Analyses Bivariées

Sinclair Tsana

21 Aout 2019

Cette activité est éffectuer dans le cadre des cours certifiants d’**OpenClassroom**.Comme son titre l’indique, elle porte sur **L’analyse Bivariées**.Nous devons répondre une série de quatre questions.Pour ce faire, nous avons préférer utiliser **Rmarkdown** pour sa facilité d’utilisation.

Nous avons divisé notre travail en deux parties.Dans une première partie, nous allons préparer nos données en les nettoyant, et en les visualisant.Dans une seconde partie, nous allons procéder à la réponse aux questions posées en analysant ces données.

# 2ème partie: Réponses aux Questions

## Question 1: Calcul des coéfficients de corrélation linéaire

### petal\_width en fonction de petal\_length

x = iris\_dna$petal\_length  
y = iris\_dna$petal\_width  
model1 = lm(y ~ x)  
summary(model1)$r.squared

## [1] 0.8871412

### sepal\_width en fonction de petal\_width

x = iris\_dna$petal\_length  
y = iris\_dna$sepal\_width  
model2 =lm(y ~ x)  
summary(model2)$r.squared

## [1] 0.1940728

## Question 2: Commentaires sur les résultats de la première question

* Nous avons obtenu la valeur **0.8871412**, soit **88,71%** comme coéfficient de corrélation linéaire entre les variables **petal\_width** et **petal\_length**.Ceci signifie que le lien linéaire entre ces deux variables est **fort**.La
* Puis **0.1940728** soit **19,4%** commme coéfficient de corrélation entre les variables **sepal\_width** et **petal\_width**.Donc le lien entre ces deux variables est **non linéaire**.

## Question 3: Régressions linéaires

* **cas 1:** petal\_width = a\*petal\_length + b + résidu

x = iris\_dna$petal\_length  
y = iris\_dna$petal\_width  
fit = lm(y ~ x)  
coeffs1 = coef(fit)  
sprintf(" Estimations obtenues: a = %f et b = %f",coeffs1[2], coeffs1[1])

## [1] " Estimations obtenues: a = 0.385994 et b = 0.509268"

* **cas 2:** sepal\_width = a\*petal\_width + b + résidu

x = iris\_setosa$petal\_width  
y = iris\_setosa$sepal\_width  
fit = lm(y ~ x)  
coeffs2 = coef(fit)  
sprintf("Estimations obtenues: a = %f et b = %f",coeffs2[2], coeffs2[1])

## [1] "Estimations obtenues: a = 1.787501 et b = 1.459327"

* **cas 3:** sepal\_width = a\*petal\_width + b + résidu

x = iris\_virginica$petal\_width  
y = iris\_virginica$sepal\_width  
fit = lm(y ~ x)  
coeffs3 = coef(fit)  
sprintf("Estimations obtenues: a = %f et b = %f",coeffs3[2], coeffs3[1])

## [1] "Estimations obtenues: a = 0.699251 et b = 1.037743"

* **cas 4**: sepal\_width = a\*petal\_width + b + résidu

x = iris\_versicolor$petal\_width  
y = iris\_versicolor$sepal\_width  
fit = lm(y ~ x)  
coeffs4 = coef(fit)  
sprintf("Estimations obtenues: a = %f et b = %f",coeffs4[2], coeffs4[1])

## [1] "Estimations obtenues: a = 0.973570 et b = 0.809489"

## Question 4: Imputation des valeurs manquantes par les valeurs prédites

coeffs = list(  
 "cas 1" = list('a'=coeffs1[2] , 'b'=coeffs1[1]),  
 "cas 2" = list('a'=coeffs2[2] , 'b'=coeffs2[1]),  
 "cas 3" = list('a'=coeffs3[2] , 'b'=coeffs3[1]),  
 "cas 4" = list('a'=coeffs4[2] , 'b'=coeffs4[1])  
)

lignes\_modifiees = c()  
for(i in 1:nrow(iris)){ # pour chaque individu de iris,...  
 individu = iris[i,]  
 if(is.na(individu["petal\_width"])){ #... on test si individu["petal\_width"] est nul.  
 a = coeffs[["cas 1"]][['a']]  
 b = coeffs[["cas 1"]][['b']]  
 X = individu["petal\_length"]  
 Y = a\*X + b  
 iris[i,"petal\_width"] = Y # on remplace la valeur manquante par Y  
 lignes\_modifiees = c(lignes\_modifiees, i)  
 print(sprintf("On a complété petal\_width par %f a partir de petal\_length=%f",Y,X))  
 }  
 if(is.na(individu["sepal\_width"])){  
 espece = individu["species"]  
 a = coeffs[["cas 4"]][['a']]  
 b = coeffs[["cas 4"]][['b']]  
 X = individu["petal\_width"]  
 Y = a\*X + b  
 iris[i, "sepal\_width"] = Y  
 lignes\_modifiees = c(lignes\_modifiees,i)  
 print(sprintf("On a complété sepal\_width par %f a partir de l'espece %s et de petal\_width=%f",Y,espece,X))  
 }  
}

## [1] "On a complété sepal\_width par 1.785341 a partir de l'espece 1 et de petal\_width=1.002344"  
## [1] "On a complété petal\_width par 2.063990 a partir de petal\_length=4.027838"  
## [1] "On a complété sepal\_width par 3.009714 a partir de l'espece 2 et de petal\_width=2.259954"  
## [1] "On a complété sepal\_width par 4.043566 a partir de l'espece 3 et de petal\_width=3.321872"

print("Lignes modifiées:")

## [1] "Lignes modifiées:"

print(lignes\_modifiees)

## [1] 4 54 62 101