

Rapport d'étude

Papa Amadou NIANG

2024-03-08

Présentation de la base de données

Dans une étude fictive, 1000 profils ont été simulés, reflétant une diversité de caractéristiques démographiques et socio-professionnelles. Les variables incluent le sexe, la taille, le poids, et l'âge, avec des distributions normales et uniformes. Les données s'étendent sur des aspects tels que le salaire, la profession, la ville de résidence, le niveau d'éducation, le statut familial, les habitudes de vie, le bien-être, le groupe sanguin, la vaccination contre le Covid-19, la distance domicile-travail, et le type de logement. Cette base de données imaginaire offre un terrain fertile pour des analyses statistiques multidimensionnelles.

```
# Charger les packages
library(dplyr)
```

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

```
##
## 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(random)
```

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

```
# Définir le nombre d'observations
n <- 1000
```

```
# Générer des données aléatoires
data <- data.frame(
  Sexe = sample(c("Homme", "Femme"), n, replace = TRUE),
  Taille = rnorm(n, mean=170, sd=10), # Taille en cm
  Poids = rnorm(n, mean=70, sd=15), # Poids en kg
  Age = sample(18:100, n, replace = TRUE), # Age en années
  Salaire = runif(n, min=20000, max=100000), # Salaire annuel en euros
```

```

Profession = sample(c("Enseignant", "Ingénieur", "Médecin", "Avocat", "Artiste"), n, replace = TRUE),
Ville = sample(c("Dakar", "Thiès", "Kaolack", "Ziguinchor", "Touba"), n, replace = TRUE),
Education = sample(c("Secondaire", "Bac", "Licence", "Master", "Doctorat"), n, replace = TRUE),
Enfant = sample(0:5, n, replace = TRUE), # Nombre d'enfants
Fumeur = sample(c("Oui", "Non"), n, replace = TRUE),
Sportif = sample(c("Oui", "Non"), n, replace = TRUE),
ScoreSatisfaction = runif(n, 0, 100), # Score de satisfaction
GroupeSanguin = sample(c("A", "B", "AB", "O"), n, replace = TRUE), # Groupe sanguin
VaccinéCovid = sample(c(TRUE, FALSE), n, replace = TRUE), # Vacciné contre le Covid-19
DistanceTravail = rnorm(n, mean=10, sd=5), # Distance du travail en km
TypeLogement = sample(c("Maison", "Appartement", "Studio"), n, replace = TRUE) # Type de logement
)
summary(data)

```

```

##      Sexe      Taille      Poids      Age
## Length:1000    Min.    :136.7    Min.    : 19.77    Min.    : 18.00
## Class :character 1st Qu.:163.1    1st Qu.: 59.52    1st Qu.: 37.00
## Mode  :character Median :169.6    Median : 70.21    Median : 58.00
##              Mean  :169.7    Mean  : 70.31    Mean  : 58.63
##              3rd Qu.:176.8    3rd Qu.: 81.23    3rd Qu.: 80.00
##              Max.   :198.1    Max.   :114.01    Max.   :100.00
##      Salaire      Profession      Ville      Education
## Min.    :20306    Length:1000    Length:1000    Length:1000
## 1st Qu.:40411    Class :character  Class :character  Class :character
## Median :59304    Mode  :character  Mode  :character  Mode  :character
## Mean    :60124
## 3rd Qu.:80332
## Max.    :99971
##      Enfant      Fumeur      Sportif      ScoreSatisfaction
## Min.    :0.000    Length:1000    Length:1000    Min.    : 0.03526
## 1st Qu.:1.000    Class :character  Class :character  1st Qu.:24.09326
## Median :2.000    Mode  :character  Mode  :character  Median :50.49168
## Mean    :2.405
## 3rd Qu.:4.000
## Max.    :5.000
##      GroupeSanguin      VaccinéCovid      DistanceTravail      TypeLogement
## Length:1000            Mode :logical    Min.    :-6.696    Length:1000
## Class :character      FALSE:491    1st Qu.: 6.850    Class :character
## Mode  :character      TRUE :509    Median : 9.892    Mode  :character
##              Mean    :10.037
##              3rd Qu.:13.196
##              Max.    :27.513

```

Statistiques descriptives

Nous commençons par examiner les variables quantitatives de notre ensemble de données. La taille moyenne des individus est de `r mean(dataTaille)cm`, avec un écart – `typedersd(dataTaille)`. Cela indique une distribution relativement homogène autour de la moyenne. En ce qui concerne le poids, la moyenne est de `r mean(dataPoids)kg`, accompagnée d'un écart – `typedersd(dataPoids)`, reflétant une variabilité similaire à celle de la taille.

L'âge moyen des participants est de `r mean(dataAge)ans`, ce qui suggère une population d'adultes avec une présence notable de personnes âgées. Le salaire annuel moyen s'élève à `r mean(dataSalaire)euros`, avec une dispersion mesurée par un écart – `typedersd(dataAge)` ans.

typedersd(dataSalaire) euros, indiquant une disparité économique au sein de la population.

Variables quantitatives