

Formation analyse

Ines Amzert

2024-01-01

Assemblage des différents documents en fonction de l'axe formation

```
library(dplyr)
```

```
##  
## Attachement du package : 'dplyr'  
  
## Les objets suivants sont masqués depuis 'package:stats':  
##  
##      filter, lag  
  
## Les objets suivants sont masqués depuis 'package:base':  
##  
##      intersect, setdiff, setequal, union
```

```
library(purrr)
```

```
chemins <- c(  
  "../Nettoyage_2018/enquete_2018DS-2_nettoyer.csv",  
  "../Nettoyage_2019/2019.csv",  
  "../Nettoyage_2020/enquete_2020DS_nettoyer.csv",  
  "../Nettoyage_2021/enquete_2021DS-2_nettoyer.csv",  
  "../Nettoyage_2022/enquete_2022DS-2_nettoyer.csv",  
  "../Nettoyage2023/enquete_2023DS-2_NR2.csv"  
)  
  
dataframes <- map(chemins, ~ read.csv(.x, sep = ",") %>%  
  mutate(identifiant = as.character(identifiant)) %>%  
  select(identifiant, type_formation, filiere, date_diplome, emploi_formation))  
  
merged_data <- bind_rows(dataframes, .id = "SourceFile")  
  
write.csv(merged_data, "axeformation.csv", row.names = FALSE)  
  
axeformation <- read.csv("axeformation.csv", header = TRUE, stringsAsFactors = FALSE)
```

Evaluation des effets des modalités des variables qualitatives sur “remuneration_prime”

Test d'égalité des moyennes (ANOVA)

ANOVA pour la variable : type_formation

```
anova_type_formation <- aov(remuneration_prime ~ type_formation, data = axeformation)
summary(anova_type_formation)
```

```
##              Df      Sum Sq  Mean Sq F value    Pr(>F)
## type_formation    4 4.829e+09 1.207e+09   12.62 3.44e-10 ***
## Residuals       3038 2.905e+11 9.562e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1378 observations effacées parce que manquantes
```

La valeur de p (p-value) est très petite (3.44e-10), ce qui indique qu'il existe des **différences significatives entre les groupes**. Cela suggère que les différents types de formation (modalités de "type_formation") ont un effet significatif sur la "remuneration_prime".

ANOVA pour la variable : filiere

```
anova_filiere <- aov(remuneration_prime ~ filiere, data = axeformation)
summary(anova_filiere)
```

```
##              Df      Sum Sq  Mean Sq F value    Pr(>F)
## filiere         17 2.319e+10 1.364e+09   15.17 <2e-16 ***
## Residuals      3025 2.721e+11 8.996e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1378 observations effacées parce que manquantes
```

Le test montre une **différence statistiquement significative entre les groupes** (p-value < 2e-16), indiquant que les différentes filières ont un effet significatif sur la "remuneration_prime".

ANOVA pour la variable : date_diplome

```
anova_date_diplome <- aov(remuneration_prime ~ date_diplome, data = axeformation)
summary(anova_date_diplome)
```

```
##              Df      Sum Sq  Mean Sq F value    Pr(>F)
## date_diplome     1 9.420e+08 9.42e+08   9.731 0.00183 **
## Residuals       3041 2.944e+11 9.68e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1378 observations effacées parce que manquantes
```

Le test ANOVA indique une **différence significative entre les groupes** (p-value = 0.00183) entre les dates de diplomation en termes de "remuneration_prime".

ANOVA pour la variable : emploi_formation

```
anova_emploi_formation <- aov(remuneration_prime ~ emploi_formation, data = axeformation)
summary(anova_emploi_formation)
```

```
##              Df      Sum Sq   Mean Sq F value Pr(>F)
## emploi_formation    2 5.364e+08 268186807    2.766 0.0631 .
## Residuals          3040 2.948e+11  96967510
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1378 observations effacées parce que manquantes
```

L'ANOVA montre une tendance à la **différence négligeable** (p-value = 0.0631) entre les différentes catégories d'emploi pendant la formation par rapport à "remuneration_prime".

ANOVA pour la variable : satisfaction_formation

```
anova_satisfaction_formation <- aov(remuneration_prime ~ satisfaction_formation, data = axeformation)
summary(anova_satisfaction_formation)
```

```
##              Df      Sum Sq   Mean Sq F value Pr(>F)
## satisfaction_formation    1 9.051e+07 90506535    0.939 0.333
## Residuals                2991 2.884e+11 96419638
## 1428 observations effacées parce que manquantes
```

Le test ne montre **pas de différence significative** entre les niveaux de satisfaction de formation en termes de "remuneration_prime" (p-value = 0.333).

ANOVA pour la variable : alternance

```
anova_alternance <- aov(remuneration_prime ~ alternance, data = axeformation)
summary(anova_alternance)
```

```
##              Df      Sum Sq   Mean Sq F value Pr(>F)
## alternance      2 4.634e+07 23168398    0.239 0.788
## Residuals      3040 2.953e+11 97128706
## 1378 observations effacées parce que manquantes
```

L'ANOVA ne montre **pas de différence significative** entre les personnes ayant suivies une formation en alternance ou non en termes de "remuneration_prime" (p-value = 0.788).

ANOVA pour la variable : recommandation_polytech

```
anova_recommandation <- aov(remuneration_prime ~ recommandation_polytech, data = axeformation)
summary(anova_recommandation)
```

```
##              Df      Sum Sq   Mean Sq F value Pr(>F)
## recommandation_polytech    2 1.086e+08 54281145    0.559 0.572
## Residuals                3040 2.952e+11 97108237
## 1378 observations effacées parce que manquantes
```

Le test ne montre **pas de différence significative** entre le fait de recommander l'école polytechnique en termes de "remuneration_prime" (p-value = 0.572).

ANOVA pour la variable : ig_avantage_double_competence

```
anova_double_competence <- aov(remuneration_prime ~ ig_avantage_double_competence, data = axeformation)
summary(anova_double_competence)
```

```
##                                Df      Sum Sq   Mean Sq F value Pr(>F)
## ig_avantage_double_competence    2 7.700e+09 3.850e+09   40.69 <2e-16 ***
## Residuals                      3040 2.876e+11 9.461e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1378 observations effacées parce que manquantes
```

Le test indique des **différences significatives** entre les niveaux d'avantage de double compétence en termes de "remuneration_prime" (p-value < 2e-16).