Marcel Délèze
Edition 2017

Thème : Statistiques I, § 1 Statistique descriptive d'une variable discrète
Lien vers les énoncés des exercices:
https://www.deleze.name/marcel/sec2/applmaths/csud/statistique_1/1-stat_I.pdf

Corrigés des exercices du § 0

Corrigé ex 0 - 1

a) variable continue
(les données sont groupées en classes d'amplitude 1).

b) variable qualitative
(on ne donne pas de critère quantitatif pour classer les oeufs).

c) variable discrète
(s'il existe une classe d'amplitude infinie, la variable n'est pas continue).

d) variable continue
(les données ont été groupées en trois classes).

Corrigé ex 0 - 2

a) variable discrète
(il y a une session d'examens par année).

b) variable continue

Corrigés des exercices du § 1

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 1 - 1

Modalités

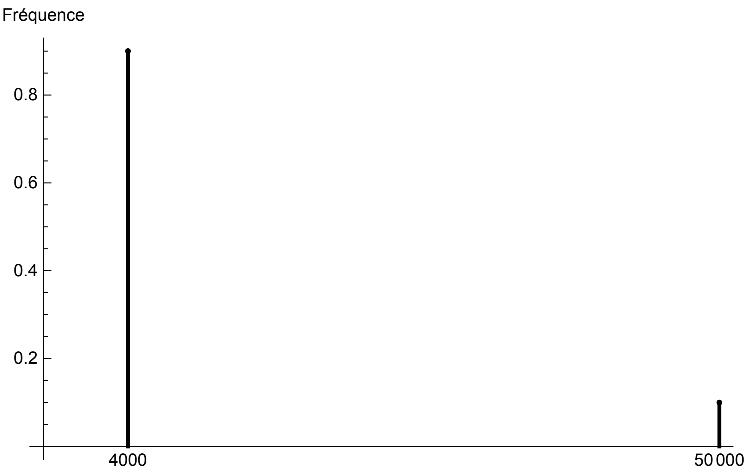
$$c_1 = 4000; \qquad c_2 = 50\,000$$

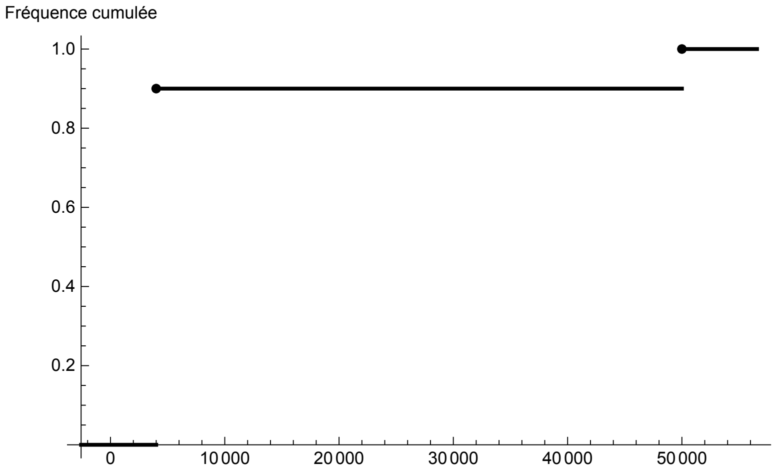
Effectif total

$$n = 9 + 1 = 10$$

Fréquences

$$f_1 = \frac{9}{10}; \qquad f_2 = \frac{1}{10}$$





Quantiles

$Q_{0.25} = 4000;$ $\quad Q_{0.5} = 4000;$ $\quad Q_{0.75} = 4000$

<u>Mesures de tendance centrale</u>

Moyenne

$$m = c_1 f_1 + c_2 f_2 = \frac{9 * 4000 + 1 * 50\,000}{10} = 8600$$

Médiane : classons les 10 personnes par ordre de salaire croissant; le salaire du 5-ème est

$$Me = Q_{0.5} = 4000$$

Mode : le salaire le plus fréquent est

$$Mo = 4000$$

<u>Mesures de dispersion</u>

$$\text{etendue} = 50\,000 - 4000 = 46\,000$$

Ecart-type

$$s = \sqrt{(c_1 - m)^2 f_1 + (c_2 - m)^2 f_2} = \sqrt{(4000 - 8600)^2 \frac{9}{10} + (50\,000 - 8600)^2 \frac{1}{10}} = 13\,800$$

Ecart interquartile

$$\text{interQuart} = Q_{0.75} - Q_{0.25} = 4000 - 4000 = 0$$

<u>Commentaires</u>

La moyenne est sensible aux valeurs extrêmes tandis que la médiane l'est beaucoup moins.
Malgré que beaucoup de personnes aient le même salaire, les écarts à la moyenne sont grands ($s = 13\,800$).
Plus de la moitié des personnes ont le salaire médian (interQuart = 0).

Corrigé de l'exercice 1 - 2

Valeurs

$$c_1 = 1,\ c_2 = 2,\ c_3 = 3,\ c_4 = 4,\ c_5 = 5,\ c_6 = 6,\ c_7 = 7$$

Taille de l'échantillon = effectif total

$$n = 48 + 72 + 96 + 64 + 39 + 25 + 3 = 347$$

Fréquences

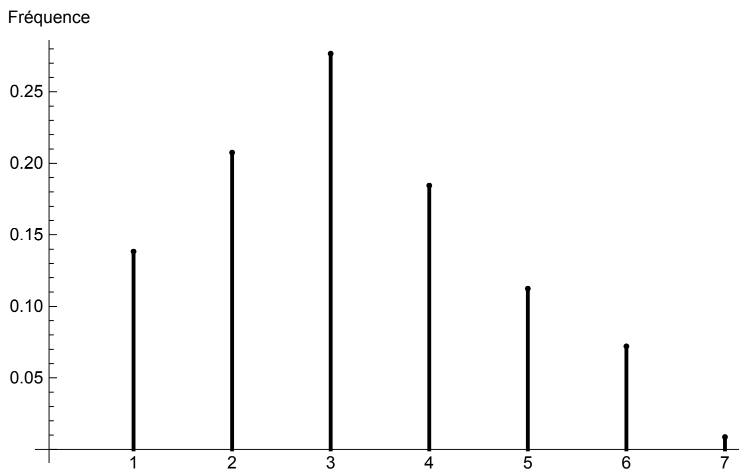
$$f_1 = \frac{48}{347}, \quad f_2 = \frac{72}{347}, \quad f_3 = \frac{96}{347},$$

$$f_4 = \frac{64}{347}, \quad f_5 = \frac{39}{347}, \quad f_6 = \frac{25}{347}, \quad f_7 = \frac{3}{347}$$

$$f_1 \simeq 0.138, \quad f_2 \simeq 0.207, \quad f_3 \simeq 0.277,$$

$$f_4 \simeq 0.184, \quad f_5 \simeq 0.112, \quad f_6 \simeq 0.072, \quad f_7 \simeq 0.009$$

Diagramme à bâtons



Fréquences cumulées

$$F_1 = f_1 = \frac{48}{347} \simeq 0.138$$

$$F_2 = f_1 + f_2 = F_1 + f_2 = \frac{48}{347} + \frac{72}{347} = \frac{120}{347} \simeq 0.346$$

$$F_3 = f_1 + f_2 + f_3 = F_2 + f_3 = \frac{120}{347} + \frac{96}{347} = \frac{216}{347} \simeq 0.622$$

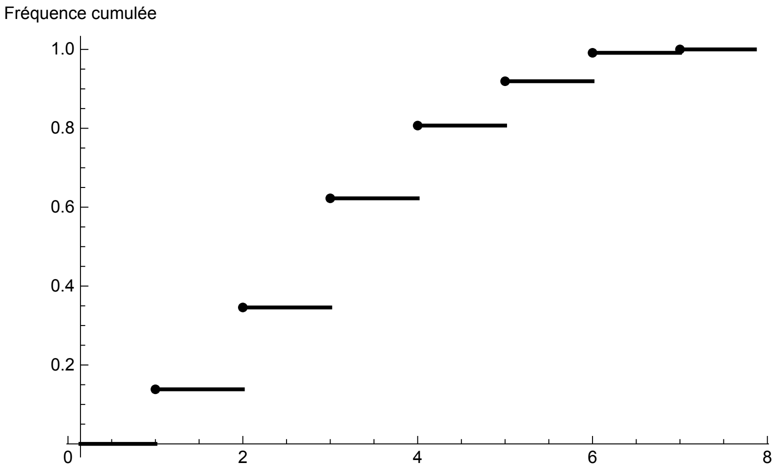
$$F_4 = f_1 + f_2 + f_3 + f_4 = F_3 + f_4 = \frac{216}{347} + \frac{64}{347} = \frac{280}{347} \simeq 0.807$$

$$F_5 = F_4 + f_5 = \frac{280}{347} + \frac{39}{347} = \frac{319}{347} \simeq 0.919$$

$$F_6 = F_5 + f_6 = \frac{319}{347} + \frac{25}{347} = \frac{344}{347} \simeq 0.991$$

$$F_7 = F_6 + f_7 = \frac{344}{347} + \frac{3}{347} = 1$$

Diagramme cumulatif



Moyenne arithmétique

$$m = \sum_{j=1}^{k} c_j f_j = c_1 f_1 + c_2 f_2 + c_3 f_3 + c_4 f_4 + c_5 f_5 + c_6 f_6 + c_7 f_7 =$$

$$1 \frac{48}{347} + 2 \frac{72}{347} + 3 \frac{96}{347} + 4 \frac{64}{347} + 5 \frac{39}{347} + 6 \frac{25}{347} + 7 \frac{3}{347} = \frac{1102}{347} \approx 3.17579$$

Mode (voir le diagramme à bâtons)

$$Mo = 3$$

Médiane (voir le diagramme cumulatif)

$$Me = Q_{\frac{1}{2}} = 3$$

Etendue (voir les valeurs données)

$$7 - 1 = 6$$

Variance (moyenne arithmétique des carrés des écarts à la moyenne)

$$s^2 = \sum_{j=1}^{k} (c_j - m)^2 f_j = (c_1 - m)^2 f_1 + (c_2 - m)^2 f_2 +$$

$$(c_3 - m)^2 f_3 + (c_4 - m)^2 f_4 + (c_5 - m)^2 f_5 + (c_6 - m)^2 f_6 + (c_7 - m)^2 f_7 =$$

$$\left(1 - \frac{1102}{347}\right)^2 \frac{48}{347} + \left(2 - \frac{1102}{347}\right)^2 \frac{72}{347} + \left(3 - \frac{1102}{347}\right)^2 \frac{96}{347} + \left(4 - \frac{1102}{347}\right)^2 \frac{64}{347} +$$

$$\left(5 - \frac{1102}{347}\right)^2 \frac{39}{347} + \left(6 - \frac{1102}{347}\right)^2 \frac{25}{347} + \left(7 - \frac{1102}{347}\right)^2 \frac{3}{347} = \frac{258\,958}{120\,409} \approx 2.15065$$

Ecart-type (moyenne quadratique des écarts à la moyenne = racine carrée de la variance)

$$s = \sqrt{\frac{258\,958}{120\,409}} \approx 1.46651$$

Premier quartile (dans le diagramme cumulatif, valeur qui correspond à la fréquence cumulée $\frac{1}{4}$)

$$Q_{\frac{1}{4}} = 2$$

Troisième quartile (dans le diagramme cumulatif, valeur qui correspond à la fréquence cumulée $\frac{3}{4}$)

$$Q_{\frac{3}{4}} = 4;$$

Intervalle interquartile

$$Q_{\frac{3}{4}} - Q_{\frac{1}{4}} = 2$$

Corrigé de l'exercice 1 - 3. Première version avec données groupées

```
c = Range[1, 7]
    plage
```

```
{1, 2, 3, 4, 5, 6, 7}
```

```
eff = {48, 72, 96, 64, 39, 25, 3}
```

```
{48, 72, 96, 64, 39, 25, 3}
```

```
n = Apply[Plus, eff]
    remp··· plus
```

```
347
```

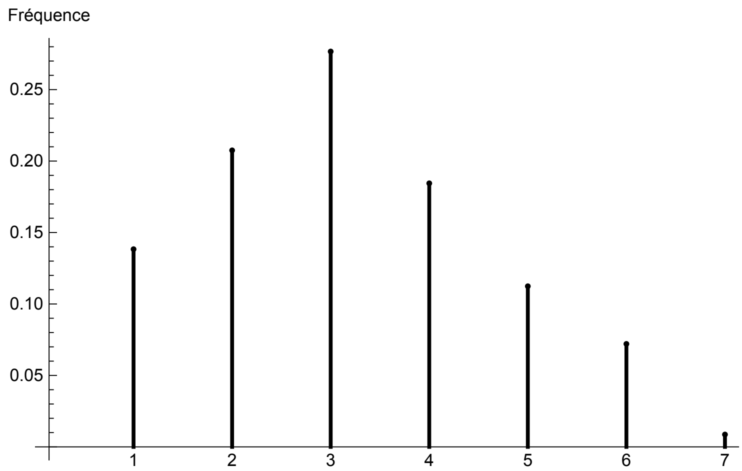
$$freq = \frac{eff}{n}$$

$$\left\{\frac{48}{347}, \frac{72}{347}, \frac{96}{347}, \frac{64}{347}, \frac{39}{347}, \frac{25}{347}, \frac{3}{347}\right\}$$

```
N[freq]
valeur numérique
```

```
{0.138329, 0.207493, 0.276657, 0.184438, 0.112392, 0.0720461, 0.00864553}
```

```
diagrammeBatons[c, freq, AxesLabel → {None, "Fréquence"}]
                         titre d'axe   aucun
```



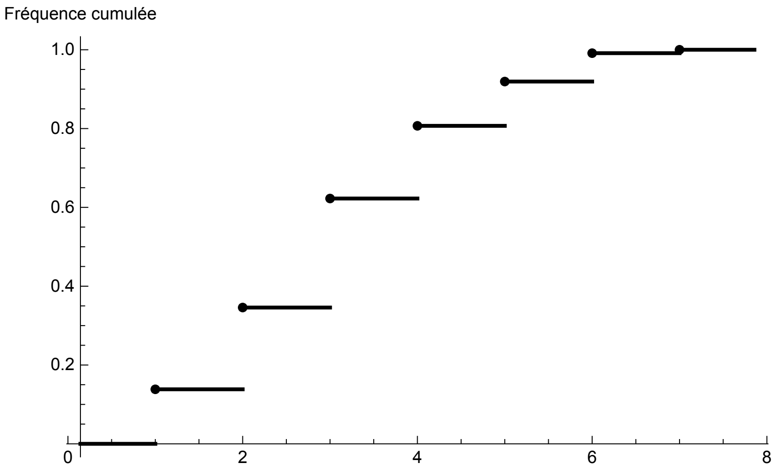

```
freqCumulees = Accumulate[freq]
               accumule
```

$$\left\{\frac{48}{347}, \frac{120}{347}, \frac{216}{347}, \frac{280}{347}, \frac{319}{347}, \frac{344}{347}, 1\right\}$$

```
N[freqCumulees]
valeur numérique
```

```
{0.138329, 0.345821, 0.622478, 0.806916, 0.919308, 0.991354, 1.}
```

```
diagrammeCumulatif[c, freq, AxesLabel → {None, "Fréquence cumulée"}]
```
titre d'axe   aucun



Taille de l'échantillon

```
n
```

347

Moyenne arithmétique

```
m = freq.c
```

$$\frac{1102}{347}$$

```
N[m]
```
valeur numérique

3.17579

<u>Mode</u> : Il s'agit de repérer, dans la liste des fréquences, le(s) rang(s) de celle(s) qui est (sont) maximale(s).

```
Mo = Position[freq, Max[freq]]
```
position   maximum

{{3}}

```
Mo = Mo[[1]][[1]]
```

3

<u>Médiane</u> Il s'agit de repérer, dans la liste des fréquences cumulées, le rang de la première qui atteint ou dépasse $\frac{1}{2}$. Pour ce faire, cherchons la position de la première valeur non négative dans la liste $\left(\text{freqCumulees} - \frac{1}{2}\right)$.

$$\text{Me} = \text{Position}\left[\text{Negative}\left[\text{freqCumulees} - \frac{1}{2}\right], \text{False}\right][[1]][[1]]$$
position   négatif?   faux

3

Étendue

```
etendue = Max[c] - Min[c]
          maximum minimum
```

6

Variance

```
var = freq.(c - m)^2
```

$\frac{258\,958}{120\,409}$

```
N[var]
valeur numérique
```

2.15065

Ecart-type

```
s = √var ;
N[s]
valeur numéric
```

1.46651

Premier quartile Il s'agit de repérer, dans la liste des fréquences cumulées, le rang de la première qui atteint ou dépasse $\frac{1}{4}$. Pour ce faire, cherchons la position de la première valeur non négative dans la liste $\left(\text{freqCumulees} - \frac{1}{4}\right)$.

```
q1 = Position[Negative[freqCumulees - 1/4], False][[1]][[1]]
     position négatif?                      faux
```

2

Troisième quartile Il s'agit de repérer, dans la liste des fréquences cumulées, le rang de la première qui atteint ou dépasse $\frac{3}{4}$. Pour ce faire, cherchons la position de la première valeur non négative dans la liste $\left(\text{freqCumulees} - \frac{3}{4}\right)$.

```
q3 = Position[Negative[freqCumulees - 3/4], False][[1]][[1]]
     position négatif?                      faux
```

4

Intervalle interquartile

```
q3 - q1
```

2

Corrigé de l'exercice 1 - 3.   Deuxième version avec données brutes

```
c = Range[1, 7]
    plage
```

{1, 2, 3, 4, 5, 6, 7}

```
eff = {48, 72, 96, 64, 39, 25, 3}
```

{48, 72, 96, 64, 39, 25, 3}

Construction de l'échantillon

```
x = Flatten[Table[Table[c[[i]], {eff[[i]]}], {i, 1, 7}]]
```
aplatis table table

```
{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7}
```

Taille de l'échantillon

```
n = Length[x]
```
longueur

```
347
```

Moyenne arithmétique

```
m = Mean[x]
```
valeur moy

$$\frac{1102}{347}$$

```
N[m]
```
valeur numérique

```
3.17579
```

Mode

```
Mo = Commonest[x]
```
le plus commun

```
{3}
```

Médiane

$$\texttt{Me = Quantile}\left[x, \frac{1}{2}\right]$$
quantile

```
3
```

Étendue

```
etendue = Max[x] - Min[x]
```
maximum minimum

```
6
```

Variance

```
var = VarianceMLE[x];
N[var]
```
valeur numérique

2.15065

Ecart-type

```
s = StandardDeviationMLE[x];
N[s]
```
valeur numérique

1.46651

Premier quartile

$$q1 = \text{Quantile}\left[x, \frac{1}{4}\right]$$
quantile

2

Troisième quartile

$$q3 = \text{Quantile}\left[x, \frac{3}{4}\right]$$
quantile

4

Intervalle interquartile

```
q3 - q1
```

2

Corrigé de l'exercice 1 - 4

$$V = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_i^2 - 2\, x_i\, m + m^2\right)$$
$$= \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2\right) + \left(\frac{1}{n} \sum_{i=1}^{n} (-2\, x_i\, m)\right) + \left(\frac{1}{n} \sum_{i=1}^{n} m^2\right) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2\right) - 2\, m \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right) + \left(m^2\right)$$
$$= \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2\right) - 2\, m\, (m) + \left(m^2\right) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2\right) - m^2$$