Lab 9

SDS348 Spring 2021

#BRO THIS LAB IS ACTUAL CANCER

#Welcome to Santhosh's Hell Journey. If you make it through this document, you will feel my cold clammy hands around your neck, strangling you in the name of Father Chungus. Let's begin.

```
fit <- glm(am ~ wt, data = mtcars, family = "binomial")
summary(fit)
```

```
##
## Call:
## glm(formula = am ~ wt, family = "binomial", data = mtcars)
##
## Deviance Residuals:
##      Min       1Q    Median        3Q       Max
## -2.11400  -0.53738  -0.08811   0.26055   2.19931
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    12.040      4.510    2.670  0.00759 **
## wt             -4.024      1.436   -2.801  0.00509 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43.230  on 31  degrees of freedom
## Residual deviance: 19.176  on 30  degrees of freedom
## AIC: 23.176
##
## Number of Fisher Scoring iterations: 6
```

#Answer to 1 is yes and the effect is positive. #answer to 1 is C. #2

```
exp(coef(fit))
```

```
##  (Intercept)          wt
## 1.694596e+05 1.788183e-02
```

```
#Answer to 2 is 0.01788183
```

```
percentDecrease <- (0.01799*100-100)
#Answer to 3 is 98.201.
```

```r
fit <- glm(am ~ wt, data = mtcars, family = "binomial")
summary(fit)
```

```
##
## Call:
## glm(formula = am ~ wt, family = "binomial", data = mtcars)
##
## Deviance Residuals:
##      Min       1Q    Median       3Q       Max
## -2.11400  -0.53738  -0.08811   0.26055   2.19931
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    12.040      4.510   2.670  0.00759 **
## wt             -4.024      1.436  -2.801  0.00509 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43.230  on 31  degrees of freedom
## Residual deviance: 19.176  on 30  degrees of freedom
## AIC: 23.176
##
## Number of Fisher Scoring iterations: 6
```

```r
model <- exp(12.040)*exp(-4.024*3)
#Answer to 4 is 0.9685.
```

```r
predict(fit, newdata = data.frame(wt = 3), type = "response")
```

```
##         1
## 0.4921156
```

```r
#Answer to 5 is 0.492.
```

```r
titanic <- read.csv("https://raw.githubusercontent.com/laylaguyot/datasets/main//TitanicSurvival.csv")
titanic <-titanic %>%
  as.data.frame %>%
  mutate(y = ifelse(survived == "yes", 1, 0)) %>%
  rename(class = passengerClass)
```

```r
sucDeezNuts <- glm(y ~ class, data = titanic, family = "binomial")
summary(sucDeezNuts)
```

```
##
## Call:
## glm(formula = y ~ class, family = "binomial", data = titanic)
##
## Deviance Residuals:
```

```
##       Min        1Q     Median        3Q        Max
## -1.3896   -0.7678   -0.7678    0.9791     1.6525
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.4861     0.1146    4.242 2.21e-05 ***
## class2nd      -0.7696     0.1669   -4.611 4.02e-06 ***
## class3rd      -1.5567     0.1433  -10.860  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1741.0  on 1308  degrees of freedom
## Residual deviance: 1613.3  on 1306  degrees of freedom
## AIC: 1619.3
##
## Number of Fisher Scoring iterations: 4
```

```
#1st class is the reference group (Answer to Question 6).
```

```
#For First class
probFirstClass <- exp(0.4861)/(1-exp(0.4861))
probThirdClass <- (exp(0.4861) * exp(-1.5567))/(1-(exp(0.4861)*exp(-1.5568)))
ratio <- probFirstClass/probThirdClass
#Ratio is about 4.743, so go figure. ( answer to question 7 is 4.743).
```

```
#Question 8
full_fit <- glm(y ~ class + age + sex, data = titanic, family = "binomial")
summary(full_fit)
```

```
##
## Call:
## glm(formula = y ~ class + age + sex, family = "binomial", data = titanic)
##
## Deviance Residuals:
##       Min        1Q     Median        3Q        Max
## -2.6399   -0.6979   -0.4336    0.6688     2.3964
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.522074   0.326702   10.781  < 2e-16 ***
## class2nd     -1.280570   0.225538   -5.678 1.36e-08 ***
## class3rd     -2.289661   0.225802  -10.140  < 2e-16 ***
## age          -0.034393   0.006331   -5.433 5.56e-08 ***
## sexmale      -2.497845   0.166037  -15.044  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1414.62  on 1045  degrees of freedom
## Residual deviance:  982.45  on 1041  degrees of freedom
```

```
##   (263 observations deleted due to missingness)
## AIC: 992.45
##
## Number of Fisher Scoring iterations: 4
```

```
predict(full_fit,newdata = data.frame(age = 50, sex = "male", class = "2nd"),type = "response")
```

```
##         1
## 0.1217457
```

#Answer to Question 8 is 0.121.

```
titanic$prob <- predict(full_fit, newdata = titanic, type = "response")
titanic$predicted <- ifelse(titanic$prob > .8, 1, 0)
table(actual = titanic$y, prediction = titanic$predicted)
```

```
##        prediction
## actual   0   1
##      0 612   7
##      1 262 165
```

#previous values were 0.704918 and 0.8400646
#Sensitivity decreases and specificity increases for Question #9
sensitivity
```
mean(titanic[titanic$y == 1,]$prob > .8, na.rm = TRUE)
```

```
## [1] 0.3864169
```

specificity
```
mean(titanic[titanic$y == 0,]$prob <= .8, na.rm = TRUE)
```

```
## [1] 0.9886914
```

#Answer to Question 10 is 0.8, the higher the alpha, the higher the power, the less likelihood of commi

I hope you enjoyed the ride. Feel my hands erase your soul and suck you up like a buttercup . Amen to WongChungus