

SEANCE DE REVISION DE BIOSTATIQUES

2015/2016

Avec le professeur BENOLIEL ! 😊

Présentation de la séance

- ▶ I. Questions des étudiants
- ▶ II. QCMs d'entraînement
- ▶ III. Correction détaillée

Question 1

A propos des **intervalles de confiance** :

Voici donc ce que veulent savoir les étudiants :

"Risque de 1^{ère} espèce α dans l'estimation important \rightarrow Intervalle resserré \rightarrow **Estimation améliorée (car l'intervalle se resserre)** et précision augmentée (aussi car l'intervalle se resserre)

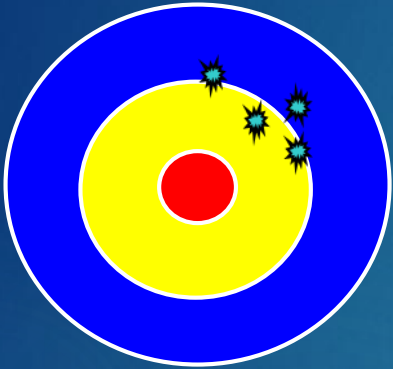
ou bien

Risque de 1^{ère} espèce α dans l'estimation important \rightarrow Intervalle resserré \rightarrow **Estimation amoindrie (du fait risque α de manquer l'estimation très élevé)** et précision augmentée (aussi car l'intervalle se resserre)"

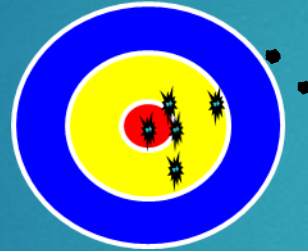
A retenir :

Précision de l'estimation

Intervalle de confiance peut être vu comme une cible



α petit donc IC large
donc plus de chances
de l'atteindre, et
mauvaise précision
de l'estimation



α grand donc IC
resserré donc plus de
risques de rater,
meilleure précision de
l'estimation

Indice qui permet de calculer la précision
de l'estimation de μ :

$$i = \frac{\varepsilon S}{\sqrt{n}}$$

Cette valeur est la largeur de l'intervalle de
confiance.

Indice petit = meilleure précision

$$\mu \in \left[m \pm \frac{\varepsilon S}{\sqrt{n}} \right] \Rightarrow \text{Intervalle au risque } \alpha$$

Suite :

« Risque de 1^{ère} espèce α dans l'estimation important \rightarrow Intervalle resserré \rightarrow Estimation améliorée (car l'intervalle se resserre) et précision augmentée (aussi car l'intervalle se resserre) »

- ✓ Estimation améliorée : terme non correct
- ✓ Pour un n fixé, α grand donc ε petit donc i diminue donc intervalle resserré donc précision améliorée. Cette affirmation est juste.
- ✓ En échange, α grand donc risque plus grand de rater la bonne estimation.

« Risque de 1^{ère} espèce α dans l'estimation important \rightarrow Intervalle resserré \rightarrow Estimation amoindrie (du fait risque α de manquer l'estimation très élevé) et précision augmentée (aussi car l'intervalle se resserre) »

- ✓ Estimation amoindrie : terme non correct
- ✓ Pour un n fixé, α grand donc ε petit donc i diminue donc intervalle resserré donc précision améliorée. Cette affirmation est juste.
- ✓ En échange, α grand donc risque plus grand de rater la bonne estimation

Plus l'intervalle de confiance est resserré plus il y a de risque de rater mais meilleure est l'estimation

Question 2

QCM : On compare la diminution du taux de cholestérol produite par un nouveau tt T à celle produite par le tt de référence R.

Après tt dans le groupe T ($n_T=400$) $IC_{95\%} = [1,10 ; 1,20]$ $m_T = 1,15$ mmol/l

Après tt dans le groupe R ($n_R=400$) $IC_{95\%} = [1,30 ; 1,50]$ $m_R = 1,4$ mmol/l

Parmi les propositions suivantes, choisir celle(s) qui est (sont) exacte(s) :

- A) Le nouveau traitement est plus efficace que celui de référence
- B) Différence statistiquement significative car $m_R \notin IC_{95\%}(m_T)$
- C) Les intervalles de confiance permettent de conclure que $m_T \neq m_R$
- D) Différence significative entre les 2 traitements.
- E) On ne peut rien conclure, car on ne sait pas si la distribution des valeurs est normale.

Réponses

Réponse : B, C, D

Les 2 IC ne se recouvrent pas. Les 2 moyennes sont donc significativement différentes au risque 5%.

T est moins efficace que R pour la baisse des taux de cholestérol ($1,15 < 1,4$) (A)

Avec des échantillons de 400 sujets ($n > 30$), on ne se pose pas la question (E)

Question 2

« Pourquoi l'item A est faux ? »

Rappelons que l'on s'intéresse à la diminution du taux de cholestérol

Le nouveau traitement ne fait baisser en moyenne le cholestérol

QUE DE 1,15 mmol/l

alors que l'ancien le fait baisser de 1,4 mmol/l.

« Pourquoi l'item B est juste ? »

Après tt dans le groupe T ($n_T=400$) $IC_{95\%} = [1,10 ; 1,20]$ $m_T = 1,15$ mmol/l

Après tt dans le groupe R ($n_R=400$) $IC_{95\%} = [1,30 ; 1,50]$ $m_R = 1,4$ mmol/l

Les 2 IC sont totalement disjoints donc il existe une différence statistiquement significative entre les moyennes. En d'autres termes, il n'existe aucune chance que, même par hasard, une des 2 moyennes se trouve dans l'IC de l'autre !

Cela justifie pleinement le « car » qui apparaît dans l'item B.

Question 3

Un sondage est effectué auprès de 1200 malades atteints par la fièvre Ebola pour voir s'ils ont été à l'hôpital le mois précédant l'infection. Seulement 400 personnes ont répondu au sondage, dont 200 ont été à l'hôpital. Donnez les réponses vraies :

Item : Le sondage reste valable même si tous n'ont pas répondu, du moment que les personnes ont été tirés au sort (**compté faux**)

La question est : Peut-on réellement dire qu'un sondage est "invalide" du fait de non-réponses ? Dans la vie courante, un sondage fait toujours face à des non-répondants, doit-on forcément le considérer non valable pour ce fait ?

- 1) **Le sondage reste valable ... terme non correct.**
- 2) **...qu'un sondage est "invalide" ...terme non correct.**
- 3) **...doit-on forcément le considérer non valable... terme non correct**

400 personnes répondent sur 1200 interrogées. 800 non répondants. On doit tout tenter pour obtenir des réponses.. (vu en cours).

On effectue les calculs sur les répondants, on tire des %, y compris des non répondants (dans les sondages effectués tous les jours, il y a le % « ne se prononcent pas.. »)

ATTENTION, On ne sait pas si TAS, il y a trop de non répondants, on ne sait donc rien de leur histoire, donc on conclut uniquement sur les 400. Ce sondage sera interprété sur ces 400 malades.

Question 4

Variation ordinale peut transformer une variable qualitative ordinale en variable quantitative

Ex : variation de la douleur sur une échelle de 0 à 10 : à considérer comme variable quantitative pour les tests et non comme variable qualitative

Certains traitements et analyses sur variables qualitatives nécessitent, voire exigent, que ces dernières présentent une forme « pseudo quantitative ». Exemple :

5 = Très bon, 4 = Bon, 3 = Moyen, 2 = Mauvais, 1 = Très mauvais

Il est à noter que les nombres affectés aux modalités qualitatives n'ont pas de signification et ne peuvent faire l'objet d'opérations arithmétiques comme par exemple le calcul d'une somme ou d'une moyenne.

Par contre, il est possible de chercher à corrélér ces valeurs à d'autres valeurs purement quantitatives (Corrélation note de biostat/rang de classement au concours..)

Rappel : le test r' de Spearman est appelé « test de corrélation des rangs.. »

Question 5

Les tests non paramétriques peuvent s'utiliser sur des grands échantillons, malgré leur manque de robustesse pour les grands effectifs.

Les étudiants veulent savoir s'il est important de retenir cette notion pour le concours ou bien si celle-ci leur est seulement donnée à titre d'ouverture des connaissances.

ATTENTION : Oui, on peut les utiliser même pour les grands échantillons.

Ces tests sont plus robustes mais moins puissants !

La robustesse d'un test est la qualité de rester relativement insensible à certaines modifications du modèle : on constatera que les tests sur les moyennes sont robustes.

La puissance du test = $(1 - \beta)$ est la probabilité d'accepter H_1 en ayant raison. C'est la probabilité de discerner une différence lorsqu'elle existe. Notion assez complexe.

ouverture des connaissances.

MÉTHODOLOGIE D'UTILISATION DES TESTS

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

Question 6

" L'échantillon doit comporter un nombre minimal de personnes pour être représentatif "

FAUX !

Un échantillon est représentatif si il a été constitué par tirage au sort : **VRAI**

La précision de l'estimation est directement liée à la taille de l'effectif : **VRAI**

Question 7

QCM : 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

Les réponses données en correction sont AB.

*Cependant les étudiants ne comprennent pas pourquoi la D est fausse sachant que le risque α est utilisé dans la table pour trouver le Z théorique. **D est juste !!***

Source de ce QCM ? Fiabilité de cette source? Faute de frappe, bug ?

Question 8

Item qui a posé problème :

“Dans le cas du r' de Spearman, si r' calculé et r' lu dans la table **sont proches (en valeur absolue) alors on rejette H_0** ”

Il est compté vrai avec cette explication : “Si r' calculé et r' lu dans la table sont **proches** on peut en conclure que les deux séries sont liées (on rejette donc H_0)”

Les étudiants s'interrogent puisqu'ils ont appris qu'on rejette H_0 si $r'_{\text{calc}} > r'_{\text{th}}$. Ils ont du mal à saisir la notion de proximité des valeurs absolues avec le fait de rejeter H_0 .

r' calculé et r' théorique sont proches : Notion inconnue en biostatistiques. (< ou > ou =)

Vous avez raison de vous interroger : source de ce QCM ? Fiabilité de cette source? Faute de frappe, bug ?

D'autres questions ?

Questions posées en direct ;)

- ▶ Le professeur Lupi-Pégurier n'a pas le même point de vue quant aux non-répondants lors d'un sondage ...

Le professeur Lupi-Pégurier est en charge d'enseigner l'EPIDEMIOLOGIE, ce qui fait que certaines notions peuvent différer de celles enseignées en statistiques déductives. Le mieux étant de voir avec elle par le biais des tuteurs.

Questions posées en direct :)

- Imaginons que l'on TIRE AU SORT 300 personnes dans le cadre d'une étude. Seulement 100 personnes y répondent. Peut-on parler de représentativité de l'étude ?

Si seuls 100 personnes répondent, on est amené à se poser la question qui est de savoir pourquoi les 200 autres n'ont pas répondu. On décèle donc une faille dans notre étude. Il faut se remettre en question par rapport à cela (ex: Sont-ils morts des suite de l'administration du médicament ? L'étude a-t-elle été menée comme il le fallait ? ...). La question est de savoir si l'on conserve la représentativité. Initialement les 300 personnes tirées au sort sont représentatives de notre population d'origine. Mais les constats post-test doivent nous faire réfléchir sur ce qui a pu pousser les individus à ne pas répondre.

Questions posées en direct ;)

- ▶ Item : '' Pour un effectif > 30 et en présence de variables quantitative et qualitative, les tests à utiliser sont le U de Mann & Whitney, le t de Student et le test de comparaison des moyennes ''
- ▶ Item : '' Pour un effectif > 30 et en présence de variables quantitative et qualitative, les tests à utiliser est le test de comparaison des moyennes ''

Le professeur ne demandera jamais cela le jour du concours. Il vous l'a dit et vous le redit.

Questions posées en direct ;)

- ▶ La catégorisation d'une variable quantitative le transforme t-elle en variable qualitative ?

Ex : Âge fractionné en : Inférieur à 50 ans et Supérieur à 50 ans est à considérer comme qualitative (jeune / vieux).

Tout à fait !

Questions posées en direct :)

- ▶ Que signifie " $p < 0,05$ " à l'issu d'un test ?

Il s'agit de la probabilité post-test de se tromper en admettant que H1 est vraie (en considérant que la probabilité pré-test $\alpha = 5\%$).

Ici, la probabilité de faire une erreur en admettant H1 est faible. On choisit donc d'admettre H1 à l'issu du test.

- ▶ Que signifie "les deux variables étudiées sont comparables" ?

Comparable signifie qu'entre les deux variables étudiées il n'y a pas de différence significative, c'est-à-dire qu'on admet H0.

Place aux QCMS ♥♥

Le tutorat est gratuit toute reproduction ou vente est interdite

QCM 1 : On réalise une étude sur la qualité de l'érection chez l'homme lorsqu'il prend du Viagra. On mesure la qualité de l'érection à l'aide d'un questionnaire à choix multiple : nulle, moyenne, bien, très bien. On constitue 2 groupes de 400 personnes choisies au hasard. Le premier prend du viagra. Le second ne prend pas de viagra. Donner la ou les proposition(s) correcte(s) :

- A) On peut utiliser le test du χ^2
- B) On étudie dans cette étude deux variables quantitatives
- C) L'hypothèse H_0 correspond à « le Viagra modifie la qualité de l'érection »
- D) L'hypothèse H_1 correspond à « le Viagra ne modifie pas la qualité de l'érection »
- E) Les propositions A, B, C et D sont fausses

Réponses :

QCM 1 : A

A) Vrai

B) Faux: deux variables qualitatives

C) Faux : H_0 « Il n'y a pas de différence entre les 2 groupes pour ce qui est de la qualité de l'érection »

D) Faux: H_1 « Il existe une différence significative entre les 2 groupes pour ce qui est de la qualité de l'érection »

E) Faux

QCM 2 : A propos des tests. Donner la ou les proposition(s) correcte(s) :

- A) Pour l'étude de la liaison entre deux variables qualitatives on peut utiliser le test du χ^2
- B) Pour l'étude de la liaison entre un caractère qualitatif et un quantitatif on utilise le test de t de student si et seulement si n_1 et $n_2 < 30$
- C) Dans le même cas si n_1 et $n_2 > 30$ on utilisera la comparaison des moyennes
- D) Pour deux variables qualitatives on utilise le coefficient r de corrélation
- E) Les propositions A, B, C et D sont fausses

Réponses :

QCM 2 : AC

- A) Vrai
- B) Faux : n_1 OU n_2
- C) Vrai
- D) Faux : pour deux variables quantitatives
- E) Faux

QCM 3 : Un médecin cherche à prouver qu'il existe un lien entre le fait de vivre en montagne et l'augmentation du nombre d'hématies dans le sang. Pour cela, il constitue, par TAS, 1 groupe de 33 personnes qui ne vivent pas en montagne et un groupe de 15 personnes vivants en montagne. On réalise une numération de la formule sanguine et on ne s'occupe que du nombre d'hématies. Le paramètre calculé est de 3,3 et le paramètre théorique est de 2,042 au risque $\alpha = 5\%$. Donner la ou les proposition(s) correcte(s) :

- A) On utilise un test de coefficient de corrélation
- B) On peut utiliser un test t de Student
- C) On peut accepter H_0 au risque de 5%
- D) On peut rejeter H_1 au risque de 5%
- E) Les propositions A, B, C et D sont fausses

Réponses :

QCM 3 : B

- A) Faux : une variable qualitative et une variable quantitative
- B) Vrai : Une variable qualitative et une variable quantitative **et** n_1 ou $n_2 < 30$
- C) Faux : t calculé $>$ t théorique rejet de H_0 et acceptation de H_1 au risque de 5%
- D) Faux :
- E) Faux

QCM 4 : Un chirurgien obstétrique veut étudier le lien entre la taille d'une tumeur de l'utérus et le caractère bénin ou malin au risque $\alpha = 5\%$. Pour cela il constitue deux échantillons de 35 patients représentatifs du service de gynécologie-obstétrique du CHU de Nice. Un groupe est constitué de patients porteurs de tumeurs bénignes, et l'autre de tumeurs malignes. Pour chaque patient dans les 2 groupes, on mesure la taille de la tumeur.

La valeur du Z calculée est de 2,1. Donner la ou les proposition(s) correcte(s) :

- A) On utilise le test de t de comparaison de pourcentage
- B) L'hypothèse H_0 est « il n'y a pas de lien significatif entre la taille d'une tumeur et le caractère bénin/malin »
- C) On peut conclure qu'il n'y a pas de lien significatif entre la taille de la tumeur et le caractère bénin/malin de la tumeur
- D) On peut conclure qu'il existe une différence de pronostic (bénin/malin) suivant la taille de la tumeur.
- E) Les propositions A, B, C et D sont fausses

Réponses :

QCM 4 : BD

A) Faux : test de comparaison de **moyennes**

B) Vrai :

C) Faux : χ calculé > χ théorique (**1,96**) rejet H_0 « il n'y a pas de lien significatif entre la taille d'une tumeur et le caractère bénin/malin »

D) Vrai

E) Faux

QCM 5 : Relation entre une maladie (classée en 5 niveaux D1, ...D5)) et la mesure de 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

- A. Etude d'une liaison entre caractères quantitatifs (Taux de transa) et qualitatifs (niveaux de la maladie)
- B. On peut calculer un coefficient de corrélation
- C. On peut effectuer un test de comparaison de pourcentage
- D. On peut effectuer un test du χ^2
- E. Les propositions A, B, C et D sont fausses

Réponses :

QCM 5 : A

A) Vrai

B) Faux : voir A

C) Faux

D) Faux

E) Faux

QCM 6 : Lors de la réalisation d'un essai clinique, une étude est menée pour évaluer la clairance (élimination par l'organisme) d'une nouvelle molécule en regard d'une ancienne molécule utilisée de façon très répandue en clinique. L'essai comparatif porte sur 400 patients. On mesure la concentration plasmatique de la molécule dans les 2 heures qui font suite à l'administration des médicaments.

On cherche à savoir s'il existe une différence de clairance notable entre ces deux médicaments au compromis universel. Quelle(s) est(sont) la(les) proposition(s) exacte(s) parmi les suivante(s) ?

- A) On peut réaliser le test de comparaison des moyennes
- B) On peut réaliser le test de comparaison des pourcentages
- C) On pose comme hypothèse nulle H_0 : "La clairance des deux médicaments est comparable"
- D) La différence étant positive, on peut affirmer avec un risque de 1^{ère} espèce de 5% que la clairance diffère entre les deux molécules testées
- E) Les propositions A, B, C et D sont fausses

	Concentration plasmatique
Groupe A : Nouvelle molécule – 200 patients	IC95% = [1,95 +/- 0,15]
Groupe B : Ancienne molécule – 200 patients	IC95% = [1,70 +/- 0,20]

Réponses :

QCM 6 : AC

A) Vrai : On compare deux molécules (ancienne / récente) – variable qualitative – et leur clairance respective (concentration plasmatique) – variable quantitative. De plus, l'effectif est supérieur à 30.

B) Faux

C) Vrai

D) Faux : IC95% groupe A = $[1,80 - 2,10]$; IC95% groupe B = $[1,50 - 1,90]$. Les intervalles de confiance se chevauchent, on ne peut donc pas admettre de différence significative.

E) Faux

QCM 7 : Une étude s'attache à analyser les résultats obtenus suite à la mise en place de deux traitements hygiéno-diététiques pour des patients en surcharge pondérale tirés au sort. On note 54% d'amélioration de l'état de santé dans le groupe suivant des séances de kinésithérapie et 78% dans le groupe suivant des séances de sport. Le résultat de la comparaison est accompagné de " $p < 0,05$ ". Que conclure ?

- A) Il y a moins de 5 chances sur 100 que la différence observée soit due au hasard
- B) On rejette l'hypothèse alternative H_1
- C) Les deux traitements diffèrent significativement pour ce qui est de l'amélioration de l'état de santé
- D) Il manque des données pour pouvoir conclure
- E) Les propositions A, B, C et D sont fausses

Réponses :

QCM 7 : AC

- A) Vrai
- B) Faux
- C) Vrai
- D) Faux
- E) Faux

QCM 8 : Un laboratoire pharmaceutique s'intéresse à l'effet d'un nouveau médicament hypoglycémiant. Il constitue par tirage au sort un groupe A de 5 personnes diabétiques et un groupe B de 4 personnes diabétiques. On donne au groupe A le médicament en question et au groupe B un placebo. Au bout d'une semaine de traitement on mesure leur glycémie capillaire.

Les concentrations relevées pour chaque sujets sont (en g/L)

Groupe A : 0,99 – 1,09 – 1,14 – 1,32 – 1,44

Groupe B : 0,97 – 1,26 – 1,31 – 1,42

On donne $U_{th} = 4$, $U_{ab} = 10$ et $U_{ba} = 10$

- A) On étudie le lien entre deux variables qualitatives
- B) On utilise le U de Mann et Withney et on conclut à une différence significative entre les deux groupes
- C) H_0 : « les sujets du groupe A ont une glycémie plus élevée que le groupe B »
- D) H_1 : « il n'y a pas de différence significative de concentration de glucose entre les deux groupes »
- E) Les propositions A, B, C et D sont fausses

Réponses :

QCM 8 : E

A) Faux : on étudie ici la relation entre une variable qualitative (sujets ayant reçu le médicament ou non) et une variable quantitative (concentration de glucose en g/L) De plus les effectifs sont compris entre 4 et 12 on utilise donc le test du U de Mann & Withney

B) Faux on trouve $U_{calculé} > U_{théorique}$ donc on rejette H_1 et on accepte H_0

C) Faux : ce n'est ni H_1 ni H_0

D) Faux : c'est H_0

E) Vrai

QCM 9 : On s'intéresse à la concentration d'hémoglobine dans le sang en fonction du sexe de l'individu. Pour ceci on tire au sort un groupe A de 13 hommes et un groupe B de 16 femmes. On réalise ensuite les mesures des concentrations sanguines.

On donne $Z_{théorique} = 2,052$ à 5%, et $Z_{calculé} = 1,3$

- A) On utilise un test du t de student avec 27 ddl
- B) Au risque $\alpha = 5\%$ on accepte H_0
- C) Au risque $\alpha = 1\%$ on accepte H_1
- D) On positionnera notre $Z_{calculé}$ dans une table de l'écart-réduit
- E) Les propositions A, B, C et D sont fausses

Réponses :

QCM 9 : AB

- A) Vrai : on étudie la liaison entre une variable qualitative (femme/homme) et une variable quantitative (concentration d'hémoglobine). De plus nous avons n_A et/ou $n_B < 30$ on utilisera donc le test du t de student
- B) Vrai : Dans la table du t de student on lit l'intersection entre α et notre ddl, on trouve donc $Z_{calculé} = 1,3 < Z_{théorique} 2,052$: on accepte H_0
- C) Faux voir réponse B)
- D) Faux : table du test de student
- E) Faux

QCM10 : La première étape de mise en œuvre d'un test d'hypothèse correspond à l'une des propositions suivantes, laquelle ?

- A) Choix du risque alpha
- B) Définition de l'hypothèse nulle et de l'hypothèse alternative
- C) Recueil des données
- D) Définition du test en fonction du type de données étudiées
- E) Les réponses A, B, C, D sont fausses

Réponses :

QCM 10 : B

- A) Faux
- B) Vrai
- C) Faux
- D) Faux
- E) Faux

FIN

