ESTIMATION

I. Introduction :

1°) Vocabulaire :

Définitions :

Lorsqu'on étudie une partie de la population, on dit qu'on étudie un **échantillon**.

Le nombre d'individus formant l'échantillon est appelé **taille de l'échantillon**. Nous le noterons n.

2°) Introduction :

Introduction :

Dans ce chapitre, on est dans le cas où on ne connaît pas les caractéristiques (proportion et/ou moyenne) d'une population et on cherche à en donner une estimation à partir de l'étude d'échantillons.

II. Estimation ponctuelle :

Propriété :

Connaissant la moyenne $\overline{x}$ et l'écart-type σ_s d'un échantillon de taille n, il s'agit d'estimer la moyenne m et l'écart-type σ de la population totale. L'estimation ponctuelle consiste à confondre la moyenne de l'échantillon avec la moyenne de la population totale. On dira que $\overline{x}$ est une estimation ponctuelle de la moyenne m.

En choisissant la variance σ_s^2 d'un échantillon prélevé au hasard comme estimation ponctuelle de la variance inconnue σ^2 d'une population, on sous-estime la variance de la population.

On choisit donc:

- la moyenne $\overline{x}$ comme estimation ponctuelle de la moyenne m ;
- l'écart type $\sqrt{\dfrac{n}{n-1}}\sigma_s$ comme estimation ponctuelle de l'écart-type inconnu σ de cette population.

Remarque :

Le nombre $\sqrt{\dfrac{n}{n-1}}\sigma_s$ est fourni directement par la calculatrice : Sx ou $\sigma n - 1$.

III. Estimation par intervalle de confiance :

1°) Estimation d'une proportion, dans le cas d'une loi binomiale approximable par une loi normale :

Propriété :

Nommons p la proportion dans la population (donc inconnue ici) et f la proportion obtenue dans l'échantillon.

Pour tout réel α dans $]0\,;\,1[$, l'intervalle $\mathrm{I}=\left[f-u_\alpha\dfrac{\sqrt{f(1-f)}}{\sqrt{n}}\,;\,f+u_\alpha\dfrac{\sqrt{f(1-f)}}{\sqrt{n}}\right]$ où u_α désigne le nombre réel tel que $P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1 - \alpha$ lorsque Z suit la loi normale $\mathcal{N}(0\,;\,1)$ est un intervalle de confiance de la proportion p au seuil de $1 - \alpha$.

Remarques :

- $1 - \alpha$ est appelé le coefficient de confiance ;
- Si $\alpha = 0{,}05$, c'est-à-dire 5 %, cela signifie que dans 95 % des cas la proportion p cherchée est dans I ;
- Cet intervalle est centré en f.

Exemple, Méthode d'utilisation :

On voudrait estimer la proportion de téléspectateurs ayant regardé hier soir une émission sur un célèbre mathématicien, avec un coefficient de confiance de 97 %. Pour ce faire, on interroge 900 téléspectateurs tirés au hasard. Sur ces 900 personnes, 240 déclarent avoir vu l'émission.

On obtient donc $f \approx 0{,}267$.

On cherche ensuite $u_{0,03}$. $P(-u_{0,03} \leqslant Z \leqslant u_{0,03}) = 0{,}97$, donc $P(Z \leqslant u_{0,03}) = 0{,}985$. En tapant FracNormal(0.985,0,1) à la calculatrice, on obtient $u_{0,03} \approx 2{,}17$.

Ce qui donne $\mathrm{I} = \left[0{,}267 - 2{,}17\frac{\sqrt{0{,}267(1-0{,}267)}}{\sqrt{900}} ; 0{,}267 + 2{,}17\frac{\sqrt{0{,}267(1-0{,}267)}}{\sqrt{900}}\right] \approx [0{,}235 ; 0{,}299]$.

La proportion de téléspectateurs ayant regardé cette émission est donc entre 23,5 % et 29,9 % au seuil de confiance de 97 %.

Remarque :

Deux seuils de confiance fréquemment utilisés sont ceux de 95 % et 99 %, il peut donc être intéressant de connaître par cœur $u_{0,05} \approx 1{,}96$ et $u_{0,01} \approx 2{,}58$.

2°) Estimation d'une moyenne, dans le cas d'une loi normale quand l'écart type de la population est connu :

Propriété :

Nommons m la moyenne dans la population (donc inconnue ici) et $\bar{x}$ la moyenne obtenue dans l'échantillon.

Pour tout réel α dans $]0 ; 1[$, l'intervalle $\mathrm{I} = \left[\bar{x} - u_\alpha\frac{\sigma}{\sqrt{n}} ; \bar{x} + u_\alpha\frac{\sigma}{\sqrt{n}}\right]$ où u_α désigne le nombre réel tel que $P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1 - \alpha$ lorsque Z suit la loi normale $\mathcal{N}(0 ; 1)$ est un intervalle de confiance de la moyenne m au seuil de $1 - \alpha$.

3°) Précision de l'intervalle de confiance :

Exemple :

Reprenons l'exemple des téléspectateurs précédent, mais cette fois on se demande combien de personnes on devrait interroger pour que l'erreur de notre estimation soit de moins de 1 %, c'est-à-dire pour que l'amplitude de l'intervalle de confiance soit inférieure à 0,01. Rappelons que l'intervalle de confiance est $\mathrm{I} = \left[0{,}267 - 2{,}17\frac{\sqrt{0{,}267(1-0{,}267)}}{\sqrt{n}} ; 0{,}267 + 2{,}17\frac{\sqrt{0{,}267(1-0{,}267)}}{\sqrt{n}}\right]$, car n est inconnue cette fois. Comme $2{,}17\sqrt{0{,}267(1-0{,}267)} \approx 0{,}96$, on obtient $\mathrm{I} = \left[0{,}267 - \frac{0{,}96}{\sqrt{n}} ; 0{,}267 + \frac{0{,}96}{\sqrt{n}}\right]$.

Son amplitude est de $\frac{2 \times 0{,}96}{\sqrt{n}} = \frac{1{,}92}{\sqrt{n}}$, on veut donc $\frac{1{,}92}{\sqrt{n}} < 0{,}01$, soit $\frac{1{,}92}{0{,}01} < \sqrt{n} \Leftrightarrow 192 < \sqrt{n}$, et donc, comme $n > 0$, $n > 192^2$ soit $n > 36\,864$.

Il faudrait donc interroger au moins 36 864 téléspectateurs pour que l'erreur de l'estimation soit inférieure à 1 %.