# Construction d'intervalles de confiance

Mathieu Fauvernier

Equipe Biostatistique-Santé, UMR CNRS 5558, Université Lyon 1
Service de Biostatistique, Hospices Civils de Lyon

Section 1

# Intervalle de confiance d'une moyenne

# Rappel : théorème central limite

Soient $X_1, \ldots, X_n$ des v.a.r indépendantes, de même loi, et admettant une variance.

On note $\mu = E(X_1)$ et $\sigma^2 = Var(X_1)$.

Alors,

$$\hat{\mu} = \bar{X}_n \underset{n \to \infty}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

# Intervalle de fluctuation

Autrement dit, on a la relation suivante

$$P\left(\mu - 1.96\frac{\sigma}{\sqrt{n}} < \hat{\mu} < \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = 95\%$$

Il s'agit ici d'un **intervalle de fluctuation** d'une variable aléatoire

# Intervalle de fluctuation

Autrement dit, on a la relation suivante

$$P\left(\mu - 1.96\frac{\sigma}{\sqrt{n}} < \hat{\mu} < \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = 95\%$$

Il s'agit ici d'un **intervalle de fluctuation** d'une variable aléatoire

Or, de notre côté nous souhaitons un encadrement du paramètre inconnu $\mu$

# Intervalle de confiance de la moyenne

En inversant le relation précédente, on peut montrer qu'un encadrement de $\mu$ est donné par

$$P\left(\hat{\mu} - 1.96\frac{\hat{\sigma}}{\sqrt{n}} < \mu < \hat{\mu} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right) = 95\%$$

On appelle cet encadrement l'**intervalle de confiance** à 95% de $\mu$.

# Intervalle de confiance : illustration 1

On dispose d'un échantillon représentatif de 50 individus avec pour taille moyenne $\hat{\mu}$=175cm et pour écart-type $\hat{\sigma}$=50cm.

L'intervalle de confiance à 95% de la vraie moyenne $\mu$ (celle de la population) est

$$IC_{95\%}(\mu) = \left[\hat{\mu} - 1.96\frac{\hat{\sigma}}{\sqrt{n}}; \hat{\mu} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

$$IC_{95\%}(\mu) = [161; 189]$$

# Intervalle de confiance : illustration 2

On dispose d'un échantillon représentatif de **500** individus avec pour taille moyenne $\hat{\mu}$=175cm et pour écart-type $\hat{\sigma}$=50cm.

L'intervalle de confiance à 95% de la vraie moyenne $\mu$ (celle de la population) est

$$IC_{95\%}(\mu) = \left[\hat{\mu} - 1.96\frac{\hat{\sigma}}{\sqrt{\boldsymbol{n}}}; \hat{\mu} + 1.96\frac{\hat{\sigma}}{\sqrt{\boldsymbol{n}}}\right]$$

$$IC_{95\%}(\mu) = [\mathbf{171}; \mathbf{179}]$$

# Intervalle de confiance : illustration 3

On dispose d'un échantillon représentatif de **5000** individus avec pour taille moyenne $\hat{\mu}$=175cm et pour écart-type $\hat{\sigma}$=50cm.

L'intervalle de confiance à 95% de la vraie moyenne $\mu$ (celle de la population) est

$$IC_{95\%}(\mu) = \left[\hat{\mu} - 1.96\frac{\hat{\sigma}}{\sqrt{\boldsymbol{n}}}; \hat{\mu} + 1.96\frac{\hat{\sigma}}{\sqrt{\boldsymbol{n}}}\right]$$

$$IC_{95\%}(\mu) = [\mathbf{174}; \mathbf{176}]$$

Section 2

# Intervalle de confiance d’une proportion

# Estimation d'une proportion

On cherche à estimer la proportion $p$ de malades dans une population ($p$ est appelée la prévalence).

On dispose d'une échantillon aléatoire i.i.d $X_1, \ldots, X_n$ avec $X_i \sim \mathcal{B}(p)$

Ainsi, pour tout $i$, $E(X_i) = p$ et $Var(X_i) = p(1-p)$.

# Estimation d'une proportion

On cherche à estimer la proportion $p$ de malades dans une population ($p$ est appelée la prévalence).

On dispose d'une échantillon aléatoire i.i.d $X_1, \dots, X_n$ avec $X_i \sim \mathcal{B}(p)$

Ainsi, pour tout $i$, $E(X_i) = p$ et $Var(X_i) = p(1-p)$.

$p$ n'est rien d'autre qu'une moyenne de 0 et de 1. Ainsi, le théorème central limite nous donne

$$\hat{p} = \bar{X}_n \underset{n\to\infty}{\sim} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

# Intervalle de confiance d'une proportion

En reprenant les mêmes arguments que ceux exposés plus haut, on peut montrer qu'un encadrement de $p$ est donné par

$$P\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 95\%$$

On appelle cet encadrement l'**intervalle de confiance** à 95% de $p$.

# Intervalle de confiance : illustration 4

On dispose d'un échantillon représentatif de 50 individus avec pour proportion de malades $\hat{p}$=0.15.

L'intervalle de confiance à 95% de la vraie proportion $p$ (celle de la population) est

$$IC_{95\%}(p) = \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$IC_{95\%}(p) = [0.05; 0.25]$$

# Intervalle de confiance : illustration 5

On dispose d'un échantillon représentatif de **500** individus avec pour proportion de malades $\hat{p}$=0.15.

L'intervalle de confiance à 95% de la vraie proportion $p$ (celle de la population) est

$$IC_{95\%}(p) = \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{\boldsymbol{n}}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{\boldsymbol{n}}}\right]$$

$$IC_{95\%}(p) = [\mathbf{0.12}; \mathbf{0.18}]$$

# Intervalle de confiance : illustration 6

On dispose d'un échantillon représentatif de **5000** individus avec pour proportion de malades $\hat{p}$=0.15.

L'intervalle de confiance à 95% de la vraie proportion $p$ (celle de la population) est

$$IC_{95\%}(p) = \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{\boldsymbol{n}}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{\boldsymbol{n}}}\right]$$

$$IC_{95\%}(p) = [\mathbf{0.14}; \mathbf{0.16}]$$

# Exercice 1

Un sondage réalisé sur un échantillon de 1000 votants avant le deuxième tour de la présidentielle donne le candidat A vainqueur à $\hat{p}$=51%.

L'écart avec le candidat B est-il statistiquement significatif (la proportion estimée est-elle statistiquement différente de 0.5) ?

# Exercice 1

Un sondage réalisé sur un échantillon de 1000 votants avant le deuxième tour de la présidentielle donne le candidat A vainqueur à $\hat{p}$=51%.

L'écart avec le candidat B est-il statistiquement significatif (la proportion estimée est-elle statistiquement différente de 0.5) ?

L'intervalle de confiance à 95% de la vraie moyenne $p$ (celle de la population de votants) est

$$IC_{95\%}(p) = \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$IC_{95\%}(p) = [0.48; 0.54]$$

# Exercice 1

Un sondage réalisé sur un échantillon de 1000 votants avant le deuxième tour de la présidentielle donne le candidat A vainqueur à $\hat{p}$=51%.

L'écart avec le candidat B est-il statistiquement significatif (la proportion estimée est-elle statistiquement différente de 0.5) ?

L'intervalle de confiance à 95% de la vraie moyenne $p$ (celle de la population de votants) est

$$IC_{95\%}(p) = \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$IC_{95\%}(p) = [0.48; 0.54]$$

**L'intervalle contient la valeur 0.5 donc l'écart n'est pas significatif**

# Exercice 2

Un sondage réalisé sur un échantillon de 1000 votants avant le deuxième tour de la présidentielle donne le candidat A vainqueur à $\hat{p}$=55%.

La proportion estimée est-elle statistiquement différente de 0.5 ?

# Exercice 2

Un sondage réalisé sur un échantillon de 1000 votants avant le deuxième tour de la présidentielle donne le candidat A vainqueur à $\hat{p}$=55%.

La proportion estimée est-elle statistiquement différente de 0.5 ?

L'intervalle de confiance à 95% de la vraie moyenne $p$ (celle de la population de votants) est

$$IC_{95\%}(p) = \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$IC_{95\%}(p) = [0.52; 0.58]$$

# Exercice 2

Un sondage réalisé sur un échantillon de 1000 votants avant le deuxième tour de la présidentielle donne le candidat A vainqueur à $\hat{p}$=55%.

La proportion estimée est-elle statistiquement différente de 0.5 ?

L'intervalle de confiance à 95% de la vraie moyenne $p$ (celle de la population de votants) est

$$IC_{95\%}(p) = \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$IC_{95\%}(p) = [0.52; 0.58]$$

**La proportion estimée est statistiquement différente de 0.5**

# Exercice 3

On dispose d'un échantillon représentatif de 100 individus avec pour taux moyen de triglycérides $\hat{\mu}$=0.38g/L et pour écart-type $\hat{\sigma}$=0.5g/L.

Le taux moyen estimé est-il statistiquement anormal (en dehors des bornes 0.45 et 1.75 g/L) ?

# Exercice 3

On dispose d'un échantillon représentatif de 100 individus avec pour taux moyen de triglycérides $\hat{\mu}$=0.38g/L et pour écart-type $\hat{\sigma}$=0.5g/L.

Le taux moyen estimé est-il statistiquement anormal (en dehors des bornes 0.45 et 1.75 g/L) ?

L'intervalle de confiance à 95% de la vraie moyenne $\mu$ est

$$IC_{95\%}(\mu) = \left[\hat{\mu} - 1.96\frac{\hat{\sigma}}{\sqrt{n}}; \hat{\mu} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

$$IC_{95\%}(\mu) = [0.28; 0.48]$$

# Exercice 3

On dispose d'un échantillon représentatif de 100 individus avec pour taux moyen de triglycérides $\hat{\mu}$=0.38g/L et pour écart-type $\hat{\sigma}$=0.5g/L.

Le taux moyen estimé est-il statistiquement anormal (en dehors des bornes 0.45 et 1.75 g/L) ?

L'intervalle de confiance à 95% de la vraie moyenne $\mu$ est

$$IC_{95\%}(\mu) = \left[\hat{\mu} - 1.96\frac{\hat{\sigma}}{\sqrt{n}}; \hat{\mu} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

$$IC_{95\%}(\mu) = [0.28; 0.48]$$

**Le taux moyen estimé n'est pas statistiquement anormal**