Récapitulatif des exposés

Khadidiatou DIAKHATE

2024-04-27

## Récapitulatif de l’expose sur le traitement des valeurs manquantes et aberrantes par Khadidiatou Diakhaté et Ahmadou Niass

Plusieurs méthodes permettent le traitement des valeurs manquantes à savoir l’imputation par la moyenne ( cette méthode convient pour les données faibles et dont l’analyse est peu sensible à la variabilité et elle consiste à remplacer les valeurs manquantes par la moyenne de la variable); l’imputation par la médiane (utile pour les données faibles, variables non symétriques dont l’analyse est peu sensible à la précision de l’imputation , et elle consiste à remplacer les valeurs manquantes par la médiane de la variable); la suppression( à utiliser avec précaution pour les analyses robustes); la régression( qui consiste en la sélection de variables explicatives pertinentes puis à la construction d’un modèle de régression pour prédire les valeurs manquantes) et enfin méthode d’imputation HOT DECK(qui consiste à remplacer les valeurs manquantes par des valeurs similaires d’autres observations) Pour la détection des valeurs aberrantes, la méthode de Tukey (méthode de l’intervalle interquartile est celle mise en exergue), elle consiste à considérer pour valeurs aberrantes celles qui s’éloignent de 1,5 fois de l’IQR des quartiles. Pour le traitement de ces valeurs, on peut soit transformer les valeurs aberrantes trouvées en valeurs manquantes ou alors procéder par winsorisation(qui plafonne les valeurs extrêmes à un certain percentile). Toutefois ces méthodes présentent de nombreuses limites à savoir le risque de biais, la perte d’information ou encore une complexité accrue.

## Récapitulatif de l’exposé sur les statistiques descriptives et visualisation des variables catégorielles avec ggplot2 par Khadidiatou Coulibaly et Samba Dieng :

Dans cet exposé ,la nature des différentes variables sur R à savoir : vecteurs, double, float, character, les facteurs a été décrite puis l’importance de la visualisation des données et l’utilisation des données et l’utilisation de ggplot2 pour créer des graphiques informatifs et esthétiques dans le cadre de l’analyse des statistiques catégorielles a été mise en exergue. L’utilisation de ggplot2, une bibliothèque puissante de visualisation de données en R, est présentée comme une approche structurée basée sur la grammaire graphique pour créer une variété de graphiques de manière cohérente et intuitive. Les thèmes de tracé et d’axe sont explorés pour permettre un contrôle précis des éléments visuels des graphiques, tandis que différentes techniques de visualisation des données catégorielles sont illustrées à travers des exemples pratiques. Cet exposé met l’accent sur l’importance de la visualisation des données pour une communication claire et une prise de décision éclairée et permet de ce fait de renforcer nos connaissances sur les méthodes et les outils essentiels pour l’analyse et la représentation visuelle des variables catégorielles.

## Récapitulatif sur ANOVA et tests non paramétriques par Celina Nguemfouo et Raherinasolo Rayan

Cet exposé sur l’ANOVA et les tests non paramétriques met en exergue une analyse détaillée des outils statistiques essentiels pour comparer les moyennes de différents groupes. Il commence par présenter l’ANOVA à un facteur, illustrant comment cette méthode est utilisée pour examiner les relations entre les notes des candidats et des variables telles que le sexe et le correcteur. Ensuite, il explore l’ANOVA à deux facteurs pour étudier les interactions entre plusieurs variables simultanément. L’exposé détaille les étapes clés de ces analyses, notamment la formulation des hypothèses, le calcul des moyennes empiriques et la réalisation des tests statistiques pour évaluer les différences significatives entre les groupes. Enfin, il met en évidence les multiples applications pratiques de l’ANOVA et des tests non paramétriques dans des domaines variés tels que la médecine, le marketing, l’éducation, l’industrie, la psychologie, l’agriculture, la sociologie et d’autres sciences sociales. Ainsi, cet exposé permet donc d’approfondir nos connaissances sur l’utilisation de ces méthodes statistiques pour l’analyse de données et la prise de décision dans divers contextes.