



Etude observationnelle en épidémiologie clinique



Contexte

Dans ce projet, j'ai traité une base de données réelle. Le jeu de données analysé provient d'une étude observationnelle multicentrique qui a été réalisé aux États-Unis dans cinq centres hospitaliers. Elle portait sur 5735 patients adultes (âgés de 18 ans ou plus) admis en unité de soins intensifs (USI) entre juin 1989 et janvier 1994. Ces patients ont été suivis pendant une période de six mois.

Mon objectif était de lier l'impact qu'ont les scores cliniques sur les décès. Afin de répondre à la problématique posée qui était « Quel est l'impact des variables sélectionnées sur le décès des patients ? ».

J'ai réalisé ce projet avec trois autres personnes pendant une période de 4 mois dans le cadre de mon école

La démarche suivie

Nettoyage de la base des données

J'ai tout d'abord nettoyé la base de données et l'ai rendu plus compréhensible en renommant par exemple les variables, en supprimant les valeurs aberrantes ou encore en supprimant des colonnes inexploitable, car elles contenaient trop de données manquantes.

Création d'une base de données

Pour créer une base de données propre j'ai crée de nouvelles variables me permettant d'effectuer des analyses statistiques. J'ai aussi sélectionné 18 variables au lieu des 67 de la base de données brute.

Les tests

Pour pouvoir effectuer des analyses statistiques il a fallu faire des tests statistiques. J'ai choisi 3 tests qui s'adaptait le mieux en fonction du type de données qu'il fallait analyser (Test du CHI 2, Corrélation et Student).

Représentation

Pour finir, j'ai réalisé des graphiques statistiques pour représenter visuellement des données statistiques de manière à rendre l'information plus compréhensible et accessible. Mon groupe et moi même avons par la suite, présenté ce projet lors d'une soutenance oral.



Conclusion



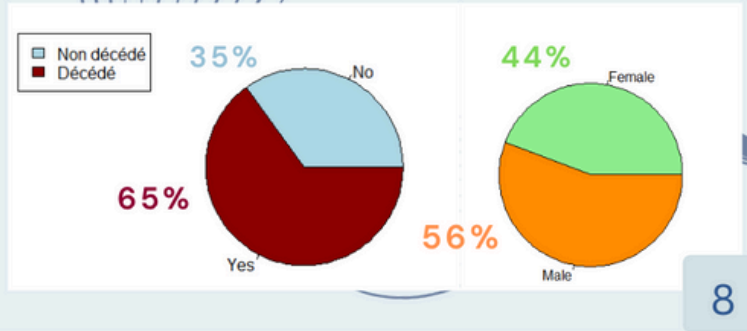
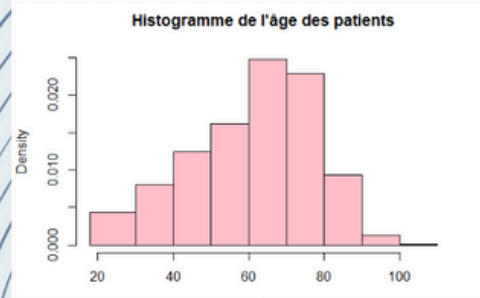
En résumé, cette analyse approfondie a permis de mettre en évidence certaines associations et différences importantes entre les variables étudiées. Ces résultats contribuent à une meilleure compréhension de la recherche et du logiciel utilisé (R Studio) et m'a aidé à prendre des décisions basées sur des connaissances statistiques. J'ai pu acquérir des compétences d'analyse et de modélisation, une meilleure cohésion de groupe ainsi qu'une meilleure gestion du temps de travail.

Exemples

Page de la présentation Orale:

ANALYSE UNIVARIÉE

- La plupart des variables comme l'âge sont des variables quantitatives continues
 - Les variables "DEATH" et "SEX" sont qualitatives nominale
 - Définition des variables
 - Analyse descriptives de ces variables
1. Minimum
 2. 1er quartile
 3. Médiane
 4. Moyenne
 5. Maximum
 6. Nombre de NA



8

Extrait du programme avec graphique statistique:

```
146 #Personnes pas decede et decede au bout de 2 mois
147 epi$DECES_2mois = ifelse(epi$Temps_Décès < 62, 1, 0)
148 epi$DECES_2mois = ifelse(is.na(epi$DECES_2mois), 0, epi$DECES_2mois)
149
150 mean(epi$SURVIE À 2 MOIS[epi$DECES_2mois == 1])
151 mean(epi$SURVIE À 2 MOIS[epi$DECES_2mois == 0])
152
153
154 x = c(mean(epi$SURVIE À 2 MOIS[epi$DECES_2mois == 1]),
155       mean(epi$SURVIE À 2 MOIS[epi$DECES_2mois == 0]))
156 barplot(table(epi$SURVIE À 2 MOIS, epi$DECES_2mois))
157 table(x[epi$DECES_2mois])
158 barplot(x, main = "Graphique de la moyenne de la probabilité de survie à 2 mois
159          en fonction de si le patient est décédé ou non",
160         , ylab = "Probabilité moyenne de survie à 2 mois"
161         , xlab = "Etat du patient au bout de 2 mois"
162         , col = c("orange", "red"))
163
164 legend("topleft", legend = c("Non décédé", "Décédé"), fill = c("red", "orange"))
165 table(epi$DECES_2mois)
166
167
168.1 (Untitled) R Script
```

```
> table(x[epi$DECES_2mois])
0.496593106535363
2234

> barplot(x, main = "Graphique de la moyenne de la probabilité de survie à 2 mois
+          en fonction de si le patient est décédé ou non",
+         , ylab = "Probabilité moyenne de survie à 2 mois"
+         , xlab = "Etat du patient au bout de 2 mois"
+         , col = c("orange", "red"))
+ legend("topleft", legend = c("Non décédé", "Décédé"), fill = c("red", "orange"))
+ table(epi$DECES_2mois)
```

```
0 1
3274 2234
>
```

