

## TP n° 4 : Tendence linéaire

On reprend le fichier “GPA.csv” utilisé dans le TP n° 2. On rappelle que ce fichier contient des notes d’études secondaires et universitaires pour des diplômés en informatique dans une école publique locale. Notre objectif est de déterminer la droite des moindres carrés permettant d’expliquer linéairement la note universitaire d’un étudiant par sa note secondaire.

1. Quelle est la variable indépendante (explicative) ? la variable dépendante (à expliquer) ?
2. Stocker la variable `high_GPA` (note secondaire) dans un vecteur  $x$  et la variable `univ_GPA` (note universitaire) dans un vecteur  $y$ .
3. Tracer le nuage de points des données. Commenter.
4. Compléter la fonction suivante permettant de calculer le coefficient de corrélation linéaire entre deux variables quantitatives.

```
correlation= function(u,v){
  cov= cov(x,y)
  denom= sqrt(var(u) *.....)
  corr= ...../.....
  return(.....)
}
```

5. En déduire la valeur de ce coefficient entre les deux variables étudiées puis retrouver cette valeur avec la fonction `cor`.
6. Déterminer en utilisant la méthode des moindres carrés les coefficients de la droite de régression  $y = a + bx$ .
7. Utiliser la fonction `abline` pour superposer au nuage de points des données la droite de régression obtenue.
8. Déterminer les valeurs ajustées et les erreurs.
9. La fonction `lm` (linear model) permet d’effectuer une régression linéaire. Vérifier les résultats des questions 6 et 8 en exécutant les lignes de commandes suivantes:

```
model<-lm(y~x)
a<-model$coefficients[1]
b<-model$coefficients[2]
yajustes<-model$fitted.values
erreurs<-model$residuals
```

10. Représenter les résidus (les erreurs) en fonction des valeurs ajustées. Les résidus ont -ils une structure particulière ?
11. Un diplômé a eu 2.5. Donner une prédiction de sa note universitaire.