

Logiciels Spécialisés 3 – Feuille TP 5

STID 2-ième année (2019/2020)

Compléments R

Partie I – Création d'un fichier HTML

Sauvegarder dans un dossier les fichiers suivants : Rapport_TP.html, R2HTML.css ainsi que les 4 figures, présents sur le portail. Puis ouvrir le fichier HTML à l'aide d'un navigateur.

1- A partir du fichier de données « decathlon_2016.txt », écrire le script R qui permet d'obtenir, à l'aide du package R2HTML, le rapport « Rapport_TP.html » (Les figures sont évidemment à refaire dans le script R).

2- Certaines figures ne sont pas très réussies. Comment pourrait-on les modifier ?

Partie II – Connexion à une base de données

1- Récupérer sur le portail la base de données « accidents ». Cette base comporte plusieurs tables dont la principale est la table MAccident.

2- A partir du panneau de configuration, créer un identifiant ODBC associant le fichier Access et le pilote Access (voir cours).

3- Utiliser la librairie RODBC pour récupérer les tables MAccident et MDate, cette dernière apportant des informations sur la date des accidents. Afficher les 6 premières lignes des deux tables.

4- En utilisant la fonction **sqldf** de la librairie sqldf, effectuer la jointure des tables MAccident et MDate, puis nommer le dataframe jointure1, puis afficher les 6 premiers accidents.

5- On veut étudier l'influence de la luminosité sur les accidents.

a) Récupérer la table MLuminosite, puis vérifier qu'elle permet 2 niveaux d'étude :

libelle_luminosite : 2 niveaux

libelle : 5 niveaux

b) En utilisant la librairie sqldf, effectuer la jointure des tables jointure1 et MLuminosite, puis nommer le dataframe jointure2, puis afficher les 6 premiers accidents.

c) Calculer le nombre d'accidents par année et par libelle_luminosite. Y-a-t'il un lien entre ces deux variables ?

d) Etudier d'autres liens.

Partie III – Données multivariées et représentation graphique (PARTIE FACULTATIVE)

On s'intéresse à la loi normale multidimensionnelle (ici, en dimension 2). Cette loi est paramétrée par un vecteur μ de dimension 2 (le point central de la distribution) et une matrice Σ qui est la matrice de variance-covariance (matrice 2x2). La fonction de densité de cette loi est définie par :

$$f(x) = \frac{1}{2\pi \cdot |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right),$$

où $|\cdot|$ représente le déterminant, et $x=(x_1, x_2)$ est un point de \mathbb{R}^2 . On travaillera pour commencer sur la loi normale multidimensionnelle (ou loi multinormale) de paramètres $\mu=(0,0)$ et $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

- 1- Calculer la valeur de la densité au point de coordonnée $x=(1,0)$.
- 2- Représenter le graphe (on utilisera la fonction plot) de la fonction f lorsque la première composante x_1 varie de -4 à 4 et que la deuxième est fixée à 0 ($x_2=0$). Rajouter sur le même graphique les graphes de f lorsque $x_2=1$ et $x_2=2$.
- 3- Représenter, en trois dimensions, la fonction de densité de la loi multinormale de paramètres μ et Σ .
- 4- Représenter la densité de la loi multinormale par lignes de niveau.
- 5- Charger le package MASS. La fonction « mvrnorm » (consulter l'aide) permet d'obtenir des réalisations de la loi multinormale. Générer 1000 réalisations de la loi multinormale de paramètres μ et Σ , et représenter ces réalisations dans un nuage de points.
- 6- Répéter les étapes précédentes en faisant varier les valeurs de μ et Σ . A quoi ressemblent les lignes de niveaux de la densité ?

Pour les questions 7- et 8-, on repart sur les paramètres $\mu=(0,0)$ et $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

- 7- Générer un grand nombre de réalisations ($N=10000$) de la loi multinormale et tracer un nuage de points. Sur le graphique précédent, rajouter la ligne de niveau $l=0.008$.
- 8- Vérifier que la proportion de points à l'intérieur de la ligne de niveau est d'environ 95%.
- 9- Comment peut-on obtenir, approximativement, la ligne de niveau qui correspond à une « région de dispersion » de 95% pour la loi multinormale $\Sigma = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$?