PSC 103B Homework 7

Answer Key

```r
wine <- read.csv("../lab/data/wine.csv")
```

Question 1

```r
wine$quality_binary = ifelse(wine$quality == "good", 1, 0)

mod1 = glm(quality_binary ~ residual.sugar, data = wine, family = "binomial")
summary(mod1)
```

```
##
## Call:
## glm(formula = quality_binary ~ residual.sugar, family = "binomial",
##     data = wine)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     0.146858   0.103255   1.422    0.155
## residual.sugar -0.003071   0.035549  -0.086    0.931
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2209  on 1598  degrees of freedom
## Residual deviance: 2209  on 1597  degrees of freedom
## AIC: 2213
##
## Number of Fisher Scoring iterations: 3
```

Question 2

$$\log(Odds) = 0.15 - 0.003 \times \text{ResidualSugar}$$

The intercept of 0.15 means that a wine with 0 residual sugar has a log-odds of 0.15 of being rated good.

The slope of -0.003 means that for every 1-unit increase in the residual sugar, the log-odds of the wine being rated good decrease by 0.003

Question 3

$$\text{Odds} = e^{0.15 - 0.003 \times \text{ResidualSugar}}$$

The intercept is 1.16, so when a wine has 0 residual sugar, the odds of it being rated good are 1.16 (it is 1.16 times as likely to be rated good as it is to be rated bad).

The slope is 0.997, so for every 1-unit increase in residual sugar, the odds of the wine being rated good change by a factor of 0.997.

Question 4

```
wine$residual.sugar_c = wine$residual.sugar - mean(wine$residual.sugar)
mod2 = glm(quality_binary ~ residual.sugar_c, data = wine, family = "binomial")
summary(mod2)
```

```
##
## Call:
## glm(formula = quality_binary ~ residual.sugar_c, family = "binomial",
##     data = wine)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.139061   0.050137   2.774  0.00554 **
## residual.sugar_c -0.003071   0.035549  -0.086  0.93116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2209  on 1598  degrees of freedom
## Residual deviance: 2209  on 1597  degrees of freedom
## AIC: 2213
##
## Number of Fisher Scoring iterations: 3
```

Question 5

The intercept in the log-odds is 0.14, which means that a wine with an average amount of residual sugar has a log-odds of 0.14 of being rated good.

The intercept in terms of the odds is 1.15, which means that a wine with an average amount of residual sugar is 1.15 times as likely to be rated good as it is to be rated bad.

Question 6

- $\log(Odds) = 0.15 - 0.003 \times 13 = 11$
- $\text{Odds} = e^{0.11} = 1.12$
- $P = \frac{1.12}{1+1.12} = 0.53$

A wine with a residual sugar content of 13 has a log-odds of 0.11 of being rated good, is 1.12 times as likely to be rated good as it is to be rated bad, and has a 53% probability of being rated good.

Question 7

```
mod3 = glm(quality_binary ~ residual.sugar + citric.acid, data = wine, family = binomial)
summary(mod3)
```

```
##
## Call:
## glm(formula = quality_binary ~ residual.sugar + citric.acid,
##     family = binomial, data = wine)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.22714    0.11936  -1.903    0.057 .
## residual.sugar -0.03688    0.03634  -1.015    0.310
## citric.acid     1.71381    0.26852   6.382 1.74e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2209  on 1598  degrees of freedom
## Residual deviance: 2167  on 1596  degrees of freedom
## AIC: 2173
##
## Number of Fisher Scoring iterations: 4
```

Question 8

A wine with 0 residual sugar and 0 citric acid has a log-odds of -0.23 of being rated good.

Every 1-unit increase in residual sugar, holding citric acid constant, decreases the log-odds by 0.04.

Every 1-unit increase in citric acid, holding residual sugar constant, increases the log-odds by 1.71.

Question 9

A wine with 0 residual sugar and 0 citric acid is 0.79 times as likely to be rated good as it is to be rated bad.

For every 1-unit increase in residual sugar, holding citric acid constant, the odds of a wine being rated good change by a factor of 0.96.

For every 1-unit increase in citric acid, holding residual sugar constant, the odds of a wine being rated good change by a factor of 5.53.

Question 10

```
library(palmerpenguins)
```

```
## Warning: package 'palmerpenguins' was built under R version 4.3.2
```

```r
dummymodel = lm(bill_length_mm ~ island, data = penguins)
summary(dummymodel)
```

```
##
## Call:
## lm(formula = bill_length_mm ~ island, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.0677  -3.8559   0.2958   3.8175  14.3425
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      45.2575     0.3897 116.127  < 2e-16 ***
## islandDream      -1.0897     0.5970  -1.825   0.0688 .
## islandTorgersen  -6.3065     0.8057  -7.827 6.44e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.036 on 339 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.154,  Adjusted R-squared:  0.149
## F-statistic: 30.86 on 2 and 339 DF,  p-value: 4.86e-13
```

Question 11

The intercept is 45.26, which represents the average bill length of the reference group (Biscoe Island penguins).

The slope of -1.09 means that Dream Island penguins have an average bill length that is 1.09 mm shorter than the average bill length of Biscoe Island penguins. But this difference is not significant.

The slope of -6.31 means that Torgersen Island penguins have an average bill length that is 6.31 mm shorter than the average bill length of Biscoe Island penguins. This difference is significant.