

Organisation des TP du module "Introduction aux statistiques avec R"

1. Séance 1 (3h) : introduction à R , statistique descriptives avec R , tests avec R
2. Séance 2 (3h) : tests avec R (tests du chi-deux, test d'égalité d'espérances, etc....). Modèle ANOVA à un facteur.
3. Séance 3 (1,5h) : ANOVA et régression linéaire
4. Séance 4 (3h) : Régression linéaire et modèle ANOVA à deux facteurs.

Evaluation : Projet noté ou lecture de sorties R lors de l'examen de "modélisation statistique".

M1- MINT / M1 MA

TP 1 : Modèle linéaire : le modèle d'analyse de la variance à un facteur

Contexte : Etude du diamètre des arbres en fonction de leur statut de dominance.

On s'intéresse au diamètre des arbres dans une forêt en fonction de leur statut de dominance. On considère qu'il y a trois types d'arbres, les dominants, les co-dominants et les dominés. Les données sont dans le fichier *Arbres.txt*. La première colonne représente les coordonnées géographiques de l'arbre, la troisième colonne le diamètre et la quatrième le statut de dominance. Pour la lecture des données, on utilisera la commande *R* :

```
dat=read.table('Arbres.txt')
colnames(dat)=c('X','Y','Diametre', 'Statut')
```

1. Préciser la variable "expliquée", la variable explicative ainsi que leur type (qualitative, quantitative,...)
2. Représentez sous forme de box-plot le diamètre des arbres en fonction du statut de dominance.
3. Calculez le diamètre moyen et la médiane en fonction du statut.
4. Calculez la variance du diamètre en fonction du statut.
5. Emettez des hypothèses sur la base de ces graphes.
6. Comparez avec un test statistique le diamètre des arbres dominants et dominés.
7. Même question pour les arbres codominant et dominés.

On souhaite comparer les trois types d'arbres simultanément. Le modèle le plus simple pour cela est un modèle d'analyse de la variance à un facteur (ANOVA 1).

8. Ecrire le modèle correspondant
9. Mettre en oeuvre ce modèle d'ANOVA 1 avec R.
10. Comparez ce modèle au modèle vide avec la fonction *anova*. Vous décrierez le tableau obtenu. Que représente la colonne F-value. Comment est-elle calculée à partir des colonnes précédentes.
11. A l'aide du tableau précédent conclure sur l'effet du statut de l'arbre sur son diamètre.
12. Faites un graphe pour représenter la dispersion des arbres dans l'espace.
13. Tracer le graphe du diamètre des arbres en fonction de leur abscisse et de leur ordonnée. Que constatez-vous ?

M1- MINT / M1 MA

TP 2 : Modèle linéaire : modle de régression linéaire

Contexte : prédiction de la hauteur d'un arbre à partir de sa circonférence

Lorsqu'un forestier évalue la vigueur d'une forêt, il considère souvent la hauteur des arbres qui la compose. Plus les arbres sont hauts, plus la forêt ou la plantation produit. Si l'on cherche à quantifier la production par le volume de bois il est nécessaire d'avoir la hauteur de l'arbre pour calculer le volume du bois grce à une formule du type "tronc-cône". Cependant la mesure de la hauteur d'un arbre d'une vingtaine de mètres n'est pas aisée. Il est alors nécessaire d'estimer la hauteur grâce à une mesure simple, la mesure de la circonférence à 1 mètre 30 du sol. Les données sont consituées de $n = 1429$ mesures couples circonférences-hauteur, mesures obtenues sur une parcelle d'eucalyptus agés de 6 ans (age de rotation avant la coupe). Ces données sont dans un fichier *eucalyptus.txt*. Nous souhaitons donc trouver la relation qui lie la circonférence à la hauteur de façon à prédire la hauteur d'un arbre à partir de sa circonférence.

1. Importer les données à l'aide de la commande *read.table*
2. Préciser la variable "expliquée", la variable explicative ainsi que leur type (qualitative, quantitative,...)
3. Tracer le nuage de point de la hauteur en fonction de la circonférence.
4. On propose dans un premier temps de modélisation la relation entre hauteur et circonférence par un modèle de régression linéaire simple. Rappeler le modèle est les hypothèses associées.
5. En utilisant *R*, calculer les estimateurs des coefficients de la droite.
6. Quels sont les tests mis en oeuvre par le logiciel *R*.
7. Tracer sur un même graphe les valeurs prédites par le modèle linéaire et les valeurs observées. Commenter.
8. On cherche maintenant à modéliser la relation par un modèle linéaire reliant la racine carrée de la circonférence à la hauteur. Expliciter le modèle et le mettre en oeuvre avec *R*.
9. On considère maintenant un modèle linéaire faisant intervenir deux variables, la racine carrée de la circonférence à la hauteur et la circonférence elle même. Expliciter le modèle et le mettre en oeuvre avec *R*.
10. Considérons maintenant un modèle de régression linéaire multiple faisant intervenir des puissances entieres de la circonférence. Ecrire les modèles associés et les mettre en oeuvre avec *R*.
11. Comparer ces différents modèles à l'aide de la fonction "anova" de *R*. Proposer le meilleur modèle parmi tous les modèles testés.

TP JDTP ML