

# Test de chi-deux et test de student

Visseho Adjiwanou, PhD.

01 April 2021

## Question 1

Supposons que vous ayez deux groupes d'étudiants, un groupe de majors en mathématiques et un groupe de majors en philosophie. Vous voulez en savoir plus sur leurs capacités de pensée critique. Vous demandez à chaque groupe de passer un test. Les résultats montrent que les 28 majors en mathématiques ont en moyenne un score de 88 avec un écart type de 6 et que les 19 majors en philosophie ont en moyenne un score de 90 avec un écart type de 3. Utilisez ces informations pour déterminer ce qui suit: 1. Énoncer votre hypothèse nulle 2. L'erreur type de la différence entre les moyennes 3. La statistique t 4. La différence entre les majors en mathématiques et les majors en philosophie est-elle statistiquement significative au niveau de signification de 1%? Expliquer.

## Question 2

Dans cet exercice, nous examinons les différences transnationales dans les attitudes envers la violence domestique et l'accès à l'information. Nous explorons l'hypothèse selon laquelle il existe une association à un niveau global, entre la mesure dans laquelle les individus dans un pays ont accès à la connaissance et aux nouvelles informations, à la fois, à travers la scolarisation formelle et par les médias, et leur probabilité de condamner les actes de violence conjugale. Cet exercice est en partie basé sur:

Pierotti, Rachel. (2013). "Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion." *American Sociological Review*, 78: 240-265. **Un beau texte à lire si vous souhaitez.**

Nous utilisons les données des enquêtes démographiques et de santé (EDS), qui représentent un ensemble de plus de 300 enquêtes représentatives à l'échelle nationale, régionale et résidentielle menées dans des pays en développement du monde entier depuis 1992. Les enquêtes utilisent un plan d'échantillonnage en grappes, stratifié à deux degrés. Dans la première étape, les zones de dénombrement (ZD) sont tirées des fichiers de recensement. Dans la deuxième étape, dans chaque ZD, un échantillon de ménages est tiré d'une liste actualisée de ménages. En outre, les enquêtes ont des questionnaires identiques et des formations identiques pour les enquêteurs, permettant de comparer directement les données d'un pays avec celles collectées dans d'autres pays. Il est important de noter que différents groupes de pays sont interrogés chaque année.

Dans l'étude, l'auteur a utilisé ces données pour montrer que «les femmes ayant un meilleur accès aux scripts culturels mondiaux grâce à la vie urbaine, à l'enseignement secondaire ou à l'accès aux médias étaient plus susceptibles de rejeter la violence entre partenaires intimes». Les données se trouvent dans le fichier csv `dhs_ipv.csv`. Les noms et descriptions des variables sont les suivants:

Name	Description
<code>beat_goesout</code>	Pourcentage de femmes dans chaque pays qui pensent qu'un mari est justifié de battre sa femme si elle sort sans le lui dire.
<code>beat_burnfood</code>	Pourcentage de femmes dans chaque pays qui pensent qu'un mari a le droit de battre sa femme si elle brûle sa nourriture.

Name	Description
no_media	Pourcentage de femmes dans chaque pays qui ont rarement accès un journal, une radio ou une télévision.
sec_school	Pourcentage de femmes dans chaque pays ayant un niveau d'éducation secondaire ou supérieur.
year	Année de l'enquête
region	Région du monde
country	pays

Notez qu'il existe deux indicateurs d'*attitudes envers la violence domestique*: `beat_goesout` et `beat_burnfood`. Ce sont ces variables que nous voulons expliquer/comprendre. On les appelle des variables dépendantes ou variables à expliquer ou des *outcomes*. Il existe également deux indicateurs d'*accès à l'information*: `sec_school` et `no_media`. Celles-ci sont appelées des variables explicatives.

Comme toujours, il faut prévoir un premier *chunk* où vous installez vos packages, les chargez et chargez la bse de données. C'est une procédure qu'il faut toujours suivre. Sachez aussi que vous devez ouvrir ce fichier RMarkdown et travailler directement dedans.

```
rm(list = ls())      # Permet d'effacer l'environnement

# Chargement

library(tidyverse)

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

## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.0.6      v dplyr  1.0.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 3.6.2
## Warning: package 'tibble' was built under R version 3.6.2
## Warning: package 'tidyr' was built under R version 3.6.2
## Warning: package 'readr' was built under R version 3.6.2
## Warning: package 'purrr' was built under R version 3.6.2
## Warning: package 'dplyr' was built under R version 3.6.2
## Warning: package 'forcats' was built under R version 3.6.2

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

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

## For best results, restart R session and update pander using devtools:: or remotes::install_github('r
```

```
##
## Attaching package: 'summarytools'
## The following object is masked from 'package:tibble':
##
##      view
dhs <- read_csv("dhs_ipv.csv")

## Warning: Missing column names filled in: 'X1' [1]
##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   beat_burnfood = col_double(),
##   beat_goesout = col_double(),
##   sec_school = col_double(),
##   no_media = col_double(),
##   country = col_character(),
##   year = col_double(),
##   region = col_character()
## )
```

## T test

Est-ce que la violence diffère selon les régions ?

## T test

Comparer l'Afrique et l'Asie. Est-ce que la différence observée est due au hasard? Testé cela au niveau de signification de 5% ? Prenez soin de bien présenter votre hypothèse nulle et votre hypothèse alternative.

- Faites cela à la main et ensuite avec R.

## Question 3: degré de fondamentalisme par niveau d'éducation

Le tableau suivant donne la relation entre l'éducation et le degré du fondamentalisme

	libéral	Modéré	Fondamentaliste
Moins de 12 années	171	311	282
12 années	317	452	387
Plus de 12 années	839	947	628

1. Quelle est la valeur du chi-carré?
2. Est-ce que la relation observée est-elle valable au sein de la population?
3. Calculer les mesures d'association suivantes et justifier pourquoi on peut les calculer ou pas. Énumérer leurs limites éventuelles. 3.1. Le coefficient de contingence de Pearson 3.2. Le phi  $\phi$  3.3. Le Lambda 3.4. Le Dxy de Somer