

Exercice du thème 1

Ouvrir la feuille de données : « intro.csv » disponible dans le dossier DATA. Créer un nouveau jeu de données en conservant uniquement les répondants qui ont moins de 40 ans et nommer le « intro2 ».

Travailler par la suite à partir du nouveau fichier intro2 uniquement.

1. Trier tous les individus du fichier intro2 selon le nombre de visites au cours de la dernière année en ordre décroissant et décrire l'individu qui est allé le plus souvent au cinéma.

2. Créer une variable binaire, appelée « vis_bin », en attribuant les valeurs :

1 si le nombre de visites est >5

0 sinon

À partir de l'onglet « Variable view », attribuer une étiquette appropriée à chacune des valeurs de la variable « vis_bin », soit 0 = « 5 ou moins » et 1 = « plus de 5 ».

- Obtenir le tableau de distribution de fréquences de la variable « vis_bin » et déterminer la proportion des répondants de moins de 40 ans qui sont allés plus de 5 fois au cinéma au cours de la dernière année.

3. Obtenir la moyenne, la médiane et l'écart type, les premiers et 3e quartiles, un histogramme et le diagramme boîte à moustaches du nombre de visites effectuées par les répondants de moins de 40 ans et interpréter.

4. Construire un tableau croisé pour analyser la répartition des répondants par rapport aux variables « sexe » et « vis_bin » de façon à pouvoir déterminer les proportions suivantes :

- proportion de femmes parmi les individus qui sont allés plus de 5 fois au cinéma;
- proportion d'individus qui sont allés plus de 5 fois au cinéma chez les hommes.

5. Estimer au moyen d'un intervalle de confiance de niveau 95% la proportion

d'individus qui sont allés plus de 5 fois au cinéma dans la population considérée.

6. Estimer au moyen d'un intervalle de confiance de niveau 95% le nombre moyen de visites effectuées dans la population de référence.

7. Obtenir les diagrammes en boîte du nombre de visites selon le sexe et comparer les distributions observées.