Chapitre 3

Les intervalles de confiance

1) Introduction

- Comment détermine-t-on dans quelles limites se situe le nombre de globules rouges par litre de sang chez un individu en bonne santé ?
- Ce type d'évaluations est déduit de modèles probabilistes par les techniques statistiques d'estimation paramétrique.
- La notion importante est celle **d'intervalle de confiance**, qui permet d'évaluer la précision d'une **estimation ponctuelle**.

2) Intervalle de confiance d'une moyenne

Cas de variance σ^2 connue.

- On suppose que la variable $\boldsymbol{X}$ étudiée sur les individus d'une population **suit une loi normale $N(\mu, \sigma^2)$.**
- **Ou n>30** si on n'a pas la normalité

- On peut en général considérer que la population est normale lorsque les données de l'échantillon aléatoire simple **de taille n**, de **moyenne $\overline{x}$** et **d'estimateur de variance**

$$\boldsymbol{s^2 = \frac{n}{n-1} s_{ech}^2}$$

confirment qu'il n'y a pas des valeurs extrêmes et que l'histogramme a une forme qui n'est pas éloignée de celle d'une loi normale.

- L 'objectif est d'estimer μ par un intervalle.
- **$IC_\mu = [\overline{x}-E , \overline{x}+E]$**est **l'intervalle de confiance** de la moyenne μ au **niveau de confiance 1-α (ou au risque d'erreur α)** si

  $P(\mu \in IC_\mu) = 1 - \alpha$.

- $\boldsymbol{E = t_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}}$ la marge d'erreur sur la moyenne
- $\boldsymbol{t_{\frac{\alpha}{2}}}$ la valeur critique de **N(0,1)**, c'est-à-dire

$$P\left(U > t_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \quad \text{où } U \rightsquigarrow \mathrm{N}(0,1).$$

*interpréter un intervalle de confiance

- Si on a par exemple α =5% , on dit que nous avons confiance à 95% que l'intervalle IC_μ contienne la vraie valeur de μ .

- Cela signifie que si on sélectionnait de nombreux échantillons de même taille et qu'on construisait les intervalles de confiance correspondants, 95% d'entre eux contiendraient la vraie valeur.

Exemple 1 « Température du corps humain ».

Pour un échantillon des températures corporelles, on a n=106 et $\bar{x}$=36.78°C.

Supposer que l'échantillon est un échantillon aléatoire simple et que σ est connue et vaut $0.34^{\circ}C$.

➢ Donner l'intervalle de confiance à 95% de la température corporelle.

Il faut d'abord vérifier les conditions requises pour estimer μ

On a la normalité de la population et n>30.

n>30 est suffisante même si on n'a pas la normalité.

$$E = t_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} = 1.96 \frac{0.34}{\sqrt{106}} = 0.06$$

IC_{μ} =[36.72, 36.84]

Cas de variance σ^2 inconnue

Conditions requises pour estimer μ quand σ est inconnu

L'échantillon est aléatoire simple

Soit la population est normalement distribuée, soit n>30.

$$E = t_{\frac{\alpha}{2};\, n-1} * \frac{s}{\sqrt{n}}$$

$t_{\frac{\alpha}{2};\, n-1}$ la valeur critique de la loi de Student $t(n-1)$

Degré de liberté $\nu = n - 1$.

Exemple 2 Considérons l'exemple de la température corporelle mais avec l'écart type estimé s=0.34.

Les conditions requises pour l'estimation de μ sont satisfaites.

(n étant >30 , on n'a pas besoin de vérifier la normalité)

pour α=0.05, n-1=105 on a $t_{\alpha/2,n-1}$=1.984

$$E = t_{\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} = 1.984 \frac{0.34}{\sqrt{106}} = 0.06$$

IC_{μ}= [36.72, 36.84]

Interprétation à partir des résultats de l'échantillon on peut être sûr à 95% que l'intervalle [36.72; 36.84] contient effectivement la vraie valeur de la moyenne de la population.

- **σ est connu**

Si $\boldsymbol{X} \sim \mathrm{N}(0{,}1)$ **ou n>30,** E est calculé avec $t_{\alpha/2}$ (de la table de la loi normale).

- **σ de la population n'est pas connu. Si n>30** ou **la population suit une loi normale,** E est calculé avec $t_{\alpha/2,n-1}$ (de la loi de Student)
- **n<30 et la population ne suit pas une loi normale,** E ne devrait pas être calculé par les méthodes paramétriques

3) intervalle de confiance d'une proportion

*Notations pour les proportions

$\boldsymbol{p} =$ proportion de sucés dans toute la population.

$\widehat{\boldsymbol{p}} = \frac{\boldsymbol{x}}{\boldsymbol{n}}$= proportion de x succès dans un échantillon de taille n.

$\widehat{\boldsymbol{q}} = \mathbf{1} - \widehat{\boldsymbol{p}}$ proportion d'échecs dans un échantillon de taille n.

*Conditions requises pour l'estimation de p

* l'échantillon est aléatoire simple

* **np≥5 et nq ≥ 5**

*Intervalle de confiance pour p de la population

IC_p=[$\widehat{p}$-E, $\widehat{p}$ +E]

$$\boldsymbol{E} = \boldsymbol{t}_{\frac{\alpha}{2}} * \sqrt{\frac{\widehat{\boldsymbol{p}}\widehat{\boldsymbol{q}}}{\boldsymbol{n}}}$$

$\boldsymbol{t}_{\frac{\alpha}{2}}$ est la valeur critique de **N(0,1)**

Exemple Gregor Mendel a mené ses fameuses expériences génétiques avec des pois, un des échantillons des croisements a été obtenu en croisant **des pois à gousses vertes** et **des pois à gousses jaunes**. Cette lignée comportait **580 pois**. Parmi ces pois, **428 avait des gousses vertes** et **152 gousses jaunes**.

A partir de sa théorie des gènes, Mendel s'attendait à ce que **25%** des pois aient **des gousses jaunes**. Le pourcentage **de gousses jaunes est de 26.2%.**

a. Trouver la marge d'erreur qui correspond à un intervalle de confiance à 95%
b. Trouver l'intervalle de confiance à 95% de p de la population
c. A partir de ces résultats, que pouvons- nous conclure sur la théorie de Mendel qui déclare que le pourcentage de pois à gousse jaune devrait être de 25%?

Solution

On a: **n$\widehat{p}$=152 ≥ 5, n$\widehat{q}$=428≥ 5**

a. **E=1.96** $\sqrt{\frac{0.262*0.738}{580}}$ **= 0.036**

b. **IC$_p$=[0.226, 0.298]** cet intervalle est décrit comme suit :

Le pourcentage de pois à gousse jaune est estimé à 26.2% avec une marge d'erreur de plus ou moins 3.6%.

A partir de ces résultats nous sommes sûr à 95% que les limites 22.6% et 29.8% contiennent le vrai pourcentage de pois à gousses jaunes.

Le vrai pourcentage peut être vraisemblablement n'importe quelle valeur entre ces deux limites. Comme cet intervalle contient la valeur 25%, la valeur de Mendel ne peut pas être considéré comme fausse.

4) Intervalle de confiance d'une variance

*conditions requises pour estimer σ^2

L 'échantillon est aléatoire simple

La population doit avoir une distribution normale (même si n>30).

*Intervalle de confiance de σ^2

$$IC_{\sigma^2} = \left[\frac{(n-1)s^2}{{\chi^2}_D}, \frac{(n-1)s^2}{{\chi^2}_G}\right]$$

- s^2 estimateur de la variance de la population
- ${\chi^2}_G$ est la valeur qui se localise à la ligne (le degré de liberté) de la colonne gauche et la colonne 1- α/2 au niveau des en-têtes.
- ${\chi^2}_D$ est la valeur qui se localise à la ligne (le degré de liberté) de la colonne gauche et la colonne α/2 au niveau des en-têtes.

Exemple : « Températures corporelles »

On liste 106 températures corporelles prises par des chercheurs. Supposer que c'est un échantillon simple et utiliser les caractéristiques suivantes pour construire un **intervalle de confiance à 95%** de **l'écart type** pour les températures corporelles de l'ensemble de la population.

a. D'après la représentation graphique (histogramme) de l'échantillon, **Il n'y a pas de valeurs extrêmes** et les **données semblent suivre une loi normale**.

$\bar{x}$ **= 36.78°C,** $s = $ **0.34 °C,** $n = 106$

Solution

1. La condition de normalité est satisfaite.

n=106, le ddl=106-1=105 . Si ce ddl n'est pas dans la table, nous prendrons la valeur la plus proche. Par exemple **ddl=100.**

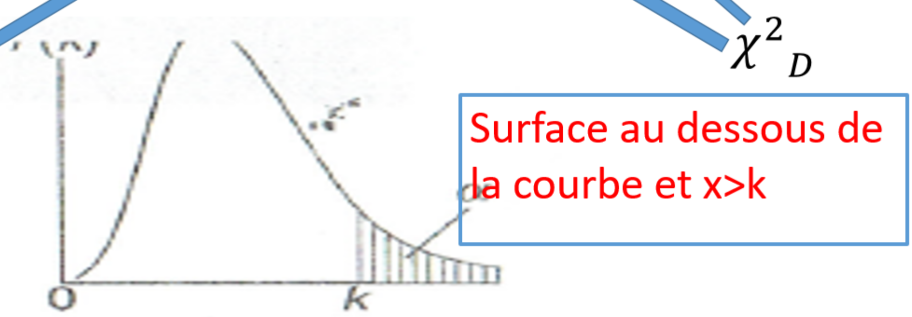
Pour un niveau de confiance 95%, on se réfère aux valeurs 0.975 et 0.025 comme en- têtes de colonnes, On a pour **ddl=105** :$\chi^2{}_G$= **78,54**, $\chi^2{}_D$=**135,2**

Pour **ddl=100** : $\chi^2{}_G$= 74.22, $\chi^2{}_D$=129.561

α=0,05 | 1- α/2 | α/2

	0,995	0,99	0,98	0,975	0,95	0,9	0,8	0,03	0,025	0,02	0,01	0,005
100	68,15	70,06	73,14	74,22	77,93	82,36	87,95	128,2	129,6	131,1	135,8	140,2
105	71,43	74,25	77,42	78,54	82,35	86,91	92,65	133,9	135,2	136,9	141,6	146,1

$\chi^2{}_G$ $\chi^2{}_D$



$$IC_{\sigma^2} = \left[\frac{(n-1)s^2}{\chi^2{}_D}, \frac{(n-1)s^2}{\chi^2{}_G}\right]$$

$$IC_{\sigma^2} = \left[\frac{(106-1)*(0.34)^2}{135.2}, \frac{(106-1)*(0.34)^2}{78.54}\right]$$

$\boldsymbol{IC_{\sigma^2}}$ =**[0.0898, 0.155]**

Si on prend la racine carré **IC $_\sigma$=[0.3, 0.39]**