CORRIGE DES EXERCICES : Estimation ponctuelle et estimation par intervalle

Exercice 1

$\mathcal{P}$={étudiants}
X= résultat au test de QI, variable quantitative de moyenne μ inconnue et d'écart-type σ =13 connu dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=30 sur lequel on observe $\overline{x} = 111$ qui est l'estimation ponctuelle de la moyenne inconnue μ.

1) X suit une loi $\mathcal{N}(\mu, \sigma = 13)$ donc quel que soit n, $\overline{X}_n$ suit une loi normale $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}} = \frac{13}{\sqrt{n}}\right)$;

pour n=30 $\frac{\sigma}{\sqrt{n}} = \frac{13}{\sqrt{30}} = 2{,}37$

- l'estimation par intervalle de confiance au niveau 95% (au risque α=5%) de μ dans $\mathcal{P}$ s'écrit :

$IC_{95\%}(\mu) = \left[\overline{x} \pm z_{0,975} \frac{\sigma}{\sqrt{n}}\right] = \left[111 \pm z_{0,975} \frac{13}{\sqrt{30}}\right] = [111 \pm 1{,}96 \times 2{,}37] \approx [111 \pm 4{,}7] = [106{,}3;\ 115{,}7]$ où $z_{1-(\alpha/2)} = z_{0,975}$ = 1,96 est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

➨ l'estimation par intervalle de confiance au niveau 95% du résultat moyen des étudiants est d'environ 106,3 à 115,7 ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 4,7.

- l'estimation par intervalle de confiance au niveau 90% (au risque α=10%) de μ dans $\mathcal{P}$ s'écrit :

$IC_{90\%}(\mu) = \left[\overline{x} \pm z_{0,95} \frac{\sigma}{\sqrt{n}}\right] = \left[111 \pm z_{0,95} \frac{13}{\sqrt{30}}\right] = [111 \pm 1{,}645 \times 2{,}37] \approx [111 \pm 3{,}9] = [107{,}1;\ 114{,}9]$ où $z_{1-(\alpha/2)} = z_{0,95}$ =1,645 est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

➨ l'estimation par intervalle de confiance au niveau 90% du résultat moyen des étudiants est d'environ 107,1 à 114,9 ; la précision (ou marge d'erreur) de l'estimation à 90% est d'environ 3,9.

- l'estimation par intervalle de confiance au niveau 99% (au risque α=1%) de μ dans $\mathcal{P}$ s'écrit :

$IC_{99\%}(\mu) = \left[\overline{x} \pm z_{0,995} \frac{\sigma}{\sqrt{n}}\right] = \left[111 \pm z_{0,995} \frac{13}{\sqrt{30}}\right] = [111 \pm 2{,}575 \times 2{,}37] \approx [111 \pm 6{,}1] = [104{,}9;\ 117{,}1]$ où $z_{1-(\alpha/2)} = z_{0,995}$ = 2,575 est le quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$.

➨ l'estimation par intervalle de confiance au niveau 99% du résultat moyen des étudiants est d'environ 104,9 à 117,1 ; la précision (ou marge d'erreur) de l'estimation à 99% est d'environ 6,1.

remarque : $IC_{99\%}(\mu)$ contient $IC_{95\%}(\mu)$ qui contient $IC_{90\%}(\mu)$

2) • Pour n=50 $\frac{\sigma}{\sqrt{n}} = \frac{13}{\sqrt{50}} = 1{,}838$:

- l'estimation par intervalle de confiance au niveau 95% (au risque α=5%) de μ dans $\mathcal{P}$ s'écrit :

$IC_{95\%}(\mu) = \left[111 \pm z_{0,975} \frac{13}{\sqrt{50}}\right] = [111 \pm 1{,}96 \times 1{,}838] \approx [111 \pm 3{,}6] = [107{,}4;\ 114{,}6]$ où $z_{1-(\alpha/2)} = z_{0,975}$ = 1,96 est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

- l'estimation par intervalle de confiance au niveau 90% (au risque α=10%) de μ dans $\mathcal{P}$ s'écrit :

$IC_{90\%}(\mu) = \left[111 \pm z_{0,95} \frac{13}{\sqrt{50}}\right] = [111 \pm 1{,}645 \times 1{,}838] \approx [111 \pm 3] = [108;\ 114]$ où $z_{1-(\alpha/2)} = z_{0,95}$ =1,645 est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

- l'estimation par intervalle de confiance au niveau 99% (au risque α=1%) de μ dans $\mathcal{P}$ s'écrit :

$IC_{99\%}(\mu) = \left[111 \pm z_{0,995} \frac{13}{\sqrt{50}}\right] = [111 \pm 2{,}575 \times 1{,}838] \approx [111 \pm 4{,}7] = [106{,}3;\ 115{,}7]$ où $z_{1-(\alpha/2)} = z_{0,995}$ = 2,575 est le quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$.

- Pour n=100 $\dfrac{\sigma}{\sqrt{n}} = \dfrac{13}{\sqrt{100}} = 1{,}3$:
  - l'estimation par intervalle de confiance au niveau 95% (au risque α=5%) de μ dans $\mathcal{P}$ s'écrit :

    $IC_{95\%}(\mu) = \left[111 \pm z_{0,975}\dfrac{13}{\sqrt{100}}\right] = [111 \pm 1{,}96 \times 1{,}3] \approx [111 \pm 2{,}5] = [108{,}5\,;\,113{,}5]$ où $z_{1-(\alpha/2)} = z_{0,975} = 1{,}96$ est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.
  - l'estimation par intervalle de confiance au niveau 90% (au risque α=10%) de μ dans $\mathcal{P}$ s'écrit :

    $IC_{90\%}(\mu) = \left[111 \pm z_{0,95}\dfrac{13}{\sqrt{100}}\right] = [111 \pm 1{,}645 \times 1{,}3] \approx [111 \pm 2{,}1] = [108{,}9\,;\,113{,}1]$ où $z_{1-(\alpha/2)} = z_{0,95} = 1{,}645$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.
  - l'estimation par intervalle de confiance au niveau 99% (au risque α=1%) de μ dans $\mathcal{P}$ s'écrit :

    $IC_{99\%}(\mu) = \left[111 \pm z_{0,995}\dfrac{13}{\sqrt{100}}\right] = [111 \pm 2{,}575 \times 1{,}3] \approx [111 \pm 3{,}3] = [107{,}7\,;\,114{,}3]$ où $z_{1-(\alpha/2)} = z_{0,995} = 2{,}575$ est le quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$.

  remarque : plus la taille n augmente plus les intervalles de confiance pour un même niveau de confiance sont étroits (meilleure précision).

3) La demi-longueur de l'intervalle $IC_{95\%}(\mu)$, correspondant à la marge d'erreur dans l'estimation du résultat moyen à 95%, est de 2,5 pour un échantillon de taille n=100 ; pour obtenir une marge d'erreur (demi-longueur) plus faible, égale à 1, il faudra augmenter la taille de l'échantillon n. Pour n inconnu, σ=13 et α=5% connus, la demi-longueur de l'intervalle $IC_{95\%}(\mu)$ s'écrit : $z_{0,975}\dfrac{\sigma}{\sqrt{n}} = 1{,}96\dfrac{13}{\sqrt{n}}$

   on cherche n tel que : $1{,}96\dfrac{13}{\sqrt{n}} \le 1$ c'est à dire $1{,}96 \times 13 \le \sqrt{n}$ d'où $n \ge (1{,}96 \times 13)^2 = 649{,}23$

   ➨ on choisira donc une taille d'échantillon au moins égale à 650 pour que demi-longueur de l'intervalle de confiance à 95% (la marge d'erreur dans l'estimation du résultat moyen à 95%) soit inférieure à 1.

Exercice 2

$\mathcal{P}$={enfants fréquentant la maternelle}
X= score au test de Pensée Créative de Torrance, variable quantitative de moyenne μ et d'écart-type σ inconnus dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=30

1) L'estimation ponctuelle du score moyen μ est donnée par la moyenne observée $\bar{x} = \dfrac{639}{30} = 21{,}3$

   ➨ le score moyen des enfants de maternelle est estimé à 21,3 (points de score).

2) L'estimation ponctuelle sans biais de la variance σ² est donnée par la variance observée sans biais :

   $s^{*2} = \dfrac{14\,691 - 30 \times 21{,}3^2}{29} = \dfrac{1080{,}3}{29} = 37{,}25$

   (autre calcul : $s^2 = \dfrac{14\,691}{30} - 21{,}3^2 = 36{,}01$ et $s^{*2} = \dfrac{30}{29} \times s^2 = 1{,}034 \times 36{,}01 = 37{,}25$)

   l'estimation ponctuelle sans biais de l'écart-type σ est donnée par l'écart-type observé sans biais $s^* = \sqrt{37{,}25} = 6{,}1$

   ➨ la variance du score des enfants de maternelle est estimée à 37,25 et son écart-type à 6,1 (points de score).

3) La loi de X étant quelconque et n=30≥30, $\overline{X}_n$ suit approximativement une loi normale $\mathcal{N}\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$ et σ est estimé par s*. L'estimation par intervalle de confiance au niveau 95% (au risque α=5%) de μ dans $\mathcal{P}$ s'écrit :

   $IC_{95\%}(\mu) \approx \left[\bar{x} \pm z_{0,975}\dfrac{s^*}{\sqrt{n}}\right] = \left[21{,}3 \pm 1{,}96\dfrac{6{,}1}{\sqrt{30}}\right] \approx [21{,}3 \pm 2{,}18] \approx [19{,}1\,;\,23{,}5]$ où $z_{1-(\alpha/2)} = z_{0,975} = 1{,}96$ est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

   ➨ l'estimation par intervalle de confiance au niveau 95% du score moyen des enfants de maternelle est d'environ 19,1 à 23,5 (points de score) ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 2,2 (points de score).

Exercice 3

$\mathcal{P}$={individus âgés de 20 à 30 ans}
X= temps nécessaire pour reproduire 16 modèles (mesuré en secondes), variable quantitative de moyenne μ et d'écart-type σ inconnus dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=60

1) L'estimation ponctuelle du temps moyen μ est donnée par la moyenne observée $\bar{x} = \frac{24\,056}{60} = 400{,}9$ secondes

➨ le temps moyen des individus âgés de 20 à 30 ans est estimé à 400,9 secondes.

2) L'estimation ponctuelle sans biais de la variance σ^2 est donnée par la variance observée sans biais :

$$s^{*2} = \frac{10\,253\,632 - 60 \times 400{,}9^2}{59} = 10\,345{,}5$$

(autre calcul : $s^2 = \frac{10\,253\,632}{60} - 400{,}9^2 = 10173{,}06$ et $s^{*2} = \frac{60}{59} \times s^2 = 1{,}0169 \times 10\,173{,}06 = 10\,345{,}5$)

l'estimation ponctuelle sans biais de l'écart-type σ est donnée par l'écart-type observé sans biais $s^* = \sqrt{10\,345{,}5} = 101{,}7$ secondes

➨ la variance du temps des individus âgés de 20 à 30 ans est estimée à 10 345,5 et son écart-type à 101,7 secondes.

3) Estimation par intervalle de confiance au niveau $1-\alpha$ ou au risque α du temps moyen μ dans $\mathcal{P}$:

La loi de X étant quelconque et n=60≥30, $\overline{X}_n$ suit approximativement une loi normale $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ et σ inconnu est estimé par s* d'où : $IC_{1-\alpha}(\mu) \approx \left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{s^*}{\sqrt{n}}\right] = \left[400{,}9 \pm z_{1-\frac{\alpha}{2}} \frac{101{,}7}{\sqrt{60}}\right] \approx \left[400{,}9 \pm z_{1-\frac{\alpha}{2}} \times 13{,}13\right]$

- Pour $1-\alpha = 90\%$ $\alpha = 10\%$ $z_{1-(\alpha/2)} = z_{0,95} = 1{,}645$ quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$

$$IC_{90\%}(\mu) \approx \left[400{,}9 \pm 1{,}645 \frac{101{,}7}{\sqrt{60}}\right] \approx [400{,}9 \pm 21{,}6] = [379{,}3\,;\ 422{,}5]$$

- Pour $1-\alpha = 95\%$ $\alpha = 5\%$ $z_{1-(\alpha/2)} = z_{0,975} = 1{,}96$ quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$

$$IC_{95\%}(\mu) \approx \left[400{,}9 \pm 1{,}96 \frac{101{,}7}{\sqrt{60}}\right] \approx [400{,}9 \pm 25{,}7] = [375{,}2\,;\ 426{,}6]$$

- Pour $1-\alpha = 99\%$ $\alpha = 1\%$ $z_{1-(\alpha/2)} = z_{0,995} = 2{,}575$ quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$

$$IC_{99\%}(\mu) \approx \left[400{,}9 \pm 2{,}575 \frac{101{,}7}{\sqrt{60}}\right] \approx [400{,}9 \pm 33{,}8] = [367{,}1\,;\ 434{,}7]$$

remarque : $IC_{99\%}(\mu)$ contient $IC_{95\%}(\mu)$ qui contient $IC_{90\%}(\mu)$

Exercice 4

$\mathcal{P}$={étudiants d'une promotion} X= temps de mémorisation d'un texte (mesuré en mn), variable quantitative de moyenne μ et d'écart-type σ inconnus dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=37 pour lequel $\bar{x} = 25$ et s=5

La loi de X étant quelconque et n=37≥30, $\overline{X}_n$ suit approximativement une loi normale $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Le temps moyen μ inconnu est estimé par la moyenne observée $\bar{x} = 25$ mn et l'écart-type du temps σ inconnu est estimé par l'écart-type observé sans biais $s^* = \sqrt{\frac{37}{36}} \times s = \sqrt{\frac{37}{36}} \times 5 = 5{,}07$ mn

L'intervalle de confiance au risque α=5% (au niveau 95%) du temps moyen μ dans $\mathcal{P}$ s'écrit :

$IC_{95\%}(\mu) \approx \left[\bar{x} \pm z_{0,975} \frac{s^*}{\sqrt{n}}\right] = \left[25 \pm 1{,}96 \frac{5{,}07}{\sqrt{37}}\right] \approx [25 \pm 1{,}96 \times 0{,}834] \approx [25 \pm 1{,}6] \approx [23{,}4\,;\ 26{,}6]$ où $z_{1-(\alpha/2)} = z_{0,975} = 1{,}96$ est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

➨ l'estimation par intervalle de confiance au risque 5% (au niveau 95%) du temps moyen de mémorisation d'un texte par les étudiants d'une promotion est d'environ 24,3 à 26,6 mn ; la précision (ou marge d'erreur) de l'estimation au risque 5% (à 95%) est d'environ 1,6 mn.

Exercice 5

$\mathcal{P}$={sujets} X= temps de parcours d'un labyrinthe (mesuré en mn), variable quantitative de moyenne μ et d'écart-type σ inconnus dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=100 pour lequel $\overline{x} = 8{,}76$ et s*=2,3

1) L'estimation ponctuelle du temps de parcours moyen μ est donnée par la moyenne observée $\overline{x} = 8{,}76$ mn

➨ le temps moyen de parcours du labyrinthe des sujets est estimé à 8,76 mn.

2) La loi de X étant quelconque et n=100≥30, $\overline{X}_n$ suit approximativement une loi normale $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ et l'écart-type du temps σ inconnu est estimé par l'écart-type observé sans biais s*=2,3 mn

L'intervalle de confiance au niveau 90% (au risque α=10%) du temps moyen μ dans $\mathcal{P}$ s'écrit :

$IC_{90\%}(\mu) \approx \left[8{,}76 \pm z_{0,95} \frac{2{,}3}{\sqrt{100}}\right] \approx [8{,}76 \pm 1{,}645 \times 0{,}23] \approx [8{,}76 \pm 0{,}38] \approx [8{,}38 ; 9{,}14]$ où $z_{1-(\alpha/2)} = z_{0,95} = 1{,}645$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

➨ l'estimation par intervalle de confiance au niveau 90% du temps moyen de parcours d'un labyrinthe des sujets est d'environ 8,38 à 9,14 mn ; la précision (ou marge d'erreur) de l'estimation à 90% est d'environ 0,38 mn.

3) La marge d'erreur dans l'estimation du temps moyen à 90%, donnée par la demi-longueur de l'intervalle $IC_{90\%}(\mu)$, est de 0,38 mn pour un échantillon de taille n=100 ; pour obtenir une marge d'erreur (demi-longueur) plus faible, de 0,3 mn, il faudra augmenter la taille de l'échantillon n. Pour n inconnu, σ=s*=2,3 et α=10% connus, la demi-longueur de l'intervalle $IC_{90\%}(\mu)$ s'écrit : $z_{0,95} \frac{s^*}{\sqrt{n}} = 1{,}645 \frac{2{,}3}{\sqrt{n}}$

on cherche n tel que : $1{,}645 \frac{2{,}3}{\sqrt{n}} \leq 0{,}3$ c'est à dire $\frac{1{,}645 \times 2{,}3}{0{,}3} \leq \sqrt{n}$ d'où $n \geq \left(\frac{1{,}645 \times 2{,}3}{0{,}3}\right)^2 = 159{,}05$

➨ on choisira donc une taille d'échantillon au moins égale à 160 pour que la marge d'erreur dans l'estimation du temps moyen à 90% soit inférieure à 0,3 mn.

Exercice 6

$\mathcal{P}$={handicapés mentaux}
X= résultat à un test de dextérité manuelle, variable quantitative de moyenne μ et d'écart-type σ inconnus dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=32

La loi de X étant quelconque et n=32≥30, $\overline{X}_n$ suit approximativement une loi normale $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Le résultat moyen μ inconnu est estimé par la moyenne observée $\overline{x} = \frac{2272}{32} = 71$ et l'écart-type du résultat σ inconnu est estimé par l'écart-type observé sans biais s* où $s^{*2} = \frac{161664 - 32 \times 71^2}{31} = 11{,}35$ et $s^* = \sqrt{11{,}35} = 3{,}37$

(autre calcul : $s^2 = \frac{161664}{32} - 71^2 = 11$ et $s^{*2} = \frac{32}{31} \times s^2 = 1{,}0323 \times 11 = 11{,}35$)

L'intervalle de confiance à 99% (au risque α=1%) du résultat moyen μ dans $\mathcal{P}$ s'écrit :

$IC_{99\%}(\mu) \approx \left[71 \pm z_{0,995} \frac{3{,}37}{\sqrt{32}}\right] \approx [71 \pm 2{,}575 \times 0{,}596] \approx [71 \pm 1{,}5] = [69{,}5 ; 72{,}5]$ où $z_{1-(\alpha/2)} = z_{0,995} = 2{,}575$ est le quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$.

➨ l'estimation par intervalle de confiance au niveau 99% du résultat moyen des handicapés mentaux est d'environ 69,5 à 72,5 ; la précision (ou marge d'erreur) de l'estimation à 99% est d'environ 1,5.

Exercice 7

$\mathcal{P}$={nouveaux-nés prématurés (nés avant 30 semaines de gestation)}
X= score d'Apgar à 5 mn, variable quantitative de moyenne μ et d'écart-type σ inconnus dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=70

1) L'estimation ponctuelle de μ est donnée par la moyenne observée $\overline{x} = \frac{567}{70} = 8{,}1$

➨ le score d'Apgar moyen des nouveaux-nés prématurés est estimé à 8,1.

2) L'estimation ponctuelle sans biais de la variance σ² est donnée par la variance observée sans biais $s^{*2} = \frac{4817 - 70 \times 8{,}1^2}{69} = \frac{224{,}3}{69} = 3{,}25$

l'estimation ponctuelle sans biais de l'écart-type σ est donnée par l'écart-type observé sans biais $s^* = \sqrt{3{,}25} = 1{,}8$

➨ la variance du score d'Apgar des nouveaux-nés prématurés est estimée à 3,25 et son écart-type à 1,8.

3) L'estimation par intervalle de confiance au niveau 90% (au risque α=10%) de μ dans $\mathcal{P}$ s'écrit :

$IC_{90\%}(\mu) \approx \left[8{,}1 \pm z_{0,95} \frac{1{,}8}{\sqrt{70}}\right] \approx [8{,}1 \pm 1{,}645 \times 0{,}215] \approx [8{,}1 \pm 0{,}35] \approx [7{,}75;8{,}45]$ où $z_{1-(\alpha/2)}=z_{0,95}=1{,}645$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée car n=70≥30 donc la moyenne empirique $\overline{X}_n$ suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 90% du score moyen des nouveaux-nés prématurés est d'environ 7,75 à 8,45 (points de score) ; la précision (ou marge d'erreur) de l'estimation à 90% est d'environ 0,35 (point de score).

Exercice 8

$\mathcal{P}$={enfants atteints d'otite moyenne avec écoulement (OME) bilatérale chronique}
X= score de Reynell, variable quantitative de moyenne μ et d'écart-type σ inconnus dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=77

1) L'estimation ponctuelle de μ est donnée par la moyenne observée $\overline{x} = \frac{-27}{77} = -0{,}35$

➨ le score de Reynell moyen des enfants atteints d'OME bilatérale chronique est estimé à –0,35.

2) L'estimation ponctuelle sans biais de la variance σ² est donnée par la variance observée sans biais $s^{*2} = \frac{86 - 77 \times (-0{,}35)^2}{76} = \frac{76{,}53}{76} = 1{,}007 \approx 1$

l'estimation ponctuelle sans biais de l'écart-type σ est donnée par l'écart-type observé sans biais $s^* = \sqrt{1} = 1$

➨ la variance du score de Reynell des enfants atteints d'OME bilatérale chronique est estimée à 1 et son écart-type à 1.

3) L'estimation par intervalle de confiance au niveau 95% (au risque α=5%) de μ dans $\mathcal{P}$ s'écrit :

$IC_{95\%}(\mu) \approx \left[-0{,}35 \pm z_{0,975} \frac{1}{\sqrt{77}}\right] \approx \left[-0{,}35 \pm 1{,}96 \times \frac{1}{8{,}775}\right] \approx [-0{,}35 \pm 1{,}96 \times 0{,}114] \approx [-0{,}35 \pm 0{,}22] \approx [-0{,}57;-0{,}13]$ où

$z_{1-(\alpha/2)} = z_{0,975}=1{,}96$ est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée car n=77≥30 donc la moyenne empirique $\overline{X}_n$ suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 95% du score de Reynell moyen des enfants atteints d'OME bilatérale chronique est d'environ –0,57 à –0,13 (points de score) ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 0,22 (point de score).

4) La marge d'erreur dans l'estimation du score moyen à 95%, donnée par la demi-longueur de l'intervalle $IC_{95\%}(\mu)$, est de 0,22 pour un échantillon de taille n=77 ; pour obtenir une marge d'erreur (demi-longueur) plus faible, de 0,1, il faudra donc plus d'enfants. Pour n inconnu, s*=1 et α=5% connus, la demi-longueur de l'intervalle $IC_{95\%}(\mu)$ s'écrit : $z_{0,975} \frac{s^*}{\sqrt{n}} = 1{,}96 \frac{1}{\sqrt{n}}$

on cherche n tel que : $1{,}96 \frac{1}{\sqrt{n}} \leq 0{,}1$ c'est à dire $\frac{1{,}96}{0{,}1} \leq \sqrt{n}$ d'où $n \geq \left(\frac{1{,}96}{0{,}1}\right)^2 = 384{,}16$

➨ on choisira donc une taille d'échantillon au moins égale à 385 pour que la marge d'erreur dans l'estimation du score moyen à 95% soit inférieure à 0,1.

Exercice 9

$\mathcal{P}$={électeurs}

1) X= intention de vote pour le candidat A, variable qualitative dichotomique : oui, non
p = proportion d'intentions de vote pour le candidat A dans $\mathcal{P}$, p inconnue dans $\mathcal{P}$
Echantillon de taille n=100 de X issu de $\mathcal{P}$ dont 52 intentions de vote pour le candidat A

La proportion p d'intentions de vote pour le candidat A dans $\mathcal{P}$ est estimée par la fréquence observée $f = \frac{52}{140} = 0,52 = 52\%$

➨ la proportion d'intentions de vote pour le candidat A chez les électeurs est estimée à 52%.

2) L'estimation par intervalle de confiance à 95% (au risque α=5%) de p dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p) \approx \left[0,52 \pm z_{0,975}\sqrt{\frac{0,52\times(1-0,52)}{100}}\right] \approx [0,52 \pm 1,96\times 0,05] \approx [0,52 \pm 0,0979] \approx [52\% \pm 9,8\%] = [0,422\,;\,0,618]$$

où $z_{1-(\alpha/2)}=z_{0,975}=1,96$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=100≥30 et n×(1−0,618)=n×0,382=38,2≥5 donc n×0,422≥5, n×(1−0,422)≥5 et n×0,618≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 95% de la proportion d'intentions de vote pour le candidat A chez les électeurs est de 42,2% à 61,8% environ ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 9,8%.

3) Echantillon de taille n=1 000 de X issu de $\mathcal{P}$ avec une fréquence (proportion) observée f =52% (520 intentions de vote pour le candidat A) : la proportion p d'intentions de vote pour le candidat A chez les électeurs est encore estimée à 52%

L'estimation par intervalle de confiance à 95% (au risque α=5%) de p dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p) \approx \left[0,52 \pm z_{0,975}\sqrt{\frac{0,52\times(1-0,52)}{1000}}\right] \approx [0,52 \pm 1,96\times 0,0158] \approx [0,52 \pm 0,031] \approx [52\% \pm 3,1\%] = [0,489\,;\,0,551]$$

où $z_{1-(\alpha/2)}=z_{0,975}=1,96$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=1 000≥30 et n×(1−0,551)=n×0,449=449≥5 donc n×0,489≥5, n×(1−0,489)≥5 et n×0,551≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 95% de la proportion d'intentions de vote pour le candidat A chez les électeurs est de 48,9% à 55,1% environ ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 3,1% (la précision augmente avec la taille de l'échantillon).

4) La précision de l'intervalle de confiance à 95% est donnée par sa demi-longueur. Pour une taille d'échantillon de n=100, la demi-longueur de l'intervalle $IC_{95\%}(p)$ est de 9,8% (cf question 2) et pour une taille d'échantillon de n=1 000, la demi-longueur de l'intervalle $IC_{95\%}(p)$ est de 3,1% (cf question 3); pour obtenir une demi-longueur plus faible, de 0,5%, il faudra donc plus de 1 000 sélecteurs. Pour n inconnu, f=0,52 et α=5% connus, la demi-longueur de l'intervalle $IC_{95\%}(p)$ s'écrit : $z_{0,975}\sqrt{\frac{f(1-f)}{n}} = 1,96\sqrt{\frac{0,52(1-0,48)}{n}}$. On cherche n tel que : $1,96\sqrt{\frac{0,52\times 0,48}{n}} \le 0,005$ c'est à dire $\frac{1,96}{0,005}\sqrt{0,52\times 0,48} \le \sqrt{n}$ d'où $n \ge \left(\frac{1,96}{0,005}\right)^2 \times 0,52\times 0,48 = 38\,354,5$

➨ on choisira donc une taille d'échantillon au moins égale à 38 355 pour que la demi-longueur de l'intervalle de confiance à 95% (précision à 95%) soit inférieure à 0,5%.

Exercice 10

$\mathcal{P}$={patients lombalgiques}

1) X= sexe féminin, variable qualitative dichotomique : oui, non
p = proportion de femmes dans $\mathcal{P}$, p inconnue dans $\mathcal{P}$
Echantillon de taille n=262 de X issu de $\mathcal{P}$ dont 136 femmes

La proportion p de femmes dans $\mathcal{P}$ est estimée par la fréquence observée $f = \frac{136}{262} \approx 0,519 = 51,9\%$

➨ la proportion de femmes chez les patients lombalgiques est estimée à 51,9%.

2) L'estimation par intervalle de confiance à 95% (au risque α=5%) de p dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p) \approx \left[0,519 \pm z_{0,975}\sqrt{\frac{0,519\times(1-0,519)}{262}}\right] \approx [0,519 \pm 1,96\times 0,031] \approx [0,519 \pm 0,061] \approx [0,458\,;\,0,58]$$

où $z_{1-(\alpha/2)}=z_{0,975}=1,96$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=262≥30 et n×(1−0,58)=n×0,42=110,04≥5 donc n×0,458≥5, n×(1−0,458)≥5 et n×0,58≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 95% de la proportion de femmes chez les patients lombalgiques est de 45,8% à 58% environ ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 6,1%.

3) X= trouble psychologique, variable qualitative dichotomique : oui, non

p = proportion de troubles psychologiques dans $\mathcal{P}$, p inconnue dans $\mathcal{P}$

Echantillon de taille n=262 de X issu de $\mathcal{P}$ dont 99 présentant un trouble psychologique

La proportion (fréquence) des troubles psychologiques p dans $\mathcal{P}$ est estimée par la fréquence observée $f = \dfrac{99}{262} \approx 0{,}378 = 37{,}8\%$

➨ la proportion (fréquence) des troubles psychologiques chez les patients lombalgiques est estimée à 37,8%.

4) L'estimation par intervalle de confiance à 99% (au risque α=1%) de p dans $\mathcal{P}$ s'écrit :

$$IC_{99\%}(p) \approx \left[0{,}378 \pm z_{0{,}995}\sqrt{\frac{0{,}378 \times (1-0{,}378)}{262}}\right] \approx [0{,}378 \pm 2{,}575 \times 0{,}03] \approx [0{,}378 \pm 0{,}077] \approx [0{,}3\,;\,0{,}455]$$

où $z_{1-(\alpha/2)}=z_{0,995}=2{,}575$ est le quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$

cette approximation est justifiée puisque n=262≥30 et n×0,3=262×0,3=78,6≥5 donc n×0,455≥5, n×(1−0,3)≥5 et n×(1−0,455)≥5, donc la fréquence empirique F_n suit approximativement une loi normale

➨ l'estimation par intervalle de confiance au niveau 99% de la proportion des troubles psychologiques chez les patients lombalgiques est de 30% à 45,5% environ ; la précision (ou marge d'erreur) de l'estimation à 99% est d'environ 7,7%.

Exercice 11

$\mathcal{P}$={patients souffrant de dépression résistante traités par antidépresseur pendant 28 jours}

X= efficacité du traitement antidépresseur, variable qualitative dichotomique : oui, non

p = proportion d'efficacité dans $\mathcal{P}$, p inconnue dans $\mathcal{P}$

Echantillon de taille n=1 197 de X issu de $\mathcal{P}$

1) L'estimation ponctuelle de la proportion d'efficacité du traitement p est donnée par la fréquence observée $f = \dfrac{957}{1\,197} \approx 0{,}799 = 79{,}9\%$

➨ la proportion d'efficacité du traitement antidépressif chez les patients souffrant de dépression résistante est estimée à 79,9%.

2) L'estimation par intervalle de confiance au niveau 90% (au risque α=10%) de p dans $\mathcal{P}$ s'écrit :

$$IC_{90\%}(p) \approx \left[0{,}799 \pm z_{0{,}95}\sqrt{\frac{0{,}799 \times (1-0{,}799)}{1197}}\right] \approx [0{,}799 \pm 1{,}645 \times 0{,}0116] \approx [0{,}799 \pm 0{,}019] \approx [0{,}78\,;\,0{,}818]$$

où $z_{1-(\alpha/2)}=z_{0,95}=1{,}645$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=1197≥30 et n×(1−0,818)=1197×0,182=217,9≥5 donc n×0,78≥5, n×0,818≥5 et n×(1−0,78)= n×0,22≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 90% de la proportion d'efficacité du traitement chez les patients souffrant de dépression résistante traités par antidépresseur pendant 28 jours est d'environ 78% à 81,8% ; la précision (ou marge d'erreur) de l'estimation à 90% est d'environ 1,9%.

3) La demi-longueur de l'intervalle précédent $IC_{90\%}(p)$ est de 0,019=1,9% donc inférieure à 2% : donc la taille de l'échantillon observé n=1197 convient pour obtenir une marge d'erreur plus grande; pour obtenir une demi-longueur de 2% exactement, il faudrait diminuer la taille de l'échantillon. Pour n inconnu, f=0,799 et α=10% connus, la demi-longueur de l'intervalle $IC_{90\%}(p)$ s'écrit : $z_{0,95}\sqrt{\dfrac{f(1-f)}{n}} = 1{,}645\sqrt{\dfrac{0{,}799(1-0{,}799)}{n}}$. On cherche n tel que : $1{,}645\sqrt{\dfrac{0{,}799 \times 0{,}201}{n}} \le 0{,}02$ c'est à dire $\dfrac{1{,}645}{0{,}02}\sqrt{0{,}799 \times 0{,}201} \le \sqrt{n}$ d'où $n \ge \left(\dfrac{1{,}645}{0{,}02}\right)^2 \times 0{,}799 \times 0{,}201 = 1086{,}46$

➨ on choisirait donc une taille d'échantillon au moins égale à 1 087 pour que la demi-longueur de l'intervalle de confiance à 90% soit inférieure à 2%.

Exercice 12

$\mathcal{P}$={personnes de plus de 66 ans consommateurs de TCA}

X= survenue d'une fracture de hanche, variable qualitative dichotomique : oui, non

p = proportion de fracture de hanche dans $\mathcal{P}$ = fréquence de fracture de hanche dans $\mathcal{P}$, p inconnue dans $\mathcal{P}$

Echantillon de taille n=5 838 de X issu de $\mathcal{P}$

1) L'estimation ponctuelle de p est donnée par la fréquence observée $f = \frac{1\,493}{5\,838} = 0{,}2557 \approx 0{,}256 = 25{,}6\%$

➨ la fréquence des fractures de hanche chez les personnes de plus de 66 ans consommateurs de TCA est estimée à 25,6%.

2) L'estimation par intervalle de confiance à 99% (au risque α=1%) de p dans $\mathcal{P}$ s'écrit :

$$IC_{99\%}(p) \approx \left[0{,}256 \pm z_{0{,}995}\sqrt{\frac{0{,}256 \times (1-0{,}256)}{5838}}\right] \approx [0{,}256 \pm 2{,}575 \times 0{,}006] \approx [0{,}256 \pm 0{,}015] \approx [0{,}241;0{,}271]$$

où $z_{1-(\alpha/2)}=z_{0,995}=2{,}575$ est le quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=5838≥30 et n×0,241=5838×0,241=1407≥5 donc n×0,271≥5, n×(1−0,241)≥5 et n×(1−0,271)≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 99% de la fréquence des fractures de hanche chez les personnes de plus de 66 ans consommateurs de TCA est de 24,1% à 27,1% environ ; la précision (ou marge d'erreur) de l'estimation à 99% est d'environ 1,5%.

Exercice 13

$\mathcal{P}$ ={adolescents français de 12 à 20 ans} de taille N inconnue
X= consommation de psychotropes, variable qualitative dichotomique : oui, non
p = proportion de consommateurs de psychotropes dans $\mathcal{P}$
= fréquence de consommation de psychotropes dans $\mathcal{P}$ p inconnue dans $\mathcal{P}$

Echantillon de taille n=3 279 de X issu de $\mathcal{P}$:

1) L'estimation ponctuelle de p est donnée par la fréquence observée $f = \frac{689}{3\,279} = 0{,}21 = 21\%$

➨ la proportion d'adolescents français de 12 à 20 ans consommateurs de psychotropes est estimée à 21%.

2) L'estimation par intervalle de confiance à 95% (au risque α=5%) de p dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p) \approx \left[0{,}21 \pm z_{0{,}975}\sqrt{\frac{0{,}21\times(1-0{,}21)}{3279}}\right] \approx [0{,}21 \pm 1{,}96 \times 0{,}007] \approx [0{,}21 \pm 0{,}014] \approx [0{,}196;0{,}224]$$

où $z_{1-(\alpha/2)}=z_{0,975}=1{,}96$ est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=3279≥30 et n×0,196=3279×0,196=642,7≥5 donc n×0,224≥5, n×(1−0,196)≥5 et n×(1−0,224)≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 95% de la proportion d'adolescents français de 12 à 20 ans consommateurs de psychotropes est de 19,6% à 22,4% environ ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 1,4%.

3) • $\mathcal{P}_1$={adolescents français de 12 à 20 ans, de sexe féminin} de taille inconnue

X= consommation de psychotropes, variable qualitative dichotomique : oui, non

p_1 = proportion de consommatrices de psychotropes dans $\mathcal{P}_1$ p_1 inconnue dans $\mathcal{P}_1$

Echantillon de taille n_1=1 728 de X issu de $\mathcal{P}_1$:

l'estimation ponctuelle de p_1 est donnée par la fréquence observée $f_1 = \frac{475}{1728} = 0{,}275 = 27{,}5\%$

L'estimation par intervalle de confiance à 95% (au risque α=5%) de p_1 dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p_1) \approx \left[0{,}275 \pm z_{0{,}975}\sqrt{\frac{0{,}275\times(1-0{,}275)}{1728}}\right] \approx [0{,}275 \pm 1{,}96 \times 0{,}011] \approx [0{,}275 \pm 0{,}021] \approx [0{,}254;0{,}296]$$

où $z_{1-(\alpha/2)}=z_{0,975}=1{,}96$ est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=1728≥30 et n×0,254=1728×0,254=438,9≥5 donc n×0,296≥5, n×(1−0,254)≥5 et n×(1−0,296)≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 95% de la proportion d'adolescentes françaises de 12 à 20 ans consommatrices de psychotropes est de 25,4% à 29,6% environ ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 2,1%.

• $\mathcal{P}_2$={adolescents français de 12 à 20 ans, de sexe masculin} de taille inconnue

X= consommation de psychotropes, variable qualitative dichotomique : oui, non

p_2 = proportion de consommateurs de psychotropes dans $\mathcal{P}_2$ p_2 inconnue dans $\mathcal{P}_2$

Echantillon de taille n_2=3 279−1 728=1 551 de X issu de $\mathcal{P}_2$:

l'estimation ponctuelle de p_2 est donnée par la fréquence observée $f_2 = \frac{689-475}{1\,551} = \frac{214}{1\,551} = 0{,}138 = 13{,}8\%$

L'estimation par intervalle de confiance à 95% (au risque α=5%) de p_2 dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p_2) \approx \left[0{,}138 \pm z_{0{,}975}\sqrt{\frac{0{,}138\times(1-0{,}138)}{1551}}\right] \approx [0{,}138 \pm 1{,}96\times 0{,}0088] \approx [0{,}275 \pm 0{,}017] \approx [0{,}121;0{,}155]$$

où $z_{1-(\alpha/2)}=z_{0,975}$=1,96 est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=1551≥30 et n×0,121=1551×0,121=187,7≥5 donc n×0,155≥5, n×(1−0,121)≥5 et n×(1−0,155)≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 95% de la proportion d'adolescents français masculins de 12 à 20 ans consommateurs de psychotropes est de 12,1% à 15,5% environ ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 1,7%.

4) La précision de l'intervalle de confiance à 95% est donnée par sa demi-longueur. Pour une taille d'échantillon de n=3 279, la demi-longueur de l'intervalle $IC_{95\%}$(p) est de 1,4% (cf question 2); pour obtenir une demi-longueur plus faible, de 1%, il faudra donc plus de 3 279 sujets. Pour n inconnu, f=0,21 et α=5% connus, la demi-longueur de l'intervalle $IC_{95\%}$(p) s'écrit : $z_{0,975}\sqrt{\frac{f(1-f)}{n}} = 1{,}96\sqrt{\frac{0{,}21(1-0{,}21)}{n}}$. On cherche n tel que : $1{,}96\sqrt{\frac{0{,}21\times 0{,}79}{n}} \le 0{,}01$ c'est à dire $\frac{1{,}96}{0{,}01}\sqrt{0{,}21\times 0{,}79} \le \sqrt{n}$ d'où $n \ge \left(\frac{1{,}96}{0{,}01}\right)^2 \times 0{,}21\times 0{,}79 = 6373{,}2$

➨ on choisira donc une taille d'échantillon au moins égale à 6 374 pour que la demi-longueur de l'intervalle de confiance à 95% soit inférieure à 1%.

Exercice 14

$\mathcal{P}$={femmes norvégiennes âgées de 20 à 49 ans}
X= présence d'une dépression , variable qualitative dichotomique : oui, non
p = proportion de dépression dans $\mathcal{P}$ = fréquence de dépression dans $\mathcal{P}$, p inconnue dans $\mathcal{P}$
Echantillon de taille n=3 103 de X issu de $\mathcal{P}$

1) L'estimation ponctuelle de p est donnée par la fréquence observée $f = \frac{571}{3\,103} = 0{,}184 = 18{,}4\%$

➨ la proportion de femmes norvégiennes âgées de 20 à 49 ans dépressives est estimée à 18,4%.

2) L'estimation par intervalle de confiance à 95% (au risque α=5%) de p dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p) \approx \left[0{,}184 \pm z_{0{,}975}\sqrt{\frac{0{,}184\times(1-0{,}184)}{3103}}\right] \approx [0{,}184 \pm 1{,}96\times 0{,}007] \approx [0{,}184 \pm 0{,}014] \approx [0{,}170;0{,}198]$$

où $z_{1-(\alpha/2)}=z_{0,975}$=1,96 est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$.

Cette approximation est justifiée puisque n=3103≥30 et n×0,17=3103×0,17=527,51≥5 donc n×0,198≥5, n×(1−0,17)≥5 et n×(1−0,198)≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

➨ l'estimation par intervalle de confiance au niveau 95% de la fréquence de la dépression chez les femmes norvégiennes âgées de 20 à 49 ans est de 17% à 19,8% environ ; la précision (ou marge d'erreur) de l'estimation à 95% est d'environ 1,4%.

3) La demi-longueur de l'intervalle précédent $IC_{95\%}$(p) est de 1,4% ; pour obtenir une demi-longueur plus faible, de 1%, il faudra donc plus de 3103 sujets. Pour n inconnu, f=0,184 et α=5% connus, la demi-longueur de l'intervalle $IC_{95\%}$(p) s'écrit : $z_{0,975}\sqrt{\frac{f(1-f)}{n}} = 1{,}96\sqrt{\frac{0{,}184(1-0{,}184)}{n}}$. On cherche n tel que : $1{,}96\sqrt{\frac{0{,}184\times 0{,}816}{n}} \le 0{,}01$ c'est à dire $\frac{1{,}96}{0{,}01}\sqrt{0{,}184\times 0{,}816} \le \sqrt{n}$ d'où $n \ge \left(\frac{1{,}96}{0{,}01}\right)^2 \times 0{,}184\times 0{,}816 = 5767{,}93$

➨ on choisira donc une taille d'échantillon au moins égale à 5 768 pour que la demi-longueur de l'intervalle de confiance à 95% soit inférieure à 1%.

Exercice 15

$\mathcal{P}$={personnes ayant des rides}

1) • X= diminution de la profondeur des rides après un mois, variable qualitative dichotomique : oui, non

p = proportion de diminution de la profondeur des rides après un mois dans $\mathcal{P}$, p inconnue dans $\mathcal{P}$

Echantillon de taille n=30 de X issu de $\mathcal{P}$ pour lequel la fréquence observée f =93% donc l'effectif observé de volontaires ayant une diminution de la profondeur des rides après un mois est : n×f=30×0,93=27,9 donc 28 volontaires et $\frac{28}{30} = 0{,}9333 \approx 93\%$

• X= amélioration du micro relief cutané après trois mois, variable qualitative dichotomique : oui, non

p = proportion d'amélioration du micro relief cutané après trois mois dans $\mathcal{P}$, p inconnue dans $\mathcal{P}$

Echantillon de taille n=30 de X issu de $\mathcal{P}$ pour lequel la fréquence observée f =84% donc l'effectif observé de volontaires ayant une amélioration du micro relief cutané après trois mois est : n×f=30×0,84=25,2 donc 25 volontaires or $\frac{25}{30} = 0{,}8333 \approx 83\%$ et $\frac{26}{30} = 0{,}8667 \approx 87\%$ mais pas 84%

2) L'estimation par intervalle de confiance à 95% (au risque α=5%) de la proportion de diminution de la profondeur des rides après un mois, p dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p) \approx \left[0{,}93 \pm z_{0{,}975}\sqrt{\frac{0{,}93\times(1-0{,}93)}{30}}\right] \approx [0{,}93 \pm 1{,}96\times 0{,}0466] \approx [0{,}93 \pm 0{,}091] \approx [93\% \pm 9{,}1\%] = [0{,}839\,;\mathbf{1{,}021}]$$

où $z_{1-(\alpha/2)}=z_{0{,}975}=1{,}96$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

Mais cette approximation n'est pas justifiée même si n=30≥30 puisque la borne supérieure de l'intervalle dépasse 1 on ne peut donc pas vérifier les autres conditions pour cette borne (et n×(1−0,839)=n×0,161=4,83 <5), donc la fréquence empirique F_n **ne** suit **pas** approximativement une loi normale.

➨ il faut utiliser la vraie distribution de la fréquence empirique F_n, la loi binomiale, pour calculer les intervalles de confiance de p

3) L'estimation par intervalle de confiance à 95% (au risque α=5%) de la proportion d'amélioration du micro relief cutané après trois mois, p dans $\mathcal{P}$ s'écrit :

$$IC_{95\%}(p) \approx \left[0{,}83 \pm z_{0{,}975}\sqrt{\frac{0{,}83\times(1-0{,}83)}{30}}\right] \approx [0{,}83 \pm 1{,}96\times 0{,}0686] \approx [0{,}83 \pm 0{,}134] \approx [83\% \pm 13{,}4\%] = [0{,}696\,;0{,}964]$$

où $z_{1-(\alpha/2)}=z_{0{,}975}=1{,}96$ est le quantile d'ordre 0,95 de la loi $\mathcal{N}(0,1)$.

Mais cette approximation n'est pas justifiée même si n=30≥30 puisque la condition n×(1−0,964)=n×0,036=1,08 <5 n'est pas vérifiée c'est à dire que la fréquence empirique F_n **ne** suit **pas** approximativement une loi normale.

➨ il faut utiliser la vraie distribution de la fréquence empirique F_n , la loi binomiale, pour calculer les intervalles de confiance de p

Exercice 16

$\mathcal{P}$ ={personnes de plus de 18 ans ayant arrêté de fumer}
X= prise de poids, variable quantitative de moyenne μ et d'écart-type σ inconnus dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=38 sur lequel on observe $\bar{x} = 1{,}5$ kg et s*=3 kg

1) L'estimation ponctuelle de la moyenne inconnue μ est donnée par $\bar{x} = 1{,}5$.

- les réponses *a* et *b* sont fausses : la prise de poids moyenne dans $\mathcal{P}$ est égale à μ inconnue
- les réponses *c* et *f* sont justes
- la réponse *d* est fausse : F_n est la variable aléatoire représentant l'ensemble des valeurs de la fréquence observée *f* sur tous les échantillons de taille n de X issus de $\mathcal{P}$, utilisée pour estimer une proportion pour une variable qualitative
- la réponse *e* est fausse : $\overline{X}_n$ est la variable aléatoire représentant l'ensemble des valeurs de la moyenne observée $\bar{x}$ sur tous les échantillons de taille n de X issus de $\mathcal{P}$
- la réponse *g* est fausse : p désigne un paramètre représentant une proportion dans $\mathcal{P}$

2) - l'assertion *a* est fausse
- l'assertion *b* est vraie
- les assertions *c* et *d* sont vraies : un intervalle de confiance quelconque (parmi tous les intervalles de confiance possibles d'un niveau donné) de la prise de poids moyenne contient la vraie valeur de la moyenne, prise de poids

moyenne dans $\mathcal{P}$ égale à μ, avec une probabilité donnée, égale à $1-\alpha$ pour un intervalle de confiance au niveau de confiance $1-\alpha$, égale à 95% pour un intervalle de confiance au niveau de confiance 95%

- les assertions *e* et *g* sont vraies : pour $n \geq 30$ $IC_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{s^*}{\sqrt{n}}\right]$ où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$; plus le niveau de confiance est élevé plus $1-\frac{\alpha}{2}$ est élevé plus $z_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1-\frac{\alpha}{2}$ l'est, donc plus l'amplitude de l'intervalle $2 \times z_{1-\frac{\alpha}{2}} \frac{s^*}{\sqrt{n}}$ est grande ;
  plus la variance observée s* est élevée, plus l'amplitude de l'intervalle est grande
- l'assertion *f* est fausse : plus la taille de l'échantillon n est élevée plus $\sqrt{n}$ l'est, plus $\frac{1}{\sqrt{n}}$ est faible donc plus l'amplitude de l'intervalle (la marge d'erreur) est petite (la précision augmente avec la taille de l'échantillon)
- l'assertion *h* est fausse : l'intervalle de confiance est centré sur la valeur observée de la moyenne $\bar{x}$, contrairement à l'intervalle de variation qui est centré sur la vraie valeur de la moyenne μ
- l'assertion *i* est vraie : environ 95 intervalles de confiance sur 100 (95%) calculés à partir de 100 échantillons de taille n de la variable étudiée issus de la population, contiennent la vraie valeur de la moyenne μ

3) L'estimation par intervalle de confiance au niveau 95% (au risque α=5%) de la prise de poids moyenne μ dans $\mathcal{P}$ s'écrit :

$IC_{95\%}(\mu) = \left[1{,}5 \pm z_{0{,}975} \frac{3}{\sqrt{38}}\right] = [1{,}5 \pm 1{,}96 \times 0{,}487] \approx [1{,}5 \pm 0{,}95] \approx [1{,}5 \pm 1] = [0{,}5\,;\,2{,}5]$ où $z_{1-(\alpha/2)} = z_{0{,}975} = 1{,}96$ est le quantile d'ordre 0,975 de la loi $\mathcal{N}(0,1)$. Cette approximation est justifiée car n=38≥30 donc la moyenne empirique $\overline{X}_n$ suit approximativement une loi normale.

- les réponses *a* et *c* sont justes
- les réponses *b* et *f* sont fausses : elles correspondent au calcul de l'intervalle de confiance à 90% ; de plus dans la réponse b "dans l'échantillon observé" ne veut rien dire
- la réponse *d* est fausse : il n'est pas fait mention du risque d'erreur ou de la confiance que l'on peut associer à cet intervalle
- la réponse *e* est fausse : c'est la définition de l'intervalle de variation à 95% de la prise de poids moyenne
- la réponse *g* est juste
- la réponse *h* est fausse : elle associe à l'intervalle de confiance une probabilité de 95% de contenir la moyenne observée $\bar{x}$ qui est toujours le milieu de l'intervalle de confiance (100% de chances de lui appartenir)
- les réponses *i* et *j* résultent d'un abus de langage : en effet, une fois que l'intervalle de confiance a été calculé, ses bornes sont fixes (non aléatoires). Puisque ni la vraie moyenne μ ni l'intervalle [0,5 ; 2,5] ne sont aléatoires, soit l'intervalle [0,5 ; 2,5] contient la vraie moyenne μ, soit il ne la contient pas, ce que l'on ignore puisque μ est inconnue; en toute rigueur on ne peut plus associer à cet intervalle une probabilité de 95% de contenir, ou 5% de ne pas contenir, la vraie moyenne μ. Ces formulations sont cependant souvent utilisées et tolérées, la réponse *j* étant toutefois mieux formulée.
- les réponses *k* et *l* sont fausses : la probabilité 95% porte sur les personnes (individus) de $\mathcal{P}$, ce qui serait un intervalle de variation à 95% de la variable prise de poids, et non pas de la prise de poids moyenne.

Exercice 17

$\mathcal{P}$ ={patients dépressifs âgés de 18 à 65 ans traités par psychothérapie pendant 16 semaines}
X= guérison, variable qualitative, proportion de guérison p inconnue dans $\mathcal{P}$
Echantillon de X issu de $\mathcal{P}$ de taille n=55 sur lequel on observe f = 36/55 = 0,655=65,5%

1) L'estimation ponctuelle de la proportion de guérison inconnue p est donnée par f = 0,655=65,5%.
   - la réponse *a* est fausse : F_n est la variable aléatoire représentant l'ensemble des valeurs de la fréquence observée f sur tous les échantillons de taille n de X issus de $\mathcal{P}$
   - les réponses *b* et *c* sont fausses : la proportion de guérison dans $\mathcal{P}$ n'est pas égale à f mais est égale à p inconnue
   - la réponse *d* est juste
   - la réponse *e* est fausse : $\bar{x}$ représente une moyenne observée utilisée pour estimer la moyenne d'une variable quantitative

2) - l'assertion *a* est vraie
   - les assertions *b* et *c* sont vraies : l'intervalle de confiance aléatoire (au risque 1% ou au niveau de confiance 99%) de la proportion de guérison contient la vraie valeur de la proportion de guérison dans $\mathcal{P}$ égale à p, avec une certaine probabilité (de 99%)
   - l'assertion *d* est fausse : la vraie valeur de la proportion de guérison dans $\mathcal{P}$ égale à p étant une valeur fixe (non aléatoire, déterministe) la probabilité pour qu'elle appartienne à un intervalle est donc soit nulle, si elle n'appartient pas à l'intervalle, soit égale à 1 si elle lui appartient

- les assertions *e* et *g* sont vraies : l'intervalle de confiance au risque 1% ou au niveau de confiance 99% s'écrit $IC_{99\%}(p) = \left[f \pm z_{0,995}\sqrt{\frac{f(1-f)}{n}} \right] = \left[f \pm 2{,}575\sqrt{\frac{f(1-f)}{n}} \right]$ où $z_{1-(\alpha/2)} = z_{0,995} = 2{,}575$ est le quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$
  cet intervalle est centré sur *f* la fréquence (proportion) observée de guérison ;
  plus la taille de l'échantillon n est élevée plus $\sqrt{n}$ l'est, plus $\frac{1}{\sqrt{n}}$ est faible donc plus l'amplitude de l'intervalle (la marge d'erreur) est petite (la précision augmente avec la taille de l'échantillon)
- l'assertion *f* est vraie : en effet, pour $n \geq 30$
  $IC_{1-\alpha}(p) = \left[f \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} \right]$ où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$; plus le risque α est élevé plus $1-\frac{\alpha}{2}$ est faible, plus $z_{1-\frac{\alpha}{2}}$ l'est, donc plus l'amplitude de l'intervalle $2 \times z_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}$ est faible
- l'assertion *h* est vraie : environ un intervalle de confiance sur 100 (99%) calculés à partir de 100 échantillons de taille n=55 de la variable étudiée issus de la population, ne contient pas la vraie valeur de la proportion de guérison

3) L'estimation par intervalle de confiance au niveau 99% (au risque α=1%) de la proportion de guérison p dans $\mathcal{P}$ s'écrit :

$$IC_{99\%}(p) \approx \left[0{,}655 \pm z_{0,995}\sqrt{\frac{0{,}655 \times 0{,}345}{55}} \right] \approx [0{,}655 \pm 2{,}575 \times 0{,}064] \approx [0{,}655 \pm 0{,}165] \approx [0{,}49 ; 0{,}82]$$

où $z_{1-(\alpha/2)} = z_{0,995} = 2{,}575$ est le quantile d'ordre 0,995 de la loi $\mathcal{N}(0,1)$; conditions d'application : cette approximation est justifiée puisque n=55≥30 et n×(1−0,82)=55×0,18=9,9≥5 donc n×0,49≥5, n×(1−0,49)≥5 et n×0,82≥5, donc la fréquence empirique F_n suit approximativement une loi normale.

- la réponse *a* est fausse : il n'est pas fait mention du risque d'erreur ou de la confiance que l'on peut associer à cet intervalle
- les réponses *b*, *g* et *h* sont justes
- la réponse *c* est fausse : la proportion de guérison dans $\mathcal{P}$ égale à p inconnu et la proportion dans l'échantillon observé est égale à *f* = 0,655 et ne varie pas dans l'échantillon
- les réponses *d, f* et *i* sont fausses : elles correspondent au calcul de l'intervalle de confiance à 95%
- la réponse *e* est fausse : c'est la définition de l'intervalle de variation à 99% de la proportion de guérison
- la réponse *j* est fausse : les conditions d'application de la formule approximative de l'intervalle de confiance d'une proportion sont vérifiées

Exercice 18

Définition	Notation
fréquence empirique	F_n
moyenne observée	$\bar{x}$
moyenne empirique	$\bar{X}_n$
moyenne observée sur l'échantillon	$\bar{x}$
taille de l'échantillon	n
fréquence observée sur l'échantillon	f
écart-type observé	s
écart-type observé sans biais	s^*
écart-type empirique	S_n
moyenne dans la population	μ
écart-type dans la population	σ
proportion dans la population	p
intervalle de confiance à 95% de la moyenne	$IC_{95\%}(\mu)$
intervalle de confiance au risque 5% de la proportion	$IC_{95\%}(p)$
intervalle de variation à 95% de la fréquence observée	$I_{95\%}(F_n)$
intervalle de fluctuation au risque 5% de la moyenne observée	$I_{95\%}(\bar{X}_n)$
proportion observée	f
écart-type empirique sans biais	S_n^*