Résumé : Statistique 2

Module 1

Rappels

Quelques rappels de statistique descriptive :

- **Moyenne** : $\overline{x} = \frac{1}{n}\sum r_i.x_i$
- **Variance (Unités²)** : $Var(x) = \overline{x^2} - \overline{x}^2 = \frac{1}{n}.\sum r_i.(x_i)^2 - (\overline{x})^2 = \frac{1}{n}\sum r_i.(x_i - \overline{x})^2$
- **Ecart-Type** : $\sigma(x) = \sqrt{Var(x)}$

Estimation de paramètres

Cette partie du cours est destinée à utiliser un estimateur(échantillon) pour une information sur la population.

Un estimateur est d'autant plus efficace que sa variance est petite.

Un estimateur est **absolument correct** s'il est **sans biais** ($E(\hat{\theta}) = \theta$) et **consistant** (au plus l'échantillon de personnes interrogées grandit, au plus $\hat{\theta}$ se rapproche de θ.

Notations

Variable	Estimateur
p	f
m	$\overline{x}$
σ^2	S_n^2

> N: *taille de la population*
>
> n: *taille de l'échantillon*
>
> m : *moyenne pour la population*
>
> $\overline{x}$: *moyenne pour l'échantillon*
>
> p: *proportion pour la population*
>
> f : *fréquence pour l'échantillon*
>
> σ^2: *variance pour la population*
>
> S_n^2 : *variance pour l'échantillon*

Non-Exhaustif

- $E(\overline{x}) = m$
- $Var(\overline{x}) = \frac{\sigma^2}{n}$
- $E(f) = p$

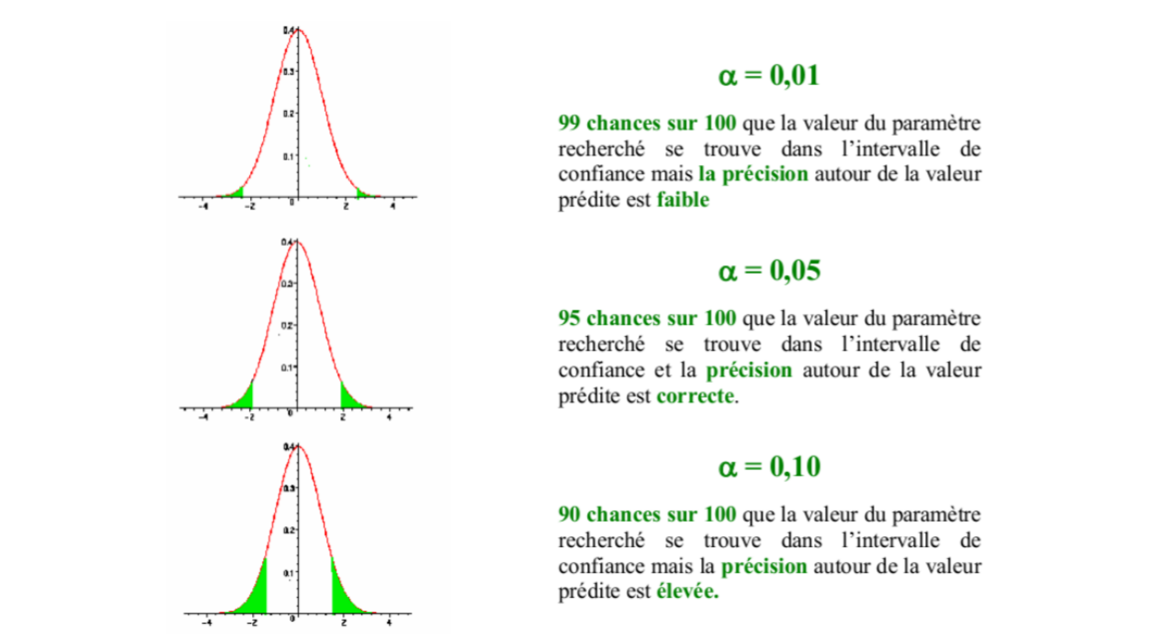
- $Var(f) = \frac{p.(1-p)}{n}$
- $E(S_n^2) = \frac{n-1}{n}.\sigma^2$
- $E(S_{n-1}^2) = \sigma^2 \quad (S_{n-1}^2 = \frac{n}{n-1}.S_n^2)$

Exhaustif

- $E(\overline{x}) = m$
- $Var(\overline{x}) = \frac{\sigma^2}{n}.\frac{N-n}{N-1}$
- $E(f) = p$
- $Var(f) = \frac{p(1-p)}{n}.\frac{N-n}{N-1}$
- $E(S_n^2) = \frac{n-1}{n}.\frac{N}{N-1}.\sigma^2$
- $E(\frac{N-1}{N}.S_{n-1}^2) = \sigma^2$

Estimation par intervalle de confiance

On souhaite donc utiliser cet estimateur pour .... estimer (LUL) un paramètre. Cet estimation à un coût en termes de précision. Le but est donc de déterminer un intervalle dans lequel on estime que le paramètre à $(1-\alpha)$ [1] chances de se trouver.



La zone colorée représente la **marge d'erreur tolérée**.

$Pr\{L_i \leq \theta \leq L_s\} = 1 - \alpha$

L'intervalle (si $n.p > 5$ et $n > 30$) vaut

> g est la valeur dans la table de gauss de $1 - \frac{\alpha}{2}$ (α de 0.05 donne un g de 1.96)

Echantillon non exhaustif

$[f - g.\sqrt{\frac{p(1-p)}{n}}; f + g.\sqrt{\frac{p(1-p)}{n}}]$

Echantillon exhaustif

$\left[f - 1.96.\sqrt{\frac{N-n}{N-1}}.\sqrt{\frac{p.(1-p)}{n}}; f + 1.96.\sqrt{\frac{N-n}{N-1}}.\sqrt{\frac{p.(1-p)}{n}}\right]$

Remplacer f

f remplacé par 0.5 pour maximiser le risque d'erreur.

Niveau de confiance n'existe pas dans la table

$a \approx g^{-} + \frac{a-a^{-}}{a^{+}-a^{-}}.(g^{+} - g^{-}) = g$

Intervalle de confiance d'une moyenne

Grande taille (n>30)

Non-Exhaustif

$$\left[\overline{x} - g.\frac{\sigma}{\sqrt{n}}; \overline{x} + g.\frac{\sigma}{\sqrt{n}}\right]$$

Exhaustif

$$\left[\overline{x} - g.\sqrt{\frac{N-n}{N-1}}.\frac{\sigma}{\sqrt{n}}; \overline{x} + g.\sqrt{\frac{N-n}{N-1}}.\frac{\sigma}{\sqrt{n}}\right]$$

Petite taille (n<30)

« **moyenne standardisée des échantillons** » **suit une loi t de Student à (n – 1) degrés de liberté**.

$E(t_v) = 0, Var(t_v) = \frac{v}{v-2}$ avec $(v > 2)$

$v = n - 1$

En sachant que $E(\overline{x}) = m$ et $var(\overline{x}) = \frac{\sigma^2}{n}$

On obtiens la fonction suivante

$$Pr(\overline{x})\left\{\overline{x} - SudentVal * \sqrt{\frac{\sigma^2}{n}} \leq E(\overline{x}) \leq \overline{x} + SudentVal * \sqrt{\frac{\sigma^2}{n}}\right\}$$

Echantillon exhaustif

$$Pr(\overline{x})\left\{\overline{x} - SudentVal * \sqrt{\frac{N-n}{N-1}} * \sqrt{\frac{\sigma^2}{n}} \leq E(\overline{x}) \leq \overline{x} + SudentVal * \sqrt{\frac{N-n}{N-1}} * \sqrt{\frac{\sigma^2}{n}}\right\}$$

Marge d'erreur associée à un intervalle de confiance

La marge d'erreur associée à un intervalle correspond à la **demi amplitude de cet intervalle**.

De façon générale, la marge d'erreur vaut :

- Pour l'intervalle de confiance d'une proportion : $a.\sqrt{\frac{p.(1-p)}{n}}$
- Pour un intervalle de confiance de moyenne : $a.\frac{\sigma}{\sqrt{n}}$

Pour diminuer cette marge d'erreur, nous pouvons soit **augmenter la taille de l'échantillon**, soit **augmenter le niveau d'incertitude**.

Marge d'erreur Maximale

La marge d'erreur maximale est obtenue lorsque $p = 0.5$.

---

1. Niveau de Confiance ↩