

# Analyse de données- TP1

## 3.1 Analyse préalable Chargement des données

### Q1)Taper et comprendre les commandes suivantes :

`clear` % nettoie les figures et les variables présentes dans matlab en les détruisant.

`load('fisheriris.mat')` % charge le fichier de base qui contient la matrice de l'iris et ses données

%(meas : 4 colonnes sepal length, sepal width, petal length et petal width. species 1 colonne toutes les deux avec 150 lignes)

`whos` % liste les variables par ordre alphabétique, tailles et types

De plus, la commande `whos` permet d'afficher l'équivalent des informations utiles sur les variables en mémoire. Ces informations sont aussi disponibles dans la sous fenêtre 'workspace'.

Matlab affiche après chargement :

Name	Size	Bytes	Class	Attributes
------	------	-------	-------	------------

meas	150x4	4800	double	
------	-------	------	--------	--

species	150x1	19300	cell	
---------	-------	-------	------	--

### Q2)Quelle commande vous donne le nombre d'inconnues statistiques ?

```
[unitmeas, varmeas]=size(meas)
```

```
[unitspecies, varspecies]=size(species)
```

La commande `size` donnera les unités ou individus statistiques ainsi que les variables attachées à ces individus. Ici pour la matrice `meas` nous avons 150 individus et 4 variables (sepal length, sepal width, petal length et petal width). Nous avons donc des données multivariées. D'autre part, `species` a 150 individus et 1 variable (`species`).

On peut aussi obtenir avec `length` le nombre d'unités statistiques.

### Q3)Trouver les variables qualitatives et leurs modalités associées (utiliser la commande `unique`). Sont-elles nominales ou ordinales ?

Les variables qualitatives correspondent à variable de la table `species` ( avec les modalités `setosa`, `versicolor` et `virginica`) avec 50 fois l'occurrence pour chaque. Les variables sont des catégories de fleurs donc nominales. Elles ne suivent pas un ordre.

### Q4)Trouver les variables quantitatives. Sont-elles continues ou discrètes ?

Les variables quantitatives correspondent aux colonnes de la matrice meas (sepal length et width, petal length et width). En effet, ces variables sont des mesures en cm. Les variables sont continues car nous avons un grand nombre de valeurs possibles.

**Q5) Séparer la matrice en quatre variables sepalLength, sepalWidth, petalLength, petalWidth**

```
sepalLength=meas(:,1)
```

```
sepalWidth=meas(:,2)
```

```
petalLength=meas(:,3)
```




```
petalWidth=meas(:,4)
```

## 3.2 Etude de la variable species

**Q6) Quels sont les effectifs de chaque modalité ? (indication : il faut recoder les variables textes en numérique avec la commande strmatch par exemple pour éviter une boucle)**

```
indexSetosa=strmatch('setosa', species)
indexVersicolor=strmatch('versicolor', species)
indexVirginica=strmatch('virginica', species)
lengthSetosa=length(indexSetosa)
lengthVersicolor=length(indexVersicolor)
lengthVirginica=length(indexVirginica)
```

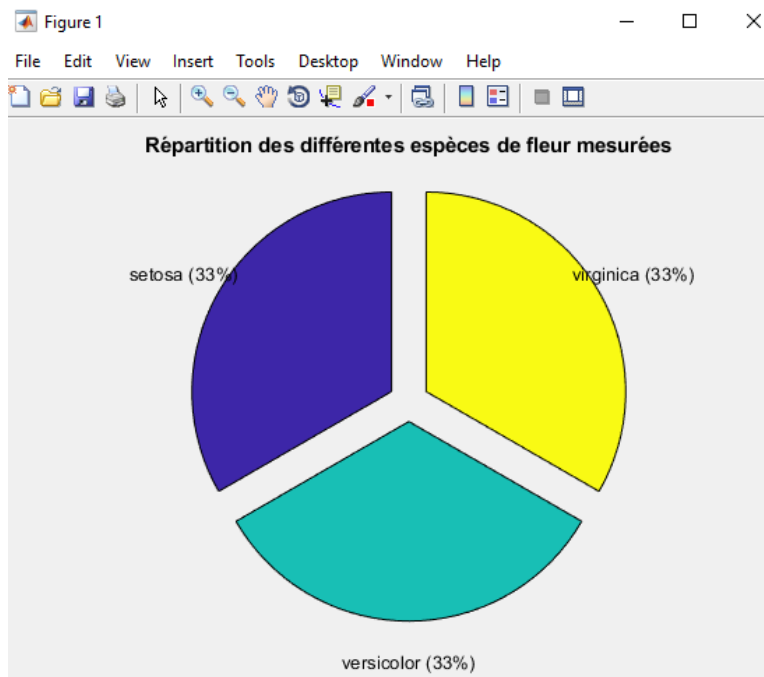
On obtient 50 fleurs pour chacune des modalités.

 lengthSetosa	50
 lengthVersicolor	50
 lenathViainica	50

**Q7) Les représentations graphiques classiques liées aux variables qualitatives sont la représentation en secteurs ou camembert (pie), la représentation en bâtons (hist). Représenter ces graphiques.**

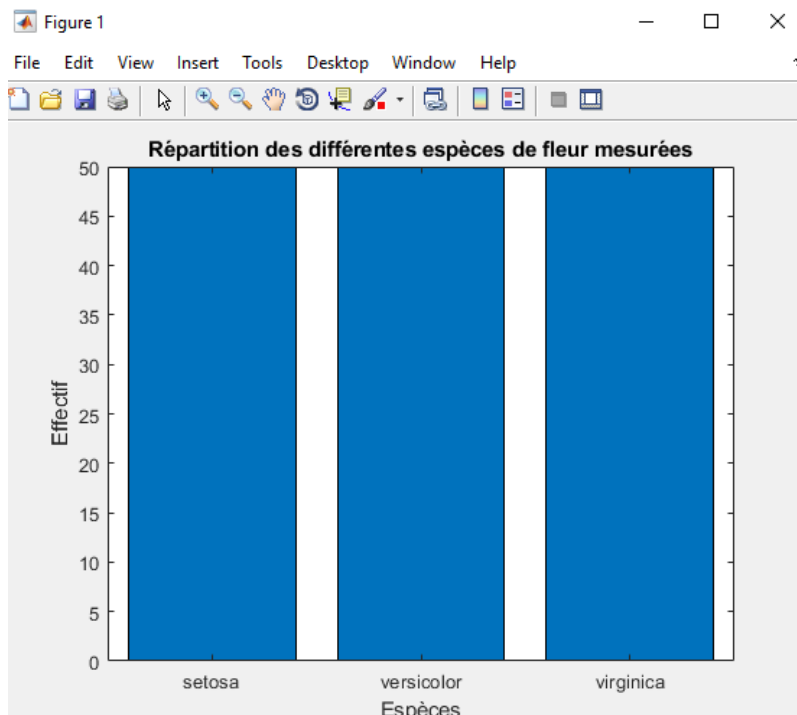
Avec pie :

```
cat = categorical({'setosa', 'versicolor', 'virginica'});  
pie(cat, [lengthSetosa,lengthVersicolor,lengthViginica]);  
title("Répartition des différentes espèces de fleur mesurées");
```



Avec hist :

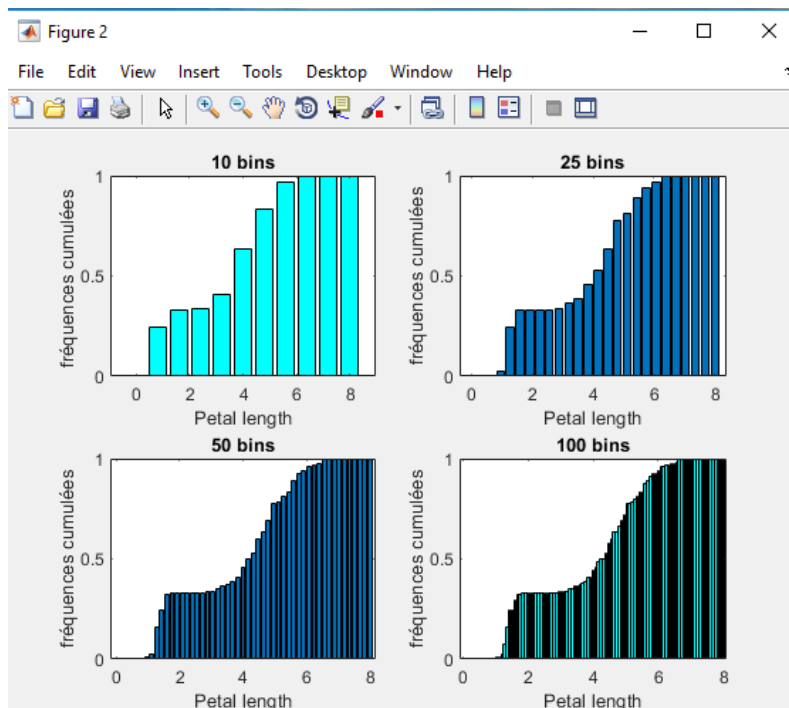
```
cat = categorical({'setosa', 'versicolor', 'virginica'});  
bar(cat, [lengthSetosa,lengthVersicolor,lengthViginica]);  
title("Répartition des différentes espèces de fleur mesurées");  
xlabel("Espèces");  
ylabel("Effectif");
```



### 3.3 Etude de la variable petalLength

Première approche : graphique

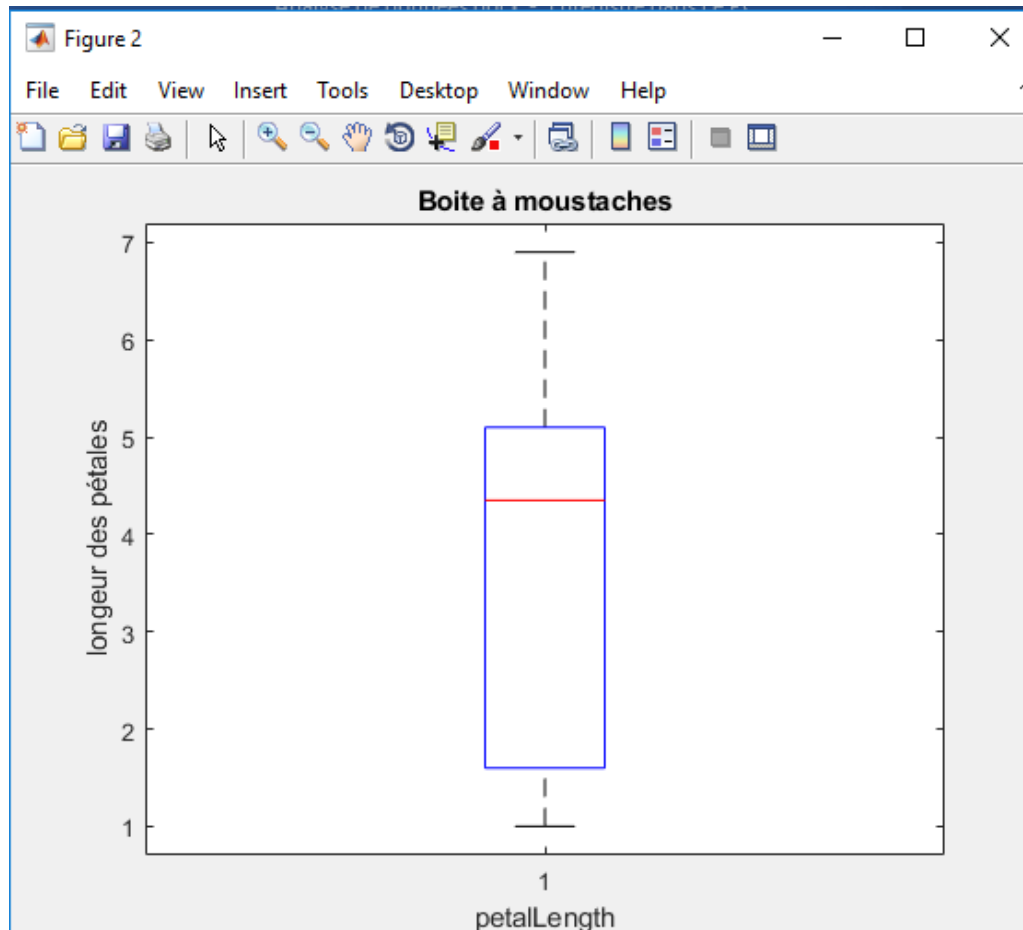
**Q8) Tracer l'histogramme en fréquences relatives et l'histogramme des fréquences relatives cumulées (histc). Faire varier le nombre de classes de l'histogramme (cf. aide).**



**Q9) Décrire les caractéristiques de l'histogramme et analyser ces caractéristiques en fonction du nombre de classes.**

On remarque que plus le nombre de classes est grand plus l'histogramme sera précis. Il y aura de plus en plus de rectangles. La pente sera moins carrée. Nous avons une courbe qui apparaît. Alors qu'avec moins de classe nous avons trop d'angles.

**Q10) Tracer la boîte à moustaches (boxplot) et rappeler les différents éléments la constituant.**



Voici une boîte à moustaches qui permet de visualiser sommairement la répartition des données de la série et les données « extrêmes ». Ici elle est simplifiée car je ne sais pas comment mettre les valeurs au niveau du schéma. Elle est composée de :

- D'un diagramme en boîte qui est un rectangle de bornes  $Q_1$ ,  $Q_3$  coupé au niveau de la médiane des longueurs des pétales  $m$  représentée par la barre horizontale rouge.
- De moustaches : deux segments joignant  $Q_1$  à  $X(1)$  pour l'un et  $Q_3$  à  $X(N)$  pour l'autre

(issue du cours)

## Deuxième approche : résumés numériques

**Q11) Calculer les résumés numériques de localisation (moyenne et médiane) et ceux de dispersion : (écart-type, variance et quartiles). Retrouver en particulier, les valeurs des éléments**

**de la boîte à moustache**

```
moyenne=mean (petalLength)
mediane=median (petalLength)
ecarttype=std (petalLength)
variance=var (petalLength)
quartiles=quantile (petalLength, 3)
```

<b>Moyenne</b>	<b>3.7580</b>	<b>Le pétale mesure en moyenne 3.7580 unités.</b>
<b>Médiane</b>	<b>4.3500</b>	<b>Il y a la moitié des fleurs qui ont une longueur inférieure à 4.3500 unités et l'autre partie supérieur.</b>
<b>Ecart type</b>	<b>1.7653</b>	<b>L'écart type correspond à la dispersion qu'il y aura autour des valeurs.</b>
<b>Variance</b>	<b>3.1163</b>	<b>La variance correspond au carré de l'écart type.</b>
<b>Quartiles</b>	<b>1.6000 4.3500 5.1000</b>	<b>La première valeur correspond au premier quartile, le deuxième la médiane et le second quartile. Et enfin le dernier le 3<sup>ième</sup> quartile.</b>