U.F.R. SPSE Master 1
PMP STA 21 Méthodes statistiques pour l'analyse des données en psychologie 2011-2012

TD1 Analyse descriptive des données
Tests de normalité

Fichier de données

On utilise le fichier de données `salaires.sta`[1], téléchargeable à l'adresse suivante :

`http://coursenligne.u-paris10.fr`,

`UFR SPSE, Niveau M, Méthodes statistiques pour l'analyse des données en psychologie.`

Ce fichier contient les données observées sur un échantillon de 474 employés tirés au sort dans une entreprise canadienne. Les variables étudiées sont les suivantes :

- `salary` : salaire brut actuel, en $/an ;
- `salbegin` : salaire de départ, en $/an ;
- `jobtime` : nombre de mois depuis l'entrée dans l'entreprise ;
- `prevexp` : expérience professionnelle antérieure (nombre de mois de travail avant l'entrée dans l'entreprise) ;
- `educ` : nombre d'années d'étude ;
- `minority` : appartenance à une minorité (0 = Non, 1 = Oui) ;
- `sex` : sexe (0 = Homme, 1 = Femme).

1re partie : démarrage de Statistica, fichier de données

1. Définir la population étudiée. Indiquer quelles sont les variables quantitatives discrètes, quantitatives continues et qualitatives.
2. Démarrer Statistica. Fermer la fenêtre de bienvenue. Si une feuille de données vierge (ou toute autre feuille) s'ouvre automatiquement au démarrage, fermer la feuille.
3. Paramétrer le logiciel pour que le fichier de données et les résultats de toutes les analyses effectuées sur ce fichier de données soient placés automatiquement dans un même `classeur`.
4. Créer un nouveau classeur qui va contenir la feuille de données et tous les résultats des analyses statistiques effectuées.

   Ouvrir le fichier `salaires.sta`. Insérer la feuille de données dans le classeur. Rendre la feuille de données `salaires.sta` active.
5. Affecter les valeurs-texte aux codes numériques des variables `minority` et `sex`.

2e partie : étude descriptive des données pour une variable

1. Pour la variable `appartenance à une minorité` :
   (a) Déterminer la distribution des effectifs (*fréquences*) et la distribution des proportions (*pourcentages*) observées sur l'échantillon. Donner le mode de la distribution.
   (b) Représenter la distribution à l'aide d'un camembert et d'un diagramme en barres.
2. Résumer les données quantitatives en déterminant pour chaque variable, les valeurs minimum et maximum, la moyenne, la variance, l'écart-type et les trois quartiles.

   Relever :
   (a) le salaire actuel moyen des employés de l'échantillon.
   (b) le salaire actuel en dessous duquel se situent 50% des employés de l'échantillon ;
   (c) le salaire actuel au-dessus duquel se situent 25% des employés de l'échantillon ;
   (d) le salaire minimal des employés de l'échantillon.

   Donner une estimation ponctuelle du salaire actuel moyen dans l'entreprise.

1. source : fichier salaire.prn ; http ://www.stat.ucl.ac.be/cours/stat2430/exercices.html

3. Visualisations de la distribution des données pour la variable `salaire actuel` :
   (a) Représenter la distribution des données à l'aide d'un histogramme. Quel est l'intervalle modal ? La distribution est-elle unimodale ? Est-elle symétrique, étalée vers la gauche, vers la droite ?
   (b) Relever les coefficients d'asymétrie et d'aplatissement de la distribution.
   (c) Représenter les données à l'aide d'une boîte à moustaches.
   (d) Au vu des observations, peut-on considérer que la variable `salaire actuel` se distribue selon une loi normale dans la population étudiée ?
4. Tester séparément la normalité de chacune des deux variables `salaire actuel` et `salaire à l'embauche`. On fera les tests au niveau $\alpha = 5\%$.

3e partie : étude descriptive conjointe pour une variable qualitative et une variable quantitative

On veut observer ici l'influence de la variable `sexe` sur la variable `salaire actuel`. Pour cela, on va comparer sur l'échantillon d'employés la distribution des salaires actuels des hommes et la distribution des salaires actuels des femmes.

1. Résumer séparément les salaires actuels des hommes et des femmes : pour chacun des deux groupes, relever l'effectif, la moyenne, l'écart-type et la médiane.
2. Sur un même graphique représenter l'histogramme des salaires actuels des hommes et celui des salaires actuels des femmes.
3. Sur un même graphique représenter les boîtes à moustaches des salaire actuels pour les hommes et pour les femmes.
4. Décrire les différences de distribution des salaires selon le sexe observées sur l'échantillon des salariés de l'entreprise.