

# POINTS IMPORTANTS DES EXPOSÉS DÉJÀ PRÉSENTÉS

ONANENA AMANA Jeanne De La Flèche

2024-04-27

## TRAITEMENT DES VALEURS MANQUANTES ET DES VALEURS ABERRANTES AVEC R

Après avoir parcouru en détail cet exposé, il ressort que les valeurs manquantes ont une certaine typologie qui peut être :

- MCAR : Manquantes Complètement Aléatoirement.
- MAR : Manquantes Aléatoirement Conditionnelles.
- MNAR : Manquantes Non Aléatoires.

Après avoir détecté les valeurs manquantes, il faudrait penser à les traiter pour être sûr que les analyses qui seront faites renseignent effectivement sur la situation étudiée. Les méthodes de traitement des valeurs manquantes sont multiples, et entre autres nous avons :

Les méthodes d'Imputation : Elles sont diverses et nous avons :

- Imputation par la suppression : Cette méthode semble simple mais elle peut entraîner une perte de données.
- Imputation par la moyenne : Elle consiste à remplacer les valeurs manquantes par la moyenne.
- Imputation par la médiane : Elle est utilisée pour les données non symétriques.
- Imputation par régression : Plus sophistiquée, cette méthode utilise un modèle prédictif.

Les valeurs manquantes ayant été traitées, intéressons-nous aux valeurs aberrantes.

Le traitement des valeurs aberrantes implique :

- La suppression de ces dernières : Cette méthode semble simple mais elle peut fausser les résultats.
- La transformation des valeurs aberrantes : Cette méthode réduit l'influence sans éliminer les données.
- La winsorisation des valeurs aberrantes : Elle plafonne les valeurs extrêmes à un certain percentile.

Ces méthodes peuvent être les meilleurs moyens pour faire des analyses de nos bases de données, mais elles comportent de nombreuses limites telles que le risque de biais, la perte d'information, et la complexité accrue selon la méthode choisie.

## STATISTIQUES DESCRIPTIVES ET VISUALISATION DES VARIABLES CATÉGORIELLES AVEC ggplot2()

Les statistiques descriptives sont essentielles lorsqu'il s'agit de proposer des analyses sur plusieurs sujets donnés. Elles comprennent les paramètres de positions, de dispersion et de tendances centrales. C'est pour explorer les paramètres de position que cet exposé nous a été proposé.

Au début de cet exposé, les exposants nous ont présenté les variables dans R qui peuvent être qualitatives ou quantitatives et nous ont expliqué que ces variables sur R sont représentées en tant que vecteurs.

Après quoi, ils nous ont montré comment on peut faire des visualisations avec ggplot2 qui est un outil puissant pour créer des graphiques basés sur la grammaire graphique.

De plus, ils nous ont montré comment utiliser les composantes esthétiques de ggplot2 telles que aes, color, fill, shape, size et alpha, qui permettent de personnaliser l'apparence des graphiques.

Ils nous ont enfin présenté les différentes composantes de ggplot2 y compris les geoms pour différents types de graphiques, les facettes pour diviser les graphiques en panneaux, et les thèmes pour contrôler l'apparence des graphiques.

Nous retenons de cet exposé que ggplot2 est un outil puissant pour la visualisation des données sur R.

## **ANOVA ET TESTS NON PARAMÉTRIQUES**

Cet exposé nous a permis de faire une introduction à l'ANOVA (Analyse de la Variance). L'ANOVA a été développée par Ronald Fisher et est applicable dans divers domaines notamment pour comparer les moyennes de groupes et détecter les interactions entre facteurs de variabilité et une variable quantitative principale.

Comme toute méthode scientifique, l'ANOVA présente certaines limites :

- Elle suppose l'égalité des variances et une distribution normale des données.
- Elle est sensible aux valeurs aberrantes et nécessite des tests post-hoc pour identifier les groupes significativement différents.

Afin de surmonter ces limites, nous faisons appel aux Tests Non Paramétriques.

Ces tests offrent une alternative lorsque les données ne suivent pas une distribution normale, avec des exemples comme le test de Wilcoxon, le khi-carré et le test U de Mann-Whitney.

Que retenir de l'ANOVA et des tests non paramétriques?

L'ANOVA et les tests non paramétriques sont des outils essentiels pour faire des études statistiques qui se complètent.