Homework 3

Joe Kingsbury

Question 1

Fit a logistic regression model that assumes the probability of success is an additive function of variables x1 and x2.

```
dat <- read.csv("Homework 3 Data.csv")
fit <- glm(y ~ x1 + x2, data = dat, family = binomial)
summary(fit)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2, family = binomial, data = dat)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.43687  -1.00592   0.01992   1.00818   1.50526
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.5948     0.3950   1.506   0.1321
## x1           -0.1680     0.3648  -0.460   0.6452
## x2b          -0.9679     0.4136  -2.340   0.0193 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 138.63  on 99  degrees of freedom
## Residual deviance: 132.59  on 97  degrees of freedom
## AIC: 138.59
##
## Number of Fisher Scoring iterations: 4
```

Question 2

Interpret the effect of variable x1 on the log odds of success. Verify your interpretation in R.

```
betas <- coef(fit)
p_2 <- plogis(betas[1] + betas[2] * 2 + betas[3] * 0)
p_2
```

```
## (Intercept)
##   0.5643653
```

```r
p_3 <- plogis(betas[1] + betas[2] * 3 + betas[3] * 0)
p_3
```

```
## (Intercept)
##   0.5227148
```

```r
log((p_3 / (1 - p_3)) / (p_2 / (1 - p_2)))
```

```
## (Intercept)
##   -0.167976
```

Question 3

Interpret the effect of variable x2 on the log odds of success. Verify your interpretation in R.

```r
p_4 <- plogis(betas[1] + betas[3] * 0)
p_4
```

```
## (Intercept)
##   0.6444772
```

```r
p_5 <- plogis(betas[1] + betas[3] * 1)
p_5
```

```
## (Intercept)
##   0.4077983
```

```r
log((p_5 / (1 - p_5)) / (p_4 / (1 - p_4)))
```

```
## (Intercept)
##  -0.9679244
```

Question 4

Duplicate the Wald Test and p-values for variables x1 and x2 performed by the glm() function. Do you reject or fail to reject your null hypothesis?

```r
#Test statistic for x1
x1_ts <- betas[2] / summary(fit)[['coefficients']]['x1', 'Std. Error']
#P-value for x1
2 * pnorm(-1 * abs(x1_ts), mean = 0, sd = 1)
```

```
##        x1
## 0.6451594
```

```r
#Test statistic for x2
x2_ts <- betas[3] / summary(fit)[['coefficients']]['x2b', 'Std. Error']
#P-value for x2
2 * pnorm(-1 * abs(x2_ts), mean = 0, sd = 1)
```

```
##         x2b
## 0.01926485
```

For the variable x1 we fail to reject the null hypothesis because the p-value is quite large, > 0.05. However, for the variable x2 we reject the null hypothesis because the p-value is less than our citical alpha level, 0.05 > 0.019.

Question 5

Predict and plot the mean probability of success over the range of values of x1.

```
#Log odds of success
x1 <- seq(from = min(dat$x1), to = max(dat$x1),length.out = 100)
y <- betas[1] + (betas[2] * x1) + (betas[3] * 1)
plot(x = x1, y = plogis(y), ylab = 'Probability of Success',
xlab = 'x1', cex.axis = 1.5, cex.lab = 1.5, type = 'l')
```