TP 3

Statistiques descriptives univariées

Compétences attendues.

- ✓ Regrouper une série statistique par modalités ou par classes.
- ✓ Connaître les indicateurs de position (moyenne, médiane, quartiles) et les commandes associées.
- ✓ Connaître les indicateurs de dispersion (écart-type, étendue, distance inter-quartile) et les commandes associées.
- ✓ Représenter graphiquement une série statistique.

Objectifs. L'objet des statistiques descriptives univariées (ou unidimensionnelles) est de fournir des résumés synthétiques, graphiques et numériques, de séries de valeurs observées sur une population ou un échantillon. On présente ici les indicateurs les plus couramment employés pour décrire une série statistique.

Mathieu Mansuy - Professeur de Mathématiques en ECS2 au Lycée Louis Pergaud (Besançon)
`mansuy.mathieu@hotmail.fr`

1 Principales notions en statistiques descriptives

1.1 Présentation des données

Soit $\Omega = \{\omega_1, \dots, \omega_n\}$ un ensemble fini. Un tel ensemble sera appelé *population* en statistique descriptive. On appellera ses éléments ω des *individus*, et son cardinal n l'*effectif de la population*.

Exemple. $\Omega =$ l'ensemble de la population française, $\Omega =$ l'ensemble des voitures immatriculées en France.

On étudie un *caractère* de cette population.

Définition.

Un *caractère* (ou *variable*) sur la population Ω est une application $X : \Omega \to E$, où E désigne un ensemble quelconque.

Si E est un ensemble de nombres, on dit que X est un caractère *quantitatif*. Dans le cas contraire, on parle de caractère qualitatif.

Exemple. Un caractère possible sur la population française est la taille (caractère quantitatif) ou encore la couleur des yeux (caractère qualitatif).

Nous ne traiterons que du cas des caractères quantitatifs.

Définition.

- On appelle *série statistique* de la population Ω pour le caractère X la donnée de la liste $X(\Omega) = (X(\omega_1), \dots, X(\omega_N))$ des valeurs prises par X sur Ω.
- Les valeurs prises par X sont appelées *modalités*.

  L'*effectif d'une modalité* est le nombre de fois où cette modalité apparait dans la série statistique.

  La *fréquence d'une modalité* est son effectif divisé par l'effectif total.

Le saviez vous ?

Les statistiques sont nées en Angleterre, au début du $XVII^e$ siècle pour décompter les décès lors d'une épidémie de peste. Ce n'était à l'époque que des données numériques, sans outil théorique pour les analyser. Il faut attendre le XIX^e siècle pour voir l'apparition de méthodes mathématiques pour l'étude de telles données. Ce n'est qu'à la fin du XIX^e siècle que la statistique devient une discipline à part entière des mathématiques sous l'impulsion des savants anglais Karl Pearson et Udny Yule.

Exemple. On considère la série statistique suivante :

$$2, 11, 7, 2, 15, 4, 5, 5, 5, 13, 5, 15, 7, 7, 8, 10, 10, 10, 11, 13, 7, 2, 15, 15$$

L'ensemble des modalités est $\{2, 4, 5, 7, 8, 10, 11, 13, 15\}$. L'effectif de la modalité 2 est $n_2 = 3$ et sa fréquence est $\dfrac{n_2}{n} = \dfrac{3}{24}$.

Représentation informatique. Sous `Scilab`, nous représenterons les séries statistiques par des vecteurs (lignes ou colonnes). L'effectif de la série est obtenue à l'aide de la commande `length`.

Regroupement par modalités

Pour présenter les données, on regroupe la série statistique par *modalités - effectifs*, c'est-à-dire qu'on donne :

- la liste (x_i) des modalités du caractère X,
- les effectifs (n_i) correspondants : pour tout i, n_i est le nombre d'individus ω de l'échantillon tels que $X(\omega) = x_i$.

On peut aussi choisir de présenter cette série regroupée par *modalité - fréquence*, en donnant les modalités (x_i) et les fréquences des modalités (f_i) correspondantes.

Pour passer d'une série statistique *brute* à une série statistique groupée par modalités ordonnées, on peut utiliser l'instruction `tabul`.

Définition.

Si `x` est un vecteur, `y=tabul(x,"i")` est une matrice à deux colonnes, la première contenant les valeurs prises par les composantes de `x` rangées dans l'ordre croissant (décroissant sans `"i"`) et la seconde contenant le nombre d'occurrences de chaque valeur.

Exercice 1

En utilisant `Scilab`, grouper la série statistique de l'exemple précédent par modalités - effectifs, puis par modalités - fréquences.

On utilise les commandes suivantes :

```
--> x = [2, 11, 7, 2, 15, 4, 5, 5, 5, 13, 5, 15, 7, 7, 8, 10, 10, 10, 11, 13, 7, 2, 15, 15
--> y = tabul(x,'i')
y =
    2.     3.
    4.     1.
    5.     4.
    7.     4.
    8.     1.
    10.    3.
    11.    2.
    13.    2.
    15.    4.
--> z = y ; z(:,2) = z(:,2)/length(x); z
z =
    2.     0.125
    4.     0.0416667
    5.     0.1666667
    7.     0.1666667
    8.     0.0416667
    10.    0.125
    11.    0.0833333
    13.    0.0833333
    15.    0.1666667
```

Définition.

On appelle *fréquence cumulée d'une modalité* la somme de toutes les fréquences des modalités qui lui sont inférieures.

Exercice 2

En utilisant la commande `cumsum`, déterminer le vecteur des fréquences cumulées de la série statistique précédente.

> Il suffit d'utiliser la commande :
>
> ```
> -->z(:,2) = cumsum(z(:,2))
>  z  =
>
>     2.     0.125
>     4.     0.1666667
>     5.     0.3333333
>     7.     0.5
>     8.     0.5416667
>     10.    0.6666667
>     11.    0.75
>     13.    0.8333333
>     15.    1.
> ```

Regroupement par classes

Lorsque $x = (x_1, \dots, x_n)$ est une série statistique dont le nombre de modalités est très grand, plutôt que de conserver toutes les valeurs, il est plus intéressant de les regrouper par classes :

- on considère une suite de réels $c = (c_0 < \cdots < c_k)$ définissant les *classes* $I_1 = [c_0, c_1]$, $I_2 =]c_1, c_2]$, $\dots$, $I_k =]c_{k-1}, c_k]$, l'*amplitude* de la classe I_i étant $c_i - c_{i-1}$;
- On note n_i le nombre d'éléments de x appartenant à l'intervalle I_i pour $1 \leq i \leq k$.

On se ramène ainsi à une série statistique de taille k, dont les modalités sont les milieux $y_i = \dfrac{c_{i-1} + c_i}{2}$ des classes et d'effectifs correspondants les n_i.

Pour grouper une série par classes, on peut utiliser l'instruction `dsearch` (*dichotomic search*).

Définition.

Soit `x` un vecteur, `c` un vecteur constitué d'une suite de réels strictement croissante.
L'instruction `[a,b]=dsearch(x,c)` crée deux vecteurs :

- `a` qui est de même taille que `x`, et qui à la position x_i indique le numéro j de l'intervalle I_j auquel il appartient ;
- `b` qui est de taille k (où $k+1$ est la taille de `c`) et qui contient en j-ème position le nombre d'éléments de `x` dans l'intervalle I_j.

Exercice 3

La commande `grand(q,r,'unf',a,b)` permet de simuler une matrice de taille $q \times r$ dont les entrées suivent la loi uniforme continue $\mathcal{U}([a,b])$.

À l'aide de la commande `grand`, simuler 10000 nombres suivant la loi uniforme sur $[0, 10]$. Regrouper cette série statistique par modalités, puis par classes (choisir 5 classes de même amplitude).
Quelle présentation de la série statistique préférez vous ? Que peut on dire de l'effectif de chaque classe ? Était ce prévisible ?

> Pour créer notre série statistique et la trier par modalités - effectifs, on utlise les commandes :
>
> ```
> -->v = grand(1,10000,'unf',0,10); w = tabul(v,'i')
> ```
>
> On obtient dans `w` un tableau à 10000 lignes et deux colonnes, avec les modalités triées par ordre croissant dans la colonne de gauche, et les effectifs de chaque modalité dans la colonne de droite. Sauf que tous ces effectifs sont de 1 : la loi étant continue, la probabilité d'obtenir deux fois le même nombre est en effet nulle. Ce tri par modalité n'a donc servi à rien, puisque notre série statistique est toujours à l'état brut dans la colonne de gauche de `w`, triée dans l'ordre croissant.
>
> Pour effectuer un tri par classes de notre série, on utilisera la commande suivante.
>
> ```
> -->[a,b] = dsearch(v, 0:2:10); b
>  b  =
>
>
>     2015.    1946.    1992.    2010.    2037.
> ```
>
> Il y a donc par exemple dans `v` 1992 modalités appartenant à l'intervalle $[4, 6]$. On préfèrera bien sûr cette présentation de la série, car elle classe de façon « lisible » les modalités. On notera aussi que les effectifs de chaque classe sont proches de $2000 = \frac{10000}{5}$. On pouvait s'y attendre, puisque la série est constituée de réalisations de la loi $\mathscr{U}([0, 10])$. La probabilité pour une variable aléatoire suivant une telle loi d'appartenir à un intervalle du type $[i, i + 2]$ (avec $i = 0, 2, 4, 6, 8$) est de $\frac{2}{10} = \frac{1}{5}$. La série étant de grande taille, on s'approche de cette probabilité théorique, avec environ $\frac{1}{5}$ de nos modalités dans chacune des classes.

1.2 Indicateurs de position

Définition.

On appelle *moyenne* de la série statistique $x = (x_1, \dots, x_n)$ le réel :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Remarque. Si la série statistique est groupée par modalités - effectifs (modalités $(m_1, \dots, m_p)$ d'effectifs $(n_1, \dots, n_p)$) alors on a :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_i \cdot m_i.$$

Définition.

La *médiane* d'une série statistique ordonnée un réel m partageant la série en deux séries d'effectifs égaux. Si $(x_1 \leq x_2 \leq \dots \leq x_n)$ est la série statistique ordonnée, m est défini par :

- si $n = 2p - 1$ est impaire, $m = x_p$ (la valeur du milieu) ;
- si $n = 2p$ est paire, $m = \dfrac{x_p + x_{p+1}}{2}$.

Définition.

Soit x un vecteur.

- `mean(x)` donne la moyenne du vecteur x.
- `median(x)` donne une médiane du vecteur x (non nécessairement ordonné).

Exercice 4

Déterminer la moyenne et la médiane de la série statistique.

```
-->mean(x), median(x)
 ans  =

    8.5
 ans  =

    7.5
```

Définition.

Le *premier quartile* q_1 d'une série statistique x est la plus petite modalité de x ayant une fréquence cumulée supérieure ou égale à $\frac{1}{4}$. Autrement dit, c'est la plus petite valeur de la série telle que 25 % des valeurs lui soient inférieures ou égales.

Le *troisième quartile* q_3 est la plus petite modalité de la série telle que 75 % des valeurs de la série lui soient inférieures ou égales.

De même, on définit les *déciles* et les *centiles* d'une série statistique.

Exercice 5

À l'aide du vecteur des fréquences cumulées de la série statistique, déterminer le premier quartile, le troisième quartile et le huitième décile.

Rappelons le vecteur des fréquences cumulées :

```
-->z
 z  =

    2.     0.125
    4.     0.1666667
    5.     0.3333333
    7.     0.5
    8.     0.5416667
    10.    0.6666667
    11.    0.75
    13.    0.8333333
    15.    1.
```

On en déduit que le premier quartile est 5, le troisième quartile est 11, et le huitième décile est 13.

Définition.

On appelle *mode* d'une série statistique toute modalité pour laquelle l'effectif est maximal (il peut y en avoir plusieurs).

Exercice 6

En utilisant l'instruction `find` (dont le fonctionnement a été expliqué lors du TP1), déterminer le(s) mode(s) de la série statistique.

Rappelons que `y` contient le tri par modalité - effectif de `x`. On peut utiliser les commandes suivantes :

```
-->find(y(:,2)==max(y(:,2)))
 ans  =

    3.    4.    9.

-->y(ans,1)
 ans  =

    5.
    7.
    15.
```

La première commande déterminer les indices des lignes où l'effectif est maximal, en cherchant dans la deuxième colonne de `y`. La deuxième commande extrait les modalités correspondantes, et donc les modes de `x`, à partir de la première colonne de `y`.

1.3 Indicateurs de dispersion

Définition.

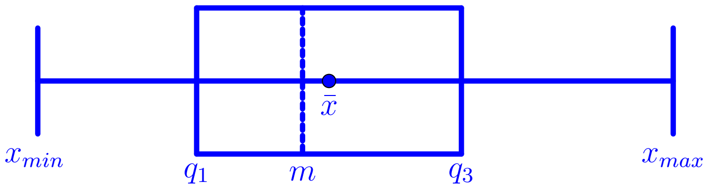
On appelle *écart-type* d'une série statistique $x = (x_1, \ldots, x_n)$ le réel :

$$\sigma_{n-1}(x) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

Définition.

- On appelle *étendue* d'une série statistique la différence entre la plus grande et la plus petite modalité.
- On appelle *distance inter-quartile* le réel $q_3 - q_1$.

Remarque. La distance inter-quartile est un indicateur de dispersion : c'est la longueur de l'*intervalle inter-quartile* $[q_1, q_3]$, lequel contient la moitié des valeurs de la série, réparties autour de la médiane. On représente parfois la boîte à moustache de la série statistique :



Définition.

Soit `x` un vecteur.

- `stdev(x)` (pour standard deviation) donne l'écart-type du vecteur `x`.
- `max(x)-min(x)` donne l'étendue du vecteur `x`.

Exercice 7

Déterminer l'écart-type de la série statistique, et représenter son diagramme à moustache.

> Il suffit d'utiliser la commande :
>
> ```
> -->stdev(x)
>  ans  =
>
>     4.3439113
> ```
>
> On a alors toutes les informations pour représenter le diagramme à moustache de `x`.

2 Représentation graphique

Aux notions de statistiques descriptives introduites dans la section précédente, on associe des représentations graphiques de trois types :

- *Diagramme en bâtons / diagramme cumulatif en bâton.*

  On représente la série statistique **groupée par modalités** en plaçant sur l'axe horizontal les modalités et en dressant à la verticale de chacune un bâton de hauteur égale à son effectif ou sa fréquence (resp. sont effectif cumulé ou sa fréquence cumulée).

- *Histogramme.*

  On représente la série statistique **groupée par classes** en plaçant les c_i sur un axe horizontal et en traçant à la verticale un rectangle de base $[c_i, c_{i+1}]$ d'aire égale à la fréquence de la classe correspondante.

- *Diagramme circulaire.*

  Chaque modalité (ou classe) est représentée par un secteur circulaire dont l'angle au centre est égal à la fréquence de la modalité (ou de la classe) multipliée par 360°.

Pour toutes ces représentations graphiques, on donne l'instruction `Scilab` à utiliser.

Définition.

- *Diagramme en bâtons.* On utilise l'instruction `bar`.

  Si `a` et `b` sont des vecteurs, `bar(a,b)` trace un diagramme en bâtons d'abcisse `a` et d'ordonnée `b`.

  On suppose que le vecteur `x` contient une série statistique brute. Pour obtenir un diagramme en bâtons, on peut alors utiliser la syntaxe :

  `y=tabul(x,"i")`

  `bar(y(:,1),y(:,2))` (effectifs) / `bar(y(:,1),y(:,2)/length(x))` (fréquences)

- *Histogramme.* On utilise l'instruction `histplot`.

  `histplot(c,x)` trace l'histogramme des données `x` où `c` est un vecteur aux composantes strictement croissantes définissant les classes.

  `histplot(n,x)` trace l'histogramme des données `x` où `n` désigne le nombre de classes (qui sont alors équiréparties entre la plus petite et la plus grande valeur de `x`).

- *Diagramme circulaire.* On utilise l'instruction `pie`.

  On suppose que le vecteur `x` contient une série statistique brute. Pour obtenir un diagramme circulaire, on peut alors utiliser la syntaxe :

  `y=tabul(x,"i")`

  `pie(y(:,2))` ou `pie(y(:,2),['x1','x2',...,'xp'])` si on veut rajouter des légendes.

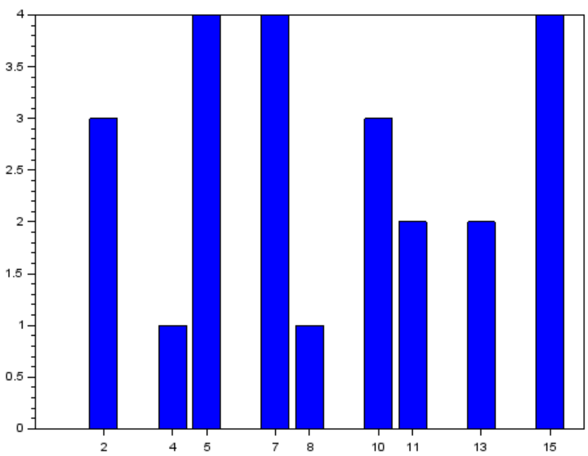
Exercice 8

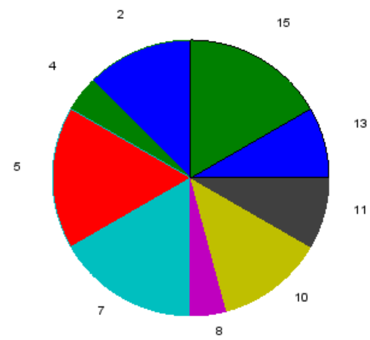
Représenter le diagramme en bâtons des effectifs de la série statistique, ainsi que le diagramme circulaire par modalité.

On utilise les commandes suivantes :

```
-->bar(y(:,1),y(:,2))
-->pie(y(:,2),['2','4','5','7','8','10','11','13','15'])
```

On obtient les diagrammes suivants :





- On retrouve en particulier facilement les modes sur le diagramme en bâtons.

Exercice 9

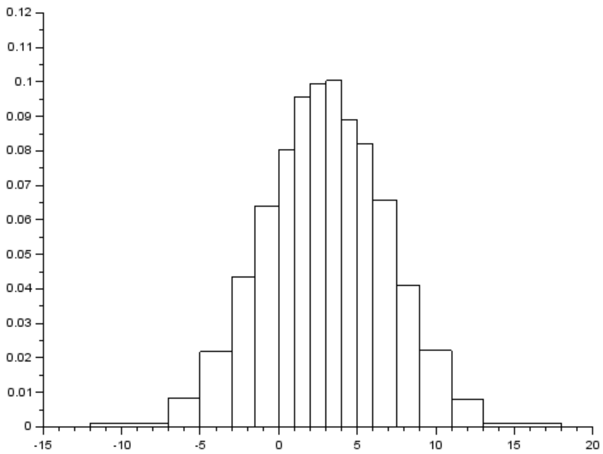
La commande `grand(q,r,'nor',`μ`,`σ`)` permet de simuler une matrice de taille $q \times r$ dont les entrées suivent la loi normale $\mathcal{N}(\mu, \sigma)$.

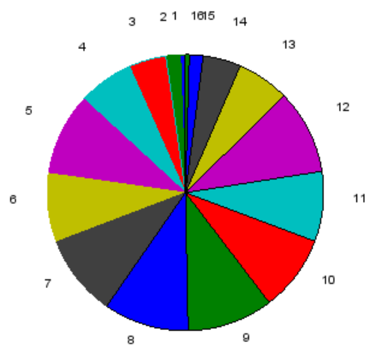
À l'aide de la commande `grand`, simuler 10000 nombres suivant la loi normale $\mathcal{N}(3, 4)$. Regrouper cette série par classes (à vous de choisir des classes appropriées), puis la représenter à l'aide d'un diagramme circulaire et d'un histogramme.

La distribution statistique va suivre une loi normale, c'est à dire une gaussienne (courbe en cloche) centrée en la moyenne 3. On choisit donc des classes avec une étendue de plus en plus grande lorsqu'on s'éloigne de 3. On propose par exemple les instructions suivantes :

```
s = grand(1,10000,'nor',3,4)
c = [-12,-7,-5,-3,-1.5,0,1,2,3,4,5,6,7.5,9,11,13,18]
histplot(c,s)
scf()
[a,b] = dsearch(s,c)
pie(b,['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','1
6'])
```

On obtient les diagrammes suivants :





3 Étude de la population mondiale

À partir de mon site `mathieu-mansuy.fr/ecs2`, télécharger et exécuter le fichier `population_mondiale.sce`, qui définit plusieurs matrices lignes :

- `pays` contient les noms des pays ;
- `superficie` contient la surface terrestre en milliers de km^2 de chaque pays ;

- `population` contient le nombre d'habitants en millions de chaque pays ;
- `naissance` contient le nombre de naissances sur 1000 habitants ;
- `mort` contient le nombre de décès sur 1000 habitants ;
- `homme` contient l'espérance de vie des hommes ;
- `femme` contient l'espérance de vie des femmes ;

à partir des données contenues dans l'étude 2017 de l'Institut National d'Études Démographiques (disponible également sur mon site). Pour les exercices qui suivent, vous trouverez en annexe de ce TP la liste des pays et de leurs index.

Exercice 10 (Interrogation de la base de données)

Écrire une fonction `donnees(n)` qui affiche le nom, la superficie, le nombre d'habitants et la densité de population du pays d'index n.

> On peut utiliser la fonction suivante :
>
> ```
> function [] = donnees(n)
>     disp(pays(n),'pays :')
>     disp(superficie(n),'superficie :')
>     disp(population(n),'population :')
>     disp(naissance(n),'naissance :')
>     disp(mort(n),'mort :')
>     disp(homme(n),'homme :')
>     disp(femme(n),'femme :')
> endfunction
> ```

Exercice 11

1. Calculer la surface terrestre mondiale, le nombre d'habitants mondial et la densité moyenne d'habitants au km^2.
2. Calculer la surface terrestre, le nombre d'habitants et la densité moyenne d'habitants au km^2 pour chaque continent.
3. Représenter la densité moyenne d'habitants au km^2 pour chaque continent en utilisant un diagramme en bâtons (on mettra en abscisse des entiers de 1 à 5).
4. Représenter la répartition de la surface terrestre puis du nombre d'habitants par continent sous la forme de diagrammes en camembert à l'aide de l'instruction `pie`.

   On pourra utiliser la commande `scf()` pour ouvrir une nouvelle fenêtre graphique.

> 1. On peut utiliser les instructions suivantes :
>
> ```
> -->S=sum(superficie)*1000, P = sum(population)*1000000, P/S
>  S  =
>
>     1.340D+08
>  P  =
> ```

```
    7.535D+09
 ans  =

    56.250259
```

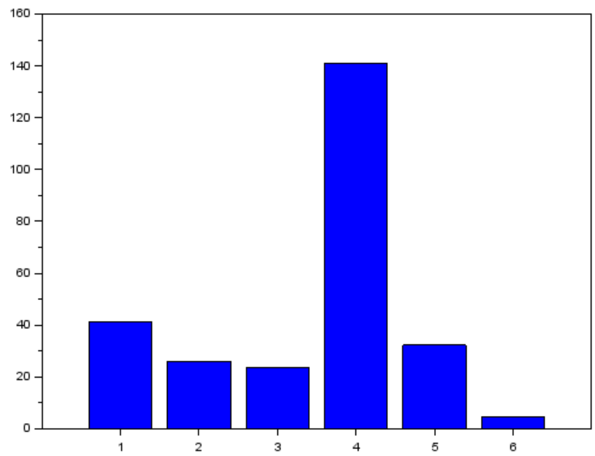
2. En copiant ce qu'on vient de faire, et en se reportant à l'index des pays pour déterminer l'étendue de chaque somme, on obtient les commandes :

```
//Surface de chaque continent
AfS = sum(superficie(1:57))*1000
AmNS = sum(superficie(58:86))*1000
AmSS = sum(superficie(87:99))*1000
AsieS = sum(superficie(100:150))*1000
EurS = sum(superficie(151:194))*1000
OcS = sum(superficie(195:208))*1000
//Population de chaque continent
AfP = sum(population(1:57))*1000000
AmNP = sum(population(58:86))*1000000
AmSP = sum(population(87:99))*1000000
AsieP = sum(population(100:150))*1000000
EurP = sum(population(151:194))*1000000
OcP = sum(population(195:208))*1000000
//Vecteur densité de population de chaque continent
D = [AfP,AmNP,AmSP,AsieP,EurP,OcP]./[AfS,AmNS,AmSS,AsieS,EurS,OcS]
```

3. On utilise la commande `bar`, avec pour abscisse les entiers de 1 à 6 (pour chacun des continents) :

```
--> bar(1:6,D)
```

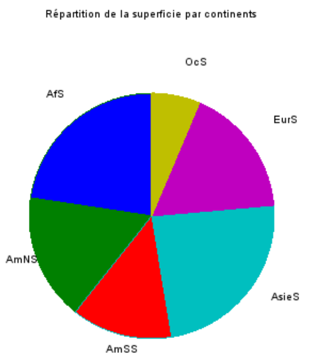
On obtient le diagramme suivant :

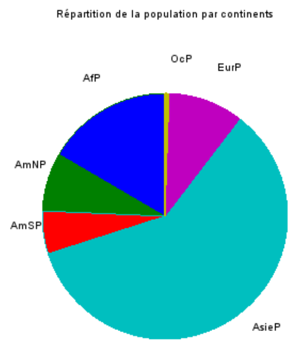


4. On utilise les instructions :

```
scf()
pie([AfP,AmNP,AmSP,AsieP,EurP,OcP],['AfP','AmNP','AmSP','AsieP','EurP','OcP'])
title('Répartition de la population par continents')
scf()
pie([AfS,AmNS,AmSS,AsieS,EurS,OcS],['AfS','AmNS','AmSS','AsieS','EurS','OcS'])
title('Répartition de la superficie par continents')
```

On obtient les diagrammes en camembert suivants :





Exercice 12

On considère l'espérance de vie des hommes (ou des femmes) par pays.

1. Calculer la moyenne sur l'ensemble des pays. Ce résultat correspond-il à l'espérance de vie mondiale des hommes (ou des femmes) ?
2. Calculer l'écart-type et la médiane.
3. Calculer les espérances de vie minimale et maximale en précisant les pays correspondant à ces valeurs extrémales (utiliser l'instruction `find`).
4. Représenter l'histogramme de l'espérance de vie des hommes sur l'intervalle $[0, 100]$ avec 20 classes. Quelle est la classe modale de l'espérance de vie des hommes ?
5. Trier le tableau `homme` par ordre croissant et en déduire :
   (a) les valeurs du premier et du troisième quartile ainsi que l'écart inter-quartile ;
   (b) les valeurs du premier et du neuvième décile ainsi que la liste des pays dont l'espérance de vie est inférieure au premier décile ou supérieure au neuvième décile.

1. Calculons la moyenne du vecteur `femme` :

```
-->mean(femme)
 ans  =

    74.846154
```

Ce résultat ne correspond pas cependant à l'espérance de vie moyenne des femmes dans le monde : en effet, la Chine et la France par exemples comptent chacune pour 1, alors qu'il faudrait pondérer par le nombre de femme dans chaque pays. Mais nous n'avons pas accès à ces données malheureusement.

2. On procède ainsi :

```
tdev(femme), median(femme)
 ans  =

    8.3594018
 ans  =

    77.
```

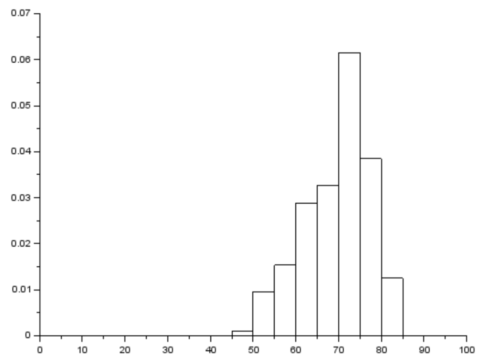
3. On peut utiliser l'instruction suivante :

```
-->min(femme), max(femme)
 ans  =

    52.
 ans  =

    89.

-->pays(find(femme==min(femme)))
 ans  =

 Sierra Leone

-->pays(find(femme==max(femme)))
 ans  =

!Monaco  Saint-Marin  !
```

4. On utilise l'instruction suivante :

```
histplot(0:5:100,homme)
```

On obtient l'histogramme :



La classe modale de l'espérance de vie des hommes est donc $[70, 75]$.

5. On effectue les instructions suivantes :

```
-->y = tabul(homme,'i');
-->y(:,2) = cumsum(y(:,2))/length(homme)
 y  =

    50.    0.0048077
    51.    0.0192308
    52.    0.0288462
    54.    0.0384615
    55.    0.0528846
    56.    0.0625
    57.    0.0721154
    58.    0.0865385
    59.    0.1201923
    60.    0.1298077
    61.    0.1538462
    62.    0.1826923
    63.    0.2163462
    64.    0.2548077
    65.    0.2740385
    66.    0.3028846
    67.    0.3317308
    68.    0.375
    69.    0.4134615
    70.    0.4375
    71.    0.5
    72.    0.5480769
    73.    0.6201923
    74.    0.6875
    75.    0.7451923
    76.    0.7884615
    77.    0.8269231
    78.    0.8557692
    79.    0.8990385
    80.    0.9375
    81.    0.9903846
    85.    1.
```

La valeur du premier quartile est donc de 64 ans, du troisième quartile est de 76 ans, et l'écart inter-quartile est donc de $76 - 64 = 12$ ans.

Les valeurs du premier et du neuvième déciles sont 59 ans et 80 ans respectivement. Pour trouver la liste des pays dont l'espérance de vie est inférieure au premier décile ou supérieure au neuvième décile, on peut procéder ainsi :

```
-->pays(find(homme<=59))
 ans  =


         column 1 to 6
```

```
!Bénin  Burkina Faso  Côte d Ivoire  Guinée  Guinée-Bissau  Mali  !

         column  7 to 13

!Niger  Nigeria  Sierra Leone  Togo  Burundi  Mozambique  Somalie  !

         column 14 to 18

!Sud-Soudan  Zambie  Zimbabwe  Angola  Cameroun  !

         column 19 to 24

!Centrafricaine (République)  Congo  Congo (rep)  Gabon  Tchad  Lesotho  !

         column 25

!Swaziland  !

-->pays(find(homme>=80))
 ans  =


         column 1 to 6

!Chypre  Israël  Singapour  Chine - Hong Kong  Chine - Macao  Japon  !

         column  7 to 13

!Islande  Norvège  Suède  Liechtenstein  Luxembourg  Monaco  Pays-Bas  !

         column 14 to 20

!Suisse  Andorre  Espagne  Italie  Malte  Saint-Marin  Australie  !

         column 21

!Nouvelle-Zélande  !
```

Exercice 13

On rappelle que le taux d'accroissement naturel est la différence entre la natalité et la mortalité.

1. Quels sont les accroissements minimaux et maximaux ? Préciser les pays.
2. Faire afficher la liste des pays pour lesquels l'accroissement est négatif.
3. Déterminer l'accroissement mondial moyen.
4. Dans ses projections, l'INED prévoit une population mondiale de 9731 millions d'habitants en 2050. Cela est-il conforme à l'hypothèse d'un taux d'accroissement constant ?

1. On procède ainsi :

```
-->accroissement = naissance-mort;
-->min(accroissement), max(accroissement)
 ans  =

  - 6.
 ans  =

    38.
-->pays(find(accroissement==min(accroissement)))
 ans  =

 Bulgarie

-->pays(find(accroissement==max(accroissement)))
 ans  =

 Niger
```

2. Toujours avec la commande `find` :

```
->pays(find(accroissement<=0))
 ans  =


         column 1 to 7

!Japon  Estonie  Finlande  Lettonie  Lituanie  Allemagne  Monaco  !

         column  8 to 14

!Biélorussie  Bulgarie  Hongrie  Moldavie  Pologne  Roumanie  Russie  !

         column 15 to 21

!Ukraine  Bosnie-Herzégovine  Croatie  Espagne  Grèce  Italie  Portugal  !

         column 22 to 24

!Saint-Marin  Serbie  Slovénie  !
```

3. On utilise la commande `mean` :

```
->mean(accroissement)
 ans  =

    13.100962
```

4. Partons de l'hypothèse d'un taux d'accroissement constant. Notons m le taux d'accroissement de la population mondiale ramené à un habitant, et p la population mondiale. Alors chaque année, la population mondiale change et devient $p = p + m \times p$. On effectue une boucle `for` pour répéter cela jusqu'en 2050 :

```
m = mean(accroissement)/1000
p = sum(population)*1000000
for k = 2017:2050
    p = p+m*p
end
disp(p)
```

On trouve une population de `1.173D+10`, soit plus de 1173 millions d'habitants en 2050. L'INED n'a donc surement pas pris l'hypothèse d'un taux d'accroissement constant.

Annexe. liste des pays et de leurs index

Afrique

Afrique septentrionale

1. Algérie
2. Égypte
3. Libye
4. Maroc
5. Sahara occidental
6. Soudan
7. Tunisie

Afrique occidentale

8. Bénin
9. Burkina Faso
10. Cap-Vert
11. Côte d'Ivoire
12. Gambie
13. Ghana
14. Guinée
15. Guinée-Bissau
16. Liberia
17. Mali
18. Mauritanie
19. Niger
20. Nigeria
21. Sénégal
22. Sierra Leone
23. Togo

Afrique orientale

24. Burundi
25. Comores
26. Djibouti
27. Érythrée
28. Éthiopie
29. Kenya
30. Madagascar
31. Malawi
32. Maurice
33. Mayotte
34. Mozambique
35. Ouganda
36. Réunion
37. Rwanda
38. Seychelles
39. Somalie
40. Sud-Soudan
41. Tanzanie
42. Zambie
43. Zimbabwe

Afrique centrale

44. Angola
45. Cameroun
46. Centrafricaine (Rép.)
47. Congo
48. Congo(Rép. dém.)
49. Gabon
50. Guinée équatoriale
51. Sao Tomé-et-Principe
52. Tchad

Afrique australe

53. Afrique du Sud
54. Botswana
55. Lesotho
56. Namibie
57. Swaziland

Amérique

Amérique septentrionale

58. Canada
59. États Unis

Amérique centrale

60. Belize
61. Costa Rica
62. Guatemala
63. Honduras
64. Mexique
65. Nicaragua
66. Panama
67. Salvador

Caraïbes

68. Antigua-et-Barbuda
69. Aruba
70. Bahamas
71. Barbade
72. Cuba
73. Curaçao
74. Dominicaine (Rép.)
75. Dominique
76. Grenade
77. Guadeloupe
78. Haïti
79. Jamaïque
80. Martinique
81. Porto Rico
82. Sainte Lucie
83. St Vincent Grenadines
84. St.Kitts-et-Nevis
85. Trinité-et-Tobago
86. Vierges (Îles)

Amérique du sud

87. Argentine
88. Bolivie
89. Brésil
90. Chili
91. Colombie
92. Équateur
93. Guyana
94. Guyane (française)
95. Paraguay
96. Pérou
97. Surinam
98. Uruguay
99. Venezuela

Asie

Asie occidentale

100. Arabie saoudite
101. Arménie
102. Azerbaïdjan
103. Bahreïn
104. Chypre
105. Émirats arabes unis

106. Géorgie
107. Irak
108. Israël
109. Jordanie
110. Koweït
111. Liban
112. Oman
113. Palestine (Territoires)
114. Qatar
115. Syrie
116. Turquie
117. Yémen

Asie centrale

118. Kazakhstan
119. Kirghizistan
120. Tadjikistan
121. Turkménistan
122. Ouzbékistan

Asie du sud

123. Afghanistan
124. Bangladesh
125. Bhoutan
126. Pakistan
127. Inde
128. Iran
129. Maldives
130. Népal
131. Sri Lanka

Asie du sud-ouest

132. Brunei
133. Cambodge
134. Indonésie
135. Laos
136. Malaisie
137. Myanmar (Birmanie)
138. Philippines
139. Singapour
140. Thaïlande
141. Timor-Est
142. Viêt Nam

Asie orientale

143. Chine
144. Chine-Hong Kong
145. Chine-Macao
146. Corée du Nord
147. Corée du Sud
148. Japon
149. Mongolie
150. Taïwan

Europe

Europe septentrionale

151. Danemark
152. Estonie
153. Finlande
154. Irlande
155. Islande
156. Lettonie
157. Lituanie
158. Norvège
159. Royaume-Uni
160. Suède

Europe occidentale

161. Allemagne
162. Autriche
163. Belgique
164. France (métropolitaine)
165. Liechtenstein
166. Luxembourg
167. Monaco
168. Pays-Bas
169. Suisse

Europe orientale

170. Biélorussie
171. Bulgarie
172. Hongrie
173. Moldavie
174. Pologne
175. Roumanie
176. Russie
177. Slovaquie
178. Tchèque (République)
179. Ukraine

Europe méridionale

180. Albanie
181. Andorre
182. Bosnie-Herzégovine
183. Croatie
184. Espagne
185. Grèce
186. Italie
187. Kosovo
188. Macédoine
189. Malte
190. Monténégro
191. Portugal
192. Saint-Marin
193. Serbie
194. Slovénie

Océanie

195. Australie
196. Fidji
197. Guam
198. Kiribati
199. Marshall (Îles)
200. Micronésie (États fédérés de)
201. Nouvelle-Calédonie
202. Nouvelle-Zélande
203. Papouasie-Nouvelle Guinée
204. Polynésie française
205. Salomon (Îles)
206. . Samoa occidentales
207. Tonga
208. Vanuatu