12/21/2022

Enzo Thiebaud, Marc Yrius

Travaux Pratique 2022

Statistique

Table of Contents

[Exercice 1 Statistiques 2](#_Toc122382024)

[A. Informations générales 2](#_Toc122382025)

[B. Intervalles de confiance pour les probabilités de dépasser un seuil de Pollution aux particules 3](#_Toc122382026)

[C. Une Régression Linéaire 3](#_Toc122382027)

[Exercice 2 Simulation 4](#_Toc122382028)

## Exercice 1 Statistiques

### Informations générales

1. Quel est le type de variable statistique de chacune des variables (nominale, ordinale,  
   quantitative discrète ou continue) ?

D’après la fonction « dtypes » toutes les variables seraient de type quantitatif continue sauf les variables ‘Clermont-Fd A71 Dioxyde d'azote’, ‘Clermont-Fd A71 Monoxyde d'azote’ qui seraient des variables quantitatives continues et la variable ‘Date’ qui serai un type ‘object’. Mais nous pensons que toutes les variables sont de type quantitatif continue sauf la date qui est de type temporel.

1. Quel est le nombre de jours d’observation de l’échantillon ? Quel est le nombre de jours où  
   les particules PM10 sont mesurées à toutes les stations ?

Le premier échantillon réaliser a été fait le 01/01/2015 et le dernier échantillon réaliser a été fait le 31/12/2016 soit il y a eu des mesures pendant 730 jours.

On trouve que les particules de PM10 ont été mesurées à toutes les stations sur 598 jours.

1. Créez deux nouvelles variables booléennes PMObs, AzoteObs ayant, chaque jour, pour  
   valeur « VRAI » respectivement si toutes les mesures de particules ont été observées ce  
   jour-ci ou si tous les autres polluants ont été observés. Donnez la table de contingence de  
   ces deux variables. Quel est le nombre de jours où tous les polluants sont observés ?

Graphical user interface, text, application

Description automatically generatedTable de Contingence :

Il y a eu 532 jours au tous les polluants ont été observer.

1. En ignorant les jours de non-observation des particules (en utilisant par exemple la fonction (dropna), trouvez la moyenne empirique, variance empirique, variance empirique non biaisée et le quartile à 25 % de la variable mesurant les particules PM10 sur l’A7 à Lyon.

La moyenne empirique = 31.009

La variance empirique = 242.936

La variance empirique non biaisée = 243.294

Le quartile 25% = 20.000

### Intervalles de confiance pour les probabilités de dépasser un seuil de Pollution aux particules

1. Créer deux variables valant 0 ou 1 telles que :

— la première vaut 1 si et seulement si le seuil d’information pour les particules PM10 de 50 microgrammes par mètre cube d’air est dépassé sur l’A7 à Lyon,

— la seconde vaut 1 si et seulement si le seuil d’alerte pour les particules PM10 de 80 microgrammes par mètre cube d’air est dépassé sur l’A7 à Lyon. Quel est la loi théorique de ces variables ?

La loi de ces variables est la loi binomiale.

1. Avec la fonction st. binomtest(). proportion\_ci(), trouver un intervalle de confiance exact pour les probabilités p que le seuil d’information pour les particules PM10 soit dépassé sur l’A7 à Lyon, séparément pour un jour typique des années 2015 et 2016. Même question pour le seuil d’alerte.

L’intervalle de confiance exact pour le seuil d’information est p [0.077, 0.121]

L’intervalle de confiance exact pour le seuil d’Alerte est p [0.007, 0.026]

1. Depuis le premier janvier 2005, le seuil d’information pour les particules PM10 ne doit pas être dépassé plus de 35 jours par an, soit une probabilité journalière maximum de dépassement de p0 = 35/365. Un riverain de l’A7 veut tester s’il y a un risque que ce seuil ne soit pas respecté. Tester pour lui l’hypothèse alternative que ce seuil a été respecté, soit p à p0, séparément pour les années 2015 et 2016.
2. Asd

### Une Régression Linéaire

1. Dans la suite, on considère ensemble toutes les particules (on ajoute les chiffres pour les particules PM10 et PM2.5) et tous les oxydes d’azote ensemble (on ajoute les chiffres du dioxyde et du monoxyde). Créez un data.frame dfs contenant les 7 mesures obtenues ainsi (on ne garde les mesures que si les deux mesures sont disponibles) et oubliant tous les jours où il manque un des mesures concernées. Dans la suite, on travaille sur les données du data.frame dfs.

Graphical user interface, application

Description automatically generatedCalculer la matrice de covariance de dfs.

1. On cherche à savoir si la pollution en Oxydes d’Azote à Grenoble y est significativement corrélée à celle sur l’A7 à Lyon x. Effectuez un test de corrélation entre y’ = ln(y) et x’=ln(x). Quelle hypothèse doit-être vérifiée pour effectuer ce test de corrélation (de Pearson) ? Que concluez-vous ?

Tout d’abord le test de Pearson correspond à un indice qui indique une relation linéaire entre deux variables continues donc si le coefficient est entre 0 exclu et 1 il y une corrélation entre les deux variables si le coefficient est égal a 0 la corrélation est nulle sinon il n’y pas de corrélation entre les deux variables.

Cor de x’ et y’ est de 0.699

On peut donc alors conclure qu’il y a bien une corrélation entre les variables x’ et y’

## Exercice 2 Simulation

1. En comparant un histogramme de S et une densité bien choisie, émettez une hypothèse sur la loi de S.
2. F
3. F
4. F
5. F
6. F