Regression Modeling and Analysis on Earthquakes
Kecheng Xu

Introduction:

Earthquakes are natural disasters that have major economic, political, and health consequences on the regions of the world they affect [1]. Two variables that play a key role in determining the human impact of an earthquake are earthquake magnitude and earthquake depth. Earthquake magnitude is a measure of the amount of energy that an earthquake releases [2]. Earthquake depth is a measure of how deep under the surface of the earth an earthquake occurred [2]. Earthquakes that occur at shallower depths or with larger magnitudes are more likely to cause major damage.

Understanding the relationship of earthquake magnitude to depth can help us characterize earthquakes and their potential for damage. Here we performed an analysis to determine if there was a significant association between earthquake size and earthquake depth. Using exploratory analysis and standard multiple regression techniques we show that there is a significant relationship between earthquake magnitude and depth, even after adjusting for important confounders such as the latitude and longitude at which those earthquakes occur. Our analysis suggests that increased earthquake magnitude is associated with increased depth. Our results suggest that there are relatively few earthquakes that have both the damaging characteristics of large magnitude and shallow depth.

Methods:

Data Collection

For our analysis we used the data on all earthquakes recorded in the U.S. Geological Survey data set from Feb 27 2015 to Mar 19 2015. The data were downloaded from data.gov on Mar 19 2015 using the R programming language [3].

Exploratory Analysis

Exploratory analysis was performed by examining tables and plots of the observed data. We identified transformations to perform on the raw data on the basis of plots and knowledge of the scale of measured variables. Exploratory analysis was used to (1) identify missing values, (2) verify the quality of the data, and (3) determine the terms used in the regression model relating earthquake magnitude to earthquake depth.

Statistical Modeling

To relate earthquake magnitude to earthquake depth we performed a standard multivariate linear regression model [4]. Model selection was performed on the basis of our exploratory analysis and prior knowledge of the relationship between geography and earthquake magnitude. Coefficients were estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations [5].

Reproducibility

All analyses performed in this manuscript are reproduced in the R markdown file regression_model_analysis_earthquakes.Rmd [6]. To reproduce the exact results presented in this manuscript the cached version of the analysis must be performed, as the data available from data.gov changes based on the date.

Results:

The earthquakes data used in this analysis contains information on the time of the earthquake(TIME), the number of sites that measured the earthquake (NST), the location - longitude (Lon) and latitude (Lat), the magnitude (EM, measured on the Richter scale) and the depth (ED, measured in kilometers from the surface) [7]. We identified no missing values in the data set we collected and all measured variables were observed to be inside the standard ranges. Earthquakes in this data set also did not seem to show major patterns over time in magnitude or depth.

Most earthquakes had a small (less than 3 on the Richter scale – 97.88% of quakes) or medium (3-5 on the Richter scale – 1.86% of quakes) magnitude. The distribution of earthquake depths was heavily right skewed. Based on the distribution of the earthquake depths we recognized that a transform was necessary to improve the performance of linear regression techniques; we performed a log base 10 transform of the earthquake depths. Subsequent analyses focus on this transformed depth variable.

We first fit a regression model relating earthquake magnitude to earthquake depth. The residuals showed patterns of non-random variation. We attempted to explain those patterns by fitting models including potential confounders. Our final regression model was:

$$EM = b_0 + b_1 \log10(ED) + f(Lat) + g(Lon) + h(NST) + e$$

where b_0 is an intercept term and b_1 represents the change in earthquake magnitude on the Richter scale associated with a change of 1 unit in log base 10 kilometers for earthquakes at the same latitude, longitude, and measured by the same number of stations. The terms f(Lat), g(Lon), and h(NST) represent factor models with 5 different levels each for latitude, longitude, and number of measurement sites. The error term e represents all sources of unmeasured and unmodeled random variation in earthquake

magnitude. Our final regression model appeared to remove most of the non-random patterns of variation in the residuals.

We observed a highly statistically significant (P < 2e-16) association between earthquake magnitude and earthquake depth. A change of one unit in log base 10 kilometers of depth corresponded to a change of $b_1 = 0.81$ on the Richter scale (95% Confidence Interval: 0.73, 0.90). So for example, for two earthquakes at the same latitude, longitude, and measured at the same number of stations, we would expect an earthquake measured at a depth of 10 kilometers to be 0.81 units higher on the Richter scale than one measured at a depth of 1 kilometer. We would expect the same difference between earthquakes measured at depths of 10 and 100 kilometers and so forth.

Conclusions:

Our analysis suggests that there is a significant, positive association between earthquake magnitude and earthquake depth. Our analysis estimates the relationship using a linear model relating log base 10 depth to magnitude. Therefore, our estimated association is not linear on the depth scale. Even so, there appears to be a strong relationship between the two variables. We also observed that other variables such as latitude, longitude and number of observing stations are associated with both earthquake magnitude and earthquake depth. Including these variables in the regression model relating magnitude to depth improves the model fit, but does not remove the significant positive relationship between the variables.

While our analysis is an interesting first step it is based on a limited sample of earthquakes from only one month in February and March 2015. A larger collection of representative earthquakes may be more appropriate for understanding the relationship between earthquake magnitude and depth. Our analysis may be of interest to scientists seeking to better understand earthquakes. But for policy makers, an important - and difficult - area of future research would be to develop accurate predictive models of earthquake depth and magnitude that can be used to evaluate the potential human consequence of these natural disasters.

References

1. Wikipedia "Earthquake" Page. URL: http://en.wikipedia.org/wiki/Earthquake. Accessed 03/19/2015.
2. USGS "Measuring the size of an earthquake" Page. URL: http://earthquake.usgs.gov/learn/topics/measure.php. Accessed 03/19/2015
3. R Core Team (2012). "R: A language and environment for statistical computing." URL: http://www.R-project.org
4. Seber, George AF, and Alan J. Lee. *Linear regression analysis*. Vol. 936. Wiley, 2012.
5. Ferguson, Thomas S. *A Course in Large Sample Theory: Texts in Statistical Science*. Vol. 38. Chapman & Hall/CRC, 1996.
6. R Markdown Page. URL: http://www.rstudio.com/ide/docs/authoring/using_markdown. Accessed 03/19/2015
7. USGS Real Time Data and Feeds Page. URL: http://earthquake.usgs.gov/earthquakes/feed/. Accessed 03/19/2015.