Stat 3155 Homework 5, Logistic Regression

Thomas Maltese

April 24th, 2018

Problem 1

Part A.

```
b = c(0.11, 0.16, -0.57, -0.66, 0.47, -1.67)

x = c(0, 1, 0, 0, 1)
logit = b[1] + sum(b[2:length(b)] * x)
p = exp(logit) / (1 + exp(logit))
cat(p, '\n')
```

```
## 0.106215
```

```
x = c(1, 0, 1, 1, 0)
logit = b[1] + sum(b[2:length(b)] * x)
p = exp(logit) / (1 + exp(logit))
cat(p, '\n')
```

```
## 0.5199893
```

Therefore, the probability that a male, protestant, and republican supports laws legalizing abortion is approximately 10.62%, and the probability that a femaile, catholic, and democrat supports laws legalizing abortion is approximately 51.99%

Part B.

$b_1 = 0.16$ means that if the subject is Female, the odds of the subject supporting laws legalizing abortion change by a multiplicative factor of about:

```
cat(exp(0.16), '\n')
```

```
## 1.173511
```

which is an increase of about 17.35% percent, when compared with male subjects, ceteris paribus.

and b_2 = -0.57 means that if the subject is Protestant, the odds of the subject supporting laws legalizing abortion change by a multiplicative factor of about

```
cat(exp(-0.57), '\n')
```

```
## 0.5655254
```

which is a decrease of about 43.44% when compared with non-protestant subject, ceteris paribus.

Part C.

let SE(b_1) = 0.0064 The confidence interval is computed with:

```
b1 = 0.16
s = 0.0064
z = abs(qnorm(0.05 / 2))
upper = b1 + z * s
lower = b1 - z * s
cat(lower, upper, '\n')
```

```
## 0.1474562 0.1725438
```

then: we say that we are 95% confident that when the subject is Female, ceteris paribus, the odds that the subject suppors laws legalizing abortion are increased by a multiplicative factor between $e^{0.1474} \approx 15.88$ percent and $e^{0.1725} \approx 18.82$ percent

Part D.

```
b1 = 0.16
s = 0.0064
z = b1 / s
zz = abs(qnorm(0.05 / 2))
if (z > zz) {
    cat("We reject the null hypothesis\n");
} else {
    cat("We fail to reject the null hypothesis\n");
}
```

```
## We reject the null hypothesis
```

Part E.

Let $\text{SE}(b_2) = 0.38$ The confidence interval is computed with:

```
b2 = -0.57
s = 0.38
z = abs(qnorm(0.05 / 2))
upper = b2 + z * s
lower = b2 - z * s
cat(lower, upper, '\n')
```

```
## -1.314786 0.1747863
```

we say that we are 95% confident that when the subject is is Protestant, ceteris paribus, the odds that the subject supports laws legalizing abortion are increased by a multiplicative factor between $e^{-1.3147} \approx 17.47$ percent and $e^{0.1747} \approx 119.09$ percent; these multiplicative factors correspond with an approximate 82.53% decrease and 19.09% increase, respectively.

Since we get conflicting answers (being Protestant could mean an increase or decrease) in the probability!) we conclude that it doesn't have much predictive power.

Problem 2

```
data =read.table("data.txt", sep=" ", header=T)
print(data)
```

```
##    Race Gender Yes  No
## 1 white   male  42 134
## 2 white female  26 169
## 3 black   male  29  23
## 4 black female  22  36
```

```
attach(data)
model = glm(cbind(Yes, No) ~ factor(Gender) + factor(Race), family=binomial)
summary(model)
```

```
##
## Call:
## glm(formula = cbind(Yes, No) ~ factor(Gender) + factor(Race),
##     family = binomial)
##
## Deviance Residuals:
##         1          2          3          4
## -0.009857   0.011753   0.015565  -0.015083
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -0.4884     0.2226  -2.194  0.02825 *
## factor(Gender)male    0.7159     0.2246   3.188  0.00143 **
## factor(Race)white    -1.3859     0.2379  -5.824 5.73e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.3249e+01  on 3  degrees of freedom
## Residual deviance: 7.0506e-04  on 1  degrees of freedom
## AIC: 25.128
##
## Number of Fisher Scoring iterations: 3
```

Part A.

The estimates for β_1 and β_2 are 0.7159 and -1.3859 respectively.

Part B.

The probability of observing a response 'Yes' from a black female is

```
logit = predict(model, newdata=data.frame(Gender="female", Race="black"))
cat(exp(logit) / (1 + exp(logit)), '\n')
```

```
## 0.3802716
```

Which is approximately 38%

The probability observing a response 'Yes' from a white male is

```
logit = predict(model, newdata=data.frame(Gender="male", Race="white"))
cat(exp(logit) / (1 + exp(logit)), '\n')
```

```
## 0.2389532
```

Which is approximately 23.89%

Part C.

```
b1  = 0.7159
sb1 = 0.2246
z = abs(qnorm(0.05 / 2))
upper = b1 + z * sb1
lower = b1 - z * sb1
cat(lower, upper, '\n')
```

```
## 0.2756921 1.156108
```

We are 95% confident that, when the subject is male, ceteris paribus, the odds of the subject having had intercourse increases by a multiplicative factor somewhere between $e^{0.2756}$ and $e^{1.1561}$, which corresponds with an increase of around 131.73% and an increase of around 317.75%, respectively.

Part D.

```
b1  = 0.7159
sb1 = 0.2246
z = b1 / sb1  # could also get from glm() output...
zz = abs(qnorm(0.05 / 2))
if (z > zz) {
    cat("We reject the null hypothesis\n");
} else {
    cat("We fail to reject the null hypothesis\n");
}
```

```
## We reject the null hypothesis
```