

TP régression linéaire multiples « tortues de galapagos ».

L'objectif est d'étudier la diversité d'espèces de tortues sur les îles Galapagos en fonction de certaines variables. Les données sont dans la table gala.

Les variables sont

- variables
- Species : Nombre d'espèces de tortues trouvé dans l'île
- Endemics : nombre d'espèces endémiques (ne se trouvant que sur l'île)
- Elevation : élévation maximale de l'île en mètre
- Nearest : distance de l'île la plus proche
- Scruz : distance de l'île de Santa Cruz en km
- Adjacent : La surface de l'île adjacente (km²)

On cherche à expliquer la variable Species par les variables Endemics, Elevation, Nearest, Scruz et Adjacent.

1. Regarder les corrélations deux à deux . Faire des graphes deux à deux.
2. Regarder la significativité globale du modèle.
3. Quelles sont les variables les plus influents ?
4. Construire un modèle avec ses deux variables.
5. Vérifier qu'il ajuste aussi bien que le modèle avec toutes les variables.
6. Prédire la valeur de species lorsque Elevation=100 et Adjacent=50.
7. Donner un intervalle de confiance de cette prédiction au niveau 0,95.
8. Vérifier par un choix de modèle que l'on retrouve bien les deux variables retenues.