

Module : R for data analysis - IASM1EL
Exercice : Exercices
Durée : 2h

CONSIGNE :

Rappel : vous pouvez consulter vos supports de cours, et utiliser toutes les fonctionnalités du logiciel R et toutes les librairies.

Vous est fourni une base de données, au format CSV et quelques informations sur la source et la construction de cette base de données.

Vous devez tout au long de ce contrôle :

- construire une problématique d'analyse autour de cette base de données.
- mettre en œuvre des techniques d'analyse quantitatives.
- construire votre argumentation, en mobilisant les techniques quantitatives utilisées en R.

Il est possible d'aller faire des recherches externes sur le même sujet afin de nourrir votre argumentation. Pour ce partiel, vous serez testé sur les compétences suivantes :

- Étude de différentes variables catégorielles : fréquences, proportions, graphiques.
- Étude de différentes variables quantitatives : statistiques descriptives, graphiques.
- Étude d'associations entre les variables catégorielles.
- Étude d'associations entre les variables quantitatives et les variables catégorielles.
- Réalisation des tests pour comparer par paires des variables.
- Effectuer un ANOVA, un KHI2 etc.
- Effectuer une régression multiple

Merci de déposer votre travail sur classroom.

1 STATISTIQUES DESCRIPTIVES À UNE VARIABLE

Consigne générale : pour tous ces exercices on mémorisera l'ensemble des commandes dans un fichier «.R» et on commentera l'ensemble des réponses.

1.1 LES DONNÉES :

Importer les données du fichier « pole_emploi_bmo_2019_metier_secteur.csv » à l'aide de la commande `read.csv2()`. Pour être sûr du répertoire courant vous pouvez utiliser « `getwd()` » ou `setwd` dans la console.

1.2 CARACTÈRES QUALITATIFS :

Étudier de façon complète (graphiques et calculs) l'ensemble des caractères qualitatifs et interprétez vos résultats.

1.3 CARACTÈRE QUANTITATIF :

Étudier de façon complète (graphiques et calculs des paramètres de position et dispersion) les caractères quantitatifs de la base de données».

2. STATISTIQUES DESCRIPTIVES À DEUX VARIABLES

Consigne générale : pour tous ces exercices on mémorisera l'ensemble des commandes dans un fichier «.R»

2.1 REGRESSION MULTIPLE :

Effectuer la régression linéaire multiple du caractère « Nombre de projets saisonniers » (toujours extrait du précédent .csv) en fonction des autres variables » On demande donc le calcul du coefficient de corrélation, l'équation de régression, le nuage de points avec et sans l'ajout de la droite. Interprétez vos résultats.

3. STATISTIQUES INFÉRENTIELLES

3.1 INTERVALLE DE CONFIANCE :

Donner un intervalle de confiance de la proportion de « Nombre de projets saisonniers» aux niveaux 95%.

3.2 TEST DU KHI-DEUX :

Tester l'indépendance des caractères qualitatifs présents dans la base de données.

3.3 TEST ANOVA :

Faire un Test d'indépendance ANOVA de deux caractères que vous aurez choisies (toujours extrait du.csv)

3.4 TEST DE NORMALITÉ :

Identifier les variables quantitatives qui suivraient une loi normale ou d'autres loi (poisson, exponentielle etc).

3.5 TEST DE CONFORMITÉ : BONUS :

Peut-on dans ce dataset, faire un test de conformité ? Si oui lequel ?