L17Ex_PneumoconiosisData

Rick Davila

6/14/2020

```
knitr::opts_chunk$set(echo = TRUE)

#Sys.setenv(JAVA_HOME='C:\\Program Files\\Java\\jdk-14.0.1') # for 64-bit version
#library(rJava)
library(MASS)

# Example 13.1 (p. 426-427) ###
# Create data frame with problem data

years <- c(5.8,15.0,21.5,27.5,33.5,39.5,46.0,51.5)
cases <- c(0,1,3,8,9,8,10,5)
miners <- c(98,54,43,48,51,38,28,11)
ymat <- cbind(cases,miners-cases)

p_data <- data.frame(ymat, years)

# Perform logistic regression using glm(). Compare output to Table 13.2 on p. 427.

model.131 <- glm(ymat ~ years, binomial(link = "logit"))
model.131
```

```
##
## Call:  glm(formula = ymat ~ years, family = binomial(link = "logit"))
##
## Coefficients:
## (Intercept)        years
##    -4.79648      0.09346
##
## Degrees of Freedom: 7 Total (i.e. Null);  6 Residual
## Null Deviance:       56.9
## Residual Deviance: 6.051     AIC: 32.88
```

```
# Example 13.2 (p. 428)
# Calculate odds ratio for the regressor variable; compare to p. 428

odds_ratio <- exp(model.131$coefficients[2])
odds_ratio
```

```
##   years
## 1.09797
```

odds ratio = 1.0979699

```
# Example 13.3 (p. 434-436)
# Add a quadratic term to the model

years_sq <- years*years
model.133 <- glm(ymat ~ years + years_sq, binomial(link = "logit"))
model.133
```

```
##
## Call:  glm(formula = ymat ~ years + years_sq, family = binomial(link = "logit"))
##
## Coefficients:
## (Intercept)        years     years_sq
##   -6.710791     0.227607    -0.002079
##
## Degrees of Freedom: 7 Total (i.e. Null);  5 Residual
## Null Deviance:       56.9
## Residual Deviance: 3.282     AIC: 32.11
```

```
# Example 13.4 (p. 437) ###
# Test each regression coefficient for significance automatically using the summary() command

summary(model.133)
```

```
##
## Call:
## glm(formula = ymat ~ years + years_sq, family = binomial(link = "logit"))
##
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
## -0.9118  -0.2109   0.3056   1.0209  -0.3351  -0.9298   0.1314   0.5254
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.710791   1.535226  -4.371 1.24e-05 ***
## years        0.227607   0.092756   2.454   0.0141 *
## years_sq    -0.002079   0.001361  -1.527   0.1267
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 56.9028  on 7  degrees of freedom
## Residual deviance:  3.2816  on 5  degrees of freedom
## AIC: 32.108
##
## Number of Fisher Scoring iterations: 4
```

```
# Example 13.5 (p. 438) ###
# Calculate confidence intervals automatically using confint()
confint.default(model.133)
```

```
##                    2.5 %        97.5 %
## (Intercept) -9.719779184 -3.7018034417
## years        0.045808498  0.4094051509
## years_sq    -0.004746846  0.0005891249
```

```
# Calculate by hand using Equation (13.25)
Z_crit <- qnorm(1-.05/2)
se_beta_0 <- coef(summary(model.133))[, "Std. Error"][1]
se_beta_1 <- coef(summary(model.133))[, "Std. Error"][2]
se_beta_11 <- coef(summary(model.133))[, "Std. Error"][3]

OR_Upper_0 <- model.133$coefficients[1] + Z_crit*se_beta_0
OR_Lower_0 <- model.133$coefficients[1] - Z_crit*se_beta_0
OR_Upper_1 <- model.133$coefficients[2] + Z_crit*se_beta_1
OR_Lower_1 <- model.133$coefficients[2] - Z_crit*se_beta_1
OR_Upper_11 <- model.133$coefficients[3] + Z_crit*se_beta_11
OR_Lower_11 <- model.133$coefficients[3] - Z_crit*se_beta_11
```

approximate 95% confidence intervals calculated by hand

$$\hat{\beta}_0 - Z_{0.025}\,\mathrm{se}(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + Z_{0.025}\,\mathrm{se}(\hat{\beta}_0)$$

$$-9.7197792 \le \beta_0 \le -3.7018034$$

$$\hat{\beta}_1 - Z_{0.025}\,\mathrm{se}(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_0 + Z_{0.025}\,\mathrm{se}(\hat{\beta}_1)$$

$$0.0458085 \le \beta_1 \le 0.4094052$$

$$\hat{\beta}_{11} - Z_{0.025}\,\mathrm{se}(\hat{\beta}_{11}) \le \beta_{11} \le \hat{\beta}_{11} + Z_{0.025}\,\mathrm{se}(\hat{\beta}_{11})$$

$$-0.0047468 \le \beta_{11} \le 5.8912492 \times 10^{-4}$$

```
# Example 13.6 (p. 438) ###
# On the original model, calculate a 95% confidence interval on the odds ratio using Equation (13.26)

odds_ratio <- exp(coef(model.133))
# 95% CI on odds ratio
Z_crit <- qnorm(1-.05/2)
OR_Upper <- exp(coef(model.133) + Z_crit*sqrt(diag(vcov(model.133))))
OR_Lower <- exp(coef(model.133) - Z_crit*sqrt(diag(vcov(model.133))))
```

$$exp[\hat{\beta}_j - Z_{\alpha/2}\mathrm{se}(\hat{\beta}_j)] \le O_R \le exp[\hat{\beta}_j + Z_{\alpha/2}\mathrm{se}(\hat{\beta}_j)]$$

$$6.0083266 \times 10^{-5} \le O_R = \hat{\beta}_0 \le 0.024679$$

$$1.0468739 \leq O_R = \hat{\beta}_1 \leq 1.5059217$$

$$0.9952644 \leq O_R = \hat{\beta_{11}} \leq 1.0005893$$

Example 13.7 (p. 439-440)

Create a 95% confidence interval on the predicted probability of Years = 40.

```
# Calculate predicted probability using predict(); compare to p. 439
# note, type = "response" gives the predicted probabilities
predict(model.131,type="response",se.fit=TRUE,newdata=data.frame(years=40))
```

```
## $fit
##         1
## 0.2576988
##
## $se.fit
##          1
## 0.03637976
##
## $residual.scale
## [1] 1
```

```
# calculate the predicted linear predictor using predict(); compare to p. 439
# note, type - default is on the scale of the linear predictors
predict(model.131,se.fit=TRUE,newdata=data.frame(years=40))
```

```
## $fit
##         1
## -1.057964
##
## $se.fit
## [1] 0.1901811
##
## $residual.scale
## [1] 1
```

```
# Define vector for the new observation

x_0 <- c(1,40)

# Calculate the variance of the new observation. The function vcov() automatically calculates the inverse of X'VX.

XtVX <- vcov(model.131)

# Calculate CI on linear predictor using Equation (13.27)

Z_crit <- qnorm(1-.05/2)
Ux_0 <- t(x_0) %*% coef(model.131) + Z_crit*sqrt(t(x_0) %*% XtVX %*% x_0)
Lx_0 <- t(x_0) %*% coef(model.131) - Z_crit*sqrt(t(x_0) %*% XtVX %*% x_0)

# Convert to probabilities using Equation (13.28)
LowerP <- exp(Lx_0)/(1+exp(Lx_0))
UpperP <- exp(Ux_0)/(1+exp(Ux_0))
```

Calculate CI on linear predictor using Equation (13.27)

$$x_0'\hat{\beta} - Z_{\alpha/2}\sqrt{x_0'(X'VX)^{-1}x_0} \leq x_0'\hat{\beta} \leq x_0'\hat{\beta} + Z_{\alpha/2}\sqrt{x_0'(X'VX)^{-1}x_0}$$

$$-1.4307119 \leq x_0'\hat{\beta} \leq -0.6852157$$

Convert to probabilities using Equation (13.28)

$$\frac{exp[L(x_0)]}{1 + exp[L(x_0)]} \leq \pi_0 \leq \frac{exp[U(x_0)]}{1 + exp[U(x_0)]}$$

$$0.1929878 \leq \pi_0 \leq 0.3350982$$