

Résumé des exposé de traitement statistiques sur R

I-INTRODUCTION

Ce document est un résumé de ce que j'ai compris sur les différents exposé qui on passé qui sont:

- Visualisation des données
- Traitement des valeur manquante et aberrantes
- ANOVA et test non paramétriques

II-RESUME

II- 1. valeur manquantes

Les valeurs aberrantes et manquantes nuisent au différent test et analyse statistiques que nous auront à faire. Détecter ces valeurs implique des méthodes spécifiques, tandis que leur traitement varie selon le contexte et les objectifs de l'étude pour préserver la qualité et la précision des résultats.

1-1.Traitement des valeurs aberrantes et manquantes

Il existe trois types de valeurs manquantes existent : • MCAR (Donnée Manquante Complètement Aléatoire)

- MAR (Donnée Manquante Aléatoire Conditionnelle)
- MNAR (Donnée Manquante Non Aléatoire At Random)

Pour traiter ces valeurs manquantes, les méthodes suivantes sont appliquées : • Méthode de suppression

- Méthode d'imputation par la moyenne
- Méthode d'imputation par la médiane
- Méthode d'imputation par régression
- Méthode d'imputation par HOT DECK
- Méthode d'imputation par KNN
- Méthode d'imputation par LOCF

3.Méthodes de traitement

D'abord pour detecter les valeurs aberrantes il existe 3 groupe de méthodes : • Détection suivant le domaine des valeurs

- Détection graphique
- Autres méthodes statistiques de détection

Pour traiter ces valeurs aberrantes, les méthodes suivantes sont appliquées :

- Méthode de suppression des valeurs aberrantes

- Méthode de transformation des valeurs aberrantes :
- les méthodes d'imputation
- Transformation logarithmique
- Méthode de winsorisation

4. quelque limites

- Biais potentiel
- Perte d'information
- Difficulté d'interprétation
- Dépendance du choix de la méthode
- Complexité accrue # Statistiques descriptives et visualisation des variables catégorielles avec ggplot2()

La visualisation des données, cruciale en analyse, est simplifiée par ggplot2 de R. Ce package, partie du tidyverse, permet de créer des graphiques variés et intuitifs grâce à une grammaire graphique. Lancé en 2006 et amélioré en 2007, ggplot2 est l'œuvre de Hadley Wickham, architecte du **tidyverse**.

II-2 Les différentes parties de ggplot()

1. La composante esthétique (aes)

- Ajout d'éléments de Base (titre)
- le mappage color
- le mappage shape
- le mappage size
- le mappage alpha
- Le mappage fill

2. La composante theme

- les thèmes complets
- Les éléments de thème
- Les fonctions d'élément

II-3 ANOVA ET TESTS NON PARAMETRIQUES

L'**ANOVA** compare les moyennes de plusieurs groupes, tandis que les tests non paramétriques sont utilisés quand les données dévient de la normalité. Le langage R permet d'effectuer ces analyses avec précision et efficacité dans divers domaines.

Les différents types d'ANOVA

ANOVA à 1 facteur

ANOVA à 2 facteurs

Les tests non paramétriques

Les tests non paramétriques ne supposent pas une distribution spécifique des données, utilisant des statistiques basées sur l'ordre ou des tableaux de contingence. Ils sont utiles lorsque les hypothèses des tests paramétriques ne sont pas satisfaites.

Types de tests non paramétriques:

- Test de rang signé de Wilcoxon
- Test du khi-carré pour la qualité d'ajustement
- Test U de Mann-Whitney (Wilcoxon/Mann-Whitney)
- Test de Kruskal-Wallis