

Intervalles de Confiance

A.B. Dufour & T. Jombart

L'objectif est de représenter les intervalles de confiance d'une moyenne, d'une proportion.

Table des matières

1	Intervalle de confiance de la moyenne	2
1.1	Cas où la variance de la population est connue	2
1.2	Cas où la variance de la population est inconnue	3
1.3	Exemple	4
2	Intervalle de confiance d'une proportion	4
2.1	Rappel de la définition	4
2.2	Exemple	5
3	Exercices	5
3.1	Taille des basketteurs américains	5
3.2	Distance entre domiciles des époux	6
3.3	Pied d'impulsion au saut en hauteur	6
3.4	Présence d'une personne handicapée dans l'entourage	7

1 Intervalle de confiance de la moyenne

Soit X une variable aléatoire continue de moyenne μ et de variance σ^2 . Soit n observations de X . La moyenne de cet échantillon \bar{x} est distribuée autour de μ .

1.1 Cas où la variance de la population est connue

La dispersion autour de la moyenne est appelée erreur standard et est définie par $\frac{\sigma}{\sqrt{n}}$.

L'intervalle de confiance à 95% est :

$$\bar{x} + \frac{\sigma}{\sqrt{n}} \epsilon_{0.025} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} \epsilon_{0.975}$$

où $\epsilon_{0.025}$ et $\epsilon_{0.975}$ sont respectivement les quantiles 2.5% et 97.5% de la loi normale centrée réduite.

Comme la loi normale est symétrique, $\epsilon_{0.025} = -\epsilon_{0.975}$ et l'intervalle s'écrit :

$$\bar{x} \pm \frac{\sigma}{\sqrt{n}} \epsilon_{0.975}$$

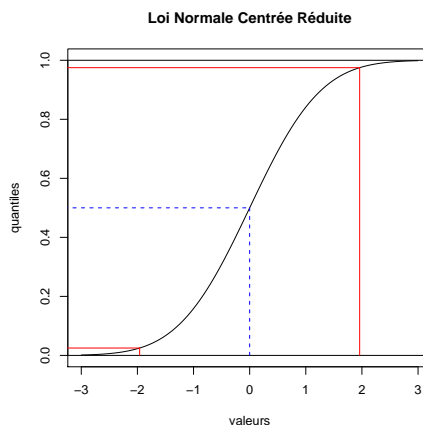
```
qnorm(0.025)
```

```
[1] -1.959964
```

```
qnorm(0.975)
```

```
[1] 1.959964
```

$\epsilon_{0.975}$ est notée $\Phi^{-1}(0.975)$ où Φ est la fonction de répartition de la loi normale (`pnorm`).



1.2 Cas où la variance de la population est inconnue

Comme la variance de la population est inconnue, on l'estime à partir de l'échantillon. On a alors la relation suivante :

$$\frac{\widehat{\sigma^2}}{n} = \frac{s^2}{n-1}$$

La dispersion autour de la moyenne est alors définie par $\frac{s}{\sqrt{n-1}}$ et l'intervalle de confiance à 95% devient :

$$\bar{x} + \frac{s}{\sqrt{n-1}} t_{0.025} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n-1}} t_{0.975}$$

où $t_{0.025}$ et $t_{0.975}$ sont respectivement les quantiles 2.5% et 97.5% de la loi de Student à $n-1$ degrés de liberté (cf tdr21).

Prenons le cas d'un échantillon de taille $n = 10$. Le degré de liberté associé à la loi de Student est alors $n-1 = 9$.

Comme la loi de Student est symétrique, $t_{0.025} = -t_{0.975}$ et l'intervalle s'écrit :

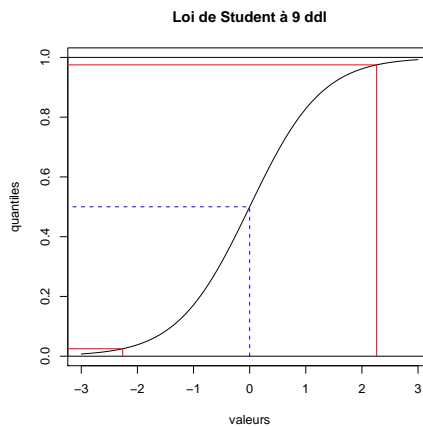
$$\bar{x} \pm \frac{s}{\sqrt{n-1}} t_{0.975}$$

```
qt(0.025, 9)
```

```
[1] -2.262157
```

```
qt(0.975, 9)
```

```
[1] 2.262157
```



1.3 Exemple

Dans un centre médico-sportif, on a mesuré la taille de 71 footballeurs du département du Rhône. La moyenne est $\bar{x} = 177\text{cm}$ pour une variance $s^2 = 31.98$.

Les valeurs associées aux quantiles 2.5% et 97.5% pour 70 ddl sont :

```
qt(0.025, 70)
```

```
[1] -1.994437
```

```
qt(0.975, 70)
```

```
[1] 1.994437
```

L'intervalle de confiance à 95% a pour borne inférieure :

```
177 + qt(0.025, 70) * sqrt(31.98/70)
```

```
[1] 175.6519
```

et pour borne supérieure :

```
177 + qt(0.975, 70) * sqrt(31.98/70)
```

```
[1] 178.3481
```

Pour éviter de réaliser les calculs à la main, une fonction peut être créée :

```
icmoyenne <- fonction(emoy, evar, eeff, nivconf) {
  int <- (1 - nivconf)/2
  binf <- emoy + qt(int, eeff - 1) * sqrt(evar/eeff)
  bsup <- emoy + qt(1 - int, eeff - 1) * sqrt(evar/eeff)
  bornes <- c(binf, bsup)
  return(bornes)
}
icmoyenne(177, 31.98 * 71/70, 71, 0.95)
```

```
[1] 175.6519 178.3481
```

2 Intervalle de confiance d'une proportion

Soit la variable F , proportion d'individus présentant le caractère A dans une population. La proportion théorique dans la population est notée p . La proportion observée dans un échantillon de taille n est notée f .

2.1 Rappel de la définition

L'intervalle de confiance à 95% est :

$$f + \sqrt{\frac{f(1-f)}{n}} \epsilon_{0.025} \leq p \leq f + \sqrt{\frac{f(1-f)}{n}} \epsilon_{0.975}$$

où $\epsilon_{0.025}$ et $\epsilon_{0.975}$ sont respectivement les quantiles 2.5% et 97.5% de la loi normale centrée réduite.

2.2 Exemple

A la question "Etes-vous favorable à l'implantation d'une nouvelle piscine sur le campus de la Doua?", 340 étudiants sur 400 ont répondu OUI. Les 400 étudiants interrogés ont été choisis au hasard sur le campus. La proportion observée d'étudiants favorables à l'implantation d'une nouvelle piscine est : $f = 0.85$. L'estimation ponctuelle d'étudiants favorables à l'implantation d'une nouvelle piscine dans la population est $\hat{p} = 0.85$.

L'intervalle de confiance au niveau 95% est :

```
0.85 + qnorm(0.025) * sqrt(0.85 * (1 - 0.85)/400)
```

```
[1] 0.8150076
```

```
0.85 + qnorm(0.975) * sqrt(0.85 * (1 - 0.85)/400)
```

```
[1] 0.8849924
```

Comme précédemment, on peut écrire une fonction permettant de calculer la borne inférieure et la borne supérieure de l'intervalle.

```
icproportion <- function(eprop, eeff, nivconf) {
  int <- (1 - nivconf)/2
  binf <- eprop + qnorm(int) * sqrt(eprop * (1 - eprop)/eeff)
  bsup <- eprop + qnorm(1 - int) * sqrt(eprop * (1 - eprop)/eeff)
  bornes <- c(binf, bsup)
  return(bornes)
}
icproportion(340/400, 400, 0.95)
```

```
[1] 0.8150076 0.8849924
```

3 Exercices

3.1 Taille des basketteurs américains

On connaît la taille (en m) de 12 basketteurs américains (in *Mondial Basket*, juillet-août 1994).

```
taibask <- c(2.08, 2.01, 2.03, 2.1, 1.98, 2.08, 1.85, 2.03, 2.16,
            2.01, 1.91, 1.88)
```

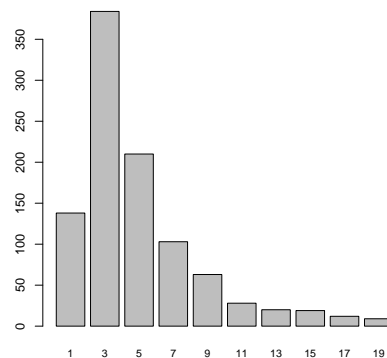
- Donner la moyenne, la variance estimées de la population à partir de l'échantillon.
- Donner les intervalles de confiance de la moyenne de la population aux niveaux de confiance 0.95 et 0.99.
- On suppose que la taille suit une loi normale dont les paramètres ont été estimés dans la première question. Quelle est la probabilité pour un basketteur d'avoir une taille supérieure à 2.05 m.

3.2 Distance entre domiciles des époux

Une enquête concernant les distances (en km) entre domiciles des époux au moment du mariage a donné, dans le Finistère, les résultats suivants.

Distances	Nombre de couples
0-2	138
2-4	384
4-6	210
6-8	103
8-10	63
10-12	28
12-14	20
14-16	19
16-18	12
18-20	9

- Représenter graphiquement les données et commenter.
- Calculer la moyenne et la variance estimées de la distance entre les domiciles des époux au moment du mariage.
- Donner l'intervalle de confiance de la moyenne de la population au niveau de confiance 0.95.



3.3 Pied d'impulsion au saut en hauteur

Dans un lycée, le professeur d'éducation physique a noté le pied d'impulsion au saut en hauteur. 517 lycéens ont le pied gauche comme pied d'impulsion ; 131 lycéens ont le pied droit comme pied d'impulsion.

En supposant que ces lycéens constituent un échantillon de la population française, donner l'intervalle de confiance (niveau 0.95) de la proportion d'individus prenant l'appel du pied droit.

```
icproportion(eprop, eeff, 0.95)
```

3.4 Présence d'une personne handicapée dans l'entourage

Après une enquête sur un échantillon de 45 étudiants de licence et maîtrise APA, on a constaté que 29 d'entre eux avaient, dans leur entourage, une personne présentant un handicap. Estimer, par intervalle de confiance au niveau 0.95, la proportion d'étudiants ayant, dans leur entourage, une personne handicapée.