Pediaditis Georgios

2610368

Data Management and Analytics using SAS

Academic Year 2001-21

DMAS Assessment 3: Lab Report

Introduction

Earthquakes are devastating events that often cause structural damage and loss of lives. Being able to understand the characteristics of earthquakes can lead to better forecasting, to allow us to prepare and react more effectively. Richter magnitude scale was developed in order to quantify the strength of an earthquake. Richter magnitude is just one measure of earthquake strength.

In this report we will examine a dataset that contains 23741 records from earthquakes over several years. We assume earthquakes are independent events.

Our dataset contains the following variables.

id - ID of record (Numeric)

lat - Latitude of earthquake (Numeric)

Long - Longitude of earthquake (Numeric)

dist - Distance travelled by earthquake in a particular direction (Numeric)

depth - Depth of earthquake (Numeric)

md - Magnitude of earthquake, estimated from the duration of seismic wave-train (Numeric)

richter - Intensity of earthquake (Numeric)

mw - Moment magnitude scale value of earthquake (Numeric)

ms - Surface wave magnitude scale value of earthquake (Numeric)

mb - Bodywave magnitude value, measured using P-waves (Numeric)

country - Country of earthquake (Character)

direction - Direction of earthquake (Character)

The first step is to view our data from earthquake dataset. We notice that the variable *mw* has a lot of empty values. We assume that the observations with missing values are random and we omit these observations.

For the purpose of this report, we will also need a new variable xm which is the maximum of md, mw, ms, mb and richter. We add a new column to our dataset called xm and our new summarized dataset is.

Summary of earthquake dataset

Variable	Label	Minimum	Mean	Median	Maximum
id	**ID**	21.00	12258.71	10939.50	23741.00
lat	**Latitude**	30.29	38.00	38.18	45.75
long	**Longitude**	18.34	30.47	27.85	48.00
dist	**Distance (km)**	0.10	3.05	2.30	81.60
depth	**Depth (km)**	1.00	27.36	17.05	225.00
md	**Magnitude**	0.00	2.91	4.30	7.20
richter	**Richter**	0.00	4.34	4.50	7.20
mw	**Moment Magnitude**	0.00	4.48	4.70	7.70
ms	**Surface-wave Magnitude**	0.00	2.89	4.20	7.90
mb	**Bodywave Magnitude**	0.00	2.95	4.40	7.10
xm	**Maximum Magnitude**	3.40	4.63	4.70	7.90

As we can see from the graph the data from the dataset give us the following information:

- Latitude
  Our data are limited from 30 to 46 degrees of latitude.
- Longitude
  Our data are limited from 18 to 48 degrees of longitude. In comparison with Latitude our data cover south and South east Europe to West Asia.
- The variables md, ms, mb range from 0 to less than 8.

Our main objectives for this report are:

a. We should investigate if there is evidence that the average value of xm is different than 4.1? (xm is the maximum of md mw ms mb and Richter.)
b. To find out if there is a difference between countries on the average moment magnitude scale.
c. Create a model for Richter based on other attributes of earthquake.
d. Classify an earthquake as serious or not.
e. Compare model from d) with a model with response variable serious and explanatory variable xm.

Exploratory analysis

The CORR Procedure

10 Variables:	lat long dist depth md richter mw ms mb xm

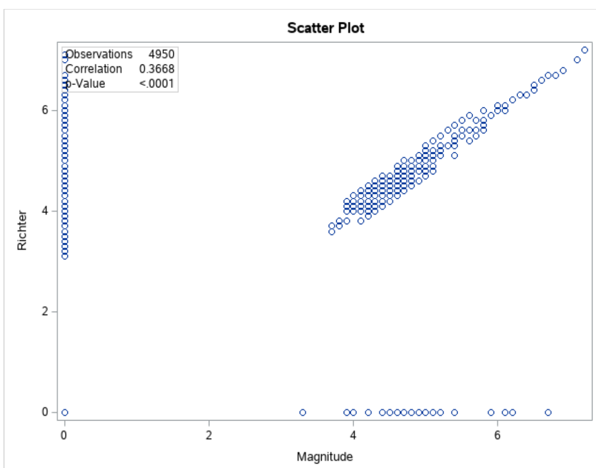
Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
lat	4950	37.99586	2.45495	188080	30.29000	45.75000	Latitude
long	4950	-0.71670	0.10461	-3548	-1.18286	-0.51893	Longitude
dist	1731	3.04697	4.17998	5274	0.10000	81.60000	Distance (km)
depth	4950	27.35826	27.09753	135423	1.00000	225.00000	Depth (km)
md	4950	2.90602	2.35268	14385	0	7.20000	Magnitude
richter	4950	4.34327	0.92002	21499	0	7.20000	Richter
mw	4950	4.47758	1.04875	22164	0	7.70000	Moment Magnitude
ms	4950	2.89269	2.35988	14319	0	7.90000	Surface-wave Magnitude
mb	4950	2.95248	2.34734	14615	0	7.10000	Bodywave Magnitude
xm	4950	4.63489	0.67495	22943	3.40000	7.90000	Maximum Magnitude

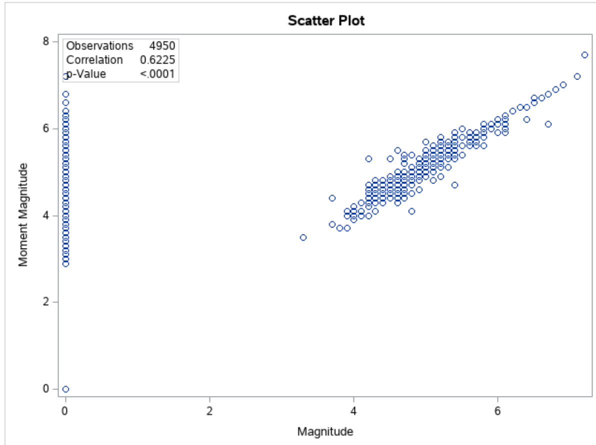
Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations										
	lat	**long**	**dist**	**depth**	**md**	**richter**	**mw**	**ms**	**mb**	**xm**
lat **Latitude**	1.00000 4950	0.28815 <.0001 4950	0.00977 0.6847 1731	-0.17720 <.0001 4950	0.09940 <.0001 4950	0.09491 <.0001 4950	0.04190 0.0032 4950	0.10291 <.0001 4950	0.09813 <.0001 4950	0.06653 <.0001 4950
long **Longitude**	0.28815 <.0001 4950	1.00000 4950	0.12425 <.0001 1731	0.00100 0.9439 4950	0.14142 <.0001 4950	0.10593 <.0001 4950	0.12292 <.0001 4950	0.13601 <.0001 4950	0.14272 <.0001 4950	0.09767 <.0001 4950
dist **Distance (km)**	0.00977 0.6847 1731	0.12425 <.0001 1731	1.00000 1731	0.01436 0.5503 1731	0.02580 0.2834 1731	0.02787 0.2465 1731	-0.01180 0.6238 1731	0.01423 0.5540 1731	0.02010 0.4032 1731	0.03309 0.1688 1731
depth **Depth (km)**	-0.17720 <.0001 4950	0.00100 0.9439 4950	0.01436 0.5503 1731	1.00000 4950	0.31697 <.0001 4950	0.11167 <.0001 4950	0.21437 <.0001 4950	0.31695 <.0001 4950	0.32161 <.0001 4950	0.23645 <.0001 4950
md **Magnitude**	0.09940 <.0001 4950	0.14142 <.0001 4950	0.02580 0.2834 1731	0.31697 <.0001 4950	1.00000 4950	0.36681 <.0001 4950	0.62248 <.0001 4950	0.97209 <.0001 4950	0.96499 <.0001 4950	0.70894 <.0001 4950
richter **Richter**	0.09491 <.0001 4950	0.10593 <.0001 4950	0.02787 0.2465 1731	0.11167 <.0001 4950	0.36681 <.0001 4950	1.00000 4950	0.44387 <.0001 4950	0.41411 <.0001 4950	0.38686 <.0001 4950	0.59681 <.0001 4950
mw **Moment Magnitude**	0.04190 0.0032 4950	0.12292 <.0001 4950	-0.01180 0.6238 1731	0.21437 <.0001 4950	0.62248 <.0001 4950	0.44387 <.0001 4950	1.00000 4950	0.63981 <.0001 4950	0.62789 <.0001 4950	0.78142 <.0001 4950
ms **Surface-wave Magnitude**	0.10291 <.0001 4950	0.13601 <.0001 4950	0.01423 0.5540 1731	0.31695 <.0001 4950	0.97209 <.0001 4950	0.41411 <.0001 4950	0.63981 <.0001 4950	1.00000 4950	0.98036 <.0001 4950	0.74343 <.0001 4950
mb **Bodywave Magnitude**	0.09813 <.0001 4950	0.14272 <.0001 4950	0.02010 0.4032 1731	0.32161 <.0001 4950	0.96499 <.0001 4950	0.38686 <.0001 4950	0.62789 <.0001 4950	0.98036 <.0001 4950	1.00000 4950	0.71876 <.0001 4950
xm **Maximum**	0.06653 <.0001 4950	0.09767 <.0001 4950	0.03309 0.1688 1731	0.23645 <.0001 4950	0.70894 <.0001 4950	0.59681 <.0001 4950	0.78142 <.0001 4950	0.74343 <.0001 4950	0.71876 <.0001 4950	1.00000 4950

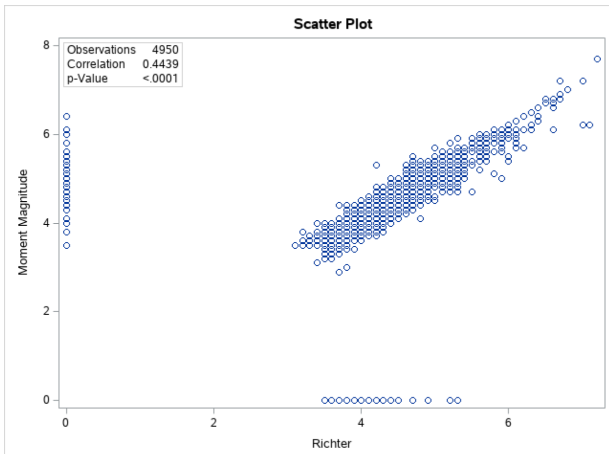
Magnitude										

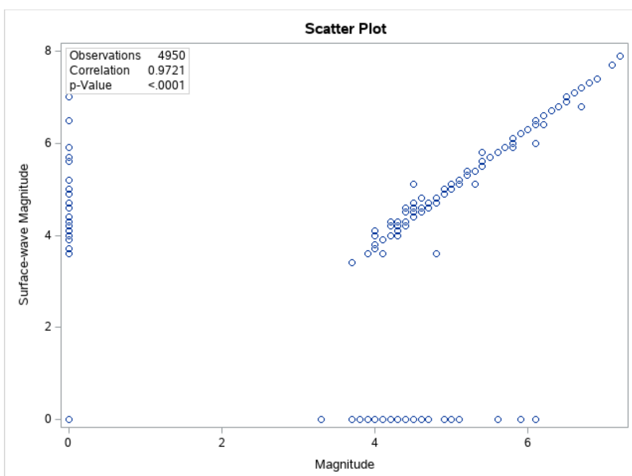
From the above table we can see that we can reject that correlation is equal to 0 in the following pairs:

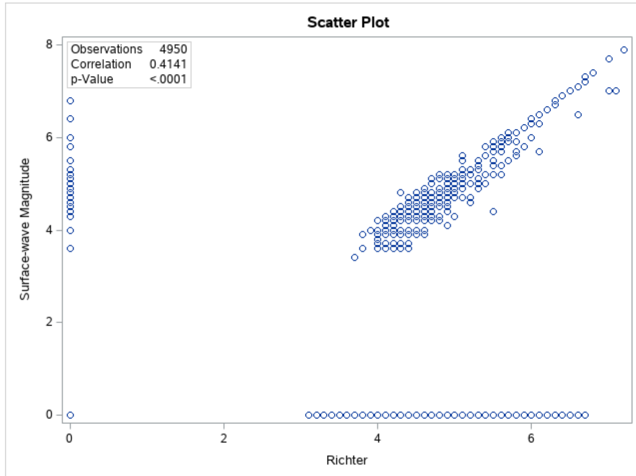
- Lat (Latitude) with Longitude, Depth, Magnitude, Richter, Moment Magnitude, Surface-wave Magnitude, Bodywave Magnitude, Maximum Magnitude.
- Long (Longitude) with Latitude, Distance, Magnitude, Richter, Moment Magnitude, Surface-wave Magnitude, Bodywave Magnitude, Maximum Magnitude.
- Distance (km) with Longitude,
- Depth (km) with Latitude, , Magnitude, Richter, Moment Magnitude, Surface-wave Magnitude, Bodywave Magnitude, Maximum Magnitude.
- Magnitude with Latitude, Longitude, Depth, Richter, Moment Magnitude, Surface-wave Magnitude, Bodywave Magnitude, Maximum Magnitude.
- Richter with Latitude, Longitude, Depth, Magnitude, Moment Magnitude, Surface-wave Magnitude, Bodywave Magnitude, Maximum Magnitude.
- Moment Magnitude (mw) with Latitude, Longitude, Depth, Magnitude, Richter, Surface-wave Magnitude, Bodywave Magnitude, Maximum Magnitude.
- Surface-wave Magnitude (ms) with Latitude, Longitude, Depth, Magnitude, Richter, Moment Magnitude, Bodywave Magnitude, Maximum Magnitude.
- Bodywave Magnitude (mb) with Latitude, Longitude, Depth, Magnitude, Richter, Moment Magnitude, Surface-wave Magnitude, Maximum Magnitude.
- Maximum Magnitude (xm) with Latitude, Longitude, Depth, Magnitude, Richter, Moment Magnitude, Surface-wave Magnitude, Bodywave Magnitude.

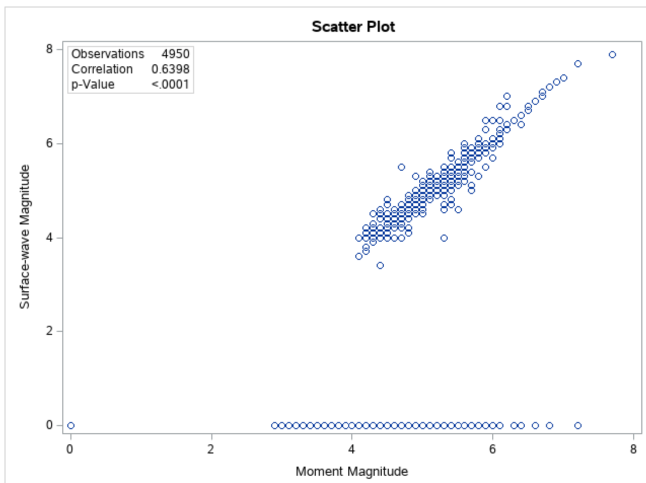


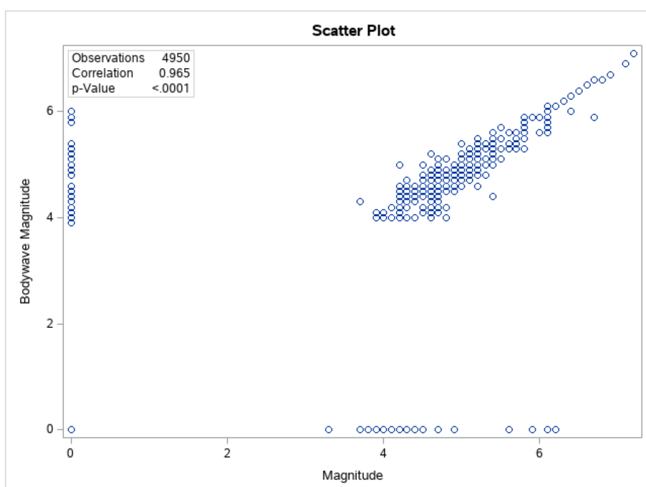

Scatter Plot
Observations 4950
Correlation 0.6225
p-Value <.0001
Moment Magnitude
Magnitude
0
2
4
6
8

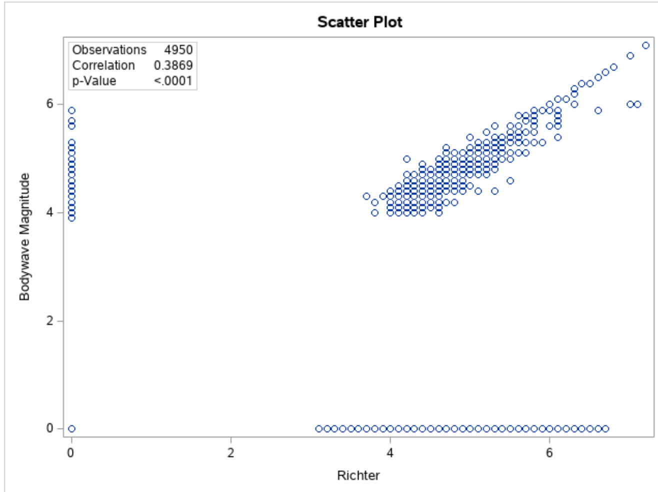

Scatter Plot
Observations 4950
Correlation 0.4439
p-Value <.0001
Moment Magnitude
Richter
0
2
4
6
8

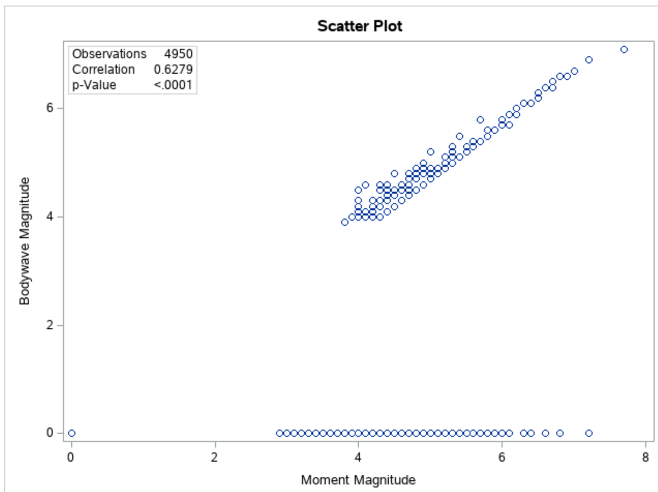

Scatter Plot
Observations 4950
Correlation 0.9721
p-Value <.0001
Surface-wave Magnitude
Magnitude
0
2
4
6
8

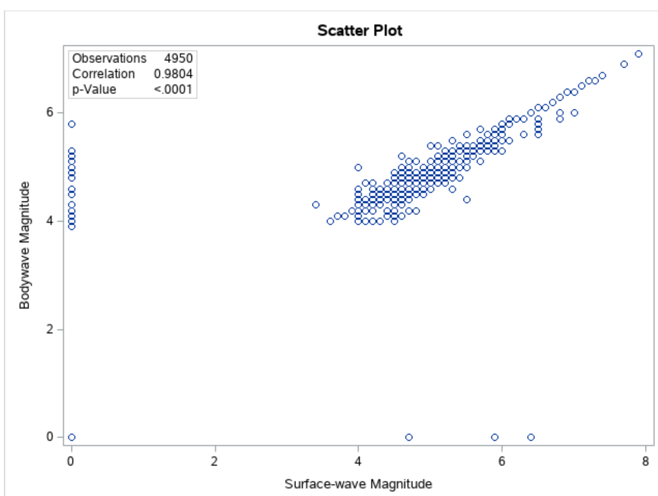

Scatter Plot
Observations 4950
Correlation 0.4141
p-Value <.0001
Surface-wave Magnitude
Richter
0
2
4
6
8

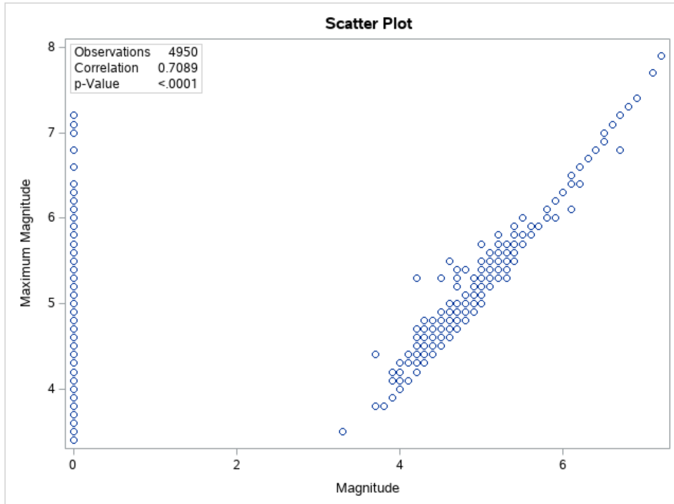

Scatter Plot
Observations 4950
Correlation 0.6398
p-Value <.0001
Surface-wave Magnitude
Moment Magnitude
0
2
4
6
8

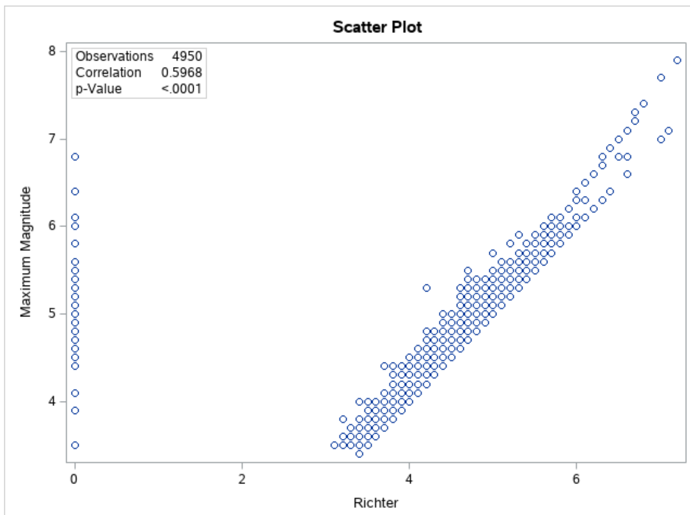

Scatter Plot
Observations 4950
Correlation 0.965
p-Value <.0001
Bodywave Magnitude
Magnitude
0
2
4
6

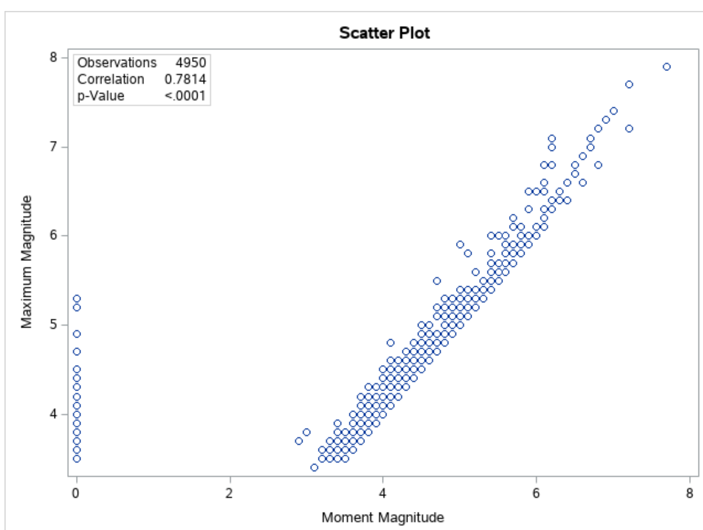

Scatter Plot
Observations 4950
Correlation 0.3869
p-Value <.0001
Bodywave Magnitude
Richter

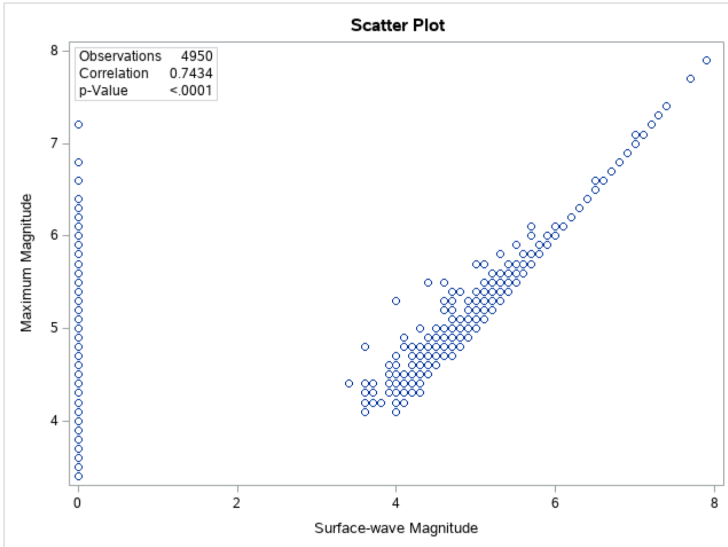

Scatter Plot
Observations 4950
Correlation 0.6279
p-Value <.0001
Bodywave Magnitude
Moment Magnitude

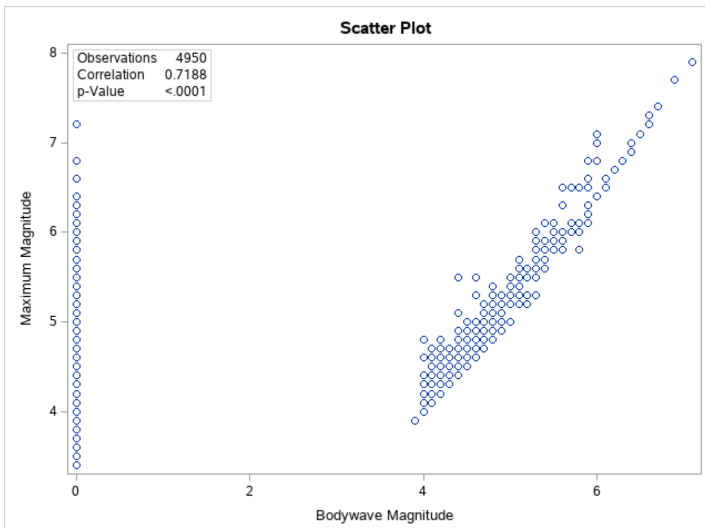

Scatter Plot
Observations 4950
Correlation 0.9804
p-Value <.0001
Bodywave Magnitude
Surface-wave Magnitude


Scatter Plot
Observations 4950
Correlation 0.7089
p-Value <.0001
Maximum Magnitude
Magnitude
0
2
4
6
8
7
6
5
4


Scatter Plot
Observations 4950
Correlation 0.5968
p-Value <.0001
Maximum Magnitude
Richter
0
2
4
6
8
7
6
5
4


Scatter Plot
Observations 4950
Correlation 0.7814
p-Value <.0001
Maximum Magnitude
Moment Magnitude
0
2
4
6
8
8
7
6
5
4

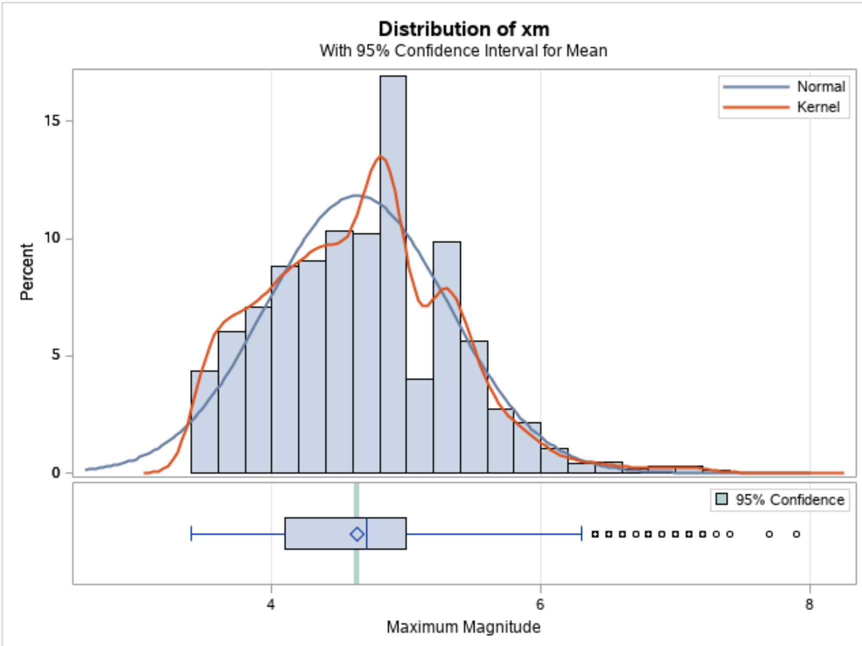


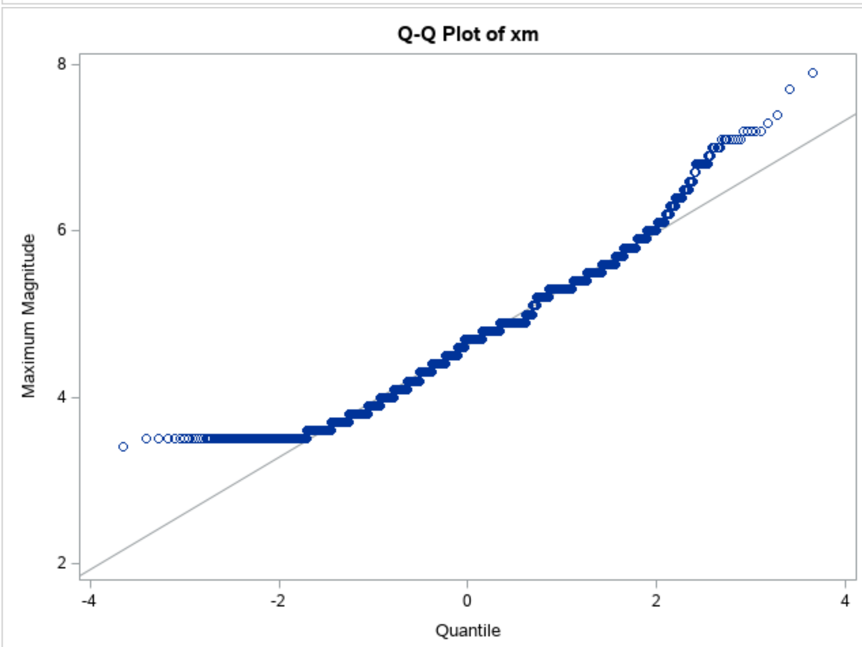


From the above plots and from the correlation matrix we can see variables that seems to corelate with each other. On most of the above plots we can also see a number of points that seems to have no correlation between the two variables of the plot. That might be a mistake of our data or outliers.

Formal Analysis

a.We start by performing a one sample t-test on xm with Null hypothesis that xm=4.1





Before we proceed with the t-test we make the following assumptions.

1. Observations are independent
2. Xm follows a normal distribution.

As we can see from the plots in the previous page xm seems to follow a normal distribution so our assumption about independence are true based on our observations.

One sample t-test. Is xm different to 4.1?

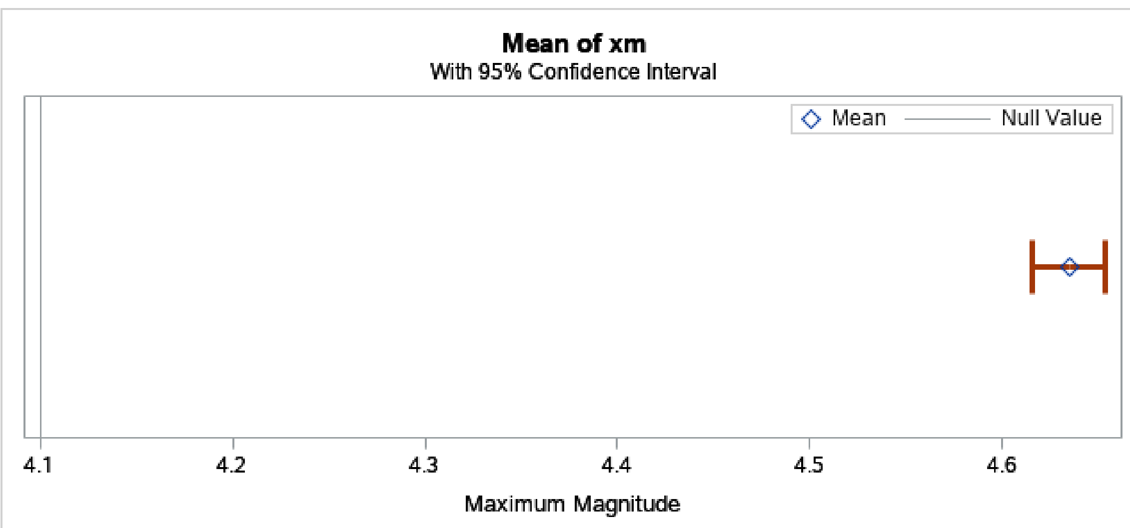
The TTEST Procedure

Variable: xm (Maximum Magnitude)

N	Mean	Std Dev	Std Err	Minimum	Maximum
4950	4.6349	0.6750	0.00959	3.4000	7.9000
Mean	**95% CL Mean**		**Std Dev**	**95% CL Std Dev**	
4.6349	4.6161	4.6537	0.6750	0.6619	0.6885
DF		**t Value**		**Pr > |t|**	
4949		55.76		<.0001	

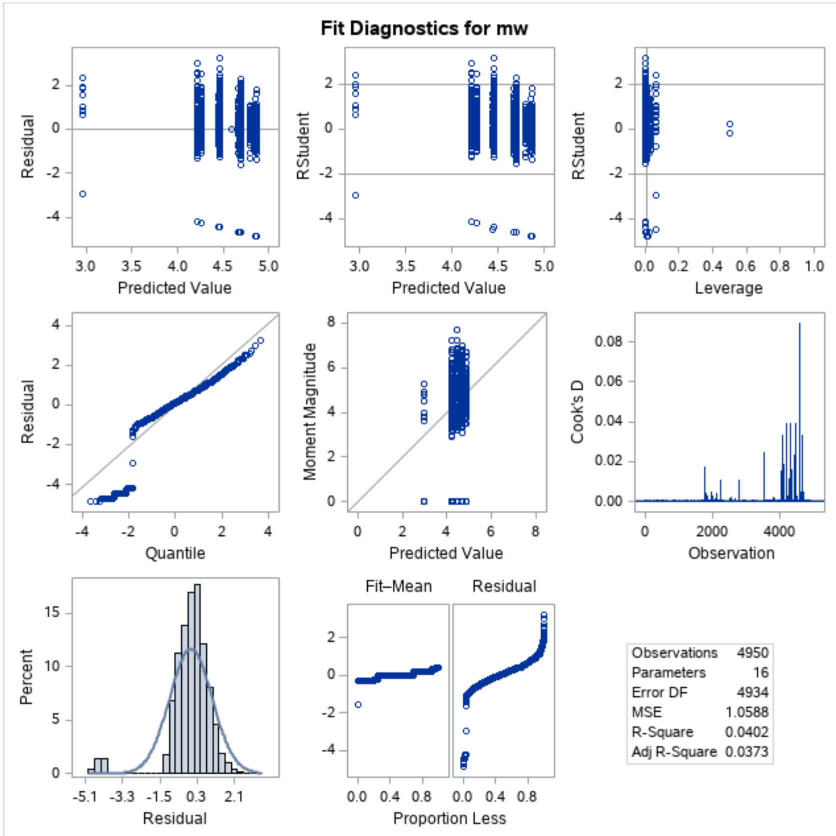
As we can see from the previous table we can reject H0 hypothesis that xm average is 4.1.

We can also see that from the 95% confidence interval.



b.In order to find if there is a difference in the average moment magnitude scale value of an earthquake between countries we can apply an anova.

First we check our assumptions



Based on the plots on the previous page we can see that the distribution is a normal distribution and the Q-Q plot lie on a diagonal line although it has some outliers.

Applying one way ANOVA on our data we have.

One Way ANOVA with Country as explanatory

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
country	16	aegean_sea albania azerbaijan blacksea bulgaria egypt georgia greece iran iraq macedonia mediterranean romania russia syria turkey

One Way ANOVA with Country as explanatory

The GLM Procedure

Dependent Variable: mw Moment Magnitude

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	218.931125	14.595408	13.78	<.0001
Error	4934	5224.339784	1.058845		
Corrected Total	4949	5443.270909			

R-Square	Coeff Var	Root MSE	mw Mean
0.040221	22.98123	1.029002	4.477576

Source	DF	Type I SS	Mean Square	F Value	Pr > F
country	15	218.9311248	14.5954083	13.78	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	15	218.9311248	14.5954083	13.78	<.0001

From the tables and with significance level 5% there is enough evidence to reject our null hypothesis that across different countries mw is equal.

c. Next, we try to fit a regression with Richter as a response and all the other variables as potential explanatory variables (excluding id and xm)

The GLMSELECT Procedure

Forward Selection Summary					
Step	**Effect Entered**	**Number Effects In**	**Number Parms In**	**AIC**	**SBC**
*** Optimal Value of Criterion**					
0	**Intercept**	1	1	1315.9747	-411.5689
1	**mw**	2	2	850.2743	-871.8128
2	**ms**	3	3	814.5323	-902.0983
3	**md**	4	4	731.2843	-979.8898
4	**lat**	5	5	723.9075	-981.8102
5	**long**	6	6	717.7797*	-982.4816*

The selected model is the model at the last step (Step 5).

Effects:	Intercept lat long md mw ms
Root MSE	0.74455
Dependent Mean	4.41236
R-Square	0.2963
Adj R-Sq	0.2942
AIC	717.77967
AICC	717.84468
SBC	-982.48160

Parameter Estimates				
Parameter	**DF**	**Estimate**	**Standard Error**	**t Value**
Intercept	1	0.587929	0.574883	1.02
lat	1	0.042659	0.014445	2.95
long	1	0.009208	0.003232	2.85
md	1	-0.244064	0.025751	-9.48
mw	1	0.369401	0.025703	14.37
ms	1	0.295062	0.026465	11.15

Richter = 0.587929 + 0.042659*lat + 0.009208*long - 0.24406*md + 0.369401*mw + 0.295062*ms

d. An earthquake with Richter over 5 is considered a serious earthquake. We wish to make a regression model with serious as a response variable and with the other variables as potential explanatory variables. (excluding id, mw, richter, xm)

Summary of Forward Selection						
Step	**Effect Entered**	**DF**	**Number In**	**Score Chi-Square**	**Pr > ChiSq**	**Variable Label**
1	**ms**	1	1	151.1198	<.0001	Surface-wave Magnitude
2	**md**	1	2	38.7273	<.0001	Magnitude
3	**mb**	1	3	22.9398	<.0001	Bodywave Magnitude
4	**depth**	1	4	3.5886	0.0582	Depth (km)
5	**long**	1	5	3.1335	0.0767	Longitude

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
long	1	3.1236	0.0772
depth	1	4.2655	0.0389
md	1	26.6632	<.0001
ms	1	66.3388	<.0001
mb	1	31.0878	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-5.7708	1.5772	13.3879	0.0003
long	1	1.8177	1.0285	3.1236	0.0772
depth	1	-0.00754	0.00365	4.2655	0.0389
md	1	-0.6810	0.1319	26.6632	<.0001
ms	1	3.3795	0.4149	66.3388	<.0001
mb	1	-2.2714	0.4074	31.0878	<.0001

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
long	6.158	0.820	46.222
depth	0.992	0.985	1.000
md	0.506	0.391	0.655
ms	29.355	13.017	66.202
mb	0.103	0.046	0.229

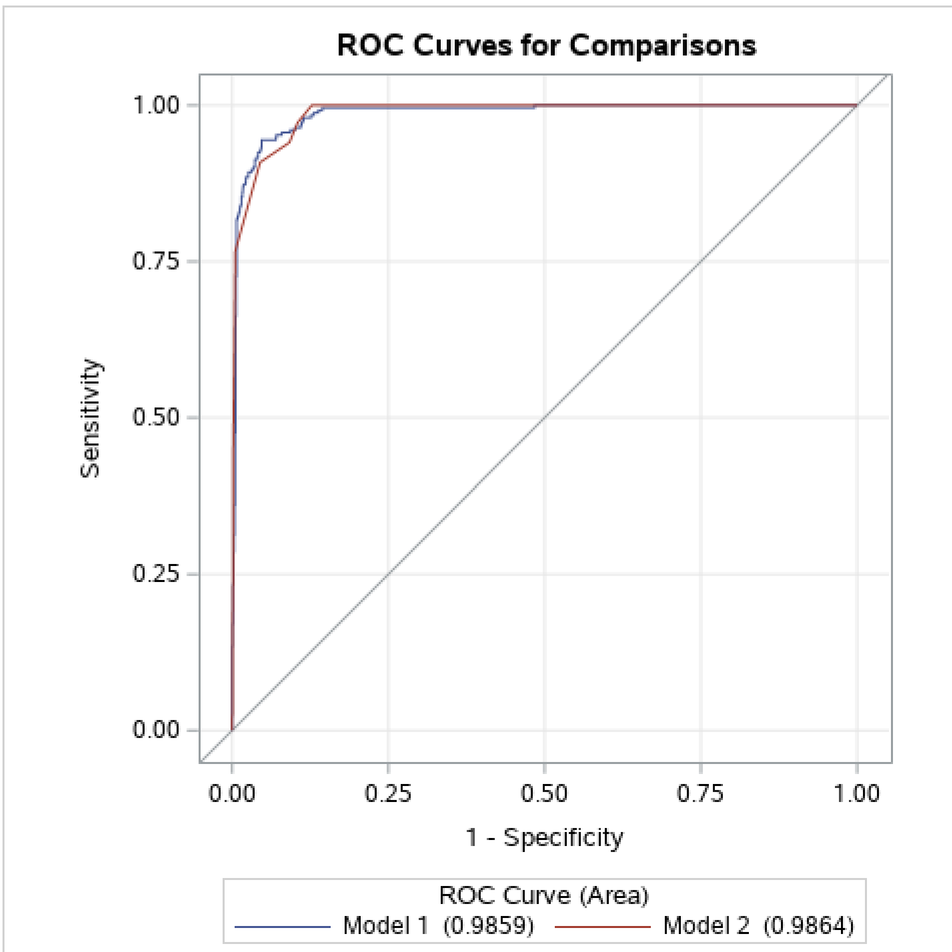
Association of Predicted Probabilities and Observed Responses			
Percent Concordant	84.2	**Somers' D**	0.684
Percent Discordant	15.8	**Gamma**	0.684
Percent Tied	0.0	**Tau-a**	0.198
Pairs	433808	**c**	0.842

Serious = -5.7708 + 1.8177*long -0.00754*depth -0.6810*md+ 3.3795*ms -2.2714*mb e.We will fit now a regression with serious as the response and xm an the only explanatory variable and compare our model to the model from objective d.

Parameter Estimates and Wald Confidence Intervals			
Parameter	**Estimate**	**95% Confidence Limits**	
Intercept	-53.3001	-57.9254	-48.6748
xm	10.0719	9.1945	10.9494

Serious = -53.3001+ 10.0719*xm

In order to compare the two models from “d” and "e” we start by dividing our dataset to separate training and test sets.



ROC Association Statistics							
ROC Model	**Mann-Whitney**				**Somers' D**	**Gamma**	**Tau-a**
	Area	**Standard Error**	**95% Wald Confidence Limits**				
Model 1	0.9859	0.00298	0.9801	0.9918	0.9719	0.9719	0.2742
Model 2	0.9864	0.00231	0.9818	0.9909	0.9727	0.9814	0.2745

ROC Contrast Test Results			
Contrast	**DF**	**Chi-Square**	**Pr > ChiSq**
Comparing Models	1	0.0322	0.8576

From the plot we can see that the two models are almost the same. As we can see from the last table the null hypothesis that the curves are the same is not rejected. Hence, we don't have evidence that Model 1 (model with multiple variables) is better than Model 2. So, we prefer model 2 as our model since its simpler.

Conclusions

Earthquakes are devastating events that often cause structural damage and loss of lives. Being able to predict the characteristics of earthquakes would be helpful in order to minimize the damage. I this report we rejected the assumption that the average maximum value of earthquakes parameters is 4.1 and we also found that earthquakes magnitude differs between countries. We proceeded with trying to find a model for the intensity of the earthquakes. Our model tried to predict Richter based on latitude, longitude, Magnitude(md), Moment Magnitude, and Surface-wave Magnitude with low success. (R adj=0.29). Finally, we tried to make a model that could predict a serious earthquake (Richter>5). Our first model contained longitude, depth, Magnitude(md), Surface-wave Magnitude, and Bodyweight Magnitude. Our second model was simpler. It contained only the maximum of the earthquake parameters (Magnitude(md), Moment Magnitude, Surface-wave Magnitude, Bodyweight Magnitude, and Richter). Comparing the two models we found them to be almost identical. We decided to keep our second model since it was simpler.