

Cours n° 4 : L'INTERVALLE DE CONFIANCE.

La notion d'intervalle de confiance renvoie au degré de précision d'une moyenne ou d'un pourcentage. On travaille sur un échantillon et l'on souhaite connaître la fiabilité que l'on peut accorder aux valeurs observées par rapport aux valeurs réelles de la population totale.

A : L'intervalle de confiance d'une moyenne :

Par exemple, avec une population mère (cad totale) de plus de 500 personnes, nous avons calculé à partir d'un échantillon de 30 personnes la moyenne d'âge qui est de 25 ans (avec un écart-type de 2 ans). La moyenne d'âge de l'ensemble des personnes diffère probablement de cette moyenne observée. Il y a une marge d'erreur que nous pouvons estimer avec le calcul de l'intervalle de confiance (avec 5% de risque d'erreur ou 95 % de certitude ou de confiance)

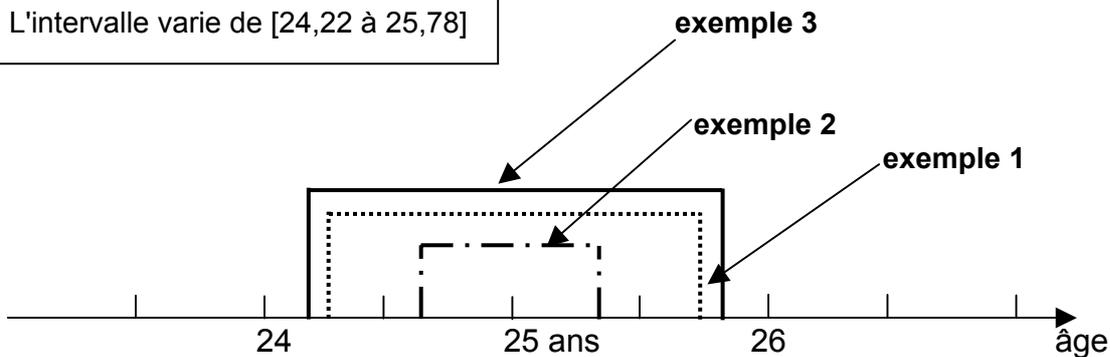
Définition: l'observation d'une moyenne m sur un échantillon de n personnes permet de calculer une moyenne inconnue située dans l'intervalle défini par:

$$m \pm \frac{1,96 \times \sigma}{\sqrt{n}}$$

<p>Exemple 1 : Avec $n = 30$ $m = 25$ $\sigma = 2,0$</p> $25,0 \pm \frac{1,96 \times 2}{\sqrt{30}}$ $25,0 \pm \frac{3,92}{5,48}$ $25,0 \pm 0,72$ <p>L'intervalle varie donc de [24,28 à 25,72]</p>	<p>Exemple 2 : idem mais avec $n = 100$</p> $25,0 \pm \frac{1,96 \times 2}{\sqrt{100}}$ $25,0 \pm \frac{3,92}{10}$ $25,0 \pm 0,39$ <p>L'intervalle varie donc de [24,61 à 25,39]</p>
--	--

L'intervalle se resserre au fur et à mesure que la taille de l'échantillon augmente (sans aller jusqu'à la population totale : dans ce cas, on a la moyenne et donc pas besoin d'intervalle de confiance).

<p>Exemple 3 : avec $\sigma = 4$ et $n = 100$</p> $25,0 \pm 0,78$ <p>L'intervalle varie de [24,22 à 25,78]</p>



n.b.: si nous avions voulu prendre 1 % de risque d'erreur (ou une estimation basée sur 99 % de confiance) nous aurions pris :

$$m \pm \frac{2,576 \times \sigma}{\sqrt{n}}$$

nb: les valeurs des coefficients 1,96 et 2,576 proviennent de la table de l'écart-réduit pour un risque d'erreur respectivement de 5 % et 1 %.

B : L'intervalle de confiance d'un pourcentage

Soit P_o le pourcentage observé et n le nombre d'observations (nombre de personnes à qui on a posé une question)

P_o = pourcentage observé, et Q_o = le pourcentage complémentaire de P_o .

Exemple : $P_o = 25$ % de filles $\implies Q_o = 100 - 25 = 75$ % de garçons.

Définition: l'observation d'un pourcentage P_o sur un échantillon de n personnes permet de calculer une pourcentage inconnu qui sera située dans l'intervalle défini par:

$$P_o \pm 1,96 \times \sqrt{\frac{P_o \times Q_o}{n}}$$

Exemple 1 :

Lors d'un cours, sur un échantillon de 32 personnes, j'observe 8 fumeurs. A combien puis-je estimer le pourcentage de fumeurs parmi l'ensemble des étudiants de la faculté ?

Solution : $n = 32$ $P_o = 25$ % de fumeurs

1) Calcul de $Q_o = 100 - P_o = 100 - 25 = 75$

2) **Formule** : $25 \pm 1,96 \times \sqrt{\frac{25 \times 75}{32}}$

3) $P_o \times Q_o = 25 \times 75 = 1875$

4) $\frac{P_o \times Q_o}{n} = \frac{1875}{32} = 58,59$

5) $\sqrt{58,59} = 7,65$

6) $1,96 \times 7,65 = 15,00$

Donc l'intervalle de confiance est de : $25\% \pm 15,0$ soit IC : [10,0 % à 40,0 %]

d'où un intervalle compris entre [10,0 % et 40,0 %]

Cet intervalle est immense : il "mesure" **30** %. Est-ce dû à la faible taille de l'échantillon ?

Nous allons prendre le même pourcentage en admettant qu'il a été observé sur 100 personnes (et non 32).

Exemple 2 :

n = 100 25% de fumeurs **Formule :** $25 \pm 1,96 \times \sqrt{\frac{25 \times 75}{100}}$

On arrive à $25\% \pm 8,49\%$. On aura un intervalle IC : [16,51 % à 33,49 %]
On a gagné en précision avec cet échantillon n = 100 : l'intervalle se resserre.

Remarque : la plupart des sondages d'opinion sont faits sur un échantillon n = 1000. Quel est l'intervalle de confiance que l'on peut accorder à ces pourcentages ?

d'où un intervalle compris entre [16,51 % et 33,49 %]

Avec n = 100, l'intervalle passe de 30 % à **17 %** environ.

n.b.: dans les sondages, quand on parle par exemple de cote de popularité en hausse ou en baisse de 1%, ce n'est pas parfaitement fiable car s'ils reposent sur un nombre satisfaisant de personnes (de l'ordre de 1000), l'intervalle de confiance dépend de la proportion P_o / Q_o comme nous allons le voir ci-dessous.

Exemple 3 :

Nous allons faire varier au maximum le rapport P_o / Q_o afin de mesurer l'impact de ce rapport sur l'intervalle de confiance.

1er cas : $P_o = 1\%$ avec n = 1000 personnes.

$$1\% \pm [1,96 \times \sqrt{\frac{1 \times 99}{1000}}]$$

$$1\% \pm [1,96 \times 0,31]$$

$$1\% \pm 0,62 \quad \boxed{\text{d'où un intervalle compris entre [0,38 % à 1,62 %]}}$$

Cet intervalle "mesure" **1,23 %**.

2ème cas : $P_o = 50\%$ avec n = 1000 personnes.

$$50\% \pm 1,96 \times \sqrt{\frac{50 \times 50}{1000}}$$

$$50\% \pm [1,96 \times 1,58]$$

$$50\% \pm 3,10 \quad \boxed{\text{d'où un intervalle compris entre [46,9 % à 53,1 %]}}$$

Cet intervalle "mesure" **6,20 %**.

* * *