Binomial and multinomial regression models

Aylin Mumcular

21 November 2017

Challenger disaster

The 1986 crash of the space shuttle Challenger was linked to failure of O-ring seals in the rocket engines. Data was collected on the 23 previous shuttle missions. The launch temperature on the day of the crash was 31C (J. Faraway citing Presidential Commission on the Space Shuttle Challenger Accident, Vol. 1, 1986: 129-131.)

Let's investigate if the failure of the orings is related to temperature at the time of the mission departure.

#Explore

```
## Parsed with column specification:
## cols(
##   temp = col_double(),
##   damage = col_double()
## )
```

```
## # A tibble: 23 x 2
##     temp damage
##    <dbl>  <dbl>
##  1  9.33      5
##  2 11.1       1
##  3 11.6       1
##  4 13.8       1
##  5 15.1       0
##  6 15.6       0
##  7 15.6       0
##  8 15.6       0
##  9 16         0
## 10 16.4       0
## # ... with 13 more rows
```

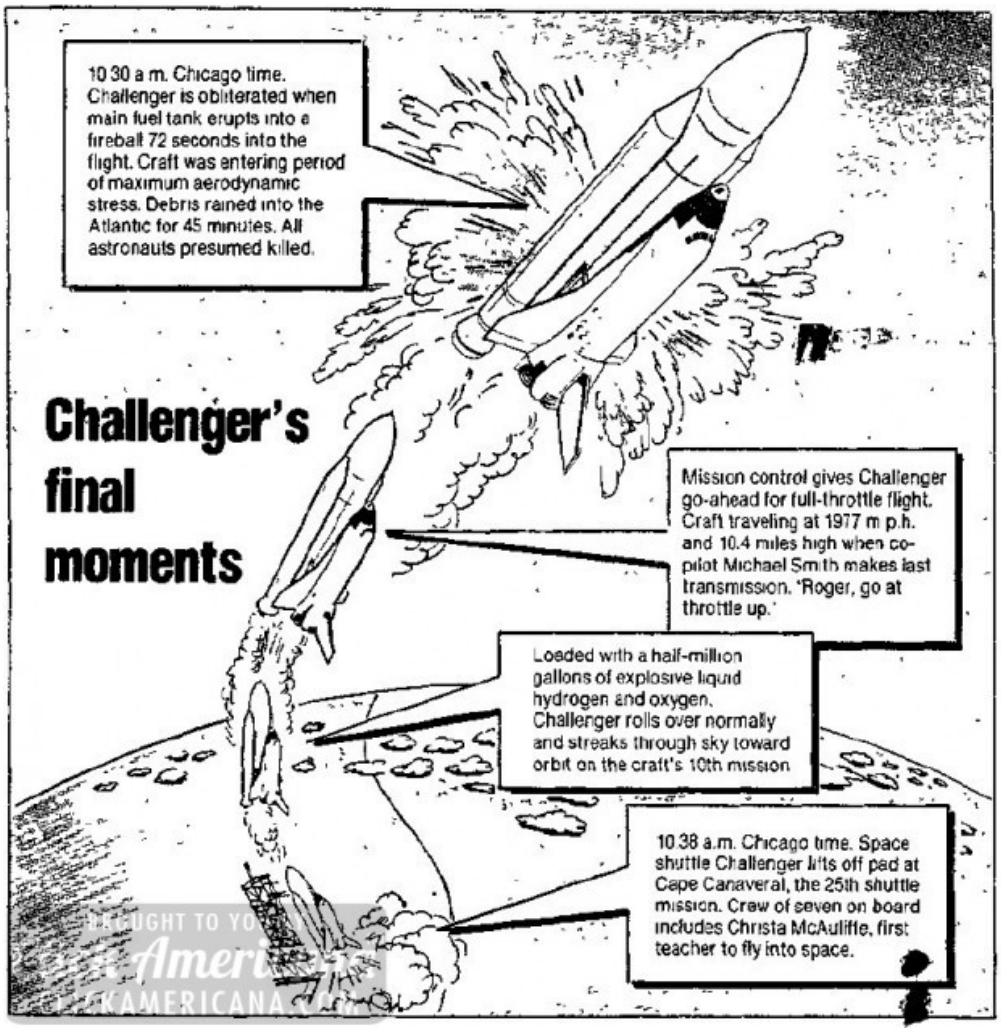


Figure 1: What happened

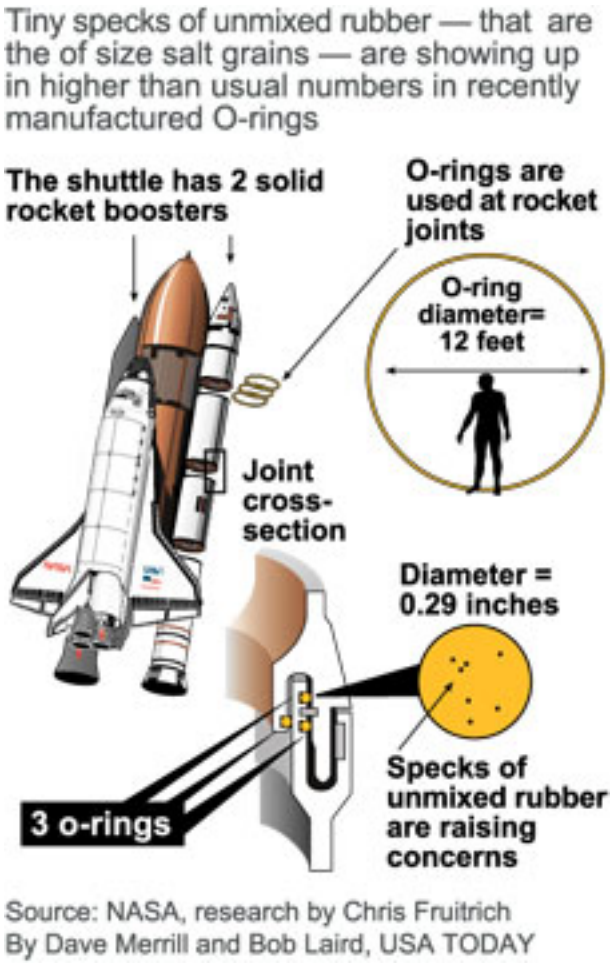
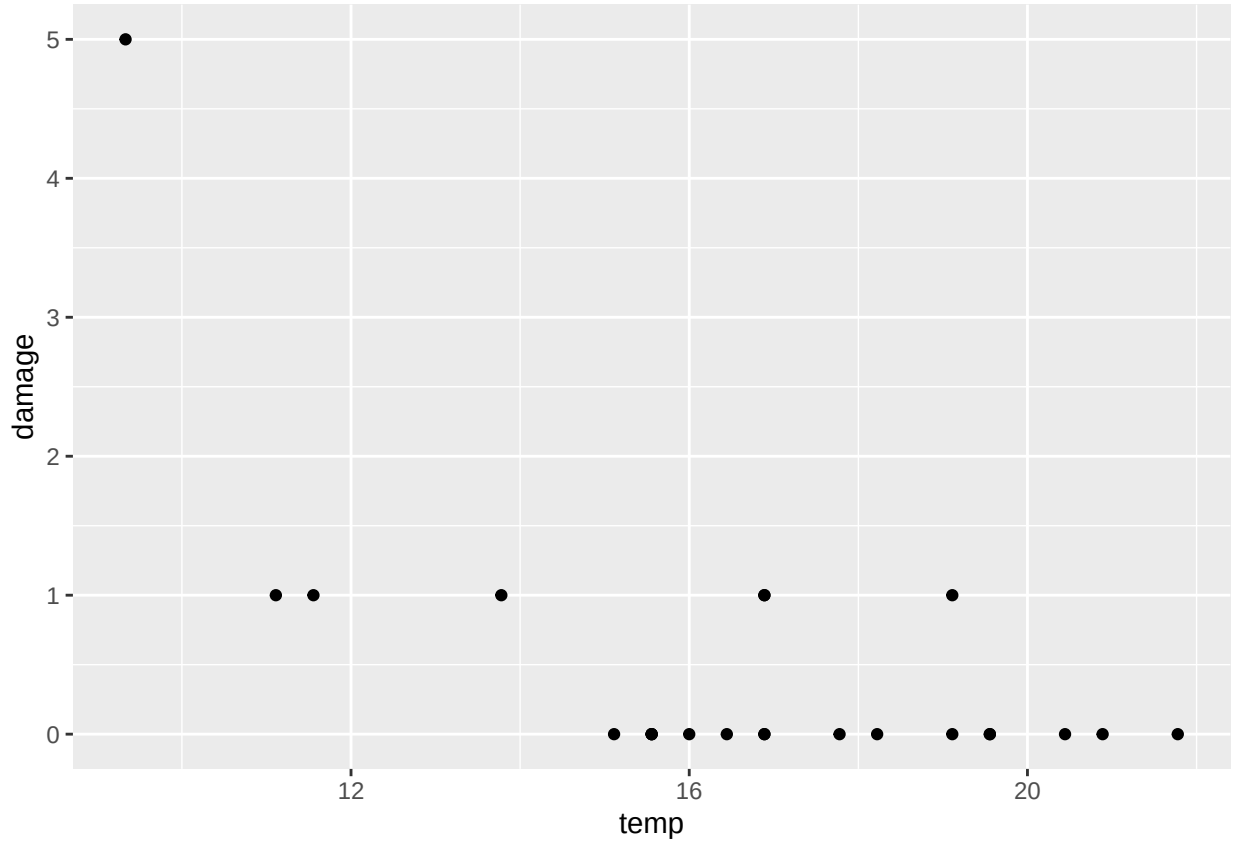


Figure 2: Questioning the role of oring failure in Challenger disaster



Let's fit a binomial regression to the data. To predict the probability of damage, use the following:

```
##
## Call:  glm(formula = cbind(damage, 6 - damage) ~ temp, family = "binomial",
##     data = d)
##
## Coefficients:
## (Intercept)         temp
##      4.7435      -0.4865
##
## Degrees of Freedom: 22 Total (i.e. Null);  21 Residual
## Null Deviance:       38.9
## Residual Deviance: 16.91     AIC: 33.67
```

```
##
## Call:
## glm(formula = cbind(damage, 6 - damage) ~ temp, family = "binomial",
##     data = d)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9529  -0.7345  -0.4393  -0.2079   1.9565
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.7435     1.6154   2.936  0.00332 **
## temp         -0.4865     0.1196  -4.066 4.78e-05 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 38.898  on 22  degrees of freedom
## Residual deviance: 16.912  on 21  degrees of freedom
## AIC: 33.675
##
## Number of Fisher Scoring iterations: 6
```

I checked that deviance equals 16.912 with 21 d.o.f. The pvalue turns out to be 0.7164267.

```
## [1] 0.7164267
```

Because pvalue is greater than 5%, I am unable to reject the current model. (H0:Current model is correct)

Test $H_0 : \beta\text{temp} = 0$. Using Wald test

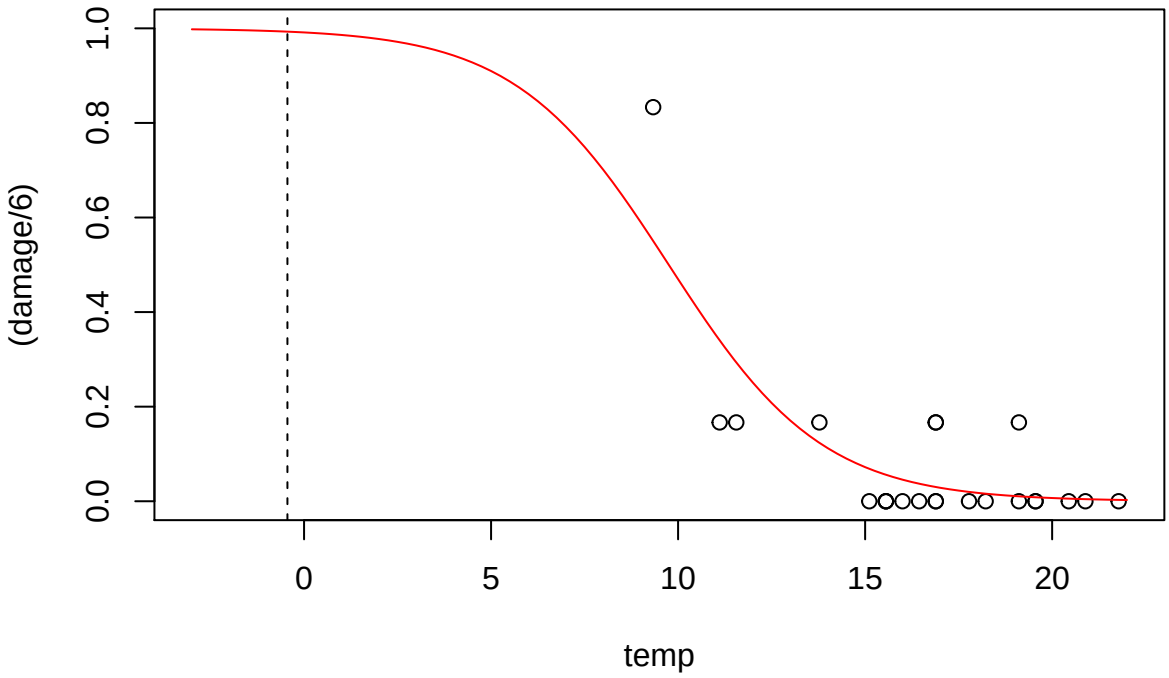
```
##
## Call:
## glm(formula = cbind(damage, 6 - damage) ~ temp, family = "binomial",
##     data = d)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9529  -0.7345  -0.4393  -0.2079   1.9565
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.7435     1.6154   2.936  0.00332 **
## temp         -0.4865     0.1196  -4.066 4.78e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 38.898  on 22  degrees of freedom
## Residual deviance: 16.912  on 21  degrees of freedom
## AIC: 33.675
##
## Number of Fisher Scoring iterations: 6
```

Because pvalue (4.78e-05) of z-test (Wald test) is practically zero, I reject the null hypothesis, which implies that temperature and number of damages are associated. (Coefficient is negative, negatively associated)

```
##         1
## 0.9930342
```

Let's plot the failure probability as a function of temperature.

To model probability of damage with binomial regression, I should feed to glm cbind(damage(success), failures) in other words cbind(damage, 6-damage).



Let's calculate the confidence interval on the damage probability

```
## [1] 0.8355008 0.9930342 0.9997501
```