Estimations

Problématique.
A partir d'observations faites sur un échantillon, on se propose de tirer des conclusions sur la population toute entière. Ainsi contrairement à la logique déductive, qui va du général au particulier, qui manipule de concepts généraux et qui permet de prouver des propositions , la démarche statistique inductive est beaucoup plus téméraire et ambitieuse puisqu'elle a pour but de fournir des réponses précises (et de portée générale) à partir d'informations imparfaites (partielles).
On dispose d'un échantillon de taille n, de moyenne m et d'écart type s .
Cet échantillon provient d'un population « parente » .
On cherche à donc « inférer », (à prévoir) la moyenne μ et l'écart type σ de la population parente.

A. Estimation d'une moyenne

Estimation « ponctuelle » :

a. si la moyenne de l'échantillon est m,
on inférera que la moyenne μ de la population parente est aussi m .

On démontre que m est un estimateur sans biais de la moyenne de la population

Estimation « par intervalle » :

a. si la moyenne de l'échantillon est m,
on inférera que la moyenne μ de la population parente appartient au seuil de 5% à l'intervalle $m - 1{,}96.\frac{s'}{\sqrt{n}} < \mu < m + 1{,}96.\frac{s'}{\sqrt{n}}$

Théorie:

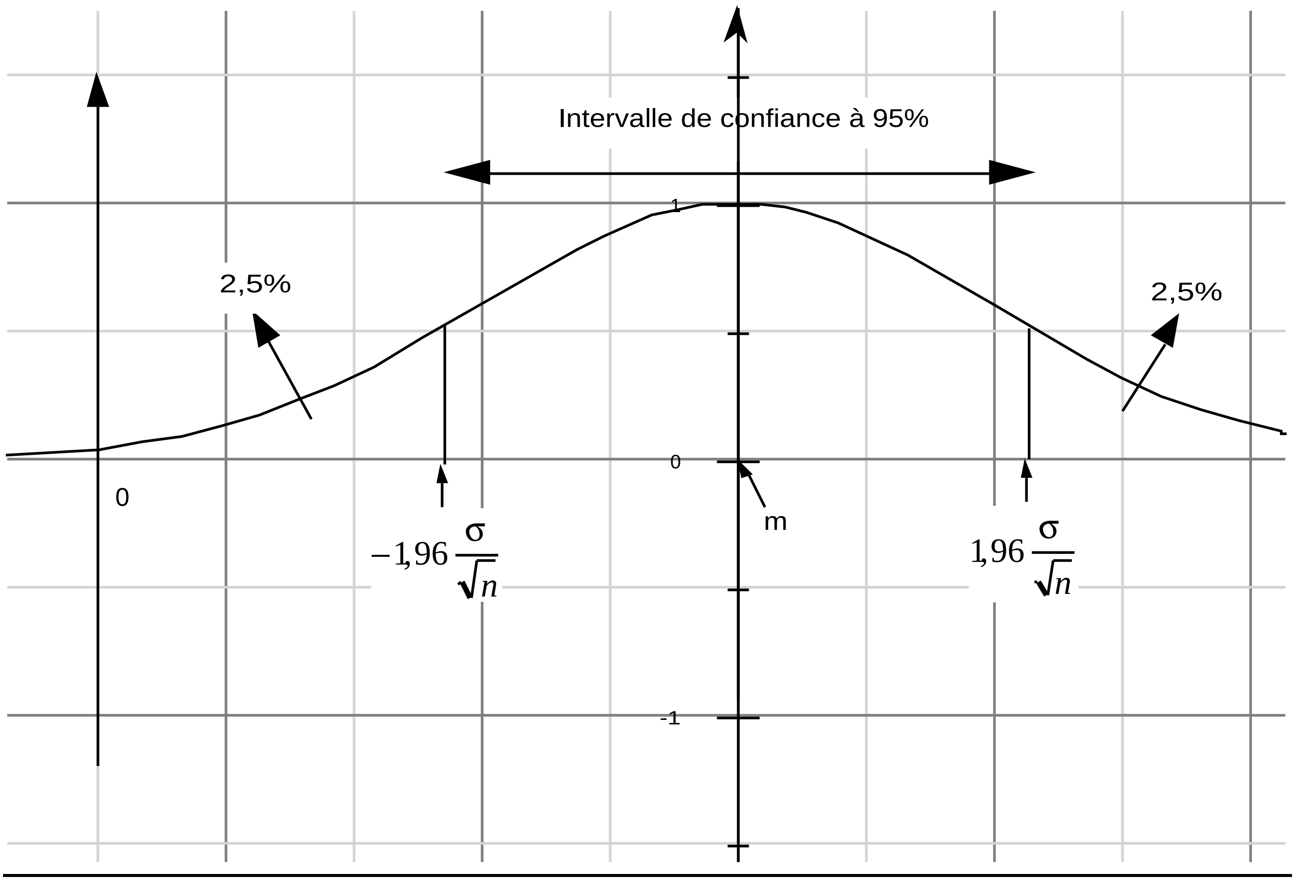
Les Moyennes des échantillons suivent une loi normale :

$$\overline{X} \; de \text{ moyenne } \mu_{\overline{X}} = m \text{ et d'écart-type } \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Remarques:

• si l'écart type σ de la population n'est pas connu on remplacera σ par l'estimateur $s' = s.\sqrt{\frac{n}{n-1}}$

• la valeur 1,96 est obtenue par lecture dans la table de la loi normale réduite.
Elle correspond au seuil de 5%
• la valeur 2,57 est obtenue par lecture dans la table de la loi normale réduite.
Elle correspond au seuil de 1%
• la valeur 0,66 est obtenue par lecture dans la table de la loi normale réduite.
Elle correspond au seuil de 50% ...



Exercice n° 1:

Dans une usine employant 20000 ouvriers, on a fait un sondage portant sur 100 ouvriers. On a trouvé un salaire moyen horaire de 30 F, avec un écart type de 1,50F.
Estimer la moyenne vraie.

$$n = 100 \quad s' = 1{,}5.\sqrt{\frac{100}{99}} \approx 1{,}5.1{,}005 = 1{,}507$$

$$30 - 1{,}96.\frac{1{,}507}{\sqrt{100}} < \mu < 30 + 1{,}96.\frac{1{,}507}{\sqrt{100}}$$

$$29{,}70 < \mu < 30{,}29$$

Au seuil de 5% la moyenne apprtient à l'intervalle $]29{,}70 \ : 30{,}29[$

B. Estimation d'une proportion ou d'un pourcentage

Nous disposons d'un échantillon de taille n ou la fréquence de la caractéristique (proportion des éléments ayant la caractéristique donnée) est f.

Estimation « ponctuelle » :

Si la fréquence d'apparition de la caractéristique est f,

on inférera que la proportion p est égale à f
On a ainsi un estimateur sans biais et convergent .

Estimation « par intervalle » :

Si la fréquence d'apparition de la caractéristique est f,
on inférera que la proportion p appartient à l'intervalle

$$f-1{,}96.\sqrt{\frac{f.(1-f)}{n}} < p < f+1{,}96.\sqrt{\frac{f.(1-f)}{n}}$$

Théorie:

La distribution des fréquences suit une loi normale

$$\overline{F} \; de \text{ moyenne } \mu_{\overline{F}} = p \text{ et d'écart-type } \sigma_{\overline{F}} = \sqrt{\frac{p.(1-p)}{n}}$$

Remarques:
• La proportion théorique p n'étant pas connue on remplacera p par f si les conditions suivantes sont réalisées:

- la taille de l'échantillon est « grand » --> n>30
- le produit p.(1-p)<0,25 ou bien (p <0,1 ou p>0,9)

Dans ce cas la distribution des fréquences suit une loi normale
• la valeur 1,96 est obtenue par lecture dans la table de la loi normale réduite.
Elle correspond au seuil de 5%
• la valeur 2,57 est obtenue par lecture dans la table de la loi normale réduite.
Elle correspond au seuil de 1%
• la valeur 0,66 est obtenue par lecture dans la table de la loi normale réduite.
Elle correspond au seuil de 50% ...

Exercice n° 3:

Dans un lot de 50 pièces, on a trouvé 12 pièces de « qualité extra ».
Estimer au seuil de 5%, la qualité de la fabrication.

$$f = \frac{12}{50} = 0{,}24 \quad \text{d'où}$$

$$0{,}24 - 1{,}96.\sqrt{\frac{0{,}24.(1-0{,}24)}{50}} < p < 0{,}24 + 1{,}96.\sqrt{\frac{0{,}24.(1-0{,}24)}{50}}$$

Au seuil de 5% on peut estimer que la proportion p de pièces de qualité extra est comprise entre 0,12 et 0, 36

Exercice n° 4:

Une machine a été observée 325 fois et on a noté: en marche 260 fois arrêt 65 fois.

Estimer sur cette base la qualité de la fabrication.

On peut remplacer p par f car " n grand" et pq $< 0,25$

$$f = \frac{260}{325} = 0,8 \quad \text{d' où}$$

$$0,8 - 1,96.\sqrt{\frac{0,8.0,2}{325}} < p < 0,8 + 1,96.\sqrt{\frac{0,8.0,2}{325}}$$

Au seuil de 5% on peut estimer que le taux d' utilisation de cette machine est comprise entre $0,757$ et $0,843$

Exercice n° 5:

Dans une population de 10000 élèves, 20% étudient l'espagnol.
On prend un échantillon de 100 élèves, quel est la probabilité pour que le pourcentage d'élèves de l'échantillon soit compris entre 18% et 22%?

$p = 0,2$ n et "grand" on va considérer que le pourcentage suit une loi normale

On effectue un test " bilatéral" car $0,20 - 0,02 < 0,20 < 0,20 + 0,02$

$$0,2 - a.\sqrt{\frac{0,2.0,8}{100}} < X < 0,2 + a.\sqrt{\frac{0,2.0,8}{100}}$$

$0,2 - a.0,04 = 0,18 \quad -- > a = 0,5$

d' où $\text{Prob}(0,18 < X < 0,22) = 0,5$

C. Estimation d'une variance

Estimation « ponctuelle » :

a. si la variance de l'échantillon est s^2,

on inférera que la variance ρ^2 de la population parente est :

$$\rho^2 = s^2.\frac{n}{n-1}$$

on a ainsi un estimateur sans biais et convergent.

D. Estimation d'un écart type

Estimation « ponctuelle » :

a. si la variance de l'échantillon est s^2,

on inférera que la variance ρ de la population parente est :

$$\rho = s.\sqrt{\frac{n}{n-1}}$$

on a ainsi un estimateur avec biais

Remarque:
Tn est un estimateur sera dit "non biaisé" A si l'espérance mathématique de Tn est nulle.