

Labo 12: Comparaison de moyenne / Analyse de la variance

Visseho Adjiwanou, PhD.

07 April 2021

Question 1

Supposons que vous souhaitiez en savoir plus sur la façon dont le fait de vivre sur le campus pendant la première année influence la satisfaction globale à l'égard de l'expérience universitaire. Vous échantillonnez 30 personnes: 10 qui vivaient sur le campus, 10 qui vivaient près du campus et 10 qui ont fait la navette vers le campus. Vous demandez ensuite aux répondants d'indiquer leur satisfaction globale à l'égard de leur expérience universitaire sur une échelle de 1 à 10. Vos données donnent les résultats suivants:

$$SC_{Inter} = 93.3 \quad SC_{Intra} = 29.7$$

Utiliser ces informations pour répondre aux questions suivantes:

1. Quelles sont les variables à l'étude et quels sont leur type?
 2. Quelle est la valeur du carré moyen intergroupe (encore appelé variance intergroupe)?
 3. Quelle est la valeur du carré moyen intragroupe (encore appelé variance intragroupe)?
 4. Quelle est la valeur de F?
 5. Quelle est la valeur de E^2 carré (E^2)?
 6. Est-ce que les résultats sont significatifs au seuil de 5%? 1%? (Que comprenez-vous pas résultat significatif?)
- Présenter votre hypothèse nulle

Question 2: Analyse de variance avec vrai données

Baumann et Jones (Moore, D. S. and McCabe, G. P. (1993) Introduction to the Practice of Statistics, Second Edition. New York: Freeman) ont conduit une étude expérimentale dans laquelle 66 enfants ont été assignés de manière aléatoire dans trois groupes différents. Les groupes représentent différentes méthodes d'enseignement: une méthode standard appelée **Basal** et deux nouvelles méthodes appelées **DTRA** et **Strat**. Les données sont présentées dans la base de donnée **Baumann** qui se trouve dans le package **carData**. Cette base de données contient les informations suivantes:

Variables	Description
group	Le groupe de l'enfant (Basal, DTRA, Strat)
pretest.1	Pré-test 1 de compréhension de lecture
pretest.2	Pré-test 2 de compréhension de lecture
post.test.1	Post-test 1 de compréhension de lecture
post.test.2	Post-test 2 de compréhension de lecture
post.test.3	Post-test 3 de compréhension de lecture

Remarque: Chaque package contient un certain nombres de bases de données pour permettre de tester

facilement les fonctions du package. Le code présenté ci-dessous vous permet de charger ces données.

```
rm(list = ls())
#install.packages("carData")

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr 0.3.3
## v tibble 2.1.3      v dplyr 1.0.2
## v tidyr 1.0.2       v stringr 1.4.0
## v readr 1.3.1       v forcats 0.4.0

## Warning: package 'ggplot2' was built under R version 3.6.2
## Warning: package 'dplyr' was built under R version 3.6.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(carData)
library(summarytools)
```

```
## Registered S3 method overwritten by 'pryr':
##   method      from
##   print.bytes Rcpp

## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE):
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/Resources/
## library/tcltk/libs//tcltk.so'' had status 1
```

```
##
## Attaching package: 'summarytools'

## The following object is masked from 'package:tibble':
##
##   view
```

```
#data(package = "carData")      # Permet de savoir l'ensemble des bases de données inclus dans le pack
data("Baumann", package = "carData") # Charger la base de données Baumann
```

```
summary(Baumann)
```

```
##   group      pretest.1      pretest.2      post.test.1      post.test.2
## Basal:22  Min.   : 4.000   Min.   : 1.000   Min.   : 1.000   Min.   : 0.000
## DRTA :22  1st Qu.: 8.000   1st Qu.: 3.250   1st Qu.: 5.000   1st Qu.: 5.000
## Strat:22  Median : 9.000   Median : 5.000   Median : 8.000   Median : 6.000
##           Mean   : 9.788   Mean   : 5.106   Mean   : 8.076   Mean   : 6.712
##           3rd Qu.:12.000   3rd Qu.: 6.000   3rd Qu.:11.000   3rd Qu.: 8.000
##           Max.   :16.000   Max.   :13.000   Max.   :15.000   Max.   :13.000
## post.test.3
## Min.   :30.00
## 1st Qu.:40.00
## Median :45.00
## Mean   :44.02
## 3rd Qu.:49.00
## Max.   :57.00
```

1. Créer une nouvelle variable nommée **group1** où vous combinez ensemble les groupes **DRTA** et **Strat**. Ces deux groupes deviennent **DTST**.
2. On va s'intéresser par la suite aux variables `pretest.1` et `post.test.3`. Calculer la moyenne, la variance et l'écart-type des scores de compréhension avant le nouveau programme (`pretest.1`) et après le programme (`post.test.3`) pour les deux nouveaux groupes.
3. Présenter le diagramme à moustache (boxplot) de ces deux scores selon les catégories du **group1**.
4. Est-ce que le nouveau programme améliore l'apprentissage des élèves? Tester cela au seuil de 5%.
5. Reprendre les questions 2 à 4 maintenant en travaillant avec les trois groupes distincts de la variable **group**. La différence entre cette question et les questions précédentes est que cette fois-ci vous avez 3 groupes (**Basal**, **DTRA**, et **Strat**) alors que précédemment vous avez deux groupes (**Basal** et **DTST**)

Question 3 (A faire par vous-mêmes)

Supposons que vous souhaitiez en savoir plus sur l'isolement social. Vous émettez l'hypothèse que le fait de vivre dans des zones densément peuplées est associé à une socialisation plus fréquente avec des amis. Vous collectez des données auprès de 17 personnes des zones rurales (6), suburbaines (5) et urbaines (5). Le tableau ci-dessous contient vos données.

	Rural	Suburbain	Urbain	Total
Nombre de rencontres	3	1	3	
par mois avec des amis	4	3	2	
	2	1	4	
	1	3	1	
	1	4	5	
	2	2		

Utilisez ces informations pour répondre aux questions suivantes:

1. Quelles sont les variables à l'étude et quels sont leur type?
2. Calculer la somme des carrés intergroupe
3. Calculer la somme des carrés intragroupe
4. Calculer la statistique F
5. Est-ce que les résultats sont significatifs au seuil de 5%? et de 1%? (Que comprenez-vous pas résultat significatif?)
6. Quelle est la valeur de Eta carré (E^2)?
7. Maintenant, considérer uniquement deux groupes:
 - ceux qui vivent en milieu urbain et
 - ceux qui ne vivent pas en milieu urbain (Rural + Suburbain)
- 7.1. Dresser un nouveau tableau qui présente ces résultats.
- 7.2. Existe-il une relation significative au niveau de signification de 5% entre le fait de vivre en milieu urbain et le nombre de rencontres avec les amis par mois?