ch8-6

Harry Woo

2020 6 16

rm(list = ls())  
  
library(MASS) # for menarche  
library(dplyr)  
library(ggplot2)  
library(gt)

data(menarche)  
head(menarche) %>% gt()

Age

Total

Menarche

9.21

376

0

10.21

200

0

10.58

93

0

10.83

120

2

11.08

90

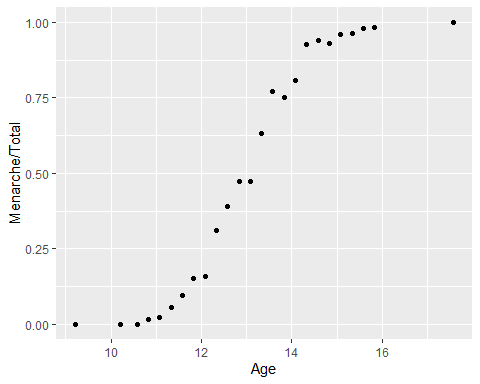
2

11.33

88

5

mnc\_plot <- ggplot(data = menarche, aes(x = Age, y = Menarche/Total)) +  
 geom\_point(pch = 19)  
mnc\_plot



# S-shape : 로지스틱 회귀모형 적합 타당

mnc\_glm <- glm(cbind(Menarche, Total - Menarche) ~ Age, family = binomial, data = menarche)  
summary(mnc\_glm)

##   
## Call:  
## glm(formula = cbind(Menarche, Total - Menarche) ~ Age, family = binomial,   
## data = menarche)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0363 -0.9953 -0.4900 0.7780 1.3675   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -21.22639 0.77068 -27.54 <2e-16 \*\*\*  
## Age 1.63197 0.05895 27.68 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3693.884 on 24 degrees of freedom  
## Residual deviance: 26.703 on 23 degrees of freedom  
## AIC: 114.76  
##   
## Number of Fisher Scoring iterations: 4

# Age 유의  
  
exp(mnc\_glm$coefficients["Age"])

## Age   
## 5.113931

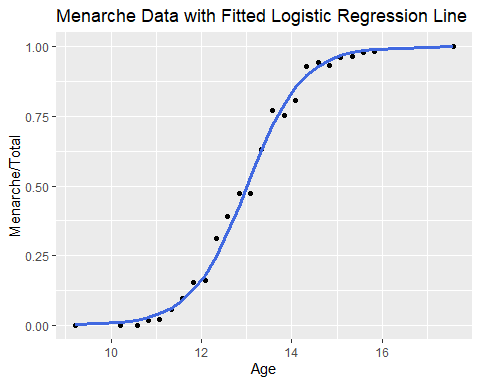
1 - pchisq(mnc\_glm$deviance, mnc\_glm$df.residual)

## [1] 0.2687953

mnc\_glm$fitted

## 1 2 3 4 5 6 7   
## 0.00203349 0.01031285 0.01870339 0.02786353 0.04132099 0.06087114 0.08881411   
## 8 9 10 11 12 13 14   
## 0.12783822 0.18061043 0.24894906 0.33264793 0.42843455 0.52990205 0.62895659   
## 15 16 17 18 19 20 21   
## 0.71823740 0.79310223 0.85216954 0.89657280 0.92875389 0.95146398 0.96719083   
## 22 23 24 25   
## 0.97793995 0.98522119 0.99012343 0.99942675

mnc\_plot + geom\_line(aes(x = Age, y = mnc\_glm$fitted), lwd = 1.2, colour = "royalblue") +  
 ggtitle(label = "Menarche Data with Fitted Logistic Regression Line")



prediction <- ifelse(predict(mnc\_glm, menarche, type = "response") > 0.5, TRUE, FALSE)  
actual <- ifelse(menarche$Menarche/menarche$Total > 0.5, TRUE, FALSE)  
  
confu\_mat <- table(as.numeric(prediction), as.numeric(actual),  
 dnn = c("Prediction", "Actual"))  
confu\_mat

## Actual  
## Prediction 0 1  
## 0 12 0  
## 1 1 12

error <- 1 - sum(diag(confu\_mat))/sum(confu\_mat)  
error

## [1] 0.04

# table(as.numeric(prediction), as.numeric(as.logical(mnc\_glm$y)))  
# menarche2 <- menarche %>% mutate(ratio = ifelse(Menarche/Total < 0.5, "No", "Yes"), fit = prediction)  
# confu\_mat <- table(menarche2$ratio, menarche2$fit)  
# mnc\_glm$y == menarche$Menarche/menarche$Total