

# Eléments de statistique

## Répétition 1

2 octobre 2018

### Statistique descriptive

L'objet de la statistique descriptive est de décrire un ensemble relativement important de données, pour résumer en quelques caractéristiques l'information qu'elles contiennent.

### Exercices

- 1.1. On a demandé à 50 familles combien d'enfants sont à charge dans le ménage. On a obtenu les données brutes suivantes :

0	2	1	2	1	1	3	2	2	3
1	2	0	1	1	2	0	1	2	1
2	0	2	2	0	3	2	1	2	0
0	2	1	2	2	1	3	0	2	1
2	1	3	0	2	1	2	1	0	2

- Réalisez un tableau résumant les données brutes. Indiquez notamment les effectifs observés, les fréquences observées et les fréquences cumulées observées.
  - Déterminez le pourcentage de familles ayant 1 enfant à charge.
  - Déterminez le pourcentage de familles ayant au maximum 2 enfants à charge.
  - Réalisez le diagramme en bâtonnets des fréquences observées.
  - Réalisez le diagramme en escalier des fréquences cumulées.
  - Calculez la moyenne et la médiane.
  - Calculez la variance et l'écart-type.
  - Peut-on considérer qu'une famille ayant deux enfants à charge est "normale" ? (au sens de la loi normale)
- 1.2. On a observé la durée de vie en heures d'un très grand nombre de transistors venant d'un même fabricant, ce qui a donné les résultats suivants :

Durée observée	Quantité
]1700,1800]	4
]1800,1900]	44
]1900,1950]	40
]1950,2000]	62
]2000,2050]	58
]2050,2100]	46
]2100,2200]	38
]2200,2300]	8

- Réalisez l'histogramme des fréquences relatives non cumulées.
- Réalisez le polygone des fréquences relatives cumulées.
- Montrez et évaluez sur ces graphiques la proportion des transistors dont la durée de vie est comprise entre 1850 et 2100.
- Calculez la durée de vie atteinte ou dépassée par 20% des transistors
- Calculez la moyenne et la médiane.
- Calculez la variance et l'écart-type.
- Quelle est la proportion de transistors ayant une durée de vie "normale" ? (au sens de la loi normale)
- L'aspect du polygone des fréquences non cumulées vous fait sans doute penser à la densité de probabilité d'une loi connue. Laquelle ?

## Exercice suggéré

- 1.3. Dans un village de plus ou moins 300 habitants, le presse régionale a comme gros titre que le poids moyen des habitants du village a augmenté de 2 kg entre 2009 et 2010. Face à une telle hausse, le bourgmestre décide de créer une prime pour aider les gens qui souhaitent commencer la pratique d'un sport d'endurance. Ayant eu accès aux résultats bruts, vous décidez de construire un boxplot des données de 2009 et 2010, dont le résultat est la figure 1.

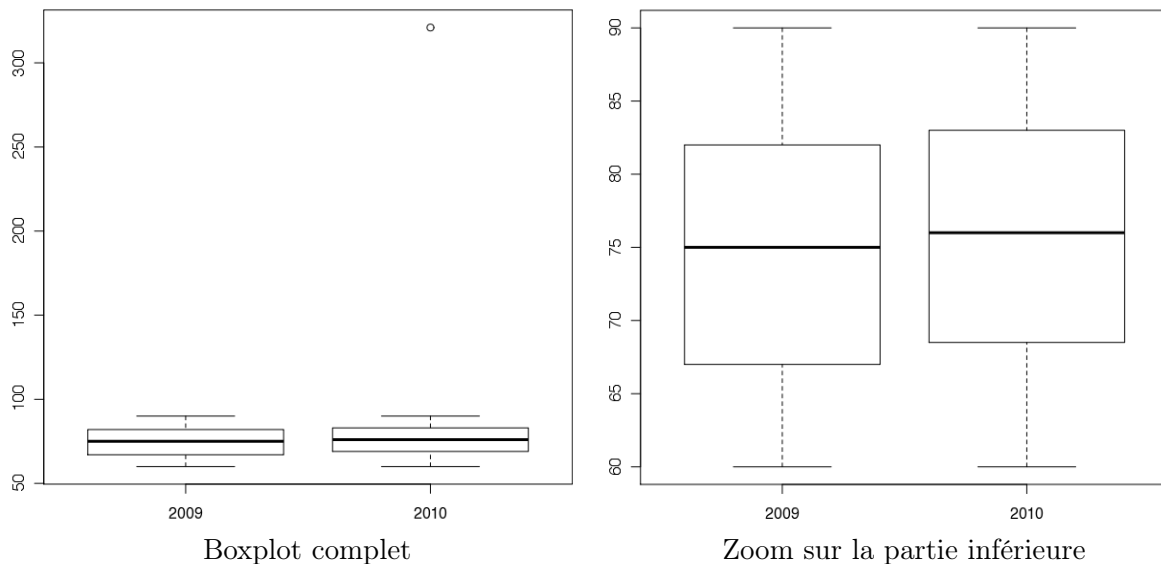
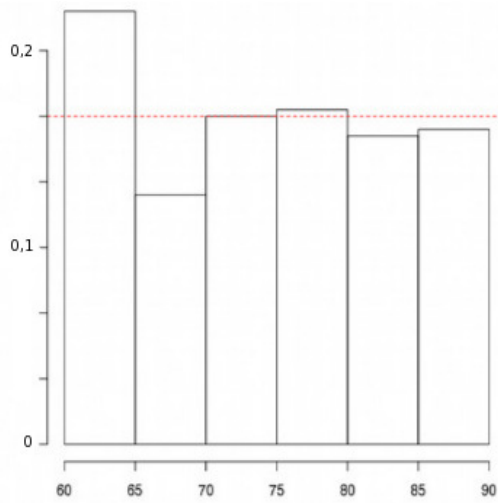
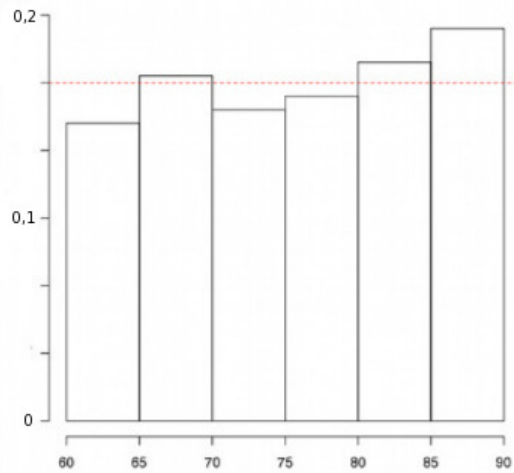


FIGURE 1 – Boxplot des poids des habitants du village en 2009 et 2010.

- Que constatez-vous de particulier sur le boxplot complet ?
- Que représentent les différentes barres ? Peut-on en déduire quelque chose par rapport au changement de poids effectif de la population ?



Année 2009



Année 2010 (sans la donnée aberrante)

FIGURE 2 – Histogrammes des poids des habitants du village

#### 1.4. EXAMEN JANVIER 2014

Une station de ski s'intéresse à l'âge de ses clients. Elle réalise une enquête auprès de 300 skieurs qui la fréquentent et obtient les résultats suivants :

Tranche d'âge	Nombre de répondants
$[0, 10[$	7
$[10, 20[$	45
$[20, 30[$	58
$[30, 40[$	59
$[40, 50[$	52
$[50, 60[$	36
$[60, 70[$	23
$[70, 80[$	14
$[80, 90[$	6

- Réalisez l'histogramme des fréquences relatives non cumulées. (**4 points**)
- Réalisez le polygone des fréquences relatives cumulées. (**4 points**)
- Montrez et évaluez sur ces graphiques la proportion de skieurs ayant un âge compris entre 23 et 45 ans. (**4 points**)
- Calculez la moyenne. (**2 points**)
- Calculez les quartiles. (**3 points**)
- Calculez la variance. (**2 points**)
- Calculez l'écart-type. (**1 point**)

## Solution des exercices suggérés

1.3

- a) Une **donnée abberante** est apparue en 2010. Soit une personne pesant entre 300 et 350 kg a rejoint le village (ça existe!), soit une faute de frappe a eu lieu lors de l'acquisition du poids d'un des habitants, soit autre chose.

Une donnée est abberante si elle est plus petite que

$$Q^1 - 1.5(Q^3 - Q^1)$$

ou plus grande que

$$Q^3 + 1.5(Q^3 - Q^1)$$

Cette donnée abberante a tiré la moyenne vers le haut et l'augmentation hors donnée aberrante du poids moyen des habitants est largement inférieur à 2 kg.

- b) La barre centrale indique le **poids médian**.

2009  $\Rightarrow$  75 kg

2010  $\Rightarrow \pm 76kg$

Les bords des rectangles représentent **les quartiles**.

2009  $\Rightarrow Q^1 = \pm 67kg$  et  $Q^3 = \pm 82$

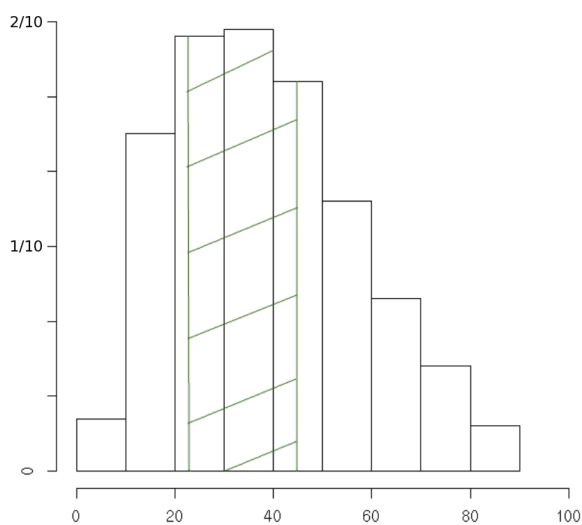
2010  $\Rightarrow Q^1 = \pm 68,5kg$  et  $Q^3 = \pm 83$ .

Les "moustaches" sont **les valeurs max et min** (hors données abberantes).

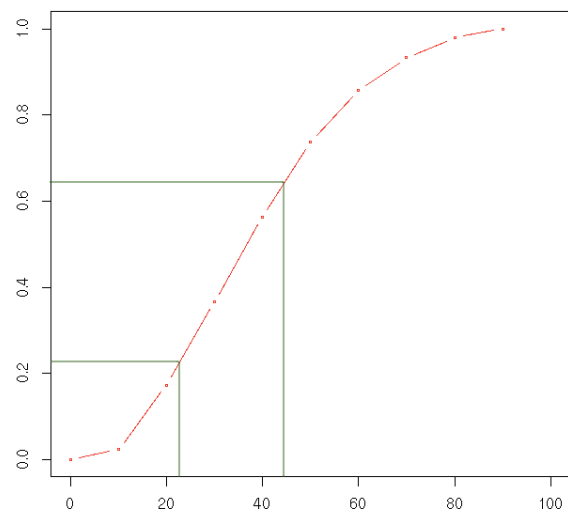
2009 et 2010  $\Rightarrow min = 60kg$  et  $max = 90 kg$ .

**Les boxplots ne nous apprennent rien sur la moyenne.**

1.4. a)



b)



c) 41%.

d) 38,67.

e)  $Q1 = 23,97$ . Médiane = 36,78.  $Q3 = 51,11$ .

f)  $s^2 = 345,9$ .

g) 18,6.