

LMAFY1101 - Exercices - Série 8

Régression et corrélation

Exercice 1

Pour cet exercice, nous allons utiliser le jeu de données `logement` disponible [ici](#). Pour charger ces données dans votre session R, il suffit d'ouvrir le fichier `logement.rda` dans RStudio.

`logement` contient deux variables:

- P: prix de vente, en euro, d'un bien immobilier.
 - S: la superficie, en m^2 , du bien.
1. Faites une analyse descriptive de cet échantillon. À la vue de cette analyse, pouvons-nous soupçonner un lien linéaire entre ces deux variables? Répondez à cette question en utilisant à la fois des chiffres et un ou plusieurs graphiques.
 2. Calculez et *interprétez* la corrélation entre la surface et le prix.
 3. Ajustez une régression linéaire à ces données, càd. donnez les coefficients de la droite des moindres carrés et l'équation du modèle.
 4. Interprétez les coefficients estimés.

Pour faciliter les interprétations, nous suggérerons, dans tout ce qui suit, de remplacer la variable `S` par `Sm = S - 75` (75 est la surface minimale observée).

5. Soit le modèle linéaire $P \sim Sm$: $P = \alpha_0 + \alpha_1 \times Sm + \epsilon$. En se basant *uniquement* sur l'écriture de ce modèle et l'analyse effectuée précédemment, trouvez les estimateurs de moindres carrés de α_0 et α_1 manuellement (sans R). Donnez les détails de vos calculs. Vérifiez vos estimations à l'aide de R puis interprétez-les.
6. Visualisez graphiquement le résultat de votre modèle (avec `Sm` comme variable explicative). Un modèle linéaire semble-t-il adéquat pour modéliser la relation entre le prix et la surface?
7. Suggérez un modèle qui soit plus performant que celui étudié jusqu'à présent. Justifiez votre choix. Donnez l'équation de votre modèle, représentez-le graphiquement, et analysez sa qualité d'ajustement numériquement et graphiquement.

Exercice 2

Nous souhaitons analyser la relation entre la vitesse des voitures et la distance de freinage avant l'arrêt du véhicule. Pour cela, nous allons utiliser le jeu de données `cars`, qui est disponible par défaut en R. `cars` contient 2 colonnes :

- `speed`: la vitesse des voitures en miles par heure (*mph*)
 - `dist`: la distance de freinage avant l'arrêt du véhicule en pieds (*ft*).
1. Convertissez les vitesses en *km/h* ($1\text{miles/h} = 1.609344\text{km/h}$) et les distances en mètres ($1\text{ft} = 0.3048\text{m}$).
 2. Ajustez un modèle de régression linéaire à ces données. Ajoutez la droite de régression sur le nuage de points.
 3. Calculez un intervalle de confiance à 95% pour la pente. Interprétez le.
 4. La validité de l'intervalle calculé ci-dessus nécessite un certain nombre d'hypothèses sur le modèle. Nommez-les et vérifiez-les.
 5. Des chercheurs affirment que la relation entre la vitesse et la distance peut être exprimée par la relation suivante : $\text{dist} = 0.8(\text{speed} - 7)$. Que pensez-vous de cette affirmation ? Devrions-nous la rejeter ou l'adopter ? Justifiez votre réponse.
 6. Utilisez le modèle pour prédire les distances de freinage pour les vitesses suivantes: 19.312, 24.784 et 30.578. Accompagnez vos estimations d'intervalles de confiance à 95%.
 7. Calculez les trois sommes des carrés. Vérifiez que $SCT = SCR + SCE$. Utilisez vos résultats pour obtenir le coefficient de détermination.

Exercice 3

Un scientifique désire étudier l'influence d'un antibiotique sur une culture bactérienne. Il répartit dans 10 tubes des volumes égaux de culture additionnée d'une quantité X (en *mL*) d'antibiotique, et il mesure (après incubation) la densité optique Y (grandeur sans unité). Il a noté que $\bar{x} = 0.6$, $\bar{y} = 4.15$, $s_x = 0.298$ et $s_y = 0.906$.

1. D'après les résultats d'expériences très similaires, on sait que lorsqu'il n'y a pas d'antibiotique, la densité optique moyenne est estimée à 2.335. Trouver l'équation de la droite de régression.
2. Sachant que la somme des carrés résiduelle du modèle est de 0.0645, calculer la proportion de variation de la densité optique expliquée (linéairement) par la quantité d'antibiotique.
3. Pour des raisons pratiques, on souhaite modifier l'échelle de X et travailler avec $Z = 10 \times X$. Refaire les deux dernières questions en utilisant Z à la place de X .
4. De combien estimez-vous l'augmentation dans la densité optique moyenne lorsque la quantité d'antibiotique augmente de 10 *mL*.