Université Claude Bernard Lyon 1

Tutorat Lyon Est

Unité d'Enseignement 4

BANQUE DE QCM

2013 - 2021

PRINCIPE D'UN TEST STATISTIQUE ET COMPARAISON DE PROPORTIONS

QUESTIONS – REPONSES

2013-2022

Question 1 :

A Lyon 80% des étudiants âgés de 18 à 24 déclarent sortir au moins un samedi soir sur deux. On souhaite savoir si en première année de médecine à Lyon Est, les étudiants sortent moins pour se concentrer sur leur travail.
Sur 400 étudiants de PACES interrogés aléatoirement, 304 déclarent sortir au moins un samedi soir sur deux.

A. Les conditions d'approximation de la loi de la statistique de test par la loi Normale ne sont pas vérifiées.
B. Au risque de première espèce $\alpha = 5\%$, vous rejetez l'hypothèse nulle d'une prévalence de 80% dans la population des étudiants en première année à Lyon-Est.
C. Vous pouvez conclure que la prévalence est moins élevée pour les P1 au seuil de significativité $p < 0{,}01$.
D. Vous pouvez conclure que la prévalence est moins élevée pour les P1 au seuil de significativité $p < 0{,}001$.
E. Le seuil de significativité du test p, donne : $0{,}04 < p < 0{,}06$.

Question 1 : B

Tout d'abord il est important de bien cerner le problème. Il s'agit ici de la comparaison d'une proportion observée à une probabilité théorique. Posons l'hypothèse nulle et l'hypothèse alternative:

- H0 est l'hypothèse nulle : la prévalence dans la population des étudiants de PACES de Lyon-Est est de $\pi = \pi_0 = 0{,}8$
- H1 : il s'agit ici d'une hypothèse alternative unilatérale (« les étudiants sortent moins »), $\pi < \pi_0$

On vérifie les conditions d'application de la loi Normale :
$n^* \pi_0 = 400^*0.8 = 320 > 5$
$n^*(1-\pi_0) = 400^*0.2 = 80 > 5$
Les conditions d'application de la loi Normale sont donc vérifiées : **réponse A fausse.**

Petite astuce : pour gagner quelques secondes, vous pouvez très bien ne faire qu'un calcul en ne calculant que $n^*\pi_0$ ou $n^*(1-\pi_0)$, à savoir celui pour lequel π_0 ou $1-\pi_0$ est le plus petit ! Car si en multipliant le plus petit des deux par n, cela est supérieur à 5, alors le plus grand le sera aussi, mais si c'est inférieur, alors les conditions ne sont pas validées !

Ensuite on fixe le α, or ici d'après les questions suivantes, $\alpha = 5\ \%$

On calcule à présent la statistique de test Z :
Avec $\pi_0 = 0.8$ (80 %), $n = 400$ et $f = 304/400 = 0.76$

$$Z = \frac{f - \pi}{\sqrt{\frac{\pi 0 \times (1 - \pi 0)}{n}}} = \frac{0{,}76 - 0{,}8}{\sqrt{\frac{0{,}8 \times 0{,}2)}{400}}} = \frac{-0{,}04}{\sqrt{\frac{0{,}4}{20}}} = -2$$

Il s'agit d'un test unilatéral donc : $Z_{seuil} = Z_{\alpha}$ (alors que pour un test bilatéral on prend $Z_{\alpha/2}$

On cherche dans la table et on trouve : $Z_\alpha = 1.645$ (par interpolation linéaire)

Attention il est important de bien arrondir la valeur au supérieur pour ne pas risquer de rejeter une valeur qui ne devrait pas l'être !

→ $|Z_{calculé}| > |Z_{seuil}|$
Donc on rejette H0 au risque 5% : **réponse B vraie**

Calculons à présent p :
$p = P(Z > |Z_{calculé}|)$
$= P(Z > 2)$
$= 1 - P(Z < 2)$

= 1 – 0.9772
= 0.0228

→ Réponses **C et D fausses** et réponse **E fausse** aussi !

Attention ne vous laissez pas influencer si vous ne trouvez pas de réponses proches de ce que vous avez trouvé comme ici ! Soyez sûr de votre résultat !:)

NB : La E aurait été juste si le test avait été un test bilatéral.

Énoncé commun aux questions 2 et 3 :

Un nouveau laboratoire décide de tester l'efficacité d'un nouveau traitement antihypertenseur. Les patients ont été assignés soit dans un bras recevant ce nouveau traitement, soit dans le bras du traitement de référence, après randomisation. On mesure la PAS (Pression Artérielle Systolique) pour évaluer l'efficacité du traitement. L'écart type de la PAS est de 25 et on souhaite mettre en évidence un écart de PAS de 20mmHg avec un risque de seconde espèce (beta) de 10% et un risque de première espèce (alpha) bilatéral de 5%.
Pour les calculs on arrondira les chiffres au dixième, et on donne $3{,}3^2 \approx 10$

Question 2 :

A. La puissance de l'étude est de 90%.
B. La randomisation est le seul moyen de rendre les groupes comparables.
C. Toutes choses étant égales par ailleurs, plus la différence que l'on souhaite mettre en évidence est petite, plus la puissance est grande.
D. Si le risque de première espèce augmente alors la puissance augmente toutes choses étant égales par ailleurs.
E. La puissance d'un test correspond à la probabilité de ne pas rejeter H_0 alors qu'elle était fausse.

Question 2 : ABCD

A VRAI En effet on a : puissance = 1- bêta = 1-0,1 = 0,9

B VRAI Cet item est récurrent au concours et c'est une notion très importante. La randomisation permet de répartir « équitablement » les facteurs de confusions et ainsi de rendre les groupes comparables.

C VRAI C'est assez logique : plus la différence que l'on souhaite mettre en valeur est grande, moins il faudra de sujets dans l'étude : il faut moins de sujets pour une différence de PAS de 10mmHg que pour une différence de pression de 0,1mmHg

D VRAI Si alpha augmente, alors on « prend plus de risques » de rejeter H0 à tort mais du coup la puissance augmente parce qu'on a « plus de chance » de rejeter H_0 alors qu'elle était vraiment fausse.

E FAUX Attention ! c'est la définition du risque beta ! Or la puissance est son complément à 1, c'est-à-dire la probabilité de rejeter H_0 alors qu'elle était effectivement fausse.

Question 3 :

Concernant le nombre de sujets à inclure dans cette étude,

A. Il faut environ 32 sujets dans cette étude.
B. Si alpha diminue alors il faudra inclure plus de sujets dans cette étude toutes choses étant égales par ailleurs.
C. Il est divisé par 4 si on souhaite mettre en évidence une différence deux fois plus faible.

D. Il est multiplié par 4 si on souhaite mettre en évidence une différence deux fois plus faible
E. Si le test est bilatéral il faut inclure plus de sujets que si le test est unilatéral.

Question 3 : BDE

A FAUX Calculons le nombre de sujets à inclure dans cette étude :

$$
\begin{aligned}
n &= \frac{2\sigma^2}{\delta^2}\,(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 \\
&= \frac{2 \times 25^2}{20^2} \times (2{,}0 + 1{,}3)^2 \\
&= \frac{625}{200} \times 3{,}3^2 \\
&= 3{,}125 \times 10 \\
&= 31{,}25
\end{aligned}
$$

Mais ATTENTION ! Il y a deux groupes donc il faut deux fois plus de sujets ! Il faut donc environ 64 sujets en tout (il faut arrondir 31,25 à 32 puis le multiplier par 2)

B VRAI Si alpha diminue alors $Z_{1\text{-alpha}/2}$ augmente, donc il faudra inclure plus de sujets dans notre étude.

C FAUX Si on souhaite mettre en évidence une différence deux fois plus faible, alors comme delta est au carré et comme il est au dénominateur, le nombre de sujets à inclure sera multiplié par 4 (2 au carré !).

D VRAI

E VRAI

Si le test est unilatéral, on ne prendre plus $Z_{1\text{-alpha}/2}$ mais $Z_{1\text{-alpha}}$ qui sera du coup plus petit. Donc le nombre de sujets à inclure sera moins élevé pour un test unilatéral que pour un test bilatéral.

Question 4 :

Un élève de deuxième année a décidé de comparer la proportion de célibataires parmi les étudiants de médecine de première année de Lyon Sud à la proportion de célibataires parmi les étudiants de médecine de première année de Lyon Est. Deux échantillons aléatoires indépendants issus de ces deux populations ont été interrogés. Sur 200 étudiants de Lyon Sud, 150 étaient célibataires, et sur 400 étudiants de Lyon Est, 250 étaient célibataires.

On cherche à savoir si les étudiants de médecine de première année de Lyon Sud sont plus célibataires que les étudiants de première année de Lyon Est. On prendra si besoin : $\sqrt{6} = 2{,}4$

A. Les conditions d'approximation de la Loi Normale sont vérifiées.
B. Au risque de première espèce alpha = 5% on rejette l'hypothèse nulle d'égalité des proportions.
C. Au risque de première espèce alpha = 1% on ne rejette pas l'hypothèse nulle d'une égalité des proportions.
D. On peut donc en conclure que les étudiants de Lyon Est sont plus en couple que les étudiants de Lyon Sud, au risque alpha = 1%.
E. Il aurait également été possible de réaliser un test du Chi2 pour résoudre cet exercice.

Question 4 : ABDE

A VRAI Cherchons tout d'abord p_0 :

$$p_0 = \frac{n_A f_A + n_B f_B}{n_A + n_B} = \frac{150+250}{200+400} = \frac{400}{600} = \frac{2}{3}$$

Vérifions tout d'abord les conditions d'approximation de la Loi Normale :
$n_A \times p_O = 200 \times 2/3 > 5$
$n_B x p_0 = 400 \times 2/3 > 5$
$n_A \times (1 - p_0) = 200 \times 1/3 > 5$
$n_B \times (1 - p_0) = 400 \times 1/3 > 5$
Les conditions d'approximation de la Loi Normale sont donc bien validées.

B VRAI Il s'agit ici d'un test unilatéral de comparaison de deux proportions observées.
Calculons notre valeur test :

$$Z = \frac{fA - fB}{\sqrt{p0(1-p0)\left(\frac{1}{nA}+\frac{1}{nB}\right)}}$$

$$= \frac{\frac{150}{200} - \frac{250}{400}}{\sqrt{\frac{2}{3} \times \frac{1}{3} \times \left(\frac{1}{200} + \frac{1}{400}\right)}}$$

$$= \frac{\frac{50}{400}}{\sqrt{\frac{2}{3} \times \frac{1}{3} \times \frac{3}{400}}}$$

$$= \frac{\frac{1}{8}}{\sqrt{\frac{1}{3 \times 200}}}$$

$$= \frac{1}{8} \times \sqrt{600}$$

$$= \frac{1}{8} \times 10 \times 2{,}4$$

$$= 1{,}25 \times 2{,}4$$

$$= 3{,}0$$

Calculons maintenant le petit p :

$$\begin{aligned} p &= P(Z > 3{,}0) \\ &= 1 - P(Z < 3{,}0) \\ &= 1 - 0{,}99865 \\ &= 0{,}00135 \end{aligned}$$

On a donc p< 5% et p< 1% : on peut donc bien rejeter l'hypothèse nulle à la fois au risque 5% et au risque 1%.

C FAUX Comme vu dans l'item C, on peut rejeter l'hypothèse nulle au risque de 1%.

D VRAI La réponse D est vraie également : attention à ne pas se tromper car ici on parle bien des personnes en couple et pas des célibataires, donc ce sont bien les étudiants de Lyon Est qui sont plus en couple.

E VRAI : Attention cette notion est tombée l'année dernière au concours. Le Pr Roy vous l'expliquera plus en détail lors de vos EDs, il est possible de réaliser un test unilatéral avec un chi2, le tout étant de savoir se servir de la table.

Question 5 :

Pour comparer l'efficacité de deux méthodes chirurgicales dans le traitement d'une maladie sévère, 400 patients atteints de cette maladie sont randomisés entre ces deux traitements : 200 seront traités par la méthode A (chirurgie classique par cœlioscopie) et 200 par la méthode B (nouvelle méthode chirurgicale testée). Parmi ceux qui ont subi la méthode A, 140 guérissent, alors qu'ils sont 160 pour la méthode B.
Soit α, le risque de première espèce.

A. α est le risque de ne pas rejeter l'hypothèse nulle alors qu'elle est fausse.
B. Le Chi 2 calculé est supérieur à 5.
C. Si on prend α= 5%, on rejette l'hypothèse nulle.
D. Si on prend α= 1%, on rejette l'hypothèse nulle.
E. On aurait pu utiliser la loi normale N(0, 1) pour résoudre ce problème.

Question 5 : BCE

Oi (Ei)	Méthode A	Méthode B	Totaux
Guéris	**140 (150)**	160 (150)	**300**
Non guéris	60 (50)	40 (50)	100
Totaux	**200**	200	**400**

Cet exercice met en jeu un calcul de Chi 2 ; vous devez comprendre la logique de ce type d'exercice car cela vous aidera à retenir la méthode et ainsi à réussir le jour du concours.

A FAUX. Cette définition désigne le risque β. Le risque de première espèce α est le risque de rejeter l'hypothèse nulle alors qu'elle est vraie : c'est le seuil de probabilité que l'on va choisir pour pouvoir rejeter ou non l'hypothèse nulle. Ici l'hypothèse nulle que l'on veut tester est la suivante : les probabilités de guérison avec les méthodes chirurgicales A et B sont égales. Ainsi, si la probabilité p calculée (probabilité d'un écart à l'hypothèse nulle au moins aussi important que celui observé dans l'étude) est inférieur à α, on rejettera l'hypothèse nulle.

B VRAI. Avant de calculer votre Chi 2, je vous conseille de faire un tableau avec tous les effectifs observés et les effectifs attendus que vous allez calculer.

Calcul des effectifs attendus Ei : je prends l'exemple de la case soulignée pour que vous compreniez. On multiplie l'effectif total de la méthode A (200) par la proportion de guéris dans la population entière (méthode A et B confondues) : Ei = 200 x $\frac{300}{400}$ = 150. Pour faire simple, quand vous cherchez les Ei d'une case, vous multipliez les totaux correspondants à cette case et vous divisez le résultat obtenu par l'effectif total. Dans le cas présent, il y a un seul degré de liberté : une fois l'un des effectifs attendus calculé, vous pouvez obtenir les trois autres par différence par rapport aux effectifs marginaux correspondants (on peut trouver l'effectif attendu des guéris par la méthode B en faisant : 300-150=150).

Une fois que vous avez calculé vos effectifs attendus, vous pouvez calculer votre Chi 2 :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(140-150)^2}{150} + \frac{(160-150)^2}{150} + \frac{(60-50)^2}{50} + \frac{(40-50)^2}{50}$$
$$= \frac{200}{150} + \frac{200}{50} = 5{,}3$$

C VRAI. Il nous reste alors à calculer le petit p : c'est la probabilité qu'un Chi 2 à un degré de liberté prenne une valeur supérieure à la valeur du Chi 2 calculée.

p= $P_{H0\ vraie}(X^2 \geq 5{,}3) = 1 - P_{H0\ vraie}(X^2 < 5{,}3)$

(on passe par le complément à 1 de $P_{H0\ vraie}(X^2 \geq 5{,}3)$ car la probabilité que l'on va lire dans la table du Chi 2 est de la forme $P(X^2 < a)$)

On lit dans la table du Chi 2, à un degré de liberté, un encadrement de $P_{H0\ vraie}(X^2 < 5{,}3)$:

$0,975 < P_{H0\,vraie}(X^2 < 5,3) < 0,990$
$0,01 < P_{H0\,vraie}(X^2 \geq 5,3) < 0,025$ (on repasse à l'inverse pour obtenir la probabilité que l'on cherche)
Donc $0,01 < p < 0,025$
La probabilité est inférieure au seuil α= 5% = 0,05.
On rejette l'hypothèse nulle au risque α de 5%.

D FAUX. La probabilité p, degré de significativité du test, est supérieure au seuil α= 1% = 0,01. On ne rejette pas l'hypothèse nule si le seuil de significativité retenu est de 1%.

E VRAI. On pourrait effectivement utiliser la loi normale pour comparer 2 proportions observées, donc pour résoudre cet exercice. Cependant, les calculs sont beaucoup plus complexes et vous avez plus de chances de faire des erreurs de calculs : à moins qu'on ne vous demande d'utiliser cette méthode, je vous conseille d'utiliser un calcul de Chi 2.

Question 6 :

En France, 90% de la population a déjà contracté le virus de la varicelle. Cependant, dans une étude réalisée dans la région Provence-Alpes-Côte-D'azur, seulement 40 personnes sur les 50 tirées au sort dans cette population l'avaient contractée. On souhaite tester l'hypothèse nulle suivante : la différence de proportion d'individus ayant contracté la varicelle observée en région PACA dans cette étude est due au hasard. On effectuera un test bilatéral.

A. Les effectifs calculés sous l'hypothèse nulle sont supérieurs ou égaux à 5.
B. On souhaite utiliser la loi normale N(0,1) pour résoudre ce problème. La valeur de la grandeur test calculée est environ égal à - 1,6.
C. Le petit p calculé est environ égal à 0,01.
D. Si on prend α = 5%, on ne peut pas rejeter l'hypothèse nulle.
E. Si on prend α = 5%, on peut conclure que la probabilité d'avoir contracté la varicelle est plus basse en PACA qu'en France, et fournir une estimation ponctuelle de celle-ci : 40/50=80%.

Question 6 : AE

Lorsque qu'un énoncé comme celui-ci vous est présenté, il faut analyser ce qu'il vous est demandé pour savoir quelle formule appliquer. Ici, on souhaite vous faire comparer une proportion OBSERVEE (la fréquence observée dans l'échantillon d'individu de la région PACA) à une proportion THEORIQUE (la proportion dans la population française).

A VRAI. Les effectifs sous l'hypothèse nulle sont np et n(1-p) : il faut qu'ils soient suppérieurs ou égaux à 5 pour que l'on puisse appliquer la loi Normale. 50x0,90 = 45 et 50x0,1 = 5, donc on peut appliquer la loi Normale.

B FAUX. Pour résoudre cet exercice avec la loi Normale, il faut appliquer la formule du cours. La grandeur test est la variable aléatoire suivante : $Z = \frac{(f-\pi)}{\sqrt{\frac{\pi(1-\pi)}{n}}}$

On doit donc calculer, à partir des données de l'énoncé, la fréquence f observée dans l'échantillon : f = 40/50 = 0,8. On en déduit la valeur prise par la grandeur test sur l'échantillon :
Z = (0,8 – 0,9) / $\sqrt{0,9\ x\ 0,1/50}$
Z = - 0,1$\sqrt{50}$ / $\sqrt{0,09}$
Z = - 0,1 x 7 / 0,3 (pensez à arrondir quand vous le pouvez)
Z = - 2,3

C FAUX. Une fois qu'on a calculé le Z, on va calculer le degré de significativité du test. / ! \ Ici, le test est bilatéral, on va donc calculer la probabilité :

$P_{Ho\ vraie}$ (|Z|>2,3) = $P_{Ho\ vraie}$ (|Z|< -2,3) et non $P_{Ho\ vraie}$ (Z < -2,3), ce qui correspondrait à un test unilatéral.

p = $P_{Ho\ vraie}$ (|Z|> 2,3) = $P_{Ho\ vraie}$ (Z < -2,3) + $P_{Ho\ vraie}$ (Z > 2,3)
p = 2 x $P_{Ho\ vraie}$ (Z > 2,3)
p = 2 x (1 – Φ (2,3))

D'après la table de la fonction de répartition de la loi normale centrée réduite:
p = 2 x (1 – 0,9893)
p = 0,0214

D FAUX. p <α, donc on rejette l'hypothèse nulle au risque α=5%.

E VRAI. L'hypothèse nulle testée était : la différence de proportion d'individu ayant eu la varicelle observée en région PACA par rapport à la proportion française est due au hasard. Ainsi, rejeter l'hypothèse nulle induit que la différence n'est pas due au hasard. On peut donc calculer une estimation ponctuelle de la probabilité d'avoir contracté la varicelle dans la région PACA à partir de notre échantillon : 40/50 = 80% (celle-ci est plus faible que dans la population française).

Question 7 :

Pour comparer l'efficacité des traitements A et B, un essai thérapeutique comparatif randomisé est réalisé, incluant 50 patients dans chaque bras. Les rémissions observées sont de 20 dans le bras traité A et de 8 dans le bras B.
Le protocole de l'essai prévoyait la réalisation d'un test du Chi-2 au seuil de significativité α=0,05.
Aide au calcul : 50/7 ≈7,14

A. On peut rejeter l'hypothèse nulle pour un risque de première espèce de 5%
B. On peut rejeter l'hypothèse nulle pour un risque de première espèce de 1%
C. La lecture dans la table du Chi-2 donne 0,001 ≤ p ≤ 0,01
D. Le test étant bilatéral, il n'est pas possible de conclure en faveur de l'un des deux traitements.
E. Toutes les réponses sont justes.

Question 7 : ABC

A VRAI et B VRAI : Ici on nous donne des pourcentages, on va donc faire un tableau en prenant 100 de total. On vérifie également les conditions d'application np≥5 n(1-p)≥5 :

	A	B	Total
Guérison	20 **(14)**	8 **(14)**	28
Echec	30 **(36)**	42 **(36)**	72
	50	50	100

Effectif attendu : la proportion que l'on aurait sous l'hypothèse nulle.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(20-14)^2}{14} + \frac{(8-14)^2}{14} + \frac{(30-36)^2}{36} + \frac{(42-36)^2}{36}$$
$$= 2 * \frac{6^2}{14} + 2 * \frac{6^2}{36} = \frac{72}{14} + 2 = \frac{72}{14} + \frac{28}{14} = \frac{100}{14} = \frac{50}{7} = 7{,}14$$

On regarde alors où se situe notre valeur dans la table du Chi 2 a un 1ddl :

6,6349 ≤ 7,14 ≤ 10,8276
0,990 ≤ X^2 ≤ 0,999
0,001 ≤ p ≤ 0,01 ≤ 0,05
Donc on rejette H0 au risque α=5% et au risque α=1%

C VRAI

P / ddl	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,999
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,1015	0,4549	1,3233	2,7055	3,8415	5,0239	6,6349	10,8276

D FAUX Même si le test est bilatéral, comme Pa>Pb on peut conclure en faveur du traitement A.
E FAUX

Question 8 :

Dans la population adulte, on admet que la prévalence du diabète est de 10%. On cherche à savoir si la prévalence du diabète est plus élevée dans une région particulière. On constitue un échantillon aléatoire de 400 individus de cette région, parmi lesquels 60 souffrent de diabète.

Approximation : 1/3=0,333

A. Les conditions d'application de la loi Normale sont réunies.
B. Si on utilise la loi Normale, la statistique de test calculée est environ égale à 3,33.
C. Si on prend un risque de première espèce de 5%, on rejette l'hypothèse nulle.
D. Si on prend un risque de première espèce de 10%, on rejette l'hypothèse nulle.
E. Si on prend un risque de première espèce de 1%, on peut considérer que la proportion de diabétiques dans la région d'intérêt est plus importante que celle généralement admise dans la population adulte.

Question 8 : ABCDE

A Vrai : np = 40 >5 et n(1-p) = 360 > 5

B Vrai : Pour calculer la statistique de test, on utilise la formule suivante, avec f la fréquence dans l'échantillon et p la proportion dans la population :

$$Z = \frac{f - \mathrm{p}}{\sqrt{\frac{\mathrm{p}(1-\mathrm{p})}{n}}} = \frac{0{,}15 - 0{,}10}{\sqrt{\frac{0{,}10 \text{ x } 0{,}90)}{400}}} = \frac{0{,}05 \text{ x } 20}{\sqrt{0{,}09}} = \frac{0{,}05 \text{ x } 20}{0{,}3} = \frac{10}{3} \approx 3{,}33$$

C Vrai : Car la différence observée est dans le sens de l'hypothèse alternative

$H_0 : \pi = 0{,}10 = \pi_0$

$H_1 : \pi > 0{,}10$

On calcule le petit p :

p = P(Z>3,33) = 1 – P(Z<3,33) = 1-0,99957 = 0,00043

p<α, on peut rejeter l'hypothèse nulle d'une proportion de diabétique dans la région d'intérêt égale à 10% pour retenir l'hypothèse alternative d'une proportion de diabétiques supérieure.

D Vrai E Vrai

Question 9 :

Pour savoir s'il existe un lien entre la spécialité choisie en terminale S et la réussite aux concours des grandes écoles, une étude est réalisée sur un échantillon aléatoire de 200 étudiants candidats.

	Réussite	Échec	Total
Spécialité Mathématique	25	75	100
Spécialité Physique	15	85	100
Total	40	160	200

On teste l'hypothèse nulle d'une probabilité de succès indépendante de la spécialité. On réalise un test du χ^2.

A. On peut rejeter l'hypothèse nulle au risque $\alpha = 5\%$
B. On peut rejeter l'hypothèse nulle au risque $\alpha = 10\%$

C. On peut rejeter l'hypothèse nulle au risque $\alpha = 15\%$

D. X^2=3.125

E. $X^2 = 3.925$

Question 9 : BCD

	Réussite (R)	Echec (E)	Total
Spécialité Mathématique	25 **20**	75 **80**	100
Spécialité Physique	15 **20**	85 **80**	100
Total	40	160	200

Pour effectuer un test du Khi-2, on commence par calculer les effectifs attendus pour chaque case du tableau sous l'hypothèse nulle d'une probabilité de réussite égale pour les spécialités mathématique et physique. Ces probabilités sont calculées à partir des marges du tableau

Au total : P(R) = $\frac{40}{200} = 0{,}2$ et P(E)= $\frac{160}{200} = 0{,}8$

En admettant l'hypothèse nulle, on aurait 100 x 0,2 = 20 réussites pour chaque spécialité et 80 échecs pour chaque spécialité. Dans le tableau, on indique les effectifs attendus sous l'hypothèse nulle en **bleu.** On remarque que ces 4 effectifs attendus sous l'hypothèse nulle sont tous supérieurs ou égaux à 5. On peut donc réaliser un test du chi-2.

Il suffit ensuite d'utiliser la formule du test du Chi-2 (indiquée dans le formulaire…) :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(25-20)^2}{20} + \frac{(15-20)^2}{20} + \frac{(75-80)^2}{80} + \frac{(85-80)^2}{80}$$

NB : l'ordre des termes n'importe pas, la valeur absolue est la même (dans le cadre d'un tableau avec 2 lignes et 2 colonnes) et le signe est toujours positif du fait du ²

$$\chi^2 = \frac{50}{20} + \frac{50}{80} = \frac{200}{80} + \frac{50}{80} = \frac{250}{80} = 3{,}125$$

Donc item D vrai et item E faux.

Pour les autres items, il faut lire dans le tableau des valeurs de la fonction de répartition du Chi-2 pour 1 degré de liberté (ddl) :

A 1 ddl : 2,7055 < 3,125 < 3,8415

D'où 0,90 < P(χ^2 < 3,125) < 0,95 et donc 0,05 < P(χ^2 > 3,125) < 0,10

On ne peut pas rejeter l'hypothèse nulle au risque $\alpha = 5\%$

Mais on peut la rejeter aux risques $\alpha = 10\%$ et $\alpha = 15\%$

Donc item A faux, item B vrai et item C vrai.

Question 10 :

Un médecin militaire compare l'efficacité des opérations chirurgicales menées sur 3 types de blessure durant la bataille de Diên Biên Phu

A : blessure abdominale

B : blessure tête et cou

C : blessure aux membres.

Il dresse le tableau suivant :

	Décès	Survie	Total
A	20	40	60
B	10	30	40
C	20	30	50

Total	50	100	150

A. On admet que les 3 groupes sont équilibrés au risque $\alpha = 5\%$

B. La statistique du test réalisé pour comparer l'efficacité des opérations chirurgicales menées sur les 3 types de blessure prend la valeur suivante : (bien sûr le calcul est juste !)

$$\chi^2 = \frac{\left(10 - \frac{40}{3}\right)^2}{40/3} + \frac{\left(20 - \frac{50}{3}\right)^2}{50/3} + \frac{\left(30 - \frac{80}{3}\right)^2}{80/3} + \frac{(40 - \frac{100}{3})^2}{100/3} = 2{,}25$$

C. On peut rejeter l'hypothèse nulle d'un taux de réussite égal pour les 3 types d'intervention au risque $\alpha = 5\%$

D. On peut rejeter l'hypothèse nulle d'un taux de réussite égal pour les 3 types d'intervention au risque $\alpha = 10\%$

E. On aurait pu utiliser la loi normale pour comparer les probabilités de survie des 3 indications chirurgicales

Question 10 : AB

A VRAI : La question posée ici est celle de la comparaison d'une distribution observée à une distribution théorique. Cette question n'est pas une étape préliminaire du test d'indépendance entre deux variables (questions B,C,D, et E) qui peut être bien entendu être réalisé lorsque les effectifs des modalités de l'une ou l'autre des variables sont déséquilibrés !

Observés	60	40	50	150
Attendus	150/3=50	150/3=50	150/3=50	150

On réalise un test du Chi-2 (tous les effectifs attendus sous l'hypothèse nulle de l'équilibre des groupes sont supérieurs à 5).

On compare une distribution observée 3 modalités à une distribution théorique. On a donc un Khi-2 à 3-1=2 degrés de liberté.

On peut calculer la valeur de la grandeur test :

$$\chi^2 = \frac{(60 - 50)^2}{50} + \frac{(40 - 50)^2}{50} + \frac{(50 - 50)^2}{50} = 4$$

On lit dans la table du Khi-2 à 2ddl :

$$2.7726 < 4 < 4.6052$$

D'où on a $0.750 < P(X^2 < 4) < 0.900$ et finalement $0.250 > P(X^2 > 4) > 0.100$

Au risque $\alpha = 5\%$, il n'existe pas de différence significative dans l'effectif des 3 groupes.

B VRAI, **C FAUX et D FAUX** : Pour B, C et D, on réalise un Khi-2 d'indépendance entre 2 variables à 2 ddl car les effectifs comparés sont sur 3 lignes et 2 colonnes (l-1)*(c-1)=(3-1)*(2-1)=2

	Décès (D)	Survie (S)	Total
A	20 **20**	40 **40**	60
B	10 **40/3**	30 **80/3**	40
C	20 **50/3**	30 **100/3**	50
Total	50	100	150

Pour effectuer un test du Khi-2, on commence par calculer les effectifs attendus pour chaque case du tableau sous l'hypothèse nulle d'une probabilité de réussite égale pour 3 groupes. Ces probabilités sont calculées à partir des marges du tableau

Au total : P(D)=1/3 et P(S)=2/3

Dans le tableau, on indique les effectifs attendus sous l'hypothèse nulle en **rouge**. On remarque que les effectifs attendus sous l'hypothèse nulle sont tous supérieurs ou égaux à 5. On peut donc réaliser un test du chi-2.

Il suffit ensuite d'utiliser la formule du test du Chi-2 (indiquée dans le formulaire…) :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{(20-20)^2}{20} + \frac{\left(10-\frac{40}{3}\right)^2}{40/3} + \frac{\left(20-\frac{50}{3}\right)^2}{50/3} + \frac{(40-40)^2}{40} + \frac{\left(30-\frac{80}{3}\right)^2}{80/3} + \frac{(40-\frac{100}{3})^2}{100/3} = 2{,}25$$

En enlevant les 2 fractions nulles, on retrouvait la formule de

$$1.38 < 2.25 < 2.77$$

D'où on a $0.500 < P(X^2 < 2.25) < 0.750$ et finalement $0.500 > P(X^2 > 2.25) > 0.250$

On ne peut donc pas rejeter l'hypothèse nulle au risque $\alpha = 5\%$.

Pour $\alpha = 10\%$, le $X_{seul}{}^2 = 4.6052$, ce qui est inférieur aux khi-2 calculé, donc on ne peut rejeter d'hypothèse nulle.

E FAUX : On peut passer de la loi normale à celle du Khi-2 et inversement uniquement pour un test du Khi-2 à 1 ddl, ce qui n'est pas le cas ici.

Pour votre culture, le taux de mortalité opératoire à la bataille de Dien Bien Phu a été de «seulement » 12.3%(entre 3000 et 3500 opérations menées), ce qui peut constituer une prouesse au vue des conditions : bombardement intensif, pertes parmi le corps médical… Si ces histoires vous intéressent, renseignez-vous sur le médecin-lieutenant Jacques Gindrey.

Question 11 :

WW est un chimiste remarquable ; il fabrique un produit dont la pureté est supérieure à 99% (critère de pureté retenu) dans 95% des cas. Son ami JP lui rapporte l'existence d'un produit concurrent ayant des caractéristiques organoleptiques comparables ! WW, très rigoureux, souhaite savoir si la probabilité que ce nouveau produit vérifie le critère de pureté retenu est inférieure à celle de son propre produit. Sur un échantillon de taille 100 du produit concurrent, il analyse la proportion d'exemplaires du produit concurrent vérifiant le critère de pureté retenu. Très exigeant, il fixe a priori $\alpha = 2\%$, et réalise un test unilatéral. Le spectromètre de masse indique que 90 exemplaires du produit concurrent vérifient le critère de pureté retenu.

Aide aux calculs : $\frac{0.05}{\sqrt{\frac{0.95 \times 0.05}{100}}} \approx 2.3$

A. Les conditions d'application du test statistique sont remplies (on prendra les conditions avec « $\geq$ »)
B. On réalise un test de comparaison de deux proportions observées
C. Au risque $\alpha = 2\%$, on ne peut pas rejeter H_0, WW a simplement raté une « fournée »
D. Au risque $\alpha = 2\%$ on peut rejeter H_0, WW va devoir s'entretenir avec son concurrent
E. Si WW avait réalisé le même test, mais bilatérale cette fois-ci, on n'aurait pas pu rejeter H_0

Question 11 : ADE

A VRAI Et **B FAUX** : Il s'agit d'un test de comparaison d'une proportion théorique (95%) à une proportion observée (90%) sur un échantillon de taille 100

Les conditions de validité d'un tel test sont $n * \pi_0 \geq 5$ et $n(1-\pi_0) \geq 5$ et sont bien remplies

On réalise donc un test de comparaison **unilatéral** d'une proportion observée à une proportion théorique :

C FAUX et **D VRAI**

$$P(F \leq 0.90) = P\left(\frac{F - 0.95}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \leq \frac{0.90 - 0.95}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}\right) = P\left(Z \leq \frac{-0.05}{\sqrt{\frac{0.95 \times 0.05}{100}}}\right) = P(Z \leq -2.3)$$

$$P(Z \leq -2.3) = P(Z \geq 2.3) = 1 - P(Z \leq 2.3)$$

On lit $P(Z \leq 2.3)$ dans la table 1 de la fonction de répartition de la loi normale centrée réduite

Et on obtient :

$$P(Z \leq -2.3) = 1 - 0.9893 = 0.0107 < 2\%$$

On peut donc rejeter H_0 au risque $\alpha = 2\%$.

E VRAI : Si on réalise un test bilatéral, on multiplie par 2 le « petit p », on aurait alors : $p' = 2 \times 0.0107 > 2\%$
Dans ce cas, on ne peut pas rejeter H_0 au risque $\alpha = 2\%$.

Question 12 :

Un laboratoire pharmaceutique organise une étude randomisée en double-insu pour vérifier si son médicament A, est plus efficace que le médicament de référence B actuellement sur le marché, pour guérir une maladie X. Pour cela, elle suit 100 patients atteints de la maladie X répartis de façon égale en deux bras d'étude : le bras A sera traité avec le médicament A, et le bras B sera traité avec le médicament B. A la fin du test, 20 patients sont guéris dans le bras A, et 10 dans le bras B. On réalise un test du Chi-2 au risque α=5%, on admet que les conditions pour faire le Chi-2 sont réunies.

Aides aux calculs : $\frac{3}{5} + \frac{5}{7} = 2{,}38$ $\qquad \frac{5}{15} + \frac{5}{35} = 0{,}47$

A. Les effectifs attendus sont 15, 35, 15 et 35.
B. Le Chi-2 obtenu est de 0,94.
C. Le Chi-2 obtenu est de 4,76.
D. Le test est significatif.

On suppose maintenant que la pharmacie a commis une erreur quant à l'attribution des traitements dans les deux bras A et B de l'étude : elle a malheureusement donné le placebo à tous les patients de l'étude. L'essai étant réalisé en double-aveugle, ni les malades, ni les médecins ne se sont aperçus de la supercherie. Le statisticien lui-même fait ses calculs sans être mis au courant du problème de délivrance du médicament A.

E. Il y a 5 chances sur 100 pour que le test statistique final conduise à déclarer que A et B ont des effets différents.

Question 12 : ACDE

A VRAI : H_0 = la proportion de guéris est égale dans les deux groupes. Les groupes A et B ont les mêmes effectifs. On répartit donc de façon égale nos 30 guéris (15 et 15) et nos 70 non-guéris (35 et 35).

B FAUX Et **C VRAI** :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{25}{15} + \frac{25}{15} + \frac{25}{35} + \frac{25}{35} = \frac{5}{3} + \frac{5}{3} + \frac{5}{7} + \frac{5}{7} = 2{,}38 * 2 = 4{,}76$$

D VRAI : On calcule le nombre de ddl : (c-1)(l-1) = 1

On cherche dans la table la valeur seuil du Chi-2 telle que $P(X_1{}^2{}_{ddl} > X_{seuil}{}^2) = \alpha = 0{,}05$, c'est-à-dire telle que $P(X_1{}^2{}_{ddl} > X_{seuil}{}^2) = 1 - 0{,}05 = 0{,}95$. La valeur seuil lue dans la table de la fonction de répartition de la loi du Chi-2 à 1 degré de liberté est 3,8415. Dans l'exercice, $X_{observ}{}^2 > X_{seuil}{}^2$ d'où $p < \alpha$. Le test est donc significatif, on rejette H_0 au risque alpha de le faire à tort (définition du risque de première espèce alpha). Ce qui signifie que la différence est significative, qu'elle n'est pas due au hasard mais au traitement.

E VRAI : Dans le cas où tous les malades ont reçu le même traitement (le type de traitement placebo ou nouveau traitement n'a pas d'importance ici), on est en fait dans le cas où l'hypothèse nulle H_0 est vraie : il n'y a pas de différence entre les deux bras d'étude. Donc si H_0 est vrai, on a un risque alpha de rejeter celle-ci, soit ici 5%.

Question 13 :

Lors de la bataille de Diên Biên Phu en 1954, durant la guerre d'Indochine, malgré des conditions précaires, 3500 opérations chirurgicales ont été menées, avec une probabilité de survie après chirurgie abdominale estimée à 60% à partir d'un échantillon aléatoire de 25 opérés dans le camp. On cherche à savoir si la probabilité de survie après chirurgie abdominale à Diên Biên Phu était significativement différente de la probabilité de survie après chirurgie abdominale en France pour les mêmes opérations. Cette probabilité de survie étant de 0,8 en France dans les années 1950. Le risque de première espèce a été fixé à α = 5%.

A. On peut utiliser un test du Chi-2 pour tester H_0.

B. Les conditions d'application sont réunies pour pouvoir comparer ces probabilités.

C. La valeur prise par la statistique du test Chi-2 vaut 6,25.

D. Le test est statistiquement significatif avec 0,01 < p < 0,025.

E. On en déduit que les chirurgiens de Diên Biên Phu n'ont pas pu maintenir les mêmes conditions sanitaires que celles pratiquées dans les hôpitaux français à la même époque.

Question 13 : ABCE

A VRAI : On est dans le cas où on compare une proportion théorique à une proportion observée. On peut soit utiliser la loi normale, soit un test du Chi-2.

B VRAI : Les conditions d'application sont n*$\pi_0 \geq 5$ et $n(1-\pi_0) \geq 5$.

Ici, $n * \pi_0 = 0.8 * 25 = 20 \; et \; n * (1-\pi_0) = 0.2 * 25 = 5$

Oui, les conditions d'applications sont respectées.

C VRAI : Les effectifs observés (15=25*0.6 et 10=25*0.4) sont comparés aux effectifs attendus sous l'hypothèse nulle (respectivement 20=25*0.8 et 5=25*0.2).

$$\chi^2 = \frac{(15-20)^2}{20} + \frac{(10-5)^2}{5} = \frac{25}{20} + \frac{25}{5} = \frac{25+100}{20} = \frac{125}{20} = 6{,}25$$

D FAUX : On cherche dans la table du Chi-2 à 1 ddl.

5,0239 < 4,45 < 6.6349

D'où 0.01 < p < 0.025

E VRAI : Si le test est significatif, on en déduit effectivement que l'échantillon n'est pas représentatif de la population, donc que les conditions n'étaient probablement pas les mêmes. Donc item E vrai.

Question 14 :

En Ecosse, la prévalence du daltonisme est de 10% chez les hommes. Un chercheur a tenté d'éclaircir les répercussions de l'isolement sur cette prévalence. Une étude de prévalence a été réalisée sur un échantillon aléatoire de 200 hommes résidants sur l'île de Mainland, au nord de l'Ecosse, afin de savoir si les hommes de Mainland sont plus fréquemment daltoniens que les écossais en général. Le protocole de l'étude fixait le risque de première espèce à 5%, et prévoyait l'utilistisation d'un test du Chi-2.

Résultat : La prévalence observée dans l'étude était de 14%.

A. La valeur calculée du Chi-2, arrondie à la deuxième décimale, est Chi-2 = 3,56.
B. Le niveau de significativité du test est 0,05 < p < 0,1.
C. Au risque de première espèce α = 5%, vous pouvez déclarer la prévalence du daltonisme chez les hommes de Mainland significativement supérieure à la prévalence du daltonisme dans la population de référence.
D. Au risque de première espèce α = 1%, vous pouvez déclarer la prévalence du daltonisme chez les hommes de Mainland significativement supérieure à la prévalence du daltonisme dans la population de référence.
E. La loi Normale aurait pu être utilisée dans cette étude.

Question 14 : ACE

Dans cet exercice, on a une population générale dans laquelle la probabilité de daltonisme chez les hommes est de $\boldsymbol{\pi 0}$ =10%. On dispose d'un échantillon de 200 hommes (n=200) du Mainland dans lequel la fréquence (ou proportion) observée vaut $\boldsymbol{\pi}$ = **14**%. Il s'agit donc d'une comparaison entre une proportion théorique et une proportion observée.

On pose $H0$ l'hypothèse nulle qui est que la probabilité de daltonisme dans le Mainland est de π = 0.10 = π, et l'hypothèse alternative $\boldsymbol{H1}$: $\boldsymbol{\pi} > \boldsymbol{\pi 0}$. L'hypothèse alternative est donc ici unilatérale, l'énoncé précisant le sens de la différence recherchée.

Les effectifs observés (Oi) de sujets daltoniens et non daltoniens dans l'échantillon sont respectivement de 28 et 172. Les valeurs attendues Ei sous l'hypothèse nulle sont respectivement n* $\boldsymbol{\pi 0}$ = 200*0.10 = 20 daltoniens et n*(1- $\boldsymbol{\pi 0}$)=200*(1-0.10)=180 non daltoniens.
On vérifie les conditions d'utilisation :
Les Ei (20 et 180) sont tous supérieurs à 5 : OK !

A VRAI Les conditions d'applications étant vérifiées on applique la formule :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(28-20)^2}{20} + \frac{(172-180)^2}{180}$$
$$= \frac{64}{20} + \frac{64}{180} \approx 3,2 + 0,36 = 3,56$$

B FAUX Pour avoir le degré de significativité on reporte la valeur du Chi-2 calculée aux valeurs fournies par une table de Chi-2 à 1ddl.
2,7055 < 3,56 < 3,8415.
On en déduit que :
0,900 < 1-p < 0,950 d'où 0,05 < p < 0,10
Or notre test est unilatéral alors que le Chi-2 nous donne une probabilité pour un test bilatéral, pour avoir le niveau de significativité pour un test unilatéral on fait :
0,05 < 2 x p < 0,10 **soit 0,025 < p < 0,050**

C VRAI (Rappel : **0,025 < p < 0,050)**
Pour α=5% → **p < α donc l'hypothèse nulle est rejetée.** On conclut que la prévalence du daltonisme chez les hommes de Mainland est significativement supérieure à celle des hommes de la population générale au risque **α = 5%.**

D FAUX (Rappel : **0,025 < p < 0,050)**

Pour **α=1% → p > α donc on ne pourrait pas rejeter H0 pour un risque de première espèce fixé à 1%**. Le test serait non significatif.

La prévalence du daltonisme chez les hommes du Mainland n'est donc pas significativement supérieure à la prévalence du daltonisme dans la population de référence avec α = 1%.

E VRAI Il est en effet possible d'utiliser la loi normale pour résoudre cet exercice car les conditions d'approximations des données du test statistique à la loi normale sont vérifiées :
n x $\boldsymbol{\pi 0}$ ≥5 ? →200*0.10=20≥5

n x (1-$\boldsymbol{\pi 0}$)≥5 ? →200*0.9=180≥5

Question 15 :

A propos des tests statistiques :

A. Petit p est la probabilité d'observer sous l'hypothèse nulle une valeur au moins aussi éloignée de la valeur de référence que celle observée dans l'échantillon.
B. α correspond au degré de significativité du test car il détermine un seuil à partir duquel le test sera significatif. Il est fixé avant l'essai.
C. α est aussi le risque de première espèce, c'est-à-dire le risque de ne pas rejeter l'hypothèse nulle alors qu'elle est fausse.
D. Lorsque l'on ne rejette pas l'hypothèse nulle, c'est toujours parce que l'hypothèse nulle est vraie.
E. Il n'est pas possible de réaliser un test du Chi-2 pour un test unilatéral car un test du Chi-2 est par définition bilatéral.

Question 15 : A

A VRAI : C'est la définition du petit p.

B FAUX : α correspond au seuil de significativité du test et petit p au degré de significativité du test. α sera fixé avant le début de l'étude et peut être défini comme étant aussi le risque consenti à rejeter l'hypothèse nulle H_0 alors qu'elle est vraie. Une fois le test réalisé, on comparera notre petit p trouvé à notre seuil α pour conclure de la significativité ou de la non significativité de la différence observée.

C FAUX : D'après l'item B, α est donc le risque de rejeter l'hypothèse nulle alors qu'elle est vraie. Il est aussi appelé risque de première espèce.

D FAUX : Il existe deux situations conduisant à ne pas rejeter l'hypothèse nulle : soit parce que l'hypothèse nulle est vraie, soit parce que notre étude n'est pas assez puissante pour la rejeter alors qu'elle est fausse (pas assez de sujets dans l'échantillon par exemple).

E FAUX : Lors d'un test du Chi-2, le petit p trouvé dans la table est celui d'un test bilatéral car le Chi-2 est par définition bilatéral **MAIS** il est possible de trouver le petit p d'un test unilatéral en divisant par 2 le petit p de la table. (voir méthode de résolution n°2 qui est un Chi-2 unilatéral pour une explication plus précise).

Question 16 :

La 141116 est une arme bactériologique mise au point par les petits Gris ou « Zétas », un sous-groupe d'extraterrestres ayant pour projet de détruire la race humaine. Cette arme fut identifiée lors d'une épidémie d'encéphalopathies aigües le 13 décembre 2278. L'évolution de la maladie est très rapide pouvant provoquer le décès en quelques jours. Un chercheur indépendant décide de lancer une expérience sur trois groupes de 75 humains chacun. Les sujets sont randomisés dans l'un des trois bras et obtiennent un traitement suivant leur affectation : chirurgie cérébrale, cure de Prokaryota Defensus ou placebo. Dans le premier bras, on observe 40 améliorations et 20 décès, dans le deuxième on observe 20 améliorations et 15 absences d'amélioration, et dans le dernier groupe on observe 15 absences d'amélioration et 30 décès. Le critère de jugement était recueilli chez tous les sujets. Les probabilités d'évolution de la maladie sont-elles comparables ? Le test effectué sera effectué en fixant le risque de première espèce α=0,05.

A. Pour résoudre cet exercice, on utilise les tables de distribution de la loi normale.
B. Le test du Chi-2 effectué est à 3 degrés de liberté.
C. Le niveau de significativité du test est 0,001<p<0,01.
D. On rejette l'hypothèse nulle d'égalité d'évolution de la maladie, et on observe une évolution plus favorable dans le bras « traitement par chirurgie cérébrale » et moins favorable dans le bras recevant la cure.
E. On ne peut pas rejeter l'hypothèse nulle car p > α.

<u>Question 16 :</u> CD

A FAUX : Nous ne savons pas tester l'hypothèse nulle d'égalité de plus de deux probabilités avec la loi normale, nous utiliserons donc un test du Chi-2.
Pour la suite de l'exercice, nous construisons le tableau suivant :

	Traitement chirurgical	Cure	Placebo	Total
Amélioration	40 30	20 30	30 30	90
Absence d'amélioration	15 15	15 15	15 15	45
Décès	20 30	40 30	30 30	90
Total	75	75	75	225

En noir, les données de l'énoncé, puis en bleu les données obtenues par différences et sommes, et en rouges les effectifs attendus sous l'hypothèse nulle. REMARQUE : tous les effectifs attendus sous l'hypothèse nulle sont supérieurs ou égaux à 5 : le test du Chi-2 peut être appliqué !
B FAUX : le test du Chi-2 est à (nb colonnes-1)(nb lignes-1) ddl, soit (3-1)(3-1)=4 ddl.
C VRAI : On utilise donc la formule du Chi-2 :
Sous H0 :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(40-30)^2}{30} + \frac{(20-30)^2}{30} + \frac{(30-30)^2}{30} + \frac{(15-15)^2}{15} + \frac{(15-15)^2}{15} + \frac{(15-15)^2}{15} + \frac{(20-30)^2}{30} + \frac{(40-30)^2}{30} + \frac{(30-30)^2}{30} = \frac{10}{3} + \frac{10}{3} + \frac{10}{3} + \frac{10}{3} = \frac{40}{3} \approx 13{,}33$$

avec E_{ij} les effectifs attendus sous l'hypothèse nulle et O_{ij} les effectifs observes
On regarde ensuite dans la table de la fonction de répartition du Chi-2 à la ligne 4 ddl :

ddl \ p	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,999
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,1015	0,4549	1,3233	2,7055	3,8415	5,0239	6,63[illegible]9	10,8276
2	0,0100	0,0201	0,0506	0,1026	0,2107	0,5754	1,3863	2,7726	4,6052	5,9915	7,3778	9,210[illegible]	13,8155
3	0,0717	0,1148	0,2158	0,3518	0,5844	1,2125	2,3660	4,1083	6,2514	7,8147	9,3484	11,3449	16,2662
4	0,2070	0,2971	0,4844	0,7107	1,0636	1,9226	3,3567	5,3853	7,7794	9,4877	11,1433	13,2767	18,4668
5	0,4117	0,5543	0,8312	1,1455	1,6103	2,6746	4,3515	6,6257	9,2364	11,0705	12,8325	15,0863	20,5150

Notre valeur test du Chi-2 se trouve entre 13,2767 et 18,4668, on obtient donc :
0,990 < P(x^2 < 13,33) < 0,999

$0,001 < P(x^2 > 13,33) < 0,010$

On a donc un niveau de significativité qui se situe entre 0,1% et 1% de calculer une valeur test du Chi-2 sous l'hypothèse nulle qui soit supérieure à 13,33.

D VRAI et **E FAUX** : D'après l'énoncé, α=0,05 et p<α donc on rejette l'hypothèse nulle.

Question 17 :

Les poissons rouges sont originaires des rivières, lacs et étangs de Chine où leur domestication est déjà mentionnée en 970 av. J.-C. Depuis, de nombreuses variétés ont vu le jour suite à de multiples croisements et mutations génétiques. Leurs congénères sont de nos jours très largement commercialisés. Une étude a été réalisée sur un échantillon aléatoire de 180 familles, respectivement 150 familles françaises et 30 familles chinoises, a cherché à mettre en évidence une différence de probabilité de posséder des poissons rouges dans les familles françaises et chinoises. Le risque de première espèce est fixé à 5%.

Les résultats observés sont recueillis dans le tableau suivant :

	Familles françaises	Familles chinoises	Totaux
Possession d'au moins un poisson rouge dans la famille	27	9	36
Absence de possession d'un poisson rouge	123	21	144
Totaux	150	30	180

A. Les conditions d'approximation par la loi Normale sont vérifiées.
B. La statistique de test calculée avec l'approximation par la loi Normale est de 2.
C. Les probabilités de posséder des poissons rouges dans les familles françaises et chinoises ne sont pas significativement différentes, le niveau de significativité du test étant 06 < p < 0,07.
D. Les probabilités de posséder des poissons rouges dans les familles françaises et chinoises ne sont pas significativement différentes, le niveau de significativité du test étant 0,13 < p < 0,14.
E. L'estimation ponctuelle de la différence des probabilités de possession de poissons rouges est de 0,12, avec plus de poissons dans les familles chinoises.

Question 17 : ADE

On note n_F le nombre de familles françaises et n_C le nombre de familles chinoises.

A VRAI : Pour comparer deux proportions observées, il faut que :

- $n_F \times \pi_0 \geq 5$
- $n_F \times (1-\pi_0) \geq 5$

- $n_C \times \pi_0 \geq 5$
- $n_C \times (1-\pi_0) \geq 5$

D'après le tableau dans l'énoncé, on a n_F=150 et n_C=30.

On peut calculer f_F la proportion de familles françaises possédant un poisson rouge et f_C la proportion de familles chinoises possédant un poisson rouge.

Ces proportions se définissent comme le nombre de familles françaises/chinoises possédant un poisson rouge sur le nombre de familles françaises/chinoises dans l'étude.

Donc $f_F = \frac{27}{150} = \frac{9}{50} = 0{,}18$ et $f_C = \frac{9}{30} = \frac{3}{10} = 0{,}3$

On peut désormais calculer π_0 qui est la proportion théorique de familles possédant un poisson rouge sous l'hypothèse nulle d'égalité des de probabilité de posséder des poissons rouges dans les familles françaises et chinoises.

π_0 peut être calculé en général grâce à la formule $\pi_0 = \frac{n_F f_F + n_C f_C}{n_F + n_C}$

Cette formule revient à faire le rapport entre le nombre de familles possédant un poisson rouge dans l'étude sur le nombre de familles étudiées, soit $\pi_0 = \frac{36}{180} = \frac{6}{30} = \frac{1}{5} = 0{,}2$

Donc : $n_F\pi_0 = 150 \times 0{,}2 = 30$; $n_F \times (1 - \pi_0) = 150 \times 0{,}8 = 120$
$n_C\pi_0 = 30 \times 0{,}2 = 6$; $n_C \times (1 - \pi_0) = 30 \times 0{,}8 = 24$

Tous sont supérieurs ou égaux 5, donc les conditions d'approximation par la loi Normale sont vérifiées.

B FAUX On a donc $z = \frac{F_F - F_C}{\sqrt{p_0(1-p_0)(\frac{1}{n_F}+\frac{1}{n_C})}}$ *suit approximativement une loi N (0 ;1) sous l'hypothèse nulle*

$$Z = \frac{0{,}18 - 0{,}3}{\sqrt{(0{,}3 \times (1-0{,}2)(\frac{1}{150}+\frac{1}{30})}} = \frac{-0{,}12}{\sqrt{0{,}16\,(\frac{1}{150}+\frac{1}{30})}} = \frac{-0{,}12}{\sqrt{0{,}16 \times \frac{6}{150}}} = \frac{-0{,}12}{\sqrt{0{,}16 \times \frac{1}{25}}}$$
$$= \frac{-0{,}12}{\sqrt{0{,}16 \times 0{,}04}} = \frac{-0{,}12}{0{,}4 \times 0{,}2} = \frac{-0{,}12}{0{,}08} = -1{,}5$$

C FAUX et **D VRAI** :

$P(Z \leq -1{,}5) = P(Z \geq 1{,}5) = 1 - P(Z \leq 1{,}5) = 1 - 0{,}9332 = 0{,}0668$

Nous sommes dans le cas d'un test bilatéral car l'étude ne précise pas le sens de la différence à observer donc le petit p que nous recherchons est $P(|Z| \geq 1{,}5) = 2 \times P(Z \geq 1{,}5) = 2 \times 0{,}0668$ $= 0{,}1336$.

Nous pouvons donc déclarer que les probabilités de posséder des poissons rouges dans les familles françaises et chinoises ne sont pas significativement différentes, le niveau de significativité du test étant p = 0,1336 (0,13 < p < 0,14).

E VRAI : $f_F = 0{,}18$ et $f_C = 0{,}3$ donc l'estimation ponctuelle de la différence de proportion de possession de poissons rouges est égale à 0,3-0,18 = 0,12, avec une plus grande proportion de poissons rouges dans les familles chinoises.

Question 18 :

L'asthme est une maladie chronique qui peut être traitée par des bronchodilatateur de type β2-mimétique. Une étude a été réalisée pour répondre à la question d'une prévalence plus élevée de l'asthme en France que dans le reste de l'Europe où 10% des enfants de moins de quinze ans sont atteints. Au cours d'une étude réalisée en France, un test de dépistage de l'asthme a été réalisé sur 300 enfants d'un échantillon aléatoire de moins de quinze ans. Parmi eux, 39 sont positifs au test. Le risque de première espèce retenu est α=0,08.

On donne : $\frac{3^{3/2}}{3}$ = 1,73 $3^{3/2}$**= 5,19**

A. La valeur de la statistique de test est 5,19.
B. Sous l'hypothèse nulle, la loi de la statistique de test peut-être approximée par une loi normale
C. Vous déclarez que l'asthme est une maladie plus fréquente en **France** qu'en **Europe** avec p< 0,05
D. Vous déclarez que l'asthme est une maladie plus fréquente en **France** qu'en **Europe** avec p< 0,04
E. Vous déclarez que la différence entre les prévalences en France et dans le reste de l'Europe n'est pas significative.

Question 18 : BC

Remarque : on fait ici l'hypothèse que le test est parfait (Sensibilité =1 et Spécificité = 1) 2

A FAUX La grandeur test est une variable aléatoire qui s'écrit :

$$Z = \frac{F - \pi_0}{\sqrt{\frac{\pi_0 \times (1 - \pi_0)}{50}}} \sim N(0,1)$$

Il faut donc d'abord vérifié que la loi de la statistique de test peut être approximée par une loi normale :

- **$n \times \pi_0 \geq 5$**
- **$n \times (1 - \pi_0) \geq 5$**

Ici π0 = 0.10 donc 300*0.10>5 et 300*0.9>5 (remarque : les conditions d'application ne concernent que les probabilités théoriques !)

On peut effectuer le calcul :

$$z = \frac{0.13 - 0.10}{\sqrt{\frac{0.10 * 0.90}{300}}} = \frac{0.03\sqrt{300}}{\sqrt{0.09}} = \frac{0.03 * 10\sqrt{3}}{0.3} = \sqrt{3} = 1.73$$

- π_1 (prévalence en France)
- π_0 (prévalence dans le reste de l'Europe)

B VRAI cf A

C VRAI

H0 : $\pi_0 = \pi_1$

Il faut calculer p pour pouvoir le comparer à α.

Puisque l'on cherche à savoir **si $\pi_1 > \pi_0$**, on effectue un **test UNILATERAL.**

On a donc p = 1- ϕ(z) = 1- ϕ(1.73) = 1- 0.9582 = 0.0418 *(lecture sur la table I de la fonction de répartition de la loi normale)*

Donc, **p= 0.0418**

p<α=0.08, on rejette donc l'hypothèse nulle selon laquelle π_1 (prévalence en France) = π_0 (prévalence en Europe).

La maladie n'est pas donc plus fréquente en France que dans le reste de l'Europe.

D FAUX

H0 : $\pi_0 = \pi_1$

H1 : $\pi_0 < \pi_1$

La maladie est plus fréquente en France qu'en Europe mais avec p > 0.04

E FAUX

H0 : $\pi_0=\pi 1$
H1 : $\pi_0 \neq \pi_1$

Question 19 :

Un laboratoire pharmaceutique de Sétif, en Algérie, dit avoir trouvé un traitement (bras A) contre la maladie d'Alzheimer. Il souhaite cependant le comparer à un traitement déjà présent sur le marché (bras B) en réalisant un essai randomisé dont les résultats sont analysés à l'aide d'un test du chi-2. Le traitement Sétifien s'attaque aux plaques amyloïdes qui se forment entre les neurones durant la maladie, et aux agrégats de protéines tau formant les dégénérescences neurofibrillaires à l'intérieur des neurones. Dans ce test, 100 patients sont répartis aléatoirement entre les 2 traitements. Dans le bras A de 50 patients 30 guérisons sont observées. Dans le bras B de 50 patients 22 guérisons sont observées.

A. Le test du chi-2 est à 1ddl.
B. La formulation de l'énoncé conduit à réaliser un test bilatéral
C. La loi du Chi-2 ne permet de réaliser que des tests bilatéraux
D. Avec un risque α=0.05, les efficacités des traitements A et B ne sont pas déclarées significativement différentes
E. La différence d'efficacité ponctuelle entre les 2 traitements est estimée à 16% en faveur du traitement B.

Question 19 : ABD

On résume les données de l'énoncé dans un tableau avec les effectifs observés :

	TOUJOURS MALADE	GUERISON	
A (nouveau traitement)	20 (1)	30 (3)	50
B (traitement de ref)	28 (2)	22 (4)	50
	48	52	100

Maintenant, il faut calculer les effectifs attendus :

- **(1)** $= \frac{t_{nouveau\ traitement} * t_{toujours\ malade}}{effectif\ total\ des\ 2\ groupes} = \frac{50*48}{100} = 24$
- **(2) =** $\frac{t_{traitement\ de\ référence} * t_{toujours\ malade}}{effectif\ total\ des\ 2\ groupes} = 24$
- **(3) =**$\frac{t_{nouveau\ traitement} * t_{guérison}}{effectif\ total\ des\ 2\ groupes}$ **=26**
- **(4) =** $\frac{t_{traitement\ de\ réference} * t_{guérison}}{effectif\ total\ des\ 2\ groupes}$ **= 26**

⇨ **Tous les effectifs attendus sont supérieurs à 5 : on peut effectuer un test du chi-2.**

On se retrouve avec le tableau suivant (entre parenthèse se sont les affectifs attendus Ei) :

	TOUJOURS MALADE	GUERISON	
A (nouveau traitement)	20 **(24)**	30 **(26)**	50
B (traitement de ref)	28 **(24)**	22 **(26)**	50
	48	52	100

On calcule maintenant la valeur du Chi-2 avec la formule suivante :

$$\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{1ddl}$$

$$Chi - 2 = \frac{(20-24)^2}{24} + \frac{(28-24)^2}{24} + \frac{(30-26)^2}{26} + \frac{(22-26)^2}{26}$$

$$Chi - 2 = \frac{(4)^2}{24} + \frac{(4)^2}{24} + \frac{(4)^2}{26} + \frac{(4)^2}{26}$$

$$Chi-2 = \frac{32}{24} + \frac{32}{26}$$
$$Chi-2 = \frac{4}{3} + 2 * \frac{8}{13}$$
$$Chi-2 \approx 1.3 + 2 * 0.6 = 2.5$$

Maintenant il faut regarder dans la table du Chi-2 à 1ddl (car la table est à 2 lignes et 2 colonnes, (l-1)*(c-1)=(2-1)*(2-1)=1). Regardons les valeurs qui entourent 2.5

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

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

Exemple d'utilisation de la table

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

ddl \ p	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,999
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,1015	0,4549	1,3233	2,7055	3,8415	5,0239	6,6349	10,8276
2	0,0100	0,0201	0,0506	0,1026	0,2107	0,5754	1,3863	2,7726	4,6052	5,9915	7,3778	9,2103	13,8155
3	0,0717	0,1148	0,2158	0,3518	0,5844	1,2125	2,3660	4,1083	6,2514	7,8147	9,3484	11,3449	16,2662
4	0,2070	0,2971	0,4844	0,7107	1,0636	1,9226	3,3567	5,3853	7,7794	9,4877	11,1433	13,2767	18,4668

2.5 ∈ [1.3233 ; 2.7055]

ATTENTION : la table du chi-2 est BILATERALE. La première ligne donne des valeurs de la fonction de répartition pour un chi-2 à 1 ddl.

1-0.90   p < 1-0.750 (si ce qu'on nous demande nous renvoi à un test BILATERAL)
1-0.90 < 2p < 1-0.750 (si ce qu'on nous demande nous renvoi à un test UNILATERAL)

A VRAI Cf intro

B VRAI Cf énoncé : comparaison avec le traitement de référence sans idée a priori du sens de la différence

C FAUX Il faut distinguer la présentation de la table et la possibilité d'effectuer des tests uni ou bilatéraux.

D VRAI On est ici dans un test bilatéral puisqu'on nous demande s'il y a une différence entre les 2 traitements : (H1 : $\pi_0 \neq \pi_1$)
Donc, 1-0.90 < p < 1-0.750
0.10 < p < 0.25

p> α => on ne rejette pas l'hypothèse nulle d'égalité des efficacités entre les 2 traitements. Les 2 traitements ont des efficacités non significativement différentes au risque d'erreur α=5%

E FAUX

Efficacité traitement A : $\frac{30}{50}$ **= 0.6**
Efficacité traitement B : $\frac{22}{50}$ **= 0.44**
La différence d'efficacité entre les 2 traitements est bien de 16% mais en faveur du traitement A.

Question 20 :

A. Dans un test statistique l'hypothèse testée est toujours l'hypothèse nulle H0
B. Le risque de première espèce α est la probabilité de rejeter H0 sachant que H0 est vraie
C. Le risque de seconde espèce β est la probabilité de rejeter H0 alors qu'elle est fausse
D. Soit p le niveau de significativité du test, si p < α on rejette H1
E. Soit p le niveau de significativité du test, si p > α on rejette H0

Question 20 : AB

Cf Cours.

Question 21 :

La fente labiopalatine est une des malformations les plus fréquentes de la région cranio-faciale. En France, selon des recherches, la prévalence de la fente labiopalatine est de 20%. Sur un échantillon de 81 personnes, 30% sont atteintes de cette pathologie. On effectue un test bilatéral.

A. La proportion de sujets atteints de fente labiopalatine dans l'échantillon est significativement différente de celle de la population au risque α= 5%.
B. La proportion de sujets atteints de fente labiopalatine dans l'échantillon est significativement différente de celle de la population au risque α= 10%.
C. La proportion de sujets atteints de fente labiopalatine dans l'échantillon n'est pas significativement différente de celle de la population au risque α= 3%.
D. La proportion de sujets atteints de fente labiopalatine dans l'échantillon est significativement différente de celle de la population au risque α= 2%.
E. Les conditions ne sont pas vérifiées pour réaliser ce test.

Question 21 : AB

A VRAI On réalise ici un test de comparaison d'une proportion observée à une proportion théorique il faut utiliser la formule suivante :

$$Z = \frac{F - \pi_0}{\sqrt{\frac{\pi_0 \times (1 - \pi_0)}{n}}} \sim N(0,1)$$

Ainsi, $z = \frac{0.3-0.2}{\sqrt{\frac{0,2*0,8}{81}}} = \frac{0,1}{\sqrt{\frac{0,16}{81}}} = \frac{0,1}{\frac{0,4}{9}} = \frac{1}{4}$ *9 = 2,25

Puisque l'on effectue un test bilatéral, il faut calculer le petit p puis le comparer au risque α donné dans les items

p = 2*[1- ϕ(z)] = 2* [1- ϕ(2,25)] = 2*(1- 0,9878) = 0.0244

0,0244 < 0,05 => on RH0

Donc La proportion de sujets atteints de fente labiopalatine dans l'échantillon est significativement différente de celle de la population au risque α= 5%.

B VRAI 0,0244 < 0,1 => on RH0

Donc la proportion de sujets atteints de fente labiopalatine dans l'échantillon est significativement différente de celle de la population au risque α= 10%.

C FAUX 0,0244 < 0,03 => on RH0

Donc La proportion de sujets atteints de fente labiopalatine dans l'échantillon est significativement différente de celle de la population au risque α= 3%.

D FAUX 0,0244 > 0,02 => NRH0

Donc la proportion de sujets atteints de fente labiopalatine dans l'échantillon n'est pas significativement différente de celle de la population au risque α= 2%.

E FAUX $n \times \pi_0 = 81 \times 0,2 = 16,2 \geq 5$ et $n \times (1 - \pi_0) = 81 \times 0,8 \geq 5$, les conditions sont vérifiées pour effectuer ce test.

Question 22 :

D'après les résultats des colles, les tuteurs de Lyon Est, affirment que 60 % des PACES réussissent globalement bien les épreuves d'UE4 c'est-à-dire qu'ils ont au moins la moyenne. On souhaite savoir si la proportion d'étudiants ayant un bon ressenti correspond à ce que déclarent les tuteurs. Vous décidez d'effectuer un test bilatéral. Sur un échantillon aléatoire de 225 étudiants, 113 déclarent que leur entretien s'est bien passé.

On arrondira $\sqrt{0{,}24} \approx \sqrt{0{,}25}$

A. Les conditions d'approximation de la loi statistique par la loi normale ne sont pas vérifiées.
B. Au risque de première espèce α=5% vous rejetez l'hypothèse nulle d'une prévalence de 60 % d'étudiants ayant un bon ressenti dans la population d'étudiant à Lyon est.
C. Vous pouvez conclure que la prévalence du ressenti des étudiants est inférieure à celle affirmée par les tuteurs au seuil de significativité p< 0,01.
D. Vous pouvez conclure que la prévalence du ressenti des étudiants est inférieure à celle des tuteurs au seuil de significativité p< 0,001.
E. Le niveau de significativité du test donne 0,001<p<0,002.

Question 22 : BC

Tout d'abord il est important de bien cerner le problème. Il s'agit ici de la comparaison d'une proportion observée à une probabilité théorique.

Posons l'hypothèse nulle et l'hypothèse alternative:

H0 est l'hypothèse nulle : la prévalence dans la population des étudiants est de π= π0 =0,6
(Il n'y a pas de différence significative entre ce que déclarent les tuteurs et le ressenti des étudiants)
H1 : il s'agit ici d'une hypothèse alternative **bilatérale** (la proportion d'étudiants ayant un bon ressenti est différente) π ≠ π0.

On vérifie les conditions d'application de la loi Normale :
N*π0 =225*0.6=135>5
N*(1- π0) = 225*0.4 = 90 > 5

Les conditions d'application de la loi Normale sont donc vérifiées : **A FAUX**.
Ensuite on fixe le α, or ici d'après les questions suivantes, α = 5 %
N= 225 ; π0=0,6 et la fréquence f= 113/225 0,5

On calcule à présent la statistique de test z :

$$z = \frac{f-\pi}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}5-0{,}6}{\sqrt{\frac{0{,}6\ x\ 0{,}4}{225}}} = -\frac{0{,}1}{\sqrt{0{,}24}} \text{ x } 15 \approx -\frac{0{,}1}{0{,}5} \text{ x } 15 = -\frac{15}{3} = -3$$

($\sqrt{0{,}24} \approx \sqrt{0{,}25}$ Cf énoncé)
Il s'agit d'un test bilatéral donc : Zseuil = Z α/2
On cherche dans la table 2 (loi Normale centrée réduite) et on trouve : Zα = 1,96 (par interposition linéaire)
Attention il est important de bien arrondir la valeur au supérieur pour ne pas risquer de rejeter une valeur qui ne devrait pas l'être !

→ |-3|>|1,96| donc $|Zcalculé| > |Zseuil|$
Donc on **rejette H0 au risque 5%**, la différence entre le ressenti des étudiants et ce que déclarent les tuteurs est significative. (Le ressenti des étudiants étant sûrement biaisé par le stress) : **B VRAI**

Calculons à présent p :

p = P (Z >$|Zcalculé|$) = 2x P(Z>3)
= 2 x (1 - P (Z < 3)) (On utilise ici la Table de la fonction de répartition)
= 2x(1 – 0,99865)
= 2 x 0,00135
=0,00270

(On multiplie par 2 car il s'agit d'un test bilatéral, la fréquence calculée peut être plus grande ou plus petite que la proportion théorique et comme nous voulons juste montrer qu'elle est significativement différente on prend en compte les deux cas.)
De plus, la différence significative étant prouvée, on peut conclure sur le sens de la différence en disant que la prévalence du ressenti des étudiants est inférieure à celle affirmée par les tuteurs (0,5<0,6) même s'il s'agit d'un test bilatéral.
Donc **C VRAI** car 0,0027 < 0,01 mais **D FAUX** car 0,0027 > 0,001 et non plus petit.

E FAUX mais aurait été vrai si le test avait été unilatéral (0,001<0,00135<0,002).

Nous pouvions aussi rejeter H0 en utilisant le risque α : 0,0027 ≤ 0,05 donc **p ≤ α : rejet H0**

NB : lorsqu'on prend le risque α, on le compare à p, on peut alors se dire que la différence est significative quand le risque n'est pas dépassé : **p ≤ α : rejet H0**
A contrario qd on utilise le seuil on peut se dire que la différence est significative à partir du moment où le seuil est dépassé : $|Zcalculé| > |Zseuil|$ **: rejet de H0**

Question 23 :

Lors d'un essai comparatif randomisé, on souhaite tester l'efficacité de deux nouveaux traitements A et B contre une maladie inflammatoire chronique de l'intestin, et plus précisément la maladie de Crohn. Les 500 patients de l'essai sont randomisés en deux groupes de même effectif. Le protocole prévoit de tester l'hypothèse nulle H0 d'absence de différence de probabilité de guérison entre les 2 groupes en effectuant un test du Chi-2. Le risque de première espèce est fixé à 5%. A l'issue de l'essai, 420 guérisons sont observées : 200 avec le traitement A contre 220 avec le traitement B.

Aide au calcul : $\frac{20}{21} \cong 0{,}95$

A. Au risque de première espèce =5%, le calcul du Chi-2 entraîne le rejet de H0.
B. Vous déclarez une différence d'efficacité significative entre les traitements A et B, avec un degré de significativité 0,025<p<0,05.
C. Vous déclarez une différence d'efficacité significative entre les traitements A et B, avec un degré de significativité 0,01<p<0,025.
D. Le traitement B est significativement plus efficace que le traitement A.
E. Ce test aurait pu être effectué en utilisant la loi normale.

Question 23 : ACDE

Tout d'abord on remarque qu'il s'agit ici de faire un test du Chi-2 avec comparaison de 2 répartitions observées.
De plus on fait un test bilatéral car on cherche à savoir si l'efficacité des 2 traitements est différente ou non.
Le plus pratique avec les tests du chi-2, c'est de faire un tableau avec les effectifs observés pour ensuite plus facilement trouver les effectifs attendus.

	Traitement A	Traitement B	Totaux
Guéris	200 – 210	220 – 210	420
Non Guéris	50 – 40	30 - 40	80
Totaux	250	250	500

$E_{attendu} = \frac{Total\ lignes\ x\ total\ colonnes}{Effectif\ Total}$ ainsi pour l'effectif attendu de guéris avec le traitement A on a $E_{att\ A-Guéris} = \frac{420\ x\ 250}{500}$ = 210
On fait pareil pour les autres effectifs attendus (en bleu dans le tableau).

A présent on vérifie que les conditions d'application sont vérifiées : $E_{attendu} \geq 5$, on constate alors que nos 4 effectifs attendus sont bien 5 donc les conditions sont vérifiées.

On calcule alors la statistique de test : (O_{ij}: effectifs observés)

$$\kappa = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(200-210)^2}{210} + \frac{(220-210)^2}{210} + \frac{(30-40)^2}{40} + \frac{(50-40)^2}{40}$$

$$= \frac{100}{210} + \frac{100}{210} + \frac{100}{40} + \frac{100}{40}$$

$$= \frac{200}{210} + \frac{200}{40} = \frac{20}{21} + \frac{20}{4} \cong 0{,}95 + 5 = 5{,}95$$

On regarde alors où se situe notre valeur dans la table du Chi-2 à (c-1)(l-1) ddl soit à (2-1) x (2-1) = 1ddl

5,0239 < 5,95 < 6,6349

0,975 < χ^2< 0,990

0,01 < p < 0,025 ⟹car Degré de significativité :p = P($\chi^2 > \kappa$) et que la table nous donne p=P($\chi^2 < \kappa$)

Ainsi **B FAUX** et **C VRAI**

De plus 0,01<p<0,025< 0,05 **p<α donc rejet de l'hypothèse nulle H0 d'absence de différence entre l'efficacité des 2 traitements** donc **A VRAI**.

D VRAI : H_0 étant rejeté on peut affirmer une différence significative entre les 2 traitements, différence en faveur du traitement B qui présente plus de personnes guéries que le traitement A d'après les effectifs observés et répertoriés dans le tableau.

E VRAI : Cf cours, on peut approximer un test du Chi-2 par la loi normale. Si les conditions de validité du Chi-2 sont validés, celles de la loi normale aussi et inversement. De plus la valeur de statistique de test du Chi-2 est le carré de la valeur pour la loi normale : $\kappa = z^2$

Question 24 :

On souhaite comparer l'alcoolisation aigüe chez les 15-30 ans, entre deux régions de France et plus particulièrement entre la Bretagne et l'Alsace. On souhaite alors savoir si la prévalence de l'ivresse aigüe est significativement supérieure en Bretagne. Pour cela, 300 individus sélectionnés aléatoirement dans ces 2 régions ont été interrogés sur l'existence d'au moins une ivresse aigüe dans l'année. A l'issue de l'étude, 30 réponses positives sont observées chez 100 alsaciens, et 110 réponses positives chez 200 Bretons. Le risque de première espèce est fixé à 5%.

Aide au calcul : $0{,}47 \times 0{,}53 \cong 0{,}25$ et $\frac{0{,}25}{\sqrt{0{,}375}} \cong 0{,}409$.

A. Les conditions d'application de la loi Normale sont réunies.
B. Afin de réaliser ce test, il faudra au préalable poser $p_0 = (0{,}3 + 0{,}55)/2 = 0{,}425$
C. Vous rejetez l'hypothèse nulle au niveau de significativité p<0,0001
D. Vous ne rejetez pas l'hypothèse nulle au risque α = 5%.
E. Ce test aurait pu être effectué en utilisant le test du Chi-2.

Question 24 : ACE

On pose tout d'abord $p_0 = \frac{n_A f_A + n_B f_B}{n_A + n_B} = \frac{100 \times 0{,}3 + 200 \times 0{,}55}{100 + 200} = \frac{140}{300} \cong 0{,}4$

On peut d'ores et déjà dire que l'item **B est FAUX** (il ne suffit pas de faire la moyenne des 2 proportions observées pour trouver p_0, il faut aussi prendre en compte les effectifs respectifs n_A *et* n_B).

On vérifie ensuite les conditions d'application de la loi normale :

$n_A p_0 = 100 \times 0{,}47 = 47 \geq 5$ $\quad n_A(1 - p_0) = 100 \times 0{,}53 = 53 \geq 5$

$n_B p_0 = 200 \times 0{,}47 = 94 \geq 5$ $\quad n_B(1 - p_0) = 200 \times 0{,}53 = 106 \geq 5$

Les conditions sont donc bien vérifiées, item **A VRAI**

On calcule à présent la statistique de test z :

$$z = \frac{f_A - f_B}{\sqrt{p_0(1-p_0)(\frac{1}{n_A}+\frac{1}{n_B})}} = \frac{0,3 - 0,55}{\sqrt{(0,47 \times 0,53)(\frac{1}{100}+\frac{1}{200})}} = \frac{-0,25}{\sqrt{0,25 \times 0,015}} = \frac{-0,25}{\sqrt{0,00375}}$$
$$= \frac{-0,25 \times \sqrt{100}}{\sqrt{0,375}} = -0,409 \times 10 = -4,09$$

Il nous est demandé de réaliser un test unilatéral on a donc :
p = P(Z < -4,09) $= 1 - \phi(4,09) = 1 - 0,99998 = 0,00002$ (lecture de p dans la table de la fonction de répartition)
Item **C VRAI** et **D FAUX** car 0,00002 < 0,05 donc **p < α** donc on peut rejeter l'hypothèse nulle d'égalité des proportions observées, et donc on peut conclure à une différence significative au risque $\alpha = 5\%$.

Au risque de première espèce $\alpha = 5\%$, la proportion d'individus déclarant au moins une ivresse aigüe annuelle est significativement supérieure chez les jeunes Bretons (55%) que chez les jeunes Alsaciens (30%).

E VRAI En effet il aurait été possible d'utiliser un test du chi-2 dont les conditions d'application auraient été validées (car ce sont les mêmes conditions que celles requises pour utiliser la loi Normale).

Question 25 :

L'un des principaux effets indésirables des corticoïdes est la prise de poids. On interroge un échantillon aléatoire de 100 patients sur leur prise de poids après 15 jours de corticoïdes. A l'issue de l'étude, 14 répondent avoir pris au maximum 3kg [0 ; 3], 61 répondent avoir pris entre 3 et 7kg]3 ; 7] et enfin 25 répondent avoir pris entre 7 et 10kg]7 ; 10].
On souhaite savoir si ces résultats sont en adéquation avec ceux de la littérature médicale. Dans cette dernière, 10% des patients prennent jusqu'à 3kg [0 ; 3], 70% entre 3 et 7kg]3 ; 7] et enfin 20% entre 7 et 10kg]7 ; 10]. Vous effectuez un test du Chi-2 avec un risque de première espèce fixé à 5%.
Aide au calcul : $\frac{81}{70} \approx 1,15$

A. Les conditions d'application du test sont vérifiées car les trois effectifs observés sont bien supérieurs ou égaux à 5.
B. Vous effectuez le test du Chi-2 à 3 degrés de liberté.
C. Le test effectué est un test de comparaison d'une distribution observée à une distribution théorique à 2 modalités.
D. La valeur calculée du Chi-2 vaut 4.
E. Vous rejetez l'hypothèse nulle au degré de significativité p<0,05.

Question 25 : D

On pose tout d'abord l'hypothèse nulle : 10% des patients prennent jusqu'à 3kg, 70% entre 3 et 7kg et enfin 20% entre 7 et 10kg.
A FAUX Ce sont les effectifs attendus qui doivent être supérieurs ou égaux à 5 et non les effectifs observés. Ces conditions sont vérifiées ici car 10 ≥ 5 ; 70 ≥ 5 ; 20 ≥ 5.

B et C FAUX Ici il s'agit bien d'une comparaison d'une distribution observée à une distribution théorique, mais à 3 modalités.
De plus on a ddl = k – 1, k étant le nombre de modalités, on a ddl = 3 – 1 = 2.
Le test sera alors effectué à 2 degrés de liberté.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(14-10)^2}{10} + \frac{(61-70)^2}{70} + \frac{(25-20)^2}{20}$$
$$= \frac{16}{10} + \frac{81}{70} + \frac{25}{20} = 1,6 + 1,15 + 1,25 = 4,0.$$ Donc item **D VRAI**

Pour avoir le degré de significativité on reporte la valeur du Chi-2 calculée aux valeurs fournies par une table de Chi-2 à 2ddl.
2,7726 < 4,0 < 4,6052.

On en déduit que :
0,750 < 1-p < 0,900 (la table dont nous disposons est celle de la fonction de répartition du Chi-2)
D'où 0,1 < p < 0,25

Comme p > 0,1 > 0,05, donc **p > α,** on ne peut pas rejeter l'hypothèse nulle, et donc on ne peut pas affirmer qu'il y a une différence significative entre ces deux distributions --> **E FAUX**

Question 26 :

Les cadeaux de Noël tant attendus chaque année peuvent être ouverts le 24 au soir ou bien le 25 au matin. On souhaite alors comparer la proportion de familles qui ouvrent les cadeaux le 24 (ou bien le 25) dans trois pays : les États-Unis, l'Angleterre et la France. On constitue alors un échantillon de 200 personnes de chaque pays, et on les interroge. 50 anglais répondent les ouvrir le 24 contre 150 le 25, 60 français répondent les ouvrir le 24 contre 140 le 25, quant aux américains, 40 répondent les ouvrir le 24 contre 160 le 25. Le plan d'analyse prévoit la réalisation d'un test du Chi-2, au seuil α = 5%.

A. Sous l'hypothèse nulle, les effectifs observés autorisent l'utilisation du test du Chi-2.
B. On rejette l'hypothèse nulle H0 d'absence de différence pour un risque de première espèce α = 5%.
C. Niveau de significativité du test de comparaison entre les trois distributions : 0,05 < p < 0,1.
D. On réalise un test du chi-2 à 6 modalités donc à 5 ddl.
E. Vous déclarez qu'il n'a pas été possible de mettre en évidence une différence significative du jour d'ouverture des cadeaux entre les différents pays.

Question 26 : CE

Dans cet exercice il faut réaliser un test du chi-2 bilatéral fixé à 5% afin de comparer la proportion de personnes ouvrant les cadeaux de Noël le 24 (ou bien le 25) entre 3 pays et de voir s'il y a donc une différence significative du jour d'ouverture.
200 personnes sont incluses dans chaque bras, on a donc un échantillon de 600 au total.
On pose l'hypothèse nulle H0 : il n'y a pas de différence significative entre les 3 pays.
Il faut à présent réaliser un tableau afin de voir les résultats plus clairement :

	Ouverture le 24	Ouverture le 25	Totaux
Américains	40 (50)	160 (150)	200
Français	60 (50)	140 (150)	200
Anglais	50 (50)	150 (150)	200
Totaux	150	450	600

En bleus sont indiqués les effectifs attendus en faisant ($\sum$ lignes x $\sum$ colonnes) /Total
A FAUX Ce sont les effectifs attendus qui doivent être supérieurs à 5 et non pas les effectifs observés. Les effectifs attendus étant bien supérieurs à 5, les conditions d'application du test du chi-2 sont validées.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(40-50)^2}{50} + \frac{(60-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(160-150)^2}{150} + \frac{(140-150)^2}{150} + \frac{(150-150)^2}{150}$$

$$= \frac{100}{50} + \frac{100}{50} + 0 + \frac{100}{150} + \frac{100}{150} + 0$$

$$= 2 + 2 + \frac{2}{3} + \frac{2}{3} = 4 + \frac{4}{3} = 5{,}33$$

D FAUX On a ici 3 modalités : les anglais, les américains et les français et non pas 6.
On a donc ddl = k – 1 = 3 – 1 = 2

Ainsi pour avoir le degré de significativité on reporte la valeur du Chi-2 calculée aux valeurs fournies par une table de Chi-2 à 2ddl.
$4{,}6052 < 5{,}33 < 5{,}9915$

On en déduit que :
$0{,}900 < 1\text{-}p < 0{,}950$ (la table dont nous disposons est celle de la fonction de répartition du Chi-2 pour un test bilatéral)
D'où $0{,}05 < p < 0{,}1$ **C VRAI**
Donc $\mathbf{p > \alpha}$ (α = 0,05) ainsi la différence entre les trois traitements n'est pas significative donc l'hypothèse nulle ne peut pas être rejetée. **B FAUX**

Ainsi **E VRAI** : Au seuil fixé α = 5%, il n'a pas été possible de mettre en évidence une différence significative du jour d'ouverture des cadeaux de Noël entre les 3 pays car on n'a pas pu rejeter l'hypothèse nulle d'absence de différence.

Question 27 :

Parmi les résidents du Maroc, il y a 10% de diabétiques. On cherche à savoir si la prévalence du diabète est plus élevée dans la ville de Meknes, ville réputée pour ses gâteaux succulents et riches en sucre. On y constitue un échantillon aléatoire de 625 individus, parmi lesquels 125 sont diabétiques. On utilisera la loi normale.
Cochez la (ou les) proposition(s) juste(s) :

A. Les conditions d'application de la loi normale sont réunies
B. La valeur test vaut 9,33
C. Au risque α = 5% on rejette H_0
D. Au risque α = 10% les habitants de M ne sont pas plus diabétiques que le reste des marocains.
E. Au risque α = 1% on rejette H_1

Question 27 : AC

A VRAI. On est dans une comparaison de proportion théorique (π_0) à observée (p). Avec les données qu'on a, on ne peut qu'utiliser la loi normale. Les conditions d'applications sont : $\pi_0 \times n > 5$ et $(1-\pi_0) \times n > 5$.
Ici, on aura donc $\pi_0 = 0{,}1$ et n = 625. D'où : $\pi_0 \times n = 0{,}1 \times 625 = 62{,}5 > 5$ OK.
Et $(1-\pi_0) \times n = (1-0{,}1) \times 625 = 0{,}9 \times 625 = 562{,}5 > 5$, OK.
B FAUX. La formule pour ce type de comparaison est : $\frac{p-\pi}{\sqrt{\left(\frac{\pi(1-\pi)}{n}\right)}}$. En remplaçant par les valeurs numériques on obtient : $\frac{0{,}2-0{,}1}{\sqrt{\frac{0{,}1\times 0{,}9}{625}}} = \frac{0{,}1}{\sqrt{\frac{0{,}09}{625}}} = \frac{0{,}1}{\frac{0{,}3}{25}} = 0{,}1 \times \frac{25}{0{,}3} = 8{,}33$
C VRAI. À α= 5%, pour un test unilatéral, z_{seuil} vaut 1,64 (dans la table de la loi normale centrée réduite qui est unilatérale). Donc on voit bient que $z_{calc} > z_{seuil}$, donc on rejette effectivement H_0. Ainsi, la différence à 5% est significative.
D FAUX. Si on a pu dire que la différence était significative à 5% d'erreur, elle le sera forcément pour un risque d'erreur plus grand donc moins précis. Donc on rejette encore H_0 à α = 10%. Ainsi, la différence est significative, les habitants de Meknes sont significativement plus diabétique que le reste des marocains.
E FAUX. Cette fois-ci, puisque le α vaut 1%, on doit rechercher le z_{seuil}. À 1%, il vaut 2,32. Ainsi, $z_{seuil} < z_{calc}$. On rejette donc H_0 et on accepte par conséquent H_1.

Question 28 :

En France, certains enfants sont répondeurs au vaccin contre le tétanos, d'autres non. Dans le cadre d'une étude sur les retentissements des vaccins entre eux, on cherche à savoir si les probabilités de

réponse au vaccin contre l'hépatite B diffèrent entre les enfants répondeurs et non répondeurs au vaccin contre le tétanos.
On forme un échantillon de 140 enfants vaccinés contre le tétanos, dont un groupe A et B respectivement de 50 enfants répondeurs et 90 enfants non répondeurs. On les vaccine ensuite contre l'hépatite B et on compte 20 individus répondeurs dans le groupe A et 40 pour le groupe B.
On réalise un Chi2 au risque α=5%.

A. Les conditions d'application du test ne sont pas vérifiées.
B. Les réponses des vaccins contre le tétanos et contre l'hépatite B sont dépendantes.
C. Le test est à 2 ddl car l'étude compare deux groupes d'enfants.
D. Le test étant bilatéral, il est impossible de conclure dans le sens de la différence.
E. La valeur test vaut exactement 3.

Question 28 : Aucune réponse juste (impossible sur sides)

On dresse le tableau suivant :

	Répondeurs HB	Non répondeurs HB	
Répondeurs tétanos (A)	20 **(24)**	40 **(36)**	60
Non répondeurs tétanos (B)	40 **(36)**	50 **(54)**	90
	60	90	150

Le test est bilatéral car on ne précise pas le sens de la différence dans l'énoncé, on cherche juste à savoir si les résultats peuvent être simplement différents.
À partir du tableau, on calcule les effectifs attendus pour chaque effectif observé : pour cela, on pose la formule suivant : $\frac{total\ de\ la\ ligne \times total\ de\ la\ colonne}{total\ final}$. On obtient alors : (en violet sur le tableau).

A FAUX Les conditions d'applications du Chi2 en comparaison de plusieurs proportions observées sont les suivantes : il faut que les effectifs attendus soient tous supérieurs à 5. C'est le cas : les conditions sont vérifiées.

B FAUX La formule du calcul de la valeur test est : $\sum \frac{(Oi-Ei)^2}{Ei}$.
On aura donc :

$$\frac{(20-24)^2}{24} + \frac{(40-36)^2}{36} + \frac{(40-36)^2}{36} + \frac{(50-54)^2}{54}$$
$$= \frac{4^2}{24} + \frac{4^2}{36} + \frac{4^2}{36} + \frac{4^2}{54}$$
$$= 0,7 + 0,4 + 0,4 + 0,3$$
$$= 1,8$$

À 5% d'erreur, dans un test bilatéral, le Z_{seuil} vaut 3,84 (à regarder dans la table des fractiles du Chi2 à 1 ddl). Donc $Z_{seuil} > Z_{calc}$, ainsi on accepte H_0. Il n'y a donc pas de différence significative entre les réponses aux différents vaccins.

C FAUX L'item ne veut strictement rien dire, le test est à 1 ddl car on ne compare que deux groupes. Pour savoir à combien de ddl on effectue un Chi2 on pose la formule suivante : (nombre de colonne-1) × (nombre de lignes-1)=(2-1)×(2-1) = 1.

D FAUX La bilatéralité n'influe aucunement sur les conclusions qu'on peut tirer les études. On peut tout à fait conclure en faveur d'un test ou d'un groupe, même en ayant réalisé une étude bilatérale.

E FAUX La valeur test vaut 1,8.

Question 29 :

Lors d'une fracture du col fémoral, deux traitements sont proposés : le traitement chirurgical ou le traitement orthopédique. On veut savoir si les probabilités de décès entre les deux traitements diffèrent. On réalise une étude chez des patients suivis pendant 5 ans après leur traitement. Le tableau suivant présente les résultats de l'étude :

	Vivants	morts	
TTT chir.	180	20	200
TTT ortho.	60	40	100
	240	60	300

On utilisera l'approximation de la loi Normale au risque d'erreur α=5%.

Aide au calcul $\sqrt{2}$= 1,4 ; $\sqrt{3} = 1{,}7$; $0{,}68 = 0{,}7$

A. La valeur test calculée vaut approximativement 6
B. Au risque α=5% on rejette l'hypothèse nulle
C. Au risque α = 1% on rejette l'hypothèse alternative
D. L'estimation ponctuelle de la différence des probabilités de décès causés par les traitements est de 30%, significativement plus élevé en cas de traitement orthopédique ($p < 0.00004$).
E. La valeur du seuil de significativité peut-être fixée avant ou après le test dans le protocole de l'étude

Question 29 : ABD

L'énoncé indique la nécessité d'effectuer un test d'hypothèse bilatéral en comparant deux proportions observées pour tester l'hypothèse nulle d'égalité des 2 probabilités. L'énoncé nous indique que le test sera effectué en utilisant la loi normale, au risque d'erreur de 5%. On sait d'ores et déjà que la table de la loi normale centrée réduite est unilatérale, il faudra donc faire des adaptations.

On note Fa la proportion de décès pour le traitement chirurgical et Fb la proportion de décès pour le traitement orthopédique. Fa = 20/200 = 10% et Fb=40/100 = 40%.

A VRAI, la formule pour l'approximation de la loi normale en comparaison de proportions observées entre elles est :

$$Z = \frac{|Fa - Fb|}{\sqrt{(\pi \times (1-\pi))\left(\frac{1}{Na} + \frac{1}{Nb}\right)}}$$

Avec π = la probabilité de décès sous l'hypothèse nulle = $\frac{Fa \times Na + Fb \times Nb}{Ntotal} = \frac{60}{300} = 20\%$

Donc :

$$Z = \frac{|0{,}1 - 0{,}4|}{\sqrt{(0{,}2 \times 0{,}8)\left(\frac{1}{200} + \frac{1}{100}\right)}}$$

$$Z = \frac{0{,}3}{\sqrt{(0{,}16)\left(\frac{3}{200}\right)}} \approx \frac{0{,}3}{0{,}4 \ \times \left(\frac{1{,}7}{14}\right)} \approx \frac{0{,}3}{\frac{0{,}7}{14}} = \frac{42}{7} = 6$$

B VRAI Il faut chercher la valeur de la fonction de répartition de la loi normale centrée réduite telle que P(Z<z) = 1 - α/2 = 0,975 ou (P(Z>z) = α/2 = 2,5% (table 2). On retrouve évidemment 1,96 dans les 2 cas. Comme |Zcalc| > 1,96, on rejette H_0 au risque α = 5%. On peut calculer le niveau de significativité correspondant à la valeur calculée de la grandeur test. Petit hic, la table de la fonction de répartition de la loi normale ne fournit qu'une valeur maximum de Z de 4,09. On en déduit que la valeur de p est telle que $p < 0{,}00004$ (le double du complément à 1 de la probabilité renvoyée par la table de la fonction de répartition d'une loi normale pour la valeur de z = 4,09).

C FAUX car on sait que $p < 0{,}00004$. Par ailleurs, on ne rejette jamais H1. La conclusion du test d'hypothèse ne concerne que l'hypothèse nulle, puisque le test d'hypothèse est celui de l'hypothèse nulle. Donc soit on rejette l'hypothèse nulle (ce qui est le cas ici) et on accepte l'hypothèse alternative, soit le test n'a pas pu rejeter l'hypothèse nulle et on s'arrête là.

D VRAI l'estimation ponctuelle de la différence des proportion de décès entre le traitement chirurgical et le traitement orthopédique correspond à la différence suivante : 40%-10%.

E FAUX c'est du cours. La valeur du seuil de significativité du test est fixée dans le protocole d'étude, avant le démarrage effectif de celle-ci, et donc bien avant l'analyse des résultats.

Question 30 :

D'après la littérature, la proportion des français dormant moins de sept heures par nuit est de 25%. On cherche à tester l'hypothèse nulle que la proportion d'étudiants en première année de médecine (doublants et primants confondus) dormant moins de 7 heures par nuit est également de 25%.
On construit un échantillon aléatoire de 200 étudiants en première année de médecine, doublants et primants confondus, et on observe que la proportion de personne dormant moins de sept heures par nuit est de 35%.

A. Le Chi 2 calculé vaut $\chi^2 \approx 9{,}7$
B. Le Chi 2 calculé vaut vaut $\chi^2 \approx 10{,}7$
C. Le test ne permet pas de rejeter l'hypothèse nulle au seuil de significativité α = 5%
D. On rejette l'hypothèse nulle avec un degré de significativité $0{,}001<p<0{,}01$
E. Un Chi-2 ne peut se faire qu'avec des tests bilatéraux

Question 30 : BD

On est dans un schéma de comparaison d'une proportion observée (35%) et d'une probabilité théorique (25%). Le test à réaliser est bilatéral (puisqu'on ne nous précise rien). C'est un chi-2 classique.
A FAUX cf B
B VRAI, avant de se lancer dans le calcul, on vérifie que les conditions sont vérifiées. Il faut que les effectifs attendus Ei soient tous supérieurs à 5. Comment calculer les effectifs attendus ?
On pose donc simplement ceci :
P1 dormant moins de 7h par nuit : Oi = ? Ei = ?
P1 dormant plus de 7h par nuit : Oi = ? Ei= ?
Les effectifs observées sont ceux qu'on a réellement observés dans la population, soient 35% d'étudiants dormant moins de 7h par nuit (**70** étudiants) et 65% d'étudiants dormant plus de 7 heures par nuit (**130** étudiants).
Les effectifs attendus sous l'hypothèse nulle sont ceux correspondant aux probabilités théoriques, soient 25% d'étudiants dormant moins de 7h par nuit (**50** étudiants en P1) et 75% d'étudiants dormant plus de 7 heures par nuit (**150** étudiants de P1).
On a donc **:**
P1 dormant moins de 7h par nuit : Oi = 70 Ei = 50
P1 dormant plus de 7h par nuit : Oi = 130 E i= 150
On voit bien que nos effectifs attendus (50 et 150) sont supérieurs à 5, les conditions sont vérifiées, on peut continuer.
On pose la formule : Zcalc = $\sum \frac{(Ei-Oi)^2}{Ei} = \frac{(50-70)^2}{50} + \frac{(150-130)^2}{150} = \frac{400}{50} + \frac{400}{150} \approx 8 + 2{,}7 = 10{,}7$.
Réponse B VRAI.
C FAUX on encadre notre petit p avec la valeur de Chi2 à 1ddl. On obtient un petit p compris entre $0{,}001<p<0{,}01$. Donc p sera forcément <1%, par conséquent il est également inférieur à 5%. On peut donc rejeter l'hypothèse nul à $p<0{,}05$.
D VRAI, voir les explications précédentes.
E FAUX c'est du cours, on peut utiliser un CHI-2 à 1 test unilatéral ou un test bilatéral.

Question 31 :

Dans la population française, 25% des adolescents sont atteints de myopie. Jérémy, qui s'intéresse fortement à ces données, cherche à savoir si les adolescents qui jouent aux jeux vidéo de façon excessive ont un risque différent d'être myopes. Sur un échantillon aléatoire de 300 adolescents qui jouent aux jeux vidéo quatre à cinq heures par jour, une étude a permis le diagnostic de 90 cas de myopie. Jérémy monte une analyse statistique s'appuyant sur l'utilisation de la loi Normale. Le risque de première espèce est fixé à 5%.

A. Les conditions d'application de la loi normale sont vérifiées, Jérémy peut réaliser le test
B. La valeur de la grandeur test calculée vaut 2

C. Avec 2,2%<p<2,3%, la différence est significative
D. Avec 0,044%<p<0,046% vous rejetez l'hypothèse nulle
E. Si l'on avait utilisé la loi du Chi-2, la valeur de la statistique de test aurait été 4

Question 31 : ABE

Il s'agit de la comparaison d'une proportion observée à une probabilité théorique. La probabilité théorique est de 0,25. La proportion observée est de 0,30. L'énoncé est en faveur d'un test bilatéral : « … ont un risque différent …».

A VRAI Jérémy peut effectivement réaliser son petit test, car les conditions d'application de l'approximation de la loi normale dans le cadre
d'une comparaison de proportion observée à une probabilité théorique sont :

$n \times \pi 0 > 5$ et $n \times (1 - \pi 0) > 5$

Elles sont vérifiées ici puisque $300 \times 0{,}25 = 75 > 5$ et $300 \times 0{,}75 = 225 > 5$

B VRAI la variable aléatoire grandeur test s'écrit :

$$Z = \frac{F - \pi_0}{\sqrt{\dfrac{\pi_0 \times (1 - \pi_0)}{300}}} \sim \mathrm{N}(0{,}1)$$

Et l'application numérique qui en découle fournit la valeur prise par la variable aléatoire grandeur test sur l'échantillon d'étude

$$z = \frac{0{,}30 - 0{,}25}{\sqrt{\dfrac{0{,}25 \times 0{,}75}{300}}} = \frac{0{,}05}{\sqrt{\dfrac{0{,}25 \times 0{,}25}{100}}} = \frac{0{,}05}{\dfrac{1}{4} \times \dfrac{1}{10}} = 40 \times 0{,}05 = 2$$

C FAUX notre valeur test vaut 2. La fonction de répartition de la loi normale ϕ renvoie une probabilité de 0,9772 pour z = 2. On en déduit la valeur du niveau de significativité « p » pour un test bilatéral

$$P(|Z| > 2) = P(Z < -2) + P(Z > 2)$$
$$= (1 - \phi(2)) + (1 - \phi(2))$$
$$= 2 \times (1 - \phi(2)) = 2 \times (1 - 0{,}9772) = 0{,}0456$$

D FAUX, car p n'est PAS compris entre 0,044% et 0,046% mais entre 0,044 et 0,046. Prudence avec les % !!

E VRAI si on avait utilisé une loi de Chi-2 à 1 ddl, le calcul de la grandeur test aurait été :

$$\chi^2_{obs} = \frac{(90 - 75)^2}{75} + \frac{(210 - 225)^2}{225} = \frac{15^2}{75} + \frac{15^2}{225} = \frac{225}{75} + \frac{225}{225} = 3 + 1 = 4$$

Comme par ailleurs on sait qu'un Chi2 à 1 ddl est le carré d'une loi normale centrée réduite, il aurait été possible de répondre à la question sans effectuer de calcul !

Question 32 :

Suite à quelques crises isolées, Martin cherche activement à comparer l'efficacité de quatre traitements A, B, C, D contre l'épilepsie. Les probabilités de guérison des différents traitements sont comparées en effectuant un test du Chi-2. Les bras A, B, C, D incluent respectivement 70, 50, 80 et 100 patients. Les effectifs observés de guéris pour les traitements A, B, C, D sont respectivement de 45, 35, 50, 80 patients. On fixe le risque d'erreur de première espèce α = 5%. On note que la valeur du Chi-2 est de 8.

A. Les conditions d'application pour l'utilisation du Chi-2 sous l'hypothèse nulle sont vérifiées
B. Il est possible de conclure en faveur d'une différence d'efficacité entre traitements au niveau de significativité p < 1%
C. Il est possible de conclure en faveur d'une différence d'efficacité entre traitements, avec 2,5% < p < 5%

D. Les efficacités des traitements peuvent être alors comparées 2 à 2 avec chaque fois une probabilité critique de 5%

E. Les efficacités des traitements peuvent être alors comparées 2 à 2 avec une probabilité critique de 5%/6

Question 32 : ACE

Comparaison de k proportions observées. Le nombre de degrés de liberté (ddl) du test du Chi-2 est de (ncolonne-1) x (nligne-1) = (4-1) x (2-1) = 3 x 1 = 3 ddl.

A VRAI les effectifs calculés sous H0 sont respectivement 49, 35, 56, 70, 21, 15, 24 et 30 (expliquer le calcul, et montrer qu'il n'est nécessaire d'en calculer que 3).

B FAUX cf C

C VRAI on fournit un encadrement du complément à 1 de la probabilité donnée par la fonction de répartition d'une loi du Chi-2 à 3 ddl pour la valeur calculée du Chi-2 : Ainsi le niveau de significativité est compris entre les valeurs de probabilité 1-0,975 et 1-0,950, soit $0,025 < p < 0,05$ (ou $2,5\% < p < 5\%$).

D FAUX car 6 comparaisons sont alors réalisées (nombre de paires possibles parmi 4), rendant nécessaire la correction du risque de première espèce

E VRAI (voir fin du cours sur l'analyse de variance introduisant la correction de Bonferroni).

Question 33 – un Kiki 2 pour une star:

Dans la team biostats de Kiki, tout le monde connait Bob le chondrocyte. Néanmoins, la team UE3 se demande si sa présence dans les colles l'année dernière a suffi à le faire connaitre à tous les P2, ou si seuls les P2 qui avaient installé Discord le connaissent.

Ainsi, on prend un échantillon de 400 P2, 200 avaient Discord, et 200 n'avaient pas. Parmi ceux qui avaient Discord, 30 personnes ne le connaissent pas. Parmi ceux qui n'avaient pas Discord, 150 personnes le connaissent.

On propose un test d'hypothèse : on réalise un test du Chi-2, où l'hypothèse nulle est « Les P2 qui n'avaient pas Discord sont aussi nombreux à connaitre Bob le chondrocyte que les P2 qui avaient Discord » et l'hypothèse alternative « Les P2 qui n'avaient pas Discord ne sont pas aussi nombreux à connaitre Bob le chondrocyte que les P2 qui avaient Discord ».

A. L'hypothèse alternative est bilatérale.
B. Le Chi-2 calculé est inférieur à 6.
C. Si on prend α= 5%, on rejette l'hypothèse nulle.
D. Si on prend α= 2,5%, on rejette l'hypothèse nulle.
E. Si on prend α= 1%, on rejette l'hypothèse nulle.

Question 33 – un Kiki 2 pour une star: ACD

A VRAI Si elle avait été unilatérale, elle serait du type « Les P2 qui n'avaient pas Discord sont **moins** nombreux à connaître Bob le chondrocyte que les P2 qui avaient Discord » ou « Les P2 qui n'avaient pas Discord sont **plus** nombreux à connaître Bob le chondrocyte que les P2 qui avaient Discord ».

B FAUX On commence par faire un tableau :

- En **noir**, je mets toutes les valeurs données par la consigne, ce sont les effectifs observés.
- Entre parenthèses, en **violet**, je mets les effectifs attendus, c'est-à-dire que je fais comme si les personnes connaissant Bob ou ne le connaissant pas étaient réparties uniformément entre les 2 groupes.

	P2 ayant Discord	**P2 n'ayant pas Discord**	**Total**
P2 connaissant Bob	170 (160)	150 (160)	320

P2 ne connaissant pas Bob	30 (40)	50 (40)	80
Total	200	200	400

Maintenant, je peux calculer mon Chi-2 selon la formule :

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{1ddl}$$

Ce qui donne en appliquant les valeurs numériques :

$$X^2 = \frac{(170-160)^2}{160} + \frac{(150-160)^2}{160} + \frac{(30-40)^2}{40} + \frac{(50-40)^2}{40}$$

$$X^2 = \frac{(10)^2}{160} + \frac{(-10)^2}{160} + \frac{(-10)^2}{40} + \frac{(10)^2}{40}$$

$$X^2 = \frac{100}{160} + \frac{100}{160} + \frac{100}{40} + \frac{100}{40} = \frac{200}{160} + \frac{200}{40} = 1{,}25 + 5 = 6{,}25$$

Le Chi-2 calculé est donc supérieur à 6.

C VRAI On a trouvé un Chi-2 de 6,25. Maintenant, on doit lire dans la table la valeur de p, le degré de significativité, qui lui est associé : on lit dans la ligne de 1 ddl.

On calcule le ddl en faisant : **(lignes-1) x (colonnes-1)**

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

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

Exemple d'utilisation de la table

Soit X une variable aléatoire suivant une loi du χ^2 à 3 degrés de liberté. Soit $p = 0{,}95$ 6,25 valeur de x telle que $P(X < x) = 0{,}95$ est 7,8147.

ddl \ p	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,999
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,1015	0,4549	1,3233	2,7055	3,8415	5,0239	6,6349	10,8276
2	0,0100	0,0201	0,0506	0,1026	0,2107	0,5754	1,3863	2,7726	4,6052	5,9915	7,3778	9,2103	13,8155
3	0,0717	0,1148	0,2158	0,3518	0,5844	1,2125	2,3660	4,1083	6,2514	7,8147	9,3484	11,3449	16,2662
4	0,2070	0,2971	0,4844	0,7107	1,0636	1,9226	3,3567	5,3853	7,7794	9,4877	11,1433	13,2767	18,4668
5	0,4117	0,5543	0,8312	1,1455	1,6103	2,6746	4,3515	6,6257	9,2364	11,0705	12,8325	15,0863	20,5150
6	0,6757	0,8721	1,2373	1,6354	2,2041	3,4546	5,3481	7,8408	10,6446	12,5916	14,4494	16,8119	22,4577
7	0,9893	1,2390	1,6899	2,1673	2,8331	4,2549	6,3458	9,0371	12,0170	14,0671	16,0128	18,4753	24,3219

On a : 5,0239<6,25<6,6349

Soit : 0,01<p<0,025 (en effet, il faut faire 1-0,975 et 1-0,990)

Donc on est sûr qu'on peut rejeter l'hypothèse nulle avec **α= 5%, car p< α**

D VRAI Là encore, p<α, donc on peut rejeter l'hypothèse nulle.

E FAUX

Question 34 – Un Chi-2 hors du commun :

Lors d'un essai comparatif randomisé, on souhaite tester l'efficacité de deux nouveaux traitements A et B contre une maladie inflammatoire chronique de l'intestin, plus précisément la maladie de Crohn. Les 500 patients de l'essai sont randomisés en deux groupes de même effectif. Le protocole prévoit de tester l'hypothèse nulle H0 d'absence de différence de probabilité de guérison entre les 2 groupes en effectuant un test du Chi-2. Le risque de première espèce est fixé à 1%. A l'issue de l'essai, 420 guérisons sont observées : 200 avec le traitement A contre 220 avec le traitement B.

Aide au calcul : $\frac{20}{21} \approx 1$

A. Les conditions d'application du test sont vérifiées car les quatre effectifs observés sont bien supérieurs ou égaux à 5.

B. Au risque de première espèce fixé (1%) vous rejetez l'hypothèse nulle d'une même efficacité des traitements A et B.
C. Vous ne pouvez rejeter l'hypothèse nulle car 0,01 < p < 0,025.
D. Vous déclarez qu'il n'a pas été possible de mettre en évidence une différence significative et que la différence observée entre les guérisons des deux traitements est peut-être due au hasard.
E. La probabilité de ne pas rejeter l'hypothèse nulle sachant qu'elle est fausse est le complément à un de la puissance.

Question 34 – Un Chi-2 hors du commun : CDE

Tout d'abord on remarque qu'il s'agit ici de faire un test du Chi-2 à partir de la répartition des effectifs. De plus on fait un ***test bilatéral*** *car on nous donne aucune information justifiant la réalisation d'un test unilatéral. Le plus pratique avec les tests du chi-2, c'est de faire un tableau avec les effectifs observés pour trouver facilement les effectifs attendus.*

	Traitement A	Traitement B	Totaux
Guéris	200 I **210**	220 I **210**	420
Non Guéris	50 I **40**	30 I **40**	80
Totaux	250	250	500

A FAUX Tout est vrai dans la phrase excepté la nature de l'effectif, il s'agit des effectifs **ATTENDUS** sous l'hypothèse nulle qui sont supérieurs ou égaux à 5.

Les effectifs attendus sous l'hypothèse nulle se calcule de la manière suivante :
$E_{att\ A-Guéris} = \frac{420\ x\ 250}{500}$ = **210** et ainsi de suite pour les trois autres. Les 4 effectifs attendus sous l'hypothèse nulle étant tous supérieurs 5, les conditions d'application de la loi du Chi-2 sont vérifié.
B FAUX On doit donc calculer la statistique du test :

$$X^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(200-210)^2}{210} + \frac{(220-210)^2}{210} + \frac{(30-40)^2}{40} + \frac{(50-40)^2}{40}$$
$$= \frac{100}{210} + \frac{100}{210} + \frac{100}{40} + \frac{100}{40}$$
$$= \frac{200}{210} + \frac{200}{40} = \frac{20}{21} + \frac{20}{4} \cong 1 + 5 = 6$$

On regarde alors où se situe notre valeur dans la table du Chi-2 à (c-1).(l-1) ddl soit à (2-1) x (2-1) = 1ddl.

$$5{,}0239 < \chi^2 < 6{,}6349$$
$$0{,}975 < 1 - p < 0{,}990$$
$$0{,}01 < p < 0{,}025$$

⟹ Car le degré de significativité : p = P(X > x), la table donnant p = P($X < x$).
Avec un seuil de 1%, p > α, on ne peut pas rejeter l'hypothèse nulle H0 d'une même efficacité des 2 traitements.

C VRAI Cf l'explication de l'item B.

D VRAI Cf l'explication de l'item B.

E VRAI C'est du cours, en effet il s'agit du risque de seconde espèce β qui correspond à probabilité de ne pas rejeter l'hypothèse nulle sachant qu'elle est fausse. La puissance de l'étude étant égale au complément à 1 de β.

Question 35 – Un Cancer au premier rang :

Le cancer du sein se situe au 1er rang des cancers incidents chez la femme, nettement devant le cancer du côlon-rectum et le cancer du poumon. C'est aussi celui qui cause le plus grand nombre de décès chez la femme. On admet qu'en France, en 2018, environ 80% des femmes traitées pour cancer du sein sont vivantes et indemnes de récidive à 5 ans.
Un échantillon aléatoire de 400 femmes dont le cancer présente des caractéristiques biologiques particulières est tiré au sort afin de savoir si le pronostic de leur cancer diffère de celui de l'ensemble de la population des femmes ayant un cancer du sein.
Le protocole de l'étude (1) prévoit d'utiliser une loi normale et la réalisation d'un test d'hypothèse bilatéral, (2) fixe le risque de rejeter à tort l'hypothèse nulle d'égalité des proportions à 5%, (3) fixe le risque de ne pas rejeter l'hypothèse nulle sous une hypothèse alternative précisée à 10%.
Parmi les 400 femmes de l'échantillon d'étude, 340 sont vivantes et indemnes de récidive à 5 ans.

A. La puissance de l'étude, déterminée avant sa réalisation, est égale à 90%.
B. Les conditions d'application de la loi Normale ne sont pas vérifiées.
C. La valeur prise par la statistique du test est 2,5.
D. Au risque de première espèce consenti (5%), vous déclarez un meilleur pronostic des femmes dont le cancer présente ces caractéristiques biologiques particulières.
E. Si nous augmentons la taille de l'échantillons du nombre de femme suivis alors la densité de probabilité se resserre et par conséquent la puissance de l'étude augmente.

Question 35 – Un Cancer au premier rang : ACDE

Tout d'abord je tiens à préciser que les valeurs utilisées dans l'exercice se rapprochent de la réalité mais ont été simplifiées afin de vous permettre de calculer plus facilement ce qu'il vous est demandé.
Lorsque qu'un énoncé comme celui-ci vous est présenté, il faut analyser ce qui est demandé pour savoir quelle formule appliquer. Ici, on souhaite comparer une proportion OBSERVEE (de femmes vivantes sans récidive à 5 ans) à une proportion THEORIQUE.
Données de l'énoncé :

- Risque de première espèce = α = 5%
- Risque de seconde espèce = β = 10%
- Taille de l'échantillon = n = 400
- Probabilité théorique : $\pi_0 = 80\%$
- Probabilité Observée : $f = \frac{340}{400} = 85\%$

On pose les Hypothèses :

- **H0 : $\pi_0 = \pi_A$ = 80%**
- **H1 : $\pi_0 \neq \pi_A$** → Test Bilatérale

A VRAI La puissance de l'étude est déterminée avant de réaliser l'étude. Elle se calcule en prenant le complément à un du risque de seconde espèce β : Puissance = $1 - \beta = 1 - 0{,}1 = 90\%$.
B FAUX Les condition d'application de la loi Normale sont vérifié :
$n * \pi_0 = 400*0.80 = 320 > 5$
$n*(1-\pi_0) = 400*0.20 = 80 > 5$
C VRAI Il s'agit bien d'un exercice utilisant la loi Normale pour une comparaison de proportion observée à une théorique, il faut appliquer la formule du cours. La grandeur test est la variable aléatoire suivante : $Z = \frac{(\pi_A - \pi_0)}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ → **N(0 ; 1)** qui prend ici la valeur **z.**

$$z = \frac{(0,85 - 0,80)}{\sqrt{\frac{0,80(1-0,80)}{400}}}$$

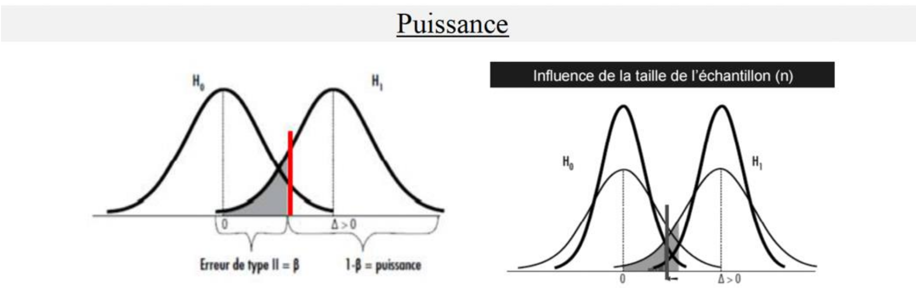
$$z = \frac{0,05}{\sqrt{\frac{0,80*0,20}{400}}}$$

$$z = \frac{0,05}{\sqrt{\frac{0,16}{400}}}$$
$$z = \frac{0,05}{\frac{0,4}{20}}$$
$$z = \frac{0,05 * 20}{0,4}$$
$$z = 2,5$$

D VRAI Puisque l'on effectue un test bilatéral, il faut calculer le petit p puis le comparer au risque α donné dans les items
p = 2*[1- ϕ(z)] = 2* [1- ϕ(2,5)] = 2*(1- 0,9938) = 0,0124

0,0124 < 0,05 => Donc **p < α ⇔ | $Z_{calculé}$| > |Z_{seuil}| : rejet de H0**
Donc la différence observée entre les deux proportions n'est pas due au hasard
On peut donc conclure que les caractéristiques biologiques particulières de ces cancers sont associées à un pronostic favorable.
E VRAI C'est du cours, en effet, à connaitre par-cœur car aussi utile en essais clinique.

- **Puissance = 1-β**



- ***Quand nous augmentons la taille de l'étude***, il y a moins de fluctuation : la densité se resserre, la puissance augmente.
- Il y a une opposition entre α et β, ***quand α augmente*** : la puissance augmente.
- ***Quand la différence attendue entre les moyennes augmente*** : la puissance augmente.

Question 36 – Merry Chi-2stmasss :

Dans le cadre de la lutinite aiguë, le Père Noël souhaite comparer l'efficacité de trois traitements TTT_A, TTT_B et le TTT_C à l'aide d'un essai randomisé en groupe parallèle, avec allocation aléatoire équilibrée de ces trois traitements. Le critère principal de jugement est la guérison clinique et sans récidive à 5ans (succès/échec). Le risque de première espèce est fixé à *α = 5%*. Dans un premier temps, le protocole prévoit la comparaison globale des trois traitements TTT_A, TTT_B et le TTT_C et la réalisation d'un test du Chi-2 bilatéral. A l'issue de l'essai, les effectifs randomisés dans les bras TTT_A, TTT_B et le TTT_C sont tous de 200 patients.
Les effectifs des patients de chaque groupe en succès (guérison) et en échec (absence de guérison) sont présentés dans le tableau ci-dessous.

Traitements	TTT_A	TTT_B	TTT_C	Total
Succès	65	117	120	302
Échec	135	83	80	298
Total	200	200	200	600

La première des statistiques du test calculée (Chi-2 ≈ 38,26) conduit à rejeter l'hypothèse nulle de probabilités de guérison identiques pour les 3 traitements. Les efficacités des traitements sont alors comparées 2 à 2 en réalisant 3 tests du Chi-2 interprétés en utilisant une correction de Bonferroni.

La valeur de la grandeur test calculée lors de la comparaison des traitements TTT_A et TTT_B vaut Chi-2 ≈ 27,26.
La valeur de la grandeur test calculée lors de la comparaison des traitements TTT_A et TTT_C vaut Chi-2 ≈ 30,42.
La valeur de la grandeur test calculée lors de la comparaison des traitements TTT_B et TTT_C vaut Chi-2 ≈ 0,09.
Aide au calcul : $\frac{0,05}{3} = 0,0167$

A. La première grandeur test calculée (Chi-2 ≈ 38,26) conduit à rejeter l'hypothèse nulle de probabilité de guérison identique pour les 3 traitements avec $p < 0,001$.
B. On peut conclure que le traitement TTT_B est significativement plus efficace que le traitement TTT_A.
C. On peut conclure que le traitement TTT_C est significativement plus efficace que le traitement TTT_A.
D. On peut conclure que le traitement TTT_C est significativement plus efficace que le traitement TTT_B.
E. Nous pouvons déclarer à la fin de l'étude que le traitement TTT_C est significativement meilleur que TTT_A et TTT_B.

Question 36 – Merry Chi-2stmasss : ABC

Cet exercice peut vous paraitre compliqué mais ne prenez pas peur !
Il est important dans un premier temps de relever toutes les informations fournies par l'énoncé :

- $N_{Tot} = 600$
- $\alpha = 5\%$
- $\text{Chi-2}_{TOT} \approx 38,26$
- $\text{Chi-2}_{A/B} \approx 27,26$
- $\text{Chi-2}_{A/C} \approx 30,42$
- $\text{Chi-2}_{B/C} \approx 0,09$

La principale difficulté de cet exercice était tout d'abord de déterminer le bon nombre de *ddl* en fonction du type de comparaison.

En effet lors d'une comparaison générale des trois traitements le calcul du nombre de *ddl* se fait par la formule suivante *ddl* = (colonne – 1)(ligne – 1) = (3 – 1)(2 – 1) = 2 *ddl.*

Or, dans la comparaison 2 à 2 le nombre de ddl = (2 – 1)(2 – 1) = 1 *ddl.*

Il est important de bien les prendre en considération en fonction de ce qui était demandé dans les items.

Aucun calcul de test n'est demandé dans cet exercice car toutes les statistiques de test sont fournies, il s'agissait principalement d'un exercice de lecture de table de la loi du Chi-2 avec le bon nombre de ddl et la bonne valeur seuil.

En effet, le protocole de l'étude prévoyait un seuil de significativité à 5%. Or ici, 3 comparaisons des traitements 2 à 2 sont réalisées. il fallait calculer le nouveau seuil $\alpha' = \alpha / Q$ avec Q le nombre de combinaison possible. Ici Q = $C_2^3 = \frac{3!}{2!(3-2)!} = 3$, soit 3 combinaisons possibles des traitements comparés 2 à 2. Donc $\alpha' = \frac{0,05}{3} = 0,016 = 1,67\%$.

A VRAI Ici Le Chi-2 calculé vaut 38,26. On regarde alors où se situe notre valeur dans la table du Chi-2 à 2ddl.

$$\chi^2 > 13,8155$$

$$1 - p > 0,999$$
$$p < 0,001$$

⟹ Car le degré de significativité : p = P(X > x), la table donnant p = P(X< x).
Avec un seuil de 5%, p < α, on peut rejeter l'hypothèse nulle H0 d'une même efficacité des 3 traitements. Il s'agit maintenant de comparer les traitements 2 à 2.
B VRAI Ici le Chi-$2_{A/B}$ ≈ 27,26. On regarde alors où se situe notre valeur dans la table du Chi-2 à 1ddl.

$$\chi^2 > 10{,}8276$$
$$1 - p > 0{,}999$$
$$p < 0{,}001$$

⟹ Car le degré de significativité : p = P(X > x), la table donnant p = P(X < x).

Avec un seuil de 1,67%, p < α', on peut rejeter l'hypothèse nulle H0 d'une même efficacité des traitements A et B. Le traitement TTT_B est significativement plus efficace que le traitement TTT_A.

C VRAI Ici Le Chi-$2_{A/C}$ ≈ 30,42 en le comparant à la valeur seuil, on regarde alors où se situe notre valeur dans la table du Chi-2 à 1ddl.

$$\chi^2 > 10{,}8276$$
$$1 - p > 0{,}999$$
$$p < 0{,}001$$

⟹ Car le degré de significativité : p = P(X > x), la table donnant p = P(X < x).

Avec un seuil de 1,67%, p < α', on peut rejeter l'hypothèse nulle H0 d'une même efficacité des traitements A et C. Le traitement TTT_C est significativement plus efficace que le traitement TTT_A.

D FAUX Ici Le Chi-$2_{B/C}$ ≈ 0,09. On regarde alors où se situe notre valeur dans la table du Chi-2 à 1ddl.

$$0{,}0158 < \chi^2 < 0{,}1015$$
$$0{,}10 < 1 - p < 0{,}25$$
$$0{,}75 < p < 0{,}90$$

⟹ Car le degré de significativité : p = P(X > x), la table donnant p = P(X < x).

Avec un seuil de 1,67%, p > α', on ne peut pas rejeter l'hypothèse nulle H0 d'une même efficacité de traitements TTT_B et TTT_C.

E FAUX Cf synthèse des résultats précédents.

Question 37 – PLAGE2 :

Alors que nos 100 P1 sont partis à la plage, ils se séparent en deux groupes de tailles identiques : le premier groupe se rend sur une plage de sable rouge et le deuxième sur une plage de sable jaune.

Dans le premier groupe, 15 P1 se font piquer par des méduses. Dans le deuxième groupe, seulement 5 P1 se sont fait piquer par des méduses. Le chef des P1 émet l'hypothèse que le deuxième groupe se soit moins fait piquer par les méduses grâce à la couleur du sable, mais le sous-chef l'arrête de suite : « tu ne peux pas dire que le deuxième groupe s'est moins fait piquer sans avoir fait de Chi-2 ! Montre-moi d'abord que la différence est bien significative au risque alpha = 5%, et qu'elle est en faveur du groupe 2, ensuite on discutera ! ».

A. Le Chi-2 calculé vaut 4,25.
B. L'hypothèse alternative est bilatérale.
C. On peut rejeter l'hypothèse nulle au risque alpha choisi par l'énoncé.
D. La différence entre les groupes n'est pas significative.
E. Si on avait choisi un risque alpha = 2%, on aurait pu rejeter l'hypothèse nulle.

Question 37 – PLAGE2 : CE

On commence par mettre toutes les infos dans un tableau :

	Piqué	Pas piqué	

Groupe 1	15	35	50
Groupe 2	5	45	50
	20	80	100

On ajoute dans le tableau les valeurs des effectifs théoriques (cela correspond aux effectifs qu'on aurait si les 20 P1 piqués étaient répartis uniformément entre les 2 groupes, donc si la probabilité de se faire piquer était la même dans les groupes).

	Piqué	Pas piqué	
Groupe 1	15|**10**	35|**40**	50
Groupe 2	5|**10**	45|**40**	50
	20	80	100

Et on applique la formule du Chi-2 :

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{1ddl}$$

Soit :

$$X^2 = \frac{(15-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(35-40)^2}{40} + \frac{(45-40)^2}{40}$$

$$X^2 = \frac{25}{10} + \frac{25}{10} + \frac{25}{40} + \frac{25}{40}$$

$$X^2 = \frac{50}{10} + \frac{50}{40} = 5 + 1{,}25 = 6{,}25$$

On doit donc maintenant regarder dans la table à quelle valeur de alpha un chi-2 de 6,25 correspond :

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

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

Exemple d'utilisation de la table

Soit X une variable aléatoire suivant une loi du χ^2 à 3 degrés de liberté. Soit $p = 0,95$ 6,25 valeur de x telle que $P(X < x) = 0,95$ est 7,8147.

ddl \ p	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,999
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,1015	0,4549	1,3233	2,7055	3,8415	5,0239	6,6349	10,8276
2	0,0100	0,0201	0,0506	0,1026	0,2107	0,5754	1,3863	2,7726	4,6052	5,9915	7,3778	9,2103	13,8155
3	0,0717	0,1148	0,2158	0,3518	0,5844	1,2125	2,3660	4,1083	6,2514	7,8147	9,3484	11,3449	16,2662
4	0,2070	0,2971	0,4844	0,7107	1,0636	1,9226	3,3567	5,3853	7,7794	9,4877	11,1433	13,2767	18,4668
5	0,4117	0,5543	0,8312	1,1455	1,6103	2,6746	4,3515	6,6257	9,2364	11,0705	12,8325	15,0863	20,5150
6	0,6757	0,8721	1,2373	1,6354	2,2041	3,4546	5,3481	7,8408	10,6446	12,5916	14,4494	16,8119	22,4577
7	0,9893	1,2390	1,6899	2,1673	2,8331	4,2549	6,3458	9,0371	12,0170	14,0671	16,0128	18,4753	24,3219

On se place à 1 ddl car le nombre de ddl correspond à (nb de lignes dans le tableau (sans compter les totaux – 1) x (nb de colonnes dans le tableau sans compter les totaux – 1) soit (2-1)x(2-1) = 1x1 = 1 ddl.

On lit donc que pour 6,25, 0,975 < p < 0,990, donc 0,01 < alpha < 0,025.

Or, ici, l'hypothèse du test est unilatérale, on divise donc le risque par 2 : soit : 0,005 < alpha < 0,0125.

Le risque alpha est inférieur à 1,25%. On peut donc rejeter l'hypothèse nulle pour alpha = 5%.

A FAUX Il vaut 6,25.

B FAUX « Montre-moi d'abord que la différence est bien significative au risque alpha = 5%, et qu'elle est en faveur du groupe 2 » signifie que l'on doit montrer que le risque est inférieur dans le 2e groupe, on donne donc un sens à la différence, l'hypothèse est unilatérale. C'est pourquoi on doit diviser le risque obtenu à la fin par 2.

C VRAI

D FAUX

E VRAI Le risque alpha est inférieur à 1,25% donc on peut aussi rejeter l'hypothèse nulle pour 2%. Attention, il fallait bien voir que le test était unilatéral !