

Echantillonnage sur des données de santé

Objectifs :

Le projet plan d'échantillonnage, qui a eu lieu en février 2024, fut réalisée en groupe de 3 personnes. Nous devions, en 1 mois, pouvoir rendre un rapport et faire une présentation orale de nos résultats. Ce projet consistait à appliquer, sur R, deux méthodes de randomisation sur des données de santé. La randomisation est le moyen d'attribuer à un patient, un traitement A ou un traitement B lors d'essai clinique.

Démarche Méthodologique

Dans un premier temps, nous avons constitué deux bases de données. La première était composée de 1000 études de 20 patients, et la deuxième de 1000 études de 50 patients. Après cela, nous avons dû assimiler et comprendre les différentes méthodes de randomisation afin de les appliquer à nos données. Une fois que la randomisation fut réalisée, nous avons généré une variable « Succès ». Cette variable permettait de déterminer, selon une certaine probabilité, si le traitement fut efficace.

Extrait de la base de données réalisée :

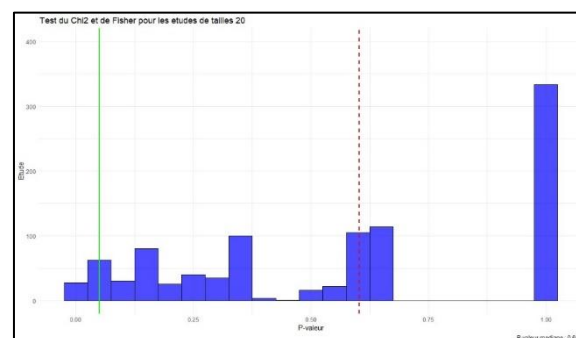
etude	idpatient	pays	age	Traitement	succes
9	1	9	1	74 A	Echec
10	1	10	2	61 B	Succès
11	1	11	2	85 A	Echec
12	1	12	1	66 B	Succès
13	1	13	1	69 A	Echec
14	1	14	3	69 A	Echec
15	1	15	3	81 A	Echec
16	1	16	2	51 B	Echec
17	1	17	3	77 A	Echec
18	1	18	1	70 B	Echec
19	1	19	3	68 A	Echec
20	1	20	1	81 B	Echec
21	2	1	2	48 B	Succès
22	2	2	2	71 A	Echec
23	2	3	2	75 A	Succès

Showing 9 to 23 of 20,000 entries. 6 total columns

Résultats

Afin de comparer nos deux méthodes de randomisation, nous avons effectué des tests et calculer les p-valeur pour vérifier leurs efficacités. Les résultats n'étaient pas ceux attendus, cela est dû au nombre de patients dans chaque étude. Pour des résultats plus concluant, il aurait fallu générer des études avec beaucoup plus que seulement 50 personnes.

P-valeur obtenues pour les études de taille 20 :



Compétences développées

La véritable difficulté de ce projet était de s'organiser et de communiquer afin d'avancer de la manière la plus efficace, car nous avions moins d'un mois pour le réaliser en autonomie et le présenter à l'oral. Nous devions donc nous répartir les tâches équitablement et nous tenir régulièrement au courant des avancées de chacun. En plus de pouvoir apprendre le principe d'échantillonnage, ce projet nous a permis de développer nos compétences en programmation R, mais aussi nos compétences de rédaction et d'expression orale dans le but de restituer nos résultats statistiques.