Marcel Délèze
Edition 2017

Thème : Statistiques I, § 2 Statistique descriptive à une variable continue
Lien vers les énoncés des exercices:
https://www.deleze.name/marcel/sec2/applmaths/csud/statistique_1/2-stat_I.pdf

Packages de l’auteur

- On peut consulter le mode d’emploi du package **Statistique**:
  https://www.deleze.name/marcel/sec2/applmaths/packages/aide/Statistique.pdf
- Avant d’utiliser le package, il faut le charger en donnant son adresse web:

```
Needs["Statistique`",
nécessite
  "https://www.deleze.name/marcel/sec2/applmaths/packages/Statistique.m"]
```

Voici la liste des instructions disponibles :

```
Names["Statistique`*"]
noms
```

```
{amplitudes, densiteContinue, densites, diagrammeBatons,
 diagrammeCumulatif, distributionContinue, distributionLissee, fctDensite,
 fctFrequenceCumulee, frequenceCumuleeContinue, frequenceCumuleeLissee,
 histogramme, InterpolatedQuantile, noeudsPolygonaux, polygoneDeDensite,
 quantileC, quantileLisse, sommesCumulees, StandardDeviationMLE, VarianceMLE}
```

- Le package **Tableaux** contient des commandes qui facilitent la présentation des données et résultats sous la forme de tableaux:

```
Needs["Tableaux`",
nécessite
  "https://www.deleze.name/marcel/sec2/applmaths/packages/Tableaux.m"]
```

```
Names["Tableaux`*"]
noms
```

```
{afficheTableau, afficheTableauTitre, arrondis, fusionneColonnes,
 fusionneLignes, fusionneTableaux, prodCart, prodCartTrans, tableauGraph}
```

- On peut consulter le mode d’emploi du package **Tableaux**:
  https://www.deleze.name/marcel/sec2/applmaths/packages/aide/Tableaux.pdf

Pour ne pas oublier d'exécuter ces instructions au début de chaque session de travail, il est conseillé de déclarer les instructions **Needs** comme étant des cellules d'initialisation. Pour ce faire, sélectionnez les cellules voulues puis passez par le menu

Cell / Cell properties / Initialization cell

Corrigé de l'exercice 2 - 1

Partie a)

Bornes des classes

$b_0 = 27.5;\quad b_1 = 37.5;\quad b_2 = 47.5;\quad b_3 = 52.5;$

$b_4 = 57.5;\quad b_5 = 62.5;\quad b_6 = 72.5;\quad b_7 = 82.5$

Effectifs des classes

$n_1 = 3;\quad n_2 = 51;\quad n_3 = 74;\quad n_4 = 112;\quad n_5 = 92;\quad n_6 = 62;\quad n_7 = 6$

Effectif total

$$n = n_1 + ... + n_7 = 400$$

Fréquences des classes

$$f_1 = \frac{n_1}{n} = \frac{3}{400};\quad f_2 = \frac{51}{400};\quad f_3 = \frac{37}{200};\quad f_4 = \frac{7}{25};\quad f_5 = \frac{23}{100};\quad f_6 = \frac{31}{200};\quad f_7 = \frac{3}{200}$$

Fréquences cumulées

$$F_1 = f_1 = \frac{3}{400} = 0.0075$$
$$F_2 = F_1 + f_2 = \frac{3}{400} + \frac{51}{400} = \frac{54}{400} = \frac{27}{200} = 0.135$$
$$F_3 = F_2 + f_3 = \frac{54}{400} + \frac{74}{400} = \frac{128}{400} = \frac{8}{25} = 0.32$$
$$F_4 = F_3 + f_4 = \frac{128}{400} + \frac{112}{400} = \frac{240}{400} = \frac{3}{5} = 0.6$$
$$F_5 = F_4 + f_5 = \frac{240}{400} + \frac{92}{400} = \frac{332}{400} = \frac{83}{100} = 0.83$$
$$F_6 = F_5 + f_6 = \frac{332}{400} + \frac{62}{400} = \frac{394}{400} = \frac{197}{200} = 0.985$$
$$F_7 = F_6 + f_7 = \frac{394}{400} + \frac{6}{400} = \frac{400}{400} = 1$$

Distribution empirique

$F_c(b_0) = 0;\quad F_c(b_1) = F_1;\quad F_c(b_2) = F_2;\quad ...;\quad F_c(b_k) = F_k$

x	27.5	37.5	47.5	52.5	57.5	62.5	72.5	82.5
***F*(*x*)**	0.	0.0075	0.135	0.32	0.6	0.83	0.985	1.

La médiane est calculée par interpolation linéaire

x	52.5	me	57.5
***F*(*x*)**	0.32	0.5	0.6

$$\frac{0.5 - 0.32}{me - 52.5} = \frac{0.6 - 0.32}{57.5 - 52.5}$$
$$\frac{0.18}{me - 52.5} = \frac{0.28}{5}$$
$$(me - 52.5)\ 0.28 = 0.9$$
$$me - 52.5 = \frac{0.9}{0.28}$$
$$me = 52.5 + \frac{0.9}{0.28} = 55.7143$$

Ecart interquartile

x	57.5	q3	62.5
$F(x)$	0.6	0.75	0.83

$$\frac{0.75 - 0.6}{q3 - 57.5} = \frac{0.83 - 0.6}{62.5 - 57.5}$$

$$\frac{0.15}{q3 - 57.5} = \frac{0.23}{5}$$

$$(q3 - 57.5)\ 0.23 = 0.75$$

$$q3 - 57.5 = \frac{0.75}{0.23}$$

$$q3 = 57.5 + \frac{0.75}{0.23} = 60.7609$$

x	47.5	q1	52.5
$F(x)$	0.135	0.25	0.32

$$\frac{0.25 - 0.135}{q1 - 47.5} = \frac{0.32 - 0.135}{52.5 - 47.5}$$

$$\frac{0.115}{q1 - 47.5} = \frac{0.185}{5}$$

$$(q1 - 47.5)\ 0.185 = 0.575$$

$$q1 - 47.5 = \frac{0.575}{0.185}$$

$$q1 = 47.5 + \frac{0.575}{0.185} = 50.6081$$

$$\text{interQuart} = q3 - q1 = 60.7609 - 50.6081 = 10.1528$$

Partie b)

Densités

$$h_1 = \frac{f_1}{b_1 - b_0} = \frac{\frac{3}{400}}{10} = \frac{3}{4000} = 0.00075$$

$$h_2 = \frac{f_2}{b_2 - b_1} = \frac{\frac{51}{400}}{10} = \frac{51}{4000} = 0.01275$$

$$h_3 = \frac{f_3}{b_3 - b_2} = \frac{\frac{74}{400}}{5} = \frac{74}{2000} = 0.037$$

$$h_4 = \frac{f_4}{b_4 - b_3} = \frac{\frac{112}{400}}{5} = \frac{112}{2000} = 0.056$$

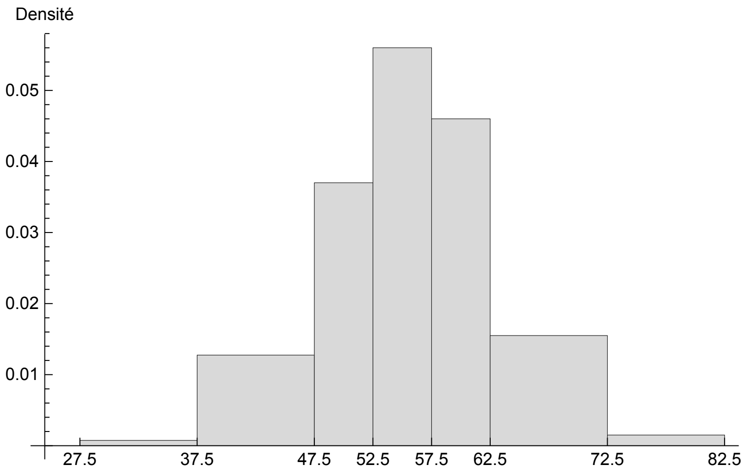
$$h_5 = \frac{f_5}{b_4 - b_3} = \frac{\frac{92}{400}}{5} = \frac{92}{2000} = 0.046$$

$$h_5 = \frac{f_5}{b_5 - b_4} = \frac{\frac{92}{400}}{5} = \frac{62}{2000} = 0.046$$

$$h_6 = \frac{f_6}{b_6 - b_5} = \frac{\frac{62}{400}}{10} = \frac{62}{4000} = 0.0155$$

$$h_7 = \frac{f_7}{b_7 - b_6} = \frac{\frac{6}{400}}{10} = \frac{6}{4000} = 0.0015$$

Histogramme



Classe modale : la classe dont la densité est maximale est la 4-ème classe, sur l'intervalle [52.5; 57.5[.

Partie c)

Centres des classes

$$c_1 = \frac{b_0 + b_1}{2} = \frac{27.5 + 37.5}{2} = 32.5$$

$$c_2 = \frac{b_1 + b_2}{2} = \frac{37.5 + 47.5}{2} = 42.5$$

$$c_3 = \frac{b_2 + b_3}{2} = \frac{47.5 + 52.5}{2} = 50$$

$$c_4 = \frac{b_3 + b_4}{2} = \frac{52.5 + 57.5}{2} = 55$$

$$c_5 = \frac{b_4 + b_5}{2} = \frac{57.5 + 62.5}{2} = 60$$

$$c_6 = \frac{b_5 + b_6}{2} = \frac{62.5 + 72.5}{2} = 67.5$$

$$c_7 = \frac{b_6 + b_7}{2} = \frac{72.5 + 82.5}{2} = 77.5$$

Moyenne arithmétique

$$m = c_1 f_1 + c_2 f_2 + c_3 f_3 + c_4 f_4 + c_5 f_5 + c_6 f_6 + c_7 f_7 =$$
$$32.5 \frac{3}{400} + 42.5 \frac{51}{400} + 50 \frac{74}{400} + 55 \frac{112}{400} + 60 \frac{92}{400} + 67.5 \frac{62}{400} + 77.5 \frac{6}{400} = 55.7375$$

Variance

$$v = (c_1 - m)^2 f_1 + (c_2 - m)^2 f_2 + (c_3 - m)^2 f_3 + (c_4 - m)^2 f_4 +$$
$$(c_5 - m)^2 f_5 + (c_6 - m)^2 f_6 + (c_7 - m)^2 f_7 = (32.5 - 55.7375)^2 \frac{3}{400} +$$
$$(42.5 - 55.7375)^2 \frac{51}{400} + (50 - 55.7375)^2 \frac{74}{400} + (55 - 55.7375)^2 \frac{112}{400} +$$

$$(60 - 55.7375)^2 \frac{92}{400} + (67.5 - 55.7375)^2 \frac{62}{400} + (77.5 - 55.7375)^2 \frac{6}{400} = 65.3623$$

Ecart-type

$$s = \sqrt{v} = \sqrt{65.3623} = 8.0847$$

Corrigé de l'exercice 2 - 2

Partie a)

Bornes des classes

```
b = {27.5, 37.5, 47.5, 52.5, 57.5, 62.5, 72.5, 82.5};
```

Effectifs des classes

```
eff = {3, 51, 74, 112, 92, 62, 6};
```

Effectif total

```
n = Apply[Plus, eff]
```

remp··· plus

```
400
```

Fréquences des classes

$$freq = \frac{eff}{n}$$

$$\left\{\frac{3}{400}, \frac{51}{400}, \frac{37}{200}, \frac{7}{25}, \frac{23}{100}, \frac{31}{200}, \frac{3}{200}\right\}$$

Fréquences cumulées

```
freqCum = Accumulate[freq]
```

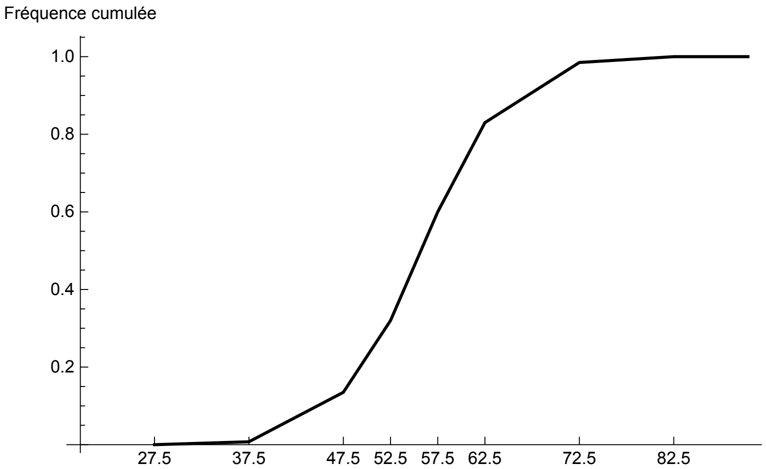
accumule

$$\left\{\frac{3}{400}, \frac{27}{200}, \frac{8}{25}, \frac{3}{5}, \frac{83}{100}, \frac{197}{200}, 1\right\}$$

Distribution empirique

```
frequenceCumuleeContinue[b, freq, AxesLabel → {None, "Fréquence cumulée"}]
```

titre d'axe aucun



Médiane

```
me = quantileC[b, freq, 1/2]
```

55.7143

Ecart interquartile

```
quantileC[b, freq, 3/4]
```

60.7609

```
quantileC[b, freq, 1/4]
```

50.6081

```
interQuart = quantileC[b, freq, 3/4] - quantileC[b, freq, 1/4]
```

10.1528

Partie b)

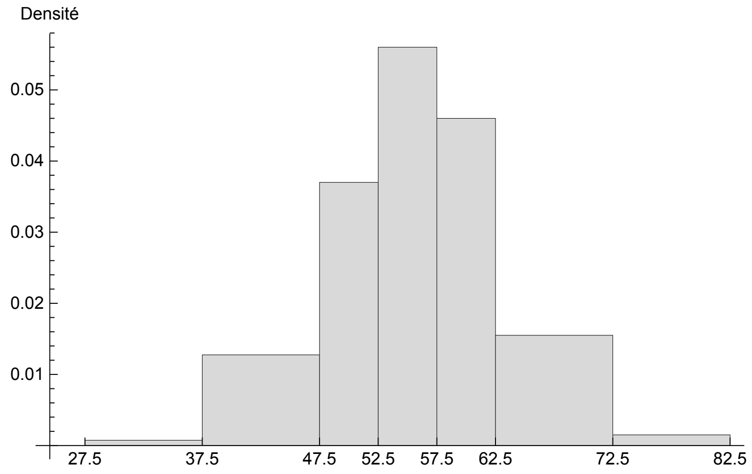
Densités

```
h = densites[b, freq]
```

{0.00075, 0.01275, 0.037, 0.056, 0.046, 0.0155, 0.0015}

Histogramme

```
histogramme[b, freq, AxesLabel → {None, "Densité"}]
```
titre d'axe aucun



Classe modale

```
Position[h, Max[h]]
```
position maximum

{{4}}

```
clMo = Position[h, Max[h]][[1, 1]]
```
position maximum

4

```
{b[[clMo]], b[[clMo + 1]]}
```

{52.5, 57.5}

Partie c)

Centres des classes

$$c = \frac{Drop[b, 1] + Drop[b, -1]}{2}$$

{32.5, 42.5, 50., 55., 60., 67.5, 77.5}

Moyenne arithmétique

```
m = c.freq
```

55.7375

Variance

$$v = (c - m)^2.freq$$

65.3623

Ecart-type

$$s = \sqrt{v}$$

8.0847

Corrigé de l'exercice 2-3

Partie a)

Première méthode : Calcul de la fréquence d'un intervalle à partir des fréquences des classes

$$f([49.5;\ 50.5[) = \frac{50.5 - 49.5}{b_3 - b_2} f_3 = \frac{1}{52.5 - 47.5} \frac{74}{400} = 0.037$$

Deuxième méthode : Calcul de la fréquence d'un intervalle à partir des densités:

$$f([49.5;\ 50.5[) = \frac{50.5 - 49.5}{b_3 - b_2} f_3 = (50.5 - 49.5) \frac{f_3}{b_3 - b_2} = (50.5 - 49.5)\ h_3 = (50.5 - 49.5)\ 0.037 = 0.037$$

$$f([50;\ 52[) = \frac{52 - 50}{b_3 - b_2} f_3 = \frac{2}{52.5 - 47.5} \frac{74}{400} = 0.074$$

$$f([50;\ 52]) = f([50;\ 52[) = 0.074$$

$$f([60;\ 80[) = f([60;\ 62.5[) + f([62.5;\ 72.5[) + f([72.5;\ 80[) = \frac{62.5 - 60}{b_5 - b_4} f_5 + f_6 + \frac{80 - 72.5}{b_7 - b_6} f_7 = \frac{62.5 - 60}{62.5 - 57.5} \frac{92}{400} + \frac{62}{400} + \frac{80 - 72.5}{82.5 - 72.5} \frac{6}{400} = 0.28125$$

$$f(]-\infty;\ 60[) = f([27.5;\ 37.5[) + f([37.5;\ 47.5[) + f([47.5;\ 52.5[) + f([52.5;\ 57.5[) + f([57.5 .5;\ 60[) = f_1 + f_2 + f_3 + f_4 + \frac{60 - 57.5}{b_5 - b_4} f_5 = F_4 + \frac{60 - 57.5}{b_5 - b_4} f_5 = 0.6 + \frac{60 - 57.5}{62.5 - 57.5} \frac{92}{400} = 0.715$$

Partie b)

Calcul des valeurs de la fonction de distribution par interpolation linéaire

x	$-\infty$	20	27.5
$F(x)$	0	y	0

$$F_c\,(20) = y = 0$$

x	37.5	40	47.5
$F(x)$	0.0075	y	0.135

$$\frac{y - 0.0075}{40 - 37.5} = \frac{0.135 - 0.0075}{47.5 - 37.5}$$

$$\frac{y - 0.0075}{2.5} = \frac{0.1275}{10}$$

$$y - 0.0075 = 0.031875$$

$$F_c\,(40) = y = 0.031875 + 0.0075 = 0.039375$$

x	57.5	60	62.5
$F(x)$	0.6	y	0.83

$$\frac{y - 0.6}{60 - 57.5} = \frac{0.83 - 0.6}{62.5 - 57.5}$$

$$\frac{y - 0.6}{2.5} = \frac{0.23}{5}$$

$$y - 0.6 = 0.115$$

$$F_c\,(60) = y = 0.715$$

x	72.5	80	82.5
$F(x)$	0.985	y	1

$$\frac{y - 0.985}{80 - 72.5} = \frac{1 - 0.985}{82.5 - 72.5}$$

$$\frac{y - 0.985}{7.5} = \frac{0.015}{10}$$

$$y - 0.985 = 0.01125$$

$$F_c\,(80) = y = 0.99625$$

x	82.5	100	∞
$F(x)$	1	y	1

$$F_c\,(100) = y = 1$$

Partie c)

$$f(]-\infty;\ 40[) = f_1 + \frac{40 - 37.5}{b_2 - b_1} f_2 = \frac{3}{400} + \frac{40 - 37.5}{47.5 - 37.5}\ \frac{51}{400} = 0.039375 = F_c\,(40)$$

$$f(]-\infty;\ 60[) = 0.715 = F_c\,(60)$$

$$F_c\,(80) - F_c\,(60) = 0.99625 - 0.715 = 0.28125 = f\,([60;\ 80[)$$

Partie d)

$f([40;\ 60[) = F_c(60) - F_c(40) = 0.715 - 0.039375 = 0.675625$

$f([40;\ 80[) = F_c(80) - F_c(40) = 0.99625 - 0.039375 = 0.956875$

Exercice 2-4

Initialisations

```
x = {62, 52, 63, 65, 56, 55, 42, 42, 60, 41, 57, 59, 59, 55, 45, 44, 59, 40, 56, 61, 63, 59, 48,
   47, 69, 73, 55, 56, 43, 71, 57, 55, 54, 55, 47, 43, 42, 47, 80, 60, 56, 55, 58, 44,
   56, 47, 57, 58, 49, 50, 55, 54, 56, 66, 56, 55, 52, 48, 58, 58, 64, 61, 58, 58, 55,
   66, 53, 61, 55, 56, 58, 61, 62, 52, 51, 55, 49, 46, 56, 51, 61, 55, 54, 54, 64, 50,
   54, 56, 42, 41, 59, 67, 70, 69, 64, 57, 58, 65, 50, 56, 45, 52, 56, 53, 48, 52, 71,
   68, 47, 40, 41, 48, 74, 60, 53, 60, 56, 56, 50, 65, 54, 49, 60, 50, 62, 64, 61, 60,
   53, 54, 63, 56, 59, 54, 49, 59, 59, 53, 61, 58, 60, 55, 54, 48, 56, 67, 57, 54, 58,
   58, 56, 45, 38, 54, 56, 50, 46, 48, 49, 69, 55, 46, 55, 67, 57, 48, 52, 54, 44, 54,
   55, 52, 46, 61, 63, 54, 65, 50, 57, 59, 51, 70, 60, 53, 51, 54, 50, 50, 60, 45, 62,
   58, 53, 76, 53, 60, 60, 51, 47, 63, 64, 53, 61, 65, 50, 48, 54, 63, 62, 59, 48, 43,
   54, 66, 49, 50, 67, 56, 56, 63, 60, 62, 63, 49, 60, 69, 46, 29, 52, 55, 48, 61, 44,
   58, 56, 46, 57, 59, 58, 58, 58, 63, 64, 49, 79, 63, 58, 56, 58, 53, 55, 58, 50, 73,
   63, 52, 58, 52, 62, 67, 54, 54, 67, 58, 46, 60, 58, 58, 47, 63, 57, 57, 48, 54, 56,
   55, 56, 65, 63, 50, 50, 60, 54, 45, 57, 50, 48, 35, 51, 64, 41, 51, 70, 50, 43, 62,
   51, 60, 46, 46, 63, 66, 56, 50, 55, 49, 65, 57, 59, 55, 52, 64, 51, 50, 57, 46, 53,
   47, 49, 66, 54, 60, 52, 41, 60, 62, 65, 54, 54, 57, 57, 56, 38, 61, 51, 51, 60, 49,
   63, 46, 67, 68, 37, 56, 61, 46, 57, 54, 53, 51, 59, 49, 56, 61, 62, 51, 58, 51, 56,
   58, 68, 54, 51, 59, 63, 45, 61, 53, 59, 56, 60, 60, 49, 67, 52, 67, 64, 61, 51, 59,
   54, 62, 59, 61, 52, 63, 56, 56, 45, 57, 61, 46, 47, 57, 54, 48, 64, 65, 47, 62};
```

Partie a)

```
mBrut = N[Mean[x]]
```
valeur moye

55.625

```
mGr = 55.7375
```

55.7375

$$\frac{mGr - mBrut}{mBrut}$$

0.00202247

L'erreur de groupement sur la moyenne est de 0.2 % environ.

Partie b)

```
sBrut = N[StandardDeviationMLE[x]]
```
valeur numérique

7.34026

```
sGr = 8.0847
```

8.0847

$$\frac{sGr - sBrut}{sBrut}$$

0.101419

L'erreur de groupement sur l'écart-type est de 10.1 % environ.

Partie c)

```
meBrut = Median[x]
```
médiane

```
56
```

```
meGr = 55.7143
```

```
55.7143
```

$$\frac{meGr - meBrut}{meBrut}$$

```
-0.00510179
```

L'erreur de groupement sur la médiane est de 0.5 % environ.

Partie d)

$$interQuartBrut = InterpolatedQuantile\left[x, \frac{3}{4}\right] - InterpolatedQuantile\left[x, \frac{1}{4}\right]$$

```
9
```

```
interQuartGr = 10.1528
```

```
10.1528
```

$$\frac{interQuartGr - interQuartBrut}{interQuartBrut}$$

```
0.128089
```

L'erreur de groupement sur l'écart interquartile est de 12.8 % environ.

Corrigé de l'exercice 2 - R

Interprétation des données

Puisque les données sont des effectifs cumulés, il suffit de les diviser par l'effectif total pour obtenir la fréquence cumulée, c'est-à-dire la fonction de distribution empirique.

```
b = {0, 40, 50, 60, 70, 80, 90, 110};

effCum = {0, 4743, 7341, 13688, 28960, 57691, 89909, 100000};

freqCum = effCum / 100000;

afficheTableau[None, {"x", "F(x)"}, Transpose[{b, N[freqCum]}]]
                aucun               transposée  valeur numérique
```

x	$F(x)$
0	0.
40	0.04743
50	0.07341
60	0.13688
70	0.2896
80	0.57691
90	0.89909
110	1.

Partie a)

Fréquence de la classe]*a*; *b*] :

$$f]a;\ b] = F(b) - F(a)$$

```
freq = Drop[freqCum, 1] - Drop[freqCum, -1];
       laisse tomber     laisse tomber
```

Densité de la classe]*a*; *b*] :

$$h]a;\ b] = \frac{f]a;\ b]}{b - a}$$

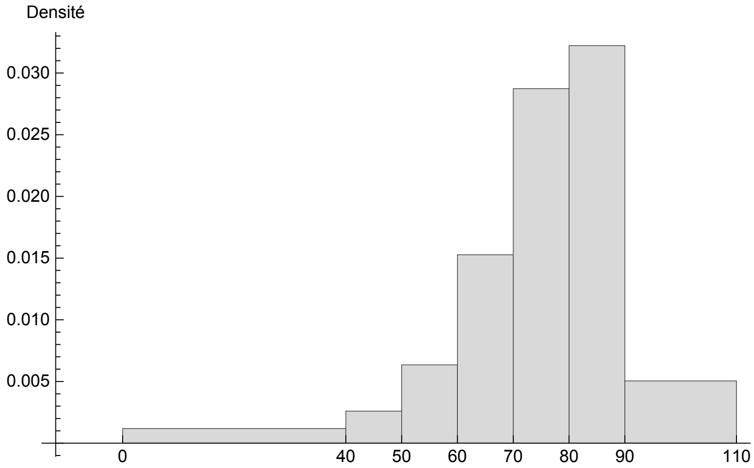
```
h = densites[b, freq];

afficheTableau[None, {"Âge", "Fréquence", "Densité"},
               aucun
 Transpose[{{"[0; 40]", "]40; 50]", "]50; 60]", "]60; 70]",
 transposée
    "]70; 80]", "]80; 90]", "]90; 110]"}, N[freq], N[h]}]]
                                          valeur n…  valeur numé
```

Âge	Fréquence	Densité
[0; 40]	0.04743	0.00118575
]40; 50]	0.02598	0.002598
]50; 60]	0.06347	0.006347
]60; 70]	0.15272	0.015272
]70; 80]	0.28731	0.028731
]80; 90]	0.32218	0.032218
]90; 110]	0.10091	0.0050455

Partie b)

```
histogramme[b, freq, AxesLabel → {None, "Densité"}]
                    titre d'axe   aucun
```



Partie c)

$$f]77;84] = \frac{3}{10} f]77;80] + \frac{4}{10} f]80;90] = \frac{3}{10} 0.28731 + \frac{4}{10} 0.32218 = 0.215065$$

Partie d)

Milieux des classes:

$$c = \frac{\text{Drop}[b, 1] + \text{Drop}[b, -1]}{2}$$

{20, 45, 55, 65, 75, 85, 100}

Moyenne:

```
m = c.freq; N[m]
             valeur
```

74.5599

Ecart-type:

$$s = \sqrt{(c - m)^2 . \text{freq}}; N[s]$$

valeur

17.3957

Partie e)

$$m = c_1 f_1 + c_2 f_2 + ... + c_7 f_7 = 20 * 0.04743 + ... + 100 * 0.10091$$

$$s = \sqrt{(c_1 - m)^2 f_1 + (c_2 - m)^2 f_2 + ... + (c_7 - m)^2 f_7} =$$

$$\sqrt{((20 - 74.5599)^2 * 0.04743 + ... + (100 - 74.5599)^2 * 0.10091)}$$

Partie f)

Repérons la classe où la fréquence cumulée franchit le cap de 0.25 :

```
afficheTableau[{"x", "F(x)"}, None, {{60, "Q", 70}, {0.13688, 0.25, 0.28960}}]
```

aucun

x	60	Q	70
$F(x)$	0.13688	0.25	0.2896

Par interpolation affine:

$$\frac{0.28960 - 0.13688}{70 - 60} = \frac{0.25 - 0.13688}{Q - 60}$$

$$(Q - 60)\ 0.15272 = 10 * 0.11312$$

$$Q = 60 + \frac{1.1312}{0.15272} = 67.407$$

Partie g)

```
mediane = quantileC[b, freq, 1/2]
```

$$\frac{2\,221\,570}{28\,731}$$

```
N[mediane]
```

valeur numérique

77.3231

```
interQuartile = quantileC[b, freq, 3/4] - quantileC[b, freq, 1/4]
```

$$\frac{552\,474\,765}{30\,752\,081}$$

```
N[interQuartile]
```

valeur numérique

17.9654