

Introduction aux logiciels pour les statistiques

Cours-TP / 17 mars 2021 / IS2A3

Préparer le support de travail

1. Créez un dossier *TP_R* et un sous-dossier *Cours* .
2. Rangez les données dans le sous-dossier.
3. Décrivez les fenêtres du logiciel Rstudio.
4. Créez un script *Mon_TP* et rangez-le dans le sous-dossier.
5. Téléchargez et/ou importez les packages *dplyr* , *DT* , *ggplot2* et *forcats*.

Interroger une table de données

1. Importez les tables données du cours dans l'environnement R.
On travaille désormais avec la table *cinema*.
2. Donnez un aperçu de la table de données.
3. Affichez des indicateurs statistiques simples.
4. Donnez le nombre de lignes et de colonnes de la table.
5. Sélectionnez la variable *FAUTEUILS*.
6. Filtrez la table pour ne conserver que les lignes où *FAUTEUILS* >1000.
7. Filtrez la table pour ne conserver que les lignes où *FAUTEUILS* >1000 et *ECRAN*>2.
8. Triez la variable *FAUTEUILS* par ordre (dé)croissant.
9. Créez une variable *DOTATION* égale à *FAUTEUILS*+5.
10. Donnez le nombre moyen de *FAUTEUILS* par *COMMUNES*.
11. Fusionnez les tables *cinema* et *écrans* en une table nommée *cinema2* (2 méthodes).
12. Renommez la variable *CODE_COMMUNE* en *CODE_POSTAL*.

Propriétés du type `data.frame`

Sélectionnez la variable *FAUTEUILS* de la table *cinema* de 3 façons différentes.

Variables qualitatives

1. Filtrez la table *cinema2* en conservant uniquement les arrondissements parisiens.
2. Donnez les occurrences de la variable *COMMUNE*.
3. Donnez le nombre de cinémas pour chaque arrondissement.
4. Donnez la proportion de cinémas par arrondissement parisien.
5. Représentez graphiquement la variable *CODE_COMMUNE* pour les 5^{ème} et 6^{ème} arrondissements.

Variables quantitatives

On travaille sur la variable *FAUTEUILS* de la table *cinema2*.

1. Représentez graphiquement la variable par un histogramme.
2. Calculez les statistiques de tendance centrale suivantes : la moyenne, la médiane et le(s) mode(s).
3. Quel est le nombre moyen d'écrans par commune ?
4. Calculez les statistiques de position : les quartiles et les déciles.
5. Représentez graphiquement la variable par un diagramme de Tuckey.
6. Calculez les statistiques de dispersion : étendue , écart interquartile, variance, écart-type et coefficient de variation.
7. Représentez graphiquement la densité de la variable.
8. Représentez graphiquement la fonction de répartition de la variable et identifiez les quartiles.

Croisement de variables quantitatives

On cherche à identifier le lien entre les nombres de fauteuils et d'écrans dans les salles de cinéma.

1. Représentez graphiquement les 2 variables par un nuage de points. Commentez.
2. Calculez le coefficient de Pearson entre les 2 variables. Commentez.
3. Effectuez une régression linéaire entre les 2 variables. Commentez.