

STATISTIQUE DESCRIPTIVE – 1^{RE} ANNÉE¹

FEUILLE DE T.P. 7.

Objectifs. Etude de la liaison entre deux variables qualitatives, ou deux variables quantitatives.

Instructions. Ce T.P. et votre compte-rendu sont à finir d'ici **dimanche 16 décembre** minuit. Le compte-rendu devra être déposé sur *moodle*. Il devra être nommé :

TP7_Prenom_Nom.pdf

Nous reprenons l'étude du jeu de données *Recensement.sta*. On commencera par créer un nouveau classeur *Statistica*, et par enregistrer à nouveau les données dans ce classeur, comme pour chaque T.P.

1 Lien entre deux variables quantitatives

Nous nous intéressons tout d'abord au lien entre le *salaire* et l'*âge*.

1. Décrire les deux variables d'intérêt. Proposer une étude des distributions marginales.
2. Proposer un graphique type "nuage de points" afin de visualiser le lien entre ces deux variables. Que remarquez-vous ?
3. Créer une nouvelle variable *Log_Wage*, qui va contenir le logarithme des salaires : dans votre feuille de données, ajouter une colonne pour la variable, puis dans le menu *Données*, choisir *Transformations*, et rentrer la formule $V9 = \log(V6)$. Tracer à nouveau le nuage de points entre la variable *Log_Wage* et l'*âge*. Commenter. Dans la suite, nous utiliserons ces deux variables.
4. Il existe des coefficients qui mesurent la corrélation entre des variables quantitatives. On s'intéresse au coefficient de corrélation de Pearson qui mesure la dépendance linéaire entre les variables. Pour des couples d'observations $(x_i, y_i)_{1 \leq i, j \leq n}$, le coefficient de corrélation linéaire a pour expression :

$$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Dans le menu *Statistiques > Statistiques élémentaires*, choisir *Matrices de corrélations*. Dans *1-liste de variables*, sélectionner les **deux** variables d'intérêt, cliquer sur *Synthèse*.

5. Essayer de calculer ces coefficients pour différents couples de variables. Pouvez-vous présenter des graphiques type nuage de points corroborant les coefficients que vous observez ?

1. Enseignant responsable des TP : G.Chagny (gaelle.chagny@parisdescartes.fr) et C.Laclau (charlotte.laclau@parisdescartes.fr)

2 Lien entre deux variables qualitatives - Révisions

Proposer une étude complète de la liaison (éventuelle) entre deux variables qualitatives de votre choix : commencer par décrire les variables, étudier les distributions marginales, jointes et conditionnelles, et terminer par le test du χ^2 .