

Fortune 1000 firms

Christophe Benavent

2024-09-25

Ce papier introduit à l'utilisation de `r` dans l'environnement `quarto` pour les analyses statistiques fondamentales de la recherche en gestion. Il sert aussi de modèle (template pour les étudiants) ...

1 Introduction

1.1 Les packages

Notre boîte à outils est `tidyverse`, de Wickham et al. (2019).

```
library(tidyverse) #outil de base(dplyr+ggplot+...)  
library(rstatix) # pour tout les test en tidy...  
library(FactoMineR) #ACP AFC et clustering  
library(factoextra)
```

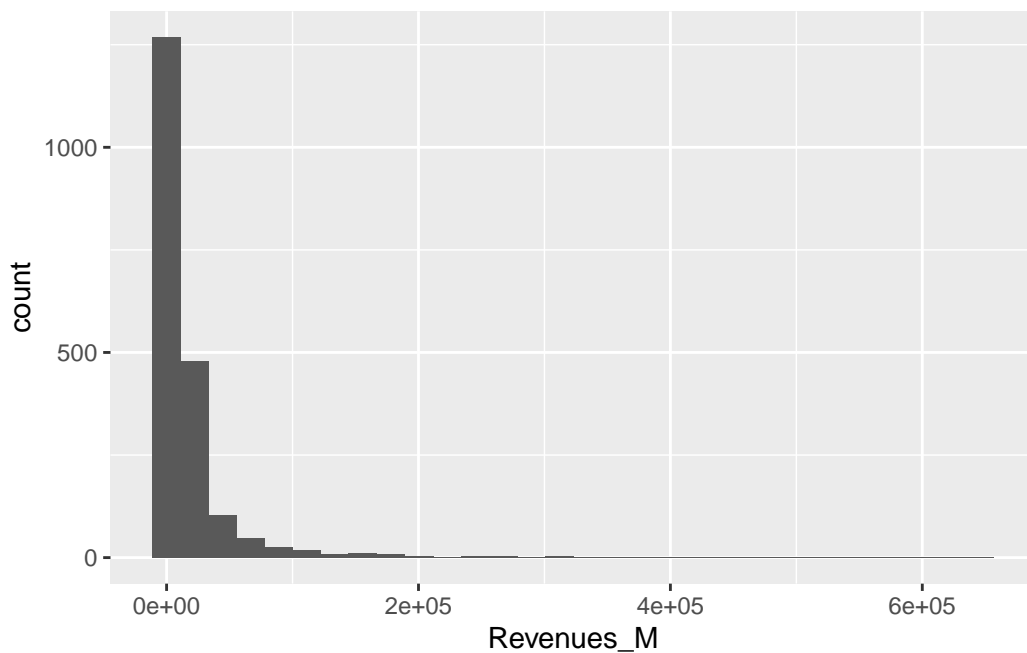
1.2 Lecture du fichier

[source des données](#) et celle de [2024](#)

2 Analyse univariée

Un simple histogramme pour la variable de revenu.

```
ggplot(df, aes(x=Revenues_M ))+  
  geom_histogram()
```



#corrélation de deux variables quanti

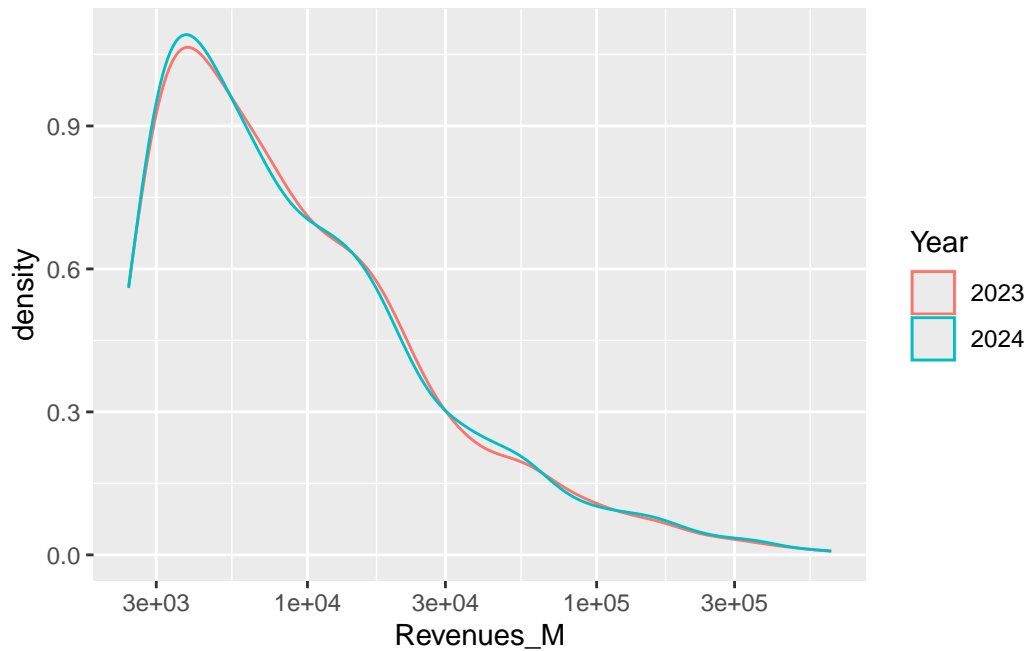
```
ggplot(df, aes(x=log10(Revenues_M), y= log10(MarketCap_March), group= Year))+  
  geom_point(alpha=.5, size=1, aes(color=Year))+  
  geom_smooth()
```



3 Comparaison plusieurs groupes sur une variable quanti

Comparer 2023 et 2024.

```
ggplot(df, aes(x=Revenues_M, group=Year ))+  
  geom_density(aes(color=Year))+  
  scale_x_log10()
```



test en t (student)

```
stat.test <- df %>%
  t_test(Revenues_M ~ Year) %>%
  add_significance()
stat.test
```

```
# A tibble: 1 x 9
  .y.      group1 group2    n1    n2 statistic    df      p p.signif
<chr>    <chr>  <chr>  <int> <int>    <dbl> <dbl> <dbl> <chr>
1 Revenues_M 2023   2024   1000  1000   -0.327 1992. 0.744 ns
```

4 Analyser deux variables quali

4.1 Tableau croisé

```
CT<-table(df1$HeadquartersState, df1$Sector)

# Calculate the chi-square statistic
chi_square_test <- chisq.test(CT)
```

```
# Print the results
print(chi_square_test)
```

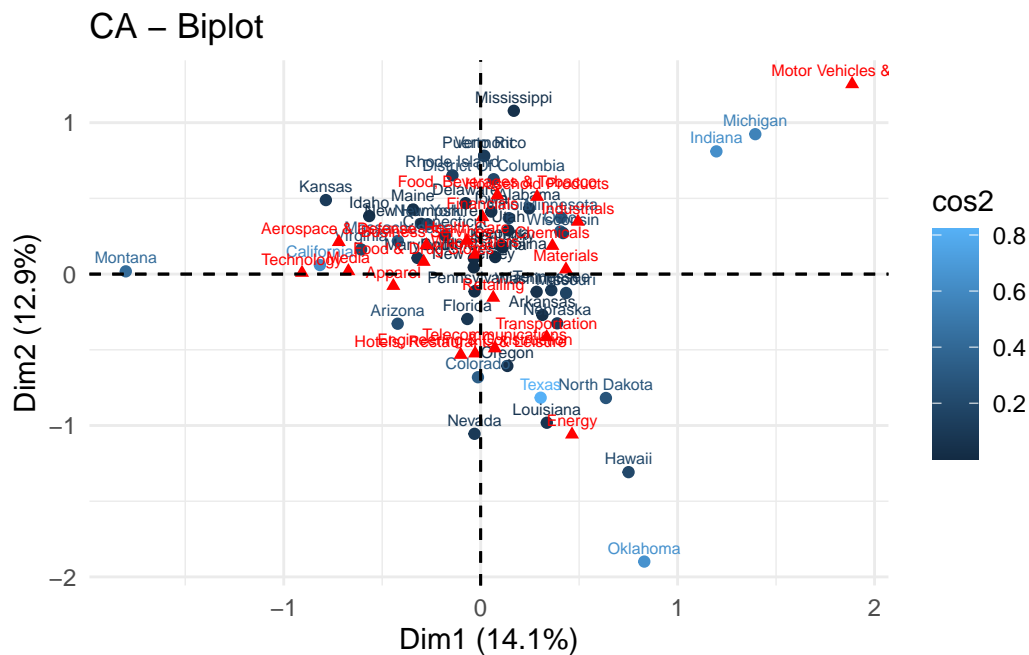
Pearson's Chi-squared test

```
data: CT
X-squared = 1800.7, df = 920, p-value < 2.2e-16
```

4.2 Analyse des correspondances

```
ca=CA(CT, graph=FALSE)

fviz_ca_biplot(ca, col.row="cos2", labels = 2) +
  theme_minimal()
```



```
ggsave("image/ca.jpg", width = 27, height = 18, units = "cm")
```

5 Analyse mutivariée descriptive

5.1 ACP

5.2 Clustering

grouper les Etats en famille similaire par leur profil d'entreprise

6 Eléments d'économétrie

6.1 régression multiple

le modèle s'écrit¹ de la manière suivante :

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{i,j} + \epsilon_i$$

6.2 modèles à décomposition d'erreurs

C'est le cas des panels par exemple: chaque individu observé sur la variable j l'est dans le pays g et au moment t . On parle aussi de modèles hiérarchique.

$$y_{i,g} = \beta_0 + \sum_{j=1}^k \beta_j x_{i,j,g} + \rho_t + \mu_g + \epsilon_i$$

6.3 autres modèles

- sur variable binaire : modèle logit
- sur variable de dénombrement : modèle de poisson
- sur variable de durée : modèle de Cox

¹Pour écrire des équations en R markdown [voir](#)

Références

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.