## Statistique descriptive – $1^{\text{re}}$ année<sup>1</sup>

### FEUILLE DE T.P. 2

**Objectifs.** Fin du T.P.1. Calculs d'indicateurs statistiques avec *STATISTICA*. Tracé de boîtes à moustaches. Graphique et calculs pour certains sous-groupes d'individus.

**Instructions.** Ce T.P. et votre compte-rendu sont à finir d'ici **vendredi 26 octobre**, minuit. Le compte rendu devra être déposé sur *moodle*. N'oubliez pas de convertir votre fichier au format ".pdf". Il devra être nommé :

#### TP2\_Prénom\_Nom.pdf

Si le compte rendu a été fait en binôme, les deux étudiants doivent le déposer sur *moodle* et spécifier **explicitement** dans le texte les deux noms.

Pensez également à (re)lire le document *Quelques conseils pour la rédaction d'un compte rendu de l'étude statistique d'un jeu de données*, et à tenir compte des remarques faites lors de la correction du T.P.1.

Nous reprenons l'étude du jeu de données Recensement 85 commencée lors de la séance précédente.

# 1 Fin du T.P.1. : fonctions de répartition empirique et diagramme en bâtons

Commencer par terminer l'étude graphique amorcée au T.P.1.

En particulier, terminer l'étude proposée aux Sections 4.1 et 4.2, vérifier que vous savez représenter les courbes des fréquences cumulées, pour une variable quantitative continue (à partir des données brutes et groupées), et pour une variable quantitative discrète, ainsi qu'un diagramme en bâtons pour une variable quantitative discrète.

## 2 Analyse graphique de la population selon la catégorie professionnelle (variables qualitatives)

Nous nous intéressons ici aux sous-populations correspondant à chacune des catégories professionnelles suivantes : « Management » (cadres), « Clerical » (employés de bureau) et « Professional » (professions libérales). Pour étudier les différences de tendance pour les variables considérées en fonction de la catégorie professionnelle des individus, nous commencerons par utiliser ici divers graphiques.

Vous pourrez par exemple représenter sur des graphiques juxtaposés et/ou étudier la distribution en effectifs et fréquences de certaines variables qualitatives qui vous semblent significatives, pour chacune des trois sous-populations.

Les questions suivantes donnent un exemple détaillé.

- Calculer le pourcentage de femmes parmi les cadres. On pourra utiliser le bouton Select Cases... de la boîte de dialogue du menu Statistiques>Statistiques Elémentaires>Tables de fréquences, pour préciser que l'on ne conserve que les individus qui ont la modalité « Management », pour la variable « Occupation ».
- 2. Effectuer également un camembert pour représenter cette distribution. On pourra utiliser le bouton Select Cases Filtres de Sélection de la boîte de dialogue du graphique concerné.

<sup>1.</sup> Enseignants responsables des TP : G. Chagny (gaelle.chagny@parisdescartes.fr) et C. Laclau (charlotte.laclau@parisdescartes.fr)

**Remarque :** en utilisant également le bouton *Par Groupes…* des boîtes de dialogue, il est possible de construire les tables de fréquences (ou de tracer un graphique), pour les cadres, les employés de bureau et les professions libérales en même temps : on spécifiera que la variable dite *de classement* est « Occupation » et on modifiera les filtres de sélection activés ci-dessus.

Poursuivre l'étude de la même manière pour les variables qualitatives autre que « Gender ».

D'après votre analyse, vous semble-t-il y avoir une différence, pour les variables étudiées, entre les différentes catégories professionnelles considérées dans la population étudiée? Justifiez votre réponse.

### 3 Indicateurs statistiques du jeu de données Recensement

### 3.1 Analyse de l'ensemble des salariés

L'objectif de cette section est de décrire la population dans son ensemble. Pour chaque variable du fichier, qualitative ou quantitative, calculer les indicateurs statistiques pertinents grâce aux procédures du module *Statistiques descriptives Statistiques>Statistiques Elémentaires*. On pourra sélectionner les indicateurs à calculer dans l'onglet *Avancé* de la boîte de dialogue.

Lorsque cela a un sens, tracer les boîtes à moustaches à l'aide de la procédure correspondante du menu Graphiques>Graph. en 2D. La variable d'intérêt doit être sélectionnée dans la rubrique Var(s) dépendante(s) de la boîte de dialogue du graphique. Représenter la médiane non pas par un point, mais par une ligne dans la boîte à moustaches. Rechercher la signification, dans le logiciel, des points dits "atypiques" et "aberrants".

#### **3.2** Analyse par catégorie professionnelle (variables quantitatives)

Dans cette section, on cherche à comparer les cadres (Management) aux employés de bureau (Clerical) d'une part et aux individus exerçant une profession libérale (Professional) d'autre part en terme de salaire et de niveau d'étude.

Calculer les indicateurs statistiques des variables « Wage » et « Education » pour chaque modalité de la variable Catégorie professionnelle (Occupation).

Pour la variable « Wage » et la variable « Education », tracer également sur un même graphique les boîtes à moustaches pour les différentes catégories professionnelles (« Management », « Clerical » et « Professional ») : la variable d'intérêt est toujours à sélectionner dans la rubrique Var(s) dépendante(s) de la boîte de dialogue du graphique, et la Var. de classement est la variable « Occupation ». Les filtres de sélection permettent de n'afficher les résultats que pour les trois modalités étudiées (pour la variable « Education », tracer aussi les diagrammes en bâtons pour chacune des catégories professionnelles).

### 3.3 Conclusion

Faire un bilan des T.P. 1 et 2, proposer un « profil » de chacune des catégories professionnelles considérées au sein de la population étudiée.