module\_12\_assignment\_rmd

### Problem 1  
df <- MPV::p13.2  
n <- 15  
set.seed(1); chosen\_rows <- sort(sample(seq(1, nrow(df)), n))  
df <- df[chosen\_rows,]  
  
# part a  
df\_fit <- glm(formula = y~x, family = binomial(), data = df)  
  
# part b  
# Reference: https://stats.stackexchange.com/questions/108995/interpreting-residual-and-null-deviance-in-glm-r  
summary(df\_fit)  
anova(df\_fit, test = 'Chi')  
df\_fit$null.deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
deviance\_D <- 2 \*   
 sum(df$y \* log(df$y / df\_fit$fitted.values),  
 (1 - df$y) \* log((1 - df$y) / (1 - df\_fit$fitted.values)),  
 na.rm = TRUE)  
deviance\_D > qchisq(p = 0.95, df = n - 2)  
deviance\_D / (n - 2)  
  
# part d  
df$x2 <- df$x^2  
df\_fit2 <- glm(formula = y ~ x + x2, family = binomial(), data = df)  
summary(df\_fit2)  
anova(df\_fit2, test = 'Chi')  
df\_fit2$null.deviance - df\_fit2$deviance >  
 qchisq(p = 0.95, df = df\_fit2$df.null - df\_fit2$df.residual)  
deviance\_D <- 2 \*   
 sum(df$y \* log(df$y / df\_fit2$fitted.values),  
 (1 - df$y) \* log((1 - df$y) / (1 - df\_fit2$fitted.values)),  
 na.rm = TRUE)  
deviance\_D > qchisq(p = 0.95, df = n - 3)  
deviance\_D / (n - 3)  
  
### Problem 2  
df <- MPV::p13.4  
n <- 9  
set.seed(1); chosen\_rows <- sort(sample(seq(1, nrow(df)), n))  
df <- df[chosen\_rows,]  
df$y <- df$r / df$n  
  
# part a  
df\_fit <- glm(formula = y~x, family = binomial(), data = df)  
  
# part b  
summary(df\_fit)  
anova(df\_fit, test = 'Chi')  
df\_fit$null.deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
deviance\_D <- 2 \*   
 sum(df$y \* log(df$y / df\_fit$fitted.values),  
 (1 - df$y) \* log((1 - df$y) / (1 - df\_fit$fitted.values)),  
 na.rm = TRUE)  
deviance\_D > qchisq(p = 0.95, df = n - 2)  
deviance\_D < qchisq(p = 0.95, df = n - 2, lower.tail = FALSE)  
deviance\_D / (n - 2)  
  
# part c  
plot(df$x, df\_fit$fitted.values, ylim = c(0,1), pch = 3,  
 main = 'Fitted and True Values against X',  
 xlab = 'x', ylab = 'y, y-hat')  
points(df$x, df$y, pch=2)  
legend("right", legend = c('Fitted Values', 'True Values'),  
 pch = c(3,2))  
  
# part d  
df$x2 <- df$x^2  
df\_fit2 <- glm(formula = y ~ x + x2, family = binomial(), data = df)  
summary(df\_fit2)  
anova(df\_fit2, test = 'Chi')  
df\_fit2$null.deviance - df\_fit2$deviance >  
 qchisq(p = 0.95, df = df\_fit2$df.null - df\_fit2$df.residual)  
  
deviance\_D <- 2 \*   
 sum(df$y \* log(df$y / df\_fit2$fitted.values),  
 (1 - df$y) \* log((1 - df$y) / (1 - df\_fit2$fitted.values)),  
 na.rm = TRUE)  
  
deviance\_D > qchisq(p = 0.95, df = n - 3)  
deviance\_D < qchisq(p = 0.95, df = n - 3, lower.tail = FALSE)  
  
deviance\_D / (n - 3)  
  
# part e  
plot(df$x, df\_fit$fitted.values, ylim = c(0,1), pch = 3,  
 main = 'Fitted and True Values against X',  
 xlab = 'x', ylab = 'y, y-hat')  
points(df$x, df$y, pch=2, col = 'blue')  
points(df$x, df\_fit2$fitted.values, pch=1, col = 'red')  
legend("right", legend = c('Model 1', 'Model 2', 'True Values'),  
 pch = c(3,1,2), col = c('black', 'red', 'blue'))  
  
# part f  
summary(df\_fit2)  
std\_errors <- sqrt(diag(summary(df\_fit2)$cov.unscaled))  
wald\_statistics <- df\_fit2$coefficients / std\_errors  
  
# part g  
round(df\_fit2$coefficients + c(1.96 \* std\_errors),4)  
round(df\_fit2$coefficients - c(1.96 \* std\_errors),4)  
  
### Problem 3  
df <- MPV::p13.7  
n <- 30  
set.seed(1); chosen\_rows <- sort(sample(seq(1, nrow(df)), n))  
df <- df[chosen\_rows,]  
  
# part a  
df\_fit <- glm(formula = y~., family = poisson(), data = df)  
summary(df\_fit)  
anova(df\_fit)  
  
# part b  
deviance\_D <- summary(df\_fit)$deviance  
deviance\_D > qchisq(p = 0.95, df = n - 5)  
deviance\_D / (n - 5)  
  
# part c  
df\_fit1 <- glm(formula = y~x2+x3+x4, family = poisson(), data = df)  
df\_fit1$deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
df\_fit2 <- glm(formula = y~x1+x3+x4, family = poisson(), data = df)  
df\_fit2$deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
df\_fit3 <- glm(formula = y~x1+x2+x4, family = poisson(), data = df)  
df\_fit3$deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
df\_fit4 <- glm(formula = y~x1+x2+x3, family = poisson(), data = df)  
df\_fit4$deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
# part d  
summary(df\_fit)  
std\_errors <- sqrt(diag(summary(df\_fit)$cov.unscaled))  
wald\_statistics <- df\_fit$coefficients / std\_errors  
alpha <- 0.05  
abs(round(wald\_statistics,4)) > qnorm(p = 1 - alpha / 2)  
  
# part e  
round(df\_fit$coefficients + c(qnorm(p = 1 - alpha / 2) \* std\_errors),4)  
round(df\_fit$coefficients - c(qnorm(p = 1 - alpha / 2) \* std\_errors),4)  
  
### Problem 5  
df <- read.csv('p13\_26.csv')  
n <- 10  
set.seed(1); chosen\_rows <- sort(sample(seq(1, nrow(df)), n))  
df <- df[chosen\_rows,]  
  
# Fit a model  
df\_fit <- glm(formula = y~., family = poisson(), data = df)  
summary(df\_fit)  
anova(df\_fit)  
  
# Calculate the deviance  
deviance\_D <- summary(df\_fit)$deviance  
deviance\_D > qchisq(p = 0.95, df = n - 4)  
deviance\_D / (n - 4)  
  
# partial deviance  
df\_fit1 <- glm(formula = y~Oil+Time, family = poisson(), data = df)  
df\_fit1$deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
df\_fit2 <- glm(formula = y~Temperature+Time, family = poisson(), data = df)  
df\_fit2$deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
df\_fit3 <- glm(formula = y~Temperature+Oil, family = poisson(), data = df)  
df\_fit3$deviance - df\_fit$deviance > qchisq(p = 0.95, df = 1)  
  
# Wald statistics  
summary(df\_fit)  
std\_errors <- sqrt(diag(summary(df\_fit)$cov.unscaled))  
wald\_statistics <- df\_fit$coefficients / std\_errors  
alpha <- 0.05  
abs(round(wald\_statistics,4)) > qnorm(p = 1 - alpha / 2)  
  
# Wald CI  
round(df\_fit$coefficients + c(qnorm(p = 1 - alpha / 2) \* std\_errors),4)  
round(df\_fit$coefficients - c(qnorm(p = 1 - alpha / 2) \* std\_errors),4)