A short Analysis of R Dataset ex2011

Bhuvnesh Kochhar

Introduction In this short report the analysis of the ex2011 data set is presented. This dataset contains 24 observations on space shuttle launch temperatures (F) as well as information on the shuttles O-ring failure status (Yes/NO) when launched at that temperature (data.temp, Appendix).

The ex2011 data set was first imported from the Sleuth3 library available in R (Appendix). The data was then modelled using a generalized linear model (glm) with a binomial family parameter. This was done to further investigate the relationship between the probability of O-ring failure (y-axis) and launch temperatures (x-axis).

The aim of this report is to:

a) To determine if a relationship exists between the likelihood of O-ring failures and the temperature at launch time.

b) To predict the probability of Failure at 31F.

Results After fitting our glm (log.temp, Appendix) to the data, we can observe a inverse relationship between Failure status and launch temperature. The temperature coefficient was -0.17132 which was statistically significant with a p-value of 0.04. Thus, giving us a equation of log (failure) = 10.87535 - 0.17132*Temperature. Using our model, the probability of failure (prob, Appendix) at 31 F was found to be 0.996 or 99.6%.

Discussion Our analysis has shown that the probability of failure at 31F was almost 100% (99.6%). Furthermore analyzing the temperature coefficient shows that the odds of O-ring failure increase by 18.7% (rate_of_change,Appendix) when temperature falls by 1F which highlights the sensitivity of these shuttle rings to low temperatures.

A major limitation of this data set is the small number of observations. Also, as we can see from Figure1, Appendix, our prediction at 31F falls well outside the range of our fitted values. This combined with the low number of observations means that extrapolation and prediction might be unreliable. The curvature shows the relationship is likely to be non-linear and therefore we cannot fully understand the true range of temperatures where the O-rings are most likely to fail.

Appendix

Import and read data.

```
library(Sleuth3)

data.temp <- ex2011
summary(data.temp)
```

```
##   Temperature    Failure
##  Min.   :53.00   No :17
##  1st Qu.:67.00   Yes: 7
##  Median :70.00
##  Mean   :69.92
##  3rd Qu.:75.25
##  Max.   :81.00
```

Fitting Generalized linear model log.temp.

```
log.temp<-glm(Failure~Temperature,data = data.temp,family = 'binomial')

summary(log.temp)
```

```
##
## Call:
## glm(formula = Failure ~ Temperature, family = "binomial", data = data.temp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2125  -0.8253  -0.4706   0.5907   2.0512
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 10.87535    5.70291   1.907   0.0565 .
## Temperature -0.17132    0.08344  -2.053   0.0400 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28.975  on 23  degrees of freedom
## Residual deviance: 23.030  on 22  degrees of freedom
## AIC: 27.03
##
## Number of Fisher Scoring iterations: 4
```

Rate_of_change

```
exp(-log.temp$coefficients)
```

```
##  (Intercept)  Temperature
## 0.0000189189 1.1868710813
```

```
Rate_of_change = (1.1868710813-1)*100
Rate_of_change
```

```
## [1] 18.68711
```

Probability Calculation at 31F

```
x<-31
y <- (-(0.17132)*(x)) + 10.87535
prob<- exp(y) / (1+exp(y))
prob
```

```
## [1] 0.9961829
```

Figure 1

```
plot(data.temp$Temperature,fitted.values(log.temp),ylab = "Probability of Failure",xlab = "Temperature"
```