



SEANCE DE REVISION

Biostatistique

2014/2015

Avec le Pr.Benoliel



Correction détaillée des qcm

QCM 1 : En 2012, on a estimé à 39.500 le nombre de nouveaux cas de cancers pulmonaires. Et l'on s'est aperçu que de 8 à 10% de ces cancers pulmonaires survenaient chez des personnes qui ne fument pas.

- A) Il s'agit d'étudier la relation entre deux variables qualitatives
- B) Il s'agit d'étudier la relation entre une variable qualitative et une variable quantitative
- C) Il s'agit de statistique descriptive
- D) 8 à 10% représente l'intervalle de confiance.
- E) Les items A, B, C et D sont faux

Correction QCM 1

- **QCM 1 : E**

Il n'y a pas 2 variables ni quantitatives, ni qualitatives

Il s'agit effectivement de description d'une situation constatée, mais mal présentée

Il n'y a pas d'intervalle de confiance.

On compare l'efficacité de deux régimes amaigrissants. Pour cela, on constitue par tirage au sort deux groupes comprenant chacun 10 personnes (G_1 = régime 1, G_2 = régime 2). On relève la masse de chacune de ces personnes avant et après le régime (après une période de 3 mois de régime). Au préalable, on a déterminé que les deux groupes étaient équivalents en termes de moyennes des masses corporelles.

- **QCM 2 : Donner les propositions exactes**
- A) Il s'agit d'étudier la relation entre deux variables qualitatives
- B) Il s'agit d'étudier la relation entre une variable qualitative et une variable quantitative
- C) Il faudra faire attention à l'effet placebo
- D) Le test adapté est une comparaison de pourcentages
- E) Les items A, B, C et D sont faux

Correction QCM 2

- QCM 2 : B
- A) Faux : Voir B.
- B) Vrai : Perte de masse = Quantitatif, régime 1 ou 2 = qualitatif
- C) Faux : Pas de placebo donc pas d'effets placebo.
- D) Faux : La comparaison de pourcentage n'est pas adapté pour une variable qualitative et une variable quantitative.
- E) Faux.

- On compare l'efficacité de deux régimes amaigrissants. Pour cela, on constitue par tirage au sort deux groupes comprenant chacun 10 personnes (G_1 = régime 1, G_2 = régime 2). On relève la masse de chacune de ces personnes avant et après le régime (après une période de 3 mois de régime). Au préalable, on a déterminé que les deux groupes étaient équivalents en termes de moyennes des masses corporelles.
- QCM 3 : Donner les propositions exactes
 - A) Les 2 séries sont indépendantes
 - B) Il s'agit d'un cas de séries appariées
 - C) H_0 est : « Le meilleur régime est celui qui fera le mieux maigrir les personnes »
 - D) H_1 est : « Ces 2 régimes sont efficaces »
 - E) Les items A, B, C et D sont faux

Correction QCM 3

QCM 3 : A

- A) Vrai : Les résultats d'un groupe ne dépendent pas des résultats de l'autre.
→ on constitue par tirage au sort **deux groupes** comprenant chacun **10 personnes** et **suivant 2 régimes différents**
- B) Faux : Les deux séries sont **indépendantes donc pas appariées**.
→ un cas de série appariée serait la **mesure d'une même variable** (glycémie, hématoците ...) **dans un même groupe avant et après une modification** (nouveau traitement, régime ...)
- C) Faux : H_0 serait « il n'y a pas de différence d'efficacité entre les deux régimes ».
- D) Faux : H_1 serait « Il existe une différence d'efficacité entre ces deux régimes ».
- E) Faux.

- **QCM 4 : On souhaite savoir si la prise de vitamines modifie la réussite en PACES. Pour cela, on interroge un groupe de 20 étudiants tirés au sort sur le nombre de médicaments de ce type qu'ils prennent chaque jour et sur leur classement au concours. On calcule ainsi un coefficient de corrélation $r_{\text{calculé}} = 0,13$. On donne également, au risque $\alpha=5\%$, la valeur de $r_{\text{théorique}} = 0,44$. Donner les propositions exactes**
- A) Il s'agit d'une étude entre deux variables quantitatives
- B) Il existe une corrélation positive entre ces deux variables au risque $\alpha=5\%$
- C) Il existe une corrélation négative entre ces deux variables au risque $\alpha=5\%$
- D) Ce test permet de rejeter H_0 au risque $\alpha=5\%$
- E) Les items A, B, C et D sont faux

QCM 4 : A

A) Vrai : Attention classement = quantitatif pour les tests.

$r_{\text{calculé}} = 0,13$ $r_{\text{théorique}} = 0,44$ (ddl = 18, $\alpha = 5\%$)

B) Faux : r calculé < r théorique donc aucune corrélation.

C) Faux : r calculé < r théorique donc aucune corrélation.

D) Faux : Pas de corrélation donc on accepte l'hypothèse nulle H_0 .

E) Les items A, B, C et D sont faux

$r_{\text{théorique}}$???????

- 20 etudiants

$$Ddl = n-2$$

$$20-2 = 18$$

« Risque $\alpha=5\%$ »:

ddl \ 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
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

QCM 5 : Un nouveau test sanguin vient d'être mis au point à Nice. Il permet de détecter les cancers du poumons AVANT leur apparition, avant d'être mis en évidence par imagerie, etc..

L'équipe de chercheurs a travaillé sur un groupe de 245 personnes au total, dont 168 gros fumeurs atteints d'une broncho-pneumopathie chronique obstructive. [...] Les participants ont tous subi le test sanguin et les examens classiques d'imagerie médicale. Grâce au test sanguin, des cellules cancéreuses dites «circulantes» ont été identifiées chez cinq patients (3%), alors que rien n'apparaissait à l'imagerie. Chez ces cinq patients, un nodule est devenu visible 1 à 4 ans après le repérage des cellules tumorales par la prise de sang. Ils ont été immédiatement opérés et l'analyse du nodule a bien révélé qu'il était cancéreux

- A) Il s'agit d'une étude entre deux variables qualitatives
- B) Les 5 patients représentent une estimation ponctuelle.
- C) Il manque un intervalle de confiance
- D) Il manque une donnée essentielle pour affirmer que ce test est très efficace.
- E) Les items A, B, C et D sont faux

Correction QCM 5

QCM 5 : BD

- Il n'y a pas 2 groupes, définis à l'avance, il n'y a pas 2 var qualitatives
- Il s'agit bien d'une estimation ponctuelle, pour laquelle il n'y a pas d'intervalle de confiance !
- il manque le nb de cancers déclarés 4 ans plus tard, non mis en évidence par le test sanguin (Faux négatifs)

- **QCM 6 : On cherche à savoir si l'exposition à l'amiante au cours de la vie favorise l'apparition de cancer du poumon. Pour cela, on lance une étude sur des anciens travailleurs d'une entreprise du bâtiment. On fait un TAS dans cette entreprise et on les classe en 4 catégories, ceux qui ont travaillé moins de 5ans dans ce type de métier, ceux entre 5-10 ans, ceux entre 10-20 ans et ceux entre 20 et 40 ans. On observe si il y a eu une survenue d'un cancer du poumon chez eux (on les classe en trois catégories : Non survenue/suspicion de survenue/Survenue). On fait un test statistique au risque $\alpha=2\%$ pour confirmer ou non la relation. On trouve un paramètre calculé $Z = 13.29$. Donnez les vraies**
- A) Ici pour répondre à la question on pourrait utiliser le test de comparaison de pourcentage
- B) On peut dire qu'on ne rejette pas l'hypothèse H_0 au niveau de l'échantillon
- C) On pourra extrapoler la conclusion de notre test a l'ensemble de la population des travailleurs en bâtiment.
- D) Ici pour répondre à la question on doit utiliser le test t de Student.
- E) Les items A, B, C et D sont faux

Correction QCM 6

QCM 6 : B

A) Faux : On a un tableau avec **4 colonne et 3 lignes** donc on ne peut pas utiliser le test de comparaison de pourcentage.

→ **Test du χ^2**

B) Vrai: Le nombre de ddl est de 6, donc en lisant la table on trouve 15,033, ce qui est **supérieur au Z calculé**.

→ $ddl = (\text{nombre de lignes} - 1) \times (\text{nombre de colonnes} - 1) = 3 \times 2 = 6$

Zthéorique > Zcalculé donc on ne rejette pas H0

C) Faux : *On prend en compte seulement une entreprise*, pour pouvoir extrapoler à toute les entreprises il **aurait fallu faire un test sur un échantillon de l'ensemble des entreprises du bâtiment**.

D) Faux : On a à faire a **2 variables qualitatives**

→ **exposition plus ou moins longue à l'amiante / survenue du cancer**

E) Faux

QCM 7 : On cherche à savoir si l'obésité a une cause génétique ou non. Pour cela on compare le poids de 5 adultes avec celui de leur mère, et on cherche à savoir s'il y a ou non un lien. On fait un test statistique et on trouve alors un paramètre calculé $Z=0,867$, au risque $\alpha=5\%$. Donnez les réponses vraies

- A) Le test utilisé ici est le coefficient r' de Spearman
- B) L'hypothèse H_1 pourrait être « il existe un lien entre le poids des mères et celui des enfants »
- C) On accepte l'hypothèse H_1
- D) La valeur théorique du Z se lit avec la valeur du risque α
- E) Les items A, B, C et D sont faux

QCM 7 : ABD

- A) Vrai: 5 adultes => $n < 12$; Variables Quant.
- B) Vrai : H_1 « Il y a une difference »
- C) Faux: Ici l'effectif est de 5, et α de 5%, donc on lit sur la loi du r' de Spearman $Z_{th} = 1$.
 $0,867 < 1$, donc on ne rejette pas H_0 .
- D) Vrai
- E) Faux

« paramètre calculé $Z=0,867$ »

On REJETTE H_0 si $Z_{calculé} > Z_{théorique}$. (Valable pour tous les tests sauf le test U de Mann whitney »

$\alpha=5\%$; $n=5$:

α	α			
	.10	.05	.02	.01
5	.900	1.000	1.000	—
6	.829	.886	.943	1.000
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.683	.783	.833
10	.564	.648	.746	.794
12	.506	.591	.712	.777
14	.456	.544	.645	.715
16	.425	.506	.601	.665
18	.399	.475	.564	.625
20	.377	.450	.534	.591
22	.359	.428	.508	.562
24	.343	.409	.485	.537
26	.329	.392	.465	.515
28	.317	.377	.448	.496
30	.306	.364	.432	.478

- **QCM 8 : Je cherche à savoir s'il y a une relation entre l'appréciation d'un restaurant (regroupé en deux classes : Médiocre/Excellent) et le nombre moyen de personne y mangeant par jour. Pour cela, j'effectue un test statistique sur 105 restaurants, avec 56 classés médiocres et 49 excellents. Je compte ensuite le nombre de personnes y mangeant et j'en fait une moyenne pour chaque restaurant. Le paramètre calculé du test vaut $Z=2,12$. Donnez les réponses vraies.**
- A) Si on prend les valeurs conventionnelles, on accepte H_1
- B) Ici on a deux variables quantitatives
- C) On utilise le test du coefficient r de corrélation
- D) Si on change le choix du risque fixé à priori α , alors la conclusion de ce test pourrait être différente pour cette valeur...
- E) Les items A, B, C et D sont faux

QCM 8 : AD

$Z_{\text{calculé}} = 2,12$

On REJETTE H_0 si $Z_{\text{calculé}} > Z_{\text{théorique}}$. (Valable pour tous les tests sauf le test U de Mann whitney »

A) Vrai :

Pour $\alpha = 5\%$, $\epsilon = 1.96$.

$Z_{\text{th}} < Z_{\text{calculé}}$: On accepte H_1

B) Faux: « Classe » => qualitative et nb de personnes.

C) Faux: Ici, on utilise le test de comparaison de moyenne.

D) Vrai: Pour $\alpha = 1\%$, $\epsilon = 2,6$ => $2,12 < 2,6$

- QCM 9 : On s'intéresse à la différence de concentration en testostérone entre 1 groupe de 7 rugby-mans (A) et un groupe de 8 footballeurs (B). Les résultats des dosages sont les suivants (en mmol/L), y a-t-il une différence significative entre les deux groupes ?
- - Groupe A : 32 – 45 – 38 – 54 – 40 – 35 – 50
- - Groupe B : 39 – 56 – 60 – 30 – 45 – 47 – 37 - 48
- A) H_0 : « Les rugby-mans ont plus de testostérone que les footballeurs »
- B) H_0 : « Il n'y a pas de différence significative de concentration de testostérone entre les deux groupes. »
- C) On utilise le U de Mann et Whitney (test non paramétrique permettant d'étudier le lien entre une variable qualitative et une variable quantitative).
- D) On calcule U_{AB} et U_{BA} puis on compare le plus grand des deux au U théorique lu dans la table.
- E) Les réponses A, B, C et D sont fausses.

Correction QCM 9

QCM 9 : BC

- A) Faux : Ce n'est **ni H_1 ni H_0** .
- B) Vrai : C'est bien H_0 .
- C) Vrai : Définition du U de Mann et Whitney.
- D) Faux : On compare le plus petit des deux au U théorique ! Si U calculée est inférieure au U théorique alors les groupes ne sont pas imbriqués et il y a une différence entre eux !!
- E) Faux.

- Rappel calcul de U_{AB} et U_{BA} :

- On commence par classer en ordre croissant les taux de testostérone (je mets en Gras les scores du groupe B et en rouge les scores du groupe A) :

30 – **32** – **35** – **37** – **38** – **39** – **40** – **45** – **46** – **47** – **48** – **50** –
54 – **56** – **60**

- Rappel calcul de U_{AB} et U_{BA} :

- On commence par classer en ordre croissant les taux de testostérone (je mets en Gras les scores du groupe B et en rouge les scores du groupe A) :

30 - **32** - **35** - **37** - **38** - **39** - **40** - **45** - **46** - **47** - **48** - **50** -
54 - **56** - **60**

- On calcule ensuite U_{BA} comme le prof le dit dans son cours : $U_{BA} = 0 + 2 + 3 + 4 + 5 + 5 + 7 + 7 = 33$;
- Pour avoir U_{AB} : $U_{AB} = N_A * N_B - U_{BA} = 8*7 - 33 = 23$

- Rappel calcul de U_{AB} et U_{BA} :

- On commence par classer en ordre croissant les taux de testostérone (je mets en Gras les scores du groupe B et en rouge les scores du groupe A) :

30 – **32** – **35** – **37** – **38** – **39** – **40** – **45** – **46** – **47** – **48** – **50** – **54** – **56** – **60**

- On calcule ensuite U_{BA} comme le prof le dit dans son cours : $U_{BA} = 0 + 2 + 3 + 4 + 5 + 5 + 7 + 7 = 33$;
- Pour avoir U_{AB} : $U_{AB} = N_A * N_B - U_{BA} = 7*8 - 33 = 23$
- On va chercher le U théorique dans le tableau avec $N_B - N_A = 1$ et $N_A = 7$ et on trouve $U = 10$
- On conclue : U calculé est supérieur à U théorique = séries intriquées = On accepte H_0 .

Table U de Mann - Whitney

n2-n1	n1									
	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	28
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	38
5	-	-	2	4	8	13	18	24	31	39
6	-	0	2	5	9	14	20	28	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

- **QCM 10 : La PACES est une année difficile ... mais pas pour vos adipocytes ! Au contraire, ils se multiplient avec joie, couvrant allègrement vos tablettes de chocolats acquises durant l'été et s'accumulant fièrement au-devant de votre abdomen ou autres parties du corps imprévues. Vos tuteurs de biostat, toujours avides d'études statistiques, veulent savoir si la réussite au concours influe sur la prise de poids (mais l'adipocyte est mignon, ne vous inquiétez pas). Pour cela, on constitue par tirage au sort 2 groupes de 15 étudiant(e)s. Le groupe 1 est constitué d'étudiants ayant réussi le concours avec une prise de poids moyenne de 3.2 kg. Le groupe 2 est constitué d'étudiants n'ayant pas réussi le concours avec une prise de poids moyenne de 2 kg.**
- A) On utilise un test étudiant le lien entre données qualitatives et données quantitatives
- B) On utilise donc un test t de Student avec un nombre de ddl = 28
- C) Le paramètre calculé = 2.6, on ne rejette pas H_0 au risque $\alpha = 1\%$
- D) De plus, on ne rejette pas H_0 au risque $\alpha = 5\%$
- E) Aucune de ces réponses n'est correcte.

Correction QCM 10

QCM 10 : ABC

- A) Vrai : lien entre **données qualitatives** (réussite ou non du concours) et **données quantitatives** (nombre de kg pris)
- B) Vrai : on utilise un **t de student** (car liaison entre qualitatif et quantitatif) avec ddl = 28 (formule des ddl pour ce test est **Nb ddl = (n1 + n2) - 2 = (15 + 15) - 2 = 28**)
- C) Vrai : on trouve un *t théorique* de 2.763 dans la table pour un *ddl=28* et $\alpha=1\% = 0.01$; donc **t calculé < t théorique** ☐ on **accepte (ne rejette pas) Ho**
- D) Faux : on trouve un *t théorique* de 2.048 dans la table pour un *ddl=28* et $\alpha=5\% = 0.05$; donc **t calculé > t théorique** ☐ on **rejette Ho / accepte H1**
- E) Faux



Questions des PACES au
professeur!!

**1 - % de la population $< \mu - 2.6\sigma$, est-ce 0,2% ou 0,5%?
Vu que pour un intervalle de $[\mu - 2.6\sigma ; \mu + 2.6\sigma]$ on a deux versions, soit 99,6% qui semble être la valeur la plus précise mais également 99%.**

Réponse :

La diapo 35 de mon cours explique dans l'intervalle $[\mu - 2,6\sigma ; \mu + 2,6\sigma]$ il y a 99,6 % de la pop. On en déduit que en dehors de cet IC il reste

0,4% donc 0,2% de chaque côté (courbe de Gauss).

Mais il est vrai que ces affirmations sont toutes probabilistes, donc il est fréquent de faire des approximations (95% au lieu de 95,4 ou 99% ; ..). Sachez aussi que la courbe de Gauss décrit « *à peu près* » 100% de la population !!

A partir de la Loi Normale ou de GAUSS, on précise :

Intervalles de confiance

- ★ $[m - 1 s ; m + 1 s]$ contient 68,2% de la population
- ★ $[m - 1,96 s ; m + 1,96 s]$ contient 95,4% de la population
- ★ $[m - 2,6 s ; m + 2,6 s]$ contient 99,6% de la population

Patient avec une glycémie de 1,2 g/l.
Quel est l'intervalle de confiance?

Intervalle de confiance au risque 5% = $[0,90 ; 1,20]$

On ne peut répondre au patient qu'en connaissant l'intervalle de confiance du dosage

2 - Que doit on considérer concernant les variables catégorielles ? Sont t-elles seulement nominales ou cela concerne t-il toutes les variables qualitatives (ordinales et nominales) ?

- **variable catégorielle** (on dit aussi **qualitative**) : les différentes valeurs que peut prendre cette variable sont appelées les catégories, modalités ou niveaux
- **variables ordinales** : il existe une relation d'ordre entre les modalités, mais ce sont des variables catégorielles

3 - La décision d'accepter H_0 :

A) est équivalente à H_0 vraie et H_1 fausse

B) est équivalente à la conclusion du test de ne pas rejeter H_0

C) traduit le fait qu'il n'y a pas d'évidence nette pour que H_0 soit fausse

D) traduit le fait qu'il n'y a pas d'évidence nette pour que H_1 soit vraie

Il faut bien comprendre que les tests d'hypothèse ne permettent pas d'accepter H_0 mais seulement de rejeter ou non H_0 , jamais à l'accepter d'emblée.

On choisit pour H_0 ce que l'on souhaite rejeter. Ne pas rejeter H_0 ne signifie pas que H_0 est vraie mais seulement que la probabilité qu'elle soit fausse est très petite (Item C)

On n'est donc en fait jamais vraiment totalement sûr de rien

4 - Suffit t-il qu'il y ait écrit TAS, ou randomisation dans un énoncé pour que l'échantillon soit représentatif de la population concernée ?

Y'a t'il un seuil d'effectif d'échantillon également ?

Si dans un QCM rien ne précise que l'échantillon est effectué par tirage au sort, est ce qu'on doit considérer que l'étude est faussée et donc que le risque alpha ne peut pas être déterminé ? Ou alors est-ce qu'on effectue l'étude MAIS on ne peut pas extrapoler (car pas de randomisation) ?

TAS, ou randomisation dans un énoncé >>échantillon représentatif de la population cible

- a) Pas de TAS : on travaille sur l'échantillon proposé.
- b) On détermine A PRIORI le risque de 1^{ère} espèce α
- c) On fait le test adapté
- d) On tire la conclusion adaptée suivant le résultat du test, au niveau de l'échantillon
- e) ON NE PEUT PAS EXTRAPOLER A LA POP CIBLE

5 - Peut t-on considérer une variable quantitative (discrète ou continue) regroupée en classe (comme les personnes entre 40 et 60kg, 61 et 80kg, 81 et 100kg, 101kg et plus) comme une variable qualitative ordinale ?

Une variable qualitative catégorielle peut résulter de la transformation d'une variable quantitative en procédant à **un découpage en classes**. Il existe alors une variable quantitative sous-jacente, mais cette information n'est pas forcément intégrée dans les traitements

6 - Le professeur a utilisé le t de student pour un effectif de 10 personnes, pourquoi n'a t'il pas utilisé le U de Mann Whitney.. ?

**« Test t Student séries appariées
ou
Mann et Whitney »**

Cette « frontière » d'effectif de 10 – 12 n'est pas fermée ..Il n'est pas interdit de faire un t de Student avec un effectif de 10, ni de faire un Mann et Whitney avec un effectif de 12..

Car, on a le droit d'utiliser un test non paramétrique à la place d'un test paramétrique. Mais au-delà de cette « frontière », l'inverse n'est pas possible !

TESTS NON PARAMÉTRIQUES

- Utilisation obligatoire si les effectifs sont trop faibles
- (>4 & < 10 ou 12), avec des caractères quantitatifs. Dans ce cas, les populations ne se distribuent pas normalement.
- Présentent une excellente robustesse

LIAISON QUANTITATIFS / QUALITATIFS

- U de Mann & Whitney

LIAISON ENTRE QUANTITATIFS

- Spearman (= Corrélation)

Récapitulatif sur les effectifs des tests et leurs utilisations

YES!

MÉTHODOLOGIE D'UTILISATION DES TESTS

Effectif	Données Quantitatives	Données Qualitatives	Données Qualitatives Quantitatives
>4 & < 10 ou 12	r' de Spearman	Comp % ou χ^2	U Mann & Withney
>=10 ou 12 & < 30	Coeff de corrélation r	Comp % ou χ^2	t Student
>= 30	Coeff de corrélation r	Comp % ou χ^2	Comp moyennes