Question 1

Joseph Froelicher

February 26, 2020

Part A

$$ln\left(\frac{p_i}{1-p_i}\right) = \mathbf{X}_j\beta$$

$$\hat{\beta}_0 = ln\left(\frac{p(y=1|x_j=0)}{1-p(y=1|x_j=0)}\right) = ln\left(\frac{(619/2416)}{1-(619/2416)}\right) = -1.065769$$

$$\hat{\beta}_1 = (\hat{\beta}_0 + \hat{\beta}_1) - \hat{\beta}_0 = ln\left(\frac{p(y=1|x_j=1)}{1-p(y=1|x_j=1)}\right) - \hat{\beta}_0 = ln\left(\frac{(355/771)}{1-(355/771)}\right) - (-1.065769) = 0.9072015$$

$$\hat{\beta}_2 = (\hat{\beta}_0 + \hat{\beta}_2) - \hat{\beta}_0 = ln\left(\frac{p(y=1|x_j=2)}{1-p(y=1|x_j=2)}\right) - \hat{\beta}_0 = ln\left(\frac{(162/731)}{1-(162/731)}\right) - (-1.065769) = -0.1905151$$

Part B

$$L(Y_i|p) = \prod_{i=1}^{N}\binom{N}{Y_i}p^{Y_i}(1-p)^{1-Y_i}$$

$$L(Y_i|p) = p^{\sum_{i=1}^{n}Y_i}(1-p)^{n-\sum_{i=1}^{n}Y_i}$$

$$ln(L(Y_i|p)) = ln\left(p^{\sum_{i=1}^{n}Y_i}(1-p)^{n-\sum_{i=1}^{n}Y_i}\right)$$

$$l(Y_i|p) = \sum_{i=1}^{n}\left(Y_i\, ln(p_i) + (n-Y_i)\, ln(1-p_i)\right)$$

$$l(Y_i, X_{ij}|\beta) = \sum_{i=1}^{n}\left(Y_i\, ln(\frac{e^{\beta_0+\beta_1X_{1j}+\beta_2X_{2j}}}{1-(e^{\beta_0+\beta_1X_{1j}+\beta_2X_{2j}})}) + (1-Y_i)\, ln(\frac{e^{\beta_0+\beta_1X_{1j}+\beta_2X_{2j}}}{1-(e^{\beta_0+\beta_1X_{1j}+\beta_2X_{2j}})})\right)$$

```r
data = data.frame(
  "strikes" = c("1", "2", "3"),
  "misconduct" = c(619, 355, 162),
  "no_misconduct" = c(1797, 416, 569)
)

model1 = glm(
  cbind(misconduct, no_misconduct) ~ strikes, data, family = binomial
)

like1 = logLik(model1)
```

Using the `logLik()` function in R, we get the value of -10.8699802 for the likelihood of model 1.

Part C

$$l(Y_i|p) = \sum_{i=1}^{n} (Y_i\, ln(p_i) + (n - Y_i)\, ln(1 - p_i))$$

$$l(Y_i, X_{ij}|\beta) = \sum_{i=1}^{n} \left(Y_i\, ln(\frac{e^{\beta_0}}{1 - e^{\beta_0}}) + (1 - Y_i)\, ln(\frac{e^{\beta_0}}{1 - e^{\beta_0}}) \right)$$

```r
model0 = glm(
  cbind(misconduct, no_misconduct) ~ 1, data, family = binomial
)

like0 = logLik(model0)
```

Using the `logLik()` function in R, we get the value of -76.4108396 for the likelihood of model 0.

Part D

$$T_{LR} = -2ln\left[\frac{L(p_{H_0})}{L(p_{MLE})}\right]$$

$$T_{LR} = -2ln\left[\frac{e^{-76.41084}}{e^{-10.86998}}\right]$$

$$T_{LR} = 131.0817$$

Part E

```r
data_numeric = data.frame(
  "strikes" = c(1, 2, 3),
  "misconduct" = c(619, 355, 162),
  "no_misconduct" = c(1797, 416, 569)
)

model2 = glm(
  cbind(misconduct, no_misconduct) ~ strikes, data_numeric, family = binomial
)

summary(model2)
```

```
##
## Call:
## glm(formula = cbind(misconduct, no_misconduct) ~ strikes, family = binomial,
##     data = data_numeric)
##
## Deviance Residuals:
##      1       2       3
## -2.903   9.647  -5.254
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.99461    0.07872 -12.635   <2e-16 ***
## strikes      0.06270    0.04439   1.413    0.158
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 131.08  on 2  degrees of freedom
## Residual deviance: 129.10  on 1  degrees of freedom
## AIC: 154.84
##
## Number of Fisher Scoring iterations: 4
```

```
like2 = logLik(model2)

p_1 = exp(summary(model2)$coefficients[1, 1] + summary(model2)$coefficients[2, 1] * 1) /
  (1 - exp(summary(model2)$coefficients[1, 1] + summary(model2)$coefficients[2, 1] * 1))

p_3 = exp(summary(model2)$coefficients[1, 1] + summary(model2)$coefficients[2, 1] * 3) /
  (1 - exp(summary(model2)$coefficients[1, 1] + summary(model2)$coefficients[2, 1] * 3))
```

The predicted probability of a misconduct violation during the first year in prison for a prisoner with 1 strike is 0.649633. The predicted probability of a misconduct violation during the first year in prison for a prisoner with 3 strikes is 0.8064297.