

# Corrigé de l'examen de Biostatistique de 1<sup>ère</sup> Candi (Session 1)

Voici une correction d'un questionnaire choisi au hasard.

**Le questionnaire contient 40 questions.**

## Question n°1

Au seuil 0,025 unilatéral droit, si sous l'hypothèse nulle, une variable suit une distribution normale de moyenne 110 et de variance 100 et que, sous l'hypothèse alternative, elle suit une autre distribution normale de moyenne 115 et de déviation standard 9, combien vaut la puissance du test de l'hypothèse nulle (moyennes identiques) si la taille d'échantillon est de 6 individus ? Justifiez votre réponse en expliquant votre démarche.

• **Réponse correcte :** 0,206991

• **Justification:** Z1 au seuil unilatéral droit sur l'hypothèse nulle = 1,96 (voir table).  $\Rightarrow 1,96 = (X - 110) / (\text{racine}(100/6)) \Rightarrow X = 118$ . Cette valeur seuil est ensuite reportée dans la seconde distribution  $\Rightarrow Z2 = (118 - 115) / (9 / \text{racine}(6)) = 0,8165$  Si on regarde dans la table pour cette valeur de Z, depuis 0 à Z il y a 29,30%  $\Rightarrow$  la puissance vaut  $50\% - 29,30\% = 20,70\%$

## Question n°2

Quelle est la puissance approximative d'un sondage effectué sur 1000 personnes pour montrer qu'un candidat dispose d'une majorité de 52% (l'hypothèse nulle est évidemment qu'il n'y a pas de majorité)? On considérera  $\alpha = 5\%$  (test bilatéral).

• **Réponse correcte :** 0,243325304

• **Justification:** L'hypothèse nulle veut du 50-50  $\Rightarrow$  si on utilise l'approximation gaussienne :  $1,96 = (X - 500) / \text{racine}(500 * 0,5) \Rightarrow X = 530,99$ . Sous la distribution alternative,  $Z = (530,99 - 520) / \text{racine}(1000 * 0,52 * 0,48) = 0,69565$ . En regardant dans la table de Z, la puissance vaut 24,333 %

## Question n°3

Les eaux peuvent être polluées par la présence d'une bactérie A ou par celle d'une bactérie B. On a constaté la présence de A dans 20% des prélèvements et la présence de B dans 10% des prélèvements, 75% étant exempts de bactéries. Sachant qu'un prélèvement est contaminé par B, quelle est la probabilité qu'il soit également contaminé par A ?

• **Réponse correcte :** 0,5

• **Justification:** On peut recréer une table des probabilités: A+ A- B+ 5% 5% 10% B- 15% 75% 90% 20% 80% 100% On voit que si B est présente, 50% ( $p = p(A+ \text{ si } B+) = p(A \text{ et } B) / p(B) = 0,05 / 0,1$ ) des prélèvements contiendront aussi A.

## Question n°4

La probabilité d'être de sexe masculin est de 0,5. La probabilité qu'un individu de sexe masculin soit de groupe sanguin A est de 0,40. La probabilité d'être d'un autre groupe sanguin que A est de

0,60. Quelle est la probabilité qu'un individu pris au hasard soit une femme de groupe sanguin A ?

• **Réponse correcte :** 0,2

• **Justification:**  $p(A \text{ si } M) = 0,4$   $p(\text{autre que } A) = 0,6 \Rightarrow p(A) = 1 - 0,6 = 0,4$   $p(A \text{ et } M) = p(M) \cdot p(A \text{ si } M) = 0,5 \cdot 0,4 = 0,2$  On peut aussi recréer la table des probabilités : M F A 0,2 0,2 0,4 pas A 0,3 0,3 0,6 0,5 0,5 1

### Question n°5

La probabilité d'être de sexe masculin est de 0,5. La probabilité qu'un individu de sexe masculin soit de groupe sanguin A est de 0,40. La probabilité d'être d'un autre groupe sanguin que A est de 0,60. Quelle est la probabilité d'être du groupe sanguin A ?

• **Réponse correcte :** 0,4

• **Justification:**  $p(A) = 1 - p(\text{autre que } A) = 1 - 0,6 = 0,4$

### Question n°6

La probabilité d'être de sexe masculin est de 0,5. La probabilité qu'un individu de sexe masculin soit de groupe sanguin A est de 0,40. La probabilité d'être d'un autre groupe sanguin que A est de 0,60. Quelle est la probabilité qu'un individu de groupe sanguin A soit de sexe masculin ?

• **Réponse correcte :** 0,5

• **Justification:** On peut recréer la table des probabilités : M F A 0,2 0,2 0,4 pas A 0,3 0,3 0,6 0,5 0,5 1  $\Rightarrow p(M \text{ si } A) = p(A \text{ et } M) / p(A) = 0,2 / 0,4 = 0,5$

### Question n°7

L'institut National de Statistiques s'est intéressé au nombre d'accidents sur la route et démontre qu'en moyenne, on observe 2 accidents par quart d'heure en pleine heure de pointe. Quelle est la probabilité de n'observer aucun accident en un quart d'heure?

• **Réponse correcte :** 0,135335

• **Justification:** Utilisation de la loi de Poisson : moyenne = 2 et  $k = 0 \Rightarrow p(0) = \exp(-2) \cdot 2^0 / \text{fact}(0) = \exp(-2) = 0,135335$

### Question n°8

L'institut National de Statistiques s'est intéressé au nombre d'accidents sur la route et démontre qu'en moyenne, on observe 2 accidents par quart d'heure en pleine heure de pointe. Quelle est la probabilité d'observer plus de 3 accidents en un quart d'heure?

• **Réponse correcte :** 0,142877

• **Justification:** Utilisation de la loi de Poisson : moyenne = 2 et  $k = 4, 5, 6, \dots \Rightarrow p(4 \text{ et plus}) = 1 - p(0 \text{ à } 3) = 1 - \exp(-2) \cdot 2^0 / \text{fact}(0) - \exp(-2) \cdot 2^1 / \text{fact}(1) - \exp(-2) \cdot 2^2 / \text{fact}(2) - \exp(-2) \cdot 2^3 / \text{fact}(3) = 1 - 0,135335 - 0,270671 - 0,270671 - 0,180447 = 0,142877$

### Question n°9

L'institut National de Statistiques s'est intéressé au nombre d'accidents sur la route et démontre qu'en moyenne, on observe 2 accidents par quart d'heure en pleine heure de pointe. Quelle est la probabilité de n'observer aucun accident en une heure?

• **Réponse correcte :** 0,000335463

• **Justification:** Utilisation de la loi de Poisson. 1ère possibilité : moyenne = 2 par 1/4 d'heure  
 $\Rightarrow p = p(0)*p(0)*p(0)*p(0) = 0,135335^4 = 0,000335463$  2e possibilité : moyenne = 8 par heure  $\Rightarrow$   
 $p = p(0) = \exp(-8) = 0,000335463$

### Question n°10

L'institut National de Statistiques s'est intéressé au nombre d'accidents sur la route et démontre qu'en moyenne, on observe 2 accidents par quart d'heure en pleine heure de pointe. Quelle est la probabilité d'observer 2 accidents en une heure?

• **Réponse correcte :** *0,0107348*

• **Justification:** Utilisation de la loi de Poisson. 1ère possibilité : moyenne = 2 par 1/4 d'heure  
 $\Rightarrow p = p(0)*p(0)*p(1)*p(1)*\text{fact}(4)/(\text{fact}(2)*\text{fact}(2))$  possibilités +  
 $p(0)*p(0)*p(0)*p(2)*\text{fact}(4)/\text{fact}(3)$  possibilités = 0,010734804 2e possibilité : moyenne = 8 par  
heure  $\Rightarrow p = p(2) = \exp(-8)*8^2/\text{fact}(2) = 0,010734804$

### Question n°11

Un examen de statistique rencontre un taux d'échec de 35%. Quelle est la probabilité que, sur 10 étudiants sélectionnés aléatoirement dans l'auditoire, il y ait plus de 2 étudiants en échec ?

• **Réponse correcte :** *0,738393*

• **Justification:** Distribution binomiale : échec - réussite  $p(\text{échec})=0,35$  et  $n = 10$ .  $r$  vaut 2,3, ..., 10.  
 $\Rightarrow p(r) = p(2 \text{ et plus}) = 1 - (p(0)+p(1)+p(2)) = 1 -$   
 $(\text{combin}(10;0)*0,35^0*0,65^{10} + \text{combin}(10;1)*0,35^1*0,65^9 + \text{combin}(10;2)*0,35^2*0,65^8) =$   
0,738393

### Question n°12

Un examen de statistique rencontre un taux d'échec de 35%. Quelle est la probabilité que, sur 10 étudiants sélectionnés aléatoirement dans l'auditoire, il y ait entre 3 et 6 étudiants (inclus) qui aient réussi ?

• **Réponse correcte :** *0,481352*

• **Justification:** Distribution binomiale : échec - réussite  $p(\text{réussite})=1-0,35=0,65$  et  $n = 10$ .  $r$  vaut  
3,4,5 et 6.  $\Rightarrow p(r) = p(3)+p(4)+p(5)+p(6) = \text{combin}(10;3)*0,65^3*0,35^7 + \dots +$   
 $\text{combin}(10;6)*0,65^6*0,35^4 = 0,21203+0,6891+0,15357+0,237668 = 0,48135$

### Question n°13

Dans le test de Chi-carré, certaines conditions doivent être remplies afin de pouvoir l'appliquer et ainsi déterminer si une différence significative existe entre les observations ( $O_i$ ) et les espérances ( $E_i$ ). Parmi les propositions suivantes, laquelle est fautive ?

#### Réponses proposées :

- Les  $O_i$  sont des observations indépendantes
- Toutes les  $E_i > 5$  ou  $= 5$
- Les  $O_i$  doivent être distribuées normalement
- Les  $O_i$  et les  $E_i$  ne sont pas des pourcentages
- Je ne sais pas

• **Réponse correcte :** *Les  $O_i$  doivent être distribuées normalement*

• **Justification:** Les données observées sont des effectifs (comptage d'individus par exemple) et celles-ci ne doivent pas nécessairement suivre une distribution normale.

### Question n°14

Qu'appelle-t-on probabilité marginale ?

#### Réponses proposées :

- Je ne sais pas
- C'est la probabilité obtenue à partir d'une table de contingence en prenant l'effectif de la case et en le divisant par le total de sa colonne
- C'est la probabilité obtenue en divisant la somme d'une rangée ou colonne de la table de contingence par le total général
- C'est la probabilité obtenue à partir d'une table de contingence en prenant l'effectif de la case et en le divisant par le total de sa rangée
- C'est la probabilité obtenue à partir d'une table de contingence en prenant l'effectif de la case et en le divisant par le total général
- C'est la probabilité obtenue à partir d'une table de contingence en divisant 1 par le total général

• **Réponse correcte :** *C'est la probabilité obtenue en divisant la somme d'une rangée ou colonne de la table de contingence par le total général*

• **Justification:**

### Question n°15

Donnez la plage de variation de z.

#### Réponses proposées :

- de -1 à +1
- de -1,96 à +1,96
- de -3 à +3
- de -3,841 à 3,841
- de moins l'infini à plus l'infini
- de moins l'infini à zéro
- de zéro à 1
- de zéro à 3,841
- de zéro à plus l'infini
- il n'y a pas de plage de variation. C'est une valeur fixe
- Aucune des propositions
- Je ne sais pas

• **Réponse correcte :** *de moins l'infini à plus l'infini*

• **Justification:** Z peut prendre toutes les valeurs possibles depuis moins l'infini à 0 (50% de la courbe) et de 0 à plus l'infini (50% de la courbe)

### Question n°16

Quelle proportion de l'aire délimitée par la courbe de Gauss et l'axe des X est comprise entre -1,96 et 1,64 ?

• **Réponse correcte :** *0,9245*

• **Justification:** de -1,96 à 0 = 47,45 % de 0 à 1,64 = 45 % => la surface totale vaut la somme des deux = 92,45%

### Question n°17

Un enfant possède 5 pièces de monnaie des valeurs différentes. Combien peut-il former de sommes d'argent différentes ?

• **Réponse correcte :** 31

• **Justification:** Chaque pièce peut-être choisie ou non => 2 cas possibles par pièce =>  $2*2*2*2*2 = 2^5 = 32$  et on exclut le cas où aucune des pièces n'est choisie => 31 possibilités. Autre solution :  $\text{combin}(5;1)+\text{combin}(5;2)+\dots+\text{combin}(5;5) = 5+10+10+5+1 = 31$

### Question n°18

Spécifiez si la donnée est discrète, continue ou indéterminable : la durée de vie des tubes de télévision fabriqués par une société.

**Réponses proposées :**

- Discrètes
- Continues
- Indéterminable
- Je ne sais pas

• **Réponse correcte :** *Continues*

• **Justification:** La durée de vie des tubes est un temps qui peut prendre n'importe quelle valeur depuis 0 jusqu'à plus l'infini. C'est donc bien une variable continue.

### Question n°19

Afin d'étudier le poids des moutons de race Texel, un échantillon de 10 brebis a été prélevé :

66	64	68	54	57	61	57	54	62	63
----	----	----	----	----	----	----	----	----	----

Donnez la valeur de la moyenne arithmétique.

• **Réponse correcte :** 60,6

• **Justification:** C'est la somme de toutes les valeurs divisée par le nombre de données soit  $606/10 = 60,6$

### Question n°20

Afin d'étudier le poids des moutons de race Texel, un échantillon de 10 brebis a été prélevé :

66	64	68	54	57	61	57	54	62	63
----	----	----	----	----	----	----	----	----	----

Donnez la valeur de la médiane.

• **Réponse correcte :** 61,5

• **Justification:** C'est la valeur telle que 50% des données lui sont inférieures (ou supérieures). Elle occupe ici la place 5,5 ( $=(10+1)*50/100$ ). Par interpolation la valeur de la place 5,5 une fois les données triées par ordre croissant vaut 61,5 (mi-chemin entre 61 et 62).

### Question n°21

Afin d'étudier le poids des moutons de race Texel, un échantillon de 10 brebis a été prélevé :

66	64	68	54	57	61	57	54	62	63
----	----	----	----	----	----	----	----	----	----

Donnez la valeur de la variance.

• **Réponse correcte :** 24,044

• **Justification:** C'est la somme des différences au carré entre la donnée et sa moyenne. Cette somme est ensuite divisée par le nombre de données moins une. Soit  $((66-60,6)^2 + (64-60,6)^2 + \dots$

$$+ (63-60,6)^2)/(10-1) = 216,4/9 = 24,044$$

### Question n°22

Afin d'étudier le poids des moutons de race Texel, un échantillon de 10 brebis a été prélevé :

66	64	68	54	57	61	57	54	62	63
----	----	----	----	----	----	----	----	----	----

Donnez la valeur de l'erreur standard.

• **Réponse correcte :** 1,551

• **Justification:** L'erreur standard est la déviation standard associée à la moyenne. Elle se calcule à partir de la variance des données: racine(variance/taille échantillon)=racine(24,044/10) = 1,55063.

### Question n°23

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

Si, en réalité, la prévalence chez les fumeurs est de 22 %, quelle est la proportion d'individus qui fument et ont des problèmes respiratoires ?

• **Réponse correcte :** 0,0814

• **Justification:**  $p(M \text{ et } F) = p(M \text{ si } F) * p(F) = 0,22 * 0,37 = 0,0814$

### Question n°24

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

Quelle est la proportion de cas dans cette population ?

• **Réponse correcte :** 0,1507

• **Justification:** On peut reconstituer la table des probabilités : M S F 0,0814 0,2886 0,37 NF 0,0693 0,5607 0,63 0,1507 0,8493 1 => la proportion de cas vaut 0,1507

### Question n°25

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

Quelle est la proportion de fumeurs parmi les cas dans cette population ?

• **Réponse correcte :** 0,540145985

• **Justification:** On peut reconstituer la table des probabilités : M S F 0,0814 0,2886 0,37 NF 0,0693 0,5607 0,63 0,1507 0,8493 1 => la proportion de fumeurs parmi les cas vaut  $0,0814/0,1507 =$

**Question n°26**

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

L'expérience porte sur 320 cas et 390 contrôles.

On a observé 183 fumeurs parmi les cas, et 131 parmi les contrôles.

Quelle est le nombre attendu de fumeurs parmi les cas dans cette population sous l'hypothèse d'indépendance entre tabac et problèmes respiratoires ?

• **Réponse correcte :** *141,5211268*

• **Justification:** On peut reconstituer une table de contingence => l'effectif attendu =  $320 * (183+131)/710 = 141,5211268$

**Question n°27**

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

L'expérience porte sur 320 cas et 390 contrôles.

On a observé 183 fumeurs parmi les cas, et 131 parmi les contrôles.

Quelle est le nombre attendu de non-fumeurs parmi les contrôles dans cette population sous l'hypothèse d'indépendance entre tabac et problèmes respiratoires ?

• **Réponse correcte :** *217,5211268*

• **Justification:** On peut reconstituer une table de contingence => l'effectif attendu =  $390 * ((320-183)+(390-131))/710 = 217,5211268$

**Question n°28**

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

L'expérience porte sur 320 cas et 390 contrôles.

On a observé 183 fumeurs parmi les cas, et 131 parmi les contrôles.

Quelle est le test statistique utilisé pour tester l'hypothèse nulle ?

**Réponses proposées :**

- Loi binomiale
- Loi hypergéométrique
- Test de khi-deux
- Test de khi-deux dans une table de contingences
- Test exact de Fisher

- Test de Mac-Nemar
- Je ne sais pas

• **Réponse correcte :** *Test de khi-deux dans une table de contingences*

• **Justification:** On souhaite comparer dans une table de contingence des effectifs observés à des effectifs attendus (calculés à partir des totaux marginaux de la table de contingence). Plus la différence entre observés et attendus est faible, plus l'hypothèse nulle a de chance d'être acceptée. Le chi-carré peut être utilisé car les effectifs attendus sont supérieurs à 5 et l'effectif total est grand.

### Question n°29

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

L'expérience porte sur 320 cas et 390 contrôles.

On a observé 183 fumeurs parmi les cas, et 131 parmi les contrôles.

Quelle est la valeur de la statistique calculée ?

• **Réponse correcte :** *39,68163222*

• **Justification:** Le chi-carré vaut :  $(183-141,521)^2/141,521+(131-172,479)^2/172,479+(137-178,479)^2/178,479+(259-217,521)^2/217,521 = 39,682$

### Question n°30

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

L'expérience porte sur 320 cas et 390 contrôles.

On a observé 183 fumeurs parmi les cas, et 131 parmi les contrôles.

Quels sont les degrés de liberté (0 si non applicable) ?

• **Réponse correcte :** *1*

• **Justification:** Table de contingence 2x2 => (2-1)\*(2-1) ddl soit 1 ddl

### Question n°31

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

L'expérience porte sur 320 cas et 390 contrôles.

On a observé 183 fumeurs parmi les cas, et 131 parmi les contrôles.

Quelle est votre décision statistique (au seuil  $\alpha=5\%$ ) ?

**Réponses proposées :**

- Valeur significative: j'accepte l'hypothèse nulle.
- Valeur significative: je rejette l'hypothèse nulle.

- Valeur non significative: je rejette l'hypothèse nulle.
- Valeur non significative: j'accepte l'hypothèse nulle.
- Je ne sais pas

• **Réponse correcte :** *Valeur significative: je rejette l'hypothèse nulle.*

• **Justification:** La valeur de chi-carré pour 1 ddl au seuil 5% vaut 3,841 (voir table). Notre valeur de 39,682 est largement supérieure à la valeur théorique => hypothèse nulle rejetée (valeur significative). Les différences entre effectifs observés et effectifs attendus sont trop importantes pour qu'il y ait indépendance entre être fumeur et être malade.

### Question n°32

Une enquête Cas-Contrôles est menée dans la population européenne pour démontrer la nocivité du tabac.

L'hypothèse de départ est que la prévalence de problèmes respiratoires graves est de 11 % chez les non-fumeurs comme chez les fumeurs.

Un échantillonnage préalable a montré que la proportion de fumeurs est de 37 % dans cette population.

L'expérience porte sur 320 cas et 390 contrôles.

On a observé 183 fumeurs parmi les cas, et 131 parmi les contrôles.

Comment interprétez vous cette décision ?

#### **Réponses proposées :**

- Les proportions de fumeurs sont identiques qu'on regarde chez les malades ou chez les sains.
- Les proportions de fumeurs sont différentes si on regarde chez les malades ou chez les sains.
- Les proportions de malades sont identiques qu'on regarde chez les fumeurs ou chez les non-fumeurs.
- Les proportions de malades sont différentes si on regarde chez les fumeurs ou chez les non-fumeurs.
- On ne sait rien conclure avec ces données.
- Je ne sais pas

• **Réponse correcte :** *Les proportions de malades sont différentes si on regarde chez les fumeurs ou chez les non-fumeurs.*

• **Justification:** Les différences entre effectifs observés et effectifs attendus sont trop importantes pour qu'il y ait indépendance entre être fumeur et être malade. Si on regarde la table cela signifie qu'il y a plus de malades chez les fumeurs que chez les non-fumeurs.

### Question n°33

Les poids, dans une race bovine, sont théoriquement distribués selon une loi normale de moyenne 800 kilos et de déviation standard 50 kilos. On ventile 1000 bovins de cette race dans 4 classes définies comme suit: poids < 700 kilos (Classe I), poids  $\geq$  700 et < 800 kilos (Classe II), poids  $\geq$  800 et < 900 kilos (Classe III) et poids  $\geq$  900 kilos (Classe IV).

Quelle est la probabilité qu'un bovin pris au hasard dans cette race soit placé dans la classe I ?

• **Réponse correcte :** *0,02275*

• **Justification:** Soit 700 le X de notre distribution normale de moyenne 800 et de déviation standard 50 =>  $z = (700-800)/50 = -2$ . Si on regarde la surface de la distribution normale standardisée allant de moins l'infini à -2, elle vaut 0,022750062.

### Question n°34

Les poids, dans une race bovine, sont théoriquement distribués selon une loi normale de moyenne 800 kilos et de déviation standard 50 kilos. On ventile 1000 bovins de cette race dans 4 classes

définies comme suit: poids < 700 kilos (Classe I), poids  $\geq$  700 et < 800 kilos (Classe II), poids  $\geq$  800 et < 900 kilos (Classe III) et poids  $\geq$  900 kilos (Classe IV).

Quelle est la probabilité qu'un bovin pris au hasard dans cette race soit placé dans la classe II ?

• **Réponse correcte :** 0,47725

• **Justification:** Soit 700 le X inférieur de notre distribution normale de moyenne 800 et de déviation standard 50 et 800 le X supérieur  $\Rightarrow Z(i) = (700-800)/50 = -2$  et  $Z(s) = (800-800)/50 = 0$ . Si on regarde la surface de la distribution normale standardisée allant de -2 à 0, elle vaut 0,5-0,022750062 soit 0,477249938.

### Question n°35

Les poids, dans une race bovine, sont théoriquement distribués selon une loi normale de moyenne 800 kilos et de déviation standard 50 kilos. On ventile 1000 bovins de cette race dans 4 classes définies comme suit: poids < 700 kilos (Classe I), poids  $\geq$  700 et < 800 kilos (Classe II), poids  $\geq$  800 et < 900 kilos (Classe III) et poids  $\geq$  900 kilos (Classe IV).

On échantillonne 1000 bovins au hasard dans cette race.

Quelle est le nombre de bovins échantillonnés attendus en classe IV ?

• **Réponse correcte :** 22,75

• **Justification:** Soit 900 le X de notre distribution normale de moyenne 800 et de déviation standard 50  $\Rightarrow Z = (900-800)/50 = 2$ . Si on regarde la surface de la distribution normale standardisée allant de 2 à plus l'infini, elle vaut 0,022750062 soit  $0,02275 * 1000$  bovins = 22,75 bovins.

### Question n°36

Les poids, dans une race bovine, sont théoriquement distribués selon une loi normale de moyenne 800 kilos et de déviation standard 50 kilos. On ventile 1000 bovins de cette race dans 4 classes définies comme suit: poids < 700 kilos (Classe I), poids  $\geq$  700 et < 800 kilos (Classe II), poids  $\geq$  800 et < 900 kilos (Classe III) et poids  $\geq$  900 kilos (Classe IV).

On échantillonne 1000 bovins au hasard dans cette race.

Quelle est le test statistique utilisé pour tester l'hypothèse selon laquelle l'échantillon provient bien de la population décrite plus-haut ?

#### Réponses proposées :

- Loi binomiale
- Loi hypergéométrique
- Test de khi-deux
- Test de khi-deux dans une table de contingences
- Test exact de Fisher
- Test de Mac-Nemar
- Je ne sais pas

• **Réponse correcte :** Test de khi-deux

• **Justification:** On compare des effectifs observés (la répartition dans les 4 classes des 1000 bovins échantillonnés) à des effectifs attendus (les proportions de la distribution normale de moyenne 800 et de déviation standard 50 dans chacune des classes). Plus les valeurs entre observés et attendus sont proches, plus la valeur du chi-carré sera petite. Les effectifs attendus sont supérieurs à 5 et l'effectif total est suffisamment important.

### Question n°37

Les poids, dans une race bovine, sont théoriquement distribués selon une loi normale de moyenne 800 kilos et de déviation standard 50 kilos. On ventile 1000 bovins de cette race dans 4 classes

définies comme suit: poids < 700 kilos (Classe I), poids  $\geq$  700 et < 800 kilos (Classe II), poids  $\geq$  800 et < 900 kilos (Classe III) et poids  $\geq$  900 kilos (Classe IV).

On échantillonne 1000 bovins au hasard dans cette race.

Si on observe, respectivement, 18, 506, 449 et 27 bovins dans les 4 classes, quelle est la valeur de la statistique calculée ?

• **Réponse correcte :** 5,189852578

• **Justification:** On calcule le chi-carré. Il faut pour cela calculer chacun des effectifs attendus selon la distribution normale dans chacune des classes puis on effectue le calcul  $(18-22,75)^2/22,75+(506-477,25)^2/477,25+(449-477,25)^2/477,25+(27-22,75)^2/22,75 = 5,189852578$

### Question n°38

Les poids, dans une race bovine, sont théoriquement distribués selon une loi normale de moyenne 800 kilos et de déviation standard 50 kilos. On ventile 1000 bovins de cette race dans 4 classes définies comme suit: poids < 700 kilos (Classe I), poids  $\geq$  700 et < 800 kilos (Classe II), poids  $\geq$  800 et < 900 kilos (Classe III) et poids  $\geq$  900 kilos (Classe IV).

On échantillonne 1000 bovins au hasard dans cette race.

Quels sont les degrés de liberté (0 si non applicable) ?

• **Réponse correcte :** 3

• **Justification:** Il existe 4 classes => k-1 ddl soit 3 ddl.

### Question n°39

Les poids, dans une race bovine, sont théoriquement distribués selon une loi normale de moyenne 800 kilos et de déviation standard 50 kilos. On ventile 1000 bovins de cette race dans 4 classes définies comme suit: poids < 700 kilos (Classe I), poids  $\geq$  700 et < 800 kilos (Classe II), poids  $\geq$  800 et < 900 kilos (Classe III) et poids  $\geq$  900 kilos (Classe IV).

On échantillonne 1000 bovins au hasard dans cette race.

Quelle est votre décision statistique (au seuil  $\alpha=5\%$ ) ?

#### Réponses proposées :

- Valeur significative: j'accepte l'hypothèse nulle.
- Valeur significative: je rejette l'hypothèse nulle.
- Valeur non significative: je rejette l'hypothèse nulle.
- Valeur non significative: j'accepte l'hypothèse nulle.
- Je ne sais pas

• **Réponse correcte :** *Valeur non significative: j'accepte l'hypothèse nulle.*

• **Justification:** Au seuil 5%, la table de chi-carré pour 3 ddl non donne la valeur théorique de , supérieure à 5,189852578 => valeur non significative et acceptation de l'hypothèse nulle.

### Question n°40

Les poids, dans une race bovine, sont théoriquement distribués selon une loi normale de moyenne 800 kilos et de déviation standard 50 kilos. On ventile 1000 bovins de cette race dans 4 classes définies comme suit: poids < 700 kilos (Classe I), poids  $\geq$  700 et < 800 kilos (Classe II), poids  $\geq$  800 et < 900 kilos (Classe III) et poids  $\geq$  900 kilos (Classe IV).

On échantillonne 1000 bovins au hasard dans cette race.

Comment interprétez vous cette décision ?

#### Réponses proposées :

- Répartition des poids dans l'échantillon conforme à la loi attendue.
- Répartition des poids dans l'échantillon non conforme à la loi attendue.

- Les poids de bovins suivent toujours une loi normale.
- On ne peut pas dire, sur base d'une telle expérience, si le poids suit une distribution attendue.
- Je ne sais pas

• **Réponse correcte :** *Répartition des poids dans l'échantillon conforme à la loi attendue.*

• **Justification:** L'hypothèse nulle a été acceptée donc la répartition des poids dans l'échantillon est bien conforme à la loi attendue (la loi normale de moyenne 800 kg et de déviation standard 50 kg).