

PROJET EN STATISTIQUE DESCRIPTIVE

Pour le vendredi 18 mars 2022

Description du projet

Un client (qui ne parle pas l'anglais) vous demande de faire une analyse descriptive du jeu de données `meteo-3MIC.txt` disponible sur la page Moodle de l'UF. Ce jeu de données comprend 688 observations météorologiques enregistrées par Météo France¹.

Sur le github de Météo France, on retrouve les descriptions (en anglais) suivantes des variables :

- Observed meteorological parameters during **the current day** :
 - `month` : the month of the current day ;
 - `season` : the season of the current day ;
 - `ff` : the wind speed (in $m.s^{-1}$) ;
 - `t` : the temperature (in Kelvin K) ;
 - `td` : the dew point (in K) ;
 - `hu` : the humidity (in %) ;
 - `precip` : the total amount of precipitation (in $kg.m^{-2}$) ;
 - `wind_dir` : the direction of the wind (1 for north, 2 for west, 3 for south and 4 for east) ;
- Forecasts of meteorological parameters for **the next day** by the *Météo France* AROME model :
 - `ws_arome` : the wind speed (in $m.s^{-1}$) ;
 - `t2m_arome` : the temperature at the vertical level of $2m$ (in K) ;
 - `d2m_arome` : the dew point at the vertical level of $2m$ (in K) ;
 - `r_arome` : the humidity (in %) ;
 - `tp_arome` : the total amount of precipitation in ($kg.m^{-2}$) ;
 - `msl_arome` : the sea level pressure (in Pa) ;

Dans un premier temps, commencez par décrire l'ensemble du jeu de données au client, en précisant bien la nature de chaque variable.

Dans un deuxième temps, menez des analyses uni- et bi-dimensionnelles du jeu de données. Observez-vous des anomalies ? Certaines variables sont-elles liées ? Une attention particulière sera portée sur le choix des représentations, et sur l'interprétation des résultats présentés.

Enfin, menez une analyse en composantes principales (ACP) sur les variables qui vous semblent pertinentes. En particulier, précisez bien, en argumentant, le type d'ACP que vous faites et définissez la matrice de travail correspondante (en posant bien toutes vos notations).

Expliquez brièvement au client le principe de l'ACP, et précisez combien de composantes principales vous lui conseillez de garder.

Enfin, pour chaque graphe, précisez ce qui est représenté et interprétez les résultats.

Remarques : Gardez en tête qu'un des objectifs principaux de la statistique descriptive est de synthétiser l'information. De plus, ne mettez en aucun cas des sorties R sans commentaire : si vous n'interprétez pas les résultats, autant ne pas les afficher.

1. Ce jeu de données a été extrait des données du Défi IA 2022.

Consignes

Vous rendrez un rapport **par trinôme** (du même groupe) au format **pdf** obtenu grâce à R Markdown, de 15 pages maximum (avec les graphes). Ce rapport doit être intitulé **gpX-Nom1-Nom2-Nom3.pdf** où **X** est à remplacer par votre groupe (A ou B). Pensez à laisser toutes les commandes R visibles dans votre rapport (option `echo = TRUE` dans les balises R).

Le rapport est à déposer **sur Moodle le vendredi 18 mars 2022 à 20h00** au plus tard (aucun retour mail ne sera accepté).

Modalités d'évaluation

Vous serez évalués sur la présentation et la rédaction du rapport, sur la pertinence des choix des représentations (à argumenter) ainsi que sur l'interprétation des différentes sorties obtenues (graphiques ou autres). Plus précisément, vous serez évalués sur les compétences suivantes.

Compétences transversales

- Savoir mener un argumentaire clair et concis.
- Savoir mener l'étude d'un jeu de données grâce au logiciel R.

Partie 1 : Statistiques descriptives unidimensionnelle et bidimensionnelle

- Savoir identifier la nature des variables.
- Maîtriser les définitions des indicateurs usuels de statistique descriptive.
- Savoir choisir les indicateurs et représentations adaptés aux données.
- Savoir mener une interprétation des graphiques usuels.

Partie 2 : Analyse en composantes principales (ACP)

- Maîtriser le vocabulaire de l'ACP.
- Maîtriser les spécificités de l'ACP centrée et l'ACP centrée réduite.
- Maîtriser le principe de l'ACP.
- Maîtriser la définition des graphiques issus de l'ACP.
- Savoir mener une interprétation des graphiques issus de l'ACP.