

Analyse de la variance (ANOVA) à un facteur

Mathilde Boissel

11/10/2020

Table of Contents

Méthode	2
Test ANOVA.....	2
Diagnostic.....	6
Test de comparaison de variance.....	11
Test de comparaison de moyennes	13
Récap'	15
Exercices	16
Exercice 1.....	16
Exercice 2.....	18
Exercice 3.....	21
Exercice 4.....	25
Exercice 5.....	27
Annexes	29
Annexe 1 : Loi normale.....	29
Annexe 2 : Loi normale (bis)	30
Annexe 3 : Loi de Student.....	31
Annexe 4 : Loi du chi-deux.....	32
Annexe 5 : Loi de Fisher I ($\alpha = 0.025$)	33
Annexe 6 : Loi de Fisher II ($\alpha = 0.05$).....	34
Annexe 7 : Valeurs de Cochran.....	35
Sources	36

Méthode

Test ANOVA

Hypothèses stochastiques

On étudie un caractère représenté par une variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Les échantillons sont issus d'une population normale (gaussienne) : on parle de test **paramétrique**.

N.B.: l'ANOVA non paramétrique, avec le Test de Kruskal–Wallis, ne sera pas abordé dans ce cours.

On suppose que la population se divise en p sous-population $\mathcal{P}_1, \dots, \mathcal{P}_n$.

Ainsi, $\forall i \in \{1, \dots, p\}$, la variable X considérée dans \mathcal{P}_i , est une variable

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2).$$

La moyenne μ , et a fortiori μ_i , sont inconnus.

Les variances conditionnelles (variances dans chaque sous-population) sont identiques :

homoscédasticité.

Les données sont constituées de, $\forall i \in \{1, \dots, n_i\}$, la valeur de X_i pour chacun des n_i individus d'un échantillon de \mathcal{P}_i . Ces valeurs sont notées $x_{i,1}, \dots, x_{i,n_i}$ pour X_i .

Les individus sont tous différents; les échantillons sont indépendants.

Enjeu

L'enjeu d'un test ANOVA est d'affirmer, avec un faible risque de se tromper, que $\mathcal{P}_1, \dots, \mathcal{P}_n$ ne sont pas homogènes quant à μ .

Pour ce faire, on compare les moyennes μ_1, \dots, μ_p via un test statistique.

Notion de facteur

$\forall i \in \{1, \dots, p\}$, \mathcal{P}_i peut être associée à une modalité A_i d'un caractère A appelé facteur.

Comparer $\mathcal{P}_1, \dots, \mathcal{P}_n$ quant à μ revient à étudier l'influence du facteur A (caractérisé par ses p modalités A_1, \dots, A_p) sur μ .

μ est ici la moyenne d'une variable (numérique) d'intérêt X .

Les données de X peuvent donc être mises sous la forme :

A1	A2	Ai	Ap
$x_{1,1}$	$x_{1,2}$	$x_{1,i}$	$x_{1,p}$
$x_{2,1}$	$x_{2,2}$	$x_{2,i}$	$x_{1,p}$
...	...	$x_{j,i}$...
...	
...
...	$x_{n_2,2}$...	x_{p,n_p}

Hypothèses

Les hypothèses associées au test ANOVA sont :

$H_0: \mu_1 = \dots = \mu_p = \mu$ (A n'influe pas sur X)

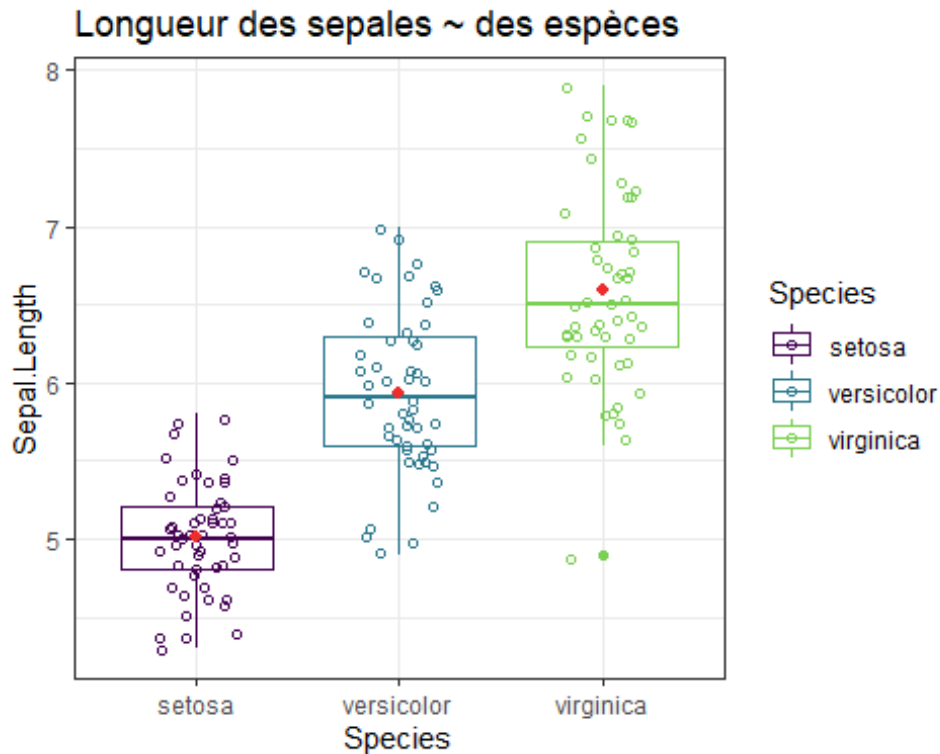
contre

$H_1: \exists j, \mu_j \neq \mu$ (il existe au moins 2 moyennes différentes, c-à-d A influe sur X).

En d'autres mots, l'hypothèse nulle indique que la moyenne de la variable dépendante (X) est la même quelque soit les groupes (A_i) définis par le facteur (A). Sous H_0 , le facteur A n'a aucune influence sur la variable dépendante X .

Illustrations

Des représentations graphiques des données peuvent aider à appréhender la solution.



Objectif

On veut décider du rejet de H_0 , au risque $\alpha/100, \alpha \in]0,1[$.

Test ANOVA

- Test statistique : On utilise alors le test ANOVA, lequel suppose que $\sigma_1^2, \dots = \sigma_p^2$. Sa construction repose sur la loi de Fisher $\mathcal{F}(v_1, v_2)$.

- Calculs : Le test ANOVA se met en œuvre en calculant :

- n : l'effectif total :

$$\sum_{i=1}^p n_i$$

- \bar{x} : la moyenne totale :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} x_{i,j}$$

- $\bar{x}_i, i \in \{1, \dots, p\}$: la moyenne de $x_{i,1}, \dots, x_{i,n_i}$:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{i,j}$$

- sce = sommes des carrés des écarts :

$$sce_T = sce_F + sce_R,$$

- ddl = degrés de liberté :

$$ddl_T = ddl_F + ddl_R,$$

- cm = carrés moyens cm_F et cm_R ,

- le f_{obs} défini par

$$f_{obs} = \frac{cm_F}{cm_R}.$$

- le réel $f_\alpha(v_1, v_2)$ vérifiant

$$\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha,$$

où $F \sim \mathcal{F}(v_1, v_2)$, $(v_1, v_2) = (ddl_F, ddl_R) = (p - 1, n - p)$.

Si $\alpha = 0.05$, ce réel est évaluable dans la table ANNEXE 6 (Loi de Fisher II).

- Le **tableau ANOVA** récapitule tout ceci :

	sce	ddl	cm	f_{obs}
Total	$sce_T = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{i,j} - \bar{x})^2$	$ddl_T = n - 1$		
Factoriel	$sce_F = \sum_{i=1}^p n_i (\bar{x}_i - \bar{x})^2$	$ddl_F = p - 1$	$cm_F = \frac{sce_F}{ddl_F}$	$f_{obs} = \frac{cm_F}{cm_R}$
Résiduel	$sce_R = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$	$ddl_R = n - p$	$cm_R = \frac{sce_R}{ddl_R}$	

Tableau ANOVA

- Règles de décision : La règle de décision associée au Si $f_{obs} \geq f_{\alpha}(v_1, v_2)$, Alors on rejette H_0 .
- p-valeurs : La p-valeur associée au test ANOVA est

$$p - valeur = \mathbb{P}(F \geq f_{obs}).$$

On peut alors déterminer le degré de significativité du rejet de H_0 .

Par exemple, si $p - valeur = \mathbb{P}(F \geq f_{obs}) < 0.001$, alors le rejet de H_0 est hautement significatif (peut-être symbolisé par ***).

Diagnostic

Hypothèses stochastiques

Dans le même contexte décrit pour le test ANOVA.

Enjeu

On s'intéresse ici aux diagnostics qui permettent d'attester que les résultats de l'anova sont valides. Quatre éléments sont à contrôler. Cela peut se faire graphiquement.

Hypothèses

On considère les hypothèses :

- **Residuals vs Fitted et Constant Leverage: Residuals vs Factor Levels.**

Les résidus sont indépendants. Les résidus ne doivent pas être corrélés entre eux. De la même façon, les résidus ne doivent pas être corrélés au facteur étudié. On peut faire le test de Dubin-Watson pour vérifier l'autocorrélation des résidus mais souvent un contrôle graphique suffit.

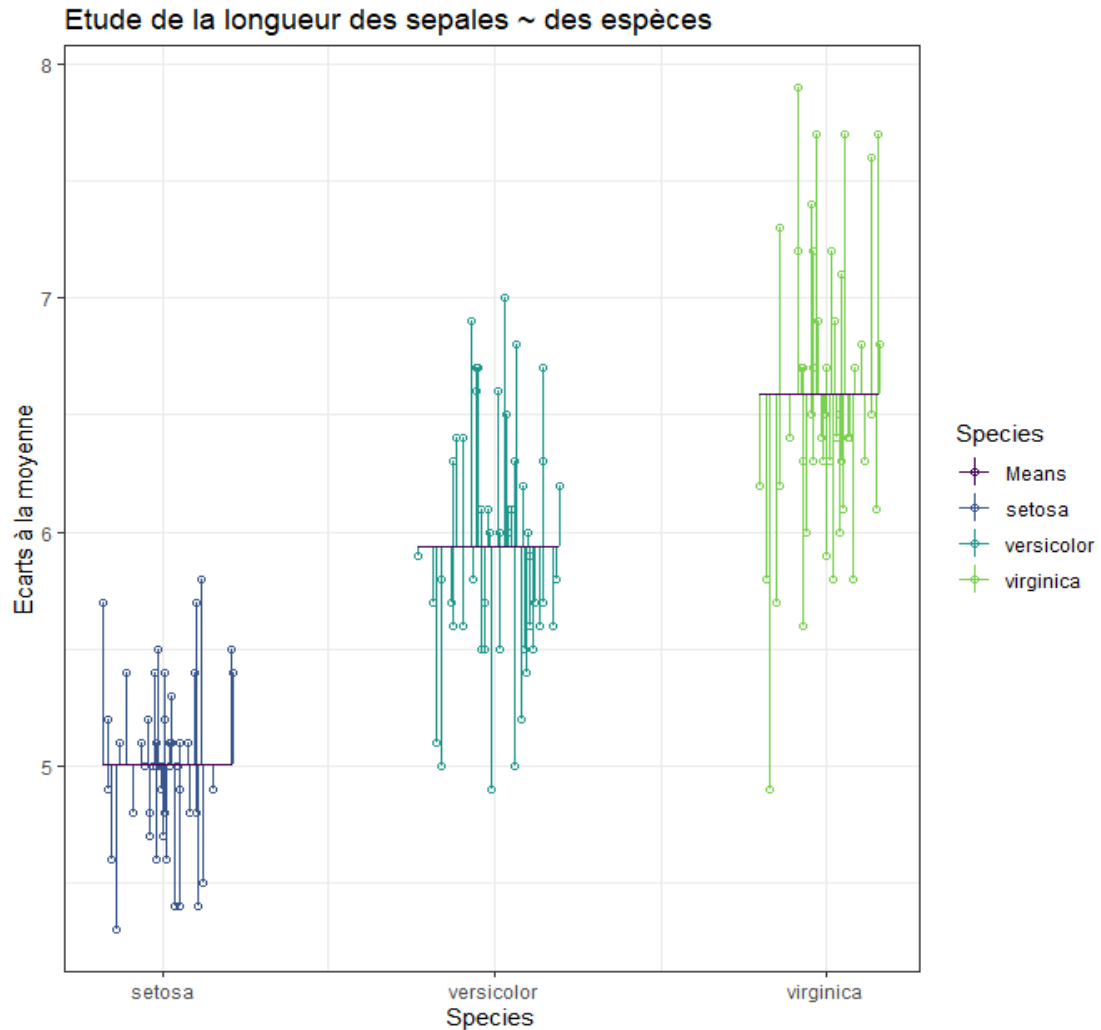
- **Normal Q-Q.**

Les résidus suivent une loi normale de moyenne 0. Pour vérifier cette hypothèse on peut faire un test de normalité comme le test de Shapiro-Wilk mais on préfère vérifier cela graphiquement avec un diagramme Quantile-Quantile (i.e. QQ-plot, graphique dans lequel les quantiles de deux distributions sont tracés l'un par rapport à l'autre).

- **Scale-Location.**

L'homogénéité des variances. Les résidus relatifs aux différentes modalités sont homogènes (ils ont globalement la même dispersion), autrement dit leur variance est constante. On peut vérifier cela graphiquement en représentant les résidus standardisés en fonction des valeurs prédites (les moyennes des différents traitements). En cas de doute on pourra aussi valider cette hypothèse avec un test statistique (Cochran, Bartlett, Levene...).

Illustrations



```
## run the anova in R
anova <- aov(formula = Sepal.Length~Species, data = iris)
## call print method by default
anova

## Call:
## aov(formula = Sepal.Length ~ Species, data = iris)
##
## Terms:
##              Species Residuals
## Sum of Squares  63.21213  38.95620
## Deg. of Freedom      2      147
##
## Residual standard error: 0.5147894
## Estimated effects may be unbalanced
```

```
## call summary method
```

```
summary(anova)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## Species      2  63.21  31.606   119.3 <2e-16 ***
```

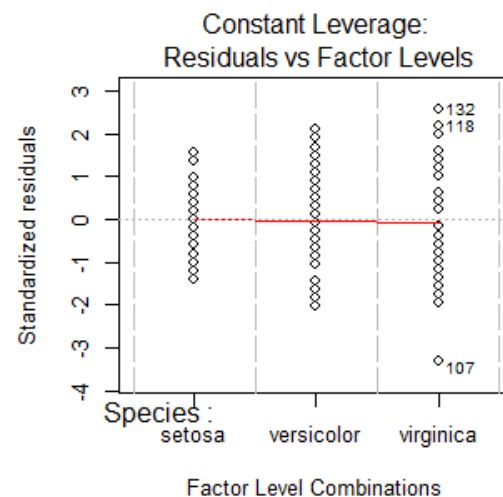
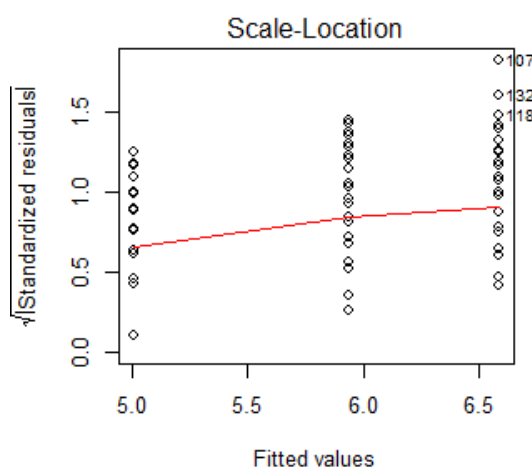
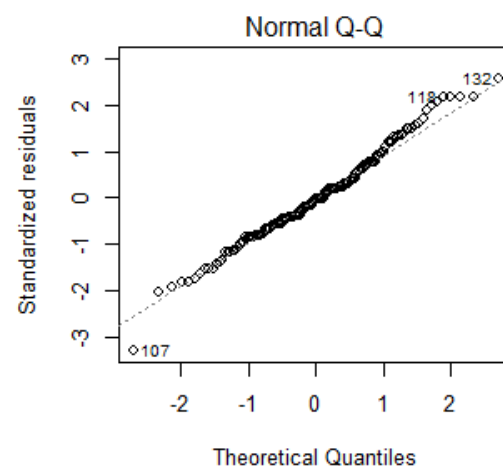
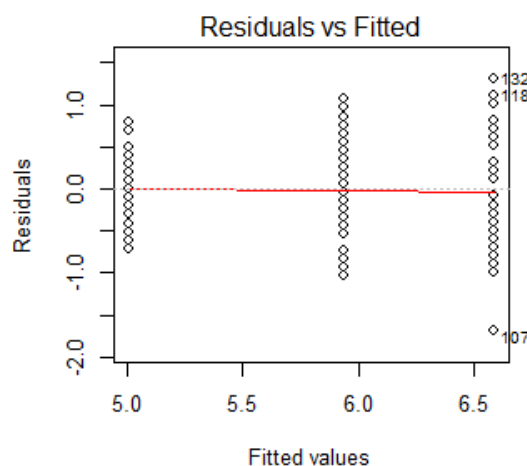
```
## Residuals  147   38.96    0.265
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## print 4 diag plots
```

```
par(mfrow=c(2, 2)); plot(anova)
```

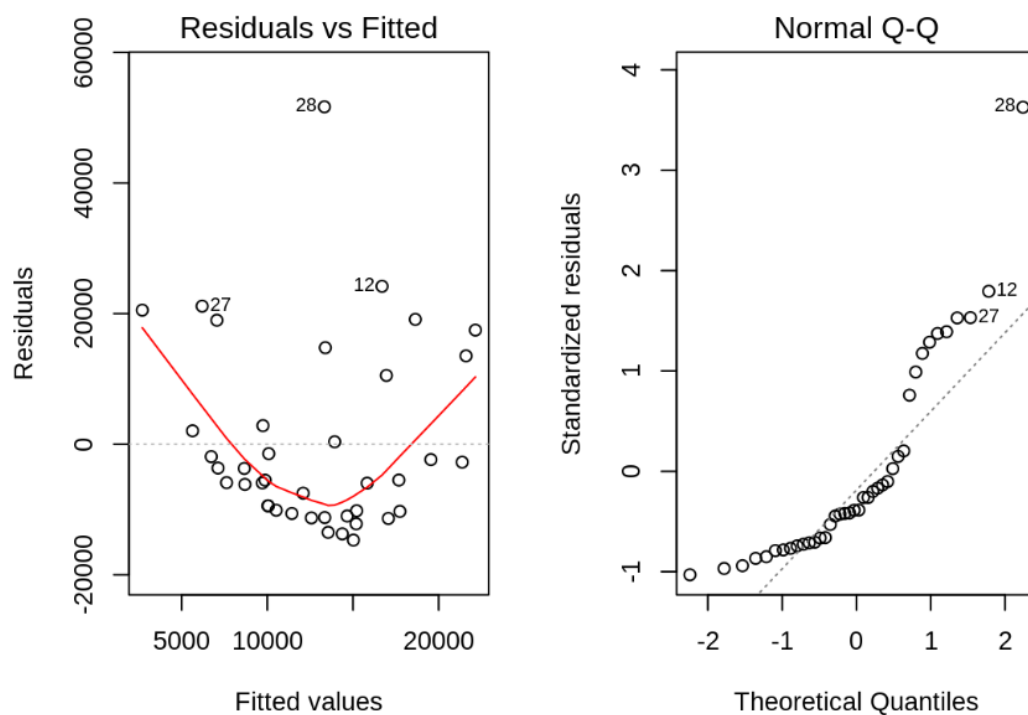


“Mauvais” diagnostique

Pour bien se rendre compte que ces précédents exemples sont signe de bon diagnostique, on peut les confronter à des graphiques qui devraient vous alarmer, signe de “mauvais” diagnostique.

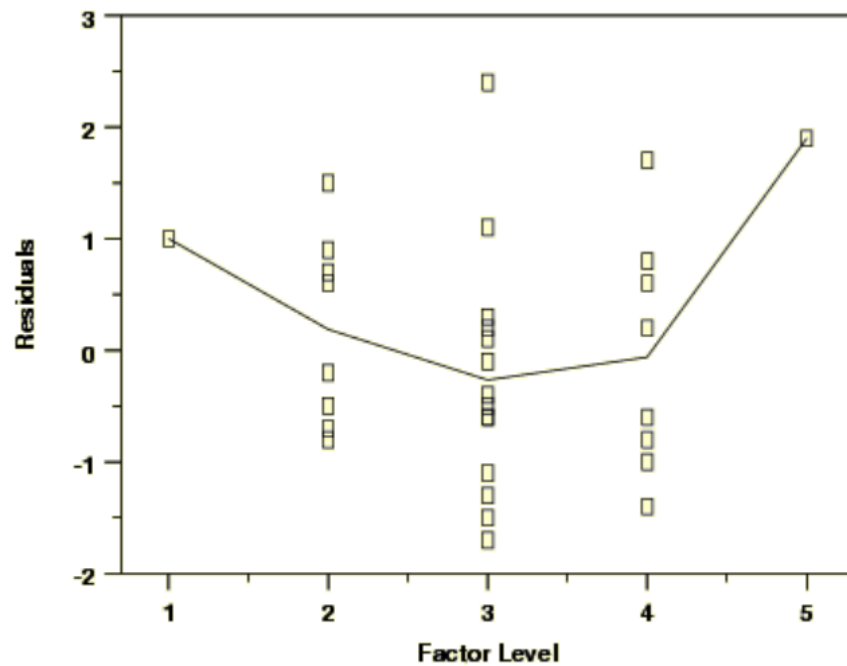
Dans le graph **Residuals vs Fitted**, la ligne rouge n’est pas une droite horizontale proche de $Y=0$.

Dans le graph **Normal Q-Q**, les points observés s’éloignent franchement de la droite théorique. (On pourrait aussi voir ce problème en faisant le graphique de densité des résidus, alors nous n’aurions certainement pas “une cloche” comme attendu avec le loi Normale).



Source : [R Cookbook](#)

Dans le graph **Residuals vs Factor Levels** ci-dessous, la valeur des résidus n'est pas rassurante quand on regarde les différents niveaux de facteur.



Source : itl.nist.gov

Dans ces cas-là, il faut s'interroger sur la nécessité d'appliquer une transformation (racine carré, log, etc) à notre prédicteur, changer de modèle, ou envisager les tests non-paramétriques.

Test de comparaison de variance

Hypothèses stochastiques

Dans le même contexte décrit pour le test ANOVA.

On s'intéresse ici à ce point en particulier : Les variances conditionnelles (variances dans chaque sous-population) sont identiques. On parle d'**homoscédasticité**.

Enjeu

L'enjeu d'un test d'homogénéité pour σ^2 est d'affirmer, avec un faible risque de se tromper, que $\mathcal{P}_1, \dots, \mathcal{P}_n$ ne sont pas homogènes quant à σ^2 .

Pour ce faire, on compare les variances $\sigma_1^2, \dots, \sigma_p^2$ via un test statistique.

Hypothèses

On considère les hypothèses :

$$H_0: \sigma_1^2 = \dots = \sigma_p^2 = \sigma$$

contre

$H_1: \exists j, \sigma_j \neq \sigma$ (il existe au moins 2 variances différentes)

Objectif

On peut décider du rejet de H_0 , au risque $\alpha/100$, $\alpha \in]0,1[$.

Dans ce cas précis, on ne veut pas rejeter H_0 au risque 5%, puisqu'on souhaite vérifier l'égalité des variances. Par convention, on admet que $\sigma_1^2 = \dots = \sigma_p^2$.

Test de Cochran

- Test statistique : Si $n_1 = \dots = n_p$, on utilise le test de Cochran.
N.B. : si $n_i \neq n_j$ voir test de Bartlett.

- Calculs : Le test de Cochran se met en œuvre en calculant :

- si, $i \in \{1, \dots, p\}$: l'écart-type corrigé de $x_{i,1}, \dots, x_{i,n_i}$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2,$$

- le c_{obs} défini par

$$c_{obs} = \frac{\max_{i \in \{1, \dots, p\}} s_i^2}{\sum_{i=1}^p s_i^2}$$

- le réel $c(m, p)$ avec $m = n_1$ (l'effectif commun).
Si $\alpha = 0.05$, ce réel est évaluable dans la table ANNEXE 7 (Valeurs de Cochran).

- Règles de décision : La règle de décision associées au test de Cochran est :
Si $c_{obs} \geq c(m, p)$,
Alors on rejette H_0 .
Par exemple si $c_{obs} < c(m, p)$ et $\alpha = 0.05$, alors on ne rejette pas H_0 ; On admet que $\sigma_1^2 = \dots = \sigma_p^2$.

Test de Bartlett

- Test statistique : Si $\min(n_1, \dots, n_p) \geq 4$, on utilise le test de Bartlett.
- Calculs : Le test de Bartlett se met en œuvre en calculant :

- si, $i \in \{1, \dots, p\}$: l'écart-type corrigé de $x_{i,1}, \dots, x_{i,n_i}$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2,$$

- le χ_{obs}^2 défini par

$$\chi_{obs}^2 = \frac{(n - p) \ln(s^2) - \sum_{i=1}^p (n_i - 1) \ln(s_i^2)}{1 + \frac{1}{3(p-1)} \left(\left(\sum_{i=1}^p \frac{1}{(n_i - 1)} \right) - \frac{1}{(n - p)} \right)}$$

- le réel $\mathbb{P}(K \geq \chi_\alpha^2(v)) = \alpha$ avec $K \sim \chi^2(v), v = p - 1$.
Si $\alpha = 0.05$, ce réel est évaluable dans la table ANNEXE 4 (Loi du Chi-deux).
- Règles de décision : La règle de décision associées au test de Bartlett est :
Si $\chi_{obs}^2 \geq \chi_\alpha^2(v)$,
Alors on rejette H_0 .

Test de comparaison de moyennes

Hypothèses stochastiques

Dans le même contexte décrit pour le test ANOVA.

On s'intéresse ici à la suite de l'étude : la réalisation d'un test "post-hoc".

Enjeu

Si le test ANOVA indique qu'au moins deux moyennes diffèrent, il est intéressant d'étudier la différence de deux d'entre elles.

Hypothèses

Soit $(k, l) \in \{1, \dots, p\}$ avec $k \neq l$. On considère le test statistique :

$H_0: \mu_k = \mu_l$

contre

$H_0: \mu_k \neq \mu_l$

Test de Bonferroni

- Test statistique : On utilise le test de Bonferroni, lequel offre plus de précision que des t-Test 2 à 2 car il prend en compte toutes les données dans sa construction (avec la présence du cm_R).
Ce test repose sur la loi de Student $\mathcal{T}(v)$.
- Calculs : Le test de Bonferroni se met en œuvre en calculant :

- $\bar{x}_i, i \in \{k, l\}$: moyenne de $x_{i,1}, \dots, x_{i,n_i}$,
- $s_R = \sqrt{cm_R}$ (évaluable dans le tableau ANOVA),
- le t_{obs} défini par

$$t_{obs} = \frac{\bar{x}_k - \bar{x}_l}{s_R \sqrt{\frac{1}{n_k} + \frac{1}{n_l}}}$$

- le réel $t_{\alpha}^{**}(v)$ vérifiant

$$\mathbb{P}(|T| \geq t_{\alpha}^{**}(v)) = \frac{2\alpha}{p(p-1)},$$

où $T \sim \mathcal{T}(v)$, $v = n - p$.

Ce réel est dans la table ANNEXE 3 (Loi de Student).

- Règles de décision : La règle de décision associée au test de Bonferroni est :
Si $t_{obs} \geq t_{\alpha}^{**}(v)$,
Alors on rejette H_0 .

- p-valeurs : La p-valeur associée au test de Bonferroni est

$$p - \text{valeur} = \mathbb{P}(|T| \geq |t_{obs}|).$$

Autres tests de comparaison de moyennes

Il existe de nombreux autres tests post-hoc pour la comparaison de moyenne 2 à 2, entre autres :

- Le test de Tukey HSD (Honest Significant Differences),
- Le test de la petite différence significative (LSD) de Fisher,
- Le test Student de Newman-Keuls,
- Dunnett
- ...

Ce sera le rôle du statisticien de se documenter sur les particularités de ces tests et de choisir le plus pertinent selon l'étude et les données.

Récap'

ANOVA à un facteur

Soient A un facteur à p modalités, A_1, \dots, A_p , et X un caractère quantitatif. Pour tout $i \in \{1, \dots, p\}$, X sous A_i est une var $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Les paramètres sont inconnus et on suppose que

$$\sigma_1^2 = \dots = \sigma_p^2.$$

On observe la valeur de X_i pour chacun des n_i individus d'un échantillon. Les échantillons sont indépendants.

On considère les hypothèses :

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_p \text{ (} A \text{ n'influe pas sur } X \text{)} \quad \text{contre} \quad H_1 : \text{"il existe au moins 2 moyennes différentes (} A \text{ influe sur } X \text{)"}$$

Pour pouvoir décider du rejet de H_0 au risque $100\alpha\%$, $\alpha \in]0, 1[$,

- on calcule (si besoin est) $n = \sum_{i=1}^p n_i$, $\bar{x} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} x_{i,j}$, $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{i,j}$.
- on dresse le tableau ANOVA à un facteur :

	sce (SS)	ddl (DF)	cm (MS)	f_{obs} (F)
Total	$sce_T = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{i,j} - \bar{x})^2$	$ddl_T = n - 1$		
Factoriel	$sce_F = \sum_{i=1}^p n_i (\bar{x}_i - \bar{x})^2$	$ddl_F = p - 1$	$cm_F = \frac{sce_F}{ddl_F}$	$f_{obs} = \frac{cm_F}{cm_R}$
Residuel	$sce_R = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$	$ddl_R = n - p$	$cm_R = \frac{sce_R}{ddl_R}$	

- on calcule le réel $f_\alpha(\nu_1, \nu_2)$ tel que

$$\mathbb{P}(F \geq f_\alpha(\nu_1, \nu_2)) = \alpha,$$

où $F \sim \mathcal{F}(\nu_1, \nu_2)$, $(\nu_1, \nu_2) = (ddl_F, ddl_R) = (p - 1, n - p)$.

Si

$$f_{obs} \geq f_\alpha(\nu_1, \nu_2),$$

alors on rejette H_0 .

Ou alors : p-valeur = $\mathbb{P}(F \geq f_{obs}) \leq \alpha \Rightarrow$ on rejette H_0 au risque $100\alpha\%$

C. Chesneau

<http://www.math.unicaen.fr/~chesneau/>

ANOVA à un facteur : compléments

On reprend les notations de "ANOVA à un facteur". Outil : $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$.

Égalité des variances : On considère des hypothèses :

$$H_0 : \sigma_1^2 = \dots = \sigma_p^2 \quad \text{contre} \quad H_1 : \text{"il existe au moins 2 variances différentes"}$$

Tests	Stat. test obs.	Valeurs	Rejet de H_0 si
Bartlett (si $\min(n_1, \dots, n_p) \geq 4$)	$\chi_{obs}^2 = \frac{(n-p) \ln(s^2) - \sum_{i=1}^p (n_i - 1) \ln(s_i^2)}{1 + \frac{1}{3(p-1)} \left(\left(\sum_{i=1}^p \frac{1}{n_i - 1} \right) - \frac{1}{n-p} \right)}$	$\mathbb{P}(K \geq \chi_\alpha^2(\nu)) = \alpha$, $K \sim \chi^2(\nu), \nu = p - 1$	$\chi_{obs}^2 \geq \chi_\alpha^2(\nu)$
Cochran (si $n_1 = \dots = n_p = m$)	$c_{obs} = \frac{\max_{i \in \{1, \dots, p\}} s_i^2}{\sum_{i=1}^p s_i^2}$	$c(m, p)$ (voir table correspondante)	$c_{obs} \geq c(m, p)$

Comparaison de 2 moyennes : test de Bonferroni : Soit $(k, \ell) \in \{1, \dots, p\}^2$ avec $k \neq \ell$. On considère les hypothèses :

$$H_0 : \mu_k = \mu_\ell \quad \text{contre} \quad H_1 : \mu_k \neq \mu_\ell.$$

Tests	Stat. test obs.	Valeurs	Rejet de H_0 si
Bonferroni	$t_{obs} = \frac{\bar{x}_k - \bar{x}_\ell}{s \sqrt{\frac{1}{n_k} + \frac{1}{n_\ell}}}$	$\mathbb{P}(T \geq t_\alpha^{**}(\nu)) = \frac{2\alpha}{p(p-1)}$ $T \sim \mathcal{T}(\nu), \nu = n - p$	$ t_{obs} \geq t_\alpha^{**}(\nu)$

C. Chesneau

<http://www.math.unicaen.fr/~chesneau/>

Exercices

Exercice 1

On souhaite comparer 4 fongicides, avec un témoin ne subissant aucun traitement, et entre eux, sur une culture de pomme de terre réalisée en 20 petites parcelles. On affecte aléatoirement les quatre types de fongicides sur 16 parcelles, les 4 parcelles restantes sont soumises au témoin. Les mesures effectuées sont :

Fong. 1	Fong. 2	Fong. 3	Fong. 4	Fong. 5 (témoin)
29.6	35.2	28.0	30.6	27.8
33.7	33.3	35.7	34.6	22.7
33.6	31.5	36.8	37.1	20.9
29.3	36.1	39.3	31.2	18.5

$\forall i \in \{1, \dots, 5\}$, le rendement en pomme de terre pour le fongicide i est une variable $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$, avec μ_i et σ inconnus. Le tableau ANOVA, incomplet, est le suivant :

	sce	ddl	cm	f_{obs}
Total				
Factoriel	411.98			
Résiduel	175.98			

Peut-on affirmer, au risque 5%, que la nature du fongicide influe sur le rendement moyen de pommes de terre ?

Exercice 1 Solution

Le facteur étudié est "type de traitement", avec pour modalités : *Fong 1*, *Fong 2*, *Fong 3*, *Fong 4* et *Fong 5*. Donc $p = 5$.

Par l'énoncé, $\forall i \in \{1, \dots, p\}$, on observe la valeur de $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$, pour chacun des n_i individus (parcelles de pomme de terre) d'un échantillon avec $n_i = 4$, et μ_i et σ inconnus. Les individus étant tous différents, les échantillons sont indépendants. On considère les hypothèses :

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ contre

H_1 : il existe au moins deux moyennes différentes.

On utilise le test ANOVA à un facteur. Complétons maintenant le tableau ANOVA.

On a $n = \sum_{i=1}^p n_i = 20$,

$$\bar{x} = \frac{29.6+33.7+33.6+29.3+35.2+33.3+31.5+36.1+28.0+35.7+36.8+39.3+30.6+34.6+37.1+31.2+27.8+22.7+20.9+18.5}{4 \times 5} = 31.275,$$

$$\bar{x}_1 = \frac{29.6+33.7+33.6+29.3}{4} = 31.55, \bar{x}_2 = \frac{35.2+33.3+31.5+36.1}{4} = 34.025,$$

$$\bar{x}_3 = \frac{28.0+35.7+36.8+39.3}{4} = 34.95, \bar{x}_4 = \frac{30.6+34.6+37.1+31.2}{4} = 33.375, \bar{x}_5 = \frac{27.8+22.7+20.9+18.5}{4} = 22.475,$$

$$sce_T = sce_F + sce_R = 411.98 + 175.98 = 587.96,$$

$$ddl_T = n - 1 = 20 - 1 = 19, ddl_F = p - 1 = 5 - 1 = 4, ddl_R = n - p = 20 - 5 = 15,$$

$$cm_F = \frac{sce_F}{ddl_F} = \frac{411.98}{4} = 102.995,$$

$$cm_R = \frac{sce_R}{ddl_R} = \frac{175.98}{15} = 11.732$$

$$\text{Et } f_{obs} = \frac{cm_F}{cm_R} = \frac{102.995}{11.732} = 8.7789.$$

D'où le tableau

	sce	ddl	cm	f_{obs}
Total	587.96	19		
Factoriel	411.98	4	102.995	8.7789
Résiduel	175.98	15	11.732	

Remarquons que $\alpha = 0.05$ (5%).

Il faut ensuite déterminer le réel $f_\alpha(v_1, v_2)$ tel que $\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha = 0.05$.

Où $F \sim \mathcal{F}(v_1, v_2)$, $(v_1, v_2) = (p - 1, n - p) = (4, 15)$.

L'annexe 6 donne $f_\alpha(4, 15) = 3.06$ et comme $f_{obs} = 8.79 > 3.06$, alors on rejette H_0 .

Ainsi, au risque 5%, on peut affirmer que la nature du fongicide influe sur le rendement moyen de pomme de terre.

Exercice 2

On dispose de 3 marques de saumons : S_1 , S_2 et S_3 . Afin de comparer la qualité du saumon de ces 3 marques, deux groupes d'individus différents, G_1 et G_2 , sont sollicités. Pour chaque marque, 4 individus de chaque groupe, tous différents, doivent goûter le saumon et lui attribuer une note entre 0 et 20. Les résultats sont :

$G_1 :$	S1	S2	S3	$G_2 :$	S1	S2	S3
	15	11	16		18	15	19
	14	9	16		16	5	13
	13	11	14		10	14	16
	14	9	14		12	6	12

On suppose que, $\forall i \in \{1,2,3\}$, la note du saumon de la marque S_i est une variable $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$, avec μ_i et σ inconnus.

Les tableaux ANOVA, incomplets, correspondant à chacun des groupes, sont :

$G_1 :$		sce	ddl	cm	f_{obs}
	Total	66			
	Factoriel				
	Residuel	10			

$G_2 :$		sce	ddl	cm	f_{obs}
	Total	208			
	Factoriel	56			
	Residuel				

- $\forall i \in \{1,2,3\}$, comparer les moyennes des notes obtenues entre G_1 et G_2 quant à S_i .

Que constatez-vous ?

Peut-on d'ores et déjà dire que le test ANOVA avec :

$H_0: \mu_1 = \mu_2 = \mu_3$ contre

H_1 : il existe au moins 2 moyennes différentes.

donnera la même conclusion pour G_1 et G_2 ? Justifier votre réponse.

- Recopier et compléter les tableaux ANOVA.
- Au final, peut-on affirmer, au risque 5%, que S_1 , S_2 et S_3 sont perçues de manière différente par les consommateurs ?

Exercice 2 Solution

1. Les moyennes sont confectionnées dans les dernières lignes de chaque tableau :

	S1	S2	S3
	15	11	16
	14	9	16
	13	11	14
	14	9	14
	14	10	15

$G1 :$

	S1	S2	S3
	18	15	19
	16	5	13
	10	14	16
	12	6	12
	14	10	15

$G2 :$

Obtenues ainsi :

$$G1: \bar{x}_1 = \frac{15+14+13+14}{4}, \bar{x}_2 = \frac{11+9+11+9}{4}, \bar{x}_3 = \frac{16+16+14+14}{4},$$

$$G2: \bar{x}_1 = \frac{18+16+10+12}{4}, \bar{x}_2 = \frac{15+5+14+6}{4}, \bar{x}_3 = \frac{19+13+16+12}{4}.$$

On remarque que, pour $i \in \{1,2,3\}$, la moyenne des valeurs correspondantes à S_i de $G1$ est égale à celle de $G2$. Toutefois, on ne peut rien dire sur la réalisation du test ANOVA car il utilise des variances (carrés moyens), non pas des moyennes.

On peut donc avoir des conclusions différentes sans qu'il n'y ait de contradiction.

2. Complétons maintenant les tableaux ANOVA :

• Tableau ANOVA $G1$:

$$\text{On a } n = \sum_{i=1}^p n_i = 12,$$

$$sce_F = sce_T - sce_R = 66 - 10 = 56,$$

$$ddl_T = n - 1 = 12 - 1 = 11, ddl_F = p - 1 = 3 - 1 = 2, ddl_R = n - p = 12 - 3 = 9,$$

$$cm_F = \frac{sce_F}{ddl_F} = \frac{56}{2} = 28,$$

$$cm_R = \frac{sce_R}{ddl_R} = \frac{10}{9} = 1.1111$$

$$\text{Et } f_{obs} = \frac{cm_F}{cm_R} = \frac{28}{1.1111} = 25.2002.$$

D'où le tableau

	sce	ddl	cm	f_{obs}
Total	66	11		
Factoriel	56	2	28	25.2002
Residuel	10	9	1.1111	

$G1 :$

- Tableau ANOVA $G2$:

On a $n = \sum_{i=1}^p n_i = 12$,

$$sce_R = sce_T - sce_F = 208 - 56 = 152,$$

$$ddl_T = n - 1 = 12 - 1 = 11, ddl_F = p - 1 = 3 - 1 = 2, ddl_R = n - p = 12 - 3 = 9,$$

$$cm_F = \frac{sce_F}{ddl_F} = \frac{56}{2} = 28,$$

$$cm_R = \frac{sce_R}{ddl_R} = \frac{152}{9} = 16.8888$$

$$\text{Et } f_{obs} = \frac{cm_F}{cm_R} = \frac{28}{16.8888} = 1.6579.$$

D'où le tableau

$G2$:

	sce	ddl	cm	f_{obs}
Total	208	11		
Factoriel	56	2	28	1.6579
Residuel	152	9	16.8888	

3. On considère les hypothèses :

$$H_0: \mu_1 = \mu_2 = \mu_3 \text{ contre}$$

$$H_1: \text{il existe au moins 2 moyennes différentes.}$$

On utilise le test ANOVA à un facteur.

Remarquons que $\alpha = 0.05$ (5%).

Il faut ensuite déterminer le réel $f_\alpha(v_1, v_2)$ tel que $\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha = 0.05$.

Où $F \sim \mathcal{F}(v_1, v_2)$, $(v_1, v_2) = (p - 1, n - p) = (3 - 1, 12 - 3) = (2, 9)$.

L'annexe 6 donne $f_\alpha(2, 9) = 4.26$.

Par le résultat de la question 2,

- Pour $G1$, on a $f_{obs} = 25.2002 > 4.26$, alors on rejette H_0 .
Ainsi, au risque 5%, on peut affirmer que $S1$, $S2$ et $S3$ sont perçues de manière différente par les consommateurs.
- Pour $G2$, on a $f_{obs} = 1.6579 < 4.26$, alors on ne rejette pas H_0 .
Ainsi, au risque 5%, on ne peut pas affirmer que $S1$, $S2$ et $S3$ sont perçues de manière différente par les consommateurs.

Au final, vu le résultat de $G1$, on peut affirmer, au risque 5%, que $S1$, $S2$ et $S3$ sont perçues de manières différente par les consommateurs.

Exercice 3

Une enquête sur la consommation annuelle des ménages est réalisée par l'I.N.S.E.E. régulièrement. Ces ménages sont répartis en 5 grandes catégories suivant leur localisation :

- $C1$: ménages en zone rurale,
- $C2$: ménages résidant dans une unité urbaine inférieure à 20000 habitants,
- $C3$: ménages résidant dans une unité urbaine comprise entre 20000 habitants et 100000 habitants,
- $C4$: ménages résidant dans une unité urbaine supérieure à 100000 habitants autre que l'agglomération parisienne,
- $C5$: ménages résidant dans l'agglomération parisienne.

Un groupement commercial s'intéresse particulièrement à la consommation annuelle des produits contenus dans la nomenclature 17 de l'I.N.S.E.E., c'est-à-dire, la consommation annuelle en mouton, agneau et chevreau et il souhaite savoir s'il y a un effet "localisation" sur la consommation annuelle moyenne des ménages pour ces produits. Le groupement commercial interroge un échantillon de 5 ménages par catégories. Les résultats, en euros, sont :

C1	C2	C3	C4	C5
56	47	55	61	69
66	50	51	62	71
54	55	59	54	55
61	46	54	54	62
56	56	59	62	53

On suppose que, $\forall i \in \{1, \dots, 5\}$, la consommation annuelle d'un ménage en euros de catégorie C_i est une variable $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$, avec μ_i et σ inconnus.

Le tableau ANOVA, incomplet, est reproduit ci-dessous :

	sce	ddl	cm	f_{obs}
Total	908.64			
Factoriel				
Résiduel	556.40			

1. Faire le test ANOVA avec :
 $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ contre
 H_1 : il existe au moins 2 moyennes différentes.
 Au risque 5%. Interpréter le résultat.
2. Peut-on affirmer, au risque 5%, que $\mu_3 \neq \mu_4$?

Exercice 3 Solution

1. Le facteur étudié est "catégorie" avec pour modalités : $C1, C2, C3, C4, C5$. Donc $p = 5$.

Par l'énoncé, $\forall i \in \{1, \dots, p\}$, on observe la valeur de $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$ pour chacun des n_i individus (ménages) d'un échantillon avec $n_i = 5$, et avec μ_i et σ inconnus. On considère les hypothèses :

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ contre

H_1 : il existe au moins 2 moyennes différentes.

On utilise le test ANOVA à un facteur. On a

$$n = \sum_{i=1}^p n_i = 25,$$

$$sce_F = sce_T - sce_R = 908.64 - 556.40 = 352.24,$$

$$ddl_T = n - 1 = 25 - 1 = 24, ddl_F = p - 1 = 5 - 1 = 4, ddl_R = n - p = 25 - 5 = 20,$$

$$cm_F = \frac{sce_F}{ddl_F} = \frac{352.24}{4} = 88.06,$$

$$cm_R = \frac{sce_R}{ddl_R} = \frac{556.40}{20} = 27.82$$

$$\text{Et } f_{obs} = \frac{cm_F}{cm_R} = \frac{88.06}{27.82} = 3.1653.$$

D'où le tableau

	sce	ddl	cm	f_{obs}
Total	908.64	24		
Factoriel	352.24	4	88.06	3.1653
Résiduel	556.40	20	27.82	

Remarquons que $\alpha = 0.05$ (5%).

Il faut ensuite déterminer le réel $f_\alpha(v_1, v_2)$ tel que $\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha = 0.05$.

Où $F \sim \mathcal{F}(v_1, v_2)$, $(v_1, v_2) = (p - 1, n - p) = (5 - 1, 25 - 5) = (4, 20)$.

L'annexe 6 donne $f_\alpha(4, 20) = 2.87$ et comme $f_{obs} = 3.1653 > 2.87$, alors on rejette H_0 .

On peut affirmer, au risque 5%, que les consommations moyennes diffèrent quant à la localisation des ménages.

2. On considère les hypothèses :

$H_0: \mu_3 = \mu_4$ contre $H_1: \mu_3 \neq \mu_4$.

On va utiliser le test de Bonferroni.

On pose $k = 3$ et $l = 4$. On a $\bar{x}_k = 55.6$, $\bar{x}_l = 58.6$, $s_R = \sqrt{cm_R} = \sqrt{27.82} = 5.2744$ et

$$t_{obs} = \frac{\bar{x}_k - \bar{x}_l}{s_R \sqrt{\frac{1}{n_k} + \frac{1}{n_l}}} = \frac{55.6 - 58.6}{5.2744 \sqrt{\frac{1}{5} + \frac{1}{5}}} = -0.8993$$

Remarquons que $\alpha = 0.05$ (5%).

Il faut ensuite déterminer le réel t_{α}^{**} tel que

$$\mathbb{P}(|T| \geq t_{\alpha}^{**}(v)) = \frac{2\alpha}{p(p-1)} = \frac{2 \times 0.05}{5(5-1)} = 0.005,$$

où $T \sim \mathcal{T}(v)$, $v = n - p = 25 - 5 = 20$.

L'annexe 3 donne l'encadrement :

$$t_{\alpha}^{**}(v) \in]2.845, 3.85[,$$

et comme

$$t_{obs} = 0.8993 < 2.845 < t_{\alpha}^{**}(v),$$

On ne rejette pas H_0 .

Par conséquent, au risque 5%, les données ne nous permettent pas d'affirmer que $\mu_3 \neq \mu_4$.

Exercice 4

Une étude a été menée en vue de déterminer la teneur en oxygène dissout dans l'eau sur quatre sites différents : *Site 1*, *Site 2*, *Site 3* et *Site 4*. Sur chaque site, 5 échantillons ont été sélectionnés aléatoirement. Les résultats sont :

Site 1	Site 2	Site 3	Site 4
5.9	6.3	5.7	6.0
6.1	6.6	6.2	6.2
6.3	6.4	6.0	6.1
6.1	6.4	5.9	5.8
6.0	6.5	6.4	6.0

On suppose que, $\forall i \in \{1, \dots, 4\}$, la teneur en oxygène dissout dans l'eau du *Site i* est une variable $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, avec μ_i et σ inconnus.

Le tableau ANOVA, incomplet, est reproduit ci-dessous :

	sce	ddl	cm	f_{obs}
Total	1.1095			
Factoriel	0.59			
Résiduel				

Peut-on affirmer, au risque 5%, qu'il y a au moins un des 4 sites qui diffère quant à la teneur moyenne en oxygène dissout dans l'eau ?

Exercice 4 Solution

Le facteur étudié est "site" avec pour modalités : *Site 1*, *Site 2*, *Site 3* et *Site 4*. Donc $p = 4$.

Par l'énoncé, $\forall i \in \{1, \dots, p\}$, on observe la valeur de $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ pour chacun des n_i individus (échantillons d'eau) d'un échantillon avec $n_i = 5$, et μ_i et σ inconnus. Les individus étant tous différents, les échantillons sont indépendants.

On considère les hypothèses :

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ contre

H_1 : il existe au moins deux moyennes différentes.

On utilise le test ANOVA à un facteur.

On a $n = \sum_{i=1}^p n_i = 20$,

$$sce_R = sce_T - sce_F = 1.1095 - 0.5195,$$

$$ddl_T = n - 1 = 20 - 1 = 19, ddl_F = p - 1 = 4 - 1 = 3, ddl_R = n - p = 20 - 4 = 16,$$

$$cm_F = \frac{sce_F}{ddl_F} = \frac{0.59}{3} = 0.1966, cm_R = \frac{sce_R}{ddl_R} = \frac{0.5195}{16} = 0.0324$$

$$\text{Et } f_{obs} = \frac{cm_F}{cm_R} = \frac{0.1966}{0.0324} = 6.0679.$$

D'où le tableau :

	sce	ddl	cm	f_{obs}
Total	1.1095	19		
Factoriel	0.59	3	0.1966	6.0679
Résiduel	0.5195	16	0.0324	

Remarquons que $\alpha = 0.05$ (5%).

Il faut ensuite déterminer le réel $f_\alpha(v_1, v_2)$ tel que $\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha = 0.05$.

Où $F \sim \mathcal{F}(v_1, v_2)$, $(v_1, v_2) = (p - 1, n - p) = (4 - 1, 20 - 4) = (3, 16)$.

L'annexe 6 donne $f_\alpha(3, 16) = 3.29$ et comme

$$f_{obs} = 6.0679 > 3.29 = f_\alpha(v_1, v_2),$$

Alors on rejette H_0 .

On peut affirmer, au risque 5%, qu'il y a au moins un des 4 sites qui diffère quant à la teneur moyenne en oxygène dissout dans l'eau.

Exercice 5

On veut comparer trois milieux de culture $M1$, $M2$ et $M3$ quant au développement de bactéries. Pour le milieu $M1$, on examine 3 boîtes de Pétri, pour le milieu $M2$ 4 boîtes Pétri et pour le milieu $M3$ 2 boîtes de Pétri. Les résultats sont :

M1	M2	M3
9	8	12
12	13	14
12	12	
	7	

On suppose que, $\forall i \in \{1,2,3\}$, le nombre de milliers de bactéries dans M_i est une variable $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ avec μ_i et σ inconnus.

La tableau ANOVA, incomplet, est reproduit ci-dessous :

	sce	ddl	cm	f_{obs}
Total	46			
Factoriel			6	
Résiduel				

Y a-t-il un effet "milieu" significatif sur le nombre moyen de bactérie ?

Exercice 5 Solutions

Le facteur étudié est “milieu de culture” avec pour modalités : $M1$, $M2$ et $M3$. Donc $p = 3$. Par l'énoncé, $\forall i \in \{1, \dots, p\}$, on observe la valeur de $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ pour chacun des n_i individus (boîtes de Pétri) d'un échantillon avec $n_1 = 3, n_2 = 4, n_3 = 2$ et μ_i et σ inconnus. Les individus étant tous différents, les échantillons sont indépendants.

On considère les hypothèses :

$H_0: \mu_1 = \mu_2 = \mu_3$ contre

H_1 : il existe au moins deux moyennes différentes.

On utilise le test ANOVA à un facteur.

On a $n = \sum_{i=1}^p n_i = 9$,

$ddl_T = n - 1 = 9 - 1 = 8, ddl_F = p - 1 = 3 - 1 = 2, ddl_R = n - p = 9 - 3 = 6$,

$sce_F = cm_F \times ddl_F = 6 \times 2 = 12$,

$sce_R = sce_T - sce_F = 46 - 12 = 34$,

$cm_R = \frac{sce_R}{ddl_R} = \frac{34}{6} = 5.6666$,

Et $f_{obs} = \frac{cm_F}{cm_R} = \frac{6}{5.6666} = 1.0588$.

	sce	ddl	cm	f_{obs}
Total	46	8		
Factoriel	12	2	6	1.0588
Résiduel	34	6	5.6666	

Remarquons que $\alpha = 0.05$ (5%).

Il faut ensuite déterminer le réel $f_\alpha(v_1, v_2)$ tel que $\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha = 0.05$.

Où $F \sim \mathcal{F}(v_1, v_2)$, $(v_1, v_2) = (p - 1, n - p) = (3 - 1, 9 - 3) = (2, 6)$.

L'annexe 6 donne $f_\alpha(2, 6) = 5.14$ et comme

$$f_{obs} = 1.0588 < 5.14 = f_\alpha(v_1, v_2),$$

Alors on ne rejette pas H_0 .

Par conséquent, en considérant un risque de 5%, les données ne permettent pas de conclure que le milieu influe sur le développement des colonies de bactéries.

Annexes

Annexe 1 : Loi normale

Soit $Z \sim \mathcal{N}(0,1)$. La table ci-dessous donne, pour un α choisi, la valeur z_α telle que $\mathbb{P}(|Z| \geq z_\alpha) = \alpha$.

α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	∞	2.576	2.326	2.170	2.054	1.960	1.881	1.812	1.751	1.695
0.10	1.645	1.598	1.555	1.514	1.476	1.440	1.405	1.372	1.341	1.311
0.20	1.282	1.254	1.227	1.200	1.175	1.150	1.126	1.103	1.080	1.058
0.30	1.036	1.015	0.994	0.974	0.954	0.935	0.915	0.896	0.878	0.860
0.40	0.842	0.824	0.806	0.789	0.772	0.755	0.739	0.722	0.706	0.690
0.50	0.674	0.659	0.643	0.628	0.613	0.598	0.583	0.568	0.553	0.539
0.60	0.524	0.510	0.496	0.482	0.468	0.454	0.440	0.426	0.412	0.399
0.70	0.385	0.372	0.358	0.345	0.332	0.319	0.305	0.292	0.279	0.266
0.80	0.253	0.240	0.228	0.215	0.202	0.189	0.176	0.164	0.151	0.138
0.90	0.126	0.113	0.100	0.088	0.075	0.063	0.050	0.038	0.025	0.013

Annexe 2 : Loi normale (bis)

Soit $Z \sim \mathcal{N}(0,1)$. La table ci-dessous donne, pour un z choisi, la valeur α telle que $\mathbb{P}(|Z| \geq z) = \alpha$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	1.00000	0.99202	0.98404	0.97607	0.96809	0.96012	0.95216	0.94419	0.93624	0.92829
0.10	0.92034	0.91241	0.90448	0.89657	0.88866	0.88076	0.87288	0.86501	0.85715	0.84931
0.20	0.84148	0.83367	0.82587	0.81809	0.81033	0.80259	0.79486	0.78716	0.77948	0.77182
0.30	0.76418	0.75656	0.74897	0.74140	0.73386	0.72634	0.71885	0.71138	0.70395	0.69654
0.40	0.68916	0.68181	0.67449	0.66720	0.65994	0.65271	0.64552	0.63836	0.63123	0.62413
0.50	0.61708	0.61005	0.60306	0.59611	0.58920	0.58232	0.57548	0.56868	0.56191	0.55519
0.60	0.54851	0.54186	0.53526	0.52869	0.52217	0.51569	0.50925	0.50286	0.49650	0.49019
0.70	0.48393	0.47770	0.47152	0.46539	0.45930	0.45325	0.44725	0.44130	0.43539	0.42953
0.80	0.42371	0.41794	0.41222	0.40654	0.40091	0.39533	0.38979	0.38430	0.37886	0.37347
0.90	0.36812	0.36282	0.35757	0.35237	0.34722	0.34211	0.33706	0.33205	0.32709	0.32217
1.00	0.31731	0.31250	0.30773	0.30301	0.29834	0.29372	0.28914	0.28462	0.28014	0.27571
1.10	0.27133	0.26700	0.26271	0.25848	0.25429	0.25014	0.24605	0.24200	0.23800	0.23405
1.20	0.23014	0.22628	0.22246	0.21870	0.21498	0.21130	0.20767	0.20408	0.20055	0.19705
1.30	0.19360	0.19020	0.18684	0.18352	0.18025	0.17702	0.17383	0.17069	0.16759	0.16453
1.40	0.16151	0.15854	0.15561	0.15272	0.14987	0.14706	0.14429	0.14156	0.13887	0.13622
1.50	0.13361	0.13104	0.12851	0.12602	0.12356	0.12114	0.11876	0.11642	0.11411	0.11183
1.60	0.10960	0.10740	0.10523	0.10310	0.10101	0.09894	0.09691	0.09492	0.09296	0.09103
1.70	0.08913	0.08727	0.08543	0.08363	0.08186	0.08012	0.07841	0.07673	0.07508	0.07345
1.80	0.07186	0.07030	0.06876	0.06725	0.06577	0.06431	0.06289	0.06148	0.06011	0.05876
1.90	0.05743	0.05613	0.05486	0.05361	0.05238	0.05118	0.05000	0.04884	0.04770	0.04659
2.00	0.04550	0.04443	0.04338	0.04236	0.04135	0.04036	0.03940	0.03845	0.03753	0.03662
2.10	0.03573	0.03486	0.03401	0.03317	0.03235	0.03156	0.03077	0.03001	0.02926	0.02852
2.20	0.02781	0.02711	0.02642	0.02575	0.02509	0.02445	0.02382	0.02321	0.02261	0.02202
2.30	0.02145	0.02089	0.02034	0.01981	0.01928	0.01877	0.01827	0.01779	0.01731	0.01685
2.40	0.01640	0.01595	0.01552	0.01510	0.01469	0.01429	0.01389	0.01351	0.01314	0.01277
2.50	0.01242	0.01207	0.01174	0.01141	0.01109	0.01077	0.01047	0.01017	0.00988	0.00960
2.60	0.00932	0.00905	0.00879	0.00854	0.00829	0.00805	0.00781	0.00759	0.00736	0.00715
2.70	0.00693	0.00673	0.00653	0.00633	0.00614	0.00596	0.00578	0.00561	0.00544	0.00527
2.80	0.00511	0.00495	0.00480	0.00465	0.00451	0.00437	0.00424	0.00410	0.00398	0.00385
2.90	0.00373	0.00361	0.00350	0.00339	0.00328	0.00318	0.00308	0.00298	0.00288	0.00279
3.00	0.00270	0.00261	0.00253	0.00245	0.00237	0.00229	0.00221	0.00214	0.00207	0.00200
3.10	0.00194	0.00187	0.00181	0.00175	0.00169	0.00163	0.00158	0.00152	0.00147	0.00142
3.20	0.00137	0.00133	0.00128	0.00124	0.00120	0.00115	0.00111	0.00108	0.00104	0.00100
3.30	0.00097	0.00093	0.00090	0.00087	0.00084	0.00081	0.00078	0.00075	0.00072	0.00070
3.40	0.00067	0.00065	0.00063	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050	0.00048
3.50	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00037	0.00036	0.00034	0.00033
3.60	0.00032	0.00031	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024	0.00023	0.00022
3.70	0.00022	0.00021	0.00020	0.00019	0.00018	0.00018	0.00017	0.00016	0.00016	0.00015
3.80	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011	0.00011	0.00010	0.00010
3.90	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00007	0.00007	0.00007	0.00007
4.00	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005	0.00005	0.00005	0.00004

Annexe 3 : Loi de Student

Soit $T \sim \mathcal{T}(v)$ La table ci-dessous donne, pour un α et un v choisis, la valeur $t_\alpha(v)$ telle que $\mathbb{P}(|T| \geq t_\alpha(v)) = \alpha$.

Si $v \geq 31$, alors $t_\alpha(v) \approx z_\alpha$ défini dans la table 1 (Loi normale).

$\nu \backslash \alpha$	0.90	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	0.158	1.000	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.142	0.816	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.137	0.765	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.134	0.741	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.727	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.718	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.711	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.706	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.703	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.700	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.697	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.695	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.694	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.692	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.691	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.690	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.689	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.688	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.688	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.687	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.686	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.686	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.685	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.127	0.685	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.684	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.684	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.684	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.683	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.683	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.683	1.055	1.310	1.697	2.042	2.457	2.750	3.646

Annexe 4 : Loi du chi-deux

Soit $K \sim \chi^2_\alpha(\nu)$. La table ci-dessous donne, pour un α et un ν choisis, la valeur $\chi^2_\alpha(\nu)$ telle que $\mathbb{P}(K \geq \chi^2_\alpha(\nu)) = \alpha$.

$\nu \backslash \alpha$	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.001
1	0.0002	0.001	0.004	0.016	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	13.82
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	16.27
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	18.47
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	20.51
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	22.46
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	24.32
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	26.12
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	27.88
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	29.59
11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	31.26
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	32.91
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	34.53
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	36.12
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	37.70
16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	39.25
17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	40.79
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	42.31
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	43.82
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	45.31
21	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	46.80
22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	48.27
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	49.73
24	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	51.18
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	52.62
26	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	54.05
27	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	55.48
28	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	56.89
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	58.30
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	59.70

Annexe 5 : Loi de Fisher I ($\alpha = 0.025$)

Soit $F \sim \mathcal{F}(v_1, v_2)$. La table ci-dessous donne, pour un v_1 et un v_2 choisis, la valeur $f_\alpha(v_1, v_2)$ telle que $\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha = 0.025$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	10	15	20	30
1	648	800	864	900	922	937	957	969	985	993	1001
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.5
3	17.4	16.0	15.4	15.1	14.9	14.7	14.5	14.4	14.3	14.2	14.1
4	12.2	10.6	9.98	9.60	9.36	9.20	8.98	8.84	8.66	8.56	8.46
5	10.0	8.43	7.76	7.39	7.15	6.98	6.76	6.62	6.43	6.33	6.23
6	8.81	7.26	6.60	6.23	5.99	5.82	5.60	5.46	5.27	5.17	5.07
7	8.07	6.54	5.89	5.52	5.29	5.12	4.90	4.76	4.57	4.47	4.36
8	7.57	6.06	5.42	5.05	4.82	4.65	4.43	4.30	4.10	4.00	3.89
9	7.21	5.71	5.08	4.72	4.48	4.32	4.10	3.96	3.77	3.67	3.56
10	6.94	5.46	4.83	4.47	4.24	4.07	3.85	3.72	3.52	3.42	3.31
11	6.72	5.26	4.63	4.28	4.04	3.88	3.66	3.53	3.33	3.23	3.12
12	6.55	5.10	4.47	4.12	3.89	3.73	3.51	3.37	3.18	3.07	2.96
13	6.41	4.97	4.35	4.00	3.77	3.60	3.39	3.25	3.05	2.95	2.84
14	6.30	4.86	4.24	3.89	3.66	3.50	3.29	3.15	2.95	2.84	2.73
15	6.20	4.76	4.15	3.80	3.58	3.41	3.20	3.06	2.86	2.76	2.64
16	6.12	4.69	4.08	3.73	3.50	3.34	3.12	2.99	2.79	2.68	2.57
17	6.04	4.62	4.01	3.66	3.44	3.28	3.06	2.92	2.72	2.62	2.50
18	5.98	4.56	3.95	3.61	3.38	3.22	3.01	2.87	2.67	2.56	2.44
19	5.92	4.51	3.90	3.56	3.33	3.17	2.96	2.82	2.62	2.51	2.39
20	5.87	4.46	3.86	3.51	3.29	3.13	2.91	2.77	2.57	2.46	2.35
22	5.79	4.38	3.78	3.44	3.22	3.05	2.84	2.70	2.50	2.39	2.27
24	5.72	4.32	3.72	3.38	3.15	2.99	2.78	2.64	2.44	2.33	2.21
26	5.66	4.27	3.67	3.33	3.10	2.94	2.73	2.59	2.39	2.28	2.16
28	5.61	4.22	3.63	3.29	3.06	2.90	2.69	2.55	2.34	2.23	2.11
30	5.57	4.18	3.59	3.25	3.03	2.87	2.65	2.51	2.31	2.20	2.07
40	5.42	4.05	3.46	3.13	2.90	2.74	2.53	2.39	2.18	2.07	1.94
50	5.34	3.98	3.39	3.06	2.83	2.67	2.46	2.32	2.11	1.99	1.87
60	5.29	3.93	3.34	3.01	2.79	2.63	2.41	2.27	2.06	1.94	1.82
80	5.22	3.86	3.28	2.95	2.73	2.57	2.36	2.21	2.00	1.88	1.75
100	5.18	3.83	3.25	2.92	2.70	2.54	2.32	2.18	1.97	1.85	1.71

Annexe 6 : Loi de Fisher II ($\alpha = 0.05$)

Soit $F \sim \mathcal{F}(v_1, v_2)$. La table ci-dessous donne, pour un v_1 et un v_2 choisis, la valeur $f_\alpha(v_1, v_2)$ telle que $\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha = 0.05$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	10	15	20	30
1	161	300	216	225	230	234	239	242	246	248	250
2	18.5	19	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.85	8.79	8.70	8.66	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96	5.86	5.80	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74	4.62	4.56	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06	3.94	3.87	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.64	3.51	3.44	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.35	3.22	3.15	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.14	3.01	2.94	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.98	2.85	2.77	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.85	2.72	2.65	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.85	2.75	2.62	2.54	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.77	2.67	2.53	2.46	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.60	2.46	2.39	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.54	2.40	2.33	2.25
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.49	2.35	2.28	2.19
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.45	2.31	2.23	2.15
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.41	2.27	2.19	2.11
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.38	2.23	2.16	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35	2.20	2.12	2.04
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.30	2.15	2.07	1.98
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.25	2.11	2.03	1.94
26	4.23	3.37	2.98	2.74	2.59	2.47	2.32	2.22	2.07	1.99	1.90
28	4.20	3.34	2.95	2.71	2.56	2.45	2.29	2.19	2.04	1.96	1.87
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.16	2.01	1.93	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.08	1.92	1.84	1.74
50	4.03	3.18	2.79	2.56	2.40	2.29	2.13	2.03	1.87	1.78	1.69
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.99	1.84	1.75	1.65
80	3.96	3.11	2.72	2.49	2.33	2.21	2.06	1.95	1.79	1.70	1.60
100	3.94	3.09	2.70	2.46	2.31	2.19	2.03	1.93	1.77	1.68	1.57

Annexe 7 : Valeurs de Cochran

Les valeurs intérieures du tableau ci-dessous donnent les coefficient $c(m, p)$ utilisé dans le test de Cochran. Le risque est fixé à 5%. Ici, m est le nombre de données par échantillon et p le nombre d'échantillons.

$p \backslash m$	2	3	4	5	6	7	8
2	0.999	0.975	0.939	0.906	0.877	0.863	0.833
3	0.967	0.871	0.798	0.746	0.707	0.677	0.653
4	0.907	0.768	0.684	0.629	0.59	0.56	0.537
5	0.841	0.684	0.598	0.544	0.507	0.478	0.456
6	0.781	0.616	0.532	0.48	0.445	0.418	0.398
7	0.727	0.561	0.48	0.431	0.397	0.373	0.354
8	0.680	0.516	0.438	0.391	0.36	0.336	0.319
9	0.639	0.478	0.403	0.358	0.329	0.307	0.29
10	0.602	0.445	0.373	0.331	0.303	0.282	0.267
12	0.541	0.392	0.326	0.288	0.262	0.244	0.23
15	0.471	0.335	0.276	0.242	0.22	0.203	0.191
20	0.389	0.271	0.221	0.192	0.174	0.106	0.15
24	0.343	0.235	0.191	0.166	0.149	0.137	0.129
30	0.293	0.198	0.159	0.138	0.124	0.114	0.106
40	0.237	0.158	0.126	0.108	0.097	0.089	0.083

Sources

- Le contenu de ce cours s'est basé sur l'enseignement et les supports écrits par [Christophe Chesneau](#). Ces supports ont été utilisés et modifiés avec son accord, dans le but de dispenser ce cours. Merci à lui pour son aide.
- Les exercices proviennent des livres du même auteur, listés sur <https://chesneau.users.lmno.cnrs.fr/>.

Pour la pratique des ANOVA dans R :

- le livre [R Cookbook, 2nd Edition, James \(JD\) Long, Paul Teetor, 2019-09-26](#)

Des lectures complémentaires pourront vous intéresser :

- Fisher RA. [Statistical Methods for Research Workers](#). Edinburgh, United Kingdom: Oliver & Boyd; 1925.
- Sur [link.springer](#) ou de nombreux autres supports de cours sur internet...