Homework 9
Syracuse University
IST 772
Summer 2021

```
# load packages
require(car)
require(dplyr)
```

Question 2

```
# use logistic regression to predict vs using gera and hp
cars_model <- glm(formula = vs ~ gear + hp,
                  data = mtcars,
                  family = "binomial")

# summarize the model
summary(cars_model)

##
## Call:
## glm(formula = vs ~ gear + hp, family = "binomial", data = mtcars)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.76095  -0.20263  -0.00889   0.38030   1.37305
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 13.43752    7.18161   1.871   0.0613 .
## gear        -0.96825    1.12809  -0.858   0.3907
## hp          -0.08005    0.03261  -2.455   0.0141 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43.860  on 31  degrees of freedom
## Residual deviance: 16.013  on 29  degrees of freedom
## AIC: 22.013
##
## Number of Fisher Scoring iterations: 7

# convert the log odds into regular odds and interpret them (only
significant)
round(exp(coef(cars_model)), 5)

##  (Intercept)         gear           hp
## 685240.25981      0.37975      0.92307
```

```
# gear is not significant, so it cannot be interpreted in the model. hp is
# significant so for a one unit increase in hp the odds of the cylinder being
# straight shaped changes by .92 : 1, so the odds go down a little bit. On
# the other hand, for a one unit decrease in hp the odds of the cylinder
being
# straight shaped go up a little bit.

# generate and interpret confidence intervals around the coefficients (only
significant)
exp(confint(cars_model))
```

```
##                  2.5 %       97.5 %
## (Intercept) 9.51464612 1.349944e+14
## gear        0.02469715 2.876797e+00
## hp          0.84508732 9.669901e-01
```

Question 5

```
# cant install baylor ed psych due to R version issues.

# package 'BaylorEdPsych' is not available for this version of R (I am using
# the most up to date version of R (4.1))
```

Question 6

```
# inspect the structure of the chile dataset
str(Chile)
```

```
## 'data.frame':    2700 obs. of  8 variables:
##  $ region    : Factor w/ 5 levels "C","M","N","S",..: 3 3 3 3 3 3 3 3 3 3
...
##  $ population: int  175000 175000 175000 175000 175000 175000 175000
175000 175000 175000 ...
##  $ sex       : Factor w/ 2 levels "F","M": 2 2 1 1 1 1 2 1 1 2 ...
##  $ age       : int  65 29 38 49 23 28 26 24 41 41 ...
##  $ education : Factor w/ 3 levels "P","PS","S": 1 2 1 1 3 1 2 3 1 1 ...
##  $ income    : int  35000 7500 15000 35000 35000 7500 35000 15000 15000
15000 ...
##  $ statusquo : num  1.01 -1.3 1.23 -1.03 -1.1 ...
##  $ vote      : Factor w/ 4 levels "A","N","U","Y": 4 2 4 2 2 2 2 2 3 2 ...
```

```
# check for missing values
apply(Chile, 2, function(x) sum(is.na(x)))
```

```
##     region population        sex        age  education     income
statusquo
##          0          0          0          1         11         98
17
##       vote
##        168
```

```
# use only observations with no missing values
ChileNew <- Chile[complete.cases(Chile),]
```

```r
# create the model
chile_model <- glm(formula = vote ~ age + statusquo,
                   data = ChileNew,
                   family = "binomial")

# summarize the model
summary(chile_model)
```

```
##
## Call:
## glm(formula = vote ~ age + statusquo, family = "binomial", data =
ChileNew)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6552   0.3104   0.3752   0.4304   0.5282
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.701599   0.219618   7.748 9.34e-15 ***
## age         0.023830   0.005967   3.994 6.50e-05 ***
## statusquo   0.175440   0.081314   2.158    0.031 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1268.2  on 2430  degrees of freedom
## Residual deviance: 1244.1  on 2428  degrees of freedom
## AIC: 1250.1
##
## Number of Fisher Scoring iterations: 5
```

```r
# interpret the results

# both the age variable and the statusquo variable are statistically
significant.
# For every one unit increase in age, the probability of voting in favor of
# Pinochet increases by a probability of 0.006, whereas for every one unit
# increase in statusquo, the probability of voting in favor of Pinochet
increases
# by a probability of 0.081.
```