

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>.

Ce TP concerne l'analyse multidimensionnelle de données qualitatives. Par conséquent, les données originales ont été en parti retraitées afin d'inclure seulement des données qualitatives. Vous pouvez récupérer ces données en téléchargeant le fichier `titanic-quali.csv` sur Moodle.

Champ	Libellé	Signification
1	Survie	0 : Décédé, 1 : survivant
2	Classe	Indique la classe du passager : 1,2 ou 3 ^{ème} classe
3	Genre	0 : féminin, 1 : masculin
4	Age	0 : Inconnu, 1 : bébé, 2 : adolescent, 3 : adulte, 4 : personne agée.
5	Frères/sœurs à bord	0 : aucun, 1 : Un, 2 : plus de un
6	Parents à bord	0 : aucun, 1 : Un, 2 : plus de un
7	Port d'embarquement	67 : Cherbourg, 81 : Queenstown, 83 : Southampton

TABLE 1 – Données du fichier `titanic-quali.csv`

1. Charger les données et visualisez les données. Contrôlez le nombre d'entrées, le nombre de champs et leur étendue.
2. N'hésitez pas à ouvrir le fichier avec votre éditeur de texte préféré (GNU/Emacs donc) pour visualiser la structure d'un fichier `csv`.

Aide :

```
— >>data=csvread('titanic-quali.csv'),
```

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. Profitez de la qualité binaire des données pour calculez le tableau de contingence par une multiplication de matrices.
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.

Aide :

— `>> help histc`

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 Profils

1. Implémentez le calcul des profils-colonnes pour une paire de descripteurs de votre choix.
2. Affichez ces profils-colonnes sous forme de graphiques (voir figure 1). Mettez en relation ce que vous observez et les hypothèses que vous avez faites

5 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.

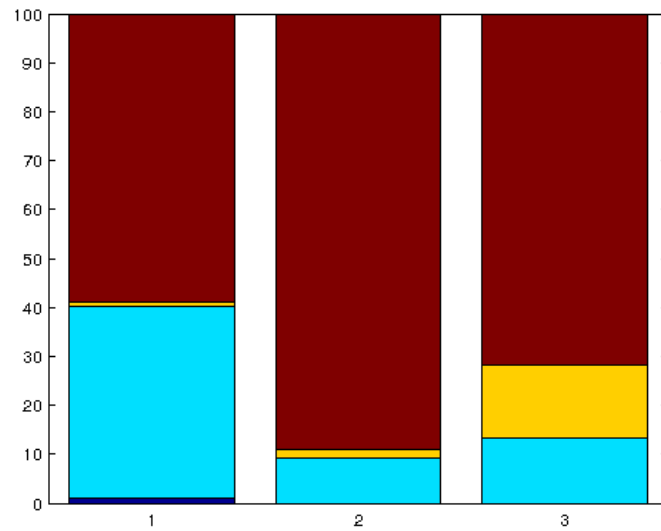


FIGURE 1 – Profils-colonnes relatifs à la table de contingence de la classe des passagers vs au port d'embarquement.