



## Université de Poitiers Département de Mathématiques

Statistique descriptive, 1er semestre, année univ. 2009-2010

Fiche 1

### Statistiques descriptives univariées

#### Exercice 1

- 1) Rappeler ce que l'on entend dans une étude statistique par les termes : population, échantillon, individu, variable.
- 2) Quels différents types de variables distingue-t-on généralement ?
- 3) Dans chacun des exemples suivants, identifier la population, l'échantillon, la variable ainsi que son type.

**Étude 1 :** au cours d'une étude sur le divorce, on s'est procuré auprès de l'INSEE le tableau ci-dessous donnant, pour les couples ayant divorcé moins de 5 ans après leur mariage, le nombre d'années écoulées entre mariage et divorce. Les données ont été obtenues après recensement de la population française.

|             |      |       |       |       |       |
|-------------|------|-------|-------|-------|-------|
| Nb d'années | 1    | 2     | 3     | 4     | 5     |
| Fréquence   | 3,96 | 18,02 | 26,92 | 25,38 | 25,72 |

**Étude 2 :** le Nouvel Observateur publie le sondage suivant donnant l'opinion des français ou des sympathisants PS à la question suivante : qui serait pour vous le meilleur dirigeant politique du Parti Socialiste pour les prochaines années ? Le sondage a été réalisé par téléphone du 20 au 22 août 2008 auprès de 1.029 personnes (échantillon des sympathisants PS : 296 personnes) représentant la population française de 18 ans et plus, selon la méthode des quotas. Les résultats sont résumés dans le tableau suivant :

|                  |        |        |        |        |        |         |        |
|------------------|--------|--------|--------|--------|--------|---------|--------|
| Réponses         | B. Del | S. Roy | M. Aub | P. Mos | J. Dra | J.L Mél | B. Ham |
| Français         | 28     | 18     | 14     | 5      | 4      | 2       | 1      |
| Sympathisants PS | 30     | 24     | 15     | 7      | 3      | 2       | 1      |

Remarque : les pourcentages ne s'additionnent pas à 100 en raison des personnes qui ne se prononcent pas.

**Étude 3 :** le service qualité d'une grande centrale téléphonique procède à une étude qualité et relève pendant une journée les intervalles de temps en minutes séparant deux appels successifs à un standard téléphonique. Voici un extrait de la liste obtenue :

|      |       |      |      |      |      |      |      |      |      |
|------|-------|------|------|------|------|------|------|------|------|
| 2,12 | 5,65  | 1,44 | 1,04 | 0,49 | 0,99 | 0,84 | 3,45 | 0,26 | 6,62 |
| 3,27 | 18,00 | 1,22 | 4,12 | 0,09 | 0,83 | 5,09 | 2,02 | 6,67 | 1,62 |
| 3,99 | 2,27  | 2,17 | 9,13 | 2,05 | 2,28 | 1,89 | 1,15 | 2,13 | 1,17 |

- 4) Pour chacun des cas, tracer la représentation graphique qui vous semble la plus appropriée pour résumer la (les) série(s) statistique(s) considérée(s). Tracer la fonction de répartition des variables quantitatives.

**Exercice 2****Indicateurs de position**

- 1) Soit un échantillon  $\{x_1, x_2, \dots, x_n\}$ . Rappeler les définitions pour sa moyenne (arithmétique), sa variance, son écart type, son coefficient de variation, sa médiane, ses quartiles, son intervalle interquartile, son étendue.
- 2) Calculer toutes ses quantités pour la série statistique présentée dans le troisième exemple de l'exercice 1.
- 3) Tracer le diagramme box-plot de cette série.

**Exercice 3****Propriété de barycentre de la moyenne arithmétique**

Pour une variable quantitative, l'échantillon  $\{x_1, x_2, \dots, x_n\}$  est obtenu. Soit  $\bar{x}$  la moyenne arithmétique de cet échantillon. Montrer que

$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

**Exercice 4**

- 1) Soit l'échantillon

4 ; 5 ; 1 ; 3, 7 ; 2, 3 ; 4 ; 1, 3 ; 2, 7 ; 5 ; 3, 3.

Déterminer la moyenne, la médiane, l'étendue, la variance, l'écart-type, l'écart interquartile. Dessiner l'histogramme avec les classes

$$\{[1; 2, 5), [2, 5; 5]\}.$$

- 2) Le tableau suivant donne le niveau de cholestérol de 1.067 hommes d'âge compris entre 25 et 34 ans.

| Niveau de cholestérol<br>(mg/100 mL) | Nombre de d'hommes |
|--------------------------------------|--------------------|
| 80 – 119                             | 13                 |
| 120 – 159                            | 150                |
| 160 – 199                            | 442                |
| 200 – 239                            | 299                |
| 240 – 279                            | 115                |
| 280 – 319                            | 34                 |
| 320 – 399                            | 14                 |
| Total                                | 1067               |

- 3) Construire l'histogramme avec les classes indiquées dans le tableau.
- 4) Déterminer l'intervalle auquel appartient la médiane.

**Exercice 5****Quantiles, "QQ-plot"**

On souhaite tester graphiquement l'adéquation de l'échantillon de l'exercice 1 étude 3 à une loi exponentielle grâce à la méthode des quantiles.

- 1) Tracer la courbe des quantiles  $\alpha \mapsto Q_x(\alpha)$  de l'échantillon  $x$ . Préciser la médiane, les quartiles, l'intervalle interquartile.
- 2) Calculer la fonction quantile  $\alpha \mapsto Q_x(\alpha)$  pour une loi exponentielle de paramètre  $\lambda > 0$ .
- 3) En supposant l'échantillon en adéquation avec une loi exponentielle, on fait l'hypothèse  $Q_x(\alpha) \approx Q_\lambda(\alpha)$  pour un certain  $\lambda_0 \in \mathbb{R}$ . Quelle est dans ce cas la nature de la courbe paramétrée :

$$\alpha \mapsto (Q_{\lambda_0}(\alpha); Q_x(\alpha))$$

appelée graphique quantile-quantile ou QQ-plot ?

- 4) À partir des données, tracer le graphe QQ-plot de l'échantillon contre la loi  $\text{Exp}(1)$ . L'adéquation avec une loi exponentielle vous paraît-elle vérifiée ? Si oui, estimer graphiquement  $\lambda$ .