20	-	3	9	17	27					

On veut étudier si le fait de fumer influence l'apparition d'un cancer du poumon?

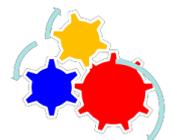
- 1) Il s'agit de lier 2 caractères qualitatifs
- 2) Il s'agit de lier 2 caractères quantitatifs
- 3) Il s'agit de lier 1 caractère quantitatif et 1 caractère qualitatif
- 4) Il faut constituer 2 groupes : « cancer » et « non cancer »
- 5) Un groupe « malades » suffira.
- 6) Il faut échantillonner
- 7) On ne dispose pas de données suffisantes

Réponse 1,4,6 : Variables qualitatives (fumer ou non et cancer ou non), 4 catégories (malades/non malades et fumeur/pas fumeur), TAS obligatoire afin d'échantillonner.

Test à faire pour répondre à la question :

Comparaison de % ou test du χ^2

	K	Non K
F	% n° 1	
Non F	% n° 2	



Pour estimer une moyenne dans une population quelconque, il faut :

- 1) T.A.S un échantillon à partir de cette population
- 2) Effectuer le calcul au niveau de la population
- 3) Disposer d'un petit effectif : écart type plus grand : meilleure précision
- 4) Disposer d'un grand effectif : écart type plus petit : meilleure précision
- 5) Le but de l'estimation est de fournir une valeur précise
- 6) Si T.A.S : le risque de première espèce est proche de 0
- 7) On ne dispose pas de données suffisantes

Réponse 1,4 : Pour estimer une moyenne il faut disposer d'un échantillon représentatif de la population par TAS, et cet échantillon doit avoir une taille suffisamment importante pour augmenter la précision

Les items 5 et 6 sont faux.



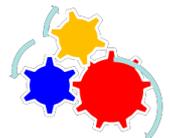
On cherche à établir une relation entre l'âge d'apparition du cancer du sein et l'âge de la puberté chez la femme. On interroge des femmes venues consulter en dépistage : on note âge, et âge de puberté. 100 femmes viennent consulter.

1- 30 femmes présentent un cancer du sein. Y a-t-il une différence entre l'âge de leur puberté et l'âge de puberté des 70 autres femmes? Quelles sont les propositions qui caractérisent ce problème ?

- a) Lien entre données qualitatives
- b) Lien entre données quantitatives
- c) Lien entre données qualitatives et quantitatives
- d) Test de comparaison de %
- e) Test de comparaison de moyennes
- f) Test de Student
- g) Coefficient de corrélation

Réponse c, e:

Données qualitatives (cancer ou non) /
Données quantitatives (Age puberté).
Test de comparaison de moyennes
(n_1 et $n_2 > 30$)



On cherche à établir une relation entre l'âge d'apparition du cancer du sein et l'âge de la puberté chez la femme. On interroge des femmes venues consulter en dépistage : on note âge, et âge de puberté. 100 femmes viennent consulter.

2- On cherche maintenant s'il existe un lien entre l'âge d'apparition du cancer et l'âge de la puberté. Quelles sont les propositions qui caractérisent ce problème ?

- a) Lien entre données qualitatives
- b) Lien entre données quantitatives
- c) Lien entre données qualitatives et quantitatives
- d) Test de comparaison de %
- e) Test de comparaison de moyennes
- f) Test de Student
- g) Coefficient de corrélation

Réponse b,g:

Données quantitatives (âge cancer) /
Données quantitatives (Age puberté)

Recherche de corrélation entre ces
variables.



28

Est il possible de déduire de cette étude le pourcentage de cancer du sein au niveau de la population féminine générale française?

A/ Oui, car sur 100 femmes, 30 diagnostics de cancer donc % cherché = 30%

B/ Non, il n'y a pas eu TAS des femmes examinées.

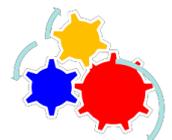
C/ Non, il manque des données.

D/ Oui, car mise en évidence d'un lien entre âge de la puberté et l'âge de l'apparition du cancer.

E/ Non, on ne connaît pas l'âge moyen de la puberté chez les femmes en France.

Réponse B:

Non seulement pas TAS, mais : dépistage.. Viennent se faire dépister des femmes avec antécédents familiaux surtout ..Biais de recrutement.



29

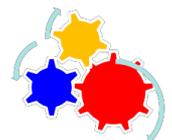
Dans un service hospitalier de 22 patients, en 1 mois on diagnostique 5 infections.

Habituellement, le nb de cas d'infections est ≤ 3 . Quelles sont les propositions qui caractérisent ce problème ?

- a) Lien entre données qualitatives
- b) Lien entre données quantitatives
- c) Test de comparaison de %
- d) Test de comparaison de moyennes
- e) Si on accepte H_0 cela signifie que seul le hasard est responsable de ces chiffres.
- f) Si on repousse H_0 cela signifie qu'il y a un problème d'hygiène dans ce service.

Réponse a, c, e, f:

Données qualitatives (mois à 5 infec / autres mois). Comparaison de %. Les items e et f sont justes tous les 2. Définitions et interprétation de H_0 .



30

Une étude nationale dans des maternités TAS, a permis de mettre en évidence les résultats suivants, concernant les poids des bébés et l'âge des mères.

	Mères < 32 ans	Mères > 32 ans
Poids moyen à la naissance	3410 g	3197 g
Ecart type	505 g	458 g
Nb de bébés	254	294

- a) Lien entre données qualitatives b) Lien entre données quantitatives
 c) Test du coefficient de corrélation d) Test de comparaison de moyennes
 e) Quel que soit le résultat du test on pourra conclure statistiquement et médicalement.
 f) Etude mal menée ou mal présentée.

Réponse d, e:

Données quantitatives (poids) et qualitatives (âge). n_1 et $n_2 > 30$.
 Comparaison de moyennes
 TAS donc résultat (H_0 ou H_1) généralisable en France.



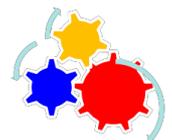
31

On désire tester à l'hôpital un nouveau protocole de radiothérapie (N) permettant de mieux réduire les masses cancéreuses de certains cancers.

- a) On attribue ce nouveau protocole systématiquement à tous les patients pendant une période donnée.
- b) TAS des patients pour N et pour l'ancien (A)
- c) On crée un groupe patients cancer et un groupe patients non cancer
- d) Le patient est son propre témoin : méthode des couples
- e) Seuls les patients avec N auront des contrôles de la masse cancéreuse
- f) Tous les patients N et A auront des contrôles de la masse cancéreuse

Réponse b, f:

Comparaison de 2 traitements : 2 groupes de patients cancer, TAS pour protocole N et protocole A, contrôle de la masse du cancer chez tous les patients.



Dans la pop française, la valeur moyenne de la glycémie est de 1g/L (5 mmol/L). On TAS un échantillon de 256 personnes dans cette pop. La valeur moyenne de la glycémie mesurée dans cet échantillon est de 1,1 g/L. Quelles sont les propositions exactes.

- A) Avec un tel résultat il est clair qu'il y a eu un biais d'échantillonnage
- B) Il n'y a aucun doute à avoir, le TAS assure d'avoir un échantillon représentatif
- C) La glycémie est une variable quantitative continue.
- D) On peut calculer l'IC directement à partir des données de l'énoncé.
- E) Aucune proposition ne convient

Réponse : B et C

A est fausse (voir B), D est fausse (manque écart type)

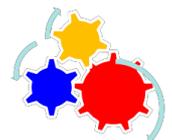


15 malaises graves ont été recensés lors de plongées profondes chez des plongeurs confirmés. Les médecins de la Fédération nationale de plongée cherchent à établir la cause de ces malaises. Ils mesurent le taux d'anexine 5 dans le sang des plongeurs, et on relève leur âge.

Age	18	22	25	26	29	31	36	38	41	42	45	53	56	62	63
Taux	5	5,1	8	7,2	12	10	18	18	9,2	25	21	26	35	28	20

2 questions se posent :

- 1 Taux d'anexine responsable des malaises?
- 2 Si oui, y a-t-il un lien entre taux d'anexine et âge des plongeurs?



34

Age	18	22	25	26	29	31	36	38	41	42	45	53	56	62	63
Taux	5	5,1	8	7,2	12	10	18	18	9,2	25	21	26	35	28	20

Taux d'anexine responsable des malaises?

- TAS autre groupe de 15 plongeurs, sans malaise, de même caractéristique que le groupe « malaise » (âge, état de santé, habitude de la plongée ..)
- Données qualitatives (malaises ou non) et quantitatives (taux anexine)
- H_0 : pas de différence entre les taux d'anexine dans les 2 groupes.
- Effectif de 15 : t de Student. Taux moyen et écart type dans les 2 groupes.
- Si on accepte H_0 : rôle de l'anexine non mis en évidence. L'étude s'arrête là, il faut chercher ailleurs.
- Si on repousse H_0 : rôle de l'anexine mis en évidence. On peut passer à l'étude 2.



Age	18	22	25	26	29	31	36	38	41	42	45	53	56	62	63
Taux	5	5,1	8	7,2	12	10	18	18	9,2	25	21	26	35	28	20

Si oui, y a-t-il un lien entre taux d'anexine et âge des plongeurs?

- Données quantitatives (âge et taux anexine)
- H_0 : pas de lien entre les taux d'anexine et l'âge des plongeurs.
- Calcul du coefficient de corrélation
- Si on accepte H_0 : autre question concernant l'anexine à se poser..
- Si on repousse H_0 : plus les plongeurs sont âgés, plus le taux d'anexine est élevé. Traitement ? Interdiction de plonger profond à partir d'un certain âge?



On veut étudier la relation entre une certaine maladie (classée en 5 niveaux) et la dégradation de la fonction hépatique (mesurée par le taux de Transaminases)

Transa	D1	D2	D3	D4	D5
N	70	21	14	9	3
N – 2N	14	21	7	3	2
2N – 3N	3	8	2	1	0
3N – 4N	3	6	1	0	0
> 4N	9	7	0	1	1

- 1) Etude d'une liaison entre caractères quantitatifs (Taux de transa) et qualitatifs (niveaux de la maladie)
- 2) On ne peut pas définir H0 car il y a trop de cas particuliers dans cette étude
- 3) On peut calculer un coefficient de corrélation
- 4) On peut effectuer un test de comparaison de pourcentage
- 5) On peut effectuer un test du χ^2
- 6) Les intervalles de confiance des taux de transa ne sont pas définis correctement. Les données sont biaisées



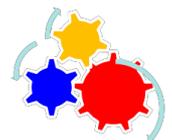
2 journaux font une enquête par sondage auprès d'étudiants en pharmacie pour connaître leurs préférences vis-à-vis de 2 projets de réforme de leurs études.

Journal 1 (Paris)	Réforme A	30%
	Réforme B	45%
	Sans opinion	25%
Journal 2 (Province)	Réforme A	37%
	Réforme B	43%
	Sans opinion	20%

Journal 1 affirme : 15 % de plus pour B, mais faible écart entre A et B

Journal 2 affirme : 6% de plus pour B donc préférence significative pour B

Les conclusions de ces 2 journaux sont elles correctes? Qui a raison ?



Il est impossible de répondre à cette question :

Dans tous les cas, les 2 journaux font une erreur d'interprétation de leur résultats..

En effet, leurs échantillons ne sont pas représentatifs de l'ensemble des étudiants en pharmacie (T.A.S ?, Paris différent de Province,...).

D'autre part, on ne connaît pas la taille des échantillons : pas de test statistique. Les différences de % (15% pour le Journal1) ne sera peut être pas significatif si l'effectif est petit, ou (6% pour le Journal 2) sera peut être significatif si l'effectif très grand..

CONCLUSION : Ne jamais conclure à la seule lecture des % observés !



Table de l'écart réduit

40

	α	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	∞	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	1,476	1,44	1,405	1,372	1,341	1,311
0,2	1,282	1,254	1,227	1,2	1,175	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	0,878	0,86
0,4	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,69
0,5	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,6	0,524	0,51	0,496	0,482	0,468	0,454	0,44	0,426	0,412	0,399
0,7	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,8	0,253	0,24	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,9	0,126	0,113	0,1	0,088	0,075	0,063	0,05	0,038	0,025	0,013

Table pour les petites valeurs de la probabilité

0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
3,2905	3,89059	4,41717	4,89164	5,32672	5,73073	6,10941

Table du Khi2

α

41

ddl	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
....									
30									

Table du t de Student

α

42

ddl	0,9	0,5	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	1	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,816	1,386	1,886	2,92	4,303	6,965	9,925	31,598
3	0,137	0,765	1,25	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,741	1,19	1,533	2,132	2,776	3,747	4,604	8,61
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	1,44	1,943	2,447	3,143	3,707	5,959
7	0,13	0,711	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,13	0,706	1,108	1,397	1,86	2,306	2,896	3,355	5,041
9	0,129	0,703	1,1	1,383	1,833	2,262	2,821	3,25	4,781
10	0,129	0,7	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	1,35	1,771	2,16	2,65	3,012	4,221
14	0,128	0,692	1,076	1,345	1,761	2,145	2,624	2,977	4,14
15	0,128	0,691	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,69	1,071	1,337	1,746	2,12	2,583	2,921	4,015
17	0,128	0,689	1,069	1,333	1,74	2,11	2,567	2,898	3,965
18	0,127	0,688	1,067	1,33	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	3,85

Table du coefficient de corrélation

ddl	α			
	0,1	0,05	0,02	0,01
1	0,9877	0,9969	0,9995	0,9999
2	0,9	0,95	0,98	0,99
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	0,576	0,6581	0,7079
11	0,4762	0,5529	0,6339	0,6835
12	0,4575	0,5324	0,612	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	0,4	0,4683	0,5425	0,5897
17	0,3887	0,4555	0,5285	0,5751
18	0,3783	0,4438	0,5155	0,5614
....				
100	0,1638	0,1946	0,2301	0,254

[re](#)
[to](#)
[ur](#)

n_1 est le plus petit des 2 effectifs, U le plus petit des 2 U calculés

n_1

$n_2 - n_1$	1	2	3	4	5	6	7	8	9	10
0	-	-	-	0	2	5	8	13	17	23
1	-	-	-	1	3	6	10	15	20	26
2	-	-	0	2	5	8	12	17	23	29
3	-	-	0	3	6	10	14	19	26	33
4	-	-	1	4	7	11	16	22	28	36
5	-	-	2	4	8	13	18	24	31	39
6	-	0	2	5	9	14	20	26	34	42
7	-	0	3	6	11	16	22	29	37	45
8	-	0	3	7	12	17	24	31	39	48
9	-	0	4	8	13	19	26	34	42	52
10	-	1	4	9	14	21	28	36	45	55
11	-	1	5	10	15	22	30	38	48	
12	-	1	5	11	17	24	32	41	50	
13	-	1	6	11	18	25	34	43		
14	-	1	6	12	19	27	36	45		
...										
18	-	2	8	16	24	33				
19	-	3	9	17	25					
20	-	3	9	17	27					

44

α	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
n									
4	0.600	1.000	1.000						
5	0.500	0.800	0.900	1.000	1.000				
6	0.371	0.657	0.829	0.886	0.943	1.000	1.000		
7	0.321	0.571	0.714	0.786	0.893	0.929	0.964	1.000	1.000
8	0.310	0.524	0.643	0.738	0.833	0.881	0.905	0.952	0.976
9	0.267	0.483	0.600	0.700	0.783	0.833	0.867	0.917	0.933
10	0.248	0.455	0.564	0.648	0.745	0.794	0.830	0.879	0.903
11	0.236	0.427	0.536	0.618	0.709	0.755	0.800	0.845	0.873
12	0.224	0.406	0.503	0.587	0.671	0.727	0.776	0.825	0.860
13	0.209	0.385	0.484	0.560	0.648	0.703	0.747	0.802	0.835
14	0.200	0.367	0.464	0.538	0.622	0.675	0.723	0.776	0.811
15	0.189	0.354	0.443	0.521	0.604	0.654	0.700	0.754	0.786
16	0.182	0.341	0.429	0.503	0.582	0.635	0.679	0.732	0.765
17	0.176	0.328	0.414	0.485	0.566	0.615	0.662	0.713	0.748
18	0.170	0.317	0.401	0.472	0.550	0.600	0.643	0.695	0.728
19	0.165	0.309	0.391	0.460	0.535	0.584	0.628	0.677	0.712
20	0.161	0.299	0.380	0.447	0.520	0.570	0.612	0.662	0.696
21	0.156	0.292	0.370	0.435	0.508	0.556	0.599	0.648	0.681
22	0.152	0.284	0.361	0.425	0.496	0.544	0.586	0.634	0.667
23	0.148	0.278	0.353	0.415	0.486	0.532	0.573	0.622	0.654
24	0.144	0.271	0.344	0.406	0.476	0.521	0.562	0.610	0.642
25	0.142	0.265	0.337	0.398	0.466	0.511	0.551	0.598	0.630
26	0.138	0.259	0.331	0.390	0.457	0.501	0.541	0.587	0.619
27	0.136	0.255	0.324	0.382	0.448	0.491	0.531	0.577	0.608
28	0.133	0.250	0.317	0.375	0.440	0.483	0.522	0.567	0.598
29	0.130	0.245	0.312	0.368	0.433	0.475	0.513	0.558	0.589
30	0.128	0.240	0.306	0.362	0.425	0.467	0.504	0.549	0.580
31	0.126	0.236	0.301	0.356	0.418	0.459	0.496	0.541	0.571
32	0.124	0.232	0.296	0.350	0.412	0.452	0.489	0.533	0.563
33	0.121	0.229	0.291	0.345	0.405	0.446	0.482	0.525	0.554
34	0.120	0.225	0.287	0.340	0.399	0.439	0.475	0.517	0.547
35	0.118	0.222	0.283	0.335	0.394	0.433	0.468	0.510	0.539
36	0.116	0.219	0.279	0.330	0.388	0.427	0.462	0.504	0.533
37	0.114	0.216	0.275	0.325	0.383	0.421	0.456	0.497	0.526
38	0.113	0.212	0.271	0.321	0.378	0.415	0.450	0.491	0.519
39	0.111	0.210	0.267	0.317	0.373	0.410	0.444	0.485	0.513
40	0.110	0.207	0.264	0.313	0.368	0.405	0.439	0.479	0.507

46

L'ensemble de ce document relève des législations française et internationale sur le droit d'auteur et la propriété intellectuelle.

Tous les droits de reproduction de tout ou partie sont réservés pour les textes ainsi que pour l'ensemble des documents iconographiques, photographiques, vidéos et sonores.

Ce document est interdit la vente ou la location par un tiers autre que l'Université Côte d'Azur (UCA).

La diffusion, la duplication, la mise disposition du public (sous quelque forme ou support que ce soit), la mise en réseau, de tout ou partie de ce document, sont strictement réservées à l'Université Côte d'Azur (UCA).

L'utilisation de ce document est strictement réservée à l'usage des étudiants inscrits aux cours et au tutorat organisés par l'UFR de Médecine de l'Université Côte d'Azur (UCA), et non destinée à toute autre utilisation privée ou collective, gratuite ou payante.

