

TD2 Statistiques multivariées et Programmation R

Variables qualitatives

- 1) Chargez les données des IRIS parisiens dans R, puis récupérez dans quatre variables distinctes, les taux de familles monoparentales, les taux d'ouvriers, les revenus moyens des IRIS parisiens, les coordonnées X des centroïdes des IRIS parisiens.
- 2) Effectuez un résumé des revenus, puis un résumé des coordonnées X.
- 3) Dans une variable nommée « data », stockez les données attributaires des IRIS parisiens. Puis tout simplement en utilisant un « \$ », créez une nouvelle colonne nommée « Riche » et affectez la valeur « Riche » par défaut à cette nouvelle colonne. Faites de même afin de créer une nouvelle colonne nommée « Ouest » et d'affecter la valeur « Ouest » par défaut à cette nouvelle colonne.
- 4) Effectuez un test permettant d'identifier les revenus inférieurs à 35 000 euros, puis utilisez ce test pour affecter la valeur « Pauvre » aux entités correspondantes dans la colonne « Riche ». Effectuez un deuxième test permettant d'identifier les coordonnées X supérieures à 651 000, puis utilisez ce test pour affecter la valeur « Est » aux entités correspondantes dans la colonne « Ouest ».
- 5) A l'aide de la fonction « table », créez un tableau de contingence issu du croisement des modalités de la variable « Riche » et de la variable « Ouest ». Puis en utilisant la fonction « prop.table » sur ce tableau de contingence, réalisez un profil en long et un profil en colonne.
- 6) Utilisez la fonction « mosaicplot » sur le tableau de contingence. Puis mettez à vrai l'argument « shade » de cette fonction.
- 7) A l'aide de la fonction « chisq.test », effectuez un test du chi-2 entre la variable « Riche » et la variable « Ouest ».
- 8) Récupérez les valeurs attendues correspondant au cas d'indépendance des deux variables qualitatives, ainsi que la statistique du chi-2 afin de calculer la contribution de chaque case du tableau de contingence à la statistique du chi-2 : $(\text{observé} - \text{attendu})^2 / \text{attendu}$ / statistique du chi-2 .
- 9) A l'aide de la fonction « tapply », comparez les moyennes des revenus moyens des IRIS de l'ouest parisien et de l'est parisien. Puis complétez cela, en utilisant la fonction « boxplot ».
- 10) Créez un modèle linéaire entre les revenus moyens des IRIS parisiens et la variable qualitative « ouest », puis appliquez la fonction « anova » sur ce modèle.
- 11) A l'aide de l'argument « type » de la fonction « plot » que vous allez mettre à « n », puis en utilisant la fonction « points », représentez en rouge dans un nuage de points les valeurs des revenus moyens des IRIS de l'ouest parisien en fonction des taux d'ouvriers. Pour les valeurs de l'est parisien représentez les en noir.
- 12) Appliquez la fonction « anova » à un modèle linéaire tentant d'expliquer les revenus par la variable qualitative « Ouest » et la variable quantitative « Tx_Ouvrier » pour effectuer une Ancova. Puis, toujours avec la fonction « anova », comparez les deux modèles étudiés.
- 13) Désormais prenez en compte dans le modèle précédent l'interaction entre la variable qualitative « Ouest » et la variable quantitative « Tx_Ouvrier ». Puis comparez, ces deux modèles.
- 14) Récupérez les coefficients de l'Ancova du dernier modèle à l'aide de la fonction « coef ». Puis représentez les régressions linéaires correspondantes sur le graphique. Calculez les moyennes ajustées.