Estimations ponctuelles et par intervalles – Exercices – Devoirs

Exercice 1 corrigé disponible

Le staff médical d'une grande entreprise réalise ses statistiques sur le nombre de kilomètres parcourus pour se rendre à son travail en réalisant un échantillonnage sur 15 employés

Nombre de kilomètres	Nombre de personnes
20	5
25	3
30	2
40	5

Aide aux calculs : $\sqrt{\frac{530}{7}} \approx 9$; $\sqrt{15} \approx 4$

1. Calculer la moyenne m et l'écart type s de cet échantillon
2. Estimer la moyenne et l'écart-type pour la population de cette entreprise
3. Déterminer un intervalle de confiance à 95,5% pour la moyenne

QCM 2 corrigé disponible

Un échantillon de taille $n = 1000$ a été tiré au sort dans une population américaine afin d'estimer la moyenne μ et la variance σ^2 de la variable « tour de taille » (notée T). L'estimation de la moyenne est égale à $m = 100$ cm et l'estimation de la variance est égale à $s^2 = 16$ cm².

Indiquer si les propositions suivantes sont vraies ou fausses.

A. L'estimation de la moyenne μ à partir d'un échantillon est une variable aléatoire

B. D'après le théorème central limite, la distribution de l'estimation de la moyenne μ à partir d'un échantillon de taille $n = 1000$ suit une loi normale centrée sur la valeur μ

C. D'après le théorème central limite, la distribution de l'estimation de la moyenne μ à partir d'un échantillon de taille $n = 1000$ suit une loi normale dont la variance est égale à σ^2

D. L'estimation par intervalle de confiance à 95% de la moyenne μ à partir des résultats de l'énoncé est :

$$\left[100 - 1{,}96 \times \sqrt{\frac{16}{1000}} < \mu < 100 + 1{,}96 \times \sqrt{\frac{16}{1000}}\right]$$

E. A partir des mêmes estimations ponctuelles m et s^2, l'estimation de la moyenne μ par un intervalle de confiance à 80% donne un intervalle plus large que son estimation par un intervalle de confiance à 95%

QCM 3 corrigé disponible

On souhaite étudier la distribution de la taille des étudiants de sexe masculin en Allemagne en 2019. On note respectivement μ_{taille} et σ_{taille} la moyenne et l'écart-type de cette distribution. Afin d'estimer la valeur μ_{taille}, un échantillon représentatif de cette population est constitué, de taille $n = 100$. Les estimations obtenues à partir de cet échantillon indiquent que la distribution de la taille a une moyenne $m = 180$ cm et un écart-type de $s = 6$ cm. On note T_i la taille d'un étudiant i, avec $i \in \{1, 2, 3, \ldots, 100\}$ dans cet échantillon.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. Pour estimer μ_{taille} à partir de cet échantillon, on peut appliquer l'estimateur $m = \frac{\sum_{i=1}^{100} T_i}{100}$ qui est sans biais

B. L'estimation ponctuelle de μ_{taille} à partir d'un échantillon représentatif de taille $n = 100$ est une variable aléatoire qui suit une loi normale centrée sur μ_{taille} et de variance $\frac{\sigma_{taille}^2}{100}$

C. A partir de cet échantillon, on peut estimer la moyenne μ_{taille} avec un intervalle de confiance à 95% dont les bornes sont d'environ [178,8 cm ; 181,2 cm]

D. Si on sélectionnait un autre échantillon représentatif de cette population, la valeur de l'estimation ponctuelle de la taille moyenne ne serait pas nécessairement égale à 180 cm

E. Si on sélectionnait un échantillon de taille $n = 1000$, on pourrait estimer la moyenne μ_{taille} avec un intervalle de confiance à 95% de plus petite amplitude que l'intervalle de confiance à 95% obtenu avec un échantillon de taille $n = 100$

QCM 4 corrigé disponible

On souhaite étudier un score de qualité de vie (noté QdV) dans une population P. Pour cela, un échantillon représentatif de la population P et de taille $n = 100$ a été sélectionné. Le score QdV peut varier entre 0 et 100. A partir des données de l'échantillon, on a estimé que la moyenne et la variance de la variable QdV étaient respectivement, $m = 72$ et $s^2 = 81$. On pourra appliquer l'approximation $1{,}96 \approx 2$.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. L'estimation non biaisée de la variance de la variable QdV dans la population P repose sur la formule suivante (où qdv_i est le score de qualité de vie d'un individu i de l'échantillon) :

$$s^2 = \frac{\sum_{i=1}^{n}(qdv_i - 72)^2}{100}$$

B. D'après le théorème central limite, la moyenne du score de qualité de vie dans la population P, estimée à partir d'un échantillon de taille $n = 100$, est une variable aléatoire qui suit une loi normale

C. D'après le théorème central limite, la moyenne du score de qualité de vie dans la population P, estimée à partir d'un échantillon de taille $n = 100$, est une variable aléatoire dont la variance peut être estimée par $\hat{\sigma}_m^2 = 81$

D. A partir de cet échantillon, l'intervalle de confiance à 95% de l'estimation de la moyenne du score de qualité de vie dans la population P est égal à [70,2 ; 73,8]

E. L'estimation à l'aide d'un intervalle de confiance à 95% correspond à une démarche inductive

QCM 5 corrigé disponible

Dans une population source, X est une variable quantitative de moyenne μ et de variance σ^2.

A propos de l'estimateur « moyenne de la variable X», correspondant à la fonction $T_n = \frac{\sum_{i=1}^{n} X_i}{n}$, obtenu à partir d'un échantillon de taille n, indiquer si les propositions suivantes sont vraies ou fausses.

A. Il s'agit d'un estimateur convergent

B. Pour que cet estimateur soit non-biaisé, il faudrait que le dénominateur soit égal à $(n - 1)$ plutôt qu'égal à n

C. Cet estimateur suit une loi normale quelle que soit la taille n de l'échantillon

D. La variance de l'estimateur T_n est égale à $\frac{\sigma^2}{n}$

E. Pour que l'estimateur T_n suive une loi normale, la variable X doit nécessairement suivre une loi normale

QCM 6 corrigé disponible

On s'intéresse à la survenue d'événements indésirables dans une population de patients traités au long cours par un médicament M. Une étude a été réalisée en sélectionnant un échantillon représentatif de ces patients. La taille de l'échantillon était $n = 200$ et les auteurs ont observé que 15% des patients avaient présenté au moins un événement indésirable au cours des 6 derniers mois de traitement. Les auteurs de l'étude ont également indiqué que les bornes de l'intervalle de confiance à 95% de ce pourcentage étaient $15\% \pm 1{,}96 \times \sqrt{\frac{0{,}15\times(1-0{,}15)}{200}} = [10{,}1\% \; ; \; 19{,}9\%]$.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. Si les auteurs faisaient une deuxième étude avec un nouvel échantillon représentatif de 200 patients, la probabilité d'observer 30 patients avec au moins un événement indésirable au cours des 6 derniers mois serait de 100%

B. En sélectionnant des échantillons de taille suffisamment élevée, l'estimateur du pourcentage de patients avec au moins un événement indésirable au cours des 6 derniers mois suit une loi normale

C. Les conditions d'application sont bien présentes pour calculer l'intervalle de confiance comme indiqué dans l'énoncé

D. Si les auteurs avaient choisi de calculer un intervalle de confiance à 90%, l'intervalle aurait été plus large que celui présenté dans l'énoncé

E. Si les auteurs avaient obtenu une estimation ponctuelle de 15% auprès d'un échantillon de taille $n = 400$, l'intervalle de confiance à 95% aurait été plus large que celui présenté dans l'énoncé

QCM 7 corrigé disponible

Dans une population source, X est une variable quantitative de moyenne μ et de variance σ^2 (avec $\sigma^2 > 0$).

A propos de l'estimateur « variance de la variable X », correspondant à la fonction $s^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n-1}$, obtenu à partir d'un échantillon de taille n, tiré au sort dans la population source, indiquer si les propositions suivantes sont vraies ou fausses.

A. Le fait d'avoir $\mathbb{E}(s^2) = \sigma^2$ signifie que cet estimateur est « convergent »

B. Dans la population source, la formule de la variance est : $\sigma^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n-1}$

C. L'estimateur s^2 tend vers 0 quand la taille d'échantillon n tend vers $+\infty$

D. Il n'est pas possible d'estimer σ^2 si la variable X ne suit pas une loi normale

E. Si $n > 30$, l'estimateur m de la moyenne μ suit une loi normale de variance $\frac{\sigma^2}{n}$

QCM 8 corrigé disponible

Dans un futur essai contrôlé randomisé évaluant l'efficacité d'un nouvel anti-hypertenseur, il est prévu de recruter un total de 100 patients qui seront exposés soit au nouvel anti-hypertenseur (bras A), soit à un anti-hypertenseur de référence (bras B). Le statisticien prévoit de faire une randomisation simple pour répartir les patients entre les bras A et B : on sait que le nombre de patients qui seront recrutés dans le bras A suivra une loi binomiale de paramètres $n = 100$ et $\pi = 0,50$. On appliquera l'approximation $1,96 \approx 2$.

Concernant le pourcentage p attendu de patients qui seront recrutés dans le groupe A, indiquer si les propositions suivantes sont vraies ou fausses.

A. En appliquant cette méthode de randomisation, il est possible que l'on n'ait pas exactement 50 patients dans le groupe A et 50 patients dans le groupe B

B. Avant de faire la randomisation, on peut calculer un intervalle de confiance autour de la proportion attendue de patients recrutés dans le groupe A

C. Les bornes de l'intervalle à 95% caractérisant les marges d'incertitude autour du pourcentage attendu de patients recrutés dans le groupe A sont d'environ 40% et 60%

D. Calculer les bornes de l'intervalle à 95% caractérisant les marges d'incertitude autour du pourcentage attendu de patients recrutés dans le groupe A correspond à une démarche inductive

E. Les conditions d'application sont bien respectées pour calculer les bornes de l'intervalle à 95% caractérisant les marges d'incertitude autour du pourcentage attendu de patients recrutés dans le groupe A

QCM 9 corrigé disponible

Pour évaluer l'intérêt d'un nouveau dispositif médical, on souhaite estimer le pourcentage de patients avec un succès thérapeutique et le pourcentage de patients ayant présenté un événement indésirable grave suite à son utilisation. Le dispositif a été évalué sur un échantillon de 100 patients, représentatifs des patients qui seront traités avec ce nouveau dispositif. On a observé que :

- **50 patients avaient un succès thérapeutique**
- **10 patients ont présenté un événement indésirable grave.**

Indiquer si les propositions suivantes sont vraies ou fausses.

A. A partir de cet échantillon, l'estimation ponctuelle de la proportion de patients avec un succès thérapeutique est de 50%

B. A partir de cet échantillon, on peut estimer la proportion de patients ayant présenté un événement indésirable grave par un intervalle de confiance

C. La variance de l'estimation de la proportion de patients ayant présenté un événement indésirable grave est estimée par l'expression $\frac{0{,}1\times 0{,}9}{100}$

D. A partir de cet échantillon, les conditions de validité ne sont pas présentes pour obtenir une estimation par intervalle (à 95%) de la proportion de patients avec un succès thérapeutique

E. L'intervalle à 95% qui encadre l'estimation de la proportion de patients avec un succès thérapeutique est plus large que l'intervalle à 95% qui encadre l'estimation de la proportion de patients ayant présenté des événements indésirables graves

QCM 10 corrigé disponible

On souhaite estimer la pression artérielle systolique moyenne d'une population P à partir d'un échantillon représentatif de cette population.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. L'estimation de la pression artérielle systolique moyenne est variable d'un échantillon à l'autre

B. L'estimation de la pression artérielle systolique moyenne à partir d'un échantillon suit une loi normale si l'échantillon est suffisamment grand (d'effectif supérieur à 30)

C. L'estimation de la pression artérielle systolique moyenne étant une variable aléatoire, on peut lui calculer une moyenne et une variance

D. L'estimateur T d'un paramètre θ est dit convergent si la variance de T augmente quand la taille d'échantillon augmente

E. L'estimateur T d'un paramètre θ est dit sans biais si l'espérance de T est égale à la valeur θ

QCM 11 corrigé disponible

On veut estimer le moyenne (noté μ) du taux de gamma-GT (gamma-glytamyl transferase) chez les hommes en population générale. Pour cela, on a mesuré le taux de gamma-GT dans un échantillon de 1000 hommes tirés au hasard dans la population générale. Dans cet échantillon, l'estimation ponctuelle de la moyenne du taux de gamma-GT est m = 40 UI/L et l'estimation de l'écart type du taux de gamma-GT est s = 10 UI/L. La marge d'incertitude de l'estimation de μ est donnée par un intervalle à 95%.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. L'intervalle à 95% qui encadre l'estimation de μ est un intervalle de pari à 95%

B. Par définition, l'intervalle à 95% qui encadre l'estimation de μ contient 95% des valeurs possibles du taux de gamma-GT en population générale

C. Parmi les conditions d'application du calcul de cet intervalle à 95%, il faut vérifier que le taux de gamma-GT suive bien une loi normale

D. Dans cet échantillon de taille n=1000, si l'écart type du taux de gamma-GT avait une valeur estimée s = 15 UI/L (plutôt que s = 10 UI/L), alors l'intervalle à 95% qui encadre l'estimation de μ serait plus large

E. Si on recommence l'expérience sur un échantillon de taille n = 100 et que l'on obtient les mêmes estimations ponctuelles (m = 40 UI/L et s = 10 UI/L), l'intervalle à 95% encadrant l'estimation de μ sera plus large que celui de l'expérience initiale avec n = 1000

QCM 12 corrigé disponible

Une étude est réalisée pour décrire la consommation de cannabis chez les lycéens. Un échantillon (de taille n=100) représentatif des lycéens de Haute-Garonne est constitué. Parmi eux, 40 déclarent en avoir déjà consommé. On souhaite estimer le pourcentage (noté π) de lycéens ayant déjà consommé du cannabis dans la population de Haute-Garonne, en donnant une marge d'incertitude de l'estimation avec un intervalle à 95%.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. La variable dénombrant le nombre de lycéens ayant déjà consommé du cannabis dans cet échantillon suit une loi de Bernoulli

B. L'intervalle à 95% centré sur l'estimation ponctuelle de π est un intervalle de confiance à 95%

C. Puisque le nombre de lycéens dans l'échantillon est supérieur à 30, il n'y a pas de condition d'application à vérifier pour le calcul des bornes de l'intervalle à 95%

D. Le calcul de cet intervalle à 95% correspond à une démarche inductive

E. En considérant que $z_{2,5\%} \approx 1{,}96$, l'intervalle à 95% encadrant l'estimation ponctuelle de π peut être estimé par

$$\left[0{,}40 - 1{,}96 \times \frac{0{,}40 \times (1-0{,}40)}{100} < \pi < 0{,}40 + 1{,}96 \times \frac{0{,}40 \times (1-0{,}40)}{100}\right]$$

QCM 13 corrigé disponible

Dans un échantillon de 100 patients, représentatif de la population des patients d'une même commune, on compte le nombre de fois où chaque patient a été hospitalisé dans le passé. On trouve 50 patients avec 0 hospitalisation, 40 patients avec 1 hospitalisation passée et 10 patients avec 2 hospitalisations passées. Après calcul, on trouve que le nombre moyen d'hospitalisations passées est $m = 0,60$ et l'écart-type est $s = 0,65$.

Dans les calculs, on approximera la valeur 1,96 par 2.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. 95% des patients des patients de la commune ont un nombre d'hospitalisations passées compris entre $m - 2s$ et $m + 2s$

B. L'estimation du nombre moyen d'hospitalisations passées par patient dans cette population est 0,6

C. Le nombre moyen d'hospitalisations passées par patient dans cette population est compris entre 0,47 et 0,73, avec un risque d'erreur de 5%

D. Si on avait pris un échantillon de 1000 patients, en supposant qu'on observe toujours $m = 0{,}60$ et $s = 0{,}65$, l'intervalle de confiance à 95% du nombre moyen d'hospitalisations passées par patient serait plus large que celui calculé sur l'échantillon de 100 patients

E. Le pourcentage de patients sans hospitalisations passées dans cette population est compris entre 40% et 60%, avec un risque d'erreur de 5%

QCM 14 corrigé disponible

On a mesuré le périmètre crânien sur un échantillon de 300 enfants de 12 mois, tirés au sort et représentatifs de l'ensemble des enfants de 12 mois d'un département français. La moyenne calculée est de 46 cms ; l'intervalle de confiance à 95% donne des valeurs comprises entre 42 et 50 cms. Indiquez si les propositions suivantes sont vraies ou fausses.

A. 46 centimètres est l'estimation ponctuelle de la moyenne du périmètre crânien des enfants de l'échantillon

B. Les enfants de 12 mois de ce département français constituent la population source de cette étude

C. 95% des enfants de l'échantillon ont un périmètre crânien compris entre 42 et 50 cms

D. Si l'échantillon dans lequel on a mesuré le périmètre crânien avait été de 3000 enfants, l'intervalle de confiance aurait été plus large

E. Si on avait choisi un risque d'erreur de 0,001, l'intervalle de confiance aurait été plus large

QCM 15 corrigé disponible

On s'intéresse à la répartition de la masse m d'un médicament dans des flacons. L'analyse de fabrication de ce médicament a permis de constater que m était normalement distribuée avec une moyenne de 125mg et un écart-type de 5mg.

Indiquez si les propositions suivantes sont vraies ou fausses.

A. m est une variable aléatoire continue

B. Dans un lot de 100 flacons, le nombre moyen de flacons contenant plus de 132mg est supérieur à 16

C. Dans un lot de 100 flacons, le nombre moyen de flacons contenant moins de 135mg est supérieur à 95

D. Dans un lot de 100 flacons, le nombre moyen de flacons contenant moins de 115mg est inférieur à 5

E. Dans un lot de 100 flacons, le nombre moyen de flacons contenant plus de 137mg est inférieur à 2,5

QCM 16 corrigé disponible

On a mesuré la fréquence cardiaque sur un échantillon de 2000 joueurs de rugby, tirés au sort et représentatifs des joueurs de rugby de la fédération française. La moyenne calculée sur les 2000 sportifs est de 54 battements par minute (b/mn) ; l'intervalle à 95% associé au résultat de l'estimation ponctuelle donne des valeurs comprises entre 50 et 58.

Indiquez si les propositions suivantes sont vraies ou fausses.

A. Les 2000 joueurs de rugby constituent la population source de cette étude
B. L'intervalle donné est un intervalle de pari à 95%
C. 95% des joueurs de l'échantillon ont une fréquence cardiaque comprise entre 50 et 58 b/mn
D. Si l'effectif de l'échantillon de joueurs avait été de 200 joueurs, l'intervalle estimé aurait été moins large
E. Si on avait choisi un risque d'erreur de 0,001, l'intervalle estimé aurait été moins large

QCM 17 corrigé disponible

On souhaite estimer la proportion π de personnes droitières (qui écrivent de la main droite) en Haute-Garonne. Pour cela, on sélectionne un échantillon de 150 personnes, représentatif de la population de Haute-Garonne. Dans cet échantillon, 120 personnes (80% de l'échantillon) écrivent de la main droite.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. Si on recommence l'expérience en tirant au sort un nouvel échantillon de 150 personnes, représentatif de la population de Haute-Garonne, on s'attend à observer exactement 120 personnes qui écrivent de la main droite
B. Si les conditions de validité se vérifient, on pourrait estimer l'intervalle de confiance à 95% de la proportion de droitiers dans la population de Haute-Garonne par la formule suivante

$$\left[0{,}80 - 1{,}96 \times \sqrt{\frac{0{,}80 \times 0{,}20}{150}} < \pi < 0{,}80 + 1{,}96 \times \sqrt{\frac{0{,}80 \times 0{,}20}{150}}\right]$$

C. Si on recommence l'expérience sur un échantillon de taille $n = 1500$ (au lieu de 150), l'écart total entre les deux bornes de l'intervalle de confiance à 95% obtenue devrait être plus grand
D. Si on effectuait un grand nombre de tirages au sort d'échantillons représentatifs de taille $n = 150$ et qu'un intervalle de confiance à 90% était calculé pour chaque nouvel échantillon, alors 90% de ces intervalles contiendraient la valeur π
E. Pour vérifier les conditions de validité du calcul de l'intervalle de confiance à 95%, on vérifiera notamment que $150 \times \left(0{,}80 - 1{,}96 \times \sqrt{\frac{0{,}80\times0{,}20}{150}}\right) \geq 5$