Université Claude Bernard Lyon 1

Tutorat Lyon Est

Unité d’Enseignement 3

Banque de QCM

2013-2022

Fluctuations d’échantillonnage – Estimations ponctuelles et par intervalle de confiance

QUESTIONS et REPONSES

2013 – 2021

Question 1 :

En France, en moyenne, les fumeurs réguliers fument 15 cigarettes par jour, avec une variance de 9. On étudie un échantillon de 81 personnes sélectionnées au hasard.
Soit μ la moyenne théorique du nombre de cigarettes fumées par jour, M l'estimateur de cette moyenne et m l'estimation de cette moyenne dans l'échantillon.

Aides au calcul : $1{,}96 \approx 2$ $\frac{2}{3} \approx 0{,}667$

A. Un intervalle de confiance à 95% de la moyenne du nombre de cigarettes fumées par jour est $ic_{0,95}(\mu) = [14{,}33; 15{,}67]$.
B. Un intervalle de fluctuation à 95% de la moyenne du nombre de cigarettes fumées par jour est $IF_{0,95}(\mu) = [14{,}33; 15{,}67]$.
C. Un intervalle de confiance à 95% de la moyenne du nombre de cigarettes fumées par jour est $ic_{0,95}(\mu) = [14{,}34; 15{,}66]$.
D. Un intervalle de fluctuation à 95% de la moyenne du nombre de cigarettes fumées par jour est $IF_{0,95}(M) = [14{,}33; 15{,}67]$.
E. Un intervalle de fluctuation à 95% de la moyenne du nombre de cigarettes fumées par jour est $IF_{0,95}(M) = [14{,}34; 15{,}66]$.

Question 1 : D

On est dans le cas où on dispose des paramètres théoriques dans la population et on aimerait estimer la moyenne du nombre de cigarettes fumées par jour dans un échantillon : il s'agit donc de calculer un intervalle de fluctuation.
Dans cet exercice, on a une variable aléatoire X modélisant le nombre de cigarettes fumées par jour par un fumeur en France. Cette variable aléatoire suit une loi quelconque :

$$X \to L(\mu = 15\,; \sigma = \sqrt{9} = 3)$$

On peut définir la variable M, modélisant le **nombre moyen** de cigarettes fumées par jour par un fumeur en France calculé sur un échantillon de 81 personnes, telle que :

$$M = \sum_{i=1}^{81} \frac{X_i}{n}$$

$n > 30$, donc par application du TCL, on conclut que M suit approximativement une loi normale de paramètres $M \to N(\mu_M = \mu = 15\,; \sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{81}} = \frac{3}{9} = \frac{1}{3})$.
On est dans un cas où µ est connue, on ne va donc pas chercher à l'estimer par un intervalle de confiance donc **A FAUX** et **C FAUX**.

De plus, comme µ est connue, on calcule un $IF_{0,95}(M)$, pas un $ic_{0,95}(\mu)$ donc **B FAUX**.

Pour connaitre $z_{\alpha/2}$, on utilise la deuxième table de la normale avec $\alpha = 0{,}05$ obtenu à partir du niveau de confiance $(0{,}95 = 1 - \alpha \leftrightarrow \alpha = 0{,}05)$: $z_{\alpha/2} = z_{0,05/2} = z_{0,025} = 1{,}96 \approx 2$
Pour déterminer $z_{\alpha/2}$, on cherche dans la deuxième table de la loi normale pour 0,025. On combine ligne et colonne pour obtenir la valeur 0,025. Ici, on se place à la ligne 0,02 et à la colonne 0,005 (l'addition de ces deux valeurs donnent 0,025). L'intersection entre cette ligne et cette colonne donne $z_{\alpha/2} = 1{,}9600$.

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	$\boxed{1,9600}$	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

Finalement, on fait le calcul :

$$IF_{1-\alpha}(M) = \mu \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

On a : $\mu = 15$ $\sigma^2 = 9$ donc $\sigma = 3$ $n = 81$ $z_{\alpha/2} = 1{,}96 \approx 2$

$$IF_{0,95}(M) = 15 \pm 2 \times \frac{3}{\sqrt{81}} = 15 \pm 2 \times \frac{3}{9} = 15 \pm \frac{6}{9} = 15 \pm \frac{2}{3} = 15 \pm 0{,}667$$

$$IF_{0,95}(M) = [14{,}333; 15{,}667]$$

Lorsqu'on doit arrondir les bornes d'un intervalle, on minore la borne inférieure et on majore la borne supérieure pour garantir le niveau de confiance de l'intervalle. On obtient donc :

$$IF_{0,95}(M) = [14{,}33; 15{,}67]$$

Donc **D VRAI**.

E FAUX $IF_{0,95}(M) = [14{,}34; 15{,}66]$: ici les bornes de l'intervalle ont mal été arrondies, on a majoré la borne inférieure et on a minoré la borne supérieure donc notre intervalle ne permet pas de garantir un niveau de confiance égal à 95%.

Question 2 :

En considérant les notations issues du cours, quelles sont la ou les affirmation(s) vraie(s) ?

A. Un estimateur est une variable aléatoire dont les réalisations sont des estimations.
B. Un paramètre est issu d'une population tandis qu'une estimation est obtenue à partir d'un échantillon.
C. M et F sont des estimateurs biaisés.
D. m est une estimation de μ, et M est son estimateur.
E. p est une estimation de F, et f est son estimateur.

Question 2 : ABD

A VRAI
B VRAI
C FAUX Au contraire, ce sont de bons estimateurs.
D VRAI Il est important de s'en tenir aux conventions de notations : M et F représentent des **estimateur** (ce ne sont pas des valeurs, ce sont des variables aléatoires), m et f représentent des **estimations** (ce sont des valeurs issues d'échantillons), et μ et p représentent des **paramètres** de la population que l'on cherche à estimer.
E FAUX La phrase correcte est : « f est une estimation de p, et F est son estimateur.

Question 3 :

On décide de tester l'efficacité de deux traitements anti-acnéiques : une crème topique et un antibiotique oral. Pour cela, deux échantillons de patients sont constitués. On obtient les résultats suivants :

	Succès	Echecs
Crème topique	80	20
Antibiotique oral	15	10

On note dans cet exercice :

- p_C et p_A les proportions théoriques de succès, respectivement de la crème topique et de l'antibiotique oral, dans les populations
- f_C et f_A les estimations de p_C et p_A calculées à partir des échantillons
- F_C et F_A les estimateurs de p_C et p_A

Aides au calcul : $1{,}96 \approx 2$ $\quad\quad \sqrt{0{,}24} \approx \sqrt{0{,}25}$

A. Un intervalle de fluctuation à 95% de la proportion théorique de succès avec l'antibiotique oral est $IF_{0{,}95}(p_A) = [0{,}4; 0{,}8]$.
B. Un intervalle de confiance à 95% de la proportion théorique de succès avec l'antibiotique oral est $ic_{0{,}95}(p_A) = [0{,}4; 0{,}8]$.
C. Un intervalle de confiance à 95% de la proportion théorique de succès avec la crème topique est $ic_{0{,}95}(p_C) = [0{,}72; 0{,}88]$.
D. La largeur de l'intervalle $[0{,}72; 0{,}88]$ est de 0,08.
E. La précision de l'intervalle $[0{,}72; 0{,}88]$ est de 0,16.

Question 3 : C

A FAUX Un intervalle de fluctuation de X se calcule lorsque que l'on connaît la loi de probabilité de X ainsi que ses paramètres (espérance et variance). Ici, ce n'est pas le cas. On dispose d'un échantillon à partir duquel on a calculé une estimation ponctuelle de la proportion dans la population. On ne peut donc pas calculer un intervalle de fluctuation. En revanche, on peut calculer un **intervalle de confiance** de p_A.

B FAUX On voudrait calculer un intervalle de confiance mais pour le faire, il faut d'abord vérifier la condition d'application : $n \geq 30$. Dans le tableau fourni, on a 15 succès et 10 échecs, soit un total de 25 sujets pour le groupe « antibiotique oral ». Puisque $n = 25 < 30$, on ne peut pas calculer d'intervalle de confiance en utilisant la méthode vue dans le cours de PASS *(pour information, il est possible de calculer cet intervalle de confiance via la loi de Student mais c'est hors programme).*

C VRAI On veut calculer un intervalle de confiance.
Condition d'application : $n = 100 \geq 30$ (on a additionné les succès (80) et les échecs (20) pour le groupe « crème topique ») donc on peut faire notre calcul.

$$ic_{0{,}95}(p_C) = f_C \pm z_{\alpha/2}\sqrt{\frac{f_C(1-f_C)}{n}}$$

$$ic_{0{,}95}(p_C) = \frac{80}{100} \pm 2\sqrt{\frac{0{,}8 \times 0{,}2}{100}} = 0{,}8 \pm 2 \times \frac{\sqrt{0{,}16}}{10} = 0{,}8 \pm \frac{2 \times 0{,}4}{10} = 0{,}8 \pm \frac{0{,}8}{10} = 0{,}8 \pm 0{,}08$$

$$ic_{0{,}95}(p_C) = [0{,}72; 0{,}88]$$

Il faut maintenant vérifier *a posteriori* les conditions de validité de l'intervalle de confiance. Soient f_1 et f_2, respectivement les bornes inférieure et supérieure de l'intervalle calculé :

- $nf_1 = 100 \times 0{,}72 = 72 \geq 5$
- $n(1 - f_1) = 100 \times 0{,}28 = 28 \geq 5$

- $nf_2 = 100 \times 0{,}88 = 88 \geq 5$
- $n(1 - f_2) = 100 \times 0{,}12 = 12 \geq 5$

Les conditions sont respectées : notre intervalle de confiance est valide.

D FAUX $Largeur = borne\ supérieure - borne\ inférieure = 0{,}88 - 0{,}72 = 0{,}16$

E FAUX $Précision = \frac{largeur}{2} = \frac{borne\ supérieure - borne\ inférieure}{2} = \frac{0{,}88-0{,}72}{2} = \frac{0{,}16}{2} = 0{,}08$

Attention donc à ne pas confondre précision et largeur d'un intervalle.

Question 4 :

Dans le cadre du suivi d'une hypercholestérolémie, on réalise un bilan lipidique dans un échantillon de 256 personnes. Le taux moyen de LDL-cholestérol estimé à partir de cet échantillon est de 2,1 g/L, avec un écart-type estimé à 0,67 g/L.

Soit μ le taux moyen théorique de LDL-cholestérol chez les personnes atteintes d'hypercholestérolémie, M l'estimateur de ce taux moyen et m l'estimation de ce taux moyen dans l'échantillon.

Aides au calcul : $1{,}6449 \approx 1{,}65$ $1{,}96 \approx 2$ $\frac{1{,}65}{16} \approx \frac{1{,}6}{16}$ $\frac{0{,}67}{8} \approx 0{,}084$

A. Un intervalle de confiance à 90% du taux moyen théorique de LDL-cholestérol chez les personnes atteintes d'hypercholestérolémie est $ic_{0,90}(\mu) = [2{,}03; 2{,}17]$.

B. Un intervalle de fluctuation à 90% du taux moyen théorique de LDL-cholestérol chez les personnes atteintes d'hypercholestérolémie est $IF_{0,90}(\mu) = [2{,}03; 2{,}17]$.

C. Un intervalle de confiance à 95% du taux moyen théorique de LDL-cholestérol chez les personnes atteintes d'hypercholestérolémie est $ic_{0,95}(\mu) = [2{,}01; 2{,}19]$.

D. Plus le niveau de confiance est élevé, plus l'intervalle de confiance est large.

E. Plus l'échantillon est grand, plus l'intervalle de confiance est étroit.

Question 4 : ACDE

Nous disposons de l'estimation du taux moyen de LDL-cholestérol dans un échantillon de personnes atteintes d'hypercholestérolémie à partir duquel nous voulons donc établir un intervalle de confiance du taux moyen théorique de LDL-cholestérol chez les personnes atteintes d'hypercholestérolémie.

$n = 256 \geq 30$, donc les conditions de validité *a priori* de l'intervalle de confiance sont respectées. Il est possible d'établir cet intervalle.

À partir de l'énoncé, on a : $m = 2{,}1$ $s = 0{,}67$ $n = 256$

A VRAI Le niveau de confiance est $1 - \alpha = 0{,}9$ donc $\alpha = 0{,}1$. On cherche dans la table de la loi normale centrée réduite la valeur de $z_{\alpha/2}$ pour $\alpha = 0{,}1$ donc on cherche la valeur de $z_{0,05}$.

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	$\boxed{1{,}6449}$	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

$$z_{0,05} = 1{,}6449 \approx 1{,}65$$

On peut maintenant établir un intervalle de confiance à 90% :

$$ic_{0,90}(\mu) = m \pm z_{0,05} \times \frac{s}{\sqrt{n}}$$
$$ic_{0,90}(\mu) = 2,1 \pm 1,65 \times \frac{0,67}{\sqrt{256}} = 2,1 \pm 1,65 \times \frac{0,67}{16}$$
$$ic_{0,90}(\mu) = 2,1 \pm \frac{1,6}{16} \times 0,67 = 2,1 \pm 0,1 \times 0,67 = 2,1 \pm 0,067$$
$$ic_{0,90}(\mu) = [2,033; 2,167] = [2,03; 2,17]$$

En effet, on minore la borne inférieure et on majore la borne supérieure pour garantir le niveau de confiance.

B FAUX On ne cherche pas un intervalle de fluctuation.

C VRAI Le niveau de confiance est $1 - \alpha = 0,95$ donc $\alpha = 0,05$. Dans la table, on trouve $z_{\alpha/2} = z_{0,025} = 1,96 \approx 2$.
On peut établir un intervalle de confiance à 95% :

$$ic_{0,95}(\mu) = m \pm z_{0,025} \times \frac{s}{\sqrt{n}}$$
$$ic_{0,95}(\mu) = 2,1 \pm 2 \times \frac{0,67}{\sqrt{256}} = 2,1 \pm 2 \times \frac{0,67}{16} = 2,1 \pm \frac{0,67}{8}$$
$$ic_{0,95}(\mu) = 2,1 \pm 0,084$$
$$ic_{0,95}(\mu) = [2,016; 2,184] = [2,01; 2,19]$$

D VRAI Plus le niveau de confiance est élevé, plus $z_{\alpha/2}$ est grand donc plus l'intervalle est large. C'est ce qu'on remarque avec les deux intervalles calculés juste avant.

E VRAI Plus l'échantillon est grand, plus n est élevé donc $\frac{s}{\sqrt{n}}$ est petit donc plus l'intervalle est étroit.

Question 5 :

Parmi les phrases suivantes, **quelles sont celles où l'on trouve (entre autres) d'estimations** ?

A. "Dans la population française, 1% des naissances révèlent un Trouble du Spectre Autistique (TSA). On s'intéresse à un échantillon de 100 naissances."
B. "Aux États-Unis, l'obésité est une pathologie très répandue. Dans un échantillon de 1200 sujets américains, 497 sont en situation d'obésité."
C. "Dans un groupe de 85 étudiants en PASS considérés représentatifs de la population générale de PASS, 64 expliquent travailler plus de 13h par jour".
D. "La mortalité routière par année en Auvergne-Rhône-Alpes peut être modélisé par une variable aléatoire gaussienne, de moyenne 396 morts/an et d'écart-type 34 morts/an."
E. "En France, 14% des patients admis aux urgences sont des personnes de plus de 70 ans. À partir d'un groupe de 36 patients de plus de 70 ans admis aux urgences, on observe que 12 d'entre eux présentent une Hypertension artérielle (HTA)."

Question 5 : BCE

Dans cet exercice (qui n'est pas du tout un exercice type), on vous demande uniquement de comprendre quelles sont les données que l'on vous fournies.

A FAUX Dans cette phrase, on nous explique que la prévalence de 1% de naissances avec TSA concerne la population française. Puisque cette donnée provient de la population, alors il s'agit d'un paramètre de la population et non pas d'une estimation. **ATTENTION !** Le fait que l'on nous parle d'un échantillon ne signifie pas qu'il s'agit forcément d'une estimation. Ce qu'il est important de savoir, c'est **d'où provient** l'information : de la population ou de l'échantillon ?

B VRAI Les 497 sujets en situation d'obésité proviennent de l'échantillon de 1200 sujets américains. Puisque la donnée provient d'un échantillon, alors il s'agit d'une estimation.

C VRAI On nous explique ici que l'on dispose d'un groupe de 85 étudiants en PASS qui eux-mêmes appartiennent à une population générale de PASS. Le groupe de 85 étudiants en PASS est donc un échantillon. De plus, 64 étudiants de cet échantillon travaillent plus de 13h par jour. Puisque cette donnée provient de l'échantillon, alors il s'agit d'une estimation.

D FAUX Ici, on nous donne les paramètres d'une variable aléatoire (sa moyenne et son écart-type). Comme leur nom l'indique ce sont des paramètres, ils proviennent donc de la population.

E VRAI Attention ici, on dispose d'un paramètre : les 14% de patients admis aux urgences de plus de 70 ans est une donnée au niveau de la population en France. Mais on dispose également d'une estimation : la proportion de sujets admis aux urgences de plus de 70 ans présentant une HTA. En effet, les 12 personnes présentant une HTA sont issues de l'échantillon de 36 sujets.

Question 6 :

On détermine l'IMC (Indice de Masse Corporelle) dans un échantillon de 100 adultes sélectionnés au hasard. On considère qu'une personne a un poids insuffisant quand son IMC est inférieur à 18,5 tandis qu'une personne est en situation de surpoids lorsque son IMC est supérieur à 25. Dans cet échantillon, 10 personnes ont un poids insuffisant tandis que 30 personnes sont en situation de surpoids.
On note dans cet exercice :

- p_I et p_S les proportions théoriques, respectivement de personnes avec un poids insuffisant et de personnes en surpoids, dans la population
- f_I et f_S les estimations de p_I et p_S dans l'échantillon
- F_I et F_S les estimateurs de p_I et p_S

Aides au calcul : $1{,}96 \approx 2$ $\sqrt{0{,}21} \approx 0{,}46$

A. Un intervalle de confiance à 95% de l'estimateur de la proportion théorique de personnes avec un poids insuffisant est $ic_{0,95}(F_I) = [0{,}04; 0{,}16]$.
B. Un intervalle de confiance à 95% de la proportion théorique de personnes avec un poids insuffisant est $ic_{0,95}(p_I) = [0{,}04; 0{,}16]$.
C. Un intervalle de confiance à 95% de la proportion théorique de personnes en surpoids est $ic_{0,95}(p_S) = [0{,}20; 0{,}40]$.
D. Un intervalle de confiance à 95% de la proportion théorique de personnes en surpoids est $ic_{0,95}(p_S) = [0{,}21; 0{,}39]$.
E. Pour obtenir une précision de 0,046 pour l'intervalle de confiance à 95% de la proportion théorique de personnes en surpoids, il faudrait un échantillon 4 fois plus grand.

Question 6 : CE

Les données dont nous disposons sont : $n = 100$ $f_I = \frac{10}{100} = 0{,}1$ $f_S = \frac{30}{100} = 0{,}3$

Nous avons les estimations des proportions de personnes avec un poids insuffisant et de personnes en surpoids dans l'échantillon : il s'agit donc ici d'établir un intervalle de confiance de la proportion théorique. Nous sommes dans le cas où $n = 100 \geq 30$ donc les conditions *a priori* sont respectées, on peut établir l'intervalle de confiance.

A FAUX On recherche un intervalle de confiance de la proportion théorique p_I (et pas de l'estimateur de cette proportion F_I) tel que $ic_{0,95}(p_I)$.

B FAUX On recherche bien $ic_{0,95}(p_I)$ et les conditions *a priori* sont respectées.

$$ic_{0,95}(p_I) = f_I \pm z_{\alpha/2} \times \sqrt{\frac{f_I(1 - f_I)}{n}}$$

$z_{\alpha/2} = 1{,}96 \approx 2$ donc :

$$ic_{0,95}(p_I) = 0,1 \pm 2 \times \sqrt{\frac{0,1 \times 0,9}{100}} = 0,1 \pm 2 \times \sqrt{\frac{0,09}{100}} = 0,1 \pm 2 \times \frac{0,3}{10} = 0,1 \pm \frac{0,6}{10} = 0,1 \pm 0,06$$
$$ic_{0,95}(p_I) = [0,04; 0,16]$$

Il faut maintenant vérifier les conditions *a posteriori* de l'intervalle (avec f_1 et f_2 respectivement les bornes inférieures et supérieures de l'intervalle) :

- $nf_1 = 100 \times 0,04 = 4 < 5$ donc cette condition n'est pas remplie (on pourrait s'arrêter ici)
- $n(1 - f_1) = 100 \times 0,96 = 96 \geq 5$
- $nf_2 = 100 \times 0,16 = 16 \geq 5$
- $n(1 - f_2) = 100 \times 0,84 = 84 \geq 5$

Les conditions à posteriori ne sont pas respectées donc $ic_{0,95}(p_I) = [0,04; 0,16]$ n'est pas un intervalle de confiance valide.

C VRAI On recherche $ic_{0,95}(p_S)$.

$$ic_{0,95}(p_S) = f_S \pm z_{\alpha/2} \times \sqrt{\frac{f_S(1 - f_S)}{n}}$$

$$ic_{0,95}(p_S) = 0,3 \pm 2 \times \sqrt{\frac{0,3 \times 0,7}{100}} = 0,3 \pm 2 \times \frac{\sqrt{0,21}}{10} = 0,3 \pm 2 \times \frac{0,46}{10} = 0,3 \pm 0,092$$
$$ic_{0,95}(p_S) = [0,208; 0,392] = [0,20; 0,40]$$

On minore la borne inférieure et on majore la borne supérieure.
On vérifie maintenant les conditions à posteriori (avec f_1 et f_2 respectivement les bornes inférieures et supérieures de l'intervalle) :

- $nf_1 = 100 \times 0,20 = 20 \ \geq 5$
- $n(1 - f_1) = 100 \times 0,80 = 80 \geq 5$
- $nf_2 = 100 \times 0,40 = 40 \geq 5$
- $n(1 - f_2) = 100 \times 0,60 = 60 \geq 5$

Les conditions à posteriori sont remplies, cet intervalle de confiance est valide.

D FAUX Dans cet item, on a majoré la borne inférieure et minoré la borne supérieure donc le niveau de confiance (0,95) n'est pas garanti.

E VRAI La précision est la demi-largeur de l'intervalle de confiance.
Si l'échantillon est 4 fois plus grand, n = 400. On a alors une précision de :

$$i = z_{\alpha/2} \times \sqrt{\frac{f_S(1 - f_S)}{n}}$$

$$i = 2 \times \sqrt{\frac{0,3 \times 0,7}{400}} = 2 \times \frac{\sqrt{0,21}}{20} = 2 \times \frac{0,46}{20} = \frac{0,46}{10} = 0,046$$

On peut aussi résoudre cet exercice en calculant le nombre de sujets minimal à inclure pour avoir une précision $i = 0,046$.

$$i = z_{\alpha/2} \times \sqrt{\frac{f_S(1 - f_S)}{n}} \leftrightarrow n = \frac{f_S(1 - f_S) \times z_{\alpha/2}^2}{i^2}$$
$$n = \frac{0,3 \times 0,7 \times 4}{0,046^2} = \frac{0,21 \times 4}{0,0021} = 400$$

Question 7 :

Un bilan hormonal permet de doser les différentes concentrations en hormones dans l'organisme dont notamment la sapinine. Il s'agit d'une hormone de l'esprit de Noël libérée dans le sang par la décoration d'un sapin de Noël. Soit la variable aléatoire X modélisant la concentration de sapinine dans la population française. On considère que X suit une loi Gaussienne de moyenne 140 mmol/L et d'écart-type 3 mmol/L. On constitue aléatoirement un échantillon de 144 individus français.
On notera μ et σ la moyenne et l'écart-type théorique dans la population ; m et s les estimations de μ et σ, réalisées à partir de l'échantillon ; et M l'estimateur de la moyenne.

Aides au calcul : $1{,}4395 \approx 1{,}5$ $\quad 1{,}6449 \approx 1{,}6$ $\quad 1{,}96 \approx 2$ $\quad 0{,}6745 \approx 0{,}67$

A. Un intervalle de fluctuation à 90% de X est $[135{,}2; 144{,}8]$.
B. Un intervalle de confiance de la concentration de sapinine moyenne théorique à 95% est $[139{,}5; 140{,}5]$.
C. Un intervalle de fluctuation à 85% de l'estimateur de la concentration moyenne de sapinine est $[139{,}7; 140{,}3]$.
D. Un intervalle de fluctuation à 85% de l'estimateur de la concentration moyenne de sapinine est $[139{,}6; 140{,}4]$.
E. Approximativement 75% des individus ont une concentration de sapinine au moins supérieure à environ 138 mmol/L.

Question 7 : ADE

$X \to N(140; 3)$. On dispose des valeurs théoriques de la concentration de sapinine dans la population, on cherche donc un intervalle de fluctuation.

A VRAI On cherche un intervalle de fluctuation de la variable aléatoire modélisant la concentration de sapinine, à la confiance 90% (donc $\alpha = 10\%$ et $z_{\alpha/2} = 1{,}6449 \approx 1{,}6$) :

$$IF_{0,9}(X) = \mu \pm z_{\alpha/2} \times \sigma = 140 \pm 1{,}6 \times 3 = 140 \pm 4{,}8 = [135{,}2; 144{,}8]$$

B FAUX On ne cherche pas un intervalle de confiance.

C FAUX On cherche un intervalle de fluctuation de l'estimateur de la concentration **moyenne** de sapinine, à la confiance 85% (donc $\alpha = 15\%$ et $z_{\alpha/2} = z_{0,075} = 1{,}4395 = 1{,}5$) :

$$IF_{0,85}(M) = \mu \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} = 140 \pm 1{,}5 \times \frac{3}{\sqrt{144}} = 140 \pm \frac{1{,}5 \times 3}{12} = 140 \pm \frac{1{,}5}{4}$$
$$IF_{0,85}(M) = 140 \pm 0{,}375 = [139{,}625; 140{,}375] = [139{,}6; 140{,}4]$$

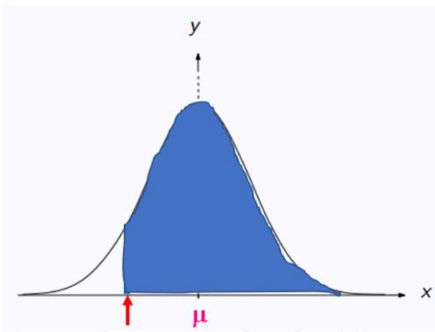
Attention à bien minorer la borne inférieure et à majorer la borne supérieure.

D VRAI Cf. C

E VRAI X suit une loi normale donc on cherche la valeur désignée par la flèche rouge ci-dessous pour laquelle la surface colorée en bleu équivaut à environ 75% des individus.

On lit dans la table la valeur de z pour un seuil de 75% :

0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0450	1,0407
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192
0,18	0,9154	0,9116	0,9078	0,9040	0,9002	0,8965	0,8927	0,8890	0,8853	0,8816
0,19	0,8779	0,8742	0,8705	0,8669	0,8633	0,8596	0,8560	0,8524	0,8488	0,8452
0,20	0,8416	0,8381	0,8345	0,8310	0,8274	0,8239	0,8204	0,8169	0,8134	0,8099
0,21	0,8064	0,8030	0,7995	0,7961	0,7926	0,7892	0,7858	0,7824	0,7790	0,7756
0,22	0,7722	0,7688	0,7655	0,7621	0,7588	0,7554	0,7521	0,7488	0,7454	0,7421
0,23	0,7388	0,7356	0,7323	0,7290	0,7257	0,7225	0,7192	0,7160	0,7128	0,7095
0,24	0,7063	0,7031	0,6999	0,6967	0,6935	0,6903	0,6871	0,6840	0,6808	0,6776
0,25	0,6745	0,6713	0,6682	0,6651	0,6620	0,6588	0,6557	0,6526	0,6495	0,6464
0,26	0,6433	0,6403	0,6372	0,6341	0,6311	0,6280	0,6250	0,6219	0,6189	0,6158
0,27	0,6128	0,6098	0,6068	0,6038	0,6008	0,5978	0,5948	0,5918	0,5888	0,5858
0,28	0,5828	0,5799	0,5769	0,5740	0,5710	0,5681	0,5651	0,5622	0,5592	0,5563
0,29	0,5534	0,5505	0,5476	0,5446	0,5417	0,5388	0,5359	0,5330	0,5302	0,5273
0,30	0,5244	0,5215	0,5187	0,5158	0,5129	0,5101	0,5072	0,5044	0,5015	0,4987
0,31	0,4959	0,4930	0,4902	0,4874	0,4845	0,4817	0,4789	0,4761	0,4733	0,4705



On cherche la valeur de la concentration de sapinine pour laquelle 75% des individus ont une valeur supérieure à celle-ci :

$$\mu - z \times \sigma = 140 - 0{,}67 \times 3 \approx 140 - 2{,}01 \approx 137{,}99 \approx 138$$

Pour répondre à cet item, on aurait aussi pu calculer la probabilité $P(X > 138)$ et on serait tombé sur 75%.

Question 8 :

On dit souvent que pour réussir en PASS, il vaut mieux être célibataire (c'est-à-dire être un Yacobien) plutôt qu'en couple (c'est-à-dire être un Kaissien). On cherche à savoir si les notes varient réellement suivant que l'on soit seul ou en couple.
Pour cela, on réalise une étude sur un échantillon de 36 Yacobiens (étudiants seuls) et sur un autre échantillon de 25 individus Kaissiens (étudiants en couple).
On note m_Y et m_K respectivement les moyennes estimées des notes des étudiants étant seuls et en couple au concours. Et on note s_Y et s_K respectivement les estimations des écarts-types.

	Moyennes des notes estimées	Ecarts-types estimés
Yacobiens	17	2
Kaissiens	10	1

Aides au calcul : $1{,}96 \approx 2$ $\quad 2{,}5758 \approx 2{,}58$

A. Un intervalle de confiance à 95% de la moyenne des notes obtenues au concours chez les Yacobiens est $ic_{0,95}(m_Y) = [16{,}34\,;17{,}66]$.
B. Un intervalle de fluctuation à 95% de la moyenne des notes obtenues au concours chez les Kaissiens est $IF_{0,95}(M_K) = [9{,}6\,;10{,}4]$.
C. Un intervalle de fluctuation à 99% de la moyenne des notes obtenues au concours chez les Yacobiens est $IF\ _{0,99}(M_Y) = [16{,}14\,;17{,}86]$.
D. Un intervalle de confiance à 99% de la moyenne des notes obtenues au concours chez les Kaissiens est : $ic_{0,99}(\mu_K) = [9{,}486\,;10{,}514]$.
E. Un intervalle de confiance à 99% de la moyenne des notes obtenues au concours chez les Yacobiens est : $ic_{0,99}(\mu_Y) = [16{,}14\,;17{,}86]$.

Question 8 : E

A FAUX Dans cet exercice, $\mu_K\,;\mu_Y\,;\sigma_K\,;\sigma_Y$ sont inconnues, nous n'avons que leurs estimations.
ATTENTION : On ne peut donc calculer que des Intervalles de confiance et donc pas d'intervalles de fluctuation !
De plus, à partir de la moyenne estimée, on cherche à calculer un intervalle qui a de grandes chances de contenir μ avec un risque d'erreur à déterminer selon ce qui est demandé (risque α).
On cherche donc des **ic $_{1-\alpha}$ (μ).**

➔ On élimine directement les items A B et C.

B FAUX Cf. A

C FAUX Cf. A

D FAUX On cherche bien un intervalle de confiance de μ, cependant pour pouvoir effectuer le calcul, il faut que n ≥ 30 (dans le cadre de la PASS), ce n'est pas le cas ici… on peut tout de suite éliminer l'item.

E VRAI On cherche un intervalle de confiance ? Oui. n≥ 30 ? oui
On peut donc effectuer le calcul de cet intervalle de confiance à la confiance 99% (donc $\alpha = 1\%$ et $z_{\alpha/2} = z_{0,005} = 2{,}5758 = 2{,}58$) avec la formule suivante :

$$ic_{0,99}(\mu_y) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}} = 17 \pm 2{,}58 \times \frac{2}{\sqrt{36}} = 17 \pm 2{,}58 \times \frac{1}{3} = 17 \pm 0{,}86$$

$$ic_{0,99}(\mu_y) = [16{,}14; 17{,}86]$$

Question 9 :

Des chercheurs américains s'intéressent à la myopathie de Duchenne. Pour leurs études, ils constituent deux échantillons aléatoires A et B. Dans l'échantillon A de 36 individus, 20% des personnes sont atteintes de la maladie. Dans l'échantillon B de 64 individus, ce pourcentage est de 50%.

On note :

- ➔ p_A et p_B, les proportions théoriques de malades dans les populations dont sont issus les échantillons A et B respectivement
- ➔ f_A et f_B, les estimations de p_A et p_B réalisées à partir des échantillons A et B respectivement
- ➔ F_A et F_B, les estimateurs des proportions p_A et p_B

Aides au calcul : $0{,}1333 \approx 0{,}13$ $1{,}96 \approx 2$

A. Un intervalle de confiance à 95% de p_A est $ic_{0,95}(p_A) = [0{,}07 ; 0{,}33]$.
B. Un intervalle de confiance à 95% de p_B est $ic_{0,95}(p_B) = [0{,}375 ; 0{,}625]$.
C. Un intervalle de confiance à 95% de f_A est $ic_{0,95}(f_A) = [0{,}07 ; 0{,}33]$.
D. Un intervalle de fluctuation à 95% de F_B est $IF_{0,95}(F_B) = [0{,}375 ; 0{,}625]$.
E. Plus le risque de première espèce est élevé, plus l'intervalle est étroit.

Question 9 : BE

Dans cet exercice, on dispose de données obtenues à partir d'échantillons, donc on ne peut calculer que des intervalles de confiance de $\boldsymbol{p}$.

De plus, la taille des 2 échantillons est supérieure à 30 donc la condition *a priori* est respectée.

A FAUX À partir de l'énoncé, on a : $\boldsymbol{n_A = 36}$ $\boldsymbol{f_A = 0,2}$. On veut calculer un intervalle de confiance à 95%, soit $\boldsymbol{z_{\alpha/2} = 1,96 \approx 2}$:

$$\boldsymbol{ic_{0,95}(p_A) = f_A \pm z_{\alpha/2} \times \frac{\sqrt{f_A(1-f_A)}}{\sqrt{n_A}} = 0,2 \pm 2 \times \frac{\sqrt{0,2 \times 0,8}}{\sqrt{36}} = 0,2 \pm 2 \times \frac{0,4}{6} = 0,2 \pm \frac{0,4}{3}}$$
$$\boldsymbol{= 0,2 \pm 0,13 = [0,07; 0,33]}$$

MAIS, il faut aussi verifier les conditions *A POSTERIORI* (avec $\boldsymbol{f_1}$ et $\boldsymbol{f_2}$ les bornes inférieure et supérieure de l'intervalle) :

- $\boldsymbol{n_A f_1 = 2,52 < 5}$: Condition non respectée
- $\boldsymbol{n_A(1-f_1) > 5}$
- $\boldsymbol{n_A f_2 > 5}$
- $\boldsymbol{n_A(1-f_2) > 5}$

L'intervalle calculé n'est pas valide.

B VRAI À partir de l'énoncé, on a : $\boldsymbol{n_B = 64}$ $\boldsymbol{f_B = 0,5}$. On veut calculer un intervalle de confiance à 95%, soit $\boldsymbol{z_{\alpha/2} = 1,96 \approx 2}$

$$\boldsymbol{ic_{0,95}(p_B) = f_B \pm z_{\alpha/2} \times \frac{\sqrt{f_B(1-f_B)}}{\sqrt{n_B}} = 0,5 \pm 2 \times \frac{\sqrt{0,5 \times 0,5}}{\sqrt{64}} = 0,5 \pm 2 \times \frac{0,5}{8} = 0,5 \pm \frac{1}{8}}$$
$$\boldsymbol{= 0,5 \pm 0,125 = [0,375; 0,625]}$$

MAIS, il faut aussi verifier les conditions *A POSTERIORI* (avec $\boldsymbol{f_1}$ et $\boldsymbol{f_2}$ les bornes inférieure et supérieure de l'intervalle) :

- $\boldsymbol{n_B f_1 > 5}$
- $\boldsymbol{n_B(1-f_1) > 5}$
- $\boldsymbol{n_B f_2 > 5}$
- $\boldsymbol{n_B(1-f_2) > 5}$

Ici, toutes les conditions sont respectées, donc l'intervalle calculé est valide.

C FAUX Un intervalle de confiance se calcule pour un paramètre et non pas pour une estimation.

D FAUX On ne peut pas calculer d'intervalle de fluctuation avec les données de l'énoncé.

E VRAI Si α augmente, la précision de l'intervalle diminue et donc l'intervalle est plus petit.

Question 10 :

La prévalence du cancer du sein chez les femmes françaises de plus de 50 ans est de $p = 10\%$. On considère un échantillon de 400 femmes correspondant à ces caractéristiques, pour lequel on mesure la fréquence f de cancers du sein. Soit F l'estimateur de p.

Aide au calcul : $1{,}96 \approx 2$

A. L'intervalle de confiance à 95% de f est $ic_{0,95}(f) = [0{,}07\,; 0{,}13]$.
B. L'intervalle de confiance à 95% de p est $ic_{0,95}(p) = [0{,}07\,; 0{,}13]$.
C. L'intervalle de fluctuation 95% de F est $IF_{0,95}(F) = [0{,}07\,; 0{,}13]$.
D. L'intervalle de fluctuation à 95% de p est $IF_{0,95}(p) = [0{,}07\,; 0{,}13]$.
E. p est une estimation de f par l'estimateur F.

Question 10 : C

Ici, on connait le paramètre p (= proportion théorique) dans la population. On peut donc calculer un intervalle de fluctuation de l'estimateur d'une proportion F. F est une variable aléatoire modélisant la proportion de femmes françaises étant atteintes du cancer du sein à partir d'un échantillon de 400 sujets.

A FAUX On ne peut pas calculer d'intervalle de confiance à partir des données de l'énoncé. De plus, un intervalle de confiance se calcule pour un paramètre et non pas une estimation.

B FAUX On ne peut pas calculer d'intervalle de confiance à partir des données de l'énoncé.

C VRAI À partir de l'énoncé, on a : $n = 400 \quad p = 0{,}1$. On calcule l'intervalle de fluctuation de F à la confiance 95%, soit $z_{\alpha/2} = 1{,}96 \approx 2$:

$$IF_{0,95}(F) = p \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} = 0{,}1 \pm 2\sqrt{\frac{0{,}1 \times 0{,}9}{400}} = 0{,}1 \pm 2 \times \frac{0{,}3}{20} = 0{,}1 \pm 0{,}03 = [0{,}07; 0{,}13]$$

D FAUX Un intervalle de fluctuation se calcule pour un estimateur et non pas pour un paramètre.

E FAUX La phrase correcte est : f est une estimation de p par l'estimateur F.

Question 11 :

On s'interroge sur le niveau scolaire des élèves français en fin de 3ème. Sur un échantillon de 36 élèves représentatifs des élèves de 3ème français, 20% ne savent pas situer la Chine sur une carte. On sait par ailleurs que les notes obtenues au brevet des collèges par les élèves de 3ème français peuvent être modélisées par une loi normale de moyenne 12 et d'écart-type de 3.

Aide au calcul : $1{,}96 \approx 2 \qquad \frac{4}{3} \approx 1{,}3$

A. Un intervalle de confiance de la proportion des élèves de 3ème français ne sachant pas situer la Chine sur une carte est $[0{,}066\,; 0{,}334]$ au risque $\alpha = 5\%$.
B. L'intervalle de confiance de la proportion des élèves de 3ème français ne sachant pas situer la Chine sur une carte n'est pas valide.
C. L'intervalle contenant 95% des notes obtenues au brevet par les élèves de 3ème français est $[11\,; 13]$.
D. L'intervalle de fluctuation de la moyenne des notes obtenues au brevet par les élèves de 3ème français sur un échantillon de taille $n = 36$ au risque $\alpha = 5\%$ est $[11\,; 13]$.

E. L'intervalle de fluctuation de la moyenne des notes obtenues au brevet par les élèves de 3ème français sur un échantillon de taille $n = 36$ au risque $\alpha = 5\%$ est $[9 ; 15]$.

Question 11 : BD

Pour cet exercice, il faut bien comprendre les données issues de l'énoncé. On peut les diviser en 2 :

- D'un côté, on nous parle de proportion, et plus précisément de la proportion d'élèves de 3ème français ne sachant pas situer la Chine sur une carte. Cette proportion vaut $f = 0{,}2$ et est obtenue à partir d'un échantillon. On peut donc calculer un intervalle de confiance d'une proportion.
- D'un autre côté, on nous parle des notes obtenues au brevet des collèges par les élèves de 3ème français. Ces notes sont modélisées par une variable aléatoire X suivant une loi normale : $X \to N(\mu_X = 12; \sigma_X = 3)$. On peut donc calculer un intervalle de fluctuation d'une variable aléatoire.

L'échantillon utilisé a une taille de $n = 36$. Puisque sa taille est supérieure à 30, la condition *a priori* est validée.

A FAUX Ici, on connait la proportion sur un échantillon et on cherche une estimation de la proportion dans la population générale : il s'agit donc de calculer un intervalle de confiance.

$$ic_{0,95}(p) = f \pm z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = 0{,}2 \pm 2\sqrt{\frac{0{,}2 \times 0{,}8}{36}} = 0{,}2 \pm 2\sqrt{\frac{0{,}16}{36}} = 0{,}2 \pm 2 \times \frac{0{,}4}{6}$$
$$= 0{,}2 \pm 0{,}13333 \ldots = [0{,}066; 0{,}334]$$

/!\ Attention à bien minorer la borne inférieure et majorer la borne supérieure /!\

Il faut maintenant vérifier les conditions *a posteriori* pour les bornes de l'intervalle que sont $f_1 = 0{,}07$ et $f_2 = 0{,}33$:

- $\boldsymbol{nf_1 = 36 \times 0,07 = 2,52 < 5}$: Condition non validée
- $\boldsymbol{n(1-f_1) > 5}$
- $\boldsymbol{nf_2 > 5}$
- $\boldsymbol{n(1-f_2) > 5}$

Les conditions ne sont pas vérifiées aux bornes de l'intervalle, on ne peut donc rien dire sur l'intervalle de confiance de p avec les connaissances au programme de PASS.

B VRAI Cf. A

C FAUX À présent, on souhaite calculer un intervalle de fluctuation de la variable aléatoire X qui modélise les notes obtenues au brevet par les élèves de 3ème français, à la confiance 95% (donc $\alpha = 5\%$ et $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$IF_{0,95}(X) = \mu_X \pm z_{\frac{\alpha}{2}} \times \sigma_X = 12 \pm 2 \times 3 = 12 \pm 6 = [6; 18]$$

D VRAI On s'intéresse maintenant à la variable aléatoire M modélisant la moyenne des notes obtenues au brevet par les de 3ème français. Grâce à la loi de l'estimateur d'une moyenne, on peut dire que M suit une loi normale : $M \to N\left(\mu_M = \mu_X; \sigma_M = \frac{\sigma_X}{\sqrt{n}}\right)$. On chercher à calculer un intervalle de fluctuation de M au risque $\alpha = 5\%$ (donc $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$IF_{0,95}(M) = \mu_X \pm z_{\alpha/2} \times \frac{\sigma_X}{\sqrt{n}} = 12 \pm 2 \times \frac{3}{\sqrt{36}} = 12 \pm 2 \times \frac{3}{6} = 12 \pm 1 = [11; 13]$$

E FAUX Cf. D

Question 12 :

Homo Floriesensis est un hominidé ayant vécu entre -95 000 et -60 000 ans avant notre ère sur l'île de Flores. On souhaite connaître sa taille moyenne. Pour cela, on mesure la taille d'un échantillon de 144 squelettes découverts sur l'île appartenant à cette espèce. On obtient une taille moyenne de 110 cm sur l'échantillon, avec un écart-type estimé de 24 cm.

Aides au calcul : $1{,}4395 \approx 1{,}44 \quad 1{,}6449 \approx 1{,}65$

A. L'intervalle de confiance de la taille moyenne de l'Homo Floriesensis au risque $\alpha = 15\%$ est $[107{,}12\,;112{,}88]$.
B. L'intervalle de confiance de la taille moyenne de l'Homo Floriesensis au risque $\alpha = 15\%$ est $[107{,}92\,;112{,}08]$.
C. L'intervalle de confiance de la taille moyenne de l'Homo Floriesensis au risque $\alpha = 10\%$ est $[106{,}72\,;113{,}28]$.
D. Si l'on veut obtenir une p récision $i = 2$ au risque $\alpha = 5\%$, il faut inclure environ 576 squelettes dans l'échantillon.
E. Si on quadruple la taille de l'échantillon, on divise par 2 la largeur de l'intervalle.

Question 12 : ACDE

A VRAI L'intervalle de confiance de la taille moyenne théorique de l'Homo Floriesensis au risque $\alpha = 15\%$ (donc $z\alpha_{/2} = z_{0,075} = 1{,}4395 \approx 1{,}44$) s'obtient par la formule :

$$ic_{0,85}(\mu) = m \pm z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} = 110 \pm 1{,}44 \times \frac{24}{\sqrt{144}} = 110 \pm 1{,}44 \times \frac{24}{12} = 110 \pm 1{,}44 \times 2$$
$$= 110 \pm 2{,}88 = [107{,}12; 112{,}88]$$

B FAUX Cf. A

C VRAI On calcule le même intervalle mais au risque $\alpha = 10\%$ (donc $z\alpha_{/2} = z_{0,05} = 1{,}6449 \approx 1{,}65$) :

$$ic_{0,90}(\mu) = 110 \pm 1{,}65 \times 2 = [106{,}7; 113{,}3]$$

D VRAI La précision i d'un intervalle correspond à la moitié de sa largeur : $i = \frac{z_{\alpha/2} \times s}{\sqrt{n}}$.

Donc $n = \left(\frac{z_{\alpha/2} \times 2}{i}\right)^2 \approx \left(\frac{24 \times 2}{2}\right)^2 \approx 24^2 \approx 576$. Donc n doit être supérieur ou égal à 576.

E VRAI Si la taille de l'échantillon est 4n :

$$Largeur(4n) = \frac{z_{\alpha/2} \times s}{\sqrt{4n}} = \frac{z_{\alpha/2} \times s}{2\sqrt{n}} = \frac{Largeur(n)}{2}$$

Question 13 :

Dans la population d'étudiants en PASS en France, les ¾ des étudiants ont une préférence pour les cookies du CROUS par rapport à ceux de leur mère. On considère un échantillon de 2500 étudiants en PASS français, chez lesquelles on mesure la prise de poids durant 3 mois. On observe que la moitié des étudiants en PASS de l'échantillon prennent plus de 5 kg durant cette période.
Soit X la variable aléatoire : « nombre d'étudiants en PASS préférant les cookies du CROUS sur l'échantillon ».
Soit Y la variable aléatoire : « nombre d'étudiants en PASS ayant pris 5 kg en 3 mois sur l'échantillon ».

Aides aux calculs : $\sqrt{3} \approx 1{,}6 \qquad 1{,}96 \approx 2 \qquad 2{,}5758 \approx 2{,}6 \qquad 2{,}3263 \approx 2{,}3$

A. X suit approximativement une loi normale $N(\mu_X = 0{,}75\,; \sigma_X = 0{,}008)$.
B. L'intervalle de fluctuation de la proportion d'étudiants préférant les cookies du CROUS sur un échantillon de taille $n = 2500$ au risque $\alpha = 5\%$ est $[0{,}59\,;0{,}91]$.
C. L'intervalle de pari de la proportion d'étudiants préférant les cookies du CROUS sur un échantillon de taille $n = 2500$ au risque alpha $\alpha = 5\%$ est $[0{,}734; 0{,}766]$.
D. L'intervalle de pari de la proportion d'étudiants ayant pris 5 kg en 3 mois sur l'échantillon au risque $\alpha = 5\%$ est $[0{,}48\,;0{,}52]$.
E. L'intervalle de confiance de la proportion d'étudiants ayant pris 5 kg en 3 mois sur l'échantillon au risque $\alpha = 1\%$ est $[0{,}477\,;0{,}523]$.

Question 13 : C

Il faut bien comprendre les différents éléments de l'énoncé :

- Concernant le nombre d'étudiant en PASS préférant les cookies du CROUS, nous savons qu'ils représentent ¾ des étudiants en France. Nous avons donc le paramètre d'une proportion (= proportion théorique) qui vaut $p = \frac{3}{4}$. Nous pouvons donc calculer des intervalles de fluctuation.
- Concernant le nombre d'étudiants en PASS ayant pris 5 kg en 3 mois, nous avons des données provenant de l'échantillon de 2500 étudiants en PASS. Nous avons donc l'estimation d'une proportion (= proportion observée) qui vaut $f = 0{,}5$. Nous pouvons donc calculer des intervalles de confiance.

A FAUX X correspond à la répétition de 2500 épreuves de Bernoulli de paramètre $p = 3/4$, X suit donc une loi binomiale $B(p = 0{,}75; n = 2500)$. On cherche ensuite si les conditions d'approximation de la loi binomiale par la loi normale sont vérifiées :

- $n = 2500 > 30$
- $np = 1875 > 5$
- $n(1-p) = 625 > 5$

On peut donc approximer cette loi binomiale par une loi normale $N(\mu_X = np = 1875\,; \sigma_X = \sqrt{npq})$.
Or $np = 1875 \neq 0{,}75$.

B FAUX Dans cet item, on veut un intervalle de fluctuation de la **proportion** d'étudiants préférant les cookies du CROUS (et non pas le nombre). On veut donc un intervalle de fluctuation de la variable aléatoire F. D'après la loi de l'estimateur d'une proportion, on peut, à partir des paramètres de la variable X, approximer F par une loi normale $N\left(\mu_F = p; \sigma_F = \sqrt{\frac{pq}{n}}\right)$. Le risque est fixé à $\alpha = 5\%$, donc $z\alpha/_2 = 1{,}96 \approx 2$.

$$IF_{0,95}(F) = \mu_F \pm z\alpha/_2 \times \sigma_F = p \pm z_{\frac{\alpha}{2}} \times \sqrt{\frac{p(1-p)}{n}} = 0{,}75 \pm 2 \times \sqrt{\frac{\frac{3}{4} \times \frac{1}{4}}{2500}}$$

$$= 0{,}75 \pm 2 \times \frac{\sqrt{3}}{\sqrt{16} \times \sqrt{2500}} = 0{,}75 \pm 2 \times \frac{1{,}6}{4 \times 50} = 0{,}75 \pm \frac{1{,}6}{100} = 0{,}75 \pm 0{,}016 = [0{,}734; 0{,}766]$$
$$= [0{,}734; 0{,}766]$$

C VRAI Cf. B

D FAUX On ne peut pas calculer d'intervalle de fluctuation de la proportion d'étudiants ayant pris 5 kg en 3 mois à partir des données de l'énoncé.

E FAUX On calcule un intervalle de confiance de la proportion d'étudiants ayant pris 5 kg en 3 mois au risque $\alpha = 1\%$, donc $z\alpha/_2 = 2{,}5758 \approx 2{,}6$:

$$ic_{0,99}(p) = f \pm z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = 0{,}5 \pm 2{,}6\sqrt{\frac{0{,}5 \times 0{,}5}{2500}} = 0{,}5 \pm 2{,}6 \times \frac{0{,}5}{50} = 0{,}5 \pm 2{,}6 \times \frac{1}{100}$$
$$= 0{,}5 \pm 0{,}026 = [0{,}474; 0{,}526]$$

Ici, on peut s'arrêter là car on voit déjà que l'item est faux. Cependant, si l'intervalle avait été celui donné dans l'énoncé, il aurait fallu en plus vérifier les conditions de validité aux bornes de l'intervalle pour pouvoir valider cet intervalle :

$$n > 30\,;\ nf_1 > 5\,;\ n(1-f_1) > 5\,;\ nf_2 > 5\,;\ n(1-f_2) > 5$$

Question 14 :

La sérotonine (5HT) est un neurotransmetteur du Système Nerveux Central (SNC). Une équipe suspecte que sa concentration sanguine serait plus élevée chez un individu autiste que chez un individu témoin.

Pour vérifier leur hypothèse, ils mesurent la concentration sanguine de sérotonine (5HT) chez 36 individus atteints d'autisme et chez 25 individus sains.

	Patients autistes	Témoins
Moyenne estimée	55	12
Variance estimée	9	1

On notera M l'estimateur de la moyenne de la concentration sanguine de sérotonine (5HT).
Le niveau de confiance est de 95%.
Aide au calcul : $1{,}96 \approx 2$

A. L'intervalle de fluctuation à 95% de l'estimateur de la moyenne de la concentration sanguine de sérotonine chez la population autiste est égal à $IF_{0{,}95}(M) = [52\,;58]$.
B. L'intervalle de confiance à 95% de l'estimateur de la moyenne de la concentration sanguine de sérotonine chez l'échantillon témoin est égal à $ic_{0{,}95}(M) = [11{,}6\,;12{,}4]$.
C. L'intervalle de confiance à 95% de l'estimateur de la moyenne de la concentration sanguine de sérotonine chez l'échantillon de patients autistes est égal à $ic_{0{,}95}(M) = [52\,;58]$.
D. L'intervalle de confiance à 95% de l'estimateur de la moyenne de la concentration sanguine de sérotonine chez l'échantillon de patients autistes est égal à $ic_{0{,}95}(M) = [54\,;56]$.
E. A partir du programme de PASS, on ne peut rien dire pour l'échantillon témoin.

Question 14 : E

Dans cet exercice, il faut bien lire l'énoncé et les items, puisqu'il n'y a pas besoin de faire le moindre calcul ! Dans l'énoncé, on nous présente les données obtenues à partir d'échantillons. On a donc affaire à des estimations, et on peut calculer des intervalles de confiance.

A FAUX Ici on cherche un intervalle de confiance. De plus, en général, on cherche un IF d'une variable aléatoire et non pas d'une estimation.

B FAUX Attention ! Pour les items B, C et D, on nous demande de calculer un intervalle de confiance de l'estimateur M. Or, un intervalle de confiance se calcule uniquement pour un paramètre tel que μ, et non pas pour un estimateur tel que M (pour un estimateur on calculera un intervalle de fluctuation). *Si vous voulez tout de même vous entraînez à calculer l'intervalle de confiance de μ à la confiance 95%, le voici :*

$$ic_{0,95}(\mu) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}} = 55 \pm 2 \times \frac{3}{\sqrt{36}} = 55 \pm \frac{6}{6} = 55 \pm 1 = [54; 56]$$

C FAUX Cf. B

D FAUX Cf. B

E VRAI La taille de l'échantillon des témoins vaut $n = 25$. Cette taille étant inférieure à 30, dans le cadre de la PASS, il n'est pas possible de calculer d'intervalles.

Question 15 :

On souhaite diminuer l'incidence des accidents cardio-vasculaires. On teste l'efficacité d'une nouvelle statine sur le taux sanguin de LDL. Notre échantillon comprend 36 sujets, tous issus de la population cible. Le taux sanguin moyen de LDL des sujets ayant reçu cette statine 1 fois/jour est estimé à 1.6 g/L avec un écart-type estimé de 0.2g/L. Dans le même temps, la concentration sanguine moyenne en troponine, témoin d'une atteinte cardiaque, est de 2 mg/L avec un écart-type de 0.1 mg/L.
Dans la population cible, la fréquence d'un taux sanguin de LDL compris entre 1.4g/L et 1.8g/L est de 85%.
Soient X, Y et F trois variables aléatoires. X correspond au taux sanguin moyen de LDL des sujets ayant reçu cette statine une fois par jour, Y à la concentration sanguine moyenne en troponine et F à la proportion de sujets ayant un taux de LDL compris entre 1.4g/L et 1.8g/L.
Données supplémentaires : E(XY) = 3.2005 ; $s_p = 0.36$

Aides au calcul : $0{,}85 \times 0{,}15 = 0{,}1275$ $\sqrt{\frac{0{,}1275}{36}} = 0{,}06$ $1{,}96 \approx 2$

A. L'intervalle de confiance du taux sanguin moyen de LDL dans la population cible est $\mathrm{ic}_{0,95}(\mu) = [1{,}53\,;1{,}67]$.

B. L'intervalle de fluctuation de la proportion de sujets ayant un taux de LDL compris entre 1,4g/L et 1,8g/L pour notre échantillon est $\mathrm{IF}_{0,95}(\mathrm{F}) = [0{,}79\,;0{,}91]$.

C. L'intervalle de fluctuation du taux sanguin moyen de LDL pour notre échantillon est $\mathrm{IF}_{0,95}(\mathrm{F}) = [1{,}54\,;1{,}66]$.

D. L'intervalle de fluctuation de la proportion de sujets ayant un taux de LDL compris entre 1,4g/L et 1,8g/L pour notre échantillon est $\mathrm{IF}_{0,95}(\mathrm{F}) = [0{,}73\,;0{,}97]$.

E. Le nombre de sujets nécessaire pour avoir une précision de l'$\mathrm{IF}_{0,95}(\mathrm{F}) \leq 1\%$ est d'au moins 5100 personnes.

Question 15 : ADE

Dans l'énoncé, on nous donne des informations sur 3 critères :

- Le taux sanguin moyen de LDL dans l'échantillon, dont on dispose des estimations obtenues à partir de ce même échantillon : $m = 1{,}6$ et $s = 0{,}2$. On peut donc calculer un intervalle de confiance à partir de ces valeurs.
- La concentration sanguine moyenne en troponine dans l'échantillon, dont on dispose les estimations obtenues à partir de ce même échantillon : $m = 2$ et $s = 0{,}1$. En réalité, aucun item ne porte sur ces données. Elles sont un peu là pour vous embrouiller (oupsi)
- La proportion du taux sanguin de LDL compris entre 1,4 g/L et 1,8 g/L dans l'échantillon, dont on dispose du paramètre dans la population : $p = 0{,}85$. On peut donc calculer un intervalle de fluctuation à partir de cette valeur

A VRAI On cherche à calculer l'intervalle de confiance du taux sanguin moyen de LDL à la confiance 95% (donc un risque $\alpha = 5\%$ et $z\alpha/_2 = 1{,}96 \approx 2$) :

$$ic_{0,95}(\mu) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}} = 1{,}6 \pm 2 \times \frac{0{,}2}{\sqrt{36}} = 1{,}6 \pm 2 \times \frac{0{,}4}{6} = 1{,}6 \pm 0{,}066 = [1{,}53; 1{,}67]$$

Attention, pour être sûr de garantir un niveau de confiance au moins égal à 95%, on minore toujours la borne inférieure et on majore toujours la borne supérieure.

B FAUX F est une variable aléatoire modélisant une proportion. Elle ne suit pas de loi particulière vu dans le cours « Variables Aléatoires ». On peut cependant approximer la variable F par une loi normale grâce à la loi de l'estimateur d'une moyenne. Mais pour cela, il faut vérifier les conditions d'approximations :

- $n = 36 > 30$
- $np = 36 \times 0{,}85 = 30{,}6 > 5$
- $n(1-p) = 36 \times 0{,}15 = 5{,}4$

Les conditions sont validées, donc on peut approximer F par la loi normale : $F \to N\left(\mu = p; \sigma = \sqrt{\frac{pq}{n}}\right)$.

On cherche à calculer l'intervalle de fluctuation de la proportion de sujets ayant un taux de LDL compris entre 1,4g/L et 1,8g/L à la confiance 95% (donc un risque $\alpha = 5\%$ et $z\alpha/_2 = 1{,}96 \approx 2$) :

$$IF_{0,95}(F) = p \pm z_{\alpha/2} \times \sqrt{\frac{pq}{n}} = 0{,}85 \pm 2 \times \sqrt{\frac{0{,}85 \times 0{,}15}{36}} = 0{,}85 \pm 2 \times 0{,}06 = 0{,}85 \pm 0{,}12$$

C FAUX Cf. A

D VRAI Cf. B

E VRAI Pour avoir le nombre de sujets nécessaires pour avoir une précision ≤1%, on utilise la formule :

$$n \geq \frac{pq \times z_{\alpha/2}^2}{i^2}$$
$$n \geq \frac{0{,}85 \times 0{,}15 \times 4}{0{,}01^2}$$
$$n \geq \frac{0{,}1275 \times 4}{10^{-4}}$$
$$n \geq 0{,}51 \times 10^4$$
$$n \geq 5100$$

Question 16 :

La Cour des Comptes souhaite savoir si le coût de production d'une centrale photovoltaïque au sol de 2,5 MWc est significativement différent selon si elle se trouve au Nord ou au Sud du pays.
Pour cela, elle réalise une étude dans 37 usines du Nord et 25 usines du Sud de la France. On notera m_n et m_s les estimations des moyennes et s_n et s_s les estimations des écart-types au Nord et au Sud du pays.
Le coût moyen estimé dans le Nord Est de 160€/GWh avec un écart-type estimé dans l'échantillon égal à 12€/GWh alors que le coût moyen estimé dans le Sud Est de 150€/GWh avec un écart-type estimé égal à 50€/GWh.
Aides au calcul : $1{,}96 \approx 2$ $2{,}5758 \approx 2{,}5$ $1{,}6449 \approx 1{,}5$

A. L'intervalle de confiance à 90% de l'estimation coût de production moyen d'une centrale photovoltaïque au sol de 2,5 MWc dans le Sud est $ic_{0,9}(m_S) = [135\,;165]$.
B. L'intervalle de confiance à 90% du coût de production moyen d'une centrale photovoltaïque au sol de 2,5MWc dans le Sud est $ic_{0,9}(\mu_S) = [135\,;165]$.
C. L'intervalle de confiance à 90% du coût de production moyen d'une centrale photovoltaïque au sol de 2,5MWc dans le Nord est $ic_{0,9}(\mu_N) = [157\,;163]$.
D. L'intervalle de confiance du coût de production moyen d'une centrale photovoltaïque au sol de 2,5MWc dans le Nord est $ic_{0,9}(\mu_N) = [142\,;178]$.
E. L'intervalle de fluctuation à 90% de l'estimateur du coût de production moyen d'une centrale photovoltaïque au sol de 2,5MWc dans le Nord est $IF_{0,9}(M_N) = [157; 165]$.

Question 16 : C

A FAUX Un intervalle de confiance se calcule pour un paramètre et non pas pour une estimation.

B FAUX A partir d'un échantillon, on définit un intervalle dans lequel on a de grandes chances de trouver la moyenne de la population = Intervalle de confiance.
Ici, μ_s et σ_s sont inconnus : l'énoncé nous donne m_s et s_s, on ne peut donc pas calculer d'IF.

- Moyenne ou proportion ? Ici, on analyse le coût de production d'électricité moyen : c'est une moyenne.
- On vérifie les conditions : n = 25 < 30. Le programme de la PASS ne permet pas de calculer d'intervalle de confiance de μ_s.

C VRAI À partir de l'énoncé, on a : $m_N = 160$ $s_N = 12$ $n_N = 36$
On cherche à calculer un intervalle de confiance de μ_N à la confiance 90% (donc $\alpha = 10\%$ et $z_{\alpha/2} = z_{0,05} = 1{,}6449 = 1{,}5$) :

$$ic_{0,90}(\mu_N) = m_N \pm z_{\alpha/2} \times \frac{s_N}{\sqrt{n_N}} = 160 \pm 1{,}5 \times \frac{12}{\sqrt{36}} = 160 \pm 1{,}5 \times \frac{12}{6} = 160 \pm 3 = [157; 163]$$

D VRAI Cf. C

E FAUX On ne peut pas calculer d'intervalle de fluctuation à partir des données de l'énoncé.

Question 17 :

On dose le taux de troponine C après un infarctus du myocarde chez 49 personnes. On obtient une moyenne de 1,8 µg/L et une variance estimée de 1 µg/L. Le risque α est égal à 0,01.

Aide au calcul : $2.57 \times \frac{1}{7} \approx 0{,}40$

On arrondira les valeurs de la table de la loi normale à 2 chiffres après la virgule.

A. Un intervalle de confiance de l'estimation du taux de troponine C après un infarctus du myocarde à la confiance 99% est $ic_{0,99}(m) = 1{,}8 \pm 0{,}40$.

B. Un intervalle de confiance du taux de troponine C après un infarctus du myocarde à la confiance 99% est $ic_{0,99}(\mu) = 1{,}8 \pm 0{,}40$.

C. Un intervalle de fluctuation de l'estimation du taux de troponine C après un infarctus du myocarde à la confiance 99% est $IF_{0,99}(m) = 1{,}8 \pm 0{,}40$.

D. Un intervalle de fluctuation du taux de troponine de C après un infarctus du myocarde à la confiance 99% est $IF_{0,99}(\mu) = 1{,}8 \pm 0{,}40$.

E. Un intervalle de fluctuation à la confiance 95% est systématiquement plus étroit qu'un intervalle de fluctuation à la confiance 99%, toutes valeurs étant égales par ailleurs.

Question 17 : BE

Ici, on dispose de la moyenne de l'échantillon : on va donc définir un intervalle dans lequel on a de grandes chances de trouver de la moyenne théorique µ de la population. On doit donc faire un intervalle de confiance **C et D FAUX**. De plus, un intervalle de confiance se calcule pour un paramètre, donc **A FAUX**.

B VRAI On cherche à calculer un intervalle de confiance à la confiance 99% (donc $\alpha = 1\%$ et $z_{\alpha/2} = z_{0,005} = 2{,}5758 \approx 2{,}58$) :

$$ic_{0,99}(\mu) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}} = 1{,}8 \pm 2{,}57 \times \frac{1}{\sqrt{49}} = 1{,}8 \pm 0{,}4$$

E VRAI En effet, si tous les éléments de la formule de l'intervalle de confiance sont égaux exceptés le risque, seul $z_{\alpha/2}$ est modifié. Il est égal à 2,57 pour une confiance à 99% contre 1,96 pour une confiance à 95%. Ainsi, pour une confiance à 95%, l'intervalle sera plus étroit.

Question 18 :

En moyenne, 20 français sur 50 pratiquent un sport. On constitue un échantillon de 36 français.
Soit F l'estimateur de la proportion, p la proportion théorique et f l'estimation de la proportion dans cet échantillon. Le risque est fixé à $\alpha = 5\%$.

Aides au calcul : $\frac{1}{6} = 0{,}17$ $\quad \frac{1}{12} = 0{,}08$ $\quad \sqrt{0{,}24} \approx \sqrt{0{,}25}$ $\quad 1{,}96 \approx 2$

A. Un intervalle de fluctuation à 95% de l'estimateur de la proportion de français pratiquant un sport est $IF_{0,95}(F) = [0{,}23\ ;\ 0{,}57]$.

B. La probabilité que la proportion de français pratiquant un sport dans cet échantillon soit strictement supérieure à 0,57 est 0,05.

C. La probabilité que la proportion de français pratiquant un sport dans cet échantillon soit strictement supérieure à 0,57 est de 0,025.

On considère à présent qu'un deuxième échantillon de français a été constitué. On obtient un intervalle de fluctuation de F au risque $\alpha = 5\%$ dont la précision a été divisée par 2.

D. 72 français ont été inclus dans cet échantillon.

E. Plus il y a de personnes incluses dans un échantillon, plus ce dernier est étroit.

Question 18 : ACE

A VRAI Dans cet exercice, on part d'une proportion donnée de la population, on a donc p. A partir de ça, on constitue un échantillon et on veut savoir quelle sera la proportion de notre échantillon, il s'agit de f. On calcule donc un intervalle de fluctuation de F à la confiance 95 %. Dans cet intervalle, on a 95% de chance de trouver la proportion de l'échantillon. On a $p = \frac{20}{50} = 0{,}4$ et $n = 36$.. On applique notre formule de l'intervalle de fluctuation pour une proportion :

$$IF_{0,95}(F) = p \pm z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} = 0{,}4 \pm 2 \times \sqrt{\frac{0{,}4 \times 0{,}6}{36}} = 0{,}4 \pm \frac{2 \times \frac{1}{2}}{6} = 0{,}4 \pm 0{,}17 = [0{,}23; 0{,}57]$$

B FAUX En effet, cet intervalle de fluctuation définit un intervalle dans lequel on a 95 % d'avoir la proportion estimée à partir de l'échantillon. Il y a donc 5 % de chance que la proportion de l'échantillon ne soit pas dans cet intervalle. Ainsi, il y a 2,5 % (= 0,025) de chance que la proportion de l'échantillon soit supérieure à la borne supérieure de l'IF et 2,5% de chance que la proportion soit inférieure à la borne inférieure l'IF.

C VRAI Cf. B

D FAUX L'effectif de cet intervalle peut aussi être trouvé de manière calculatoire, mais la façon la plus facile reste encore la logique. La population étant la même, tout comme le risque, le seul facteur qui est modifié est notre effectif n. De plus, on voit que l'intervalle que l'on obtient avec cet effectif est deux fois plus petit. On se dit donc instinctivement que l'effectif doit être doublé, puisque l'effectif est le seul diviseur. Sauf que notre effectif « se trouve sous une racine », il doit donc être quadruplé pour que le diviseur soit doublé et que la taille de l'intervalle soit divisée par deux. Donc la taille de cet échantillon est de 144.

E VRAI C'est dans la formule : $IF_{0,95}(F) = p \pm z_{\alpha/2}\sqrt{\frac{pq}{n}}$

Si n augmente, $z_{\alpha/2}\sqrt{\frac{pq}{n}}$ diminue et donc la largeur de l'intervalle de fluctuation aussi.

Question 19 :

La taille moyenne de la population française est de 1,70 m. On étudie un échantillon de 1000 villes. On note la variable aléatoire M_{1000} modélisant la taille moyenne de la population française et la variable aléatoire X modélisant la taille de la population française.

A. Il manque une donnée pour pouvoir calculer un intervalle de fluctuation de M_{1000}.
B. $M_{1000} = 1{,}70$.
C. L'espérance de M_{1000} vaut $\mu_{M_{1000}} = 1{,}70$.
D. La variable aléatoire M_{1000} correspond à la moyenne de 1000 variables aléatoires X_i suivant toutes la même loi.
E. L'espérance de X vaut $\mu_X = \frac{1{,}70}{1000}$.

Question 19 : ACD

A VRAI Pour calculer un intervalle de fluctuation de M_{1000}, on doit disposer de la taille de l'échantillon étudié, le paramètre de la moyenne (l'espérance de M_{1000}) et le paramètre de l'écart-type (l'écart-type de M_{1000}). Ici, nous ne disposons pas de l'écart-type de la population. Nous ne pouvons donc pas calculer d'intervalle de fluctuation de M_{1000}.

B FAUX Un estimateur comme M_{1000} est une variable aléatoire. Cela signifie que M_{1000} peut prendre plusieurs valeurs différentes mais n'est pas égal à une valeur précise. **Une variable aléatoire n'est jamais égale à une valeur.**

C VRAI La taille moyenne en France est de 1,70 m. On dispose donc d'un paramètre et plus précisément du paramètre de la taille moyenne qui est modélisé par M_{1000}, donc $\mu_{M_{1000}} = 1{,}70$.

D VRAI X modélise la taille de la population en France (**et non pas la taille moyenne**). M_{1000}, elle représente la moyenne de cette taille dans un échantillon de 1000 villes, donc M_{1000} est la moyenne de 1000 variables X. On peut noter ceci *(cette formule n'est pas à connaitre mais permet juste d'aider à comprendre)* :

$$M_{1000} = \frac{X_1 + X_2 + \cdots + X_{1000}}{n}$$

E FAUX L'espérance de X vaut $\mu_X = \mu_{M_{1000}} = 1{,}70$. Il faut bien comprendre que si M correspond à la moyenne de n variables aléatoires X_i, alors $\mu_X = \mu_M$ et $\sigma_X = \frac{\sigma_M}{\sqrt{n}}$.

Question 20 :

En France, la probabilité pour un homme de plus de 50 ans de présenter un cancer de la prostate est de 10 %. On étudie un échantillon constitué de n hommes français de plus de 50 ans. On calcule un intervalle de fluctuation de la proportion d'hommes ayant un cancer de la prostate dont la précision est de 0,02. Cet intervalle a été calculé avec un risque de première espèce de 5%.
Soit p la proportion d'hommes de plus de 50 ans ayant un cancer de la prostate, F l'estimateur de p et f son estimation.
Aide aux calculs : $1{,}96 \approx 2$

A. L'intervalle de fluctuation de l'estimateur de la proportion d'hommes de plus de 50 ans ayant un cancer de la prostate est $IF_{0,95}(F) = [0{,}09\,;0{,}11]$.
B. 95 % des valeurs prises par F sont comprises dans l'intervalle $[0{,}08\,;0{,}12]$.
C. Cet intervalle a été calculé pour un échantillon de taille $n = 900$.
D. Cet intervalle a été calculé pour un échantillon de taille $n = 90$.
E. La table de la loi normale est nécessaire pour résoudre cet exercice.

Question 20 : BCE

A FAUX On nous dit dans l'énoncé que l'on calcule un intervalle de fluctuation de F à la confiance 95% et que la précision de cet intervalle est de 0,02. On sait également que la probabilité pour un homme de plus de 50 ans d'avoir un cancer de la prostate vaut $p = 0{,}1$. Cela correspond à un paramètre obtenu dans la population d'hommes de plus de 50 ans. L'intervalle de fluctuation de F est donc :

$$IF_{0,95}(F) = 0{,}1 \pm 0{,}02 = [0{,}08\,;0{,}12]$$

B VRAI La façon dont est formuler l'item est une autre manière de demander un intervalle de fluctuation de F à la confiance 95%. En effet, un intervalle de fluctuation d'une variable aléatoire à la confiance $1 - \alpha\%$ contient $1 - \alpha\%$ des réalisations de cette même variable aléatoire.

C VRAI On sait que la précision de l'intervalle calculer vaut 0,02. On sait aussi que la précision vaut :

$$i = z_{\alpha/2} \times \frac{\sqrt{pq}}{\sqrt{n}} \leftrightarrow n = \frac{{z_{\alpha/2}}^2 \times pq}{i^2}$$

À la confiance 95%, $\alpha = 5\%$ et donc $z_{\alpha/2} = 1{,}96 \approx 2$. On a donc :

$$n = \frac{2^2 \times 0{,}1 \times 0{,}9}{0{,}02^2} = \frac{4 \times 0{,}09}{0{,}0004} = \frac{0{,}09}{10^{-4}} = 900$$

D FAUX Cf. C

E VRAI On a utilisé la 2e table de la loi normale pour calculer $z_{\alpha/2}$.

Question 21 :

Le cœur est un organe noble, dont l'importance pour notre santé est capitale. Une étude a été menée sur 400 sujets masculins pour détecter un lien entre la longueur du cœur à l'âge adulte et l'espérance de vie. La longueur du cœur a été mesurée par échocardiographie et est exprimée en cm (on s'intéresse à la longueur entre l'oreillette gauche et l'apex). L'espérance de vie est exprimée en année. Au sein de cet échantillon, la longueur moyenne du cœur est de 12 cm et l'espérance de vie moyenne est de 75 ans. On observe un écart type de 2 cm sur la longueur du cœur et un écart type de 10 ans sur l'espérance de vie. La covariance des variables aléatoires modélisant la longueur du cœur et l'espérance de vie est égale à 2 cm.an. De plus, les deux variables aléatoires modélisant la longueur du cœur et l'espérance de vie suivent une loi normale.

On note μ_L et μ_E les paramètres respectivement de la longueur du cœur et de l'espérance de vie, M_L et M_E leur estimateur respectif, et m_L et m_E leur estimation respective.

Aides aux calculs : $\frac{\sqrt{398}}{\sqrt{0{,}99}} \approx 20$; $\frac{\sqrt{399}}{\sqrt{1}} \approx 21$

A. L'intervalle de fluctuation à 95% de l'estimateur de la longueur moyenne du cœur est $[11{,}8 ; 12{,}2]$.

B. L'intervalle de confiance à 95% de la longueur moyenne du cœur est $[11{,}8 ; 12{,}2]$.

C. L'intervalle de confiance à 95% de l'espérance de vie est $[73 ;\ 75]$.

D. L'intervalle de fluctuation à 95% de l'estimateur de l'espérance de vie est $[73 ; 75]$.

E. m est une estimation de μ par l'estimateur M.

Question 21 : BE

Dans l'énoncé, on étudie un échantillon de 400 sujets. À partir de celui-ci, on obtient de estimations concernant la longueur du cœur et l'espérance de vie. On peut donc calculer des intervalles de confiance à partir de ces données.

A FAUX On ne peut pas calculer d'intervalle de fluctuation à partir des données de l'énoncé.

B VRAI À partir de l'énoncé, on a : $m_L = 12 \quad s_L = 2 \quad n = 400$. On cherche à calculer un intervalle de confiance à 95% (donc $\alpha = 5\%$ et $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$ic_{0,95}(\mu) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}} = 12 \pm 2 \times \frac{2}{\sqrt{400}} = 12 \pm \frac{4}{20} = 12 \pm 0{,}2 = [11{,}8; 12{,}2]$$

C FAUX À partir de l'énoncé, on a : $m_L = 75 \quad s_L = 10 \quad n = 400$. On cherche à calculer un intervalle de confiance à 95% (donc $\alpha = 5\%$ et $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$ic_{0,95}(\mu) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}} = 75 \pm 2 \times \frac{10}{\sqrt{400}} = 75 \pm \frac{20}{20} = 75 \pm 1 = [74; 76]$$

D FAUX On ne peut pas calculer d'intervalle de fluctuation à partir des données de l'énoncé.

E VRAI Il est très important de connaître qui est quoi.

Question 22 :

On étudie la justesse d'un spectroscope de masse pour le dosage d'une hormone. Anne, étudiante en PASS, obtient sur 36 analyses une moyenne de 20,10UI pour ce taux et un écart-type de 1,8 UI.

Aides au calcul : $1{,}96 \approx 2$

A. Un intervalle de confiance de la moyenne populationnelle au risque 5% est $[19{,}50 ; 20{,}70]$.

B. Un intervalle de confiance de la moyenne populationnelle au risque 5% est $[19{,}46 ; 20{,}74]$.

C. L'intervalle de fluctuation de l'estimateur de la moyenne populationnelle au risque de 5% est $[19{,}46 ; 20{,}74]$.
D. Plus la taille de l'échantillon augmente, plus l'intervalle de confiance est étroit.
E. Plus le niveau de confiance augmente, plus l'intervalle de confiance est large.

Question 22 : BDE

Il s'agit d'abord de choisir entre intervalles de confiance et de fluctuation. Un intervalle de fluctuation est centré autour de µ, la moyenne théorique, or dans ce cas-là on ne dispose que de l'estimation de la moyenne pour un échantillon donné. On ne peut donc calculer qu'un intervalle de confiance.

A VRAI On cherche à calculer un intervalle de confiance au risque $\alpha = 5\%$ (donc $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$ic_{0,95}(\mu) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}} = 20{,}1 \pm 2 \times \frac{1{,}8}{\sqrt{36}} = 20{,}1 \pm 2 \times 0{,}3 = 20{,}1 \pm 0{,}6 = [19{,}5; 20{,}7]$$

B FAUX Cf. A

C FAUX On ne peut pas calculer d'intervalle de confiance à partir des données de l'énoncé.

D VRAI On peut retrouver cette propriété en observant la formule. Si n augmente alors $\frac{s}{\sqrt{n}}$ diminue, la largeur de l'intervalle aussi.

E VRAI Augmenter le niveau de confiance, c'est augmenter la probabilité de trouver notre paramètre (µ ici) dans l'intervalle. L'intervalle est donc élargi. On peut aussi remarquer que quand α diminue, $z_{\alpha/2}$ augmente (on le voit dans la table).

Question 23 :

Un dermatologue souhaite estimer la proportion de mélanomes de la peau parmi les taches pigmentées de plus de 2cm de diamètre que ses collègues, médecins traitants, lui soumettent à l'auscultation. Parmi 100 patients ayant une tache de cette grandeur, il compte 80 mélanomes.

Aides au calcul : $1{,}6449 \approx 1{,}64$ $\quad 1{,}96 \approx 2$

A. L'intervalle de fluctuation à 90% de l'estimateur de la proportion de mélanomes est $[0{,}73 ; 0{,}87]$.
B. Un intervalle de confiance à 90% de la proportion de mélanomes est $[0{,}73 ; 0{,}87]$.
C. L'intervalle de fluctuation à 95% de l'estimateur de la proportion de mélanomes est $[0{,}72 ; 0{,}88]$.
D. Un intervalle de confiance à 95% de la proportion de mélanomes est $[0{,}72 ; 0{,}88]$.
E. Pour obtenir un intervalle deux fois plus étroit, il faut doubler le nombre de patients.

Question 23 : BD

A FAUX On a une estimation de la proportion de mélanomes dans la population à partir d'un échantillon. On cherche donc un intervalle de confiance.

B VRAI On cherche à calculer un intervalle de confiance au risque $\alpha = 10\%$ (donc $z_{\alpha/2} = 1{,}6449 \approx 1{,}64$) :

$$ic_{0,9}(p) = f \pm z_{\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} = 0{,}8 \pm 1{,}64 \sqrt{\frac{0{,}16}{100}} = 0{,}8 \pm 1{,}64 \times \frac{0{,}4}{10} = 0{,}8 \pm 0{,}0656$$

$$= [0{,}7344; 0{,}8656] = [0{,}73; 0{,}87]$$

On n'oublie pas de vérifier les conditions *a posteriori*. On a bien $nf_1 > 5$; $n(1-f_1) > 5$; $nf_2 > 5$ et $n(1-f_2) > 5$. L'intervalle que l'on a calculé est donc valide.

C FAUX On a une estimation de la proportion de mélanomes dans la population à partir d'un échantillon. On cherche donc un intervalle de confiance.

D VRAI On cherche à calculer un intervalle de confiance, mais cette fois-ci au risque $\alpha = 5\%$ (donc $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$ic_{0,95}(p) = f \pm z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = 0{,}8 \pm 2\sqrt{\frac{0{,}16}{100}} = 0{,}8 \pm 2 \times 0{,}04 = 0{,}8 \pm 0{,}08 = [0{,}72; 0{,}88]$$

On vérifie les conditions *a posteriori*. Ici aussi elles sont validées, l'intervalle est donc valide.

E FAUX Si on double le nombre de patients (n), nos bornes d'intervalle seront multipliées par $\frac{1}{\sqrt{2}}$, donc notre intervalle ne sera pas deux fois plus précis. Pour obtenir un intervalle deux fois plus précis, il faudrait quadrupler le nombre de patients.

Question 24 :

On décide de faire une étude sur le prix payé par des étudiants en médecine pour se loger lors de la première année. Pour ce faire, on interroge 49 PASS représentatifs des 1500 étudiants concernés par l'étude. L'estimation de l'écart type du prix payé par les PASS est de 28 €. De plus, un intervalle de confiance au risque α de la moyenne de la population a été réalisé et vaut [542,8 ; 557,2].

Quelle est la valeur du risque α ?

A. 3,6%.
B. 7,2%.
C. 7,5%.
D. 15%
E. Il manque une donnée pour répondre.

Question 24 : B

Cet exercice est un peu différent de ce qu'on a l'habitude de faire. Ce n'est pas un exercice type mais il permet de bien comprendre comment « jouer » avec les formules des intervalles.

Ici, la première chose à faire est d'analyser la situation dans laquelle l'étude est faite. On a un intervalle de confiance, donc on dispose d'estimations grâce à l'échantillon dont l'effectif est supérieur à 30 individus. Dans l'énoncé, on nous donne directement l'estimation de l'écart-type : $s = 28$. L'énoncé nous donne également l'estimation de la moyenne à travers l'intervalle de confiance. En effet, un intervalle de confiance d'une moyenne est centré sur celle-ci :

$$m = \frac{542{,}8 + 557{,}2}{2} = \frac{1100}{2} = 550$$

Maintenant, pour retrouver α, il va falloir calculer $z_{\alpha/2}$. Pour cela, il y a 2 moyens à partir des éléments dont on dispose : utiliser la formule de la borne inférieure ou celle de la borne supérieure. Ici, on a utilisé la formule avec la borne inférieure :

$$borne\ inférieure = m - z_{\alpha/2} \times \frac{s}{\sqrt{n}}$$

$$z_{\alpha/2} = \frac{m - borne\ inférieure}{\frac{s}{\sqrt{n}}} = \frac{550 - 542{,}2}{\frac{28}{\sqrt{49}}} = \frac{7{,}2}{\frac{28}{7}} = \frac{7{,}2}{4} = 1{,}8$$

À partir de là, on utilise la 1^{e} table de la loi normale (et non pas la 2^{e}). On utilise la 1^{e} table de la loi normale parce qu'ici on connait z et on cherche α (dans les exercices "classiques" des intervalles, on dispose de l'inverse : on connait α et on cherche z, dans ce cas-là, on utilise la 2^{e} table de la loi normale).

On cherche dans la table la valeur associée à 1,8. On obtient 0,9641. Sauf qu'attention, la 1e table de la loi normale donne des valeurs du type "$1-\alpha$" (comparé à la 2e table qui utilise directement des valeurs du type "α"). Mais il faut aussi faire attention au fait que le z que l'on a trouvé plus haut, correspond à celui de ${}^{\alpha}/_{2}$. Il faut donc multiplier ce qu'on lit dans la table par 2. On a ainsi :

$$\alpha = 2 \times (1 - 0{,}9641) = 2 \times 0{,}0359 = 0{,}0718 \approx 0{,}072 = 7{,}2\%$$ **B VRAI**

Question 25 :

Une équipe de scientifiques cherche à estimer le nombre de kilomètres parcourus par les éléphants d'Afrique en 2012. Ils connaissent pour cela quelques chiffres concernant une tribu de 196 éléphants représentatifs des éléphants d'Afrique. Ils savent en effet que dans cette tribu un quart des éléphants migre chaque année (chaque départ de la tribu étant compensé par une arrivée la même année).
On sait également que chaque éléphant migrateur de cette tribu parcourt en moyenne 104 kilomètres par an avec un écart-type estimé de 14km.

Aides aux calculs : $1{,}96 \approx 2$ $\sqrt{0{,}25 \times 0{,}75} \approx 0{,}42$ $\sqrt{196} = 14$; $\frac{\sqrt{3}}{56} \approx 0{,}03$; $\frac{\sqrt{3}}{4\times196} \approx 2.10^{-3}$; $\frac{3}{16\times14} \approx 0{,}01$

A. Un intervalle de confiance de la proportion d'éléphants d'Afrique migrant en 2012 au risque de 5% est $ic_{0,95}(p) = [0{,}19\,;0{,}31]$.
B. L'intervalle de fluctuation de la proportion d'éléphants d'Afrique migrant en 2012 au risque de 5% est $ic_{0,95}(p) = [0{,}22\,;0{,}28]$.
C. Un intervalle de confiance de la moyenne du nombre de kilomètres parcourus par chaque éléphant migrant d'Afrique en 2012 au risque de 5% est $ic_{0,95}(\mu) = [96\,;112]$.
D. L'intervalle de fluctuation de l'estimateur de la moyenne du nombre de kilomètres parcourus par chaque éléphant migrant d'Afrique en 2012 au risque de 5% est $IF_{0,95}(M) = [96\,;112]$.
E. Pour diviser la précision d'un intervalle de confiance d'une moyenne par deux, il faut multiplier l'effectif de l'échantillon par 4.

Question 25 : AE

Dans l'énoncé, on a deux informations différentes :

- La proportion d'éléphants qui migre chaque année : cette donnée provient de l'échantillon de 196 éléphants. Il s'agit donc d'une estimation : $f = 0{,}25$. On peut calculer des intervalles de confiance d'une proportion.
- La moyenne de kilomètres parcourus par an : cette donnée provient également de l'échantillon, mais attention !! Maintenant, on n'étudie plus l'échantillon de 196 éléphants d'Afrique, mais l'échantillon d'éléphants d'Afrique **migrateur**, qui représente $\frac{1}{4}$ des éléphants d'Afrique soit $\frac{196}{4} = 49$. On a donc des estimations : $m = 104$ et $s = 14$. On peut calculer des intervalles de confiance d'une moyenne.

A VRAI On cherche à calculer un intervalle de confiance de la proportion d'éléphants d'Afrique migrant en 2012 au risque de 5% (donc $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$ic_{0,95}(p) = f \pm z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = 0{,}25 \pm 2\sqrt{\frac{0{,}25 \times 0{,}75}{196}} = 0{,}25 \pm 2 \times \frac{0{,}42}{14} = 0{,}25 \pm \frac{0{,}84}{14}$$
$$= 0{,}25 \pm 0{,}06 = [0{,}19; 0{,}31]$$

On vérifie alors que les hypothèses :
$n \geq 30$ $\quad n \times f_1 \geq 5$ $\quad n \times (1 - f_1) \geq 5$ $\quad n \times f_2 \geq 5$ $\quad n \times (1 - f_2) \geq 5$

Les conditions sont validées, l'intervalle calculé est donc valide.

B FAUX On ne peut pas calculer d'intervalle de fluctuation à partir des données de l'énoncé.

C FAUX On cherche à calculer un intervalle de confiance de la moyenne du nombre de kilomètres parcourus par chaque éléphant migrant d'Afrique en 2012 au risque de 5% (donc $z\alpha_{/2} = 1{,}96 \approx 2$).

$$ic_{0,95}(\mu) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n'}} = 104 \pm 2 \times \frac{14}{\sqrt{49}} = 104 \pm 2 \times 2 = [100; 108]$$

D FAUX On ne peut pas calculer d'intervalle de fluctuation à partir des données de l'énoncé.

E VRAI Soit i la précision d'un intervalle de confiance : $i = z\alpha_{/2} \times \frac{s}{\sqrt{n}}$. Donc si on multiplie n par 4, on divise i par $\sqrt{4}$, donc par 2.

Question 26 :

On cherche à estimer la moyenne de la durée de travail avant accouchement dans la population des femmes enceintes arrivant à terme. Pour cela on a mesuré chez 100 de ces patientes le temps écoulé entre le début des contractions régulières et la naissance de leur enfant. On note x_i cette valeur chez la patiente i. On a obtenu les résultats suivants : $\sum x_i = 670$ heures et $s = 2$.

Aide aux calculs : $1{,}96 = 2$

A. L'estimation de la moyenne de la durée de travail est 6,7 heures.

B. Un intervalle de confiance à 95% de la moyenne de la durée de travail dans la population est $ic_{0,95}(\mu) = [6{,}3\,;7{,}1]$.

C. Un intervalle de confiance à 95% de la moyenne de la durée de travail dans la population est $ic_{0,95}(\mu) = [5{,}9\,;7{,}5]$.

D. Il aurait fallu observer 1600 patientes pour obtenir un intervalle de confiance à 95% de largeur 0,2.

E. Il aurait fallu observer 400 patientes pour obtenir un intervalle de confiance à 95% de largeur 0,2.

Question 26 : ABD

A VRAI Dans l'énoncé, on ne nous donne pas directement la moyenne estimée du temps entre le début des contractions et la naissance, mais le temps total de toutes les patientes. On obtient la moyenne en divisant ce temps total par le nombre de patientes, soit par 100 : $m = \frac{\sum x_i}{n} = \frac{670}{100} = 6{,}7$ heures.

B VRAI On cherche à calculer un intervalle de confiance au risque $\alpha = 5\%$ (donc $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$ic_{0,95}(\mu) = m \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}} = 6{,}7 \pm 2 \times \frac{2}{10} = [6{,}3; 7{,}1]$$

C FAUX Cf. B

D VRAI La largeur d'un intervalle correspond à 2 fois sa précision :

$$l = 2 \times z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} \quad \leftrightarrow \quad n = \frac{{z_{\alpha/2}}^2 \times s^2 \times 2^2}{l^2}$$

Avec une largeur de $l = 0{,}2$, on a :

$$n = \frac{{z_{\alpha/2}}^2 \times s^2 \times 2^2}{l^2} = \frac{2^2 \times 2^2 \times 2^2}{0{,}2^2} = \frac{64}{0{,}04} = \frac{64}{4} \times 100 = 1600$$

E FAUX Cf. D

Question 27 :

Le nombre d'heure de travail quotidien des étudiants en PASS est modélisé par une variable aléatoire X, Gaussienne telle que sa médiane vaut 10 heures et sa variance 4 heures. On considère un échantillon de 25 PASS.

Aides aux calculs : $1{,}96 \approx 2 \qquad 1{,}6449 \approx 1{,}6 \qquad 2{,}5758 \approx 2{,}5$

A. La probabilité qu'un PASS travaille plus de 12h par jour vaut 0,8413.

B. Avec les informations données dans l'énoncé, on peut calculer l'intervalle de confiance au risque $\alpha = 5\%$ de la proportion théorique de PASS dans un échantillon $n = 25$ dont le nombre d'heure de travail par jour est supérieur à 12.
C. Lorsque l'on calcule des intervalles de confiance de moyennes ou de proportions théoriques, on doit toujours vérifier des conditions d'application *a posteriori.*
D. Un intervalle de fluctuation au risque $\alpha = 10\%$ de la variable X vaut approximativement $IF_{0,9}(X) = [6,8 \,; 13,2]$
E. Un intervalle de fluctuation au risque $\alpha = 5\%$ de l'estimateur du temps de travail quotidien moyen des PASS vaut $IF_{0,95}(M) = [9,2; 10,8]$

Question 27 : DE

La variable aléatoire X modélisant le nombre d'heure de travail quotidien d'un P1 suit une loi normale d'espérance égale à 10 (la médiane est égale à l'espérance pour une loi normale) et d'écart-type = $\sqrt{\sigma^2} = \sqrt{4} = 2$. Soit $X \to N(\mu_X = 10; \sigma_X = \sqrt{4} = 2)$.

A FAUX On veut connaitre la probabilité : $P(X > 12) = P\left(Z > \frac{12-10}{2}\right) = P(Z > 1)$. En lisant la 1e table de la loi normale, on a :

$$P(Z > 1) = 1 - P(Z < 1) = 1 - 0,8413 = 0,1587 \approx 0,16$$

B FAUX Ici, nous n'avons pas d'estimations nous ne pouvons donc pas calculer d'intervalles de confiance.

C FAUX Les intervalles de confiance de moyennes et de proportions ont une condition *a priori* qui est $n > 30$. Et seuls les intervalles de confiance de proportions ont **en plus** des conditions à posteriori qui sont les suivantes :

Conditions *à posteriori* :

- $nf_1 \geq 5$
- $n(1 - f_1) \geq 5$
- $nf_2 \geq 5$
- $n(1 - f_2) \geq 5$

D VRAI On calcule l'intervalle de fluctuation de la variable aléatoire X au risque $\alpha = 10\%$ (donc $z_{\alpha/2} = 1,6449 \approx 1,6$) :

$$IF_{1-\alpha}(X) = \mu_X \pm z_{\alpha/2} \times \sigma_X = 10 \pm 1,6 \times 2 = 10 \pm 3,2 = [6,8 \,; 13,2]$$

E VRAI On nous demande cette fois ci de calculer l'intervalle de confiance du temps de travail quotidien **moyen**. Ceci peut être modélisé par l'estimateur M qui est variable aléatoire suivant une loi normale de paramètre : $M \to N\left(\mu_M = \mu_X; \sigma_M = \frac{\sigma_X}{\sqrt{n}}\right)$. Le risque est à $\alpha = 5\%$ (donc $z_{\alpha/2} = 1,96 \approx 2$) :

$$IF_{0,96}(M) = \mu_X \pm z_{\alpha/2} \times \frac{\sigma_X}{\sqrt{n}} = 10 \pm 2 \times \frac{2}{\sqrt{25}} = 10 \pm 0,8 = [9,2; 10,8]$$

Question 28 :

En France, la prévalence de cirrhose chez les chez les hommes de plus de 50 ans vaut 36%. On considère un échantillon de 625 hommes de plus de 50 ans originaires de Lyon. Soit X la variable aléatoire modélisant le nombre d'hommes de plus de 50 ans atteint de cirrhose et F la variable aléatoire modélisant la proportion d'hommes de plus de 50 ans atteint d'une cirrhose.

Soit p la proportion d'hommes de plus de 50 ans atteint de cirrhose dans la population française, F l'estimateur de cette proportion et f son estimation.

Aides au calcul : $1,6449 \approx 1,6$ $\quad 2,5758 \approx 2,5$ $\quad 1,96 \approx 2$ $\quad 0,48 \approx 0,50$ $\quad 0,24 \approx 025$

A. $IF_{0.95}(p) = [0.32 \,; 0.40]$

B. $IC_{0.95}(f) = [0.36\,;0.44]$
C. $IF_{0.95}(F) = [0.32\,;0.40]$
D. $IC_{0.99}(p) = [0.35\,;0.45]$
E. Toutes choses étant égales par ailleurs pour avoir un intervalle de confiance avec une demi-largeur deux fois moins grande, il faudra quatre fois plus de personnes dans notre échantillon.

Question 28 : CDE

A FAUX Dans cet énoncé, on dispose d'une proportion théorique : $\boldsymbol{p = 0,36}$. On peut donc dire que l'estimateur $\boldsymbol{F}$ peut être approximé (conditions à vérifier) par une loi normale :

$$\boldsymbol{F \rightarrow N\left(\mu_F = p; \sigma_F = \sqrt{\frac{pq}{n}}\right)}$$

Les conditions à vérifier sont : $\boldsymbol{n \geq 30}$; $\boldsymbol{np \geq 5}$; $\boldsymbol{nq \geq 5}$. Ici, elles sont validées.
On cherche à calculer un intervalle de fluctuation de $\boldsymbol{F}$ à la confiance 95% (soit $\boldsymbol{\alpha = 5\%}$, donc $\boldsymbol{z_{\alpha/2} = 1,96 \approx 2}$) :

$$IF_{0.95}(F) = p \pm z_{\alpha/2}\sqrt{\frac{pq}{n}} = 0,36 \pm 2 \times \sqrt{\frac{0,36 \times 0,64}{625}} = 0,36 \pm 2 \times \frac{0,6 \times 0,8}{25} = 0,36 \pm 2 \times \frac{0,48}{25}$$
$$= 0,36 \pm 2 \times \frac{0,5}{25} = 0,36 \pm 0,04 = [0,32\,;0,40]$$

B FAUX On ne peut pas calculer d'intervalle de confiance à partir des données de l'énoncé.

C VRAI Cf. A

D VRAI Cf. B

E VRAI La demi-largeur correspond à la précision : $i = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}$. Si on multiplie par 4 la taille de l'échantillon, soit n, alors on divise par $\sqrt{4} = 2$ la précision.

Question 29 :

En France, les jeunes entre 16 et 25 ans passent en moyenne 20h par semaine sur Netflix avec une variance de 16h. On posera X la variable aléatoire gaussienne modélisant le nombre d'heure passées sur Netflix par semaine par les jeunes de 16 à 25 ans en France. Soit un échantillon de 49 jeunes de 16 à 25 ans.
On pose μ la moyenne théorique, M l'estimateur de la moyenne et m l'estimation de la moyenne dans notre échantillon.
Aides au calcul : $1,96 \approx 2$ $\quad \frac{1}{7} \approx 0,143$

A. L'intervalle de fluctuation à 95% de l'estimateur de la moyenne du nombre d'heures passées sur Netflix par semaine par les jeunes de 16 à 25 ans est $IF_{0,95}(M) = [18,85\,;21,15]$.
B. Un intervalle de confiance à 95% de la moyenne du nombre d'heures passées sur Netflix par semaine par les jeunes de 16 à 25 ans est $IC_{0,95}(\mu) = [18,85\,;21,15]$.
C. Concernant les intervalles de confiance et de fluctuation, plus le niveau de confiance est grand, plus l'intervalle est étroit.
D. L'estimateur de la moyenne M est un nombre défini à partir de la population.
E. L'intervalle de fluctuation à 95% de X est $IF_{0,95}(X) = [12; 28]$.

Question 29 : A

A VRAI On a X la variable aléatoire gaussienne qui modélise le nombre d'heure passées par les jeunes de 16 à 25 ans sur Netflix par semaine, donc $X \rightarrow N\left(\mu_X = 20; \sigma_X = \sqrt{16} = 4\right)$.

D'après la loi de l'estimateur d'une moyenne, on peut poser M l'estimateur de la moyenne définit par : $M = \sum_{i=1}^{49} \frac{Xi}{n}$, donc $M \rightarrow N\left(\mu_M = \mu_X ; \sigma_M = \frac{\sigma_X}{\sqrt{n}}\right)$. Le niveau de confiance est à 95% (soit $\alpha = 5\%$, donc $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$IF_{0,95}(M) = \mu_X \pm z_{\frac{\alpha}{2}} \times \frac{\sigma_X}{\sqrt{n}} = 20 \pm 2 \times \frac{4}{\sqrt{49}} = 20 \pm \frac{8}{7} = 20 \pm 1{,}143$$
$$= [18{,}857; 21{,}143] = [18{,}85 ; 21{,}15]$$

B FAUX Nous ne sommes pas dans les conditions de calcule d'un IC. De plus, on connait déjà μ, on ne calcule pas d'intervalle de confiance de μ

C FAUX Au contraire, plus la confiance est grande, plus notre intervalle sera large.

D FAUX l'estimateur de la moyenne est une variable aléatoire. Ce n'est pas un nombre.

E VRAI On cherche à calculer un intervalle de fluctuation de X (*pour rappel :* $X \rightarrow N(\mu_X = 20; \sigma_X = 4)$, à la confiance 95% (soit $\alpha = 5\%$, donc $z_{\alpha/2} = 1{,}96 \approx 2$) :

$$IF_{0,95}(X) = \mu_X \pm z\alpha/_2 \times \sigma_X = 20 \pm 2 \times 4 = 20 \pm 8 = [12; 28]$$

Question 30 :

On étudie la proportion de naissance par césarienne dans la population. Pour ce faire, on réalise un échantillon représentatif de 65 naissances. On y compte 13 naissances par césarienne. On réalise un intervalle de confiance au risque α de la proportion théorique de naissance par césarienne et on obtient : $[0{,}125 ;\ 0{,}275]$.

Aides au calcul : $\frac{0{,}75\times 8}{4} = 1{,}5$

A. L'intervalle calculé n'est pas valable car les conditions de validité ne sont pas respectées.
B. Pour cet intervalle de confiance $\alpha = 1{,}5\%$.
C. Pour cet intervalle de confiance $6{,}6\% < \alpha < 6{,}7\%$.
D. Pour cet intervalle de confiance $13{,}2\% < \alpha < 13{,}4\%$.
E. De manière générale, contrairement aux intervalles de fluctuation, pour un risque alpha donné il existe un unique intervalle de confiance.

Question 30 : D

A FAUX Pour réaliser un intervalle de confiance d'une proportion théorique, il faut d'abord vérifier LA condition *a priori* à savoir $n \geq 30$. On a $n =$ 65 donc la condition est vérifiée.

Ensuite, sur l'intervalle donné, on vérifie les conditions a posteriori :

$$n \times f_1 = 65 \times 0{,}125 > 5$$
$$n \times f_2 = 65 \times 0{,}275 > 5$$
$$n \times (1 - f_1) = 65 \times 0.875 > 5$$
$$n \times (1 - f_2) = 65 \times 0.725 > 5$$

Nos conditions *a posteriori* sont vérifiées. Notre intervalle est tout à fait valable.

B FAUX On pose $f = \frac{13}{65} = \frac{13}{13\times 5} = \frac{1}{5} = 0{,}2$

Pour retrouver α, il va falloir calculer $z_{\alpha/2}$. Pour cela, il y a 2 moyens à partir des éléments dont on dispose : utiliser la formule de la borne inférieure ou celle de la borne supérieure. Ici, on a utilisé la formule avec la borne supérieure :

$$borne\ supérieure = f + z_{\alpha/2} \times \frac{\sqrt{f(1-f)}}{\sqrt{n}}$$

$$z_{\alpha/2} = \frac{borne\ supérieure - f}{\frac{\sqrt{f(1-f)}}{\sqrt{n}}} = \frac{0{,}275 - 0{,}2}{\frac{\sqrt{0{,}16}}{\sqrt{64}}} = \frac{0{,}075}{\frac{0{,}4}{8}} = \frac{0{,}75 \times 8}{4} = 1{,}5$$

À partir de là, on utilise la 1e table de la loi normale (et non pas la 2e). On utilise la 1e table de la loi normale parce qu'ici on connait z et on cherche α (dans les exercices "classiques" des intervalles, on dispose de l'inverse : on connait α et on cherche z, dans ce cas-là, on utilise la 2e table de la loi normale).

On cherche dans la table la valeur associée à 1,5. On obtient 0,9332. Sauf qu'attention, la 1e table de la loi normale donne des valeurs du type "$1-\alpha$" (comparé à la 2e table qui utilise directement des valeurs du type "α"). Mais il faut aussi faire attention au fait que le z que l'on a trouvé plus haut, correspond à celui de $\alpha/_2$. Il faut donc prendre le complément à 1 puis multiplier par 2 ce qu'on lit dans la table. On a ainsi :

$$\alpha = 2 \times (1 - 0{,}9332) = 2 \times 0{,}0668 = 0{,}1336 = 13{,}36\%$$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441

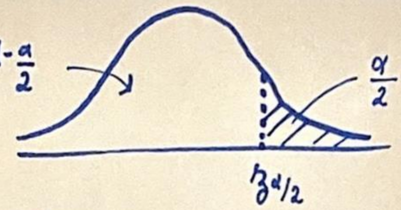
C FAUX Cf. B

D VRAI Cf. B

E FAUX On parle d'**un** intervalle de confiance pour une population donnée car il en existe une infinité tout comme il existe une infinité d'échantillon. *A contrario*, il existe un unique intervalle de fluctuation pour une population donnée et pour un risque α donnée puisque nous partons du paramètre théorique de la population qui lui est unique.

COMPRENDRE POURQUOI ON FAIT 1-P(Z>1,5) AVEC LA TABLE 1 :

On a $z_{\alpha/2} = 1,5$ Rappel



$z_{\alpha/2}$ est donc la valeur telle que $P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$

$P(Z > 1,5) = \frac{\alpha}{2}$.

Table 1 : $P(Z \leq z)$

$P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$

$P(Z \leq 1,5) = 0,9332 = 1 - \frac{\alpha}{2}$

On isole $\frac{\alpha}{2} \Rightarrow 1 - 0,9332 = 0,0668$

donc $\alpha = 0,0668 \times 2 = 0,1336$.

Question 31 :

Au service d'oncologie médicale de l'Hôpital Édouard Herriot (HEH), la probabilité pour qu'un patient admis soit atteint d'un cancer du poumon est de 36%. On s'intéresse à un groupe de 36 sujets adressés au service d'oncologie médicale à HEH. On définit X, la variable aléatoire modélisant le nombre de patient admis en oncologie à HEH avec un cancer du poumon dans un échantillon de 36 patients.
Lorsqu'un patient est admis dans le service d'Oncologie médicale, on réalise un « Test CRP » pour doser la protéine CRP (protéine de l'inflammation). Dans ce service, la concentration en CRP chez un patient admis peut être modélisé par la variable aléatoire Y suivant une loi normale de paramètre $\mu_Y = 40\,mg/L$ et $\sigma_Y = 3\,mg/L$.

Aides au calcul : $1{,}96 \approx 2$ $\quad 0{,}36 \times 36 \approx 13$ $\quad \sqrt{36 \times 0{,}36 \times 0{,}64} \approx 3$

A. X suit une loi binomiale de paramètre $B(n = 36; p = 0{,}36)$.
B. L'intervalle de fluctuation à 95% de Y est $IF_{0,95}(Y) = 40 \pm 6$.
C. L'intervalle de fluctuation à 95% de M est $IF_{0,95}(M) = 40 \pm 1$.
D. L'intervalle de fluctuation à 95% de X est $IF_{0,95}(X) = 0{,}36 \pm 0{,}16$.
E. L'intervalle de fluctuation à 95% de F est $IF_{0,95}(F) = 0{,}36 \pm 0{,}16$.

Question 31 : ABCE

Dans cet exercice, on nous demande beaucoup de compréhension mais relativement peu de calculs.
On peut déjà remarquer que nous n'avons aucune donnée issue de l'échantillon mais uniquement des données de la population : on calcule donc des intervalles de fluctuation. On essaie alors d'ordonner les informations :

- Le nombre de cas de cancer du poumon dans l'échantillon : ceci est modélisé par la variable X qui suit une loi binomiale : $X \rightarrow B(n = 36; p = 0{,}36)$. On vérifie les conditions d'approximation :
    - $n \geq 30; np \geq 5; nq \geq 5$: ici, ces conditions sont validées

On peut approximer X par une loi normale : $\boldsymbol{X \leadsto N\left(\mu_X = np \approx 13; \sigma_X = \sqrt{npq} \approx 3\right)}$ *(voir aides au calcul).*

- La concentration en CRP chez un patient admis au service d'Oncologie : ceci est modélisé par la variable Y qui suit une loi normale : $\boldsymbol{Y \rightarrow N(\mu_Y = 40; \sigma_Y = 3)}$.

À partir de ces 2 variables aléatoires, on peut en définir 2 autres :

- L'estimateur F qui modélise la proportion de cas de cancer du poumon dans l'échantillon. Par définition, $F = \frac{X}{n}$. Or, X suit approximativement une loi normale, donc F suit aussi approximativement une loi normale de paramètre $\boldsymbol{F \leadsto N\left(\mu_F = p; \sigma_F = \sqrt{\frac{pq}{n}}\right)}$.
- L'estimateur M qui modélise la concentration moyenne en CRP dans l'échantillon. Par définition, $M = \frac{\sum Y_i}{n}$. Or, Y suit une normale, donc M suit également une loi normale de paramètre $\boldsymbol{M \rightarrow N\left(\mu_M = \mu_Y; \sigma_M = \frac{\sigma_Y}{\sqrt{n}}\right)}$.

A VRAI Cf. Intro

B VRAI On cherche à calculer l'intervalle de fluctuation de Y à la confiance 95% (soit $\alpha = 5\%$ et donc $z_{\frac{\alpha}{2}} = 1{,}96 \approx 2$). Pour rappel : $Y \rightarrow N(\mu_Y = 40; \sigma_Y = 3)$.

$$IF_{0,95}(Y) = \mu_Y \pm \left(z_{\frac{\alpha}{2}} \times \sigma_Y\right) = 40 \pm (2 \times 3) = 40 \pm 6$$

C VRAI On cherche à calculer l'intervalle de fluctuation de M à la confiance 95% (soit $\alpha = 5\%$ et donc $z_{\frac{\alpha}{2}} = 1{,}96 \approx 2$). Pour rappel : $M \rightarrow N\left(\mu_M = \mu_Y; \sigma_M = \frac{\sigma_Y}{\sqrt{n}}\right)$.

$$IF_{0,95}(M) = \mu_M \pm \left(z_{\frac{\alpha}{2}} \times \sigma_M\right) = 40 \pm \left(2 \times \frac{3}{\sqrt{36}}\right) = 40 \pm 1$$

D FAUX On cherche à calculer l'intervalle de fluctuation de X à la confiance 95% (soit $\alpha = 5\%$ et donc $z_{\frac{\alpha}{2}} = 1{,}96 \approx 2$). Pour rappel : $X \leadsto N(\mu_X = 13; \sigma_X = 3)$.

$$IF_{0,95}(X) = \mu_X \pm \left(z_{\frac{\alpha}{2}} \times \sigma_X\right) = 13 \pm (2 \times 3) = 13 \pm 6$$

E VRAI On cherche à calculer l'intervalle de fluctuation de F ici à la confiance 95% (soit $\alpha = 5\%$ et donc $z_{\frac{\alpha}{2}} = 1{,}96 \approx 2$). Pour rappel : $F \leadsto N\left(\mu_F = p; \sigma_F = \sqrt{\frac{pq}{n}}\right)$.

$$IF_{0,95}(F) = \mu_F \pm \left(z_{\frac{\alpha}{2}} \times \sigma_F\right) = 0{,}36 \pm \left(2 \times \sqrt{\frac{0{,}36 \times 0{,}64}{36}}\right) = 0{,}36 \pm \left(2 \times \frac{0{,}6 \times 0{,}8}{6}\right)$$

$$0{,}36 \pm 2 \times 0{,}08 = 0{,}36 \pm 0{,}16$$

Question 32 – Les conseils de Charlie Parker :

Charlie Parker est un des plus grands saxophonistes de l'histoire du Jazz. Pour atteindre sa virtuosité, il s'est entraîné non-stop, jusqu'à 14h par jour ! Aujourd'hui, beaucoup de jeunes musiciens prometteurs s'inspirent de son style ainsi que de sa méthode de travail.

On considère que la distribution de la variable aléatoire X, modélisant le nombre d'heures d'entraînement d'un musicien est Gaussienne.

Un échantillon aléatoire de 324 musiciens est constitué pour l'étude. A partir de cet échantillon, on estime la moyenne du temps de travail d'un musicien à 12h/j et la variance à 4h/j

Aides au calcul : $1{,}2816 \approx 1{,}3$; $1{,}6449 \approx 1{,}6$; $1{,}96 \approx 2$; $2{,}3263 \approx 2{,}3$; $2{,}5758 \approx 2{,}6$
$\frac{1{,}6}{9} \approx 0{,}178$; $\frac{2}{9} \approx 0{,}222$; $\frac{2{,}6}{9} \approx 0{,}289$; $18^2 = 324$

A. Un intervalle de confiance de la moyenne théorique du nombre d'heures d'entraînement, à la confiance 0,90 vaut [11, 83 ; 12, 17].

B. Un intervalle de fluctuation à 90% de l'estimateur du nombre d'heures d'entraînement moyen vaut [11, 82 ; 12, 18].
C. Un intervalle de confiance de la moyenne théorique du nombre d'heures d'entraînement, à la confiance 0,90 vaut [11, 82 ; 12, 18].
D. Un intervalle de confiance à 95% de la moyenne théorique du nombre d'heures d'entraînement vaut [11, 82 ; 12, 18].
E. Plus le risque de première espèce est élevé, plus l'intervalle est étroit.

Question 32 – Les conseils de Charlie Parker : CE

Analyse de l'énoncé : Dans cet exercice, l'énoncé donne des estimations et non les valeurs théoriques. Par conséquent, il n'est pas possible de calculer les intervalles de fluctuation.
On voit que $n = 324$ donc $n \geq 30$, on peut alors calculer l'intervalle de confiance à l'aide de la formule ci-dessous sans avoir de conditions d'application à vérifier :

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

$$ic_{0,90}\,(\mu) = \left[m \pm z_{\frac{0,10}{2}} \times \frac{s}{\sqrt{n}}\right]$$

Trouver $z_{\frac{0,10}{2}}$ revient à trouver *z* telle que P(Z>*z*) = 0,05. Grâce à la table de la loi normale centrée réduite, on obtient : $z \;=\; 1{,}6449 \;\approx 1{,}6$

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

$$ic_{0,90}\,(\mu) = \left[12 \pm 1{,}6 \times \frac{2}{\sqrt{324}}\right]$$

$$ic_{0,90}\,(\mu) = \left[12 \pm 1{,}6 \times \frac{2}{18}\right] = \left[12 \pm 1{,}6 \times \frac{1}{9}\right] = [12 \pm 0{,}178]$$

$ic_{0,90}\,(\mu) = [11,822\,;12,178]$On minore la borne inférieure et majore la borne supérieure :

$$ic_{0,90}\,(\mu) = [11,82; 12{,}18]$$

De même, on calcule :

$$ic_{0,95}\,(\mu) = \left[m \pm z_{\frac{0,05}{2}} \times \frac{s}{\sqrt{n}}\right]$$

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 : $z \;=\; 1{,}96 \;\approx 2$

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

Donc :

$$Ic_{0,95}\,(\mu) = \left[12 \pm 2 \times \frac{2}{\sqrt{324}}\right] = \left[12 \pm 2 \times \frac{2}{18}\right] = \left[12 \pm 2 \times \frac{1}{9}\right] = [12 \pm 0{,}222]$$

$$Ic_{0,95}\,(\mu) = [11{,}778\,;12{,}222] = [11{,}77;12{,}23]$$

En fait, ce calcul n'est pas vraiment nécessaire : une fois qu'on a trouvé l'*ic* à 90 %, la proposition pour l'*ic* à 95 % ne peut pas être juste car on ne peut pas obtenir les mêmes bornes, les $z_{\frac{\alpha}{2}}$étant différents.

La proposition D est donc forcément fausse. On vous a néanmoins détaillé le calcul pour que vous puissiez vous entraîner.

A FAUX L'arrondi n'est pas juste
B FAUX
C VRAI
D FAUX
E VRAI – Plus le risque de première espèce α est élevé, plus l'intervalle est étroit.
C'est logique ! Comparez les deux résultats obtenus précédemment.
L'intervalle obtenu avec un risque de 10% est plus étroit que celui obtenu à 5%.

Question 33 – Les disciples de Parker :

On sait par ailleurs que la probabilité qu'un saxophoniste suive la stratégie de Parker vaut 0,4.
On considère un échantillon de 144 saxophonistes choisis au hasard.

Aides au calcul : $\sqrt{0{,}24} \approx \sqrt{0{,}25}$ $\quad \frac{1}{6} \approx 0{,}166 \quad$ 1,96 ≈ 2

A. *p* est une estimation de *f.*
B. Dans un échantillon statistique, les variables aléatoires X_i ne sont pas indépendantes.
C. $ic_{0,95}\,(p) = [0{,}31; 0{,}49]$.
D. $IF_{0,95}\,(F) = [0{,}31; 0{,}49]$.
E. La précision de l'intervalle de fluctuation à 95% est de 0,18.

Question 33 – Les disciples de Parker : D

A FAUX *f* est une estimation de *p.*

B FAUX L'échantillon statistique (ou n-échantillon) est un ensemble de variables aléatoires indépendantes et de même loi. Cf. Diapo du cours pour avoir plus d'info.
C FAUX Ici, on nous donne une proportion théorique des saxophonistes qui appliquent la méthode de Parker. On étudie donc les intervalles de fluctuation (ou intervalles de pari).
D VRAI

$$IF_{1-\alpha}(F) = p \pm z_{\frac{\alpha}{2}} \times \sqrt{\frac{p(1-p)}{n}}$$

$$IF_{0,95}(F) = 0,4 \pm z_{\frac{0,05}{2}} \times \sqrt{\frac{0,4 \times 0,6}{144}}$$

$$IF_{0,95}(F) = 0,4 \pm 2 \times \sqrt{\frac{0,4 \times 0,6}{144}}$$

$$IF_{0,95}(F) = 0,4 \pm 2 \times \sqrt{\frac{0,24}{144}} = 0,4 \pm 2 \times \frac{\sqrt{0,24}}{\sqrt{144}} = 0,4 \pm 2 \times \frac{\sqrt{0,24}}{\sqrt{144}}$$

$$IF_{0,95}(F) = 0,4 \pm 2 \times \frac{0,5}{12}$$

$$IF_{0,95}(F) = 0,4 \pm \frac{1}{12}$$

$$IF_{0,95}(F) = 0,4 \pm 0,083$$

$$IF_{0,95}(F) = [0,317\,;0,483]$$

On minore la borne inférieure et majore la borne supérieure :

$$IF_{0,95}(F) = [0,31\,;0,49]$$

E FAUX La précision d'un intervalle :

$$i = \frac{l}{2}$$

Avec la largeur de l'intervalle : $l = [borne\ supérieure - borne\ inférieure]$.

$$i = \frac{0,49 - 0,31}{2} = \frac{0,18}{2} = 0,09$$

En réalité, il n'y a pas besoin de la recalculer, elle est donnée dans le calcul précédent, la précision est la demi-largeur de l'intervalle de fluctuation, donc on pouvait utiliser : $IF_{0,95}(F) = 0,4 \pm 0,083$ et en déduire que i≈0,083.

Question 34 – L'intelligence artificielle et la musique :

En février 2020, L'agence digitale Space150 a élaboré un logiciel d'intelligence artificielle qui permet de composer les chansons et de les interpréter. La plus connue était *Jack Park Canny Dope Man* par TravisBott, une création inspirée du rappeur américain Travis Scott.

Pourtant, malgré l'inventivité du travail, la chanson n'était pas très bien reçue par le grand public. Les tuteurs de biostatistiques, curieux comme d'habitude, souhaitent savoir si TravisBott est un succès ou un échec au sein de deux populations distinctes : les Gen-Z et les Baby-boomers.

Voici un tableau regroupant les données qu'ils ont recueillies lors de leur enquête :

	Succès	Echec
Gen-Z (*gz*)	300	100
Baby-Boomers (*bb*)	10	40

Aide au calcul : $\sqrt{0,75 \times 0,25} \approx 0,433$ $0,8 \times \frac{0,433}{20} \approx 0,0346$ $1,96 \approx 2$ $1,645 \approx 1,6$

$\frac{0,8}{\sqrt{50}} \approx 0,11$On utilisera f_{gz}, la proportion estimée de succès de TravisBott chez les Gen-Z. p_{gz} est la proportion théorique de succès de TravisBott chez les Gen-Z. Enfin, F_{gz} est l'estimateur de la proportion de succès de TravisBott chez les Gen-Z.

On aura aussi f_{bb} la proportion estimée de succès de TravisBott chez les Baby-Boomers. P_{bb} est la proportion théorique de succès de TravisBott chez les Baby-Boomers. Enfin, F_{bb} est l'estimateur de la proportion de succès de TravisBott chez les Baby-Boomers.

A. Chez les Gen-Z, un intervalle de confiance à 95% de la proportion théorique de succès de TravisBott est : $ic_{0,95}(p_{gz}) = [0{,}707; 0{,}793]$
B. Chez les Gen-Z, un intervalle de confiance à 95% de la proportion théorique de succès de TravisBott est : $ic_{0,95}(p_{gz}) = [0{,}706; 0{,}794]$
C. Chez les Gen-Z, un intervalle de fluctuation à 95% de la proportion estimée de succès de TravisBott est : $IF_{0,95}(F_{gz}) = [0{,}706; 0{,}794]$
D. Chez les Gen-Z, un intervalle de confiance à 90% de la proportion estimée de succès de TravisBott est : $ic_{0,90}(f_{gz}) = [0{,}734; 0{,}785]$
E. Chez les Baby-Boomers, un intervalle de confiance à 95% de la proportion théorique de succès de TravisBott est : $ic_{0,95}(p_{bb}) = [0{,}09; 0{,}31]$

Question 34 – L'intelligence artificielle et la musique : B

Analyse du sujet : On a ici une étude sur deux échantillons avec leurs estimations. A partir de celles-ci, on peut construire des **intervalles de confiance** qui vont avoir (1-α) % de chances de contenir la vraie valeur du paramètre. 1-α est appelé le niveau de confiance.

- Chez les **Gen-Z** :

On va d'abord vérifier la condition sur la taille de l'échantillon : n = 400 donc n ≥ 30.
On peut alors poursuivre avec le calcul de l'intervalle de confiance.
Avec le tableau, on calcule l'estimation de la fréquence du succès. Soit : $f = \frac{300}{400} = 0{,}75$
On vous remet la formule de l'intervalle de confiance de la proportion :

$$ic_{1-\alpha}(p_{gz}) = f \pm z_{\frac{\alpha}{2}} \times \sqrt{\frac{f(1-f)}{n}}$$

$$ic_{0,95}(p_{gz}) = f \pm z_{\frac{0,05}{2}} \times \sqrt{\frac{f(1-f)}{n}}$$

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 : $z = 1{,}96 \approx 2$

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

Application numérique :

$$ic_{0,95}(p_{gz}) = 0,75 \pm 2 \times \sqrt{\frac{0,75 \times 0,25}{400}}$$
$$ic_{0,95}(p_{gz}) = 0,75 \pm 2 \times \frac{0,433}{20}$$
$$ic_{0,95}(p_{gz}) = 0,75 \pm \frac{0,433}{10}$$
$$ic_{0,95}(p_{gz}) = 0,75 \pm 0,0433$$
$$ic_{0,95}(p_{gz}) = [0,7067; 0,7933]$$
$$ic_{0,95}(p_{gz}) = [0,706; 0,794]$$

Attention : N'oublie pas de vérifier les conditions pour considérer l'intervalle de confiance d'une proportion comme valide.
Soient $f_1 et f_2$ les deux bornes de l'IC :

$$n \times f_1 \geq 5$$
$$n \times (1 - f_1) \geq 5$$
$$n \times f_2 \geq 5$$
$$n \times (1 - f_2 \geq 5$$

Dans notre situation :

$$n \times f_1 = 400 \times 0,706 \approx 280$$
$$n \times (1 - f_1) = 400 \times 0,294 \approx 120$$
$$n \times f_2 = 400 \times 0,794 \approx 320$$
$$n \times (1 - f_2 = 400 \times 0,206 \approx 80$$

Toutes les valeurs obtenues sont supérieures ou égales à 5, les conditions sont validées, l'intervalle de confiance est valide.
Conclusion : l'item B est vrai.

- Chez les **Baby-boomers** :
  De même, on commence par vérifier la condition sur la taille de l'échantillon : n = 50 soit n ≥ 30.
  On peut calculer l'intervalle de confiance avec $f = \frac{10}{50} = 0,2$

$$ic_{0,95}(p_{bb}) = f \pm z_{\frac{0,05}{2}} \times \sqrt{\frac{f(1-f)}{n}}$$
$$ic_{0,95}(p_{gz}) = 0,2 \pm 2 \times \sqrt{\frac{0,2 \times 0,8}{50}}$$
$$ic_{0,95}(p_{gz}) = 0,2 \pm 2 \times \sqrt{\frac{0,16}{50}}$$
$$ic_{0,95}(p_{gz}) = 0,2 \pm 2 \times \frac{0,4}{\sqrt{50}}$$
$$ic_{0,95}(p_{gz}) = 0,2 \pm \frac{0,8}{\sqrt{50}}$$
$$ic_{0,95}(p_{gz}) = 0,2 \pm 0,11$$
$$ic_{0,95}(p_{gz}) = [0,09; 0,31]$$
De nouveau, on ne peut pas conclure directement. Il faut vérifier les conditions
$$n \times f_1 = 50 \times 0,09 = 4,5$$
Or, notre première condition n'est pas validée : $n \times f_1 \leq 5$.
L'intervalle de confiance n'est pas valide, même si les autres conditions sont validées.
Conclusion : l'item E est faux.
A FAUX Il faut bien penser à minorer la borne inférieure et majorer la borne supérieure.
B VRAI
C FAUX On ne peut pas calculer un intervalle de fluctuation
D FAUX La notation n'est pas bonne ! On calcule un intervalle de confiance d'une proportion théorique *p* et non d'une proportion estimée *f*.
E FAUX

Question 35 – La décision collégiale :

Après 7 jours de débat intense, les tuteurs ont finalement décidé de créer un examen blanc avec seulement des exercices sur les variables aléatoires. Ils mènent une petite recherche sur les performances au sein d'un échantillon de 100 PASS qui possèdent tous une prépa, qu'ils compareront plus tard à un autre échantillon d'étudiants qui ne travaillent qu'avec le Tutorat (ce ne sera pas fait dans cet exercice.).
La moyenne des notes obtenues dans l'échantillon est estimée à 17 et l'écart type à 2.
A la fin de leur étude, ils obtiennent un intervalle de confiance de μ au risque α inconnu : $ic_{1-\alpha}(\mu) = [16,484; 17,516]$.
En effet, au cours de la communication des résultats aux tuteurs, certaines valeurs ont été mystérieusement effacées.

Aide au calcul : $1,96 \approx 2$ $1,645 \approx 1,6$ $2,5758 \approx 2,58$

A. Pour cet intervalle de confiance, le risque α est à 5%
B. Pour cet intervalle de confiance, le risque α est à 10%
C. Pour cet intervalle de confiance, le risque α est à 1%
D. Pour obtenir une précision de 0,258 sans changer le risque α, il faut multiplier le nombre de sujets par 4.
E. Pour obtenir une précision de 0,258 sans changer le risque α, il faut multiplier le nombre de sujets par 2.

<u>Question 35 – La décision collégiale :</u> CD

Analyse du sujet : il faut qu'on cherche le risque α, à partir d'un intervalle donné.

- On sait que :

La précision : $i = \frac{1}{2}(borne\ supérieure - borne\ inférieure) = z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$

On manipule la formule, pour avoir $z_{\frac{\alpha}{2}}$:

$$z_{\frac{\alpha}{2}} = \frac{\sqrt{n}}{s} \times i = \frac{\sqrt{n}}{s} \times \frac{1}{2}(borne\ supérieure - borne\ inférieure)$$

Application numérique :

$$z_{\frac{\alpha}{2}} = \frac{\sqrt{100}}{2} \times \frac{1}{2}(17{,}516 - 16{,}484) = \frac{10}{2} \times \frac{1}{2} \times 1{,}032 = 5 \times 0{,}516 = 2{,}58$$

Vous pouvez aussi choisir de lire la demi-largeur grâce à l'intervalle donné dans l'énoncé : il suffit de soustraire la moyenne à la borne supérieure pour avoir la valeur de $z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$. Ici, c'est 17,516 – 17 = 0,516 Au final, vous avez :

$$z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} = 0{,}516 \quad soit\ z_{\frac{\alpha}{2}} = 0{,}516 \times \frac{\sqrt{n}}{s} = 0{,}516 \times \frac{\sqrt{100}}{2} = 0{,}516 \times \frac{10}{2} = 0{,}516 \times 5 = 2{,}58$$

Avec la table de la fonction de répartition de la loi normale centrée réduite on lit $P(Z \leq z)$:

Soit Z une variable aléatoire suivant une loi normale centrée réduite. Pour une valeur de z donnée, la table donne la probabilité $P(Z \leq z)$

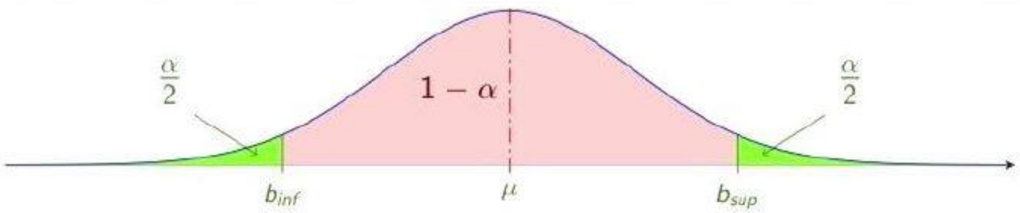
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964

On voit que P(Z ≤ 2,58) = 0,9951, or cela correspond à $1 - \frac{\alpha}{2}$ oui, mais je pense que sans dessin, ça va être dur à comprendre pour vos camarades… (mais je me trompe peut-être)

Donc : $\frac{\alpha}{2} = 1 - 0{,}9951 = 0{,}0049 \approx 0{,}005$

$\alpha = 2 \times 0{,}005 = 0{,}01$ = 1%

On vous remet le schéma du cours pour mieux visualiser :



On peut aussi utiliser la table 2 :
Au lieu de lire comme d'habitude dans la table, on regarde cette fois les valeurs qui sont dans les cases de la table et on cherche celle qui est la plus proche de 2,58. On trouve la valeur 2,5758.
Cette valeur 2,5758 est telle que P(Z> 2,5758)=0,005
donc α /2=0,005 et α = 0,01.

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
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

L'item C est vrai.

- Pour l'item D et E, on n'a pas besoin de passer par les calculs avec les valeurs numériques.

On remarque que pour n_1 = 100, i_1= 0,516.

Pour : $i_2 = 0{,}258 = \frac{0{,}516}{2} = \frac{i_1}{2}$

Avec :

$$i_1 = z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n_1}}$$
$$\sqrt{n_1} = z_{\frac{\alpha}{2}} \times s \times \frac{1}{i_1}$$
$$n_1 = z_{\frac{\alpha}{2}}^{\ 2} \times s^2 \times \frac{1}{i_1^{\ 2}}$$

Donc :

$$i_2 = z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n_2}}$$
$$\frac{i_1}{2} = z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n_2}}$$
$$\sqrt{n_2} = z_{\frac{\alpha}{2}} \times s \times \frac{2}{i_1}$$
$$n_2 = z_{\frac{\alpha}{2}}^{\ 2} \times s^2 \times \frac{1}{i_1^{\ 2}} \times 4 = 4n_1$$

Pour conclure : il faut bien multiplier le nombre de sujets par 4 si la précision est divisée par 2.