Education

Gilchrist Ouedraogo

Chargement des librairies

```
library(tidyverse)
```

```
Warning: le package 'tidyverse' a été compilé avec la version R 4.3.3
```

```
Warning: le package 'ggplot2' a été compilé avec la version R 4.3.3
```

```
Warning: le package 'dplyr' a été compilé avec la version R 4.3.3
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(questionr)
```

```
Warning: le package 'questionr' a été compilé avec la version R 4.3.3
```

Chargement de la data

```r
df <- read_csv2("data/student-mat.csv")
```

```
i Using "',' as decimal and "'.'" as grouping mark. Use `read_delim()` for more control.

Rows: 395 Columns: 33
-- Column specification --------------------------------------------------------
Delimiter: ";"
chr (17): school, sex, address, famsize, Pstatus, Mjob, Fjob, reason, guardi...
dbl (16): age, Medu, Fedu, traveltime, studytime, failures, famrel, freetime...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Les élèves qui étudient plus obtiennent-ils de meilleures notes ?

Analyse de la relation entre la variable studytime et les notes (G1, G2, G3). Une régression linéaire pourrait montrer s'il existe une corrélation positive.

```r
data <- df %>% select(sex ,age ,Medu, Fedu, Mjob, Fjob,studytime, famsup,nursery, higher, int

internet <- data %>% count(internet)

data %>%
  group_by(Mjob) %>%
  summarise(nb=n())
```

```
# A tibble: 5 x 2
  Mjob        nb
  <chr>    <int>
1 at_home     59
2 health      34
3 other      141
4 services   103
5 teacher     58
```

```r
# validé l'année
pass_g3 <- filter(data, G3 >10)
```

```r
# ajourné
fall_g3 <- filter(data, G3 <10)

influ_pjob <- pass_g3 %>%
  group_by(Mjob,Fjob) %>%
  summarise(nb=n())
```

```
`summarise()` has grouped output by 'Mjob'. You can override using the
`.groups` argument.
```

```r
influ_npjob <- fall_g3 %>%
  group_by(Mjob,Fjob) %>%
  summarise(nb=n())
```

```
`summarise()` has grouped output by 'Mjob'. You can override using the
`.groups` argument.
```

```r
# Influence de l'activité de la mère sur l'elève
lprop(table(data$Mjob, data$G3))
```

```
             0     4     5     6     7     8     9     10    11    12    13
  at_home   15.3   0.0   1.7   6.8   0.0   6.8   6.8  28.8   3.4   5.1  10.2
  health     5.9   0.0   0.0   0.0   0.0   8.8   5.9   8.8   8.8   5.9  11.8
  other      9.9   0.7   2.8   5.0   3.5   8.5   7.1  11.3  14.2   9.9   7.8
  services   8.7   0.0   1.9   2.9   1.9   8.7   3.9  10.7  15.5   9.7   5.8
  teacher    6.9   0.0   0.0   1.7   3.4   6.9  13.8  15.5  10.3   3.4   6.9
  Ensemble   9.6   0.3   1.8   3.8   2.3   8.1   7.1  14.2  11.9   7.8   7.8

             14    15    16    17    18    19    20    Total
  at_home     5.1   6.8   1.7   0.0   0.0   1.7   0.0 100.0
  health     11.8  17.6   8.8   0.0   2.9   0.0   2.9 100.0
  other       7.8   6.4   2.8   0.7   0.7   0.7   0.0 100.0
  services    4.9   8.7   2.9   4.9   7.8   1.0   0.0 100.0
  teacher     6.9   8.6   8.6   0.0   3.4   3.4   0.0 100.0
  Ensemble    6.8   8.4   4.1   1.5   3.0   1.3   0.3 100.0
```

```r
cprop(table(data$Mjob, data$G3))
```

```
          0     4     5     6     7     8     9     10    11    12    13
  at_home   23.7   0.0  14.3  26.7   0.0  12.5  14.3  30.4   4.3   9.7  19.4
  health     5.3   0.0   0.0   0.0   0.0   9.4   7.1   5.4   6.4   6.5  12.9
  other     36.8 100.0  57.1  46.7  55.6  37.5  35.7  28.6  42.6  45.2  35.5
  services  23.7   0.0  28.6  20.0  22.2  28.1  14.3  19.6  34.0  32.3  19.4
  teacher   10.5   0.0   0.0   6.7  22.2  12.5  28.6  16.1  12.8   6.5  12.9
  Total    100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0

          14    15    16    17    18    19    20    Ensemble
  at_home   11.1  12.1   6.2   0.0   0.0  20.0   0.0  14.9
  health    14.8  18.2  18.8   0.0   8.3   0.0 100.0   8.6
  other     40.7  27.3  25.0  16.7   8.3  20.0   0.0  35.7
  services  18.5  27.3  18.8  83.3  66.7  20.0   0.0  26.1
  teacher   14.8  15.2  31.2   0.0  16.7  40.0   0.0  14.7
  Total    100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0
```

```
# Influence de l'activité du père sur l'elève
lprop(table(data$Fjob, data$G3))
```

```
          0     4     5     6     7     8     9     10    11    12    13
  at_home   15.0   0.0   0.0   5.0   0.0   5.0  15.0   5.0  10.0   5.0  15.0
  health     0.0   0.0   0.0   0.0   5.6  16.7  11.1  11.1  11.1   0.0  11.1
  other      9.7   0.5   1.8   5.1   2.8   7.8   4.1  17.5  12.0   7.8   8.8
  services   9.9   0.0   2.7   0.9   0.9   9.9  11.7  11.7  12.6   9.9   6.3
  teacher   10.3   0.0   0.0   6.9   3.4   0.0   3.4   6.9  10.3   6.9   0.0
  Ensemble   9.6   0.3   1.8   3.8   2.3   8.1   7.1  14.2  11.9   7.8   7.8

          14    15    16    17    18    19    20    Total
  at_home    5.0  10.0   5.0   0.0   0.0   5.0   0.0 100.0
  health    11.1  11.1   5.6   0.0   5.6   0.0   0.0 100.0
  other      6.9   7.4   2.8   0.5   3.2   1.4   0.0 100.0
  services   5.4   9.0   4.5   2.7   0.9   0.0   0.9 100.0
  teacher   10.3  10.3  10.3   6.9  10.3   3.4   0.0 100.0
  Ensemble   6.8   8.4   4.1   1.5   3.0   1.3   0.3 100.0
```

```
cprop(table(data$Fjob, data$G3))
```

```
          0     4     5     6     7     8     9     10    11    12    13
```

```
  at_home    7.9   0.0   0.0   6.7   0.0   3.1  10.7   1.8   4.3   3.2   9.7
  health     0.0   0.0   0.0   0.0  11.1   9.4   7.1   3.6   4.3   0.0   6.5
  other     55.3 100.0  57.1  73.3  66.7  53.1  32.1  67.9  55.3  54.8  61.3
  services  28.9   0.0  42.9   6.7  11.1  34.4  46.4  23.2  29.8  35.5  22.6
  teacher    7.9   0.0   0.0  13.3  11.1   0.0   3.6   3.6   6.4   6.5   0.0
  Total    100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0

            14    15    16    17    18    19    20    Ensemble
  at_home    3.7   6.1   6.2   0.0   0.0  20.0   0.0   5.1
  health     7.4   6.1   6.2   0.0   8.3   0.0   0.0   4.6
  other     55.6  48.5  37.5  16.7  58.3  60.0   0.0  54.9
  services  22.2  30.3  31.2  50.0   8.3   0.0 100.0  28.1
  teacher   11.1   9.1  18.8  33.3  25.0  20.0   0.0   7.3
  Total    100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0
```

```
cor(data$Medu,data$G3)
```

```
[1] 0.2171475
```

```
cor(data$Fedu,data$G3)
```

```
[1] 0.1524569
```

```
cor(data$G3,data$studytime)
```

```
[1] 0.09781969
```

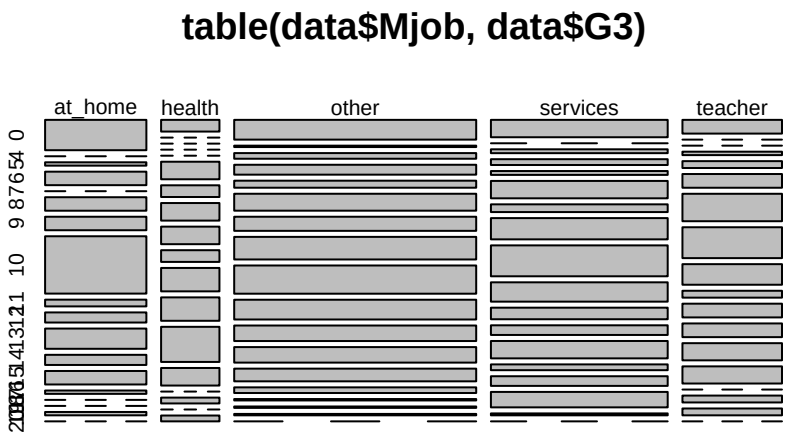
```
chisq.residuals(table(data$Mjob, data$G3))
```

```
Warning in stats::chisq.test(tab): L'approximation du Chi-2 est peut-être
incorrecte
```

```
               0     4     5     6     7     8     9    10    11    12    13
  at_home   1.40 -0.39 -0.04  1.18 -1.16 -0.36 -0.09  2.99 -1.89 -0.76  0.64
  health   -0.70 -0.29 -0.78 -1.14 -0.88  0.15 -0.26 -0.83 -0.52 -0.41  0.82
  other     0.12  1.08  0.95  0.71  1.00  0.17  0.00 -0.89  0.79  0.88 -0.02
  services -0.29 -0.51  0.13 -0.46 -0.23  0.23 -1.22 -0.94  1.07  0.67 -0.73
  teacher  -0.67 -0.38 -1.01 -0.81  0.59 -0.32  1.92  0.27 -0.34 -1.20 -0.26
```

```
             14    15    16    17    18    19    20
  at_home  -0.51 -0.42 -0.90 -0.95 -1.34  0.29 -0.39
  health    1.10  1.87  1.38 -0.72 -0.03 -0.66  3.12
  other     0.44 -0.81 -0.72 -0.78 -1.59 -0.59 -0.60
  services -0.77  0.13 -0.57  2.75  2.75 -0.27 -0.51
  teacher   0.02  0.07  1.73 -0.94  0.18  1.48 -0.38
```

```
mosaicplot(table(data$Mjob, data$G3))
```




```
lm(data$G3 ~ data$Medu)
```

```
Call:
lm(formula = data$G3 ~ data$Medu)

Coefficients:
(Intercept)    data$Medu
     7.9167       0.9088
```