

## Statistique descriptive

### TD1

#### Exercice 1.

Décrivez à la fois la population et les observations (unité statistique) ainsi que la ou les variables et leurs natures pour questions d'intérêt suivantes :

- (a) Évaluation de la satisfaction des employés d'une compagnie aérienne.
- (b) Description des notes des élèves d'un devoir.
- (c) Comparaison de deux médicaments qui traitent l'hypertension artérielle.

#### Exercice 2.

Déterminer la nature des variables suivantes :

Temps de trajet pour se rendre au travail, pointure de chaussures, parti politique préféré, prix d'un repas à la cantine, couleur des yeux, sexe, satisfaction client sur, délai de livraison d'un colis, groupe sanguin, nombre de but dans un match de foot, la taille d'un enfant.

#### Exercice 3.

Considérons les résultats des élections en Afrique du Sud en 2014 et 2009

| Party                           | Results 2014 (%) | Results 2009 (%) |
|---------------------------------|------------------|------------------|
| ANC (African National Congress) | 62.15            | 65.90            |
| DA (Democratic Alliance)        | 22.23            | 16.66            |
| EFF (Economic Freedom Fighters) | 6.35             | –                |
| IFP (Inkatha Freedom Party)     | 2.40             | 4.55             |
| COPE (Congress of the People)   | 0.67             | 7.42             |
| Others                          | 6.20             | 5.47             |

- a. Résumez les résultats des élections de 2014 dans un graphique à barres.
- b. Comment compareriez-vous les résultats des élections de 2009 et de 2014 ? Offrez une solution simple qui peut être représentée dans un seul graphique.

#### Exercice 4.

Considérons une variable X décrivant le temps jusqu'à ce que le premier but soit marqué lors des matchs de la Coupe du monde de football 2006. Seuls les matchs avec au moins un but sont pris en compte.

6 ; 24 ; 91 ; 8 ; 4 ; 25 ; 3 ; 83 ; 89 ; 34 ; 25 ; 24 ; 18 ; 6 ; 23 ; 10 ; 28 ; 4 ; 63 ; 6 ; 60 ; 5 ; 40 ; 2 ; 22 ; 26 ; 23 ; 26 ; 44 ; 49 ; 34 ; 2 ; 33 ; 9 ; 16 ; 55 ; 23 ; 13 ; 23 ; 4 ; 8 ; 26 ; 70 ; 4 ; 6 ; 60 ; 23 ; 95 ; 28 ; 49 ; 6 ; 57 ; 33 ; 56 ; 7

- 1. Décrire la population étudiée les unités statistiques la variable et sa nature
- 2. Ecrivez le tableau des fréquences de X en fonction des classes suivantes : [0, 15), [15, 30), [30, 45), [45, 60), [60, 75), [75, 90), [90, 96].
- 3. Tracer l'histogramme pour X
- 4. Tracer le graphique de la fonction cumulative

5. Considérons les données groupées en classes. Supposons maintenant que les valeurs de chaque intervalle sont distribuées uniformément. Déterminer la proportion des matchs dont les premiers buts qui se sont produits
  - a. en première mi-temps, c'est-à-dire pendant les 45 premières minutes,
  - b. dans les 10 dernières minutes ou pendant le temp additionnel,
  - c. entre la 20e et la 65e min
6. Déterminez le moment où, dans 80 % des matchs, le premier but a été marqué à ce moment ou avant

#### Exercice 5.

Un restaurant à pizza a effectué durant un mois 1266 livraisons à domicile. La distribution du nombre de pizza par livraison est donnée dans le tableau suivant :

|       |        |        |        |        |        |        |        |        |        |        |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $x_i$ | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      | 11     |
| $F_i$ | 0.1374 | 0.4092 | 0.6754 | 0.8586 | 0.9408 | 0.9779 | 0.9929 | 0.9961 | 0.9992 | 1.0000 |

1. Décrire la population étudiée les unités statistiques la variable et sa nature
2. Combien y a-t-il de livraisons de plus de 5 pizzas ?
3. Déterminer le pourcentage de livraisons entre 4 et 7 pizzas
4. Calculer les effectifs pour chaque modalité et tracer une représentation graphique appropriée

## Statistique descriptive

### TD2

#### Exercice 1.

Vrai ou faux

1. Pour une série d'observations d'une variable statistique :
  - a) On peut calculer quatre quartiles
  - b) L'intervalle interquartile contient 50 % des observations
  - c) le cinquième décile est égal à la médiane
  - d) 50 % des observations sont supérieures au premier quartile
2. Pour une variable statistique de distribution symétrique :
  - a) la moyenne est égale à la médiane
  - b) 50 % des observations sont supérieures à la moyenne
  - c)  $(Q_3 - Q_1) = 2(Q_2 - Q_1)$
3. Une étude des notes obtenues par deux classes d'une école à un test commun a fourni les résultats suivants :

|          | Effectif | Moyenne | Ecart-type | Médiane |
|----------|----------|---------|------------|---------|
| Classe 1 | 20       | 12      | 4          | 12      |
| Classe 2 | 30       | 10      | 6          | 12      |

- a. la note moyenne des deux classes réunies est égale à 11
- b. l'écart-type des notes des deux classes réunies est égal à 5
- c. la médiane des notes des deux classes réunies est égale à 12
- d. l'écart absolu moyen des notes à la médiane est inférieur ou égal à 4 pour la classe 1

#### Exercice 2.

On a relevé pendant 50 quinzaines successives les niveaux de ventes, exprimés en milliers d'unités de produit, de deux présentations notées G (Gel) et P (Poudre) d'un même produit. Les résultats sont les suivants

| Présentation G       |      |         |         |         |
|----------------------|------|---------|---------|---------|
| Niveau de vente      | < 5  | [5-10[  | [10-12[ | [12-20] |
| Nombre de quinzaines | 5    | 20      | 15      | 10      |
| Présentation P       |      |         |         |         |
| Niveau de vente      | < 10 | [10-12[ | [12-16[ | [16-20] |
| Nombre de quinzaines | 10   | 25      | 10      | 5       |

1. Déterminer les unités statistiques la variable d'étude et sa nature
2. Calculez les moyennes, écarts-types et médianes des niveaux de ventes pour chacune des deux présentations.

3. Sur l'ensemble des points de vente pour toute la période de l'étude, on disposait de 30 % du produit en gel, et de 70 % du produit en poudre. Quel a été le niveau de ventes moyen et l'écart-type pour l'ensemble des deux présentations du produit ?
4. Les niveaux de ventes étant maintenant exprimés en centaines d'unités de produit, donnez les nouvelles valeurs des moyennes, écarts-types et médianes calculées à la 2<sup>ème</sup> question.

### Exercice 3.

Afin d'étudier les disparités de salaires entre hommes et femmes, une enquête a été réalisée auprès du personnel ouvrier d'un secteur industriel. Les résultats concernant les salaires annuels nets en euros sont résumés dans les deux tableaux suivants :

Tableau 1. Hommes

|                          |        |
|--------------------------|--------|
| Effectif                 | 180    |
| Salaire moyen            | 15 400 |
| Écart-type               | 3 620  |
| 1 <sup>er</sup> décile   | 10 950 |
| 1 <sup>er</sup> quartile | 12 750 |
| Médiane                  | 14 800 |
| 3 <sup>e</sup> quartile  | 17 660 |
| 9 <sup>e</sup> décile    | 20 220 |

Tableau Femmes

| Salaire annuel<br>(en milliers d'€) | Nombre d'ouvrières |
|-------------------------------------|--------------------|
| [10 ; 12[                           | 82                 |
| [12 ; 14[                           | 34                 |
| [14 ; 16[                           | 12                 |
| [16 ; 20]                           | $n_4$              |
| Total                               | $N$                |

1. Définir la population étudiée, l'unité statistique, le caractère étudié et sa nature.
2. Proposez pour la distribution du salaire des hommes en précisant les valeurs correspondantes :
  - 2 indicateurs de tendance centrale ;
  - 2 indicateurs de dispersion ;
  - 2 indicateurs de dispersion relative.
3. Sachant que le salaire annuel moyen des femmes enquêtées est égal à 12 000 €, déterminez l'effectif  $n_4$  de la dernière classe de la distribution du salaire des femmes, ainsi que l'effectif total  $N$ . Déterminer l'écart-type.
4. Déterminez l'écart-type et le coefficient de variation de la distribution des femmes.
5. Déterminez le salaire annuel moyen et l'écart-type de l'ensemble des ouvriers hommes et femmes de l'enquête.
6. Pour la distribution des salaires des femmes, calculer les trois quartiles et tracer la boîte à moustache

### Exercice 4.

Le prix d'un Kg de bananes coûte 500UM, celui de pommes coûte 800UM, celui des mangues 300UM et celui des poires coûte 900UM.

1. Quel est le prix moyen d'un Kg d'un panier de fruits (bananes, pommes, mangues, poires)
  - a. Constituée des mêmes quantités de chaque fruit. De quelle moyenne s'agit-il ?
  - b. Constituée des mêmes valeurs pour chaque fruit. De quelle moyenne s'agit-il ?
2. Quel est le prix moyen d'un Kg d'un panier de fruits (bananes, pommes, mangues, poires)

- a. Dont 10%, 20%, 30% et 40% de la quantité sont respectivement des bananes, pommes, mangues et poires. De quelle moyenne s'agit-il ?
- b. Dont 10% , 20%, 30% et 40% de la valeur sont respectivement des bananes, pommes, mangues et poires. De quelle moyenne s'agit-il ?

**Exercice 5.**

Une boutique de mode a résumé ses ventes quotidiennes de chaussettes de créateurs en différents groupes : chaussettes pour hommes, chaussettes pour femmes et chaussettes pour enfants. Malheureusement, les données pour les chaussettes pour hommes ont été perdues. Déterminez les valeurs manquantes

|                          | N   | Moyenne arithmétique | Ecart-type     |
|--------------------------|-----|----------------------|----------------|
| Chaussettes pour femmes  | 45  | 16                   | $\sqrt{6}$     |
| Chaussettes pour hommes  | ?   | ?                    | ?              |
| Chaussettes pour enfants | 20  | 7,5                  | $\sqrt{3}$     |
| Total                    | 100 | 15                   | $\sqrt{19,55}$ |

**Exercice 6.**

Le nombre de membres d'un club était le suivant

| année  | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|--------|------|------|------|------|------|------|
| nombre | 23   | 24   | 27   | 25   | 30   | 28   |

- (a) Quel est le taux de croissance moyen du nombre de membres ?
- (b) D'après les résultats de (a), à combien de membres peut-on s'attendre en 2018 ?

## Statistique descriptive

### TD 3

#### Exercice 1.

Au total, 150 clients sont interrogés sur leur satisfaction vis-à-vis de leurs assurances auto et moto. Les résultats sont résumés ci-dessous

|                 | Satisfait | Non satisfait |
|-----------------|-----------|---------------|
| Voiture essence | 33        | 25            |
| Voiture diesel  | 29        | 31            |
| Moto            | 12        | 20            |

1. Tracer un graphique approprié qui permet de rendre compte de la liaison éventuelle de deux variables
2. Calculer le coefficient du  $\chi^2$  ainsi que le V de Cramèr
3. Combiner les deux modalités voiture essence et voiture diesel en une seule modalité et calculer à nouveau le  $\chi^2$  ainsi que le V de Cramèr

#### Exercice 2.

Il y a eu un grand débat sur l'utilité des limites de vitesse sur les voies publiques. Considérez le tableau suivant qui répertorie les limites de vitesse sur les routes de campagne (en miles/h) et les décès sur les routes (par 100 millions de km) pour différents pays en 1986 lorsque le débat était particulièrement sérieux :

| Pays        | Vitesse limite | Nombre de décès (par 1M km) |
|-------------|----------------|-----------------------------|
| Denmark     | 55             | 4.1                         |
| Japan       | 55             | 4.7                         |
| Canada      | 60             | 4.3                         |
| Netherlands | 60             | 5.1                         |
| Italy       | 75             | 6.1                         |

- (a) Dessinez le nuage de points pour les deux variables que remarque-t-on ?
- (b) Calculez les coefficients de corrélation et interpréter sa valeur
- (c) Quels sont les effets sur les coefficients de corrélation si la limite de vitesse est donnée en km/h plutôt que miles/h (1 mile/h  $\approx$  1,61 km/h) ?
- (d) Considérez une autre observation : la limite de vitesse pour l'Angleterre était de 70 miles/h et le taux de mortalité était de 3,1.
  - (i) Ajoutez cette observation au nuage de points.
  - (ii) Calculer le coefficient de corrélation compte tenu de cette observation supplémentaire

#### Exercice 3.

Le tableau suivant donne la distribution de la variable de la température (en F°) des pizzas livrées par un restaurants qui possède 3 succursales : centre, est et ouest.

| Temperature | Succursale |      |      |
|-------------|------------|------|------|
|             | Centre     | East | West |
| 40-55       | 80         | 22   | 60   |
| 55-60       | 118        | 54   | 102  |
| 60-65       | 137        | 92   | 127  |
| 65-70       | 66         | 114  | 110  |
| 70-90       | 20         | 128  | 36   |

1. Calculer les 3 quartiles pour chacune de succursales
2. Tracer les 3 boîtes à moustaches dans un même graphique et commenter

3. Calculer la température moyenne des pizzas pour les livraisons en provenance de chacune des 3 succursales
4. Calculer les variances inter et intra
5. Calculer le rapport de corrélation.