Université Claude Bernard Lyon 1

Tutorat Lyon Est

Unité d'Enseignement 3

BANQUE DE QCM

2013 - 2022

COMPARAISON DE MOYENNES ET VARIANCES

QUESTIONS ET REPONSES

2013 - 2021

Question 1 :

L'inflammation est une réaction de défense de l'organisme suite à une agression dont les principaux signes cliniques sont la douleur, la chaleur, le gonflement et la rougeur. Au cours de l'inflammation, on observe une augmentation de la synthèse de protéines dites positives de l'inflammation et notamment de la CRP.
Il s'agit d'un marqueur dont la concentration augmente précocement au cours de l'inflammation. Cependant, d'autres facteurs influent sur la concentration de CRP dont notamment l'inhalation de fumée de cigarette. On constitue ainsi aléatoirement 7 échantillons de 25 personnes chacun ayant inhalé différentes fumées toxiques dont de la fumée de cigarette. On réalise une analyse de variances pour comparer les concentrations moyennes de CRP chez ces individus.
On trouve alors une somme des carrés des écarts totaux (SCT) égale à 14 000 et une somme des carrés des écarts entre colonnes (SCC) égale à 2800.
Aides au calcul : $\frac{28}{112} = \frac{1}{4}$; $168 = 6 \times 28$

A. La grandeur test suit une distribution de Fisher $F(6\,;174)$.
B. La somme des carrés des écarts résiduelle (SCR) vaut 11 200.
C. L'estimation de la variance totale est de $\frac{14000}{174}$.
D. La valeur test est 7.
E. Au risque $\alpha = 5\%$, on peut conclure qu'au moins deux des concentrations moyennes de CRP diffèrent.

Question 1 : BCDE

Dans cet exercice, on réalise un ANOVA (ANalysis Of VAriances). Cela consiste à analyser des variances pour pouvoir comparer des moyennes. Un ANOVA est utilisé pour comparer plusieurs (3 ou plus) moyennes entre elles (si on en dispose que de 2, on réalise un test de comparaison à 2 moyennes « classique »).
Pour cela, dans un ANOVA, on pose toujours l'hypothèse nulle suivante : les moyennes comparées sont toutes égales entre elle. En rejetant cette hypothèse nulle, cela signifie que les moyennes ne sont pas toutes égales, donc il y en a au minimum 2 qui ne sont pas égales entre elles.

Ici, on a 7 échantillons différents, donc on a 7 moyennes différentes. Ce sont ces 7 moyennes que l'on cherche à comparer. On ne connait pas les valeurs de ces moyennes mais on connait leurs variances associées (ou plutôt les sommes de ces variances : SCT, SCC, SCR). On note k le nombre de groupes et N le nombre de sujet au total de l'étude.
On a :

$$k = 7 \qquad N = 7 \times 25 = 175 \qquad SCT = 14000 \qquad SCC = 2800$$

A FAUX Le test de Fisher qu'on va utiliser est à $F(k-1; N-k)$, soit un test de Fisher à $F(6\,;\ 168)$.

B VRAI $SCT = SCR + SCC$, donc $SCR = SCT - SCC = 14000 - 2800 = 11200$.

C VRAI L'estimation de la variance totale correspond à :

$$s_T^2 = \frac{SCT}{N-1} = \frac{14000}{174}$$

D VRAI La valeur test est :

$$F = \frac{s_C^2}{s^2} = \frac{\frac{SCC}{k-1}}{\frac{SCR}{N-k}} = \frac{SCC \times (N-k)}{(k-1) \times SCR} = \frac{2800 \times 168}{6 \times 11200} = \frac{2800}{11200} \times \frac{168}{6} = \frac{28}{112} \times \frac{6 \times 28}{6} = \frac{1}{4} \times 28 = 7$$

E VRAI Les paramètres du test de Fisher sont (6 ; 168). On va donc utiliser une table de Fisher. Cette table est particulière puisqu'elle est fixe pour la probabilité $\alpha = 5\%$, ainsi tous les ANOVA dans le cadre de la PASS sont au risque $\alpha = 5\%$. À partir de cette table, on va déterminer une valeur seuil F_{seuil}.

Pour lire cette table, on a besoin de deux degrés de liberté différents. Le premier est ν_1 et le second ν_2. Ici, $\nu_1 = 6$ et $\nu_2 = 168$. On se place donc à la colonne 6, et entre les lignes 100 et 200 (puisque qu'il n'y a pas de ligne 168, on se place entre les deux lignes qui l'encadre). On a ainsi :

$$2{,}144 < F_{seuil} < 2{,}191$$

Ensuite, on compare cette valeur seuil à la valeur test calculée à l'item D. On remarque que $F > F_{seuil}$. Cela revient à rejeter l'hypothèse nulle, donc au moins 2 des concentrations moyennes comparées sont différentes.

Question 2 :

Le BNP (Brain Natriuretique Peptide) est un peptide natriurétique sécrété par les cardiomyocytes. On observe une hypersécrétion du BNP en cas d'insuffisance cardiaque à cause de l'augmentation des pressions intracardiaques. Il conduit à une vasodilatation et à une diminution du retour veineux.
On cherche à montrer que chez les insuffisants cardiaques, l'obésité entraine une diminution du taux de BNP. On fixe le risque de première espèce à 5%.
Dans la population d'insuffisants cardiaques non obèses, le taux de BNP est en moyenne de 500 ng/L.
On constitue un échantillon aléatoire de 81 insuffisants cardiaques en situation d'obésité et on dose leur taux de BNP. A partir de cet échantillon, le taux moyen de BNP est estimé à 450 ng/L et l'estimation de l'écart-type est de 200 ng/L.

A. On utilise un test de Student à 78 ddl.
B. On rejette H_0 au risque α = 5%.
C. On conclut que le taux moyen de BNP chez les insuffisants cardiaques obèses est significativement supérieur au taux moyen de BNP chez les autres insuffisants cardiaques avec p inférieur à 2%.
D. On conclut que le taux moyen de BNP chez les insuffisants cardiaques obèses est significativement inférieur au taux moyen de BNP chez les autres insuffisants cardiaques avec p inférieur à 2%.
E. On conclut que le taux moyen de BNP chez les insuffisants cardiaques obèses est significativement inférieur au taux moyen de BNP chez les autres insuffisants cardiaques avec p inférieur à 1%.

Question 2 : BD

Ici, il s'agit de comparer une **moyenne observée** de BNP dans un échantillon de 81 insuffisants cardiaques en situation d'obésité avec la **moyenne théorique** de BNP chez les insuffisants cardiaques non obèses.
Dans ce type de comparaison (moyenne théorique/moyenne observée), il faut déterminer quelles sont les conditions de l'exercice :

- La loi suivie : nous ne connaissons pas la loi suivie par la variable « taux moyen de BNP ».
- L'écart-type : nous connaissons **l'estimation** de l'écart-type à partir de l'échantillon
- La taille de l'échantillon : $n = 81 > 30$.

Je vous remets le tableau du poly (p. 111) qui est très important à connaître :

	n ≥ 30	n < 30
X normale **σ connue**	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$
X normale **σ inconnue**	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$ $\to N(0;1)$	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$
X quelconque **σ connue**	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$	HORS PROGRAMME
X quelconque **σ inconnue**	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to N(0;1)$	HORS PROGRAMME

A FAUX On est dans le cas « X quelconque, σ inconnue et $n > 30$ » (case encadrée en rouge juste en haut). Donc on utilise un test ayant recours à la loi normale.

B VRAI Il s'agit d'un test **unilatéral** tel que :

$$H_0 : \mu_{\text{Insuf\_Card\_Obèses}} = \mu_{\text{Insu\_Card\_Non-obèses}}$$

et

$$H_1 : \mu_{\text{Insuf\_Card\_Obèses}} < \mu_{\text{Insuf\_Card\_Non-obèses}}$$

On calcule notre valeur test :

$$Z = \frac{m-\mu}{\frac{s}{\sqrt{n}}} = \frac{450-500}{\frac{200}{\sqrt{81}}} = \frac{-50}{\frac{200}{9}} = \frac{-50\times 9}{200} = \frac{-450}{200} = \frac{-225}{100} = -2{,}25$$

Le test est unilatéral puisque l'on veut montrer une *"diminution du taux de BNP"* (donc pas besoin de multiplier par 2). On calcule donc le degré de significativité à partir de la 1e table de la loi normale :

$$p = P(Z > |-2{,}25|) = 1 - P(Z < 2{,}25) = 1 - 0{,}9878 = 0{,}0122$$

On remarque que $\alpha = 0{,}05 > p$, donc on rejette l'hypothèse nulle.

C FAUX En rejetant l'hypothèse nulle, on conclut qu'il existe une différence significative du taux moyen de BNP. On remarque également que ce taux chez les insuffisants cardiaques obèses est inférieur par rapport aux insuffisants cardiaques non-obèses, et cela pour un degré de significativité de $p = 1{,}12\% < 2$.

D VRAI Cf. C

E FAUX Cf. C

Question 3 :

L'un des symptômes principaux du reflux gastro-œsophagien (RGO) est le pyrosis. Il s'agit d'une sensation douloureuse de brûlure rétro-sternale ascendante due aux remontées de l'acidité gastrique. De nombreuses observations semblent montrer que donner des inhibiteurs de la pompe à protons (IPP) permet de soulager le pyrosis.

On cherche donc à savoir si la douleur due au pyrosis est différente lorsque le patient prend des IPP. Le risque de première espèce α est fixé à 0,5%. Pour cela, on constitue deux échantillons indépendants de 26 personnes chacun : l'échantillon n°1 ne prend pas d'IPP tandis que l'échantillon n°2 en prend. Le

traitement est attribué par tirage au sort. On demande ensuite à chaque participant d'évaluer la douleur due au pyrosis sur une échelle allant de 0 à 20 (0 = pas de douleur ; 20 = douleur maximale possible).
On estime la douleur moyenne et l'écart-type des populations ne prenant pas et prenant des IPP à partir des échantillons correspondants. On regroupe les résultats dans le tableau suivant :

	Echantillon n°1 (pas d'IPP)	Echantillon n°2 (IPP)
Douleur moyenne estimée	12	6
Ecart-type estimé	8	6

Soit X la variable aléatoire associée au niveau de douleur : X suit une loi normale.
On suppose qu'un test d'égalité des variances a été fait au préalable et qu'il a conclu que les variances ne sont pas significativement différentes.
On pose l'hypothèse nulle d'une douleur moyenne identique dans les échantillons.
Aides au calcul : $\sqrt{26} \approx \sqrt{25}$

A. La valeur test calculée vaut environ 3.
B. Il s'agit d'un test utilisant une loi normale centrée réduite.
C. Il s'agit d'un test utilisant une loi de Student à 50 ddl.
D. On conclut que la douleur est en moyenne plus faible sous IPP, et que la différence entre la douleur moyenne avec IPP et sans IPP est statistiquement significative avec p inférieur à 0,5%.
E. On conclut que la douleur est en moyenne plus faible sous IPP, et que la différence entre la douleur moyenne avec IPP et sans IPP est statistiquement significative avec p inférieur à 0,1%.

Question 3 : ACD

Face à ce type d'exercice, il est nécessaire d'identifier la situation. Ici, il s'agit de comparer deux moyennes dans deux échantillons indépendants, donc on compare une **moyenne observée** à une autre **moyenne observée**.
Dans le programme du PASS, il n'y a que 2 possibilités de test de comparaison à 2 moyennes observées :

- Si la taille de chacun des 2 échantillons est supérieure ou égale à 30.
- Si la taille de l'un ou des 2 échantillons est inférieure à 30.

Ici, nous sommes face à la 2e situation puisque les deux échantillons ont une taille de 26. Dans ce cas-là, on utilise une table de Student à $Student(n_1 + n_2 - 2ddl)$.

Pour réaliser un test de comparaison de moyennes observées dans ce cas-ci, il faut que deux conditions soit vérifiés : les variances des deux échantillons ne sont pas significativement différentes et le critère étudié peut être modélisé par une loi normale.

A VRAI Pour le calcul de la valeur test d'une comparaison de deux moyennes observées dont l'une ou les deux sont issues d'un échantillon de taille inférieure à 30, on utilise la formule : *(celle-ci est donnée dans le formulaire lors des épreuves, donc pas apprendre par cœur)*

$$T_{test} = \frac{m_1 - m_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{12-6}{\sqrt{\frac{(26-1)\times 8^2+(26-1)\times 6^2}{26+26-2}}\sqrt{\frac{1}{26}+\frac{1}{26}}} = \frac{6}{\sqrt{\frac{25\times(64+36)}{50}}\sqrt{\frac{2}{26}}}$$
$$= \frac{6}{\sqrt{\frac{100}{2}}\sqrt{\frac{2}{26}}} = \frac{6}{\sqrt{\frac{100}{2}}\sqrt{\frac{2}{25}}} = \frac{6}{\sqrt{\frac{100\times 2}{2\times 25}}} = \frac{6}{\sqrt{\frac{100}{25}}} = \frac{6}{\sqrt{4}} = \frac{6}{2} = 3$$

B FAUX Comme dit plus haut, ce type de test a recours à la loi de Student. Dans cet exercice, la loi de Student s'utilise à $n_1 + n_2 - 2 = 26 + 26 - 2 = 50ddl$

C VRAI Cf. B

D VRAI Maintenant que l'on a calculé la valeur test, on va chercher le degré de significativité p à partir de la table de Student. Pour cela, on commence par se placer sur la ligne du ddl correspondant soit 50, puis on cherche au niveau de quelles colonnes se trouvent la valeur test 3. On remarque que celle-ci se trouve entre 2,9370 et 3,4960 correspondants aux colonnes 0,005 et 0,001. Donc :

$$0{,}001 < p < 0{,}005$$

ddl \ p	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,005	0,001
49	0,1263	0,2547	0,3876	0,5278	0,6795	0,8490	1,0475	1,2991	1,6766	2,0096	2,4049	2,6800	[illegible]	[illegible]
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,4960
60	0,1262	0,2545	0,3872	0,5272	0,6786	0,8477	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,4602

Avec un risque $\alpha = 0{,}5\% = 0{,}005$, on remarque que $p < \alpha$. Donc on rejette H_0 avec p inférieur à 0,5%. On en conclut que la différence entre la douleur moyenne avec IPP et sans IPP est statistiquement significative au seuil de 0,5% : la douleur est en moyenne plus faible sous IPP.

E FAUX On a vu à l'item que $p > 0{,}001$, donc p n'est pas inférieur à $0{,}1\%$.

Question 4 :

Un jour de Noël blanc, on mesure la hauteur de neige tombée pendant la journée dans 36 communes de Savoie et dans 36 communes de Haute-Savoie. On cherche à montrer que la quantité de neige tombée en Savoie est significativement supérieure à celle tombée en Haute-Savoie. On fixe $\alpha = 1\%$.
On estime la hauteur moyenne de neige tombée le jour de Noël dans ces communes ainsi que la variance et on regroupe les résultats dans le tableau ci-dessous :

	Savoie	Haute-Savoie
Hauteur moyenne estimée	33	31
Variance	11	14

A. Le test est significatif pour $\alpha = 1\%$.
B. La valeur test calculée est égale à 2,4.
C. On ne peut pas conclure à une différence significative entre la Savoie et la Haute-Savoie.
D. La quantité de neige tombée en Savoie est significativement supérieure à celle tombée en Haute-Savoie avec $0{,}004 < p < 0{,}005$.
E. La quantité de neige tombée en Savoie est significativement supérieure à celle tombée en Haute-Savoie avec $0{,}008 < p < 0{,}009$.

Question 4 : ABE

Face à ce type d'exercice, il est nécessaire d'identifier la situation. Ici, il s'agit de comparer deux moyennes dans deux échantillons indépendants, donc on compare une **moyenne observée** à une autre **moyenne observée**.
Dans le programme du PASS, il n'y a que 2 possibilités de test de comparaison à 2 moyennes observées :

- Si la taille de chacun des 2 échantillons est supérieure ou égale à 30.
- Si la taille de l'un ou des 2 échantillons est inférieure à 30.

Ici, nous sommes face à la 1e situation puisque les deux échantillons ont une taille de 36. Dans ce cas-là, on utilise une table de la loi normale.
C'est un test unilatéral tel que :

$$H_0 : \mu_S = \mu_{HS}$$
$$H_1 : \mu_S > \mu_{HS}$$

On peut calculer la valeur test :

$$T = \frac{m_S - m_{HS}}{\sqrt{\frac{s_S^2}{n_S} + \frac{s_{HS}^2}{n_{HS}}}} = \frac{33 - 31}{\sqrt{\frac{11}{36} + \frac{14}{36}}} = \frac{2}{\sqrt{\frac{25}{36}}} = \frac{2}{\frac{5}{6}} = \frac{12}{5} = 2{,}4$$

On cherche ensuite le degré de significativité en utilisant la 1ère table de la loi normale, sachant que l'on a un test unilatéral donc :

$$p = P(Z > 2{,}4) = 1 - P(Z < 2{,}4) = 1 - 0{,}9918 = 0{,}0082$$

A VRAI Au risque $\alpha = 1\% = 0{,}01$, on remarque que $p < \alpha$, donc le test est significatif pour $\alpha = 1\%$. On peut donc dire qu'il existe une différence significative de la hauteur de neige tombée entre la Savoie et la Haute-Savoie.

B VRAI Cf. Intro
C FAUX Cf. A
D FAUX La quantité de neige tombée en Savoie est bien significativement supérieure à celle tombée en Haute-Savoie mais $p = 0{,}0082$ n'est pas compris entre 0,004 et 0,005.

E VRAI La quantité de neige tombée en Savoie est bien significativement supérieure à celle tombée en Haute-Savoie avec $p = 0{,}0082$ donc avec $0{,}008 < p < 0{,}009$.

Question 5 :

La thyréostimuline ou TSH est une hormone sécrétée par l'antéhypophyse et qui joue un rôle dans la production d'hormones thyroïdiennes. Dans la population générale, le taux de thyréostimuline suit une loi normale de paramètres $\mu = 2{,}2$ mUI/L et $\sigma = 0{,}4$ mUI/L. On cherche à savoir si ce taux est différent en région Rhône-Alpes par rapport à la population générale. On constitue donc dans cette région un échantillon aléatoire représentatif de 100 personnes. On trouve une moyenne de 2,3 mUI/L et un écart-type de 0,2 mUI/L dans l'échantillon. On cherche à savoir si le taux en région Rhône-Alpes diffère de celui de la population générale.

A. La valeur du test t est de 4.
B. Au risque $\alpha = 5\%$, on rejette l'hypothèse nulle d'égalité des moyennes.
C. Au risque $\alpha = 1\%$, on rejette l'hypothèse nulle d'égalité des moyennes.
D. Il est nécessaire d'utiliser la table de Fisher dans cet exercice.
E. Au risque $\alpha = 1\%$, on peut dire que le taux de TSH est significativement plus élevé dans la région Rhône-Alpes.

Question 5 : B

Ici, il s'agit de comparer une **moyenne observée** du taux de TSH dans un échantillon de 100 sujets de la région Rhône-Alpes avec la **moyenne théorique** du taux de TSH dans la population générale.
Dans ce type de comparaison (moyenne théorique/moyenne observée), il faut déterminer quelles sont les conditions de l'exercice :

- La loi suivie : la variable « taux de TSH » suit une loi normale.
- L'écart-type : nous connaissons **le paramètre** de l'écart-type de la population.

- La taille de l'échantillon : $n = 100 > 30$.

	n ≥ 30	n < 30
X normale **σ connue**	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$
X normale **σ inconnue**	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$ $\to N(0;1)$	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$
X quelconque **σ connue**	$\boxed{\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)}$	HORS PROGRAMME
X quelconque **σ inconnue**	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to N(0;1)$	HORS PROGRAMME

A FAUX On calcule la valeur test :

$$t = \frac{m-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{2{,}3-2{,}2}{\frac{0{,}4}{\sqrt{100}}} = \frac{0{,}1}{0{,}4} \times 10 = 0{,}25 \times 10 = 2{,}5$$

B VRAI On calcule maintenant le degré de significativité p en utilisant la 1ère table de la loi normale sachant que le test est bilatéral (donc on multiplie par 2 ce que l'on lit dans la table) :

$$p = 2 \times P(Z > 2{,}5) = 2 \times \big(1 - P(Z < 2{,}5)\big) = 2 \times (1 - 0{,}9938) = 2 \times 0{,}0062 = 0{,}0124$$

Ainsi, au risque $\alpha = 5\% = 0{,}05$, on remarque que $p < \alpha$ donc on rejette H_0.

C FAUX Au risque $\alpha = 1\% = 0{,}01$, on remarque que $p > \alpha$ donc on ne peut pas rejeter H_0. Il n'y a pas de différence significative du taux de TSH.

D FAUX Il fallait utiliser la table de la loi normale centrée réduite dans cet exercice, car nous sommes dans le cas où X suit une loi normale dans la population générale d'écart type σ connu.

E FAUX Cf. C

Question 6 :

L'atopie désigne une susceptibilité pour un individu de présenter des signes cliniques lors du contact avec des allergènes inoffensifs pour le reste de la population. Elle est due à une production anormale d'immunoglobulines E. Bien que le caractère héréditaire ait déjà été souligné, l'environnement semble aussi être un facteur de développement de cette maladie. On constitue des échantillons aléatoires de 20 sujets atopiques dans 10 pays différents (20 patients par échantillon). Une analyse de variance est effectuée afin de comparer les niveaux moyens d'immunoglobulines E entre pays.

Après acquisition des résultats, nous obtenons une SCR (somme des carrés des écarts résiduelle) de 9500 et une SCC (sommes des carrés des écarts entre colonnes) de 1800.

A. La SCT (somme des carrés des écarts totaux) est de 11 300.
B. On utilise un test de Fisher de paramètres $F(9\,;\ 190)$.
C. La valeur test calculée vaut 4.
D. Au risque $\alpha = 5\%$, on rejette l'hypothèse nulle d'une égalité des moyennes entre les différents échantillons.
E. Au risque $\alpha = 5\%$, on peut émettre l'hypothèse qu'au moins deux des niveaux moyens d'immunoglobulines diffèrent.

Question 6 : ABCDE

Dans cet exercice, on réalise un ANOVA (ANalysis Of VAriances). Ici, on a 10 échantillons différents (10 pays différents), donc on a 10 moyennes différentes. Ce sont ces 10 moyennes que l'on cherche à comparer. On ne connait pas les valeurs de ces moyennes mais on connait leurs variances associées (ou plutôt les sommes de ces variances : SCT, SCC, SCR). On note k le nombre de groupes et N le nombre de sujet au total de l'étude.

On a : $k = 10$ $\quad N = 10 \times 20 = 200$ $\quad SCR = 9500$ $\quad SCC = 1800$

A VRAI La SCT correspond à la somme de la SCR et de la SCC : $SCT = 9500 + 1800 = 11\,300$.

B VRAI Lors d'un ANOVA, on utilise un test de Fisher dont les paramètres sont : $F(k-1\,;\,N-k)$. Soit $F(9\,;\,190)$.

C VRAI La valeur test vaut : $F = \frac{s_C^2}{s^2} = \frac{\frac{SCC}{k-1}}{\frac{SCR}{N-k}} = \frac{\frac{1800}{9}}{\frac{9500}{190}} = \frac{200}{50} = 4$

D VRAI Nous avons notre valeur test calculée qui vaut 4. Il faut alors la comparer à notre valeur seuil que l'on va trouver dans la table de Fisher à $F(9\,;\,190)$. On commence par se placer sur la colonne 9. Ensuite, pour la ligne, on se place entre 100 et 200 puisque la ligne 190 n'existe pas dans cette table. La valeur seuil se situe donc entre 1,927 et 1,975. Donc notre valeur test est supérieure à la valeur seuil : on rejette l'hypothèse nulle H0 au risque $\alpha = 5\%$.

E VRAI On rejette l'hypothèse nulle au risque $\alpha = 5\%$ donc cela signifie qu'au moins deux des niveaux moyens d'immunoglobulines diffèrent.

Question 7 :

Dans le cancer, les cellules tumorales produisent des molécules immunosuppressives qui entrainent un phénomène de tolérance qui les protègent de l'action du système immunitaire. Les chercheurs se sont intéressés à la quantité et à la qualité des cellules dendritiques présentes au niveau de mélanome. Afin d'étudier l'impact de certaines cellules dendritiques, ils ont créé 2 groupes : un groupe contrôle avec 100 patients atteints de mélanome et un groupe avec 30 patients atteints de mélanome ayant reçu un nouveau traitement.

Ils ont ensuite comparé la survie entre ces 2 groupes. La survie dans le groupe contrôle est en moyenne de 7 mois avec un écart-type estimé de 2 mois contre 22 mois avec un écart-type estimé de 5 mois ayant reçu un nouveau traitement. On effectue un test unilatéral.

Aides au calcul : $\frac{25}{30} = 0{,}84\,;\ \sqrt{0{,}88} = 1$

A. La valeur test vaut 2.
B. Pour un risque $\alpha = 5\%$, on rejette l'hypothèse nulle.
C. Pour un risque $\alpha = 0{,}1\%$, on ne rejette pas l'hypothèse nulle.
D. Au risque $\alpha = 0{,}1\%$, on déclare que le nouveau traitement entraine une amélioration de la survie des patients atteints de mélanome.
E. Il manque des données pour pouvoir calculer la valeur test et donc résoudre l'exercice.

Question 7 : BD

Il s'agit d'un test de comparaison de 2 moyennes observées dans le cas où les effectifs des 2 échantillons sont supérieurs ou égaux à 30.

A FAUX La valeur test vaut donc :

$$z = \frac{m_1 - m_2}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} = \frac{22\ -\ 7}{\sqrt{\frac{25}{30} + \frac{4}{100}}} = \frac{15}{\sqrt{0{,}88}} = 15$$

B VRAI On cherche maintenant à calculer le degré de significativité à partir de la 1ère table de la loi normale sachant que le test est unilatéral :

$$p = P(Z > 15) = 1 - P(Z < 15) = 1 - P(Z < 4{,}09) = 1 - 0{,}99998 = 0{,}00002$$

La 1ère table de la loi normale ne va pas au-delà de 4,09. Notre valeur test étant supérieure à celle-ci, on lit dans la table pour cette valeur.

Ainsi pour un risque α à 0,05, on remarque que p est très largement inférieur, on peut donc rejeter l'hypothèse nulle au risque $\alpha = 5\%$.

C FAUX Là encore p est inférieur à α qui vaut ici 0,001 donc on rejette encore l'hypothèse nulle au risque $\alpha = 0{,}1\%$.

D VRAI À l'item C, on a rejeté l'hypothèse nulle, donc on conclut en une différence significative du temps moyen de survie en faveur du groupe avec le nouveau traitement.

E FAUX

Question 8 :

De plus en plus d'études se sont penchées sur les effets de la présence des animaux sur la santé des humains. Pour de nombreuses personnes, la présence d'un animal de compagnie peut être un facteur de santé physique et psychologique très important. La zoothérapie, thérapie résultante de ces recherches, semble apporter de nombreux bénéfices, allant de la simple relaxation à la diminution d'importants facteurs de stress, en passant par le soutien social et une meilleure récupération postopératoire.
Concernant plus spécifiquement les fractures de l'avant-bras chez les enfants, on souhaite savoir si la présence d'un animal dans le programme de rééducation fonctionnelle a un impact sur le temps de rétablissement. On effectuera un test bilatéral.
Soit X, la variable aléatoire qui correspond au temps de rétablissement usuel pour une fracture de l'avant-bras. X suit une loi statistique de moyenne $\mu = 60$ et d'écart-type $\sigma = 4$. Soit M, la variable aléatoire caractérisant la moyenne des temps de rétablissement sur un échantillon de 144 enfants.
On étudie un échantillon de 144 enfants ayant eu un programme de zoothérapie lors de leur rééducation.
Résultat : La moyenne calculée du temps de rétablissement est de 58 jours dans l'échantillon.

On rejette un temps de rétablissement identique d'une fracture de l'avant-bras entre la population pédiatrique générale et les enfants ayant bénéficiés de la zoothérapie pour :

A. $\alpha = 5\%$
B. $\alpha = 1\%$
C. $\alpha = 0{,}01\%$
D. $\alpha = 0{,}003\%$
E. $\alpha = 0{,}001\%$

Question 8 : ABC

Ici, il s'agit de comparer une **moyenne observée** du temps de rétablissement dans un échantillon de 144 sujets de la région Rhône-Alpes avec la **moyenne théorique** du temps de rétablissement dans la population générale.
Dans ce type de comparaison (moyenne théorique/moyenne observée), il faut déterminer quelles sont les conditions de l'exercice :

- La loi suivie : la variable « temps de rétablissement » suit une loi quelconque.
- L'écart-type : nous connaissons **le paramètre** de l'écart-type de la population.

- La taille de l'échantillon : $n = 144 > 30$.

	$n \geq 30$	$n < 30$
X normale σ connue	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)$	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)$
X normale σ inconnue	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow Student(n-1ddl)$ $\rightarrow N(0;1)$	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow Student(n-1ddl)$
X quelconque σ connue	$\boxed{\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)}$	HORS PROGRAMME
X quelconque σ inconnue	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow N(0;1)$	HORS PROGRAMME

On calcule donc la valeur test :

$$z = \frac{m-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{58-60}{\frac{4}{\sqrt{144}}} = \frac{-2}{4} \times 12 = -6$$

On cherche ensuite le degré de significativité p à partir de la 1ère table de la loi normale sachant que le test est bilatéral :

$$p = 2 \times P(Z > |-6|) = 2 \times \big(1 - P(Z < 6)\big) = 2 \times (1 - 0{,}99998) = 2 \times 0{,}00002 = 0{,}00004$$

On compare à présent p aux différents risques α donnés dans les items. On remarque que p est inférieur à 5%, 1% et 0,01% mais pas à 0,003% et 0,001%. Donc **A, B et C VRAI**, et **D et E FAUX**.

Question 9 :

Valentin se passionne pour les éléphants. Fasciné par la variabilité des espèces, il se demande si la taille des oreilles pourrait être un bon critère de différenciation entre les éléphants d'Afrique et les éléphants d'Asie (test bilatéral). Il décide donc de monter sa propre étude et s'envole pour l'Afrique où il observe un échantillon aléatoire de 100 éléphants. La moyenne des oreilles observée dans cet échantillon est de 110 cm et la variance estimée est de 100 cm². Il se rend ensuite en Asie où il va étudier un échantillon aléatoire de 50 éléphants. La moyenne des oreilles observée est de 105 cm et la variance estimée dans cet échantillon est de 392 cm².
Valentin rentre finalement chez lui pour mettre au propre ses observations et conclure si les différences observées sont significatives ou non. Le risque de première espèce est fixé à 5%.

A. La moyenne de la taille des oreilles des éléphants d'Afrique est de 110cm.
B. Valentin ne connaît pas les lois des variables aléatoires qu'il a observées, et l'expérience s'arrête donc là.
C. Si on utilise la loi Normale, la statistique de test calculée est égale à 3.
D. Valentin rejette l'hypothèse nulle d'égalité des moyennes au niveau de significativité $0{,}04 < p < 0{,}05$.
E. Valentin ne peut pas rejeter l'hypothèse nulle d'égalité des moyennes. Le test ne permet pas de conclure à une différence significative.

Question 9 : E

Il s'agit d'un test de comparaison de 2 moyennes observées dans le cas où les effectifs des 2 échantillons sont supérieurs ou égaux à 30.

A FAUX C'est la moyenne observée dans l'échantillon qui est de 110 cm. La moyenne théorique est quant à elle inconnue.

B FAUX On pose X la taille des oreilles de l'éléphant. D'après le Théorème Centrale Limite, la moyenne de la taille des oreilles suit donc une loi normale car n est supérieur à 30.

C FAUX On calcule la valeur test :

$$z = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = Z = \frac{110 - 105}{\sqrt{\frac{100}{100} + \frac{400}{50}}} = \frac{5}{\sqrt{1+8}} = \frac{5}{3} \approx 1{,}67$$

D FAUX On cherche à présent le degré de significativité sachant que le test est bilatéral :

$$p = 2 \times P(Z > 1{,}67) = 2 \times \left(1 - P(Z < 1{,}67)\right) = 2 \times (1 - 0{,}9525) = 2 \times 0{,}0475 = 0{,}095$$

Au risque $\alpha = 5\% = 0{,}05$, on remarque que $p > \alpha$. On ne peut donc pas rejeter l'hypothèse nulle.

E VRAI Cf. D

Question 10 :

Un groupe d'amis s'intéresse à la relation entre le sexe et la taille moyenne du tour de bras. Pour cela, deux échantillons aléatoires de 11 filles et 11 garçons sont étudiés. On note X, la variable aléatoire gaussienne correspondant à la taille du tour de bras. La moyenne retrouvée chez les filles est de 27 cm avec une variance estimée de 2 cm², et la moyenne retrouvée chez les garçons est de 31 cm avec une variance estimée de 6 cm². On réalise un test bilatéral et on admet un risque $\alpha = 0{,}05$.

A. L'hypothèse nulle du test d'égalité des variances est « les variances des deux échantillons ne sont pas significativement différentes ».
B. Les variances sont significativement différentes.
C. On ne rejette pas l'hypothèse nulle d'égalité des moyennes.
D. Les conditions d'utilisation du test de Student ne sont pas remplies.
E. On rejette l'hypothèse nulle d'égalité des moyennes car la statistique de test calculée T est supérieure à t_{seuil}.

Question 10 : ABD

Il s'agit d'un test de comparaison de 2 moyennes observées dans le cas où les effectifs des 2 échantillons sont inférieurs à 30. Il faut ainsi que 2 conditions soient vérifiées : la variable que l'on étudie doit suivre ou être approximée par une loi normale (précisé dans l'énoncé) et les variances ne doivent pas être significativement différentes (non précisé).

On doit donc commencer par tester l'égalité des variances avec un test de comparaison de variances. On teste l'hypothèse nulle H_0 : $\sigma_1^2 = \sigma_2^2$ qui est donc que les variances théoriques des deux échantillons sont égales. Pour cela, on commence par calculer comme toujours la valeur test. Dans un test de comparaison de variances, la valeur test correspond tout simplement à la variance la plus grande divisée par la variance la plus petite :

$$F = \frac{s_G{}^2}{s_F{}^2} = \frac{6}{2} = 3$$

On cherche ensuite la valeur seuil en utilisant une table de Fisher. *Pour rappel, dans le cadre de la PASS, à partir d'une table de Fisher, il n'est possible de chercher que des valeurs seuils et non pas des degrés de significativité.*

Pour cela, on doit définir les degrés de libertés qui correspondent à $F(n_1 - 1; n_2 - 1)$. Pour savoir qui est n_1 et qui est n_2, il faut regarder les variances : n_1 correspond à l'échantillon d'où provient la plus grande variance (soit l'échantillon de garçons) et n_2 correspond à l'échantillon d'où provient la plus petite variance (soit l'échantillon de filles). On lit donc dans la table de Fisher à $F(10; 10)$ (ici, la taille des échantillons est identique donc savoir qui est n_1 et qui est n_2 n'a pas vraiment d'importance). On obtient ainsi $F_{seuil} = 2{,}978$.

On compare maintenant la valeur seuil et la valeur test : on remarque que $F > F_{seuil}$. Dans ce cas-là, on rejette l'hypothèse nulle d'égalité des variances, et on conclut que les variances sont significativement différentes. Ainsi, il n'est pas possible de comparer les deux moyennes, puisque la condition d'égalité des variances n'est pas validée.

A VRAI C'est l'une des conditions à vérifier pour comparer 2 moyennes issues d'échantillons de taille inférieure à 30.

B VRAI Cf. Intro

C FAUX On ne peut pas le savoir puisque l'on ne peut pas réaliser le test de comparaison de moyennes.

D VRAI Si les variances n'étaient pas significativement différentes, nous aurions réalisés le test d'égalité des moyennes en utilisant la loi de Student. Puisque les variances sont significativement différentes, on ne peut pas utiliser la loi de Student pour tester l'égalité des moyennes.

E FAUX Cf. C

Question 11 :

Un essai thérapeutique comparatif randomisé en double insu compare l'efficacité de deux traitements A et B visant à diminuer la valeur d'un paramètre biologique R (exprimée en mmol/L). A l'issue de l'étude, 100 patients ont été randomisés (50 dans chaque bras). Le risque de première espèce est fixé à 5%. Concernant le critère du jugement R, les moyennes calculées et variances estimées sont respectivement :

$$m_A = 17{,}5 \ mmol/L$$
$$m_B = 13{,}5 \ mmol/L$$
$${s_A}^2 = 95 \ mmol^2/L^2$$
$${s_B}^2 = 105 \ mmol^2/L^2$$

A. Sous l'hypothèse nulle, la variable aléatoire et grandeur test Z suit une loi normale centrée réduite.

B. La valeur prise par la variable aléatoire grandeur test est 3.

C. Vous rejetez l'hypothèse nulle d'égalité des moyennes de R dans les populations traitées par A et B au niveau de significativité $p < 0{,}01$.

D. Vous déclarez que la moyenne de R est plus basse chez les patients traités par B, au niveau de significativité $p < 0{,}05$.

E. L'estimation ponctuelle de la différence des valeurs moyennes de R chez les patients traités par A et B est de 4 mmol/L.

Question 11 : ADE

Il s'agit d'un test de comparaison de 2 moyennes observées dans le cas où les effectifs des 2 échantillons sont supérieurs ou égaux à 30.

A VRAI La grandeur test que nous utilisons dans ce type de comparaison est la variable aléatoire Z qui suit une loi normale centrée réduite.

$$Z = \frac{m_A - m_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \rightarrow N(0;1)$$

B FAUX On calcule la valeur test z :

$$Z = \frac{m_A - m_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{17{,}5 - 13{,}5}{\sqrt{\frac{95}{50} + \frac{105}{50}}} = \frac{4}{\sqrt{\frac{200}{50}}} = \frac{4}{\sqrt{4}} = 2$$

C FAUX On calcule ensuite p, le degré de significativité. Pour cela, on va chercher dans la 1ère table de la loi normale, sachant que le test est bilatéral (dans l'énoncé, on compare simplement les 2 efficacités sans préjuger d'une supériorité ou d'une infériorité) :

$$p = 2 \times P(Z > 2) = 2 \times \left(1 - P(Z < 2)\right) = 2 \times (1 - 0{,}9772) = 0{,}0456$$

Le degré de significativité n'est donc pas inférieur à 0,01.

D VRAI Au risque $\alpha = 5\% = 0{,}05$, on remarque que $p < \alpha$, donc on rejette l'hypothèse nulle d'égalité des moyennes et on conclut que la moyenne de R est plus basse chez les patients traites par B.

E VRAI L'estimation ponctuelle de la différence des valeurs moyennes de R chez les patients traités par A et B correspond à $m_A - m_B = 17{,}5 - 13{,}5 = 4\, mmol/L$.

Question 12 :

Plusieurs études ont soulevé l'hypothèse d'un risque accru de diabète chez les femmes après 50 ans ayant préalablement développé durant leurs grossesses un diabète gestationnel (pas de diabète avant ni après la grossesse) avec un taux de sucre dans le sang supérieur à 1,3g/L à jeun.

On souhaite approfondir cette question. Pour cela, on souhaite savoir si la quantité de liquide amniotique est modifiée en cas de diabète gestationnel. On étudie, sur un échantillon aléatoire de 100 femmes atteintes de diabète gestationnel, la quantité de liquide amniotique à 3 semaines du terme, exprimée en mL (soit la valeur prise par une variable X sur cet échantillon). On précise que dans la population générale, la variable X a pour moyenne 900 mL, et que son écart type vaut 400mL.

On note ici la moyenne calculée sur cet échantillon :

$$m = 1000\, mL$$

A. Les conditions de l'énoncé conduisent à l'utilisation de la loi normale.
B. Les conditions de l'énoncé conduisent à l'utilisation de la loi de Student.
C. Les conditions de l'énoncé ne permettent pas de comparer les moyennes dans le cadre de la PASS.
D. Pour un risque de première espèce à 5%, on rejette l'hypothèse nulle.
E. Pour un risque de première espèce à 1%, on ne rejette pas l'hypothèse nulle.

Question 12 : ADE

Ici, il s'agit de comparer une **moyenne observée** de la quantité de liquide amniotique dans un échantillon de 100 femmes atteintes de diabète gestationnel avec la **moyenne théorique** de la quantité de liquide amniotique de la population générale.

Dans ce type de comparaison (moyenne théorique/moyenne observée), il faut déterminer quelles sont les conditions de l'exercice :

- La loi suivie : la variable « temps de rétablissement » suit une loi quelconque.
- L'écart-type : nous connaissons **le paramètre** de l'écart-type de la population.
- La taille de l'échantillon : $n = 144 > 30$.

	n ≥ 30	n < 30
X normale **σ connue**	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$
X normale **σ inconnue**	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$ $\to N(0;1)$	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$
X quelconque **σ connue**	$\boxed{\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)}$	HORS PROGRAMME
X quelconque **σ inconnue**	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to N(0;1)$	HORS PROGRAMME

A VRAI Avec les conditions de l'énoncé (X quelconque, écart-type de la population et $n \geq 30$), on peut utiliser la loi normale (voir le tableau ci-dessus).

B FAUX Cf. A

C FAUX Cf. A

D VRAI On commence par calculer la valeur test z :

$$z = \frac{m-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1000-900}{\frac{400}{\sqrt{100}}} = \frac{100}{\frac{400}{10}} = \frac{10}{4} = 2{,}5$$

On cherche ensuite le degré de significativité p à partir de la 1ère table de la loi normale sachant que le test est bilatéral :

$$p = 2 \times P(Z > 2{,}5) = 2 \times \big(1 - P(Z < 2{,}5)\big) = 2 \times (1 - 0{,}9938) = 0{,}0124$$

Au risque $\alpha = 5\% = 0{,}05$, on remarque que $p < \alpha$, donc on rejette l'hypothèse nulle.

E VRAI Au risque $\alpha = 1\% = 0{,}01$, on remarque que $p > \alpha$, donc on ne rejette pas l'hypothèse nulle.

Question 13 :

On souhaite comparer la durée d'hospitalisation des patients atteints d'une BPCO (bronchopneumopathie chronique obstructive) dans 5 services différents afin de comparer les performances de ces services pour la prise en charge des patients. Pour cela, on observe la durée d'hospitalisation de 120 patients dans chaque service. On suppose que les conditions de l'ANOVA sont réunies.

On mesure un paramètre biologique dont on note les moyennes de temps (et quelques variances) dans ces 5 services, respectivement m_A, m_B, m_C, m_D et m_E pour les services A, B, C, D et E :

- $m_A = 8\ jours$ et $s_A = 6\ jours$
- $m_B = 9\ jours$
- $m_C = 14\ jours$
- $m_D = 10\ jours$ et $s_D = 8\ jours$
- $m_E = 7\ jours$

On cherche à savoir si les moyennes des temps d'hospitalisation diffèrent significativement au seuil de 5%. On note aussi les différents résultats de cette étude : $SCC = 3504$; $SCT = 10000$; $SCR = 6496$

Aides au calcul : $\frac{3504}{4} = 880$; $\frac{6496}{115} \approx 56$; $\frac{6496}{595} \approx 11$; $\frac{3504}{115} \approx 30$; $\frac{6496}{599} \approx 10$; $\sqrt{120} \approx 11$

A. L'analyse de variances va permettre de tester l'hypothèse d'égalité des moyennes des durées d'hospitalisation.
B. On utilise la table de Fischer à $(4\,;599)$ ddl.
C. La valeur test est égale à 5.
D. Vous rejetez l'hypothèse nulle d'égalité des moyennes au risque de première espèce α de 5%.
E. En utilisant un test unilatéral, les moyennes m_A et m_D sont significativement différentes au risque α corrigé.

<u>Question 13 :</u> AD

Dans cet exercice, on réalise un ANOVA (ANalysis Of VAriances). Ici, on a 5 échantillons différents (5 services différents), donc on a 5 moyennes différentes. Ce sont ces 5 moyennes que l'on cherche à comparer. On connait les valeurs de ces moyennes et on connait également leurs variances associées (ou plutôt les sommes de ces variances : SCT, SCC, SCR). On note k le nombre de groupes et N le nombre de sujet au total de l'étude.

On a : $k = 5 \qquad N = 5 \times 120 = 600 \qquad SCT = 10000 \qquad SCC = 3504 \qquad SCR = 6496$

A VRAI Lorsque l'on compare plusieurs moyennes entre elles, on utilise l'ANOVA qui, en comparant les variances (donc utilisation de la loi de Fisher), va permettre de comparer les moyennes.

B FAUX On va donc utiliser la loi de Fischer. Pour cela, il nous faut définir les degrés de liberté (ddl) qui sont à $(k-1; N-k)$, soit $F(4; 595)\ ddl$.

C FAUX Pour calculer la valeur test, il faut d'abord calculer les estimations des variances :

$$s_C^2 = \frac{SCC}{k-1} = \frac{3504}{5-1} = \frac{3504}{4} \approx 880$$
$$s^2 = \frac{SCR}{N-k} = \frac{6496}{595} \approx 11$$

On calcule la valeur test :

$$F = \frac{s_C^2}{s^2} = \frac{880}{11} = 80$$

D VRAI En reprenant les calculs de l'item précèdent, on lit, dans la table de Fischer, à 4 et à 595 ddl que la valeur seuil retenue est inférieure à 2,39 (valeur seuil pour un test à 4 et 500 ddl).

En comparant la valeur test et la valeur seuil, on remarque que $F > F_{seuil}$, donc on rejette l'hypothèse nulle d'égalité des moyennes. Ceci implique que l'on peut comparer les 5 moyennes deux à deux, pour savoir lesquelles sont réellement différentes entre elles.

E FAUX On demande maintenant de comparer 2 moyennes entre elles. Pour cela, on réalise un test de comparaison de deux moyennes observées avec des échantillons de tailles supérieurs ou égales à 30.
On commence par calculer la valeur test :

$$z = \frac{m_A - m_D}{\sqrt{\frac{s_A{}^2}{n} + \frac{s_D{}^2}{n}}} = \frac{8-10}{\sqrt{\frac{6^2}{120} + \frac{8^2}{120}}} = \frac{-2}{\sqrt{\frac{36+64}{120}}} = \frac{-2}{\sqrt{100}} \times \sqrt{120} = \frac{-2}{10} \times 11 = -2{,}2$$

On cherche ensuite le degré de significativité en utilisant cette fois-ci la 1ère table de la loi normale sachant que le test est unilatéral :

$$p = P(Z > |-2{,}2|) = 1 - P(Z < 2{,}2) = 1 - 0{,}9861 = 0{,}0139$$

Avant de comparer p à α, il ne faut pas oublier de corriger ce risque α. En effet, dans la comparaison des moyennes deux à deux après l'utilisation d'un ANOVA, le risque α doit être corrigé par la correction de Bonferroni. Celle-ci consiste à diviser le risque α par le nombre de paires Q possibles entre les différentes moyennes. Ici, nous avons 5 moyennes différentes, soit 10 paires possibles. Le risque α corrigé vaut donc : $\alpha' = \frac{\alpha}{Q} = \frac{0{,}05}{10} = 0{,}005$.

Le nombre de paires possibles, c'est le nombre de duos qu'on peut former avec les moyennes dont on dispose. Par exemple, dans cet item, on s'intéresse au duo A et D, mais on aurait pu s'intéresser au duo A et E ou B et C ... Au final, avec ces 5 moyennes, il y a 10 duos possibles. Globalement, vous pouvez retenir que pour 3 moyennes, il y a 3 duos possibles ; pour 4 moyennes, il y en a 6 ; et pour 5 moyennes, il y en a 10.
Au risque corrigé $\alpha' = 0{,}005$, on remarque que $p > \alpha'$. Donc on ne peut pas rejeter l'hypothèse nulle et on ne peut pas dire que les moyennes m_A et m_D sont significativement différentes.

Question 14 :

Plusieurs observations semblent affirmer un lien entre le taux d'alcoolémie et la taille. Pour cela, on étudie deux échantillons indépendants de 25 individus tirés de deux populations : un échantillon A tiré d'une population connue pour son penchant à la boisson ; et un échantillon B tiré d'une population plus raisonnable.
Soit, X la variable aléatoire associé à la taille et suivant une loi normale.
On résume donc les différentes valeurs ci-dessous :

	Groupe A	Groupe B
Moyenne	165	175
Ecart-type estimé	6	8

On suppose les variances inconnues égales à σ^2. L'hypothèse alternative correspond à une différence des tailles en faveur de la population plus raisonnable.

Aides au calcul : $\sqrt{2} \approx 1{,}4$; $\sqrt{3} \approx 1{,}7$; $\sqrt{5} \approx 2{,}2$; $\sqrt{6} \approx 2{,}4$; $\sqrt{7} \approx 2{,}5$

A. Les conditions d'application de la loi normale sont vérifiées.
B. Les conditions d'application de la loi de Student sont vérifiées.
C. Notre valeur test vaut 5.
D. Au seuil de 5%, on rejette l'hypothèse nulle.
E. Au seuil de 1%, on ne rejette pas l'hypothèse nulle.

Question 14 : BCD

Il s'agit d'un test de comparaison de 2 moyennes observées dans le cas où les effectifs des 2 échantillons sont inférieurs à 30. Il faut ainsi que 2 conditions soient vérifiées : la variable que l'on étudie doit suivre ou être approximée par une loi normale (précisé dans l'énoncé) et les variances ne doivent pas être significativement différentes (également précisé).

On peut ainsi comparer les 2 moyennes observées en utilisant une table de Student à $Student(n_1 + n_2 - 2ddl)$, soit à $Student(48ddl)$.

A FAUX Cf. Intro

B VRAI Cf. Intro

C VRAI On calcule la valeur test (la formule n'est pas à connaître) :

$$t = \frac{m_B - m_A}{\sqrt{\frac{(n_B-1)s_B^2+(n_A-1)s_A^2}{n_B+n_A-2}}\sqrt{\frac{1}{n_B}+\frac{1}{n_A}}} = \frac{175-165}{\sqrt{\frac{(25-1)8^2+(25-1)6^2}{25+25-2}}\sqrt{\frac{1}{25}+\frac{1}{25}}} = \frac{10}{\sqrt{\frac{24\times(64+36)}{48}}\sqrt{\frac{2}{25}}} = \frac{10}{\sqrt{\frac{100}{2}}\sqrt{\frac{2}{25}}}$$
$$= \frac{10}{\sqrt{\frac{100}{25}}} = \frac{10\times 5}{10} = 5$$

D VRAI On calcule à présent le degré de significativité p avec la table de Student sachant que le test est unilatéral. Pour lire dans la table de Student, on commence par se placer à la ligne correspondant à 48 ddl. Puis, sur cette ligne, on cherche à quel niveau se situe la valeur test que l'on vient de calculer.

On remarque que celle-ci se situe au-delà de 3,5051. En reportant cela au niveau des colonnes, on obtient 0,001. Cependant, le test étant unilatéral, on doit diviser cette valeur par 2. Ainsi :

$$p < 0{,}0005$$

Au risque $\alpha = 5\% = 0{,}05$, on remarque que $p < \alpha$. On rejette donc l'hypothèse nulle.

E FAUX Au risque $\alpha = 1\% = 0{,}01$, on remarque que $p < \alpha$. On rejette également l'hypothèse nulle.

Question 15 :

On souhaite savoir si la concentration X d'une nouvelle protéine P02TZa, exprimé en ng/L, est différente chez l'homme et chez la femme. On dispose de deux échantillons aléatoires de 30 hommes et 30 femmes. Les estimations de la variance de X obtenues à partir des échantillons sont respectivement de 62 ng²/L² chez l'homme et 58 ng²/L² chez la femme. Les estimations des concentrations moyennes obtenues à partir des échantillons sont 27,4 ng/L chez l'homme et 23,4 ng/L chez la femme. Le risque de première espèce consenti est fixé à 5%.

A. Vous pouvez utiliser la loi normale.
B. Le niveau de significativité du test est $p = 0{,}0228$.
C. Le niveau de significativité du test est $p = 0{,}0456$.
D. Vous rejetez l'hypothèse nulle d'égalité des concentrations moyennes au risque de première espèce consenti.
E. Vous concluez que la concentration de la protéine P02TZa est plus élevée chez l'homme que chez la femme.

Question 15 : ACDE

Il s'agit d'un test de comparaison de 2 moyennes observées dans le cas où les effectifs des 2 échantillons sont supérieurs ou égaux à 30.

A VRAI Dans le cas d'une comparaison de 2 moyennes observées où la taille des 2 échantillons est supérieur ou égale à 30, on utilise la loi normale.

B FAUX On commence par calculer la valeur test z :

$$z = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{27{,}4 - 23{,}4}{\sqrt{\frac{62}{30} + \frac{58}{30}}} = \frac{4}{\sqrt{\frac{120}{30}}} = 2$$

On cherche ensuite le degré de significativité p à partir de la 1ère table de la loi normale, sachant que le test est bilatéral :

$$p = 2 \times P(Z > 2) = 2 \times \left(1 - P(Z < 2)\right) = 2 \times (1 - 0{,}9772) = 0{,}0456$$

C VRAI Cf. B

D VRAI Au risque $\alpha = 5\% = 0{,}05$, on remarque que $p < \alpha$. On rejette donc l'hypothèse nulle et on conclut sur une différence significative de la concentration de la protéine P02TZa en faveur du groupe d'homme.

E VRAI Cf. D

Question 16 :

L'Archipel d'Okinawa au Japon est réputé pour être la région du monde avec l'espérance de vie la plus élevée de la planète. Cette moyenne élevée serait le fruit d'une alimentation et d'un mode de vie plus favorable.

On cherche à savoir si la moyenne d'espérance de vie des femmes d'Okinawa est significativement différente de l'espérance de vie des femmes des pays industrialisés, qui est de 84 ans.
On constitue un échantillon de 25 femmes de l'ile d'Okinawa, on a une moyenne d'espérance de vie de 86 ans et un écart type dans l'échantillon de 4.
On suppose que la variable aléatoire X « moyenne de l'espérance de vie dans l'échantillon » est gaussienne.

A. On peut rejeter H0 au risque α=5%.
B. On peut rejeter H0 au risque α=1%.
C. Le test est un test bilatéral.

On veut à présent comparer l'espérance de vie en fonction du sexe. Pour cela on prend le même échantillon de 25 femmes ainsi qu'un échantillon de 29 hommes, de l'ile d'Okinawa, d'espérance de vie de 78 ans et d'écart type 6. On prendra également $\alpha = 5\%$ et on étudie la même variable aléatoire que précédemment.

D. Dans ce cas, il faudra préalablement effectuer un test d'égalité des variances F(28 ;24).
E. Les conditions sont réunies pour effectuer un test de Student à 52 ddl.

Question 16 : ACD

Ici, il s'agit de comparer une **moyenne observée** de l'espérance de vie des femmes dans un échantillon de 100 femmes d'Okinawa avec la **moyenne théorique** de l'espérance de vie des femmes de la population de femme des pays industrialisés.
Dans ce type de comparaison (moyenne théorique/moyenne observée), il faut déterminer quelles sont les conditions de l'exercice :

- La loi suivie : la variable « espérance de vie » suit une loi normale
- L'écart-type : nous connaissons **l'estimation** de l'écart-type de l'échantillon (donc σ est inconnue).
- La taille de l'échantillon : $n = 25 < 30$.

	n ≥ 30	n < 30
X normale σ connue	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)$	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)$
X normale σ inconnue	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow Student(n-1ddl)$ $\rightarrow N(0;1)$	$\boxed{\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow Student(n-1ddl)}$
X quelconque σ connue	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)$	HORS PROGRAMME
X quelconque σ inconnue	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow N(0;1)$	HORS PROGRAMME

A VRAI On commence par calculer la valeur test t (on utilise une loi de Student puisque les conditions de l'énoncé nous y contraignent) :

$$t = \frac{m-\mu}{\frac{s}{\sqrt{n}}} = \frac{86-84}{\frac{4}{\sqrt{25}}} = \frac{2}{\frac{4}{5}} = 2{,}5$$

On cherche à présent le degré de significativité p à partir de la table de Student. Pour cela, on définit le ddl qui est à $n-1=25-1=24ddl$. On se place ainsi à la ligne 24 ddl sur la table et on cherche où se trouve la valeur test que l'on vient de calculer. Celle-ci se situe entre 2,4922 et 2,7969.

24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,7454
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,7251
24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,7454
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,7251

En faisant correspondre avec les colonnes et en sachant que le test est bilatéral, on obtient :

$$0{,}01 < p < 0{,}02$$

Au risque $\alpha = 5\% = 0{,}05$, on remarque que $p < \alpha$. On rejette donc l'hypothèse nulle.

B FAUX Au risque $\alpha = 1\% = 0{,}01$, on remarque que $p > \alpha$. On ne rejette donc pas l'hypothèse nulle.

C VRAI La phrase « On cherche à savoir si la moyenne d'espérance de vie des femmes d'Okinawa est significativement **différente** » nous montre que l'on ne connait pas le sens de différence.

À présent, il s'agit d'un test de comparaison de **2 moyennes observées** dans le cas où les effectifs des 2 échantillons sont inférieurs à 30. Il faut ainsi que 2 conditions soient vérifiées : la variable que l'on étudie doit suivre ou être approximée par une loi normale (précisé dans l'énoncé) et les variances ne doivent pas être significativement différentes (non précisé).

D VRAI On cherche à tester l'égalité des variances via un test de Fisher $F(n_1-1; n_2-1)$. Pour savoir qui est n_1 et qui est n_2, on regarde les variances : n_1 est l'échantillon avec la plus grande variance et n_2 celui avec la plus petite variance. L'échantillon d'homme est celui qui dispose de la plus grande variance avec une variance de 36 (on dispose de l'écart-type dans l'échantillon, que l'on met au carré pour obtenir la variance), donc $n_1 = 29$ et $n_2 = 25$.
On réalise donc un test de Fisher $F(28; 24)$.

Pour tester l'égalité des variances, on commence par calculer la valeur test :

$$F = \frac{s_2^2}{s_1^2} = \frac{6^2}{4^2} = \frac{36}{16} = \frac{9}{4} = 2{,}25$$

E FAUX Ensuite, on va déterminer la valeur seuil en lisant la table de Fisher à $F(28; 24)$. On se place ainsi à la colonne 28 ddl et à la ligne 24 ddl. Les colonnes n'allant que jusqu'à 14, nous pouvons simplement dire que la valeur seuil est inférieure à la valeur seuil correspondant à la la colonne 14 et la ligne 24. On a ainsi : $F_{seuil} < 2{,}13$.
On remarque que $F > F_{seuil}$, on rejette donc l'hypothèse nulle d'égalité des variances. Donc les variances sont significativement différentes.

La condition d'égalité des variances n'étant pas validée, on ne peut pas peut comparer les deux moyennes observées.

Question 17 :

On s'intéresse au temps moyen d'hospitalisation après une appendicectomie par coelioscopie dans 5 services différents pour 22 appendicectomies chacun. On recense alors le temps moyen d'hospitalisation pour chaque patient. Pour analyser ces données et les comparer, on s'intéresse à la somme des carrés des écarts totale, résiduelle et entre colonnes. On obtient un SCT de 20 000, une SCR de 8 000 et un SCC de 12 000.

A. La valeur prise par la statistique de test est de $\frac{315}{8} \approx 39{,}38$.
B. La valeur de la statistique de test calculée suit une loi de Fisher $F(4\,;105)$.
C. Le test est significatif au risque $\alpha = 5\%$.
D. On peut comparer les moyennes deux à deux en contrôlant le risque de première espèce, après le rejet de l'hypothèse nulle.
E. Pour contrôler le risque de première espèce on peut appliquer la correction de Bonferroni et on aura ici $\frac{\alpha}{5}$.

Question 17 : ABCD

Dans cet exercice, on réalise un ANOVA (ANalysis Of VAriances). Ici, on a 5 échantillons différents (5 services différents), donc on a 5 moyennes différentes. Ce sont ces 5 moyennes que l'on cherche à comparer. On ne connait pas les valeurs de ces moyennes mais on connait leurs variances associées (ou plutôt les sommes de ces variances : SCT, SCC, SCR). On note k le nombre de groupes et N le nombre de sujet au total de l'étude.

On a : $k = 5$ $\quad N = 5 \times 22 = 110$ $\quad SCT = 20000$ $\quad SCC = 12000$ $\quad SCR = 8000$

A VRAI Pour calculer la valeur test, on commence par calculer les estimations de variances suivantes :

$$s_C^2 = \frac{SCC}{k-1} = \frac{12000}{5-1} = \frac{12000}{4} = 3000$$

$$s^2 = \frac{SCR}{N-k} = \frac{8000}{105}$$

On calcule ensuite :

$$F = \frac{s_C^2}{s^2} = \frac{3000}{\frac{8000}{105}} = \frac{105 \times 3}{8} = \frac{315}{8} \approx 39{,}38$$

B VRAI Lorsque l'on réalise un ANOVA, on utilise un test de Fisher $F(k-1; N-k)ddl$. On utilise une loi de Fisher à $F(4; 105)$

C VRAI On cherche maintenant à déterminer la valeur seuil en utilisant la table de Fisher $F(4; 105)$. On se place ainsi à la colonne 4 ddl et entre les lignes 100 et 200 ddl. On a : $2{,}417 < F_{seuil} < 2{,}463$. En comparant la valeur seuil avec la valeur test, on remarque que $F > F_{seuil}$. On rejette donc l'hypothèse nulle, on conclut donc que les moyennes sont significativement différentes.

D VRAI Maintenant que l'on a démontré une différence significative entre les différentes moyennes, on peut les comparer 2 à 2 sont savoir lesquelles sont réellement différentes.

E FAUX Lorsque l'on compare les moyennes 2 à 2, il faut faire attention à bien corriger le risque α en appliquant la correction de Bonferroni qui consiste à diviser α par Q le nombre de paires possibles parmi toutes les moyennes.

Question 18 :

La consommation moyenne de bambous des pandas heureux de vivre en liberté est de 20 kg par jour. On suppose que les 16 pandas d'un zoo constituent un échantillon aléatoire des pandas en captivité. Ces 16 pandas consomment en moyenne 16 kg de bambous par jour. L'estimation de l'écart-type obtenu à partir des données de l'échantillon vaut $s = 5$. On souhaite savoir si la captivité entraine une modification de l'alimentation de nos pandas. On considère que la variable aléatoire X « nombre de kilos de bambous consommés par jour » est une variable aléatoire gaussienne.

A. On compare une moyenne observée à une moyenne théorique.
B. La valeur prise par la statistique de test est de -3,2.
C. On ne rejette pas l'hypothèse nulle au risque $\alpha = 5\%$ (nos pandas semblent heureux !).
D. On ne rejette pas l'hypothèse nulle au risque $\alpha = 1\%$.
E. On va lire la statistique de test dans la table de la loi Normale.

Question 18 : AB

Ici, il s'agit de comparer une **moyenne observée** de la quantité de bambous consommée par les pandas du zoo avec la **moyenne théorique** de la quantité de bambous consommée par les pandas libres.
Dans ce type de comparaison (moyenne théorique/moyenne observée), il faut déterminer quelles sont les conditions de l'exercice :

- La loi suivie : la variable « espérance de vie » suit une loi normale

- L'écart-type : nous connaissons **l'estimation** de l'écart-type de l'échantillon (donc σ est inconnue).
- La taille de l'échantillon : $n = 25 < 30$.

	n ≥ 30	n < 30
X normale σ connue	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$
X normale σ inconnue	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$ $\to N(0;1)$	$\boxed{\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)}$
X quelconque σ connue	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$	HORS PROGRAMME
X quelconque σ inconnue	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to N(0;1)$	HORS PROGRAMME

A VRAI Cf. Intro

B VRAI La valeur test est :

$$t = \frac{m-\mu}{\frac{s}{\sqrt{n}}} = \frac{16-20}{\frac{5}{\sqrt{16}}} = \frac{-4 \times 4}{5} = -3{,}2$$

C FAUX On calcule maintenant le degré de significativité p grâce un test de Student $Student(n-1ddl)$, soit un un test à $Student(15ddl)$. On se place donc à la ligne 15 ddl, et sur cette ligne, on cherche la valeur test de 3,2 (on prend la valeur absolue). Celle-ci se situe entre 2,9467 et 3,2860. En reportant ces valeurs au niveau des colonnes, on obtient un encadrement de p :

$$0{,}005 < p < 0{,}01$$

Au risque $\alpha = 5\% = 0{,}05$, on remarque que $p < \alpha$. On rejette donc l'hypothèse nulle. (*Les pandas semblent tristes…* ☹)

D FAUX Au risque $\alpha = 1\% = 0{,}01$, on remarque que $p < \alpha$. On rejette donc l'hypothèse nulle.

E FAUX Les conditions de l'énoncé vues dans l'intro imposent l'utilisation de la loi de Student.

Question 19 :

Un ingénieur de l'industrie agro-alimentaire cherche à savoir si le poids moyen d'un œuf est le même dans toutes les boites. Pour cela, il étudie 11 boîtes de 6 œufs. Dans ces douze boîtes, il pèse chaque œuf, et s'intéresse à la somme des carrés des écarts totale, et résiduelle. Il obtient une SCT de 3000 et une SCR de 1000.

A. La somme des carrés entre colonnes est de 2000.
B. La valeur prise par la statistique du test est égale à 10.
C. La valeur calculée de la statistique du test est à comparer à la valeur seuil d'une loi de Fischer $F(11\ ;\ 60)$.
D. Le test est significatif au risque α = 5%.
E. La valeur de la statistique de test calculée est le rapport de l'estimation de la variance entre colonnes à l'estimation de la variance résiduelle.

Question 19 : ABCDE

Dans cet exercice, on réalise un ANOVA (ANalysis Of VAriances). Ici, on a 12 échantillons différents (12 services différents), donc on a 12 moyennes différentes. Ce sont ces 5 moyennes que l'on cherche à comparer. On ne connait pas les valeurs de ces moyennes mais on connait leurs variances associées (ou plutôt les sommes de ces variances : SCT, SCC, SCR). On note k le nombre de groupes et N le nombre de sujet au total de l'étude.

On a : $k = 11$ $\quad N = 6 \times 11 = 66$ $\quad SCT = 3100$ $\quad SCR = 1100$

A VRAI On sait que $SCT = SCC + SCR$, donc $SCC = SCT - SCR = 3100 - 1100 = 2000$

B VRAI Pour calculer la valeur test, on commence par calculer les estimations de variances suivantes :

$$s_C^2 = \frac{SCC}{k-1} = \frac{2000}{11-1} = \frac{2000}{10} = 200$$

$$s^2 = \frac{SCR}{N-k} = \frac{1100}{66-11} = \frac{1100}{55} = 20$$

On calcule ensuite la valeur test :

$$F = \frac{s_C^2}{s^2} = \frac{200}{20} = 10$$

C VRAI On réalise un test de Fisher $F(k-1; N-k)ddl$. On a donc : $F(10; 55)$.

D VRAI On calcule la valeur test avec la table de Fisher à $F(10; 55)$. On a : $1{,}993 < F_{seuil} < 2{,}026$. EN comparant cette valeur seuil à la valeur test, on remarque que $F > F_{seuil}$. On rejette donc l'hypothèse nulle, et on conclut que les moyennes sont significativement différentes entre elles.

E VRAI Cf. B

Question 20 :

Un groupe de 25 pigeons perdus a été retrouvé vers Bruges-sur-Mer sur les côtes belges. Des chercheurs entreprennent alors de trouver d'où proviennent ces 25 pigeons pour les ramener dans leur pays natal. Grâce à la couleur caractéristique des pattes et des ailes, les chercheurs n'ont plus eu que deux possibilités : la France et l'Angleterre. L'un des scientifiques a alors une idée lumineuse : calculer la moyenne du Q.I. des pigeons du groupe et la comparer, par des tests de comparaison de moyenne unilatéraux, aux moyennes du Q.I. des pigeons français et anglais (PS : la technique du calcul du Q.I. des pigeons est gardée secrète par nos amis belges). Ils établissent alors le tableau suivant :

	Moyenne du Q.I des pigeons	Ecart-type du Q.I des pigeons
25 pigeons du groupe	25	6
Pigeons anglais	22	8
Pigeons français	26	5

Soit X, la variable aléatoire décrivant le Q.I. des pigeons français et Y, celle décrivant celui des pigeons anglais. Les variables X et Y sont supposées suivre une loi normale. Soient H_0 l'hypothèse nulle selon laquelle les pigeons sont français et H'_0 l'hypothèse nulle selon laquelle les pigeons sont anglais. Vous fixerez le risque de première espèce à 5%, sans correction malgré la réalisation de 2 tests d'hypothèse.

A. Sous les hypothèses nulles H_0 et H'_0, les statistiques des tests suivent une loi normale centrée réduite $N(0; 1)$.

B. Sous les hypothèses nulles H_0 et H'_0, les statistiques des tests suivent une loi de Student $Student(24\ ddl)$.

C. Au risque $\alpha = 5\%$, on peut rejeter les deux hypothèses nulles.

D. Au risque $\alpha = 5\%$, on peut rejeter H'_0 mais pas H_0.

E. Si les écarts-types des distributions des Q.I. des pigeons français et anglais avaient été inconnus, les hypothèses nulles H_0 et H'_0 n'auraient pu être testées.

<u>Question 20 :</u> AD

Ici, on a 2 tests différents dont les 2 consistent à comparer une **moyenne observée** du QI des pigeons du groupe avec la **moyenne théorique** du QI des pigeons soit français soit anglais.
Dans ce type de comparaison (moyenne théorique/moyenne obserSSvée), il faut déterminer quelles sont les conditions de l'exercice :

- La loi suivie : la variable « espérance de vie » suit une loi normale
- L'écart-type : nous connaissons **le paramètre** de l'écart-type de l'échantillon (on connait aussi l'estimation de l'écart-type mais puisque l'on a son paramètre, on utilise ce-dernier en priorité)
- La taille de l'échantillon : $n = 25 < 30$.

	n ≥ 30	n < 30
X normale **σ connue**	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$	$\boxed{\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)}$
X normale **σ inconnue**	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$ $\to N(0;1)$	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to Student(n-1ddl)$
X quelconque **σ connue**	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0;1)$	HORS PROGRAMME
X quelconque **σ inconnue**	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \to N(0;1)$	HORS PROGRAMME

A VRAI Les conditions de l'énoncé imposent l'utilisation de la loi normale pour les 2 tests.

B FAUX Cf. A

C FAUX On commence par le test avec la population française en calculant la statistique de test :

$$z_F = \frac{\overline{m}-\mu_F}{\frac{\sigma_F}{\sqrt{n}}} = \frac{25-26}{\frac{5}{5}} = -1$$

On cherche ensuite le degré de significativité avec la 1ère table de la loi normale sachant que le test est unilatéral :

$$p_F = P(Z > |-1|) = 1 - P(Z < 1) = 1 - 0{,}8413 = 0{,}1587$$

Au risque $\alpha = 5\% = 0{,}05$, on remarque $p_F > \alpha$. On ne peut donc pas rejeter l'hypothèse nulle.

Ensuite, on effectue le test avec la population anglaise en calculant la statistique de test :

$$z_A = \frac{\overline{m}-\mu_A}{\frac{\sigma_A}{\sqrt{n}}} = \frac{25-22}{\frac{8}{5}} = \frac{15}{8} = 1{,}875$$

On cherche ensuite le degré de significativité avec la 1ère table de la loi normale sachant que le test est unilatéral :

$$p_A = P(Z > 1{,}875) = 1 - P(Z < 1{,}875)$$

Puisque que 1,875 est compris entre 1,87 et 1,88, on obtient un encadrement de p_A :

$$1 - 0{,}9699 < p_A < 1 - 0{,}9693 \quad \leftrightarrow \quad 0{,}0301 < p_A < 0{,}0306$$

Au risque $\alpha = 5\% = 0{,}05$, on remarque $p_A < \alpha$. On rejette l'hypothèse nulle.

Il est possible de rejeter H'_0 mais pas H_0. Classiquement, en dehors du cas particulier de la comparaison d'une distribution observée à une distribution théorique, il n'est possible de conclure qu'en situation de rejet de l'hypothèse nulle. On peut néanmoins conclure que nos pigeons sont français parce qu'ils ne sont pas anglais et qu'il n'y avait ici que ces deux alternatives.

D VRAI Cf. C

E FAUX On aurait pu utiliser une loi de Student à 24 degrés de liberté.

Question 21 :

On étudie l'efficacité de deux traitements sur le stress des étudiants en PACES à l'approche du concours du S1. Ces deux traitements agissent sur le taux de S1-stressine.
Les traitements que nous souhaitons comparer sont le Tutorax et le Nadax. Nous voulons savoir si le Tutorax est plus efficace que le Nadax pour baisser la concentration sanguine de S1-stressine.
Nous réalisons un essai thérapeutique contrôlé randomisé incluant 36 patients traités par Tutorax et 36 patients par Nadax. Le risque de première espèce est fixé à $\alpha = 5\ \%$.
Les résultats suivants sont observés :

	Moyenne de S1-Stressine en mg/dL	Variance estimée en mg²/dL²
Tutorax	1,0	0,06
Nadax	1,6	0,10

A. Les conditions d'approximation par la Loi Normale sont réunies.
B. La statistique de test vaut 9.
C. Au risque de première espèce α = 5%, on a $z_{seuil} \approx 1{,}645$. On rejette l'hypothèse nulle d'égalité des moyennes.
D. Pour $\alpha = 1\%$, on n'aurait pas rejeté H0.
E. Le traitement Tutorax est plus efficace que le traitement Nadax pour diminuer la concentration sanguine en S1-Stressine aux deux risques consentis précédemment.

Question 21 : ABCE

Il s'agit d'un test de comparaison de 2 moyennes observées dans le cas où les effectifs des 2 échantillons sont supérieurs ou égaux à 30.

A VRAI Lorsque l'on compare deux moyennes observées et que la taille de chacun des deux échantillons est supérieure ou égale à 30, on utilise la loi normale.

B VRAI On calcule la valeur test :

$$z = \frac{m_N - m_T}{\sqrt{\frac{s_N^2}{n_N} + \frac{s_T^2}{n_T}}}. \quad z = \frac{1{,}6 - 1}{\sqrt{\frac{0{,}1}{36} + \frac{0{,}06}{36}}} = \frac{0{,}6}{\sqrt{\frac{0{,}16}{36}}} = \frac{0{,}6 \times 6}{0{,}4} = \frac{36}{4} = 9$$

C VRAI Dans cet item, on demande de déterminer la valeur seuil. On trouve la valeur seuil en utilisant la 2e table de la loi normale et non pas la 1ère. Pour cela, on a besoin du risque $\alpha = 0{,}05$. On se place à la ligne 0,05 et la colonne 0,000. L'intersection entre cette ligne et cette colonne correspond à la valeur seuil, soit $z_{seuil} = 1{,}6449 \approx 1{,}645$.

Remarque : *puisque le test est unilatéral, on prend le risque tel quel. Mais attention, si on cherche la valeur seuil à partir de α avec une table unilatérale comme celle de la loi normale et un test bilatéral, on* ***divise α par 2****. Mais si on cherche le degré de significativité p à partir de la valeur test avec une table unilatérale et un test bilatéral, dans ce cas on* ***multiplie p par 2****.*

On compare maintenant la valeur seuil avec la valeur test. On remarque que $z > z_{seuil}$, on rejette donc l'hypothèse nulle. On peut donc dire que le Tutorax est plus efficace que le Nadax pour diminuer la concentration sanguine de S1-Stressine.

D FAUX On cherche dans cet item la valeur seuil pour $\alpha = 1\% = 0{,}01$. En lisant dans la 2e table de la loi normale, on obtient la valeur seuil $z_{seuil} = 2{,}3263$. On remarque que $z > z_{seuil}$, on rejette l'hypothèse nulle. Pour résoudre cet item, on aurait aussi pu calculer le degré de significativité p et le comparer au risque α.

E VRAI Cf. C

Question 22 :

Tous les étudiants de France, à l'approche des examens de fin du premier semestre, sécrètent une hormone spécifique : la S1-stressine. On observe chez les étudiants de France une concentration moyenne de S1-stressine de 0,80 mg/dL. Cette hormone ne suit pas une distribution normale.
Doria N., une chercheuse spécialisée dans le domaine du stress souhaite savoir si la concentration moyenne de S1-stressine des étudiants de PASS à l'approche du concours diffère de la moyenne nationale. On sélectionne un échantillon aléatoire de 225 étudiants en PASS à l'approche de l'examen du S1. Le seuil de significativité (risque de première espèce consenti) est fixé dans cette étude à $\alpha = 10\%$.
Sur l'échantillon d'étude, on estime une moyenne de 1,8 mg/dL et un écart type de 10 mg/dL.

A. Vous réalisez un test de Student à $224\ ddl$.
B. La valeur test calculée vaut 1,5.
C. La différence entre les concentrations moyennes de S1-stressine des étudiants en PACES et des étudiants de France est significative au niveau de significativité $p = 0{,}0668$.
D. Avec un risque de première espèce $\alpha = 5\%$, on rejette l'hypothèse nulle d'égalité des moyennes de S1-stressine.
E. Le test réalisé dans cet exercice est un test de comparaison entre une moyenne observée à une moyenne théorique.

Question 22 : BE

Ici, il s'agit de comparer une **moyenne observée** de la concentration moyenne de S1-Stressine de l'échantillon de PASS avec la **moyenne théorique** de la concentration de S1-Stressine des étudiants de France.
Dans ce type de comparaison (moyenne théorique/moyenne observée), il faut déterminer quelles sont les conditions de l'exercice :

- La loi suivie : la variable « eDspérance de vie » ne suit pas une loi normale
- L'écart-type : nous connaissons **l'estimation** de l'écart-type de l'échantillon (donc σ est inconnue).
- La taille de l'échantillon : $n = 225 > 30$.

	n ≥ 30	n < 30
X normale σ connue	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)$	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)$
X normale σ inconnue	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow Student(n-1ddl)$ $\rightarrow N(0;1)$	$\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow Student(n-1ddl)$
X quelconque σ connue	$\frac{M-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0;1)$	HORS PROGRAMME
X quelconque σ inconnue	$\boxed{\frac{M-\mu_0}{\frac{S}{\sqrt{n}}} \rightarrow N(0;1)}$	HORS PROGRAMME

A FAUX Les conditions de l'exercice impose l'utilisation de la loi normale centrée réduite.

B VRAI On calcule la valeur test :

$$z = \frac{M-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{1{,}8-0{,}8}{\frac{10}{\sqrt{225}}} = \frac{1 \times 15}{10} = 1{,}5$$

C FAUX, On cherche le degré de significativité p en utilisant la 1ère table de la loi normale sachant que le test est bilatéral :

$$p = 2 \times P(Z > 1{,}5) = 2 \times \left(1 - P(Z < 1{,}5)\right) = 2 \times 0{,}0668 = \ 0{,}1336$$

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

Au risque $\alpha = 10\% = 0{,}1$, on remarque que $\alpha < p$. On ne peut donc pas rejeter l'hypothèse nulle et donc on ne peut pas conclure en une différence significative.

D FAUX Si on n'a pas pu rejeter l'hypothèse nulle avec un risque $\alpha = 10\%$, on ne peut pas la rejeter pour un risque plus petit de $\alpha = 5\%$.

E VRAI Cf. Intro

Question 23 – La Coovviiddeeww :

SARS-CoV-2, virus à ARN est un Béta-Coronavirus très proche du virus du SARS, plus éloigné du virus du MERS. Ce virus peut infecter l'homme et l'animal. Chez l'homme, l'infection aux coronavirus se traduit le plus souvent par des rhumes et des infections respiratoires rarement graves. Des tableaux respiratoires sévères sont cependant possibles avec les virus SARS-CoV et MERS-CoV. L'infection à SARS-CoV-2 peut donner des formes sévères et mortelles. Ce virus dans ses formes sévères et mortelles cause l'hospitalisation en réanimation des patients. On souhaite comparer la durée d'hospitalisation en réanimation des patients atteints de COVID-19 dans ses formes grave dans 5 pays différents, les États-Unis, l'Allemagne, la France, l'Angleterre et l'Italie afin de comparer les performances de ces services pour la prise en charge des patients. Pour cela, on observe la durée d'hospitalisation de 160 patients dans chaque service. On suppose que les conditions de l'ANOVA sont réunies.
On mesure un paramètre biologique dont on note les moyennes de temps dans ces 5 pays dans ce tableau :

Pays	***Durée moyenne d'hospitalisation en réanimation (en jours)***
États-Unis	$t_{USA} = 8$
Allemagne	$t_{ALL} = 9$
France	$t_{FR} = 14$
Angleterre	$t_{ANG} = 10$
Italie	$t_{IT} = 7$

On cherche à savoir si les moyennes des temps d'hospitalisation diffèrent significativement au seuil de 5%. On note aussi les différents résultats suite à cette étude :

SCC = 3204 ; SCT= 10000

Aides au calcul : $\frac{6796}{4} = 1699$; $\frac{6796}{795} \approx 8{,}5$; $\frac{801}{8{,}5} \approx 94$

A. La somme des carrés résiduels vaut 6796.
B. L'objectif de l'ANOVA, ou analyse de variance, est de tester l'hypothèse nulle d'égalité des moyennes des durées d'hospitalisation en réanimation.
C. La valeur de F du test de Fisher est égale à 9,4 à (4 ; 795) ddl.
D. Vous rejetez l'hypothèse nulle d'égalité des moyennes au risque de première espèce α de 5%.
E. En effectuant une correction de Bonferroni, les moyennes pourront être comparées deux à deux pour un risque d'erreur de première espèce global de 5%, en retenant une probabilité critique α' = 0,005 lors de chaque comparaison.

Question 23 – La Coovviiddeeww : ABDE

Cet exercice est un exercice classique qui n'est pas très compliqué, si on a compris le principe de l'ANOVA et assimilé les différentes formules.
On est dans une situation d'analyse de plusieurs moyennes par l'analyse de variances.
Il ne faut pas que vous vous perdiez dans toutes ces formules car, une fois acquises, l'exercice n'est plus qu'un simple exercice de lecture de table.
J'ai aussi ici, insisté sur la longueur de l'énoncé et sur les données importantes à reconnaître. Le jour du concours, en UE3 comme dans les autres UE, vous devez être en capacité de saisir l'information importante de l'information superflue et surtout de ne pas être déstabilisés par des informations inutiles.
Je vous résume donc les informations nécessaires à la réalisation de cet exercice :
- N_{tot}= 160 × 5 = 800 (effectif total)
- k = 5 (nombre de groupes)

- SCC= 3504 (somme des carrés entre colonnes)
- SCR = SCT – SCC = 6796 (somme des carrés résiduelle)
- α = 0,05

Ce test s'effectue par la formule suivante :

$$F = \frac{s_C^2}{s^2} \rightarrow F(k-1; N-k)ddl$$

Avec : $s_C^2 = \frac{SCC}{k-1}$ & $\boldsymbol{s^2} = \frac{\boldsymbol{SCR}}{\boldsymbol{N-k}}$

SCT = SCC + SCR

A VRAI La somme des carrés résiduels ou **SCR = SCT – SCC** = 10000 – 3203 = 6796.

B VRAI Pour comparer plusieurs moyennes, on doit effectuer une ANOVA. On va donc utiliser la loi de Fischer.

C FAUX D'après les formules précédentes, on va avoir :

$$s_C^2 = \frac{SCC}{k-1} = \frac{3204}{5-1} = \frac{3204}{4} = 801$$

$$s^2 = \frac{SCR}{N-k} = \frac{6796}{795} \approx 8{,}5$$

$$\text{Donc } F_{obs} = \frac{s_C^2}{s^2} \approx \frac{801}{8{,}5} \approx 94 \rightarrow F(4; 795)ddl$$

D VRAI D'après l'item C, $F_{obs} \approx 94 \rightarrow F(4; 795)ddl$ on se reporte à la table de Fisher à (4 ;795) ddl.

K − 1

N$_{tot}$ − K

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14
200	3,888	3,041	2,650	2,417	2,259	2,144	2,056	1,985	1,927	1,878	1,837	1,801	1,769	1,742
500	3,860	3,014	2,623	2,390	2,232	2,117	2,028	1,957	1,899	1,850	1,808	1,772	1,740	1,712
∞	3,842	2,996	2,605	2,372	2,214	2,099	2,010	1,939	1,880	1,831	1,789	1,752	1,720	1,692

La lecture de la table donne 2,373 $< F_{seuil} =<$ 2,390

$$F_{obs} > F_{seuil}$$

On rejette donc l'hypothèse nulle d'égalités des moyennes. On peut donc comparer, deux à deux les moyennes.

E VRAI La correction de Bonferroni permet de corriger le seuil de significativité (= probabilité critique) lors de comparaisons de plusieurs moyennes, comme dans notre cas.

Il est égal à, d'après la formule du cours, à $\frac{\alpha}{Q}$ avec Q le nombre de paires de comparaisons possibles parmi 5, soit dans notre cas, 10. Donc correction de Bonferroni = $\frac{\alpha}{Q} = \frac{0{,}05}{10} = 0{,}005$.

Pour calculer le nombre de combinaison on peut appliquer la formule n parmis k :

$$Q = C_5^2 = \frac{5!}{2!\,(5-2)!} = \frac{5.4.3.2}{2!\,.3!} = \frac{5.4.\cancel{3.2}}{2.\cancel{3.2}} = \frac{20}{2} = 10$$

Soit 10 combinaisons possible de comparaisons deux à deux. ☺

Question 24 – PASS en FUSIONNNN Saison 2 :

Pour rappel, une première expérience a été réalisée afin de doser le taux de Transferrine qui est une protéine responsable du transport du fer dans le plasma. En cas de déficit de fer, le taux de transferrine est augmenté. En France, le taux de transferrine dans le sang suit une loi normale de moyenne 1200 $\mu g.mL^{-1}$. On confirme un diagnostic d'anémie ferriprive sévère quand le taux de Transferrine est supérieur à 1600 $\mu g.mL^{-1}$.

Dans une population de 26 individus en suspicion d'anémie ferriprive sévère, le taux moyen de transferrine est de 1650 $\mu g.mL^{-1}$. L'estimation de l'écart-type de la population obtenue à partir des données de l'échantillon vaut S = 500. On souhaite savoir si le taux moyen du groupe est bien supérieur au point de les déclarer en anémie ferriprive sévère. Le risque de première espèce est fixé à 5%.

Aide aux calculs : $\sqrt{26} \approx \sqrt{25}$

A. On compare une Moyenne théorique à une moyenne observée par l'approximation de la loi Normale.
B. La valeur de la statistique du test est de 4,5.
C. On rejette l'hypothèse nulle au seuil de significativité de 5%.
D. On rejette l'hypothèse d'égalité des Moyennes avec p < 0,0005.
E. On va lire la statistique du test dans la table de Student à 25 *ddl.*

Question 24 – PASS en FUSIONNNN Saison 2 : BCDE

A FAUX Ici on compare la moyenne calculée sur notre échantillon de n=26 personnes suspectées d'anémie sévère. L'effectif de l'échantillon est inférieur à 30. L'écart type de la population est estimé à partir des données de l'échantillon : donc l'écart type de la population est **inconnu**. La variable aléatoire suivant une loi normale, on va utiliser une loi de Student à n-1 ddl.

B VRAI

- H0 : $\mu_0 = \mu_M$
- H1 : $\mu_0 < \mu_M$ → Test Unilatérale
- La variable suit une loi normale avec une estimation de l'écart-type
- N = 26 < 30

Puis une fois que le tout est vérifié, calculons la statistique du test :

$$t = \frac{M - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{1650 - 1200}{\frac{500}{\sqrt{26}}} = \frac{450}{\frac{500}{\sqrt{25}}} = \frac{450}{100} = 4{,}5$$

C VRAI Soit p = $P(T^{25} > 4.5) = P(\ |T^{25}| > 4{,}5)/2$.

En regardant dans la table de Student pour n-1 = 25 ddl on a :

3,7251 < 4,5.

$p < 0{,}001/2$ soit p < 0.0005

Ainsi, on rejette H_0 au risque α = 5%.

D VRAI Explications détaillées dans l'item C.

E VRAI La valeur statistique lors d'une comparaison de moyenne Observé/théorique en utilisant la Loi de Student se fait à (n – 1)*ddl* = 25 *ddl*.

Question 25 –SQUID PASS « Le jeu va bientôt commencer » :

Pour ce jeu les joueurs sont répartis en 2 équipes de 50 joueurs par tirage au sort. Le but du jeu est de faire en sorte de détacher une forme pré-tracée dans du sucre fondu le plus vite possible. Le groupe B devra détacher la forme du triangle (stratégie triangle), et le groupe A la forme du parapluie (stratégie parapluie). L'objectif de l'étude est de démontrer que détacher le triangle prend moins de temps en moyenne que de détacher le parapluie. Le risque de rejeter à tort l'hypothèse nulle d'égalité des moyennes est fixé à 5%.

Résultats :

L'équipe A met en moyenne 4,5 min pour détacher le parapluie, avec une estimation de la variance populationnelle de 2,5 min².

L'équipe B met en moyenne 4 min pour détacher le triangle, avec une estimation de la variance populationnelle de 2 min².

Aide au calcul : $\frac{2}{3} \approx 0{,}67$

A. Les conditions requises étant vérifiées, vous pouvez utiliser la loi normale.
B. Le risque de ne pas rejeter l'hypothèse nulle sachant qu'elle est fausse équivaut au risque alpha et vaut 5%.
C. La statistique du test vaut 1,67.
D. Au risque de première espèce alpha consenti (5%), il n'est pas possible de prouver que la stratégie triangle est plus rapide que la stratégie parapluie.
E. On peut conclure que la stratégie parapluie est plus lente que la stratégie triangle.

Question 25 –SQUID PASS « Le jeu va bientôt commencer » : ACE

A VRAI La loi de la variable aléatoire est inconnue, les deux effectifs supérieurs à 30, on applique donc l'approximation par la loi Normale.

B FAUX C'est la définition du risque de seconde espèce bêta.

C VRAI

$$Z = \frac{m_A - m_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \rightarrow N(0;1)$$

On calcule notre grandeur test en remplaçant dans la formule :

$$Z = \frac{m_A - m_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{4{,}5 - 4}{\sqrt{\frac{2{,}5}{50} + \frac{2}{50}}} = \frac{0{,}5}{\sqrt{\frac{4{,}5}{50}}} = \frac{0{,}5}{\sqrt{\frac{\cancel{5}*0{,}9}{\cancel{5}*10}}} = \frac{0{,}5}{\sqrt{0{,}09}} = \frac{0{,}5}{0{,}3} = \frac{5}{3} = 1 + \frac{2}{3} = 1{,}67$$

D FAUX On est dans le cas d'un Test **UNILATÉRAL**, et dans le cas où la différence observée est dans le sens de l'hypothèse alternative : « *L'objectif de l'étude est de démontrer que détacher le triangle prend moins de temps en moyenne que détacher le parapluie* ».
On va ensuite calculer petit p. Pour cela, on va utiliser la fonction de répartition de la loi normale centrée réduite :

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,52[illegible]9
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,56[illegible]5
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,60[illegible]4
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,64[illegible]3
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,68[illegible]8
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,71[illegible]7
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,74[illegible]6
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,77[illegible]4
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,80[illegible]8
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,83[illegible]0
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,85[illegible]7
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,87[illegible]0
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,89[illegible]0
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,91[illegible]7
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,92[illegible]2
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,94[illegible]8
1,6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,9525

La table nous donne 0,9575. D'où :

$$p = P(Z > 1{,}67) = 1 - P(Z < 1{,}67) = 1 - \phi(1{,}67) = 1 - 0.9525 = 0{,}0475$$

p = 0,0475 < 0,05, donc on rejette l'hypothèse nulle d'égalité des moyennes, le test est significatif, la différence observée des deux moyennes de temps n'est pas due au hasard et on conclut que la stratégie triangle est plus rapide que la stratégie parapluie.

E VRAI Cf les explications de l'item D.

Question 26 – BRONZE :

Nos P1 ne veulent plus se faire piquer, ils préfèrent donc se faire bronzer. Or, les capacités de bronzage ne sont vraiment pas toutes identiques ! Pour comprendre d'où cette différence vient, une des P1 décide de réaliser quelques dosages, principalement le dosage de la mélanine des cellules de l'épiderme. La moyenne théorique en Auvergne Rhône Alpes de mélanine est de 0,60 ng/g. Dans notre groupe de P1, la mélanine moyenne est de 0,50 ng/g avec un écart-type de 0,10. On souhaite savoir s'il est possible que nos P1 soient bien tous issus de la région Auvergne Rhône Alpes (on considèrera qu'une moyenne significativement différente de celle en AuRA veut dire que nos P1 ne sont pas tous de la région).

A. Le valeur test calculée vaut 10.
B. L'hypothèse alternative est bilatérale.
C. On peut rejeter l'hypothèse nulle au risque alpha = 5%.
D. p < 0,00004.
E. On peut donc dire que les P1 ne sont pas tous issus de la région AuRA.

Question 26 – BRONZE : ABCDE

On est dans le cas d'une comparaison de moyenne observée à une moyenne théorique. C'est un cas avec n>30 donc on utilise la loi normale pour faire le test. On utilise donc la formule :

$$T = \frac{M - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0{,}5 - 0{,}6}{\frac{0{,}10}{\sqrt{100}}} = \frac{0{,}1}{\frac{0{,}1}{10}} = \frac{0{,}1 * 10}{0{,}1} = \frac{1}{0{,}1} = 10$$

La valeur test vaut donc 10.

A VRAI

B VRAI L'énoncé parle d'une moyenne significativement différente de la théorique, sans donner un sens à la différence, ce qui correspond à un test bilatéral.

C VRAI On cherche donc à savoir à quelle valeur de p correspond une valeur test de 10. Pour cela, on doit se placer dans la table de la loi normale à la ligne qui vaut 10 :

Table I – *Fonction de répartition de la loi normale centrée réduite*

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

Exemple d'utilisation de la table

Considérons $z = 1,55$: la probabilité $P(Z \leqslant 1,55) = 0,9394$ est lue à l'intersection de la ligne « 1,5 » et de la colonne « 0,05 ».

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
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Or, comme vous pouvez le voir, la table s'arrête à 4 ! Ce qui signifie que la valeur que l'on pourrait lire est supérieure à 0,99998 et donc que p < 0,00002. Or, le test étant bilatéral, on fait x2 et on peut dire p<0,00004.

On rejette bien l'hypothèse nulle pour alpha = 5%.

D VRAI
E VRAI C'est la petite phrase entre parenthèse en fin d'énoncé : une moyenne significativement différente nous amène à penser que les p1 ne sont pas tous de la région, c'est bien ce qu'il se passe ici !