Guide Pratique pour le cours de Probabilité et Statistiques Appliquées 2

Contents

Ce guide pratique est un résumé du PDF sur le Cours de Probabilité et statistiques. Beaucoups de raccourcis vont être fait ici et je vous conseille de d'abord lire la version complète avant d'enreprendre la lecture de celui-ci. Pour ceux qui sont actuellement en examen, bonne chance :)

Partie 1

Moyenne

$\overline{x} = \frac{1}{n} \sum r_i.x_i$

Variance ($Unit^2$)

$Varx = \overline{x^2} - \overline{x}^2 = \frac{1}{n}. \sum r_i.(x_i)^2 - (\overline{x})^2 = \frac{1}{n} \sum r_i.(x_i - \overline{x})^2$

Ecart-type

$\sigma(x) = \sqrt{Var(x)}$

variance de l'échantillon

$S_n^2 = \frac{r_i.(C_i)^2}{n} - \overline{x}^2$

r sont le nombre d'occurence de l'échantillon et c la valeur dans l'échantillon

Biais des estimateurs

Les paramètres n et $\overline{x}$ sont des estimateurs **sans biais** tandis que S_n^2 est **biaisé** de θ^2. S_{n-1}^2 (**variance corrigée**) est un estimateur **sans biais** (non exhaustif).

Non-Exhaustif

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

Exhaustif

Qualités d'un estimateur

Un estimateur est **absolument correct** si il est **sans biais** et **consistant**

Estimation par intervalle de confiance

Il faut déterminer L_i et L_s tels que $\Pr\{L_i \leq \theta \leq L_s\}$=$1 - \alpha$

Intervalle de confiance d'une proportion

Si : $n \geq 30$ et $n.p \geq 5$ et $n.(1-p) \geq 5$ pour une loi binomiale :

il suffit de : résoudre la fonction suivante (non-exhaustive):

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

(exhaustive)

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

- f : fréquence observée dans l'échantillon = $\frac{occurence}{n}$
- N : Population Totale
- n : taille de l'échantillon
- p = f ou p = 0.5 (marge d'erreur maximale)

On lit ensuite la table de Gauss pour trouver la valeur dans le tableau qui vaut $1 - \frac{(1-\alpha)}{2}$

α est le niveau d'incertitude

- Si la valeur est dedans, on regarde combien elle vaut (eg : 0,95 -> 1,96 = g)
- Si elle n'est pas dedans, on effectue une interpolation linéaire

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

Intervalle de confiance d'une moyenne

Les échantillons de grande taille

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

Les échantillons de petite taille $(n < 30)$

Si la population est **distribuée normalement**, suit une **loi t de Student à (n − 1) degrés de liberté**. v = 20-1 que l'on remplace à la place de g dans les equations ci-dessus

Marge d'erreur associée à un intervalle de confiance

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

Partie 2

L'hypothese nulle : H_0 doit toujours (pour des raisons de simplicité) être définie avec **une égalité**

α est le risque de commettre une erreur. (par défaut : $\alpha = 0.05$)

Méthodologie de resolution Test de Valeur pour une probabilité:

Quand une différence est-elle suffisamment significative pour affirmer qu'un paramètre de la population a changé ?

1. Les hypothèses

$H_0 : p = 0.21$ (pour des raisons de simplicité, on ne travaille qu'avec des équivalences)
$H_1 : p \neq 0.21$ (test bilatéral)
$H_1 : p < 0.21$ (test unilatéral gauche)

2. Seuil de signification

$\alpha = 0.05$ (valeur utilisée en général)

3. Loi de probabilité utilisée (différente en fonction de l'échantillon, type de test, ...)

$$n \geq 30 \Rightarrow \frac{f - E(f)}{\sqrt{Var(f)}}$$

Rappel : E(f) = p et Var(f) = $\frac{p.(1-p}{n}$

4. Déterminer le seuil de rejet et la zone de non refus

`* Dans le cas d'un test bilatéral : $$Pr\{p-ME \leq f \leq p+ME \} = 1-\alpha$$`

Rappel : $ME = a.\sqrt{\frac{p.(1-p)}{n}}$ a est la valeur dans la table de gauss pour $1 - \frac{(1-\alpha)}{2}$

`* Dans le cas d'un test unilatéral gauche : On ne garde qu'une seule partie de l'inégalité c`

Rappel : $ME = a.\sqrt{\frac{p.(1-p)}{n}}$
a est la valeur dans la table de gauss pour $1 - \alpha$
$\{X_G^* \leq x_0\} = 0.05$
$\{X_G^* \geq x_0\} = 0.95$
$\{X_G^* \leq -x_0\} = 0.95$
ce qui donne après lecture directe de la table de gauss : $-x_0 \approx 1.645$
, $x_0 \approx -1.645$

5. Calculer la grandeur expérimentale

f = $\frac{occurence}{n}$

Puis on remplace dans l'équation suivante :

$$Pr\{p - a.\sqrt{\frac{p.(1-p)}{n}} \leq f \leq p + a.\sqrt{\frac{p.(1-p)}{n}}\} = 1 - \alpha$$

6. Tirer les conclusions

Soit :

- $f \in$ zone de non refus. Nous ne pouvons donc pas rejeter H_0
- $f \notin$ zone de non refus. Nous rejetons H_0

Au seuil de signification $\alpha = 0.05$, la proportion de … ne semble pas / semble avoir changé significativement.

Méthodologie de resolution Test de Valeur pour une moyenne:

quasiment idem que pour les probabilités à un détail près

3. Loi de probabilité utilisée (différente en fonction de l'échantillon, type de test, …)

$$n \geq 30 \Rightarrow \overline{x} \approx \frac{f - E(\overline{x})}{\sqrt{Var(\overline{x})}}$$

> Rappel : $E(\overline{x}) = m$
> $Var(\overline{x}) = \frac{\sigma^2}{n} \approx \frac{1}{n}S_{n-1}^2 = \frac{1}{n}.\frac{n}{n-1}.S_n^2$

4. Déterminer le seuil de rejet et la zone de non refus

```
* Dans le cas d'un test bilatéral : $$Pr\{p-ME \leq f \leq p+ME \} = 1-\alpha$$
```

> Rappel : $ME = a.\sqrt{\frac{p.(1-p)}{n}}$ a est la valeur dans la table de gauss pour $1 - \frac{(1-\alpha)}{2}$

```
* Dans le cas d'un test unilatéral gauche : On ne garde qu'une seule partie de l'inégalité c
```

> Rappel : $ME = a.\sqrt{\frac{p.(1-p)}{n}}$
> a est la valeur dans la table de gauss pour $1 - \alpha$
> $\{X_G^* \leq x_0\} = 0.05$
> $\{X_G^* \geq x_0\} = 0.95$
> $\{X_G^* \leq -x_0\} = 0.95$
> ce qui donne après lecture directe de la table de gauss : $-x_0 \approx 1.645$
> , $x_0 \approx -1.645$

5. Calculer la grandeur expérimentale

f = $\frac{occurence}{n}$

Puis on remplace dans l'équation suivante :

$$Pr\{p - a.\sqrt{\frac{p.(1-p)}{n}} \leq f \leq p + a.\sqrt{\frac{p.(1-p)}{n}}\} = 1 - \alpha$$

6. Tirer les conclusions

Soit :

- $f \in$ zone de non refus. Nous ne pouvons donc pas rejeter H_0
- $f \notin$ zone de non refus. Nous rejetons H_0

Au seuil de signification $\alpha = 0.05$, la proportion de ... ne semble pas / semble avoir changé significativement.