Lecture 15 HW

Ellen Chancey

October 28, 2018

Logistic Regression

This homework is based on problems listed in Categorical Data Analysis second edition by Alan Agresti, which can be found here

Question 5.6

Context: There were 23 space shuttle flights before the Challenger mission disaster in 1986. The dataset used here shows the temperature at the time of flight and whether at least one primary o-ring suffered thermal distress.

Read in data

ft : flight number
temp : temp at time of flight
td : thermal distress, yes (1) or no(0)

```
chal <- read.csv("challenger.csv")
chal$td <- as.factor(chal$td)
print(chal)
```

```
##    ft temp td
## 1   1   66  1
## 2   2   70  0
## 3   3   69  0
## 4   4   68  0
## 5   5   67  0
## 6   6   72  0
## 7   7   73  0
## 8   8   70  0
## 9   9   57  1
## 10 10   63  1
## 11 11   70  1
## 12 12   78  0
## 13 13   67  0
## 14 14   53  1
## 15 15   67  0
## 16 16   75  0
## 17 17   70  0
## 18 18   81  0
```

```
## 19 19    76  0
## 20 20    79  0
## 21 21    75  1
## 22 22    76  0
## 23 23    58  1
```

Plot data before modeling

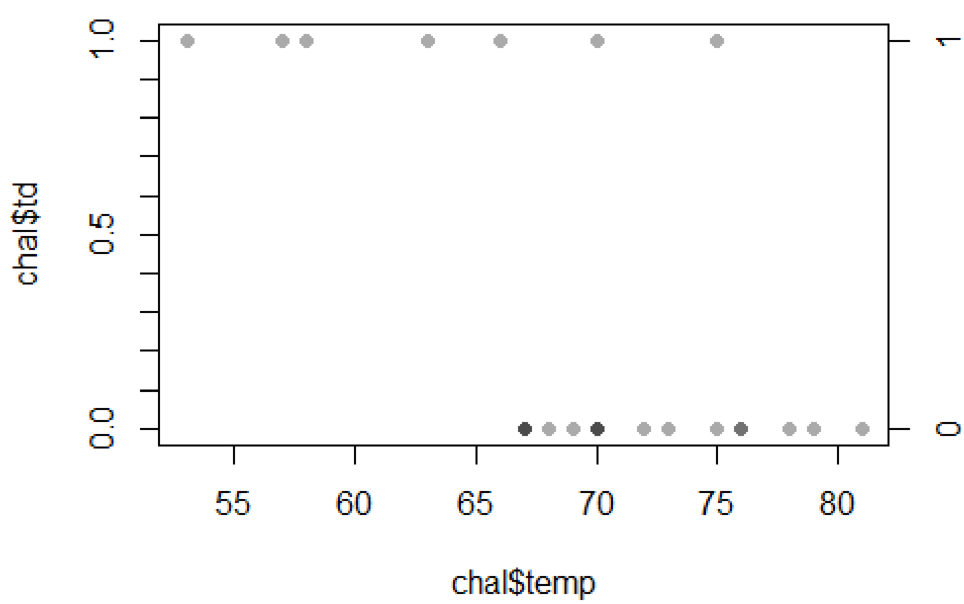
```
require("FSA")

## Loading required package: FSA

## Warning: package 'FSA' was built under R version 3.4.4

## ## FSA v0.8.20. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.

plotBinResp(chal$temp,chal$td, plot.p=FALSE)
```



Part A

Use logistic regression to model the effect of temperature on the probability of thermal distress. Plot a figure of the fitted model. Interpret.

```
model <- glm(td~temp, data = chal, family = "binomial")
summary(model)
```

```
##
## Call:
## glm(formula = td ~ temp, family = "binomial", data = chal)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9914  -0.6869  -0.3024   0.3201   2.3805
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.9598     9.1344   2.076   0.0379 *
## temp         -0.2898     0.1340  -2.162   0.0306 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28.267  on 22  degrees of freedom
## Residual deviance: 18.244  on 21  degrees of freedom
## AIC: 22.244
##
## Number of Fisher Scoring iterations: 5
```

Intercept estimate $\hat{\alpha}$ is 18.9598 with p value of 0.037. The std error, 9.1344 and z stat of 2.076 can be used to calculate confidence intervals for the estimate.

The temperature estimate $\hat{\beta}$ is -0.2898 with a p value of 0.0306. The std error 0.1340 and z stat of -2.162 can be used to calculate confidence intervals for the estimate. The sufficiently small p value indicates that there is evidence to reject the null hypothesis, $H_0: \beta = 0$, meaning there is a statistically significant relationship between temperature and thermal distress.

The model is: $\hat{\Omega}\,(x) = \hat{\alpha} + \hat{\beta}\,x$

The model using model estimates:$\hat{\Omega}\,(x) = 18.96 + 9.13x$

$\hat{\beta}$ indicates the amount that the log odds changes for each one degree increase in temperature. A $\hat{\beta}$ value of 9.13 indicates that for each 1 unit increase in temperature, the log odds increase by 9.13.

Converting to Odds Ratios

```
exp(coef(model))

##  (Intercept)         temp
## 1.714508e+08 7.484375e-01

# exponentiates the coefficients from the object model
```

This indicates that the odds of thermal distress increase by 7.48 for a one unit increase in temperature.

Part B

Estimate the probability of thermal distress at 31 degrees, the temperature at the time of the Challenger flight.

First we estimate the log odds

```
alpha <- model$coefficients[1]
beta <- model$coefficients[2]
estlo <- alpha+(beta*31)
estlo

## (Intercept)
##    9.977011
```

Next convert the log odds to the probability

```
pi <- (exp(estlo))/(1+exp(estlo))
pi

## (Intercept)
##   0.9999535
```

The estimated probability of thermal distress when temperature is 31 is 0.9999535.

Part C

Construct a confidence interval for the effect of temperature on the odds of thermal distress, and test the statistical significance of the effect.

Confidence interval

```
confint.default(model)

##                   2.5 %      97.5 %
## (Intercept)  1.056783 36.86283117
## temp        -0.552459 -0.02707618
```

The statistical signifcance of the effect of temperature on thermal distress can be found as the p value for the temp estimate, and is 0.306.

```
summary(model)

##
## Call:
## glm(formula = td ~ temp, family = "binomial", data = chal)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9914  -0.6869  -0.3024   0.3201   2.3805
##
## Coefficients:
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.9598     9.1344   2.076   0.0379 *
## temp         -0.2898     0.1340  -2.162   0.0306 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28.267  on 22  degrees of freedom
## Residual deviance: 18.244  on 21  degrees of freedom
## AIC: 22.244
##
## Number of Fisher Scoring iterations: 5
```

Extra Content

Calculating confidence intervals

There are two functions that can be used to find confidence intervals, one used standard error, while the other uses log likelihood.

```
confint(model) # uses log likelihood

## Waiting for profiling to be done...

##                  2.5 %      97.5 %
## (Intercept)  5.3056132 43.98615240
## temp        -0.6576272 -0.08965334

confint.default(model) # uses std errors

##                 2.5 %      97.5 %
## (Intercept)  1.056783 36.86283117
## temp        -0.552459 -0.02707618
```

Reference material

https://stats.idre.ucla.edu/r/dae/logit-regression/

Session Information

```
sessionInfo()

## R version 3.4.1 (2017-06-30)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 16299)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
```

```
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] FSA_0.8.20
##
## loaded via a namespace (and not attached):
##  [1] MASS_7.3-47     compiler_3.4.1  backports_1.1.0 plyr_1.8.4
##  [5] magrittr_1.5    rprojroot_1.2   tools_3.4.1     htmltools_0.3.6
##  [9] yaml_2.1.14     Rcpp_0.12.13    stringi_1.1.5   rmarkdown_1.6
## [13] knitr_1.16      stringr_1.2.0   digest_0.6.12   evaluate_0.10.1
```