

TP10

STATISTIQUES, 2^E PARTIE

I Avec numpy

On va reprendre l'analyse d'une série statistique vu en TP7 en utilisant cette fois les outils offerts par numpy (voir documentation).

Exercice 1 En utilisant numpy, calculer la moyenne, l'écart-type, la médiane et l'étendue des notes du groupe A (données fournies).

```
groupeA = [47, 78, 50, 73, 55, 39, ...]
```

```
import numpy as np
```

II Utilisation de la bibliothèque pandas

Le fichier titanic.csv est un tableur contenant les données des passagers du Titanic. Nous allons l'ouvrir et l'étudier avec python en utilisant la bibliothèque pandas (voir documentation jointe).

```
import pandas as pd
titanic = pd.read_csv('titanic.csv')
```

```
titanic
```

| PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked | |
|-------------|----------|--------|------|---------------------------------------------------|--------|-------|-------|--------|------------------|---------|----------|-----|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S |
| 3 | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1000 | C123 | S |
| 4 | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.0500 | NaN | S |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 886 | 887 | 0 | 2 | Montvila, Rev. Juozas | male | 27.0 | 0 | 0 | 211536 | 13.0000 | NaN | S |
| 887 | 888 | 1 | 1 | Graham, Miss. Margaret Edith | female | 19.0 | 0 | 0 | 112053 | 30.0000 | B42 | S |
| 888 | 889 | 0 | 3 | Johnston, Miss. Catherine Helen "Carrie" | female | NaN | 1 | 2 | W./C. 6607 | 23.4500 | NaN | S |
| 889 | 890 | 1 | 1 | Behr, Mr. Karl Howell | male | 26.0 | 0 | 0 | 111369 | 30.0000 | C148 | C |
| 890 | 891 | 0 | 3 | Dooley, Mr. Patrick | male | 32.0 | 0 | 0 | 370376 | 7.7500 | NaN | Q |

891 rows x 12 columns

- PassengerId (identifiant passager)
- Survived (0 : décédé, 1 : a survécu)
- Pclass (classe, de 1 à 3)
- Name (Nom, prénom et titre)

- Sex (male/female)
- Age (en années)
- SibSp (nombre de frère, soeur, beau-frère, belle soeur, mari ou femme)
- Parch (nombre de parents et d'enfants)
- Ticket (numéro du billet)
- Fare (prix du billet)
- Cabin (numéro de cabine)
- Embarked (port d'embarquement : C - Cherbourg, S - Southampton, Q = Queenstown)

Exercice 2 En utilisant la documentation pandas :

1. afficher les colonnes et leur type (avec info) ;
int** -> entier, float** -> nombre décimal, object -> donnée non numérique

2. afficher (un aperçu de) la liste des passagers, uniquement leur nom (Name) ;

3. afficher les statistiques de base avec describe.

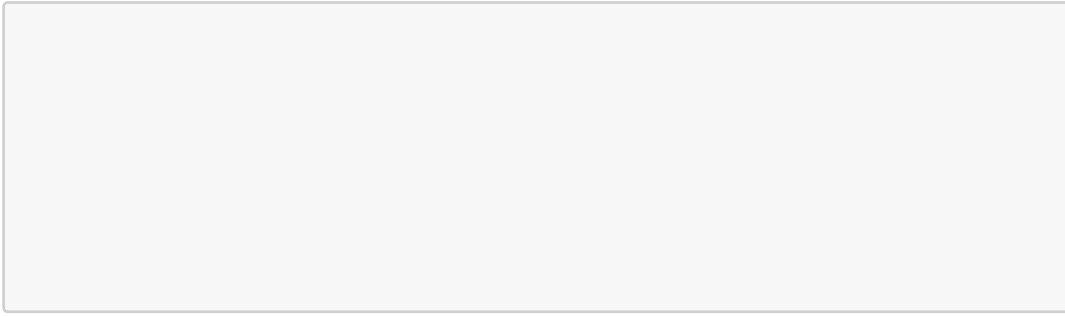
Exercice 3

1. Que fait `titanic['Name'].count()` ?
2. Calculer la proportion des passagers qui a survécu au naufrage, de deux façons différentes.

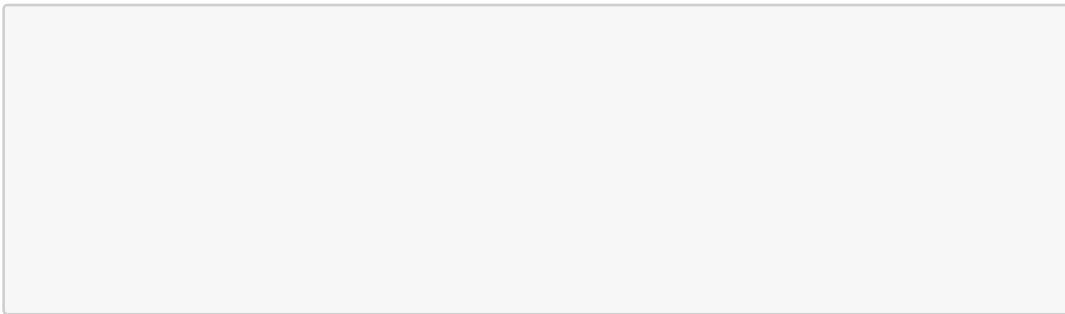
3. Calculer cette même proportion mais uniquement pour les passagers de première classe. On commence par filter les données pour n'avoir que celles correspondant aux passagers de première classe.

```
titanic1 = titanic[ _____ == 1 ]
```

4. Faire de même pour la deuxième et troisième classe.

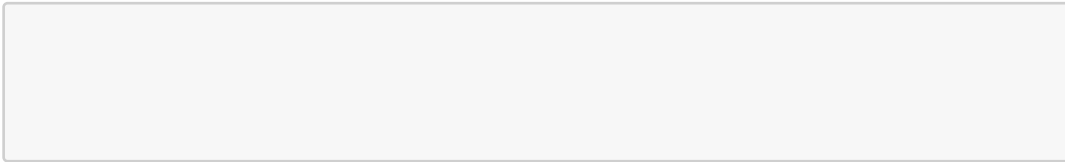


5. Comparer maintenant le taux de survie chez les hommes et les femmes.

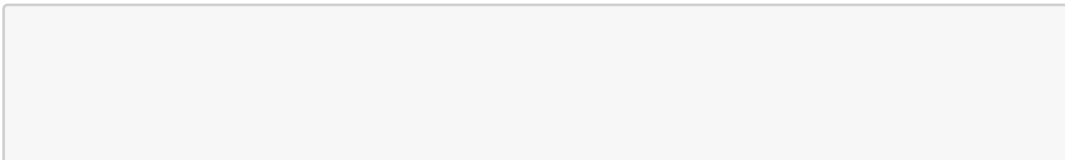


Exercice 4

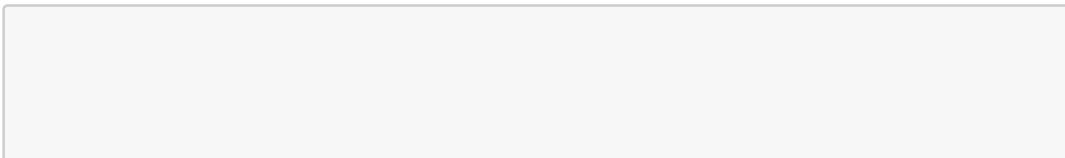
1. Tracer l'histogramme de répartition des âges dans le navire.



2. Faire de même avec le prix des billets, avec 20 intervalles (bins).

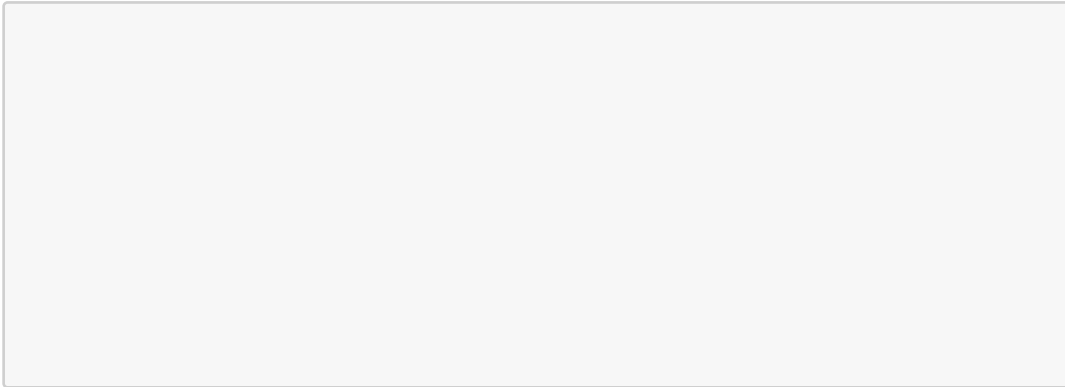


3. On remarque une valeur isolée, on va l'ignorer pour la suite. Tracer les histogrammes de prix des billets sur $[0,300]$ découpé en 20 intervalles, pour chaque classe du navire séparément.



Exercice 5

1. Déterminer le prix moyen du billet par classe de voyage ainsi que l'écart-type.



2. Mettre en avant d'autres critères discriminants (ou pas) concernant la probabilité de survie d'un passager.