Université Claude Bernard Lyon 1

Tutorat Lyon Est

Année Universitaire 2021 - 2022

Unité d'Enseignement 3

Annale Contrôle Intermédiaire 2021-22

6 pages | 9 questions | 30 minutes

Mohamed Yassine LAGHRIBI
Laura LAGRESLE
Lien Anh VO

<u>Correction rapide</u>

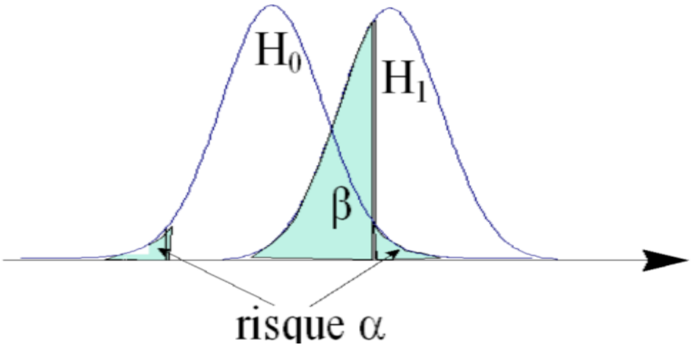
Questions	**Réponses**
1	E
2	ABD
3	AE
4	ACD
5	ABDE
6	DE
7	AE
8	BCE
9	ACD

La correction rapide est celle des professeurs. La correction détaillée néanmoins a été réalisée par vos tuteurs, elle n'est en aucun cas officielle et n'engage en rien la responsabilité des professeurs.

Question 1 – Tests statistiques: E

Indiquez la ou les réponse(s) juste (s) :

A. Le risque de seconde espèce β est la probabilité de rejeter l'hypothèse nulle quand elle est vraie.
B. L'hypothèse nulle est rejetée lorsque le risque de première espèce est inférieur au niveau de significativité.
C. La puissance est la probabilité de rejeter l'hypothèse alternative.
D. L'hypothèse nulle est rejetée lorsque le niveau de significativité excède le risque de seconde espèce.
E. Toutes choses étant égales par ailleurs, la puissance augmente avec la taille de l'étude.



Ceci est une représentation des risques alpha et beta sur la densité de la loi normale. On peut comprendre que :

- **α** est la probabilité de rejeter H0 alors qu'elle était vraie.
- **β** est la probabilité de ne pas rejeter H0 (en effet, on est en dessous du seuil α choisi) alors qu'elle était fausse.
- La puissance vaut **1- β**, c'est donc la probabilité d'accepter H1 (et rejeter H0) alors que H1 est bien vraie : plus la puissance est élevée, plus on est sûr de notre résultat. Pour être plus sûr de notre résultat et augmenter la puissance, il est possible d'augmenter le nombre de sujets dans l'étude par exemple.

Le degré (ou niveau) de significativité d'un test statistique correspond au p que l'on calcule.

C'est la probabilité que l'on puisse observer la valeur que l'on a, sachant que l'hypothèse nulle est vraie. C'est donc la probabilité d'obtenir une valeur aussi éloignée de celle attendue sous H0 que la valeur que l'on a obtenue, sachant que H0 est vraie.

On compare ensuite p au risque de première espèce α, ou seuil de significativité, qui correspond en fait à la probabilité que H0 soit vraie la plus élevée que l'on est prêt à accepter : c'est le risque de rejeter H0 à tort.

Ainsi, avec p < α, on rejette H0 et on accepte H1. Le test est significatif.

Avec p > α, on ne rejette pas H0. Le test n'est pas significatif. Attention, on n'accepte pas H0 pour autant !

A FAUX Voir au-dessus.
B FAUX Voir au-dessus.
C FAUX Voir au-dessus.
D FAUX On le compare au risque de première espèce.

E VRAI La taille de l'étude sous-entend le nombre de sujets inclus. Plus il est grand, plus nos résultats sont fiables, plus la puissance est élevée.

Question 2 – Tests statistiques (Coefficient : 2 points) : ABD

On admet que la proportion de la population des individus de plus de 12 ans vaccinés contre le COVID avec un schéma vaccinal complet (l'ensemble des doses nécessaires) est de 80%. Vous souhaitez savoir si les résultats sont différents au sein de votre canton. Vous effectuez une étude sur un échantillon aléatoire de 400 individus de plus de 12 ans du canton. Le plan d'analyse statistique prévoit d'utiliser un test basé sur la loi normale, d'effectuer un test bilatéral, en fixant le risque d'erreur de première espèce à α=5%.
Sur l'échantillon constitué de 400 individus, 280 ont bénéficié d'un schéma vaccinal complet.

A. La valeur calculée de la grandeur test est -5.
B. Vous rejetez l'hypothèse nulle et estimez la proportion d'individus vaccinés avec un schéma complet au sein du canton à 70%, valeur significativement inférieure à la valeur théorique de 80%.
C. Vous rejetez l'hypothèse nulle avec $0,001 < p < 0,01$.
D. Vous rejetez l'hypothèse nulle avec $p < 0,00004$.
E. Les conditions d'utilisation de la Normale ne sont pas vérifiées.

On peut commencer par relever les informations importantes : ici, la proportion d'individus vaccinés théorique est **π_0 = 80%.** On relève également que **n=400** et que **280 sont vaccinés**, on peut donc calculer la **proportion observée** dans l'échantillon :

$$f = \frac{280}{400} = \frac{28}{40} = \frac{7 \times 4}{10 \times 4} = 0,7$$

Nous allons donc chercher à comparer une proportion observée à une proportion théorique en utilisant la loi normale. Il est important de noter la précision dans la consigne : on réalise un **test bilatéral.**

On commence donc par vérifier que les **conditions d'utilisation** de cette dernière sont remplies : il nous faut $\mathrm{n}\pi_0 \geq 5$ et $\mathrm{n}(1-\pi_0) \geq 5$. En appliquant aux valeurs de l'exercice, nous avons $400 \times 0,8 = 320$ *et* $400 \times 0,2 = 80$. Les conditions sont donc vérifiées.

On applique donc la formule du cours :

$$Z = \frac{F - \pi_0}{\sqrt{\frac{\pi_0.(1-\pi_0)}{n}}} \sim \mathrm{N}(0;\ 1)$$

$$\mathrm{Z} = \frac{0,7-0,8}{\sqrt{\frac{0,8.0,2}{400}}} = \frac{-0,1}{\sqrt{\frac{0,16}{400}}} = \frac{-0,1}{\frac{0,4}{20}} = \frac{-0,1}{\frac{4}{200}} = \frac{-0,1}{\frac{2}{100}} = \frac{-0,1}{0,02} = -\frac{10}{2} = -5$$

La grandeur test calculée est donc égale à -5.

Nous allons maintenant pouvoir trouver le degré de significativité p auquel cela correspond. Nous cherchons donc :

$$P(Z < -5) = 1 - P(Z < 5)$$

Lorsqu'on regarde la table, on se rend compte qu'elle s'arrête à 4,09, nous pouvons donc dire que le p que nous allons lire sera supérieur au p réel du test. Ainsi, nous pouvons écrire :

$$p < 1 - P(Z < 4{,}09)$$

$$p < 1 - 0{,}99998$$

$$p < 0{,}00002$$

Or, nous sommes dans un test bilatéral ! Il faut donc multiplier la valeur du p que l'on vient de trouver par 2 : $p_{bilatéral} < 0{,}00004$. Nous pouvons donc rejeter l'hypothèse nulle au seuil alpha = 5% choisi, avec p < 0,00004.

A VRAI Voir ci-dessus.

B VRAI Voir ci-dessus.

C FAUX p est inférieur à 0,001.

D VRAI Voir ci-dessus.

E FAUX Voir ci-dessus.

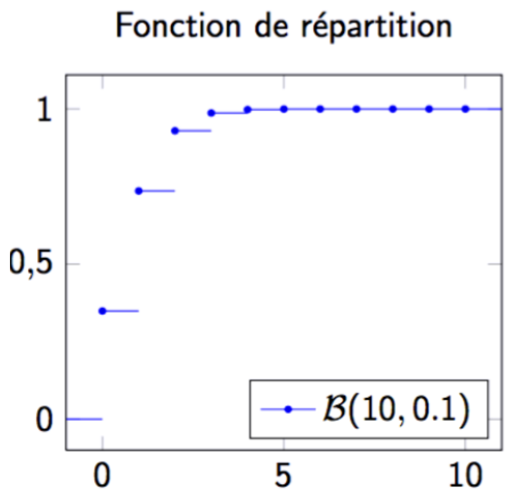
<u>Question 3 – Variables Aléatoires et lois classiques:</u> AE

Concernant la fonction de répartition d'une loi de probabilité d'une variable aléatoire X.
Indiquez la ou les réponse(s) juste(s).

A. Si la loi de probabilité est discrète, la hauteur des marches correspond aux probabilités des différentes valeurs de X.
B. Si la loi de probabilité est continue, la limite en + l'infini (+∞) de la fonction de répartition vaut 0,5.
C. La fonction de répartition d'une loi discrète est monotone croissante.
D. La densité d'une loi continue est l'intégrale de sa fonction de répartition.
E. L'aire sous la courbe de la densité d'une loi continue vaut 1.

A VRAI Voici un exemple de la fonction de répartition d'une loi binomiale de paramètres n = 10 et p = 0,1.

En abscisse, on retrouve les valeurs de X et en ordonnée la probabilité correspondante.



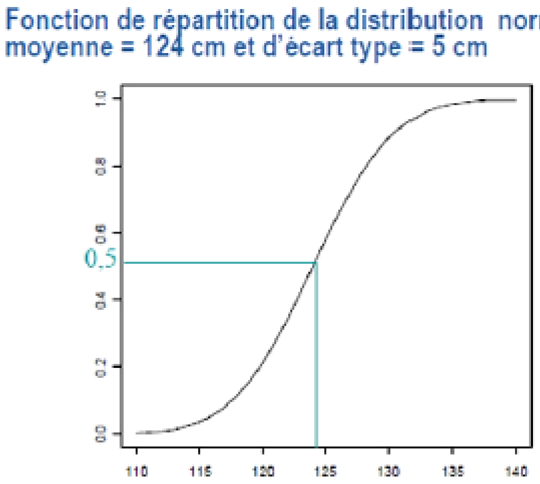
B FAUX Si la loi de probabilité est continue, la limite en + l'infini (+∞) de la fonction de répartition vaut 1.

En effet, une probabilité ne peut prendre que des valeurs entre 0 et 1. Dans une fonction de répartition, plus on prend des grandes valeurs, plus on se rapproche d'1.

Un exemple de loi de probabilité continue est la loi normale de paramètres μ et σ. Sa fonction de répartition est alors définie par :

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right)dt$$

Plus x est grand, plus P(X≤x) se rapproche d'1 (cf. Table de la fonction de répartition de la loi normale centrée réduite pour avoir une meilleure illustration).



Fonction de répartition de la distribution normale.

C FAUX on demande à la professeure et on met une correction !

D FAUX La fonction de répartition est l'intégrale de la densité d'une loi continue.

E VRAI L'aire sous la courbe de la densité d'une loi continue vaut 1.

Question 4 – Comparaison de Moyennes et de Variances (Coefficient : 2 points) : ACD

Trois traitements A, B et C d'une maladie sévère sont comparés dans le cadre d'un essai randomisé en groupes parallèles. Le critère de jugement principal est la guérison clinique (succès / échec). Dans un premier temps, le protocole prévoit la comparaison globale des trois traitements et la réalisation d'un test du Chi-2 à 2 degrés de liberté. L'hypothèse nulle est rejetée avec $0,010 < p < 0,025$.

	Traitement A	Traitement B	Traitement C	Total
Succès	50	40	60	150
Echec	50	60	40	150
Total	100	100	100	300

Les traitements sont alors comparés 2 à 2, le risque de première espèce global étant fixé à 5%. Les 3 tests réalisés sont bilatéraux.

Le Chi-2 comparant B et C vaut 8.

Concernant cette comparaison, indiquez la ou les réponse(s) juste(s) :

A. Le traitement C est significativement plus efficace que le traitement B.
B. Le niveau de significativité du test vérifie $0,001 < p < 0,01$.
C. Le niveau de significativité du test vérifie $0,003 < p < 0,03$.
D. Au risque d'erreur de première espèce fixé, vous rejetez l'hypothèse nulle.
E. Si le risque de première espèce avait été fixé à a = 0,1%, l'hypothèse nulle aurait été rejetée.

A VRAI On nous donne la valeur du Chi-2, il ne nous reste plus qu'à trouver dans la table la valeur de p associé, pour la comparer au risque de première espèce fixé.

Nous comparons ici uniquement les traitements B et C, ce qui correspond à un test à un seul degré de liberté (alors que lors de la comparaison des 3 traitements en même temps, il y a 2 degrés de liberté).

Table IV – *Fractiles de la loi du χ^2*

Soit X une variable aléatoire suivant une loi du χ^2 à n degrés de liberté. Pour une probabilité p donnée, la table donne la valeur x telle que $P(X < x) = p$.

Exemple d'utilisation de la table

Soit X une variable aléatoire suivant une loi du χ^2 à 3 degrés de liberté. Soit $p = 0{,}95$, alors la valeur de x telle que $P(X < x) = 0{,}95$ est 7,8147.

ddl \ p	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,999
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,1015	0,4549	1,3233	2,7055	3,8415	5,0239	6,6349	10,8276
2	0,0100	0,0201	0,0506	0,1026	0,2107	0,5754	1,3863	2,7726	4,6052	5,9915	7,3778	9,2103	13,8155
3	0,0717	0,1148	0,2158	0,3518	0,5844	1,2125	2,3660	4,1083	6,2514	7,8147	9,3484	11,3449	16,2662

Ainsi, comme on a $6{,}63<X^2<10{,}82$, nous pouvons en déduire que $1-0{,}999<p<1-0{,}990$ soit **$0{,}001<p<0{,}01$.**

Attention néanmoins, il faut appliquer une correction de Bonferroni. En effet, le protocole de l'étude prévoyait un seuil de significativité à 5%. Or ici, 3 comparaisons des traitements 2 à 2 sont réalisées, nous allons donc multiplier p par le nombre de combinaisons possibles Q. Q = $C_2^3 = \frac{3!}{2!(3-2)!} = 3$, soit 3 combinaisons possibles des traitements comparés 2 à 2. Donc il faut multiplier la valeur du p par 3. Ainsi, ici, nous obtenions **$0{,}003<p<0{,}03$.** Cela ne change rien au faut que le test soit significatif, par contre cela change vos réponses aux items B et C.

Nous pouvons donc rejeter l'hypothèse nulle et affirmer que le traitement C est significativement plus efficace que le traitement B. (attention, vérifiez bien que ce soit bien le traitement C le plus efficace, et pas le B, ce serait dommage de se tromper sur le sens de la comparaison.)

B FAUX Voir ci-dessus.

C VRAI Voir ci-dessus.

D VRAI Voir ci-dessus.

E FAUX Comme nous l'avons vu, en appliquant la correction de Bonferroni, cet item devient faux !

Question 5 – Probabilités (Coefficient : 2 points) : ABDE

Chaque année, en France, la gastro-entérite sévit dans les classes de maternelle. On s'intéresse à la population des enfants de moins de 6 ans dans laquelle la prévalence de la gastro-entérite est de 20 %. On considère que les enfants ne peuvent attraper qu'une seule gastro-entérite au cours de l'année et que le fait d'avoir eu une gastro-entérite une année n'influe pas sur le risque d'en avoir une l'année d'après.

Aide au calcul : 8^3=512 8^4=4096 8^5=32768

Indiquez la ou les proposition(s) juste(s).

A. La probabilité qu'un enfant de moins de 6 ans choisi au hasard ait eu une gastro-entérite en 2020 et à nouveau en 2021 vaut 0,04.
B. Les événements « avoir une gastro-entérite en 2020 » et « avoir une gastro-entérite en 2021 » sont indépendants.
C. La probabilité qu'un enfant de moins de 6 ans choisi au hasard ait eu une gastro-entérite en 2020 ou en 2021 vaut 0,4.
D. Au cours des 5 premières années de sa vie, un enfant choisi au hasard a un peu plus de 5 % de chances d'avoir eu 3 gastro-entérites.
E. Un enfant n'ayant pas eu de gastro-entérite en 2020 a 20 % de chance d'en avoir une en 2021.

Analyse du sujet :

Soit :

$P(G_{2020})$: l'événement « avoir une gastro-entérite en 2020 ».

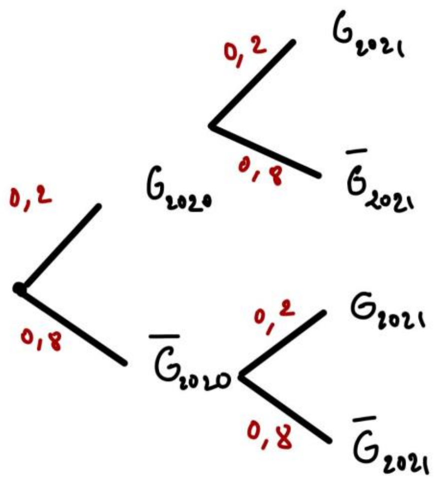
$P(G_{2021})$: L'événement « avoir une gastro-entérite en 2021 ».

$P(G_n) = 0{,}2$ avec n l'année.

« Le fait d'avoir eu une gastro-entérite une année n'influe pas sur le risque d'en avoir une l'année d'après » signifie que $P(G_{2021}|G_{2020}) = P(G_{2021})$.

Autrement dit, les événements « avoir une gastro-entérite en 2020 » et « avoir une gastro-entérite en 2021 » sont indépendants.

A VRAI Voici un arbre illustrant la situation présente.



La probabilité qu'un enfant de moins de 6 ans choisi au hasard ait eu une gastro-entérite en 2020 et à nouveau en 2021 correspond à :

$P(G_{2020} \cap G_{2021}) = P(G_{2020}) \times P(G_{2021}) = 0{,}2 \times 0{,}2 = 0{,}04.$

B VRAI Voir l'explication dans l'analyse du sujet.

C FAUX La probabilité qu'un enfant de moins de 6 ans choisi au hasard ait eu une gastro-entérite en 2020 ou en 2021 peut s'écrire :

$$P(G_{2020} \cup G_{2021}) = P(G_{2020}) + P(G_{2021}) - P(G_{2020} \cap G_{2021})$$

$$P(G_{2020} \cup G_{2021}) = 0{,}2 + 0{,}2 - 0{,}04$$

$$P(G_{2020} \cup G_{2021}) = 0{,}4 - 0{,}04$$

$$P(G_{2020} \cup G_{2021}) = 0{,}36$$

D VRAI On cherche la probabilité qu'un enfant choisi au hasard ait 3 gastro-entérites au cours des 5 premières années de sa vie.

Soit :

$P(G_1 \cap G_2 \cap G_3 \cap \overline{G_4} \cap \overline{G_5}) = P(G_1) \times P(G_2) \times P(G_3) \times P(\overline{G_4}) \times P(\overline{G_5})$

$$= 0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}8 \times 0{,}8$$

$$= 2^3 \times 10^{-3} \times 8^2 \times 10^{-2}$$

$$= 8 \times 10^{-3} \times 8^2 \times 10^{-2}$$

$$= 8^3 \times 10^{-5}$$

$$= 512 \times 10^{-5} = 0{,}00512$$

Or, on sait qu'un enfant ne peut attraper qu'une gastro-entérite au cours de l'année. De ce fait, parmi les 5 années, il y a 10 possibilités pour choisir les 3 années où l'enfant avait la gastro-entérite.

Soit :

$$G_1 \cap G_2 \cap G_3 \cap \overline{G_4} \cap \overline{G_5}$$

$$G_1 \cap G_2 \cap \overline{G_3} \cap G_4 \cap \overline{G_5}$$

$$G_1 \cap \overline{G_2} \cap G_3 \cap G_4 \cap \overline{G_5}$$

$$\overline{G_1} \cap G_2 \cap G_3 \cap G_4 \cap \overline{G_5}$$

$$\overline{G_1} \cap G_2 \cap G_3 \cap \overline{G_4} \cap G_5$$

$$\overline{G_1} \cap G_2 \cap \overline{G_3} \cap G_4 \cap G_5$$

$$\overline{G_1} \cap \overline{G_2} \cap G_3 \cap G_4 \cap G_5$$

$$G_1 \cap \overline{G_2} \cap \overline{G_3} \cap G_4 \cap G_5$$

$$G_1 \cap \overline{G_2} \cap G_3 \cap \overline{G_4} \cap G_5$$

$$G_1 \cap G_2 \cap \overline{G_3} \cap \overline{G_4} \cap G_5$$

Il faut alors multiplier notre résultat par 10.

La probabilité qu'un enfant choisi au hasard ait 3 gastro-entérites au cours des 5 premières années de sa vie :

$$P = \ 0{,}00512 \ \times 10 = 0{,}0512 = 5{,}12\%$$

Soit un peu plus de 5%.

E VRAI Un enfant n'ayant pas eu de gastro-entérite en 2020 a 20 % de chance d'en avoir une en 2021.

Il s'agit d'une conséquence directe de l'indépendance des événements « avoir une gastro-entérite en 2020 » et « avoir une gastro-entérite en 2021 ».

En effet, G_{2020} et G_{2021} sont indépendants revient au même que $\overline{G_{2020}}$ et G_{2021}.

Question 6 – Variables aléatoires et lois classiques – Intervalles (Coefficient : 2 points) : DE

Au cours d'une étude, on mesure le volume d'éjection systolique (VES) d'un échantillon de 100 patients choisis aléatoirement dans la population étudiée. L'estimation du VES moyen vaut 100 mL et l'estimation de la variance du VES vaut 400 mL^2.

On note µ la valeur de la moyenne du VES dans la population, m, son estimation et M l'estimateur de la moyenne.
Aide au calcul : 0,4125≈ 0,41 1,8250 ≈ 1,82 1,96 ≈ 2 2,1201≈ 2,12

Indiquez la ou les proposition(s) juste(s).

A. $ic_{0,966}(\,\mu\,) = [93{,}3\,;103{,}7]$.
B. $ic_{0,966}(\,\mathrm{m}\,) = [95{,}7\,;104{,}3]$.
C. L'intervalle de confiance est centré sur la valeur théorique (= de la population).
D. Plus le niveau de confiance est bas, plus l'intervalle de confiance est étroit.
E. Il faudrait inclure au moins 400 sujets pour obtenir un intervalle de confiance à la confiance 0,95 de largeur égale à 4.

Analyse du sujet :

N = 100

m(VES) = 100 mL

s²(VES) = 400 mL² donc s(VES) = 20 mL

On calcule un intervalle de confiance de la moyenne théorique du volume d'éjection systolique, à la confiance 0,966.

$$ic_{1-\alpha}(\,\mu\,) = m \pm z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$$

$$ic_{0,966}(\,\mu\,) = m \pm z_{\frac{0,034}{2}} \times \frac{s}{\sqrt{n}}$$

$$ic_{0,966}(\,\mu\,) = 100 \pm z_{\frac{0,034}{2}} \times \frac{20}{\sqrt{100}}$$

$$ic_{0,966}(\,\mu\,) = 100 \pm z_{\frac{0,034}{2}} \times \frac{20}{10}$$

$$ic_{0,966}(\,\mu\,) = 100 \pm z_{\frac{0,034}{2}} \;\times\; 2$$

Trouver $z_{\frac{0,034}{2}}$revient à trouver *z* telle que P(Z>*z)* = 0,017. Grâce à la table de la loi normale centrée réduite, on obtient : $z\;$ = 2,1201 ≈ 2,12

p	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010
0,00	∞	3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0749	2,0537
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9431	1,9268	1,9110	1,8957	1,8808
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507

$$ic_{0,966}(\,\mu\,) = 100 \pm z_{\frac{0,034}{2}} \;\times\; 2$$

$$ic_{0,966}(\,\mu\,) = 100 \pm 2{,}12 \;\times\; 2$$

$$ic_{0,966}(\,\mu\,) = 100 \pm 4{,}24$$

$$ic_{0,966}(\,\mu\,) = [95{,}76\,;104{,}24]$$

$$ic_{0,966}(\,\mu\,) = [95{,}7\,;104{,}3]$$

A FAUX

B FAUX On calcule un intervalle de confiance d'une moyenne théorique et non d'une moyenne estimée !

C FAUX L'intervalle de confiance est centré sur la valeur estimée de la population !

D VRAI Plus le niveau de confiance est bas, plus l'intervalle de confiance est étroit.

En effet, plus le niveau de confiance (1- α) est bas, plus α est élevé, plus $z_{\frac{\alpha}{2}}$ est faible (voir table) donc plus l'intervalle de confiance sera étroit. On vous remet la formule ici pour mieux visualiser.

$$ic_{1-\alpha}(\,\mu\,) = m \pm z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$$

E VRAI On cherche le nombre de sujets à inclure pour obtenir un intervalle de confiance à la confiance 0,95 de largeur (l) égale à 4.

$$\mathrm{i} = z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$$

$$i = \frac{l}{2} = z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{\mathrm{n}}}$$

$$\frac{l}{2} = z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{\mathrm{n}}}$$

$$\sqrt{\mathrm{n}} = z_{\frac{\alpha}{2}} \times s \times \frac{2}{l}$$

$$\sqrt{\mathrm{n}} = z_{\frac{0,05}{2}} \times 20 \times \frac{2}{4}$$

Trouver $z_{\frac{0,05}{2}}$ revient à trouver *z* telle que P(Z>*z)* = 0,025. Grâce à la table de la loi normale centrée réduite, on obtient : $z\; = \;1{,}96 = 2.$

Loi normale centrée réduite

Soit Z une variable aléatoire suivant une loi normale centrée réduite. Pour une probabilité *p* donnée, la table donne la valeur *z* telle que $P(Z > z) = p$

p	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010
0,00	∞	3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0749	2,0537
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9431	1,9268	1,9110	1,8957	1,8808
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507
0,04	1,7507	1,7392	1,7279	1,7169	1,7060	1,6954	1,6849	1,6747	1,6646	1,6546	1,6449
0,05	1,6449	1,6352	1,6258	1,6164	1,6072	1,5982	1,5893	1,5805	1,5718	1,5632	1,5548
0,06	1,5548	1,5464	1,5382	1,5301	1,5220	1,5141	1,5063	1,4985	1,4909	1,4833	1,4758
0,07	1,4758	1,4684	1,4611	1,4538	1,4466	1,4395	1,4325	1,4255	1,4187	1,4118	1,4051
0,08	1,4051	1,3984	1,3917	1,3852	1,3787	1,3722	1,3658	1,3595	1,3532	1,3469	1,3408
0,09	1,3408	1,3346	1,3285	1,3225	1,3165	1,3106	1,3047	1,2988	1,2930	1,2873	1,2816
0,10	1,2816	1,2759	1,2702	1,2646	1,2591	1,2536	1,2481	1,2426	1,2372	1,2319	1,2265
0,11	1,2265	1,2212	1,2160	1,2107	1,2055	1,2004	1,1952	1,1901	1,1850	1,1800	1,1750
0,12	1,1750	1,1700	1,1650	1,1601	1,1552	1,1503	1,1455	1,1407	1,1359	1,1311	1,1264
0,13	1,1264	1,1217	1,1170	1,1123	1,1077	1,1031	1,0985	1,0939	1,0893	1,0848	1,0803
0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0450	1,0407	1,0364
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986	0,9945
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581	0,9542
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192	0,9154

$$\sqrt{n} = 2 \times \cancel{2}0 \times \frac{\cancel{2}}{4} = 20$$

Soit : $n = 20^2 = 400$

Question 7 – Variables aléatoires et lois classiques - Probabilités : AE

En 2021, 14 % des étudiants ont obtenu une mention très bien au bac. On considère un échantillon constitué de 25 étudiants ayant passé le bac en 2021. On définit T, la variable aléatoire modélisant le nombre d'étudiants ayant eu une mention très bien dans l'échantillon et U, la variable aléatoire qui pour un étudiant donné prend la valeur 1 s'il a obtenu une mention très bien et 0 sinon.

Aide au calcul : : 25 x 0,14 = 3,50 25 x 0,14 x 0,86 ≈ 3 1,96 ≈ 2

A. T suit une loi binomiale de paramètres n=25 et p = 0,14.
B. L'estimateur de la moyenne du nombre de mention très bien dans l'échantillon vaut 25 x 0,14.
C. L'intervalle de confiance à la confiance 0,95 de T vaut : $ic_{0,95}(\mathrm{T}) = [3{,}50 \pm 2 \times \sqrt{3}]$.
D. On peut approximer la loi de T par une loi normale d'espérance égale à 3,50 et de variance égale à 3.
E. L'espérance de U vaut 0,14.

Analyse du sujet :

T : VA modélisant le nombre d'étudiants ayant eu une mention très bien dans l'échantillon.

U : VA qui pour un étudiant donné prend la valeur 1 s'il a obtenu une mention très bien et 0 sinon.

- $p = 0{,}14$
- $n = 25$

U suit une loi de Bernoulli, l'expérience a 2 issues : 0 (échec) ou 1 (succès).

$$U \rightarrow Bern(0{,}14)$$

T suit une loi binomiale.

$$T \rightarrow B(25\,;\ 0{,}14)$$

A VRAI T suit bien une loi binomiale de paramètres n=25 et p=0,14.

B FAUX La valeur donnée correspond à l'estimateur de la moyenne du nombre d'étudiants ayant eu une mention très bien dans l'échantillon.

E(T) = n x p = 25 x 0,14.

C FAUX On ne peut pas calculer un intervalle de confiance d'une variable aléatoire.

D FAUX On ne peut pas approximer la loi de T par une loi normale. En effet, par les conditions à valider, il faut que n soit supérieur ou égale à 30.

E VRAI L'espérance de U vaut 0,14.

En effet, pour une loi de Bernoulli, l'espérance est égale à p. Soit E(U) = p = 0,14.

Question 8 – Tests Diagnostiques (Coefficient : 2 points) : BCE

Une étude a été mise en place pour évaluer les performances d'un nouvel examen d'imagerie pour faire le diagnostic de cancer de la prostate. Pour cela 500 hommes ayant un dosage de PSA (antigène spécifique de la prostate) supérieur à 4ng/mL ont été inclus et ont eu l'examen d'imagerie. Tous les hommes ont eu des biopsies de la prostate pour déterminer leur statut vis-à-vis du cancer de la prostate. Parmi les 100 hommes ayant un résultat positif pour l'examen d'imagerie (présence d'anomalies en faveur d'un cancer), 70 avaient un cancer de la prostate. Parmi les 400 hommes ayant un résultat négatif pour l'examen d'imagerie, 50 avaient un cancer de la prostate.

Indiquez la ou les réponse(s) juste (s) :

A. La sensibilité de l'examen d'imagerie est estimée par la proportion d'hommes ayant un cancer de la prostate chez les hommes ayant un résultat d'examen d'imagerie positif.
B. La spécificité de l'examen d'imagerie est estimée par la proportion d'hommes ayant un résultat d'examen d'imagerie négatif chez les hommes n'ayant pas de cancer de la prostate.
C. La valeur prédictive positive de l'examen d'imagerie dans cette population est estimée à 70%
D. L'examen d'imagerie est plus sensible que spécifique.
E. La valeur prédictive négative est estimée par la proportion d'hommes n'ayant pas de cancer de la prostate chez les hommes ayant un résultat d'examen d'imagerie négatif.

La première chose à faire devant un exercice de tests diagnostiques est un tableau (cela aide vraiment). Je vous mets en noir les valeurs données par l'énoncé et en bleu celles que vous pouviez déduire :

	Cancer	**Pas de cancer**	**Total**
T+	70	30	100
T-	50	350	400
Total	120	380	500

A FAUX La sensibilité est la probabilité d'avoir un test positif sachant qu'on est malade. Elle est estimée par la proportion d'hommes testés positifs chez les hommes ayant un cancer de la prostate.

B VRAI La spécificité est la probabilité d'avoir un test négatif sachant que l'on n'est pas malade. Elle est estimée par la proportion d'hommes ayant un test négatif chez les hommes qui n'ont pas de cancer de la prostate.

C VRAI La valeur prédictive positive du test est la probabilité d'avoir un cancer sachant que l'on a été testé positif. L'échantillon étant représentatif de la population d'hommes avec une PSA élevée, on peut la calculer.

Ainsi, $VPP = P(M|T+) = \frac{VP}{VP+FP} = \frac{70}{70+30} = \frac{70}{100} = 0{,}7$.

D FAUX Pour affirmer cela, nous calculer la sensibilité et la spécificité.

$$Se = P(T+|M) = \frac{P(T+\cap M)}{P(M)} = \frac{VP}{VP+FN} = \frac{70}{70+50} = \frac{7}{12}$$

$$Sp = P(T-|\bar{M}) = \frac{P(T-\cap \bar{M})}{P(\bar{M})} = \frac{VN}{VN+FP} = \frac{350}{350+30} = \frac{35}{38}$$

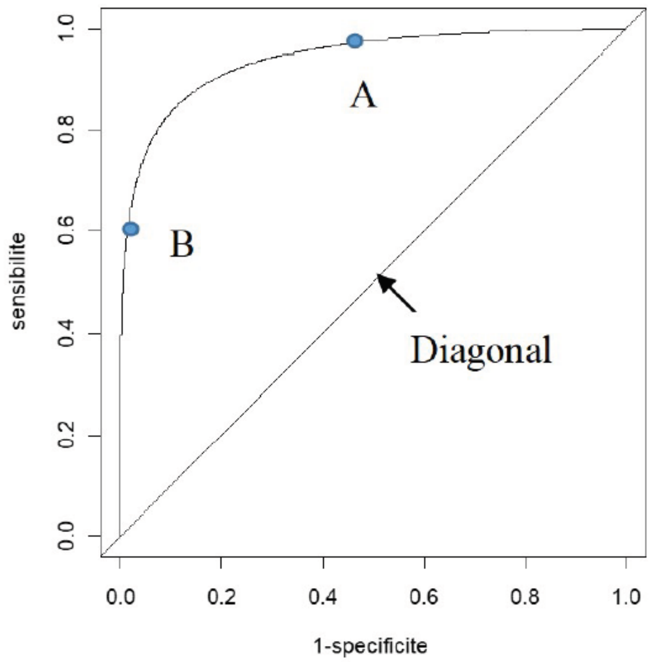
Ainsi, $\frac{35}{38} > \frac{7}{12}$ $donc\ Sp > Se$. Le test est plus spécifique que sensible, c'est-à-dire qu'il ne détecte pas très bien tous les malades, par contre, il détecte assez bien « que les malades ».

PS : j'ai décidé de vous écrire la sensibilité et la spécificité comme ça pour vous puissiez voir que c'est en réalité une application des formules vues dans le chapitre des probabilités. En plus, une fois que vous *avez votre tableau et cette réflexion, il ne vous reste plus qu'à lire aux bonnes intersections dans votre tableau.*

E VRAI La valeur prédictive négative est la probabilité de ne pas avoir de cancer sachant qu'on a été testé négatif. Ici, on l'estime donc par la proportion de personnes qui n'ont pas de cancer chez les hommes ayant été testés négatifs.

Question 9 – Tests Diagnostiques (Coefficient : 2 points) : ACD

Une étude a été mise en place pour évaluer les performances d'un marqueur sanguin pour discriminer les patients ayant un cancer du pancréas de ceux n'ayant pas de cancer. En moyenne les valeurs du marqueur sont plus élevées chez les patients ayant un cancer. La courbe ROC du marqueur pour évaluer sa capacité à discriminer les patients ayant un cancer de ceux n'en ayant pas a été construite et est présentée ci-dessous :



Indiquez la ou les proposition(s) juste(s).

A. La sensibilité du seuil correspondant au point A de la courbe ROC est plus élevée que celle du seuil correspondant au point B.
B. La spécificité du seuil correspondant au point A de la courbe ROC est plus élevée que celle du seuil correspondant au point B.
C. Le seuil correspondant au point A permet d'identifier plus de patients avec un cancer que le seuil correspondant au point B.
D. Le seuil correspondant au point B permet d'identifier plus de patients n'ayant pas de cancer que le seuil correspondant au point A.
E. Le marqueur sanguin est un test qui permet de discriminer parfaitement les patients ayant un cancer de ceux n'en ayant pas.

Pour bien lire une courbe ROC, il faut bien regarder ce que portent les axes : l'axe des abscisses correspond à **1 – spécificité** et l'axe des ordonnées correspond à la **sensibilité**. Chaque point de la courbe représente un seuil donné d'un même test.

A VRAI Le seuil correspondant au point A à une sensibilité d'environ 0,95 alors que le seuil correspondant au point B a une sensibilité d'environ 0,6.

B FAUX Le seuil correspondant au point A à une spécificité d'environ 1 – 0,5 = 0,5 alors que le seuil correspondant au point B a une spécificité d'environ 1 – 0,05 = 0,95.

C VRAI Le seuil au point A a une sensibilité plus élevée : cela signifie qu'il permet de détecter plus de patients avec un cancer que le seuil au point B.

D VRAI Le seuil au point B est celui avec la meilleure spécificité : cela signifie qu'il permet d'éviter les faux positifs et donc qu'il détecte mieux les patients n'ayant pas de cancer.

E FAUX Le marqueur sanguin dont on parle ici est la PSA. En effet, tous ces patients ont été envoyé à un test d'imagerie pour vérifier qu'ils n'avaient pas de cancer parce que leur PSA était élevée. Or, on voit bien qu'une grosse majorité d'entre eux n'avaient finalement pas de cancer, cela signifie que le marqueur sanguin au seuil choisi est un test qui fait beaucoup de faux positifs, il ne discrimine donc pas bien. Vous pouvez aussi vous douter qu'en réalité, il n'existe pas de marqueurs sanguins qui discriminent parfaitement : il n'y a pas de marqueur idéal (à l'heure d'aujourd'hui en tout cas) avec une sensibilité de 1 et une spécificité de 1.