

1 Les données

Les données que nous allons analyser dans ce TP sont extraites de <https://github.com/caesar0301/awesome-public-datasets> et concerne les caractéristiques des passagers du Titanic. La description des données originales est disponible sur cette page web : <https://www.kaggle.com/c/titanic/data>.

1. Charger les données et visualisez les données. Contrôlez le nombre d'entrées, le nombre de champs et leur étendue.
2. Enlever la variable "Fare".
3. Recoder la variable "Sex" en variable binaire (male= 0, female=1).
4. Remplacer les valeurs manquantes dans la variable age par 0.
5. Transformer les variables "Sex" et "Embarked" en entier (int)

2 Tableaux de contingence

2.1 La survie selon le genre

1. Calculez le tableau de contingence sur la survie *vs* le genre des passagers.
2. Analysez ce que vous observez et proposez une hypothèse.

2.2 La survie selon la classe et le port d'embarquement

1. Calculez le tableau de contingence sur la survie *vs* la classe des passagers.
2. Calculez le tableau de contingence sur la survie *vs* le port d'embarquement.
3. Formulez des hypothèses sur vos observations.

3 Dépendance de variables

1. Implémentez le calcul du χ^2 et des mesures issues du χ^2 :
 - Φ^2 ,
 - Coefficient de Tschuprow,
 - Coefficient de Cramer.
2. Validez ou invalidez vos hypothèses faites dans la section suivante.
3. Pour chaque mesure, calculez un tableau présentant l'interdépendance pour l'ensemble des variables utilisées.
4. Que déduisez vous sur vos hypothèses précédentes ?

4 Tableaux de Burt

1. Calculez le tableau de Burt pour l'ensemble des variables et leurs modalités.

Bonus

1. Pour les plus rapides, récupérez les données originales et effectuez une analyse plus poussée sur le profil des survivants en utilisant les données quantitatives.