```
rm(list=ls())
setwd("~/Desktop/regression_avancee")
library(faraway)

## ------------------------------------------------
## QUESTION 1

## Chargement des données
data(esoph)

## Affichage de la tête du tableau de données
head(esoph)

## pour différente catégorie de d'âge / consommation d'alcool / de
## tabac, on reporte le nombre de patiens atteints (colonne ncases) et
## le nombre de patients sains (ncontrol)

## ------------------------------------------------
## QUESTION 2

## Résumé statistique (catégorielle vs numérique)
summary(esoph)
pairs(esoph) ## check ncontrol and ncases as a function of tabacco / alcool, age
tabac <- as.table(cbind(tapply(esoph$ncases, esoph$tobgp, sum),
                        tapply(esoph$ncontrols, esoph$tobgp, sum)))
age <- as.table(cbind(tapply(esoph$ncases, esoph$agegp, sum),
                      tapply(esoph$ncontrols, esoph$agegp, sum)))
alcool <- as.table(cbind(tapply(esoph$ncases, esoph$alcgp, sum),
                         tapply(esoph$ncontrols, esoph$alcgp, sum)))
colnames(tabac) <- colnames(age) <- colnames(alcool) <- c("cases","controls")

## ------------------------------------------------
## QUESTION 3
par(mfrow=c(2,3))
## barplot pour chaque variables explicatives par groupe d'individus
barplot(t(tabac) , main="cases/control by tabacco", col=c("red","yellow"),
        legend.text=c("cases","controls"), args.legend = list(x = "topright"))
barplot(t(age)   , main="cases/control by age",col=c("red","yellow"))
barplot(t(alcool), main="cases/control by alcool",col=c("red","yellow"))
## mozaic plot pour chaque variables explicatives par groupe d'individus
plot(tabac, col=c("red","yellow") , main="tabac effect")
plot(age, col=c("red","yellow")   , main="age effect")
plot(alcool, col=c("red","yellow"), main="alcool effect")

## ------------------------------------------------

## QUESTION 4

## ajustement du modèle
## y <- esoph$ncases[order(esoph$tobgp)]  / rep(table(esoph$tobgp), table(esoph$tobgp))
## boxplot(y ~ esoph$tobgp)
modelNULL <- glm(cbind(ncases,ncontrols) ~ 1, data=esoph, family="binomial")
model0 <- glm(cbind(ncases,ncontrols) ~ tobgp, data=esoph,
              contrasts = list(tobgp='contr.treatment'), family="binomial")
par(mfrow=c(2,2))
plot(model0, which=1:4)
summary(model0)
confint(model0)

## test global sur intérêt du modèle : test du rapport des vraisemblances
pval.global <- 1-pchisq(deviance(model0), df.residual(model0))
## Très significatif -> = 1e-12

## Anova de R compare avec le modèle avec intercept
anova(model0, test="Chisq")
1-pchisq(deviance(modelNULL) - deviance(model0), 3)

## ------------------------------------------------
## QUESTION 5

## odd ratio : exp(coef(model0)), indique l'agumentation de la proba
## associé au modèle binomiale
exp(coef(model0))

## Test sur chaque paramètre : test de Wald (beta.hat_i / sd(beta.hat)_i -> N(0,1))
z.score <- coef(model0) / sqrt(diag(vcov(model0)))
pval.coef <- 2*(1-pnorm(abs(z.score)))
```

```
## ------------------------------------------
## QUESTION 6
## Recodage en fumeur / non fumeur
smoker <- factor(rep("No-Smoking",length(esoph$tobgp)), levels=c("No-Smoking", "Smoking"))
smoker[esoph$tobgp != "0-9g/day"] <- "Smoking"
esoph$smoker <- smoker
model1 <- glm(cbind(ncases,ncontrols) ~ smoker, data=esoph,
              contrasts = list(smoker='contr.treatment'), family="binomial")
summary(model1)
exp(coef(model1))

## ------------------------------------------
## QUESTION 7
## alcool
model2 <- glm(cbind(ncases,ncontrols) ~ alcgp, data=esoph,
              contrasts = list(alcgp='contr.treatment'), family="binomial")
anova(modelNULL, model2, test="Chi")

AIC(model0)
AIC(model1)
AIC(model2)

## ------------------------------------------
## QUESTION 8
model3 <- glm(cbind(ncases,ncontrols) ~ agegp * alcgp * tobgp , data=esoph,
              family="binomial")
step(model3)

modelstep <- glm(cbind(ncases,ncontrols) ~ agegp + alcgp + tobgp , data=esoph,
                 family="binomial")
AIC(modelstep)

## ------------------------------------------
## QUESTION 9
model4 <- glm(cbind(ncases,ncontrols) ~
              unclass(agegp) + unclass(alcgp) + unclass(tobgp) , data=esoph,
                 family="binomial")
AIC(model4)

## ------------------------------------------
## QUESTION 10
eta <- predict(model4, list(agegp=1, alcgp=1, tobgp=1))
exp(eta)/(1+exp(eta))

## ------------------------------------------
## QUESTION 11
par(mfrow=c(2,2))
plot(model0, which=1:4)
```