Université Claude Bernard Lyon 1

Tutorat Lyon Est

Année Universitaire 2023 – 2024

Unité d’Enseignement 3

Correction Annale Contrôle Continu

Correction détaillée

Loan FRANQUET
Léonie ROLLIN
Fanny CLAUSE
François COQUARD

Correction rapide

Questions	Réponses
1	BC
2	AD
3	BDE
4	E
5	CE
6	BDE
7	ABCDE
8	BCE
9	BCE
10	ABDE
11	ABCE
12	ACDE

Question 1 :

Concernant la p-vague, sélectionner la (les) réponse(s) correctes(s) parmi les suivantes ?

A. La p-value est comparable directement au seuil de rejet.
B. La p-value est comparable directement au risque alpha.
C. La p-value est la probabilité d'observer une statistique de test au moins aussi élevée (en valeur absolue) que la statistique estimée sur les résultats de l'expérience si l'hypothèse nulle était vraie.
D. La p-value est la probabilité que l'hypothèse nulle soit fausse.
E. La p-value est la probabilité que l'hypothèse alternative soit vraie.

A FAUX On compare la p-value au risque alpha. C'est la statistique de test qui est comparée au seuil de rejet.

B VRAI

C VRAI La p-value est définie comme la probabilité d'obtenir un résultat expérimental, si l'hypothèse nulle était vraie, au moins aussi extrême que le résultat observé dans la réalité.

D FAUX Cf C.

E FAUX Cf C.

Question 2 :

Concernant les risques admis lors de la réalisation des tests d'hypothèse, sélectionner la (les) réponse(s) correcte(s) parmi les suivantes.

A. Le risque de première espèce correspond à la probabilité de rejeter à tort l'hypothèse nulle.
B. Le risque de première espèce est déterminé depuis les données expérimentales.
C. Le risque de seconde espèce correspond à la probabilité de rejeter à tort l'hypothèse nulle.
D. Un niveau de significativité inférieur au risque de seconde espèce conduit à rejeter l'hypothèse nulle.
E. Dans le cas d'une comparaison de moyennes issues de deux échantillons indépendants, la puissance est d'autant plus faible que la différence entre les moyennes estimées est élevée, toutes choses égales par ailleurs.

A VRAI Alpha est la probabilité de rejeter H0 alors qu'H0 est vraie.

B FAUX Le risque de première espèce est fixé par l'investigateur, en amont de la réalisation du test. Il n'est donc pas déterminé depuis les données expérimentales.

C FAUX Le risque de seconde espèce est la probabilité de ne pas rejeter H0 alors qu'H0 est fausse.

D VRAI Si p < alpha : on rejette H0.

E FAUX La puissance est d'autant plus faible que la différence entre les moyennes estimées est faible, toutes choses égales par ailleurs. Pour rappel, la puissance $1-\beta$ varie en fonction :

1) de l'écart entre l'hypothèse nulle et la vraie valeur : $\mu 0 - \mu$ (même sens)

2) du risque de 1ère espèce α (même sens)

3) de l'écart-type σ (sens inverse), donc de n (même sens).

Question 3 :

Concernant la démarche générale des tests d'hypothèse, sélectionner la (les) réponse(s) correcte(s) parmi les suivantes.

A. La statistique de test est définie en tenant l'hypothèse alternative pour vraie.
B. La statistique de test est définie en tenant l'hypothèse nulle pour vraie.
C. Si on ne rejette pas l'hypothèse nulle, alors celle-ci est nécessairement vraie.
D. Les deux hypothèses sont incompatibles.
E. Pour un test bilatéral, si on définit C(A) l'événement complémentaire de A, C(H0) = H1.

A FAUX La statistique de test est définie en tenant l'hypothèse nulle pour vraie.

B VRAI

C FAUX A la suite d'un test d'hypothèse, on ne peut pas accepter H0. On peut seulement rejeter H0 ou ne pas rejeter H0.

D VRAI

E VRAI Si on rejette H0, on accepte H1.

Question 4 :

Un essai clinique randomisé évalue l'intérêt sur la qualité de vie des patients d'un nouveau traitement en cancérologie. Cet essai comporte deux bras, constitués aléatoirement (chaque bras est un échantillon, les échantillons sont indépendants, chaque individu est indépendant de tous les autres). La qualité de vie des patients a été mesurée par une échelle continue, mais l'analyse est conduite sur une échelle de qualité de vie agrégée en deux catégories : bonne et mauvaise pour l'utilisation aisée de ces résultats en pratique courante.

Les résultats de cette étude pour la totalité des patients sont les suivants :

	Groupe contrôle	Groupe intervention
Bonne	10	21
Mauvaise	75	64

Les tests sont réalisés au risque alpha de 0,05.

On souhaite savoir s'il existe une différence statistiquement significative dans la distribution de la qualité de vie (en catégories) entre les groupes contrôle et intervention. On pose H0, l'égalité des proportions de chaque niveau de qualité de vie entre les groupes contrôle et intervention.

Sachant les données de l'énoncé précédent, sélectionner la (les) réponse(s) correcte(s) parmi les suivantes.

A. On attend entre 16 et 17 patients ayant une bonne qualité de vie dans le groupe contrôle sous l'hypothèse nulle.

B. La statistique de test suit une loi du Chi-deux à 3 degrés de liberté.

C. Le seuil de rejet, tronqué à deux chiffres après la virgule, est égal à 5,02.

D. La valeur absolue de la différence entre la statistique de test estimée et le seuil de rejet est inférieure ou égale à 1.

E. Vu le résultat du test, on ne rejette pas l'hypothèse nulle au risque alpha.

On commence par calculer les effectifs attendus en rouge en faisant total ligne x total colonne / total. Ils sont tous supérieurs à 5 donc on peut appliquer un test du Chi-deux.

	Groupe contrôle	Groupe intervention	Total
Bonne	10 31x85/170=15,5	21 31x85/170=15,5	31
Mauvaise	75 139x85/170=69,5	64 139x85/170=69,5	139
Total	85	85	170

A FAUX On attend entre 15 et 16 patients dans le groupe contrôle.

B FAUX La statistique du test suit une loi du chi 2 à 1 DDL :
Nb de ddl : (nb de colonnes-1)x(nb de lignes-1): (2-1) x (2-1) = 1 ddl.

C FAUX Le seuil de rejet à 1ddl avec alpha=0,05 vaut 5,99 (dans la table du Chi-deux).

D. On utilise la formule pour calculer la statistique de test du Chi-deux:

$$S = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} \sim \chi^2_{(I-1)\times(J-1)}$$

Sobs=$\frac{(10-15,5)^2}{15,5}+\frac{(21-15,5)^2}{15,5}+\frac{(75-6\ ,5)^2}{69,5}+\frac{(64-69,5)^2}{69,5}$=4,77.

Sobs - q$1-alpha$ = 5,99 - 4,77 = 1,22 > 1.

E VRAI 4,77 < 5,99 donc Sobs <q$1-alpha$ donc on ne rejette pas H0.

Question 5 :

Un essai clinique randomisé évalue l'intérêt sur la qualité de vie des patients d'un nouveau traitement en cancérologie. Cet essai comporte deux bras, constitués aléatoirement (chaque bras est un échantillon, les échantillons sont indépendants, chaque individu est indépendant de tous les autres). 170 patients au total ont été inclus dans cette étude.

On dispose de l'âge de chaque patient de l'étude : 66 ont plus de 65 ans. Dans la population française, 19,6% des personnes ont plus de 65 ans. On souhaite savoir s'il existe une différence statistiquement significative dans la proportion de personnes de plus de 65 ans entre l'étude et la population française. Le test sera réalisé au risque alpha de 0,05.

Parmi les suivants, sélectionner la (les) réponse(s) correcte(s) parmi les suivantes.

A. La statistique de test estimée est comprise dans l'intervalle (4,5) (bornes exclues).
B. La statistique de test estimée est comprise dans l'intervalle (5,6) (bornes exclues).
C. La statistique de test estimée est comprise dans l'intervalle (6,7) (bornes exclues).
D. Le dénominateur de la statistique de test estimée est compris entre 0,029 et 0,030.
E. La statistique de test suit une loi normale de moyenne 0 et de variance 1.

On commence par vérifier les conditions d'application : 170 > 30 ; 170 x 0,196 > 5 ; 170 x (1 - 0,196) > 5. On peut effectuer un test de la loi normale pour comparer la proportion observée dans l'échantillon à la proportion de référence dans la population. On utilise la formule de la statistique de test :

Statistique de test

$$Z_{H_0} = \frac{|\hat{\pi} - \pi_0|}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim \mathcal{N}(0,1)$$

A FAUX

B FAUX

C VRAI $Zh0 = \frac{66/170 - 0{,}196}{\sqrt{\frac{0{,}196\ x\ (1-0{,}196)}{170}}} \approx 6{,}3$

D FAUX. Le dénominateur vaut 0,0304… ce qui n'est pas compris entre 0,029 et 0,030.

E VRAI La loi normale est centrée et réduite.

Question 6 :

Concernant les paramètres de position, sélectionner la (les) réponse(s) correcte(s) parmi les suivantes ?

Parmi les suivants, sélectionner la (les) réponse(s) correcte(s) parmi les suivantes.

A. La moyenne est exprimée sans unité.
B. Sur un échantillon de taille 10, la présence d'une seule valeur extrême va influencer la valeur de la moyenne.

C. Sur un échantillon de taille 10, la présence d'une seule valeur extrême va influencer la valeur de la médiane.
D. Pour un échantillon de taille n, le rang du premier quartile est (n+1) × 0,25.
E. La médiane et le 2ème quartile sont identiques.

A FAUX La moyenne est exprimée avec une unité si la variable possède une unité. Ex : moyenne du poids en kg.

B VRAI La présence d'une seule valeur extrême va influencer la valeur de la moyenne. En effet, elle sera prise en compte dans le calcul de la moyenne.

C FAUX La valeur de la médiane n'est pas influencée par la présence d'une valeur extrême puisque que la médiane correspond à la valeur qui permet de séparer la série triée en deux parties de mêmes effectifs. Or une valeur extrême se trouvera aux extrémités de la série triée.

D VRAI

E VRAI Q2 = médiane.

Question 7 :

Concernant la loi normale, sélectionner la (les) réponse(s) correcte(s) parmi les suivantes ?

A. La densité est la dérivée de la fonction de répartition.
B. La fonction de répartition est monotone croissante.
C. La densité est symétrique par rapport à la moyenne.
D. L'aire sous la densité vaut 1.
E. La densité est strictement positive sur]-∞ ; +∞ [.

A VRAI

B VRAI

C VRAI La densité de probabilité est symétrique par rapport à l'axe vertical passant par la moyenne.

D VRAI

E VRAI

Question 8 :

Soit M l'évènement « être malade » et G l'évènement « avoir le génotype CC au niveau d'un SNP »

Parmi les propositions ci-dessous relatives aux probabilités, cochez celle(s) qui est/sont vraie(s).

A. $P(G) = \frac{P(M \cap G)}{P(M)}$
B. $P(M \cap G) = P(M) + P(G) - P(M \cup G)$
C. Si M et G sont indépendants, alors $P(M/G) = P(M)$.
D. L'évènement (M|G) est un événement conditionnel.
E. Si M et G sont incompatibles alors $P(M \cap G) = 0$.

A FAUX $P(G) = \frac{P(M \cap G)}{P_G(M)}$

B VRAI

C VRAI phrase de cours

D FAUX Il n'existe pas d'événements conditionnels.

E VRAI

Question 9 :

Le volume expiratoire maximal par seconde (VEMS) correspond au volume d'air expiré pendant la première seconde d'une expiration forcée suite à une inspiration profonde. À partir d'un échantillon de 250 hommes âgés de 20 à 29 ans, on a pu estimer le VEMS moyen à 3,4 L et son écart-type à 0,6 L.

Vous ne ferez aucun arrondi dans les calculs intermédiaires. Les résultats finaux seront donnés avec 2 chiffres après la virgule.

Parmi les propositions ci-dessous, cochez celle(s) qui est/sont vraie(s).

A. La borne supérieure de l'intervalle de confiance à la confiance 0,95 du VEMS moyen dans la population des hommes âgés de 20 à 29 ans vaut 3,47 L.
B. La borne inférieure de l'intervalle de confiance à la confiance 0,968 du VEMS moyen dans la population des hommes âgés de 20 à 29 ans vaut 3,31 L.
C. La borne inférieure de l'intervalle de confiance à la confiance 0,95 du VEMS moyen dans l'échantillon des hommes âgés de 20 à 29 ans vaut 3,32 L.
D. La borne supérieure de l'intervalle de confiance à la confiance 0,90 du VEMS moyen l'échantillon des hommes âgés de 20 à 29 ans vaut 3,46 L.
E. Toutes choses égales par ailleurs, pour diminuer la largeur d'un intervalle de confiance, on peut augmenter la taille de l'échantillon étudié.

On a fait une estimation d'une grandeur dans un échantillon de 250 personnes, donc pour voir si ces estimations correspondent aux valeurs de la population, on peut réaliser des intervalles de confiance. On utilise donc la formule $ic_{1-\alpha}(\mu) = [m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}}]$. On sait que m=3,4 ; s=0,6 et n=250.

A FAUX Donc $ic_{0,95}(\mu) = [3,4 \pm 1,96 \times \frac{0,6}{\sqrt{250}}] = [3,4 \pm 0,074]$

On demande la borne supérieure, donc on arrondit au supérieur la borne. borne sup = 3,4+0,08=3,48 L.

B VRAI $ic_{0,968}(\mu) = [3,4 \pm z_{0,016} \times \frac{0,6}{\sqrt{250}}] = [3,4 \pm 2,1444 \times \frac{0,6}{\sqrt{250}}] = [3,4 \pm 0,0812]$

On demande la borne inférieure, donc on arrondit à l'inférieur la borne. borne inf = 3,4-0,09=3,31 L.

C VRAI $ic_{0,95}(\mu) = [3,4 \pm 1,96 \times \frac{0,6}{\sqrt{250}}] = [3,4 \pm 0,074]$

On demande la borne inférieure, donc on arrondit à l'inférieur la borne. borne inf = 3,4-0,08=3,32 L.

D FAUX $ic_{0,90}(\mu) = [3,4 \pm z_{0,05} \times \frac{0,6}{\sqrt{250}}] = [3,4 \pm 1,6449 \times \frac{0,6}{\sqrt{250}}] = [3,4 \pm 0,062]$

On demande la borne supérieure, donc on arrondit au supérieur la borne. borne sup = 3,4+0,07=3,47 L.

E VRAI L'intervalle sera plus précis, donc moins large si on augmente le nombre de personnes n dans l'échantillon. ("On divise par un plus grand nombre, donc on diminue plus la largeur de l'intervalle")

Question 10 :

Le BNP (Brain Natriuretic Peptide) est un peptide utilisé comme biomarqueur de l'insuffisance cardiaque. On définit la variable aléatoire X modélisant la concentration en BNP dans une population d'hommes âgés a priori sain. On considère que cette variable X suit une loi de Gauss d'espérance $\mu X = 80$ pg/mL et d'écart-type σX=30 pg/mL. On considère qu'un patient ayant une concentration de BNP supérieure à 100 pg/mL a un risque d'être atteint d'insuffisance cardiaque.

On s'intéresse à un échantillon aléatoire simple constitué de 50 hommes âgés a priori sain. On définit les variables aléatoires suivantes :

Y : « nombre de patients de l'échantillon ayant un risque d'être atteint d'insuffisance cardiaque »

Z : « proportion de patients ayant un risque d'être atteint d'insuffisance cardiaque dans l'échantillon »

Les résultats des calculs seront donnés avec 2 chiffres après la virgule.

Parmi les propositions ci-dessous, cochez celle(s) qui est/sont vraie(s)

A. La borne supérieure de l'intervalle de fluctuation, à la confiance 0,95, de la concentration moyenne en BNP vaut 88,32.
B. La probabilité qu'un individu de l'échantillon ait un risque d'être atteint d'insuffisance cardiaque est inférieure à 2 x 10 - 5.
C. La variable aléatoire Y suit une loi de Bernoulli de paramètre p environ égal à 0,25.
D. La variable aléatoire Z suit approximativement une loi normale d'espérance $\mu Z \cong 0{,}25$ et d'écart type σ$Z \cong$ 0,06.
E. La probabilité d'avoir exactement un patient à risque d'insuffisance cardiaque dans l'échantillon vaut environ : 50 x 0,25 x $0{,}75^{49}$

A VRAI $IF_{0{,}95}(X) = [80 \pm\ 1{,}96 \times \frac{30}{\sqrt{50}}] = [80 \pm 8{,}32]$ Donc la borne supérieure de l'intervalle de fluctuation de la concentration moyenne de BNP vaut 88,32.

B VRAI $P(X > 100) = P\left(Z > \frac{100}{30}\right) = P\left(Z > \frac{2}{3}\right) = 1 - 0{,}7422 = 0{,}2578$ soit environ 0,25 si on ne garde que deux chiffres après la virgule. 0,25 > 2 x 10^{-5}.

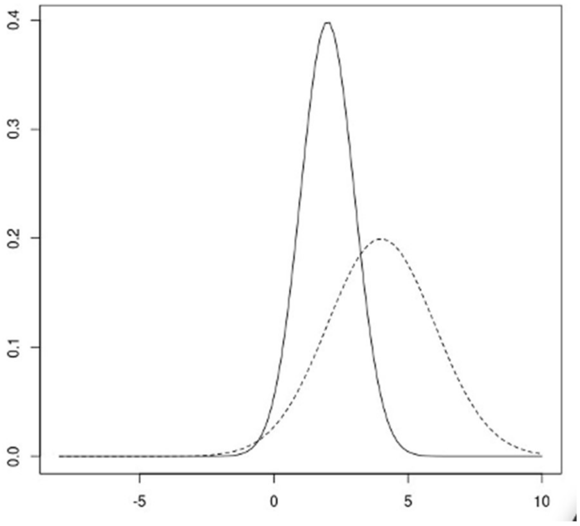
C FAUX Y suit une loi Binomiale de paramètre p = 0,25 et n = 50.

D VRAI Z suit approximativement une loi normale d'espérance : μZ = p = 0,25 et d'écart-type σZ= $\sqrt{\frac{pq}{n}} = \sqrt{\frac{0{,}25 \, x \, 0{,}75}{50}}$ = 0,06.

E VRAI P(Y =1) = $\frac{50!}{1!49!}$ x $0{,}25^1$ x $0{,}75^{49}$ = 50 x $0{,}25^1$x $0{,}75^{49}$

Question 11 :

Soient U et V, 2 variables aléatoires Gaussiennes dont les densités de probabilité sont représentées sur la figure ci-dessous, en traits pleins pour V et en pointillés pour U.



Parmi les propositions suivantes relatives aux variables aléatoires et aux intervalles de confiances, cochez celle(s) qui est/sont vraie(s).

A. On voit sur la figure que E(U) > E(V).
B. On voit sur la figure que var(U) > var(V).
C. Les bornes de l'intervalle de confiance dépendent de l'échantillon considéré.
D. Un intervalle de confiance est forcément bilatéral.
E. Soit T l'estimateur de θ, le biais de T se calcule ainsi : E(T)-θ.

A VRAI La densité de probabilité est maximale en μ. Le pic de V se trouve avant celui de U.

B VRAI La densité de probabilité de U est plus large que celle de V donc var(U) > var(V).

C VRAI La formule pour calculer les bornes d'un intervalle de confiance prend en compte la moyenne et l'écart-type observés dans l'échantillon.

D FAUX Un intervalle de confiance peut être unilatéral.

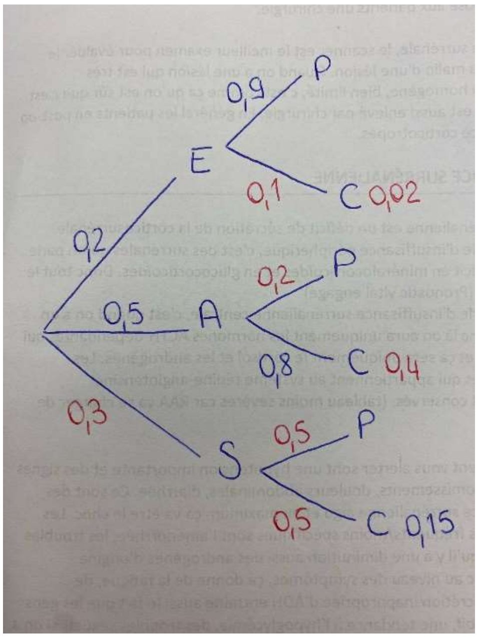
E VRAI

Question 12 :

Un médicament pouvant être pris par les enfants, les adultes ou les seniors, existe sous 2 formes : poudre à diluer ou comprimé à avaler. Parmi les patients qui prennent ce médicament, 20 % sont des enfants et 50 % sont des adultes de moins de 65 ans. 90 % des enfants qui prennent ce médicament utilisent la forme « poudre à diluer » alors que 80 % des adultes de moins de 65 ans préfèrent la forme « comprimé ». On sait aussi que 15 % des personnes prenant le médicament sont des seniors utilisant la forme « comprimé »

Parmi les propositions ci-dessous, cochez celle(s) qui est/sont vraie(s)

A. Parmi les individus prenant le médicament, 40 % sont des adultes ayant choisi la forme comprimée.
B. Parmi les individus prenant le médicament, 10 % sont des enfants ayant choisi la forme comprimée.
C. Parmi les individus prenant le médicament, 57 % ont choisi la forme « comprimé ».
D. On choisit aléatoirement un individu ayant choisi la forme « comprimé », la probabilité qu'il s'agisse d'un adulte de moins de 65 ans vaut environ 0,7.
E. La moitié des seniors ont choisi la forme « comprimé ».



A VRAI On cherche : $P(A \cap C) = P(A) \times P(C|A) = 0{,}5 \times 0{,}8 = 0{,}4$.

B FAUX On cherche : $P(E \cap C) = P(E) \times P(C|E) = 0{,}2 \times 0{,}1 = 0{,}02$. 2% des individus prenant le médicament sont des enfants qui ont choisi le comprimé.

C VRAI On cherche : $P(C) = P(E \cap C) + P(A \cap C) + P(S \cap C) = 0{,}02 + 0{,}4 + 0{,}15 = 0{,}57$.

D VRAI On utilise la formule de Bayes : $P(A|C) = \frac{P(C|A) \times P(A)}{P(C)} = \frac{0{,}8 \times 0{,}5}{0{,}57} \simeq 0{,}7$.

E VRAI On cherche : $P(C|S) = \frac{P(\cap CS)}{P(S)} = \frac{0{,}15}{0{,}3} = 0{,}5$.