UNIVERSITE DE BATNA_2 **FACULTE DE MEDECINE**

SERIE DE TD N° 8 **DE BIOSTATISTIQUES**

1ère Année Médecine, 2022/2023

TD sur la théorie de l'estimation

Exercice 1

Pour juger la teneur en magnésium d'une eau minérale, on a effectué 10 mesures (en mg pour 10 litres) :

248 246 246 247 247 249 247 250 248 245.

La teneur en magnésium étudiée est supposée être une variable aléatoire normale d'espérance μ et de variance σ^2.

Déterminer un intervalle de confiance sur μ pour un niveau de confiance de 0.95

Solution de l'exercice n° 1

Données de l'échantillon	Données de la population
Taille n = 10 < 30 ⇒ petit échantillon Moyenne de l'échantillon : $\bar{x} = 247.3$ Écart–type de l'échantillon : s = 1.494434118	Population infinie σ Écart–type de la population σ Inconnu, on l'estime par s.

Pour n < 30 la distribution suit la loi normale (condition satisfaite par hypothèse), alors on cherche le critère $t_{\alpha,n-1}$ dans la table statistique n° 3 de Student ⇒ $t_{\alpha,n-1} = t_{0.05,10-1} = 2.262$

$IC_{95\,\%} = [\bar{x} - t_{\alpha,n-1} * \frac{s}{\sqrt{n}}\ ;\ \bar{x} + t_{\alpha,n-1} * \frac{s}{\sqrt{n}}]$. On calcule tout d'abord l'erreur :

$$t_{\alpha,n-1} * \frac{s}{\sqrt{n}} = 2.262 * \frac{1.494434118}{\sqrt{10}} = 1.068979495 \approx 1.069$$

$IC_{95\,\%} = [247.3 - 1.069\ ;\ 247.3 + 1.069] = [246.23\ ;\ 248.37]$

Notez que :

Dans la littérature les critères sont symbolisés par $t_{\alpha,n-1}$ et dans le cours ils sont indiqués par $\overline{t_\alpha}$

Exercice 2

On s'intéresse au temps de mémorisation d'un texte (mesuré en minutes) par les étudiants d'une promotion. Un échantillon de 37 étudiants fournit les valeurs m = 25 et s = 5.

1) Pour un risque α = 5%, construire un intervalle de confiance du temps moyen de mémorisation μ de la promotion.
2) Si la population étudiée est celle d'une promotion de 700 étudiants, refaire la même question.

Solution de l'exercice n° 2

1) P = {les étudiants d'une promotion}

X = Variable aléatoire « temps de mémorisation d'un texte » (variable quantitative mesuré en mn) de moyenne μ et d'écart-type σ inconnus dans P.

Echantillon de X issu de P de taille n = 37 pour lequel $\bar{x}$ = 25 et s = 5.

La loi de X étant quelconque et n = 37 > 30.

Données de l'échantillon	Données de la population
Taille n = 37 > 30 Moyenne de l'échantillon : $\bar{x}$ = 25 Écart–type de l'échantillon : s = 5	Population non finie σ Écart–type de la population σ Inconnu, on l'estime par s.

X suit une loi quelconque : il faut que n > 30

On cherche le critère z_α dans la table 2 de la loi normale ⇒ $z_\alpha = z_{0.05} = 1.96$

$IC_{95\,\%} = [\bar{x} - z_\alpha * \frac{s}{\sqrt{n}} ; \bar{x} + z_\alpha * \frac{s}{\sqrt{n}}]$. On calcule tout d'abord l'erreur :

$z_\alpha * \frac{s}{\sqrt{n}} = 1.96 * \frac{5}{\sqrt{37}} = 1.611110076 \approx 1.61$

$IC_{95\,\%}$ = [25 – 1.61 ; 25 + 1.61] = [23.39 ; 26.61]

2) C'est une population finie qui est une promotion de 700 Etudiants et le tirage est exhaustif donc on ajoute le coefficient d'exhaustivité dans l'intervalle de confiance :

$IC_{95\,\%} = [\bar{x} - z_\alpha * \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} ; \bar{x} + z_\alpha * \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}]$. On calcule tout d'abord l'erreur :

$z_\alpha * \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 1.96 * \frac{5}{\sqrt{37}}\sqrt{\frac{700-37}{700-1}} = 1.56907387 \approx 1.57$

$IC_{95\,\%}$ = [25 – 1.57 ; 25 + 1.57] = [23.43 ; 26.57]

Exercice 3

La santé de 70 nouveau–nés prématurés (nés avant 30 semaines de gestation) a été mesurée par le score d'Apgar à 5 minutes. Les scores obtenus sont tels que :

$$\sum x_i = 567 \qquad \sum x_i^2 = 4\,817.$$

1) Donner une estimation ponctuelle du score d'Apgar moyen pour la population des nouveau–nés prématurés.
2) Donner une estimation ponctuelle de la variance du score d'Apgar et de l'écart–type de ce score pour la population des nouveau–nés prématurés.
3) Estimer par intervalle de confiance au niveau 90%, le score d'Apgar moyen pour la population des nouveau–nés prématurés

Solution de l’exercice n° 3

P = {Nouveau-nés prématurés (nés avant 30 semaines de gestations)}

X = Variable aléatoire quantitative « score d’Apgar à 5 mn » de moyenne μ et d’écart-type σ inconnus dans P.

Echantillon de taille n = 70, c’est un échantillon de X issu de la population P.

1) L’estimation ponctuelle de μ est donnée par la moyenne observée $\bar{x} = \frac{\sum x_i}{n} = \frac{567}{70} = 8.1$
2) L’estimation ponctuelle sans biais de la variance σ^2 est donnée par la variance corrigée S^2.

   La variance empirique de l’échantillon $\sigma^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 = \frac{4817}{70} - 8.1^2 = 3.204285714 \cong \mathbf{3.204}$

   La variance corrigée est $S^2 = 3.204285714 * \frac{n}{n-1} = 3.250724638 \cong \mathbf{3.251}$

   L’écart-type corrigé est $S = \sqrt{3.250724638} = 1.802976605 \cong \mathbf{1.803}$
3) X suit la loi quelconque : il faut que $\boldsymbol{n \geq 30}$

   On cherche z_α dans la table n° 2 de l’écart-réduit de la loi normale $\Rightarrow z_{0.1} = 1.645$

   $IC_{90\,\%} = [\bar{x} - z_\alpha * \frac{s}{\sqrt{n}} ; \bar{x} + z_\alpha * \frac{s}{\sqrt{n}}]$. On calcule tout d’abord l’erreur :

   $z_\alpha * \frac{s}{\sqrt{n}} = 1.645 * \frac{1.802976605}{\sqrt{70}} = 0.354492436 \approx 0.3545$

   $IC_{90\,\%} = [8.1 - 0.3545 ; 8.1 + 0.3545] = [7.7455 ; 8.4545]$

Exercice 4

On a administré un traitement antidépresseur à 1 197 patients souffrant de dépression résistante. Après 28 jours de traitement, le traitement a été efficace pour 957 patients.

1) Estimer ponctuellement la proportion d'efficacité du traitement.
2) Estimer par intervalle de confiance à 90% la proportion d'efficacité du traitement.
3) Quelle devrait être la taille de l'échantillon pour que la demi–longueur de l'intervalle de confiance à 90% soit inférieure à 2%, en supposant que la fréquence observée d'efficacité du traitement reste la même ?

Solution de l’exercice n° 4

P = {Population de patients souffrant de dépression résistante, traités par antidépresseur pendant 28 jours}.

X = Variable aléatoire « Efficacité du traitement antidépresseur », variable qualitative dichotomique : (les issues sont oui et non).

p = proportion d’efficacité dans la population P, la proportion p est inconnue dans P.

Echantillon de taille n = 1197 de X issu de la population P.

1) L’estimation ponctuelle de la proportion d’efficacité du traitement « p » est donnée par la fréquence observée $f = \frac{X}{n} = \frac{957}{1197} = 0.799498746 \approx \mathbf{79.95\ \%}$.
2) L’intervalle de confiance au niveau 90 % de la proportion p dans la population P est :

$$IC_{90\%}(p) = \left[f - z_\alpha\sqrt{\frac{f(1-f)}{n}}\ ;\ f + z_\alpha\sqrt{\frac{f(1-f)}{n}}\right]$$

Et les conditions nécessaires ($n * f > 5$ et $n * (1 - f) > 5$) sont satisfaites.

On calcule tout d'abord l'erreur : $z_\alpha\sqrt{\frac{f(1-f)}{n}} = 1.645\sqrt{\frac{0.7995(1-0.7995)}{1197}} = 0.0190364 \approx \mathbf{0.019}$

$IC_{90\%}(p) = [0.7995 - 0.0190\ ;\ 0.7995 + 0.0190] = \mathbf{[0.7805\ ;\ 0.8185]}$

3) On a changé l'échantillon (P n'est pas connue avant la sélection de l'échantillon, il faut trouver une valeur préalable de P, on utilise une estimation obtenue antérieurement).
Si la demi–longueur de l'intervalle de confiance à 90 % doit être inférieure à 0.02 alors :

L'erreur = $z_\alpha\sqrt{\frac{f(1-f)}{n'}} < 0.02 \Rightarrow \mathbf{n'} > z_{0.1}^2\ \frac{f(1-f)}{(0.02)^2} = 1084.442906 \Rightarrow \mathbf{n' > 1084}$

On choisira la taille de l'échantillon au moins égale à **n' ≥ 1085**

Exercice 5

On sait que les résultats à un test de QI (Quotient Intellectuel de l'Intelligence Humaine) se distribuent normalement avec un écart–type $\sigma = 13$ dans une population d'étudiants. Sur un échantillon de taille 30, on observe $\bar{x} = 111$.

1) Estimer la moyenne μ des résultats avec un intervalle de confiance à 95%, à 90%, puis à 99% et quelle est votre remarque ?

2) Quelle doit être la taille minimum de l'échantillon pour que la demi–longueur de l'intervalle de confiance à 95% n'excède pas 1 (l'unité) ?

Solution de l'exercice n° 5

P = {Population des étudiants}

X = Résultat au test de QI, variable quantitative de moyenne μ inconnue dans P et d'écart-type $\sigma = 13$ connu dans P.

Echantillon de X issu de P de taille $n = 30$ sur lequel on observe $\bar{x} = 111$ qui est l'estimation ponctuelle de la moyenne inconnue μ.

1) X suit la loi normale N (μ ; $\sigma = 13$) et comme σ est connu alors quelle que soit la valeur de n, l'estimation par intervalle de confiance de μ dans P est donnée par :
$IC_{1-\alpha}(\mu) = [\bar{x} \pm \text{erreur}] = [\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}\ ;\ \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}]$

Si $\alpha = 1$ % $\Rightarrow$ la table 2 de la loi normale donne $z_{0.01} = 2.575$

$IC_{0.99}(\mu) = [111 - 2.576\frac{13}{\sqrt{30}}\ ;\ 111 + 2.575\frac{13}{\sqrt{30}}] = [111 \pm 6.114] = [104.9\ ;\ 117.1]$

Si $\alpha = 5$ % $\Rightarrow$ la table 2 de la loi normale donne $z_{0.01} = 1.96$

$IC_{0.95}(\mu) = [111 - 1.96\frac{13}{\sqrt{30}}\ ;\ 111 + 1.96\frac{13}{\sqrt{30}}] = [111 \pm 4.652] = [106.35\ ;\ 115.65]$

Si $\alpha = 10$ % $\Rightarrow$ la table 2 de la loi normale donne $z_{0.01} = 1.645$

$IC_{0.9}(\mu) = [111 - 1.645\frac{13}{\sqrt{30}}\ ;\ 111 + 1.645\frac{13}{\sqrt{30}}] = [111 \pm 3.904] = [107.1\ ;\ 114.9]$

__Remarque__ :

En termes d'intervalles on dira que $IC_{0.99}(\mu)$ contient $IC_{0.95}(\mu)$ qui contient $IC_{0.90}(\mu)$.

En termes de longueur d'intervalles, on a :

La longueur de $IC_{0.99}$ (µ) = $2 * \text{erreur} = 2 * z_{0.01} \frac{\sigma}{\sqrt{n}} = 2 * 6.114 = 12.228$

La longueur de $IC_{0.95}$ (µ) = $2 * \text{erreur} = 2 * z_{0.05} \frac{\sigma}{\sqrt{n}} = 2 * 4.652 = 9.304$

La longueur de $IC_{0.9}$ (µ) = $2 * \text{erreur} = 2 * z_{0.1} \frac{\sigma}{\sqrt{n}} = 2 * 3.904 = 7.808$

Donc ces trois longueurs d'intervalles sont comparables.

Si la confiance $(1 - \alpha)$ diminue, la longueur de l'intervalle diminue et le risque α augmente.

2) On change l'échantillon extrait de la population de telle sorte que la demi-longueur de l'intervalle de confiance à 95 % n'excède pas l'unité sachant que la moyenne et la variance ne changent pas.

L'erreur = $z_\alpha \frac{\sigma}{\sqrt{n'}} = 1.96 \frac{13}{\sqrt{n'}} < 1 \Rightarrow$ **n' >** $(13 * 1.96)^2 = 649.2304$

On choisira un échantillon de taille au moins égale à **650** pour que la demi-longueur de l'intervalle de confiance à 95 % soit inférieure à l'unité.