Chapitre 5

ESTIMATION PAR INTERVALLES DE CONFIANCE

Chapitre 5

1. Introduction

1.1 Variable quantitative

1.2 Variable qualitative

2. Intervalles de confiance d'une moyenne

2.1 Variable normale d'écart-type connu

2.2 Variable quantitative quelconque d'écart-type connu

2.3 Variable quantitative quelconque

2.4 Taille minimum de l'échantillon pour une précision minimum

2.5 Variable normale d'écart-type inconnu

3. Intervalles de confiance d'une proportion

3.1 Rappel : intervalles de variation d'une fréquence

3.2 Intervalles de confiance d'une proportion

3.3 Précision dans l'estimation d'une proportion

3.4 Taille minimum de l'échantillon pour une précision minimum

1. Introduction (1)

X variable définie sur la population $\mathcal{P}$

échantillon de taille ***n*** de X issu de $\mathcal{P}$ ou observations $(\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n})$

- on sait donner une *estimation ponctuelle* du paramètre d'intérêt calculée sur les observations

variable **X**	*paramètre*		*estimation ponctuelle*	
qualitative	proportion de "*oui*"	p	fréquence observée	f
quantitative	moyenne	μ	moyenne observée	$\overline{x}$
	variance	σ^2	variance observée	s^2
			variance observée sans biais	s^{2*}
	écart-type	σ	écart-type observé	s
			écart-type observé sans biais	s^*

1. Introduction (2)

- l'estimation ponctuelle du paramètre n'est pas égale à la valeur du paramètre
- on cherche à intégrer dans l'estimation du paramètre la *précision* de cette estimation

⇒**cadre pratique : on suppose que la valeur du paramètre d'intérêt est *inconnue* dans la population $\mathcal{P}$**

– sur l'échantillon de taille ***n*** :

➔ évaluer la précision de l'estimation

➔ donner un intervalle (fourchette) de valeurs plausibles pour la valeur du paramètre : estimation par intervalle

1.1 Variable quantitative

- **Exemple : durée de chômage**

$\mathcal{P}$ = { *chômeurs français* } ***N*** = ?

X = *"durée de chômage" (en mois)*
variable quantitative

$$\begin{cases} \mu = \text{durée moyenne } \textbf{inconnue} \\ \sigma = \text{écart-type de la durée } \textbf{inconnu} \end{cases}$$

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n*** = *30*

➢ *la durée moyenne* μ *est estimée par la durée moyenne observée* $\bar{x} = 5{,}4$ *mois*

➢ *l'écart-type de la durée* σ *est estimé par l'écart-type observé sans biais* ***s**** = 2,4 *mois*

à partir des observations de l'échantillon de taille ***n*** *= 30 :*

① *quel intervalle contient des valeurs plausibles de la durée moyenne* μ ***inconnue*** *?*

② *quelle est la précision de l'estimation de la durée moyenne* μ *par* $\bar{x}$ *?*

③ *quelle serait la taille minimum de l'échantillon pour avoir une précision minimum de 1 mois dans l'estimation de* μ *?*

1.2 Variable qualitative

- **Exemple : efficacité d'un traitement**

$\mathcal{P}$ = { enfants atteints de troubles de l'anxiété, sous traitement } ***N*** *= ?*

X *= "amélioration clinique" : oui, non variable qualitative dichotomique*

p *= proportion d'amélioration clinique* ***inconnue*** *dans $\mathcal{P}$*

échantillon de ***X*** *issu de $\mathcal{P}$ de taille* ***n*** *= 63*

➢ *la proportion d'amélioration clinique* ***p*** *est estimée par la fréquence d'amélioration observée f = 0,762*

à partir des observations de l'échantillon de taille ***n*** *= 63 :*

① *quel intervalle contient des valeurs plausibles de la proportion d'amélioration clinique* ***p inconnue*** *?*

② *quelle est la précision de l'estimation de la proportion d'amélioration clinique* ***p*** *par* ***f*** *?*

③ *quelle serait la taille minimum de l'échantillon pour avoir une précision minimum de 5% dans l'estimation de la proportion d'amélioration clinique* ***p****?*

2. Intervalles de confiance d'une moyenne

X variable quantitative de $\mathcal{P}$ dans E

⇒deux paramètres

$$\begin{cases} \textbf{moyenne de X} = \mu & \textit{inconnue} \text{ dans } \mathcal{P} \\ \textbf{écart-type de X} = \sigma & \textit{inconnu} \text{ dans } \mathcal{P} \end{cases}$$

échantillon de X issu de $\mathcal{P}$ de taille ***n***

- rechercher un intervalle dans lequel on a une confiance de $(1-\alpha)$ de trouver la moyenne μ à partir des observations de l'échantillon de taille ***n***

⇒cet intervalle est l'**intervalle de confiance** de la moyenne μ au niveau de confiance $(1-\alpha)$ ou au risque α noté $\mathbf{IC_{1-\alpha}(\mu)}$

2.1 Variable normale d'écart-type connu (1)

X variable **quantitative** de $\mathcal{P}$ *suit le* ***modèle normal*** $\begin{cases} \text{moyenne } \mu \text{ \textbf{\textit{inconnue}}} \\ \text{écart-type } \sigma \text{ \textbf{\textit{connu}}} \end{cases}$

échantillon de **X** issu de $\mathcal{P}$ de taille ***n***

➢ la moyenne empirique $\overline{X}_n$ suit un modèle normal de moyenne μ et d'écart-type σ / √***n***

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

➢ l'intervalle de variation au niveau (1−α) ou au risque α de $\overline{X}_n$ s'écrit :

$$I_{1-\alpha}(\overline{X}_n) = \left[\mu \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = [\mu \pm \boldsymbol{e}]$$

où ***e*** marge d'erreur à (1−α)

$z_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de $\mathbf{Z} \sim \mathcal{N}(0,1)$

⇒ l'**intervalle de confiance au niveau (1−α)** ou **au risque** α de la **moyenne** μ s'écrit :

$$IC_{1-\alpha}(\mu) = \left[\overline{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = [\overline{x} \pm \boldsymbol{e}]$$

où $\overline{x}$ moyenne observée de X

Exemple variable normale (1)

- **Exemple : score de stress perçu**

  $\mathcal{P}$ = { *sujets* }

  X = *"score de stress perçu"*
  variable quantitative

  $\begin{cases} \mu = \text{moyenne du score inconnue} \\ \sigma = \text{écart-type du score} = 7{,}5 \text{ connu} \end{cases}$

  échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n*** *= 25*

  $\mathbf{X} \sim \mathcal{N}(\mu\ ;\ 7{,}5)$ donc $\overline{X}_n \sim \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{n}} = 1{,}5\right)$

  dans quel intervalle s'attend-on "raisonnablement" à observer le score moyen sur un échantillon de taille ***n*** *= 25 ?*

  ➢ *intervalle de variation à 90% (au risque 10%) du score moyen :*

  $$I_{90\%}(\overline{X}_n) = \left[\mu \pm 1{,}645 \frac{7{,}5}{\sqrt{25}}\right] = [\mu \pm 2{,}5]$$

  ➔ *90% des échantillons de* **X** *issus de* $\mathcal{P}$ *de taille* ***n*** *= 25 ont un score moyen observé compris entre* $\mu+2{,}5$ *et* $\mu-2{,}5$

  ➔ *la marge d'erreur à 90% dans l'estimation du score moyen sur les échantillons de taille* ***n*** *= 25 est de 2,5*

Exemple variable normale (2)

- **Exemple : score de stress perçu**

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*
n *= 25*

① *à partir des observations faites sur l'échantillon de taille* ***n*** *= 25 dans quel intervalle a-t-on une "grande confiance" de trouver le score moyen ?*

➢ *intervalle de confiance à 90% (au risque 10%) du score moyen sur les échantillons de taille* ***n*** *= 25 :*

$$\mathrm{IC}_{90\%}(\mu) = \left[\overline{x} \pm 1{,}645 \frac{7{,}5}{\sqrt{25}}\right] = [\overline{x} \pm 2{,}5]$$

➢ *pour le score moyen observé* $\overline{x} = 30$

intervalle de confiance à 90% (au risque 10%) du score moyen :

$$\mathrm{IC}_{90\%}(\mu) = [30 \pm 2{,}5] = [27{,}5\,;32{,}5]$$

① *on accorde une confiance de 90% au fait que l'intervalle [27,5 ; 32,5] contienne* μ *le score moyen dans* $\mathcal{P}$

Exemple variable normale (3)

- **Exemple : score de stress perçu**

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*

***n** = 25*

$$IC_{90\%}(\mu) = \left[\bar{x} \pm 1{,}645\frac{7{,}5}{\sqrt{25}}\right] = [\bar{x} \pm 2{,}5]$$

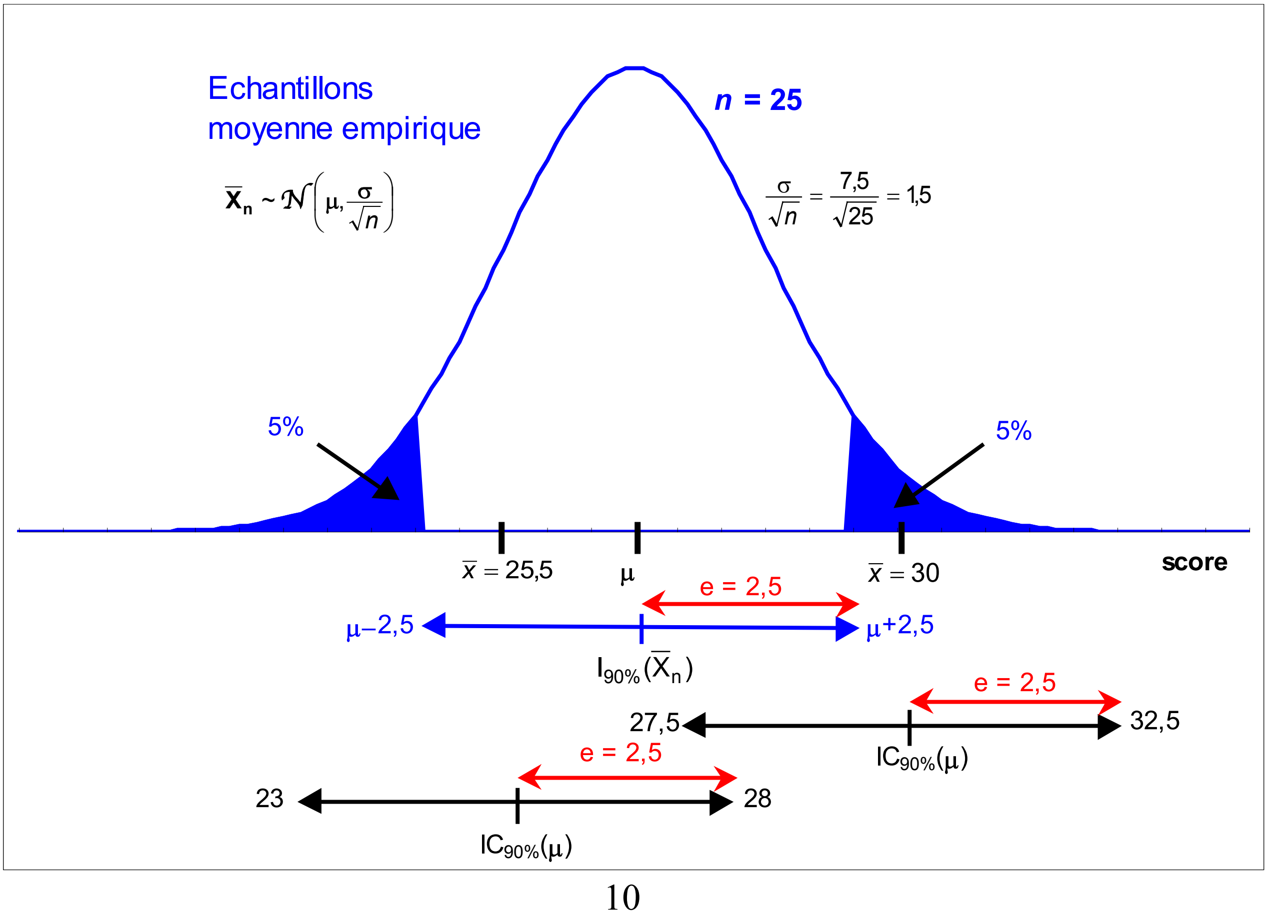
② *la marge d'erreur à 90% dans l'estimation du score moyen sur les échantillons de taille **n** = 25 est de 2,5*

➢ *pour le score moyen observé* $\bar{x} = 30$

$$IC_{90\%}(\mu) = [30 \pm 2{,}5] = [27{,}5\,;\,32{,}5]$$

➢ *pour le score moyen observé* $\bar{x} = 25{,}5$

$$IC_{90\%}(\mu) = [25{,}5 \pm 2{,}5] = [23\,;\,28]$$



Exemple variable normale (4)

- **Exemple : score de stress perçu**

$\mathcal{P}$ = *{ sujets }*

X = *"score de stress perçu"*

variable quantitative

deux paramètres dans $\mathcal{P}$

$$\begin{cases} \mu = \text{moyenne du score } \textbf{inconnue} \\ \sigma = \text{écart-type du score} = \textbf{7,5 connu} \end{cases}$$

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*

n *= 25*

$\mathbf{X} \sim \mathcal{N}(\mu \,; 7{,}5)$ donc $\overline{X}_n \sim \mathcal{N}(27 ; 1{,}5)$

➢ *intervalle de variation à 95% (au risque 5%) du score moyen observé :*

$$I_{95\%}\left(\overline{X}_n\right) = \left[\mu \pm 1{,}96 \frac{7{,}5}{\sqrt{25}}\right] = [\mu \pm 2{,}9]$$

➔ *95% des échantillons de* **X** *issus de* $\mathcal{P}$ *de taille* ***n*** *= 25 ont un score moyen observé compris entre* μ *–2,9 et* μ *+2,9*

➔ *la marge d'erreur à 95% dans l'estimation du score moyen sur les échantillons de taille* ***n*** *= 25 est de 2,9*

Exemple variable normale (5)

- **Exemple : score de stress perçu**

  *échantillon de **X** issu de $\mathcal{P}$ de taille **n** = 25*

  ➢ *intervalle de confiance à 95% (au risque 5%) du score moyen sur les échantillons de taille **n** = 25 :*

$$IC_{95\%}(\mu) = \left[\bar{x} \pm 1{,}96 \frac{7{,}5}{\sqrt{25}}\right] = [\bar{x} \pm 2{,}9]$$

  ➢ *pour un score moyen observé* $\bar{x} = 30$

  intervalle de confiance à 95% (au risque 15%) du score moyen :

$$IC_{95\%}(\mu) = [30 \pm 2{,}9] = [27{,}1\,;32{,}9]$$

  ① *on accorde une confiance de 95% au fait que l'intervalle [27,1 ; 32,9] contienne* μ *le score moyen dans* $\mathcal{P}$

  ➢ *pour un score moyen observé* $\bar{x} = 25{,}5$

  intervalle de confiance à 95% (au risque 5%) du score moyen :

$$IC_{95\%}(\mu) = [25{,}5 \pm 2{,}9] = [22{,}6\,;28{,}4]$$

  ① *on accorde une confiance de 95% au fait que l'intervalle [22,6 ; 28,4] contienne* μ *le score moyen dans* $\mathcal{P}$

Exemple variable normale (6)

- **Exemple : score de stress perçu**

$\mathcal{P}$ = *{ sujets }*

X = *"score de stress perçu"*

variable quantitative

deux paramètres dans $\mathcal{P}$

$$\begin{cases} \mu = \textit{moyenne du score } \textbf{inconnue} \\ \sigma = \textit{écart-type du score} = \textbf{7,5 connu} \end{cases}$$

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*

n *= 36*

$$\mathbf{X} \sim \mathcal{N}(\mu\ ;\ 7{,}5) \text{ donc } \overline{X}_n \sim \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{n}} = 1{,}25\right)$$

➢ *intervalle de variation à 90% (au risque 10%) du score moyen observé :*

$$I_{90\%}\left(\overline{X}_n\right) = \left[\mu \pm 1{,}645 \frac{7{,}5}{\sqrt{36}}\right] = [\mu \pm 2{,}1]$$

➢ *intervalle de confiance à 90% (au risque 10%) du score moyen sur les échantillons de taille* ***n*** *= 36 :*

$$IC_{90\%}(\mu) = \left[\overline{x} \pm 1{,}645 \frac{7{,}5}{\sqrt{36}}\right] = [\overline{x} \pm 2{,}1]$$

② *la marge d'erreur à 90% dans l'estimation du score moyen sur les échantillons de taille* ***n*** *= 36 est de 2,1*

2.1 Variable normale d'écart-type connu (2)

Remarques

l'intervalle de confiance à $(1-\alpha)$ de μ :

$$IC_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = [\bar{x} \pm \boldsymbol{e}]$$

- $\boldsymbol{e}$ la marge d'erreur à $(1-\alpha)$ dans l'estimation de la moyenne μ est *fixe*

- tous les intervalles de confiance de la moyenne μ sont centrés sur $\bar{x}$ la moyenne observée sur l'échantillon de taille $\boldsymbol{n}$

- l'intervalle de confiance dépend
  - **de la moyenne observée** $\bar{x}$
  - **de l'écart-type** σ
  - **du niveau** $(1-\alpha)$ **ou du risque** α
  - **de la taille de l'échantillon** ***n***

- l'intervalle de confiance ne dépend pas
  - **de la moyenne** μ de X dans $\mathcal{P}$

2.1 Variable normale d'écart-type connu (3)

Précision dans l'estimation de la moyenne

La demi-longueur ***e*** de l'intervalle de confiance à (1–α) représente la **précision** à (1–α) dans l'estimation de la moyenne μ sur l'échantillon de taille ***n***, ou la **marge d'erreur** au risque α

$$\boxed{\boldsymbol{e} = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{\boldsymbol{n}}}}$$

où $z_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de **Z** $\sim \mathcal{N}(0,1)$

- la marge ***e*** dépend de σ
- pour ***n*** fixé, la précision ***e*** ne change pas lorsque l'échantillon change
- pour ***n*** fixé, plus le niveau (1–α) augmente (le risque α diminue) plus la marge d'erreur ***e*** augmente
- plus la taille de l'échantillon ***n*** augmente, plus la marge d'erreur ***e*** diminue

<u>Interprétation d'un intervalle de confiance</u> (1)

- une moyenne observée $\bar{x}$ est *variable* : elle change pour chaque échantillon
- un intervalle de confiance est *variable* : ses bornes sont *variables*, changent pour chaque échantillon de taille ***n***
- les intervalles de confiance au niveau $(1-\alpha)$ (au risque α) sur les échantillons de taille ***n*** sont tels que :
  - $(1-\alpha)\%$ des intervalles contiennent la moyenne μ de X dans $\mathcal{P}$

  et

  - $\alpha\%$ des intervalles <u>ne</u> contiennent <u>pas</u> la moyenne μ de X dans $\mathcal{P}$

Interprétation d'un intervalle de confiance (2)

- les intervalles de confiance au niveau 90% (au risque 10%) sont tels que :
  - **90% des intervalles de confiance à 90% contiennent la moyenne μ de X dans $\mathcal{P}$**

  et

  - **10% des intervalles de confiance à 90% ne contiennent pas la moyenne μ de X dans $\mathcal{P}$**

- par construction, par exemple sur **100** échantillons de taille ***n*** de **X** issus de $\mathcal{P}$, en moyenne :
  - **90 intervalles de confiance à 90% contiendront la moyenne μ de X dans $\mathcal{P}$**

  et

  - **10 intervalles de confiance à 90% ne contiendront pas la moyenne μ de X dans $\mathcal{P}$**

Interprétation d'un intervalle de confiance (3)

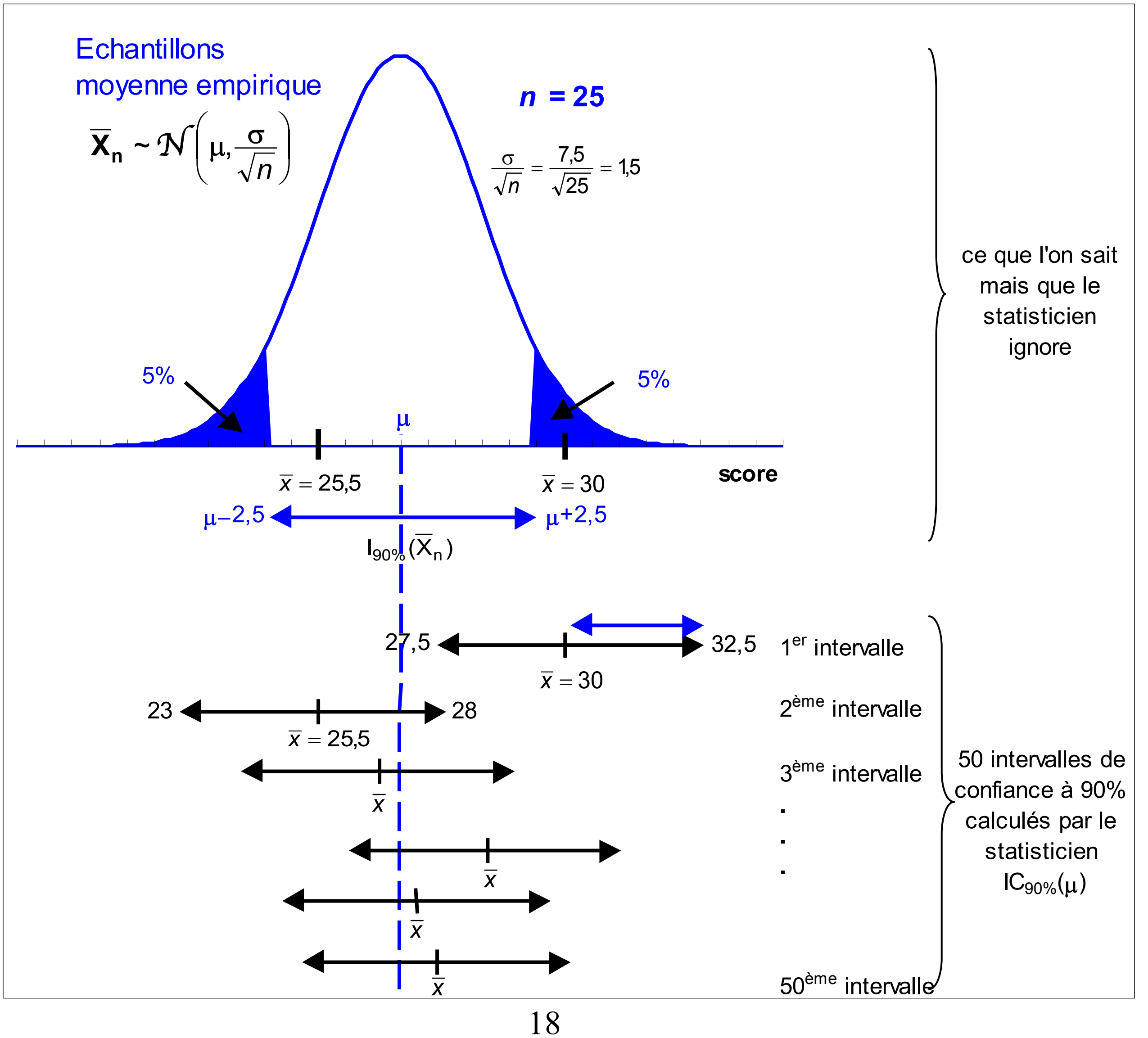
- **Exemple : score de stress perçu**

$\mathcal{P}$ = { *sujets* }

$\boldsymbol{X}$ = *"score de stress perçu"*
variable quantitative

$\begin{cases} \mu = \textit{moyenne du score } \textbf{\textit{inconnue}} \\ \sigma = \textit{écart-type du score} = \textbf{\textit{7,5 connu}} \end{cases}$

50 échantillons de $\boldsymbol{X}$ *issu de* $\mathcal{P}$ *de taille* $\boldsymbol{n} = 25$



<u>Interprétation d'un intervalle de confiance</u> (4)

pour un intervalle de confiance à $(1-\alpha)$ de la moyenne μ de X, deux possibilités :

- **soit il contient la moyenne μ de X**
- **soit il ne contient pas la moyenne μ**

mais si la moyenne μ de X est *inconnue* on ne sait pas dans quelle situation on se trouve

➔on accorde une confiance de $(1-\alpha)$ au fait que l'intervalle de confiance au niveau $(1-\alpha)$ de la moyenne μ de X contienne la moyenne μ

➔l'intervalle de confiance au niveau $(1-\alpha)$ de la moyenne μ de X contient la moyenne μ avec un niveau de confiance de $(1-\alpha)$ (avec un risque α)

2.2 Variable quantitative quelconque d'écart-type connu

X variable **quantitative**

$\begin{cases} \text{moyenne } \mu \textit{ inconnue} \text{ dans } \mathcal{P} \\ \text{écart-type } \sigma \textit{ connu} \text{ dans } \mathcal{P} \end{cases}$

échantillon de **X** issu de $\mathcal{P}$ de taille ***n* ≥ 30**

➢ la moyenne empirique $\overline{X}_n$ suit *approximativement* un modèle normal de moyenne μ et d'écart-type σ / √***n***

$$\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

⇒**l'intervalle de confiance à (1−α)** ou **au risque** α de **la moyenne** μ s'écrit :

$$IC_{1-\alpha}(\mu) \approx \left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \approx [\bar{x} \pm \boldsymbol{e}]$$

où $\bar{x}$ moyenne observée sur l'échantillon

e marge d'erreur à (1−α)

$z_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de **Z** $\sim \mathcal{N}(0,1)$

2.3 Variable quantitative quelconque (1)

X variable **quantitative**

$$\begin{cases} \text{moyenne } \mu \ \ \textbf{\textit{inconnue}} \text{ dans } \mathcal{P} \\ \text{écart-type } \sigma \ \ \textbf{\textit{inconnu}} \text{ dans } \mathcal{P} \end{cases}$$

échantillon de **X** issu de $\mathcal{P}$ de taille $\boldsymbol{n \geq 30}$

- la moyenne empirique $\overline{X}_n$ suit *approximativement* un modèle normal de moyenne μ et d'écart-type $\sigma / \sqrt{n}$

$$\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

⇒l'**intervalle de confiance à (1−α)** ou **au risque** α de **la moyenne** μ s'écrit :

$$IC_{1-\alpha}(\mu) \approx \left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{s^*}{\sqrt{n}}\right] \approx [\bar{x} \pm \boldsymbol{e}]$$

où $\bar{x}$ moyenne observée de X

$\boldsymbol{s^*}$ écart-type observé sans biais de X

$\boldsymbol{e}$ marge d'erreur à (1−α)

$z_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de **Z** $\sim \mathcal{N}(0,1)$

2.3 Variable quantitative quelconque (2)

<u>Cas particuliers</u>

- **intervalle de confiance à 90%** ou **au risque 10%** de la moyenne μ :

$$IC_{90\%}(\mu) \approx \left[\bar{x} \pm z_{0,95} \frac{s^*}{\sqrt{n}}\right] \approx \left[\bar{x} \pm 1{,}645 \frac{s^*}{\sqrt{n}}\right]$$

- **intervalle de confiance à 95%** ou **au risque 5%** de la moyenne μ :

$$IC_{95\%}(\mu) \approx \left[\bar{x} \pm z_{0,975} \frac{s^*}{\sqrt{n}}\right] \approx \left[\bar{x} \pm 1{,}96 \frac{s^*}{\sqrt{n}}\right]$$

où $\bar{x}$ moyenne observée de X
s^* écart-type observé sans biais de X

Exemple variable quantitative (1)

- **Exemple : durée de chômage**

$\mathcal{P}$ = *{ chômeurs français }* ***N*** *= ?*

X *= "durée de chômage" (en mois)*
variable quantitative

$\begin{cases} \mu = \text{durée moyenne } \textbf{inconnue} \\ \sigma = \text{écart-type de la durée } \textbf{inconnu} \end{cases}$

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*
n *= 30*

on observe :
la durée moyenne $\bar{x} = 5{,}4$ *mois*
l'écart-type sans biais ***s**** *= 2,4 mois*

n *= 30 (****n*** $\geq$ *30) alors* $\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{30}}\right)$

① *dans quel intervalle a-t-on une "grande confiance" de trouver la durée moyenne sur l'échantillon de taille* ***n*** *= 30 ?*

➢ *intervalle de confiance à 90% (au risque 10%) de la durée moyenne :*

$$IC_{90\%}(\mu) \approx \left[5{,}4 \pm 1{,}645 \frac{2{,}4}{\sqrt{n}}\right] = [5{,}4 \pm 0{,}7]$$

$$\approx [4{,}7 ; 6{,}1]$$

① *on accorde une confiance de 90% au fait que l'intervalle [4,7 ; 6,1] contienne* μ *la durée moyenne dans* $\mathcal{P}$

Exemple variable quantitative (2)

- **Exemple : durée de chômage**

  $\mathcal{P}$ = *{ chômeurs français }* ***N*** *= ?*

  X *= "durée de chômage" (en mois)*
  variable quantitative

  $\begin{cases} \mu = \text{durée moyenne } \textbf{inconnue} \\ \sigma = \text{écart-type de la durée } \textbf{inconnu} \end{cases}$

  échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*
  n *= 30*

  on observe :
  la durée moyenne $\bar{x} = 5{,}4$ *mois*
  l'écart-type sans biais ***s**** *= 2,4 mois*

  n *= 30 (*$n \geq 30$ *) alors* $\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{30}}\right)$

  ② *quelle est la précision dans l'estimation de la durée moyenne sur l'échantillon de taille* ***n*** *= 30 ?*

  ➢ *marge d'erreur à 90% dans l'estimation de la durée moyenne :*

  $$IC_{90\%}(\mu) \approx [5{,}4 \pm 0{,}7]$$

  ② *la marge d'erreur à 90% dans l'estimation de la durée moyenne est de 0,7 mois*

Exemple variable quantitative (3)

- **Exemple : durée de chômage**

$\mathcal{P}$ = { *chômeurs français* } ***N*** = ?

X = *"durée de chômage" (en mois)*
variable quantitative

$$\begin{cases} \mu = \text{durée moyenne } \textbf{inconnue} \\ \sigma = \text{écart-type de la durée } \textbf{inconnu} \end{cases}$$

*échantillon de **X** issu de* $\mathcal{P}$ *de taille* ***n*** *= 30*

on observe :
la durée moyenne $\overline{x} = 5{,}4$ *mois*
*l'écart-type sans biais **s**** *= 2,4 mois*

n *= 30 (**n*** $\geq$ *30) alors* $\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{30}}\right)$

➢ *intervalle de confiance à 95% (au risque 5%) de la durée moyenne :*

$$IC_{95\%}(\mu) \approx \left[5{,}4 \pm 1{,}96 \frac{2{,}4}{\sqrt{30}}\right] = [5{,}4 \pm 0{,}9]$$
$$\approx [4{,}5 ; 6{,}3]$$

① *on accorde une confiance de 95% au fait que l'intervalle [4,5 ; 6,3] contienne* μ *la durée moyenne dans* $\mathcal{P}$

② *la marge d'erreur à 95% dans l'estimation de la durée moyenne est de 0,9 mois*

2.3 Variable quantitative quelconque (2)

Remarques

l'intervalle de confiance à (1–α) de μ :

$$IC_{1-\alpha}(\mu) \approx \left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{s^*}{\sqrt{n}} \right] \approx [\bar{x} \pm \boldsymbol{e}]$$

- ***e*** la marge d'erreur à (1–α) dans l'estimation de la moyenne μ *varie*

- tous les intervalles de confiance de la moyenne μ sont centrés sur $\bar{x}$ la moyenne observée sur l'échantillon de taille ***n***

- l'intervalle de confiance dépend
  - **de la moyenne observée** $\bar{x}$
  - **de l'écart-type observé sans biais *s****
  - **du niveau (1–α) ou du risque α**
  - **de la taille de l'échantillon *n***

- l'intervalle de confiance ne dépend pas
  - **de la moyenne μ de X dans** $\mathcal{P}$

2.3 Variable quantitative quelconque (3)

Précision dans l'estimation de la moyenne

La demi-longueur $\boldsymbol{e}$ de l'intervalle de confiance à $(1-\alpha)$ représente la **précision** à $(1-\alpha)$ dans l'estimation de la moyenne μ sur l'échantillon de taille $\boldsymbol{n}$, ou la **marge d'erreur** au risque α

$$\boxed{\boldsymbol{e} = z_{1-\frac{\alpha}{2}} \frac{s^*}{\sqrt{\boldsymbol{n}}}}$$

où $z_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de $\mathbf{Z} \sim \mathcal{N}(0,1)$

- la marge $\boldsymbol{e}$ dépend de $\boldsymbol{s^*}$
- pour $\boldsymbol{n}$ fixé, la précision $\boldsymbol{e}$ *change* lorsque l'échantillon change
- pour $\boldsymbol{n}$ fixé, plus le niveau $(1-\alpha)$ augmente (le risque α diminue) plus la marge d'erreur $\boldsymbol{e}$ augmente
- plus la taille de l'échantillon $\boldsymbol{n}$ augmente, plus la marge d'erreur $\boldsymbol{e}$ diminue

2.4 Taille minimum de l'échantillon pour une précision minimum (1)

- Pour contrôler la marge d'erreur et obtenir une marge d'erreur à $(1-\alpha)$ ***inférieure à*** ***e*** sur l'estimation de la moyenne μ
  la longueur (amplitude) de l'intervalle de confiance au niveau $(1-\alpha)$ ou au risque α n'excède pas **2*e***

⇒**la taille de l'échantillon *n* vérifie :**

- **si l'écart-type σ est connu**

$$\boxed{n \geq \left(z_{1-\frac{\alpha}{2}} \frac{\sigma}{e} \right)^2}$$

- **si l'écart-type σ est inconnu**

$$\boxed{n \geq \left(z_{1-\frac{\alpha}{2}} \frac{s^*}{e} \right)^2}$$

→ pour une marge d'erreur maximum ***e*** (ou une précision minimum) il faut une taille d'échantillon d'au moins ***n***

2.4 Taille minimum de l'échantillon pour une précision minimum (2)

Remarques

- si l'écart-type σ est connu : le calcul est identique à celui utilisé pour les intervalles de variation

- la taille minimum ***n*** dépend de
  - **l'écart-type σ ou s***
  - **la marge d'erreur maximum e**
  - **le niveau $(1-\alpha)$ ou risque α**

- la taille minimum ***n*** augmente lorsque
  - **l'écart-type σ ou s* augmente**
  - **le niveau $(1-\alpha)$ augmente**
le risque α diminue
  - **la marge d'erreur maximum e diminue**
la précision minimum e augmente

- pour **X** quantitative quelconque
  vérifier que la taille ***n*** obtenue est
  supérieure à 30

Exemple variable normale

- **Exemple : score de stress perçu**

  $\mathcal{P}$ = *{ sujets }*

  X = *"score de stress perçu"*
  variable quantitative

  $\begin{cases} \mu = \text{moyenne du score inconnue} \\ \sigma = \text{écart-type du score} = \mathbf{7{,}5} \text{ connu} \end{cases}$

  *échantillon de **X** issu de $\mathcal{P}$ de taille **n***

  $$\mathbf{X} \sim \mathcal{N}(\mu\,;\,7{,}5) \text{ donc } \overline{X}_n \sim \mathcal{N}\left(\mu\,;\frac{7{,}5}{\sqrt{\boldsymbol{n}}}\right)$$

  ➢ *pour **n** = 25 l'intervalle de confiance à 90% (au risque 10%) du score moyen :*

  $$IC_{90\%}(\mu) = [\overline{x} \pm 2{,}5]$$

  ➔ *la marge d'erreur à 90% dans l'estimation du score moyen sur les échantillons de taille **n** = 25 est de 2,5*

  ➢ *marge d'erreur à 90% d'au plus 2 points*

  $$IC_{90\%}(\mu) = \left[\overline{x} \pm 1{,}645\frac{7{,}5}{\sqrt{n}}\right] = [\overline{x} \pm 2]$$

  $$1{,}645\frac{7{,}5}{\sqrt{\boldsymbol{n}}} \le 2 \quad \boldsymbol{n} \ge \left(1{,}645\frac{7{,}5}{2}\right)^2 = 38{,}05$$

  ➔ *pour que la marge d'erreur à 90% dans l'estimation du score moyen n'excède pas 2 points il faudra un échantillon d'**au moins** 39 sujets*

Exemple variable quantitative (1)

- **Exemple : durée de chômage**

  $\mathcal{P}$ = { chômeurs français } ***N*** *= ?*

  X *= "durée de chômage" (en mois)*
  variable quantitative

  $\begin{cases} \mu = \text{durée moyenne } \textbf{inconnue} \\ \sigma = \text{écart-type de la durée } \textbf{inconnu} \end{cases}$

  échantillon de ***X*** *issu de $\mathcal{P}$ de taille* ***n***

  si ***n*** $\geq$ *30 alors* $\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu;\frac{\sigma}{\sqrt{n}}\right)$

  on suppose que :
  l'écart-type sans biais ***s**** *= 2,4 mois*

  ➢ *pour* ***n*** *= 30 l'intervalle de confiance à 90% (au risque 10%) de la durée moyenne :*

  $$IC_{90\%}(\mu) \approx [5{,}4 \pm 0{,}7] = [\overline{x} \pm 0{,}7]$$

  ➔ *la marge d'erreur à 90% dans l'estimation du score moyen sur les échantillons de taille* ***n*** *= 30 est de 0,7 mois*

  ➢ *marge d'erreur à 90% d'au plus 1 mois*

  $$IC_{90\%}(\mu) \approx [\overline{x} \pm 1]$$

  ➔ *pour que la marge d'erreur à 90% dans l'estimation de la durée moyenne n'excède pas 1 mois un échantillon d'**au moins** 30 chômeurs convient*

Exemple variable quantitative (2)

- **Exemple : durée de chômage**

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n***

si ***n*** $\geq$ *30 alors* $\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{n}}\right)$

on suppose que :

l'écart-type sans biais ***s**** *= 2,4 mois*

➢ *marge d'erreur à 90% d'au plus 0,5 mois*

$$IC_{90\%}(\mu) \approx \left[\overline{x} \pm 1{,}645 \frac{2{,}4}{\sqrt{n}}\right] = [\overline{x} \pm 0{,}5]$$

$$1{,}645 \frac{2{,}4}{\sqrt{\boldsymbol{n}}} \leq 0{,}5 \quad \boldsymbol{n} \geq \left(1{,}645 \frac{2{,}4}{0{,}5}\right)^2 = 62{,}3$$

➔ *pour que la marge d'erreur à 90% dans l'estimation de la durée moyenne n'excède pas 0,5 mois il faut un échantillon d'**au moins** 63 chômeurs (63 ≥ 30)*

Exemple variable quantitative (3)

- **Exemple : durée de chômage**

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n***

si ***n*** $\geq$ *30 alors* $\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{n}}\right)$

on suppose que :

l'écart-type sans biais ***s**** *= 2,4 mois*

➢ *pour* ***n*** *= 30 l'intervalle de confiance à 95% (au risque 5%) de la durée moyenne :* $IC_{95\%}(\mu) \approx [\overline{x} \pm 0{,}9]$

➔ *la marge d'erreur à 95% dans l'estimation du score moyen sur les échantillons de taille* ***n*** *= 30 est de 0,9 mois*

➢ *marge d'erreur à 95% d'au plus 0,5 mois*

$$IC_{95\%}(\mu) \approx \left[\overline{x} \pm 1{,}96 \frac{2{,}4}{\sqrt{n}}\right] = [\overline{x} \pm 0{,}5]$$

$$1{,}96 \frac{2{,}4}{\sqrt{\boldsymbol{n}}} \leq 0{,}5 \qquad \boldsymbol{n} \geq \left(1{,}96 \frac{2{,}4}{0{,}5}\right)^2 = 88{,}5$$

➔ *pour que la marge d'erreur à 95% dans l'estimation de la durée moyenne n'excède pas 0,5 mois il faut un échantillon d'**au moins** 89 chômeurs (89* $\geq$ *30)*

Exemple variable quantitative (4)

- **Exemple : durée de chômage**

*échantillon de **X** issu de $\mathcal{P}$ de taille **n***

*si **n** ≥ 30 alors* $\overline{X}_n \underset{approx}{\sim} \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{n}}\right)$

on suppose que :

*l'écart-type sans biais **s*** = 2,4 mois*

➢ *marge d'erreur à 95% d'au plus 0,1 mois*

$$IC_{95\%}(\mu) \approx \left[\bar{x} \pm 1{,}96 \frac{2{,}4}{\sqrt{n}}\right] = [\bar{x} \pm 0{,}1]$$

$$1{,}96 \frac{2{,}4}{\sqrt{\boldsymbol{n}}} \leq 0{,}1 \quad \boldsymbol{n} \geq \left(1{,}96 \frac{2{,}4}{0{,}1}\right)^2 = 2212{,}8$$

➔ *pour que la marge d'erreur à 95% dans l'estimation de la durée moyenne n'excède pas 0,5 mois il faut un échantillon d'**au moins** 2213 chômeurs (2213 ≥ 30)*

2.5 Variable normale d'écart-type inconnu (1)

X variable **quantitative** de $\mathcal{P}$ *suit le* ***modèle normal*** $\begin{cases} \text{moyenne } \mu \textbf{ \textit{inconnue}} \\ \text{écart-type } \sigma \textbf{ \textit{inconnu}} \end{cases}$

échantillon de **X** issu de $\mathcal{P}$ de taille ***n***

- la moyenne empirique $\overline{X}_n$ suit une loi normale de moyenne μ et d'écart-type $\sigma / \sqrt{n}$

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

⇒ la statistique T suit une loi de **Student** à **(*n*–1) degrés de liberté**

$$T = \frac{\overline{X}_n - \mu}{\frac{S_n^*}{\sqrt{n}}} \sim T_{n-1}$$

⇒ l'**intervalle de confiance au niveau (1–α)** ou **au risque** α de la **moyenne** μ s'écrit :

$$IC_{1-\alpha}(\mu) = \left[\overline{x} \pm t_{1-\frac{\alpha}{2}} \frac{s^*}{\sqrt{n}}\right] = [\overline{x} \pm \boldsymbol{e}]$$

où $\overline{x}$ moyenne observée de X

e marge d'erreur à (1–α)

$t_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathbf{T}_{n-1}$

Exemple variable normale (1)

- **Exemple : score de stress perçu**

  $\mathcal{P}$ = { sujets }

  X *= "score de stress perçu"*

  variable quantitative

  $\begin{cases} \mu \text{ = moyenne du score } \textbf{inconnue} \\ \sigma \text{ = écart-type du score } \textbf{inconnu} \end{cases}$

  *échantillon de **X** issu de $\mathcal{P}$ de taille*

  ***n** = 25*

  on observe :

  le score moyen $\bar{x} = 30$

  *l'écart-type sans biais **s*** = 10*

  $\mathbf{X} \sim \mathcal{N}(\mu\ ;\ \sigma)$ donc $\bar{X}_n \sim \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{n}}\right)$

  et $T \sim T_{\mathbf{24}}$

  ① *à partir des observations faites sur l'échantillon de taille **n** = 25 dans quel intervalle a-t-on une "grande confiance" de trouver le score moyen?*

  ② *quelle est la précision dans l'estimation du score moyen sur l'échantillon de taille **n** = 25 ?*

Exemple variable normale (2)

- **Exemple : score de stress perçu**

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n*** *= 25*

➢ *intervalle de confiance à 90% (au risque 10%) du score moyen sur l'échantillon de taille* ***n*** *= 25 :*

$$\mathrm{IC}_{90\%}(\mu) = \left[\bar{x} \pm \mathrm{t}_{0,95} \frac{s^*}{\sqrt{n}} \right]$$
$$= \left[30 \pm 1{,}711 \frac{10}{\sqrt{25}} \right]$$
$$= [30 \pm 3{,}4] = [26{,}6\,;\,33{,}4]$$

où $\mathrm{t}_{0,95}$ = 1,711 *quantile d'ordre 95% de la loi* $\mathbf{T}_{24}$ (ν = 24 et P = 0,10)

① *on accorde une confiance de 90% au fait que l'intervalle [26,6 ; 33,4] contienne* μ *le score moyen dans* $\mathcal{P}$

② *la marge d'erreur à 90% dans l'estimation du score moyen sur l'échantillon de taille* ***n*** *= 25 est de 3,4*

Exemple variable normale (3)

- **Exemple : score de stress perçu**

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n*** *= 25*

➢ *intervalle de confiance à 95% (au risque 5%) du score moyen sur l'échantillon de taille* ***n*** *= 25 :*

$$IC_{95\%}(\mu) = \left[\overline{x} \pm t_{0,975}\frac{s^*}{\sqrt{n}}\right]$$

$$= \left[30 \pm 2{,}064\frac{10}{\sqrt{25}}\right]$$

$$= [30 \pm 4{,}1] = [25{,}9\ ;\ 34{,}1]$$

où $t_{0,975}$ = 2,064 *quantile d'ordre 97,5% de la loi* $\mathbf{T_{24}}$ (ν = 24 et P = 0,05)

① *on accorde une confiance de 95% au fait que l'intervalle [25,9 ; 34,1] contienne* μ *le score moyen dans* $\mathcal{P}$

② *la marge d'erreur à 95% dans l'estimation du score moyen sur l'échantillon de taille* ***n*** *= 25 est de 4,1*

Table de la loi de Student

Extrait de la table

ν	P=0,90	…	0,10	0,05	0,02	0,01
1	0,158		6,314	12,706	31,821	63,656
2	0,142		2,920	4,303	6,965	9,925
3	0,137		2,353	3,182	4,541	5,841
4	0,134		2,132	2,776	3,747	4,604
5	0,132		2,015	2,571	3,365	4,032
6	0,131		1,943	2,447	3,143	3,707
7	0,130		1,895	2,365	2,998	3,499
8	0,130		1,860	2,306	2,896	3,355
9	0,129		1,833	2,262	2,821	3,250
10	0,129		1,812	2,228	2,764	3,169
…						
21	0,127		1,721	2,080	2,518	2,831
22	0,127		1,717	2,074	2,508	2,819
23	0,127		1,714	2,069	2,500	2,807
24	0,127		1,711	2,064	2,492	2,797
25	0,127		1,708	2,060	2,485	2,787
26	0,127		1,706	2,056	2,479	2,779
27	0,127		1,703	2,052	2,473	2,771
28	0,127		1,701	2,048	2,467	2,763
29	0,127		1,699	2,045	2,462	2,756
30	0,127		1,697	2,042	2,457	2,750
∞	0,126		1,645	1,960	2,326	2,576

2.5 Variable normale d'écart-type inconnu (2)

Remarques

l'intervalle de confiance à (1−α) de μ :

$$\mathrm{IC}_{1-\alpha}(\mu) = \left[\bar{x} \pm t_{1-\frac{\alpha}{2}} \frac{s^*}{\sqrt{n}}\right] = [\bar{x} \pm \boldsymbol{e}]$$

- ***e*** la marge d'erreur à (1−α) dans l'estimation de la moyenne μ *varie*
- tous les intervalles de confiance de la moyenne μ sont centrés sur $\bar{x}$ la moyenne observée sur l'échantillon de taille ***n***
- l'intervalle de confiance dépend
  - **de la moyenne observée** $\bar{x}$
  - **de l'écart-type observé sans biais *s****
  - **du niveau (1−α) ou du risque** α
  - **de la taille de l'échantillon *n***
- l'intervalle de confiance ne dépend pas
  - **de la moyenne** μ **de X dans** $\mathcal{P}$

3. Intervalles de confiance d'une proportion

X variable qualitative dichotomique de $\mathcal{P}$ dans E = { oui, non }

⇒un paramètre

proportion de "*oui*" = *p* *inconnue* dans $\mathcal{P}$

échantillon de X issu de $\mathcal{P}$ de taille ***n***

- rechercher un intervalle dans lequel on a une confiance de $(1-\alpha)$ de trouver la proportion ***p*** à partir des observations de l'échantillon de taille ***n***

⇒cet intervalle est l'**intervalle de confiance** de la proportion ***p*** au niveau de confiance $(1-\alpha)$ ou au risque α noté $\mathbf{IC_{1-\alpha}}(\boldsymbol{p})$

3.1 Rappel : intervalles de variation d'une fréquence

X variable **qualitative** dichotomique de proportion de "*oui*" ***p*** ***connue*** dans $\mathcal{P}$

échantillon de **X** issu de $\mathcal{P}$ de taille $\boldsymbol{n \geq 30}$

$\boldsymbol{np \geq 5}$ et $\boldsymbol{n(1-p) \geq 5}$

- la fréquence empirique F_n suit *approximativement* un modèle normal de moyenne ***p*** et d'écart-type $\sqrt{\boldsymbol{p}(1-\boldsymbol{p})/\boldsymbol{n}}$

$$F_n \underset{approx}{\sim} \mathcal{N}\left(\boldsymbol{p}, \sqrt{\frac{\boldsymbol{p}(1-\boldsymbol{p})}{\boldsymbol{n}}}\right)$$

- l'intervalle de variation au niveau $(1-\alpha)$ ou au risque α de F_n s'écrit :

$$I_{1-\alpha}(F_n) \approx \left[\boldsymbol{p} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\boldsymbol{p}(1-\boldsymbol{p})}{n}}\right] \approx [\boldsymbol{p} \pm \boldsymbol{e}]$$

où ***e*** marge d'erreur à $(1-\alpha)$

$z_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de $\mathbf{Z} \sim \mathcal{N}(0,1)$

3.2 Intervalles de confiance d'une proportion (1)

X variable **qualitative** dichotomique
de proportion de "*oui*" $\boldsymbol{p}$ ***inconnue*** dans $\mathcal{P}$

échantillon de **X** issu de $\mathcal{P}$ de taille $\boldsymbol{n \geq 30}$

$\Rightarrow$ l'**intervalle de confiance au niveau (1−α)** ou **au risque** α de la **proportion** $\boldsymbol{p}$ s'écrit :

$$\mathrm{IC}_{1-\alpha}(\boldsymbol{p}) \approx \left[\boldsymbol{f} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\boldsymbol{f}(1-\boldsymbol{f})}{n}} \right] \approx [\boldsymbol{f} \pm \boldsymbol{e}]$$

où $\boldsymbol{f}$ fréquence observée de "*oui*"
$\boldsymbol{e}$ marge d'erreur à (1−α)
$z_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de $\mathbf{Z} \sim \mathcal{N}(0,1)$

conditions de validité

$$\mathrm{IC}_{1-\alpha}(\boldsymbol{p}) = [\boldsymbol{f}_i \,;\, \boldsymbol{f}_s]$$

$\boldsymbol{f_i}$ borne inférieure de $\mathrm{IC}_{1-\alpha}(\mathrm{p})$
$n\,\boldsymbol{f_i} \geq 5$ et $n(1-\boldsymbol{f_i}) \geq 5$

$\boldsymbol{f_s}$ borne supérieure de $\mathrm{IC}_{1-\alpha}(\mathrm{p})$
$n\,\boldsymbol{f_s} \geq 5$ et $n(1-\boldsymbol{f_s}) \geq 5$

3.2 Intervalles de confiance d'une proportion (2)

Cas particuliers

- **intervalle de confiance à 90%** ou **au risque 10%** de la proportion ***p*** :

$$\mathrm{IC}_{90\%}(\boldsymbol{p}) \approx \left[\boldsymbol{f} \pm z_{0,95} \sqrt{\frac{\boldsymbol{f}(1-\boldsymbol{f})}{n}} \right]$$

$$\approx \left[\boldsymbol{f} \pm 1,645 \sqrt{\frac{\boldsymbol{f}(1-\boldsymbol{f})}{n}} \right]$$

- **intervalle de confiance à 95%** ou **au risque 5%** de la proportion ***p*** :

$$\mathrm{IC}_{95\%}(\boldsymbol{p}) \approx \left[\boldsymbol{f} \pm z_{0,975} \sqrt{\frac{\boldsymbol{f}(1-\boldsymbol{f})}{n}} \right]$$

$$\approx \left[\boldsymbol{f} \pm 1,96 \sqrt{\frac{\boldsymbol{f}(1-\boldsymbol{f})}{n}} \right]$$

où $\boldsymbol{f}$ fréquence observée de "*oui*"

Exemple variable qualitative (1)

- **Exemple : efficacité d'un traitement**

  $\mathcal{P}$ = { enfants atteints de troubles de l'anxiété, sous traitement } ***N*** *= ?*

  X *= "amélioration clinique" : oui, non variable qualitative dichotomique*

  p *= proportion d'amélioration clinique* ***inconnue*** *dans $\mathcal{P}$*

  échantillon de ***X*** *issu de $\mathcal{P}$ de taille* ***n*** *= 63*

  ➢ *la proportion d'amélioration clinique* ***p*** *est estimée par la fréquence observée d'amélioration clinique* ***f***

  f *= 48 / 63 = 0,762*

  ① *à partir des observations de l'échantillon de taille* ***n*** *= 63 dans quel intervalle a-t-on une grande confiance de trouver la proportion d'amélioration clinique ?*

  ② *quelle est la précision de l'estimation de la proportion d'amélioration clinique* ***p*** *par la fréquence observée* ***f*** *sur l'échantillon de taille* ***n*** *= 63 ?*

Exemple variable qualitative (2)

- **Exemple : efficacité d'un traitement**

  échantillon de $\boldsymbol{X}$ *issu de* $\mathcal{P}$ *de taille* $\boldsymbol{n} = 63$ ($\boldsymbol{n} \geq 30$)

  ➢ *intervalle de confiance à 90% (au risque 10%) de la proportion d'amélioration clinique :*

  $$\mathrm{IC}_{90\%}(p) \approx \left[f \pm z_{0,95} \sqrt{\frac{f(1-f)}{n}} \right]$$

  $$\approx \left[0{,}762 \pm 1{,}645 \sqrt{\frac{0{,}762 \times 0{,}238}{63}} \right]$$

  $$\approx [0{,}762 \pm 0{,}088] = [0{,}674\,;0{,}85]$$

  conditions de validité :
  $f_i = 0{,}674$ *et* $(1 - f_s) = 0{,}15$
  $n\,(1 - f_s) = 63 \times 0{,}15 = 9{,}45 \geq 5$
  donc $n\,f_i \geq 5$ $n\,(1 - f_i) \geq 5$ et $n\,f_s \geq 5$
  et on peut supposer la normalité de F_n

  ① *on accorde une confiance de 90% au fait que l'intervalle [67,4% ; 85%] contienne* ***p*** *la proportion d'amélioration clinique dans* $\mathcal{P}$

  ② *la marge d'erreur à 90% dans l'estimation de la proportion d'amélioration clinique est de 8,8%*

Exemple variable qualitative (3)

- **Exemple : efficacité d'un traitement**

  *échantillon de **X** issu de $\mathcal{P}$ de taille **n** = 63 (**n** ≥ 30)*

  ➢ *intervalle de confiance à 95% (au risque 5%) de la proportion d'amélioration clinique :*

$$IC_{95\%}(p) \approx \left[f \pm z_{0,975} \sqrt{\frac{f(1-f)}{n}} \right]$$

$$\approx \left[0{,}762 \pm 1{,}96 \sqrt{\frac{0{,}762 \times 0{,}238}{63}} \right]$$

$$\approx [0{,}762 \pm 0{,}105] = [0{,}657\,;\,0{,}867]$$

  conditions de validité :
  f_i = 0,657 *et* (1– f_s) = 0,133
  n (1– f_s) = 63×0,133 = 8,4 ≥ 5
  donc n f_i ≥ 5 n (1– f_i) ≥ 5 et n f_s ≥ 5
  et on peut supposer la normalité de F_n

  ① *on accorde une confiance de 95% au fait que l'intervalle [65,7% ; 86,7%] contienne **p** la proportion d'amélioration clinique dans $\mathcal{P}$*

  ② *la marge d'erreur à 95% dans l'estimation de la proportion d'amélioration clinique est de 10,5%*

Exemple variable qualitative (4)

- **Exemple : efficacité d'un traitement**

*échantillon de **X** issu de $\mathcal{P}$ de taille*
$\boldsymbol{n} = 30$ ($\boldsymbol{n} \geq 30$)

➢ *la proportion d'amélioration clinique **p** est estimée par la fréquence d'amélioration observée **f***

$\boldsymbol{f} = 23 / 30 = 0{,}767$

➢ *intervalle de confiance à 95% (au risque 5%) de la proportion d'amélioration clinique :*

$$\mathrm{IC}_{95\%}(p) \approx \left[0{,}767 \pm 1{,}96\sqrt{\frac{0{,}767 \times 0{,}233}{30}} \right]$$

$$\approx [0{,}767 \pm 0{,}151] = [0{,}615\,;\,0{,}918]$$

conditions de validité :
$f_i = 0{,}615$ *et* $(1 - f_s) = 0{,}082$
$n\,(1 - f_s) = 30 \times 0{,}082 = 2{,}5 < 5$
donc on ne peut pas supposer la normalité de F_n

➔ *l'intervalle calculé n'est pas valide*

Exemple variable qualitative (5)

- **Exemple : efficacité d'un traitement**

  échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n*** *= 126 (****n*** $\geq$ *30)*

  ➢ *la proportion d'amélioration clinique* ***p*** *est estimée par la fréquence d'amélioration observée* ***f***

  f *= 96 / 126 = 0,762*

  ➢ *intervalle de confiance à 95% (au risque 5%) de la proportion d'amélioration clinique :*

$$IC_{95\%}(p) \approx \left[0{,}762 \pm 1{,}96\sqrt{\frac{0{,}762 \times 0{,}238}{126}}\right]$$

$$\approx [0{,}762 \pm 0{,}074] = [0{,}688\,;0{,}836]$$

  conditions de validité :

  $f_i = 0{,}688$ *et* $(1 - f_s) = 0{,}164$

  $n\,(1 - f_s) = 126 \times 0{,}164 = 20{,}7 \geq 5$

  donc $n\,f_i \geq 5$ $n\,(1 - f_i) \geq 5$ et $n\,f_s \geq 5$

  et on peut supposer la normalité de F_n

  ① *on accorde une confiance de 95% au fait que l'intervalle [68,8% ; 83,6%] contienne* ***p*** *la proportion d'amélioration clinique dans* $\mathcal{P}$

  ② *la marge d'erreur à 95% dans l'estimation de la proportion d'amélioration clinique est de 7,4%*

Remarques

l'intervalle de confiance à $(1-\alpha)$ de $\boldsymbol{p}$:

$$\mathrm{IC}_{1-\alpha}(\boldsymbol{p}) \approx \left[\boldsymbol{f} \pm \mathrm{z}_{1-\frac{\alpha}{2}} \sqrt{\frac{\boldsymbol{f}(1-\boldsymbol{f})}{n}} \right] \approx [\boldsymbol{f} \pm \boldsymbol{e}]$$

- $\boldsymbol{e}$ la marge d'erreur à $(1-\alpha)$ dans l'estimation de la proportion $\boldsymbol{p}$ *varie*
- tous les intervalles de confiance de la proportion $\boldsymbol{p}$ sont centrés sur $\boldsymbol{f}$ la fréquence observée sur l'échantillon de taille $\boldsymbol{n}$
- l'intervalle de confiance dépend
  - **de la fréquence observée *f***
  - **du niveau (1–α) ou du risque α**
  - **de la taille de l'échantillon *n***
- l'intervalle de confiance ne dépend pas
  - **de la proportion *p* dans $\mathcal{P}$**

Interprétation d'un intervalle de confiance (1)

- une fréquence observée f est *variable* : elle change pour chaque échantillon
- un intervalle de confiance est *variable* : ses bornes sont *variables*, changent pour chaque échantillon de taille ***n***
- les intervalles de confiance au niveau $(1-\alpha)$ (au risque α) sur les échantillons de taille ***n*** sont tels que :
  - $(1-\alpha)\%$ des intervalles contiennent la proportion ***p*** de "*oui*" dans $\mathcal{P}$

  et

  - $\alpha\%$ des intervalles ne contiennent pas la proportion ***p*** de "*oui*" dans $\mathcal{P}$

Interprétation d'un intervalle de confiance (2)

- les intervalles de confiance au niveau 90% (au risque 10%) sont tels que :
  - **90% des intervalles de confiance à 90% contiennent la proportion *p***

  et

  - **10% des intervalles de confiance à 90% ne contiennent pas la proportion *p***
- par construction, par exemple sur **100** échantillons de taille $\boldsymbol{n}$ de $\mathbf{X}$ issus de $\mathcal{P}$, en moyenne :
  - **90 intervalles de confiance à 90% contiendront la proportion *p***

  et

  - **10 intervalles de confiance à 90% ne contiendront pas la proportion *p***

Interprétation d'un intervalle de confiance (3)

pour un intervalle de confiance à $(1-\alpha)$ de la proportion ***p***, deux possibilités :

- **soit il contient la proportion *p***
- **soit il ne contient pas la proportion *p***

mais si la proportion ***p*** est *inconnue* on ne sait pas dans quelle situation on se trouve

➔ on accorde une confiance de $(1-\alpha)$ au fait que l'intervalle de confiance au niveau $(1-\alpha)$ de la proportion ***p*** contienne la proportion ***p***

➔ l'intervalle de confiance au niveau $(1-\alpha)$ de la proportion ***p*** contient la proportion ***p*** avec un niveau de confiance de $(1-\alpha)$ (avec un risque α)

3.3 Précision dans l'estimation d'une proportion

La demi-longueur $\boldsymbol{e}$ de l'intervalle de confiance à $(1-\alpha)$ représente la **précision** à $(1-\alpha)$ dans l'estimation de la proportion $\boldsymbol{p}$ sur l'échantillon de taille $\boldsymbol{n}$, ou la **marge d'erreur** au risque α

$$\boxed{\boldsymbol{e} = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\boldsymbol{f}(1-\boldsymbol{f})}{\boldsymbol{n}}}}$$

où $z_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de $\mathbf{Z} \sim \mathcal{N}(0,1)$

- la marge d'erreur $\boldsymbol{e}$ dépend de la fréquence observée $\boldsymbol{f}$
- pour $\boldsymbol{n}$ fixé, la précision $\boldsymbol{e}$ *change* lorsque l'échantillon change
- pour $\boldsymbol{n}$ fixé, plus le niveau $(1-\alpha)$ augmente (le risque α diminue) plus la marge d'erreur $\boldsymbol{e}$ augmente
- plus la taille de l'échantillon $\boldsymbol{n}$ augmente, plus la marge d'erreur $\boldsymbol{e}$ diminue

3.4 Taille minimum de l'échantillon pour une précision minimum (1)

- Pour contrôler la marge d'erreur et obtenir une marge d'erreur à $(1-\alpha)$ ***inférieure à*** ***e*** sur l'estimation de la proportion ***p***
  la longueur (amplitude) de l'intervalle de confiance au niveau $(1-\alpha)$ ou au risque α n'excède pas **2*e***

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} \leq e$$

⇒**la taille de l'échantillon *n* vérifie :**

$$\boxed{n \geq f\,(1-f)\left(\frac{z_{1-\frac{\alpha}{2}}}{e}\right)^2}$$

➔ pour une marge d'erreur maximum ***e*** (ou une précision minimum) il faut une taille d'échantillon d'au moins ***n***

3.4 Taille minimum de l'échantillon pour une précision minimum (2)

Remarques

- la taille minimum $\boldsymbol{n}$ dépend de
  - **la fréquence observée *f* de "*oui*"**
  - **la marge d'erreur maximum e**
  - **le niveau (1–α) ou risque α**
- la taille minimum $\boldsymbol{n}$ augmente lorsque
  - **le produit *f* (1–*f*) augmente**
  - **le niveau (1–α) augmente**
le risque α diminue
  - **la marge d'erreur maximum e diminue**
la précision minimum e augmente
- vérifier que pour la taille $\boldsymbol{n}$ obtenue

$$\boldsymbol{n} \geq 30 \qquad \boldsymbol{nf} \geq 5 \qquad \boldsymbol{n}(1-\boldsymbol{f}) \geq 5$$

<u>Exemple variable qualitative</u> (1)

- **Exemple : efficacité d'un traitement**

$\mathcal{P}$ = *{ enfants atteints de troubles de l'anxiété, sous traitement }* ***N*** *= ?*

X *= "amélioration clinique" : oui, non*
variable qualitative dichotomique

p *= proportion d'amélioration clinique* ***inconnue*** *dans* $\mathcal{P}$

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n***

on suppose que :
la fréquence observée ***f*** *= 0,762*

➢ *pour* ***n*** *= 63 l'intervalle de confiance à 90% (au risque 10%) de la proportion d'amélioration clinique :*

$$\mathrm{IC}_{90\%}(p) \approx [0{,}762 \pm 0{,}088]$$

➔ *la marge d'erreur à 90% dans l'estimation de la proportion d'amélioration clinique est de 8,8%*

③ *quelle est la taille minimum de l'échantillon pour avoir une précision à 90% d'au minimum 10% dans l'estimation de la proportion d'amélioration clinique ?*

Exemple variable qualitative (2)

- **Exemple : efficacité d'un traitement**

  *échantillon de **X** issu de $\mathcal{P}$ de taille **n***

  *on suppose que **f** = 0,762*

  ➢ *marge d'erreur à 90% d'au plus 10%*

  ➔ *pour que la marge d'erreur à 90% dans l'estimation de la proportion d'amélioration clinique n'excède pas 10% un échantillon d'**au moins** 63 enfants convient*

  ➢ *taille minimum requise*

$$\mathrm{IC}_{90\%}(p) \approx \left[0{,}762 \pm 1{,}645\sqrt{\frac{0{,}762 \times 0{,}238}{n}}\right]$$

$$\approx [0{,}762 \pm 0{,}1]$$

$$1{,}645\sqrt{\frac{0{,}762 \times 0{,}238}{n}} \leq 0{,}1$$

$$\boldsymbol{n} \geq 0{,}762 \times 0{,}238 \left(\frac{1{,}645}{0{,}1}\right)^2 = 49{,}1$$

$\boldsymbol{n} = 50 \geq 30$

$\boldsymbol{nf} = 50 \times 0{,}762 = 38{,}1 \geq 5$

$\boldsymbol{n}(1-\boldsymbol{f}) = 50 \times 0{,}238 = 11{,}9 \geq 5$

③ *pour que la marge d'erreur à 90% dans l'estimation de la proportion d'amélioration clinique n'excède pas 10% il faudra un échantillon d'**au moins** 50 enfants*

Exemple variable qualitative (3)

- **Exemple : efficacité d'un traitement**

  *échantillon de **X** issu de $\mathcal{P}$ de taille **n***

  on suppose que :
  *la fréquence observée **f** = 0,762*

  ➢ *marge d'erreur à 90% d'au plus 5%*

$$IC_{90\%}(p) \approx \left[0{,}762 \pm 1{,}645 \sqrt{\frac{0{,}762 \times 0{,}238}{n}} \right]$$
$$\approx [0{,}762 \pm 0{,}05]$$

$$1{,}645 \sqrt{\frac{0{,}762 \times 0{,}238}{n}} \leq 0{,}05$$

$$\boldsymbol{n} \geq 0{,}762 \times 0{,}238 \left(\frac{1{,}645}{0{,}05} \right)^2 = 196{,}3$$

$\boldsymbol{n} = 197 \geq 30$
$\boldsymbol{nf} = 197 \times 0{,}762 = 150{,}1 \geq 5$
$\boldsymbol{n}(1-\boldsymbol{f}) = 197 \times 0{,}238 = 46{,}9 \geq 5$

③ *pour que la marge d'erreur à 90% dans l'estimation de la proportion d'amélioration clinique n'excède pas 5% il faudra un échantillon d'**au moins** 197 enfants*

<u>Exemple variable qualitative</u> (4)

- **Exemple : efficacité d'un traitement**

*échantillon de **X** issu de $\mathcal{P}$ de taille **n***

on suppose que :
*la fréquence observée **f** = 0,762*

➢ *marge d'erreur à 90% d'au plus 1%*

$$\mathrm{IC}_{90\%}(p) \approx \left[0{,}762 \pm 1{,}645\sqrt{\frac{0{,}762 \times 0{,}238}{n}}\right]$$

$$\approx [0{,}762 \pm 0{,}01]$$

$$1{,}645\sqrt{\frac{0{,}762 \times 0{,}238}{n}} \leq 0{,}01$$

$$\boldsymbol{n} \geq 0{,}762 \times 0{,}238\left(\frac{1{,}645}{0{,}01}\right)^2 = 4907{,}5$$

$\boldsymbol{n} = 4908 \geq 30$
$\boldsymbol{nf} = 4908 \times 0{,}762 \geq 5$
$\boldsymbol{n}(1-\boldsymbol{f}) = 4908 \times 0{,}238 \geq 5$

③ *pour que la marge d'erreur à 90% dans l'estimation de la proportion d'amélioration clinique n'excède pas 1% il faudra un échantillon d'**au moins** 4 908 enfants*