Case Study 14.1: Decline in expected earthquake numbers with magnitude

Tou Ohone Andate - staff number 1234567

Problem

The Gutenberg-Richter law says that the expected frequency of earthquakes decreases multiplicatively with their magnitude. The formula is

$$\log_{10} N = a - bM,$$

where N is the expected number of earthquakes of magnitude M or more on the Richter scale. Here, a and b are unknown parameters.

After applying a healthy dash of calculus, this formula can be re-expressed in a form that is more familiar to us

$$E[Y|X] = \exp(\beta_0 + \beta_1 x),$$

where Y is the number of number of earthquakes of magnitude between $x - \delta$ and $x + \delta$.[1]

The variables of interest were:

- `Magnitude`: The strength of a recorded earthquake.
- `Locn`: A two-level factor which describe the location of the earthquakes.
  - It has two levels Southern California ("SC") and Washington ("WA").
- `Freq`: The number of earthquakes recorded at `Locn` with magnitude `Magnitude`.

Question of Interest

The research question is to quantify the rate of decrease in earthquake frequency (with increasing magnitude) in both California and Washington states, and to assess whether these rates are the same.

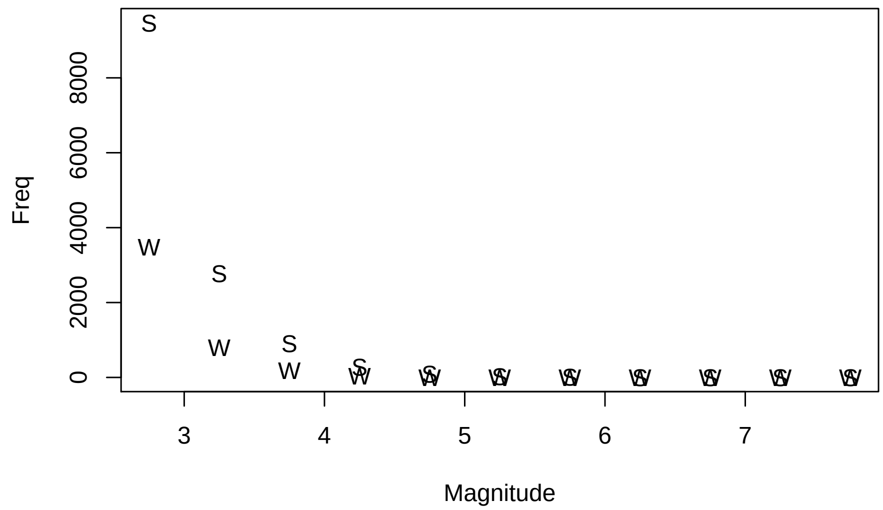
Read in and Inspect the Data

```r
Quakes.df = read.table("EarthquakeMagnitudes.txt", header = TRUE)
head(Quakes.df)
```

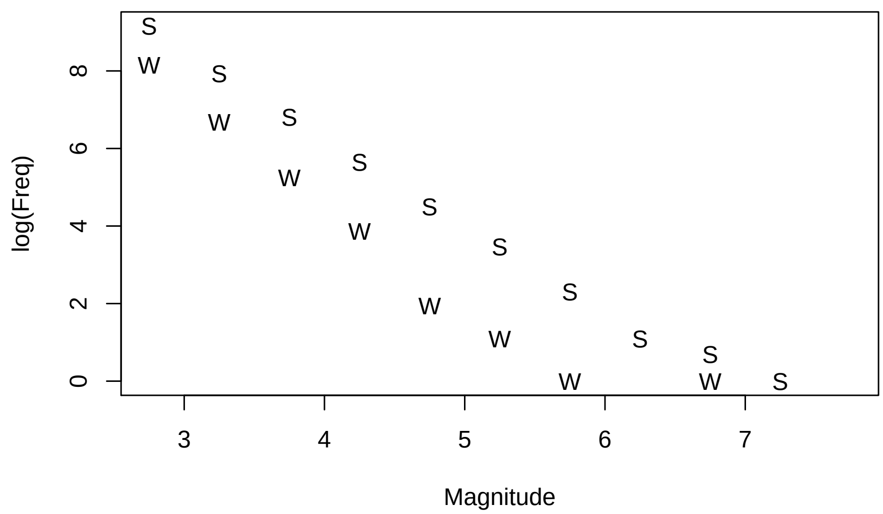
```
##   Locn Magnitude Freq
## 1   SC      2.75 9471
## 2   SC      3.25 2784
## 3   SC      3.75  912
## 4   SC      4.25  285
## 5   SC      4.75   90
## 6   SC      5.25   32
```

```r
plot(Freq ~ Magnitude, data = Quakes.df, pch = substr(Locn, 1, 1))
```

[1] In the above formula, β_0 and β_1 depend on a, b and δ in a complicated way that we are not going to concern ourselves with.



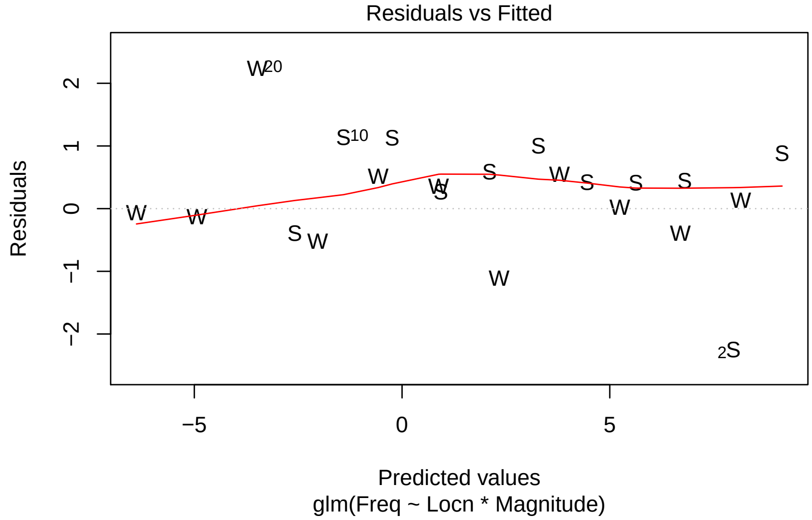

```
plot(log(Freq) ~ Magnitude, data = Quakes.df, pch = substr(Locn, 1, 1))
```



The first plot suggests that there is an exponential decreasing relationship between frequency and magnitude. After logging the frequencies, the second plot suggests a linear relationship between the log frequencies and magnitude. It also noted that there might be different slopes between states.

Model Building and Check Assumptions

```
Quake.gfit = glm(Freq ~ Locn * Magnitude, family = poisson, data = Quakes.df)
plot(Quake.gfit, which = 1, pch = substr(Quakes.df$Locn, 1, 1))
```




```r
summary(Quake.gfit)
```

```
##
## Call:
## glm(formula = Freq ~ Locn * Magnitude, family = poisson, data = Quakes.df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2416  -0.1043   0.3896   0.5838   2.2422
##
## Coefficients:
##                 Estimate Std. Error  z value Pr(>|z|)
## (Intercept)     15.59850    0.06393  243.975  < 2e-16 ***
## LocnWA           0.55610    0.15165    3.667 0.000245 ***
## Magnitude       -2.34605    0.02130 -110.126  < 2e-16 ***
## LocnWA:Magnitude -0.56321    0.05170  -10.895  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 60900.73  on 21  degrees of freedom
## Residual deviance:    17.98  on 18  degrees of freedom
## AIC: 128.41
##
## Number of Fisher Scoring iterations: 4
```

```r
1 - pchisq(17.98, 18)
```

```
## [1] 0.4569709
```

```r
exp(confint(Quake.gfit)[3, ]) # Multiplicative annual change in CA
```

```
## Waiting for profiling to be done...
```

```
##      2.5 %     97.5 %
## 0.09180560 0.09980145
```

```r
100 * (1 - exp(confint(Quake.gfit)[3, ])) # Percentage annual decrease in CA
```

```
## Waiting for profiling to be done...
```

```
##    2.5 %   97.5 %
## 90.81944 90.01986
```

Additional Output with `WA` as the Baseline Level

```r
Quakes.df$Locn = relevel(Quakes.df$Locn, ref = "WA")
Quake2.gfit = glm(Freq ~ Locn * Magnitude, family = poisson, data = Quakes.df)
exp(confint(Quake2.gfit)[3, ]) # Multiplicative annual change in WA
```

```
## Waiting for profiling to be done...
```

```
##      2.5 %     97.5 %
## 0.04965102 0.05972029
```

```r
100 * (1 - exp(confint(Quake2.gfit)[3, ])) # Percentage annual decrease in WA
```

```
## Waiting for profiling to be done...
```

```
##    2.5 %   97.5 %
## 95.03490 94.02797
```

Methods and Assumption Checks

As the response variable, Frequency, is a count, we have fitted a generalised linear model with a Poisson response distribution. We have two explanatory variables: Magnitude (Numeric) and Location (Categoric). The scatterplot of magnitude vs frequency shows an exponentially decreasing trend for both locations. A Poisson model with interaction between magnitude and location was fitted. It shows that the interaction between magnitude and location is highly significant ($P\text{-}value \approx 0$).

All model assumptions were satisfied. We can trust the results from this model ($P\text{-}value = 0.46$).

Our final model is

$$\log(\mu_i) = \beta_0 + \beta_1 \times Locn.WA_i + \beta_2 \times Magnitude_i + \beta_3 \times Locn.WA_i \times Magnitude_i,$$

where μ_i is the mean number of earthquakes in California with magnitude i and $Frequency_i$ (the number of earthquakes with magnitude i) has a Poisson distribution with mean μ_i. Also, $Locn.Wa$ is 1 if the earthquake with magnitude i was in Washington and 0 otherwise.

Executive Summary

The research question is to quantify the rate of decrease in earthquake frequency (with increasing magnitude) in both California and Washington states, and to assess whether these rates are the same.

The rate of decline in the frequency of earthquakes (with increasing magnitude) is more rapid in Washington than California.

In Washington, there is a 94.0 to 95.0% drop in the expected number of earthquakes for a one unit increase in their magnitude on the Richter scale.

In California, the decrease is between 90.0 to 90.8%.