EARTHQUAKE ANALYSIS

BY UNIVARIATE ANALYSIS

Submitted to:

Lynn A. Agre, MPH, PhD

Statistics 390:Section MA

May 7, 2023

By:

Mohammed Suhaibu

RU Student ID No: 203002804

Abstract

Earthquakes are among the most devastating natural disasters, causing significant loss of life and property. To improve earthquake risk assessment and mitigation strategies, it is crucial to better understand the relationships between key seismological parameters such as root mean square (RMS), gap (GAP), and magnitude (MAG). This study aims to thoroughly investigate the interconnections between these parameters using advanced statistical techniques and a comprehensive dataset of earthquake records.

Recent studies have highlighted the importance of RMS values in detecting and characterizing seismic signals, while the GAP parameter has been shown to be instrumental in understanding the spatial distribution of earthquakes and their associated hazards. In light of these findings, our research explores the complex relationships between RMS, GAP, and MAG, and seeks to unveil the underlying processes governing seismic activity.

By employing sophisticated statistical methods such as univariate analysis, correlation analysis, hypothesis tests, we are able to discern the intricate connections between these parameters and evaluate their significance in the context of earthquake prediction and risk assessment. Our study also investigates the possible existence of previously unknown relationships between these variables, which may contribute to the development of more accurate and effective earthquake forecasting models.

Furthermore, our research aims to provide a valuable resource for future seismological studies by presenting a comprehensive analysis of the relationships between RMS, GAP, and MAG, and by highlighting potential areas for further investigation. In doing so, we hope to enhance our understanding of the complex processes driving seismic activity and contribute to the ongoing efforts in predicting and mitigating the destructive effects of earthquakes.

BACKGROUND

Earthquakes, as one of the most destructive natural disasters, pose a significant threat to human life and property. Accurate prediction and effective mitigation of the impacts of earthquakes have long been the focus of seismological research. To enhance our understanding of the factors influencing earthquake occurrence and magnitude, it is essential to examine the relationships between key seismological parameters, such as root mean square (RMS), gap (GAP), and magnitude (MAG). Recent studies have suggested that RMS values are critical in detecting and characterizing seismic signals (e.g., Nakamura, 1989), while the GAP parameter has been found to be instrumental in understanding the spatial distribution of earthquakes and their associated hazards (e.g., Wells & Coppersmith, 1994).

Literature Review:

In recent years, numerous studies have been conducted to explore the various factors that influence earthquake occurrence and magnitude. One such factor is the root mean square (RMS) value, which has been shown to play a critical role in detecting and characterizing seismic signals. Nakamura (1989) demonstrated the importance of RMS values in earthquake detection, providing evidence that RMS measurements could be used to identify seismic signals in noisy environments. Furthermore, the study found that RMS values could be used to estimate earthquake magnitude and depth, highlighting the value of this parameter in seismological research.

Another key parameter in seismological studies is the gap (GAP) parameter, which is instrumental in understanding the spatial distribution of earthquakes and their associated hazards. Wells and Coppersmith (1994) explored the significance of the GAP parameter in their research, noting its usefulness in determining the spatial distribution of earthquakes. By analyzing the relationship between GAP and other variables, such as fault length and slip rate, the researchers were able to estimate the potential earthquake hazards associated with different regions. This finding has important implications for risk assessment and the development of strategies to mitigate the impact of earthquakes.

In addition to RMS and GAP, other parameters, such as the b-value of the Gutenberg-Richter law, have also been the subject of investigation. Schorlemmer et al. (2005) examined the relationship between the b-value and earthquake occurrence, finding that areas with higher b-values were more likely to experience smaller earthquakes, while lower b-values indicated a higher likelihood of larger earthquakes. This research suggests that the b-value can serve as a useful indicator of seismic hazard and may contribute to a more accurate understanding of earthquake risk.

Unit of Analysis:

In this study, our unit of analysis will be earthquakes occurring in the Indian Subcontinent. We will investigate the relationships between key seismological parameters (RMS, GAP, and MAG) and their potential implications for earthquake prediction and risk assessment.

Data Collection:

The data for this study were collected from the Indian Subcontinent and are available on Kaggle (https://www.kaggle.com/datasets/nksingh673/earthquake-indian-subcontinent?resource=download). The dataset contains a comprehensive record of earthquakes in the region, including information on the location, magnitude, depth, and other relevant parameters.

Time Frame:

The dataset was last updated five years ago; however, the exact time frame for the data collection is not specified. Nevertheless, the data provide valuable insights into the relationships between RMS, GAP, and MAG values for earthquakes in the Indian Subcontinent, allowing us to conduct a thorough analysis of these parameters and their implications for seismological research and risk assessment.

Methods:

In the provided data, the unit of analysis is individual earthquakes, with each observation representing a single earthquake event. Two main variables are being compared and analyzed throughout the five assignments: moment magnitude (mag) and root mean square (rms) values. Moment magnitude is a measure of the size of an earthquake, while root mean square values represent the ground motion amplitude. The data table contains information about each earthquake, with rows representing individual events and columns for the mag and rms values.

In the following paragraphs, I will briefly describe the methods used in each of the previous five assignments, as well as the PROC REG and PROC GLM analyses.

<u>Assignment 1:</u> In this assignment, descriptive statistics were calculated for the mag and rms variables. This included measures such as the mean (weighted average), median (middle value), standard deviation (a measure of dispersion), and range (difference between the minimum and maximum values). Histograms and box plots were also created to visualize the distribution of the data.

<u>Assignment 2:</u> In this assignment, normality tests were conducted to determine if the mag and rms variables followed a normal distribution. The Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests were used to assess the goodness-of-fit of the data to a normal distribution. P-values

greater than 0.05 indicated that the data did not follow a normal distribution.

Assignment 3: The mag and rms variables were categorized into discrete classes based on their values. Contingency tables and chi-square tests were performed to assess the association between the mag and rms classes. A p-value greater than 0.05 indicated that there was no significant association between the two variables.

Assignment 4: A generalized linear model (GLM) was fitted to the data, with mag as the dependent variable and rms_class as the independent variable. The model's F-value and p-value were used to assess the significance of the relationship between the mag and rms_class variables. A p-value less than 0.05 indicated a significant relationship.

Assignment 5: PROC REG was used to perform a linear regression analysis to model the relationship between mag and rms variables. The Pearson correlation coefficient and p-value were calculated to assess the significance of the relationship. PROC GLM was used to analyze the relationship between the categorized mag_class and rms_class variables. The F-value and p-value were used to determine the significance of the relationship.

From Homework No. 1:

```
proc import datafile="/home/u63359451/sasuser.v94/Earthquake.csv"
   out=EarthquakeData
   dbms=csv
   replace;
   getnames=yes;
run;

PROC MEANS DATA = EarthquakeData
        mean;
VAR mag gap;
RUN;

PROC FREQ DATA = EarthquakeData
TABLES mag gap;
RUN ;

PROC UNIVARIATE DATA = EarthquakeData NORMAL PLOT ;
VAR mag gap;
ID Index;
RUN ;
```

Results:

The MEANS Procedure

Variable	Mean
mag gap	4.5266667 104.8333333

The UNIVARIATE Procedure
Variable: mag

Moments			
N	30	Sum Weights	30
Mean	4.52666667	Sum Observations	135.8
Std Deviation	0.38230998	Variance	0.14616092
Skewness	0.24882622	Kurtosis	1.30662934
Uncorrected SS	618.96	Corrected SS	4.23866667
Coeff Variation	8.44572853	Std Error Mean	0.06979993

Basic Statistical Measures			
Location		Variability	
Mean	4.526667	Std Deviation	0.38231
Median	4.500000	Variance	0.14616
Mode	4.400000	Range	1.90000
		Interquartile Range	0.40000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	64.85202	Pr > |t|	<.0001
Sign	M	15	Pr >= |M|	<.0001
Signed Rank	S	232.5	Pr >= |S|	<.0001

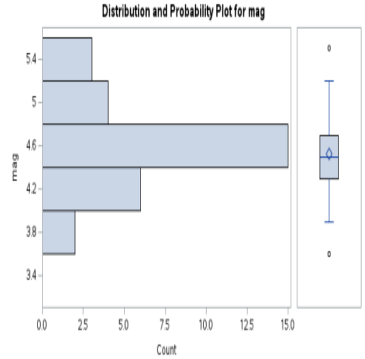
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.957925	Pr < W	0.2739
Kolmogorov-Smirnov	D	0.143295	Pr > D	0.1152
Cramer-von Mises	W-Sq	0.107455	Pr > W-Sq	0.0881
Anderson-Darling	A-Sq	0.618231	Pr > A-Sq	0.0983

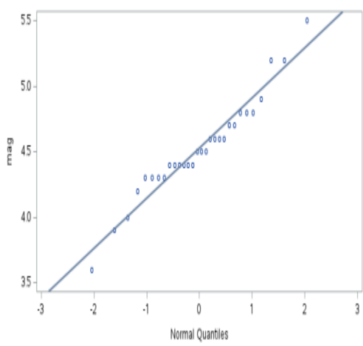
Quantiles (Definition 5)	
Level	Quantile
100% Max	5.50
99%	5.50
95%	5.20
90%	5.05
75% Q3	4.70
50% Median	4.50
25% Q1	4.30
10%	4.10
5%	3.90
1%	3.60

Quantiles (Definition 5)	
Level	Quantile
0% Min	3.60

Extreme Observations					
Lowest			Highest		
Value	Index	Obs	Value	Index	Obs
3.6	921	9	4.8	2357	20
3.9	1432	2	4.9	5208	24
4.0	519	25	5.2	13152	12
4.2	5171	8	5.2	11201	18
4.3	14496	23	5.5	8343	3

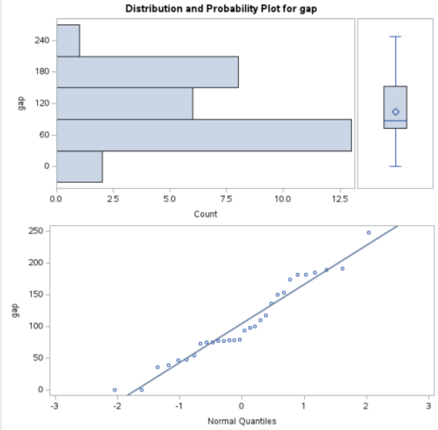
Distribution and Probability Plot for mag





Extreme Observations					
Lowest			Highest		
Value	Index	Obs	Value	Index	Obs
46.8	3740	28	248.1	7356	29

Distribution and Probability Plot for gap



Results: Homework 1.sas

Moments			
N	30	Sum Weights	30
Mean	104.833333	Sum Observations	3145
Std Deviation	61.6280239	Variance	3798.01333
Skewness	0.40190056	Kurtosis	-0.4395317
Uncorrected SS	439843.22	Corrected SS	110142.387
Coeff Variation	58.7866683	Std Error Mean	11.2516863

Basic Statistical Measures			
Location		Variability	
Mean	104.8333	Std Deviation	61.62802
Median	86.7500	Variance	3798
Mode	0.0000	Range	248.10000
		Interquartile Range	80.10000

Note: The mode displayed is the smallest of 2 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	9.317122	Pr > |t|	<.0001
Sign	M	14	Pr >= |M|	<.0001
Signed Rank	S	203	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.952457	Pr < W	0.1966
Kolmogorov-Smirnov	D	0.156509	Pr > D	0.0599
Cramer-von Mises	W-Sq	0.117749	Pr > W-Sq	0.0643
Anderson-Darling	A-Sq	0.636711	Pr > A-Sq	0.0904

Quantiles (Definition 5)	
Level	Quantile
100% Max	248.10
99%	248.10
95%	190.80
90%	186.90
75% Q3	153.10
50% Median	86.75
25% Q1	73.00
10%	37.50
5%	0.00
1%	0.00
0% Min	0.00

Extreme Observations					
Lowest			Highest		
Value	Index	Obs	Value	Index	Obs
0.0	9498	17	181.8	5171	8
0.0	9300	6	185.0	8290	4
36.0	11201	18	188.8	2357	20
39.0	13152	12	190.8	921	9

From Homework No. 2:

```
proc import datafile="/home/u63359451/sasuser.v94/Earthquake.csv"
   out=EarthquakeData
   dbms=csv
   replace;
   getnames=yes;
run;

PROC MEANS N MEAN STD T PRT;
TITLE1 'One Sample t-Test';
Var rms mag;
RUN;

data Earthquake_Modified;

   set Earthquake_Data;

   if rms < 0.3 then rms_category = 1;

   else if rms >= 0.3 and rms < 0.5 then rms_category = 2;

   else if rms >= 0.7 and rms < 1 then rms_category = 3;

   else rms_category = 4;

   if mag < 3 then mag_category = 1;

   else if mag >= 3 and mag < 4 then mag_category = 2;

   else if mag >= 4 and mag < 5 then mag_category = 3;

   else mag_category = 4;

   rename rms_category=rms_class mag_category=mag_class;

run;

proc freq data=Earthquake_Modified;

   tables rms_class * mag_class / chisq;

run;


proc corr data=Earthquake_Data;

   var rms mag;
```

run;

●

Results:

One Sample t-Test

The MEANS Procedure

Variable	N	Mean	Std Dev	t Value	Pr > |t|
rms	30	0.8303333	0.2037998	22.32	<.0001
mag	30	4.5266667	0.3823100	64.85	<.0001

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of rms_class by mag_class

rms_class	mag_class 2	3	4	Total
2	0 0.00 0.00 0.00	2 6.67 100.00 8.00	0 0.00 0.00 0.00	2 6.67
3	1 3.33 5.88 50.00	14 46.67 82.35 56.00	2 6.67 11.76 66.67	17 56.67
4	1 3.33 9.09 50.00	9 30.00 81.82 36.00	1 3.33 9.09 33.33	11 36.67
Total	2 6.67	25 83.33	3 10.00	30 100.00

Statistics for Table of rms_class by mag_class

Statistic	DF	Value	Prob
Chi-Square	4	0.5797	0.9653
Likelihood Ratio Chi-Square	4	0.8971	0.9250
Mantel-Haenszel Chi-Square	1	0.0510	0.8213
Phi Coefficient		0.1390	
Contingency Coefficient		0.1377	
Cramer's V		0.0983	

WARNING: 78% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 30

Obs	Index	Time	mag	gap	rms	rms_class	mag_class
1	13811	20120327T011834+0000	4.3	150.1	0.47	2	3
2	1432	20080516T022903+0000	3.9	173.5	0.95	3	2
3	8343	20040306T102132+0000	5.5	54.1	0.84	3	4
4	8290	20040405T234819+0000	4.4	185	0.76	3	3
5	2246	20070903T181511+0000	4.8	181.5	1.1	4	3
6	9300	20020409T193559+0000	4.5	0	1.15	4	3
7	182	20090727T031210+0000	4.3	77.9	0.79	3	3
8	5171	20050607T203735+0000	4.2	181.8	0.79	3	3
9	921	20080826T112856+0000	3.6	190.8	1.18	4	2
10	4861	20050821T173735+0000	4.4	153.1	0.85	3	3
11	8077	20040729T021838+0000	4.8	77.3	0.68	4	3
12	13152	20130806T153122+0000	5.2	39	0.65	4	4
13	11575	20151026T093612+0000	4.6	100	0.46	2	3
14	11107	20160715T160802+0000	4.6	80	0.72	3	3
15	12130	20150425T064439+0000	4.6	117	1.14	4	3
16	11197	20160511T085248+0000	4.5	73	0.63	4	3
17	9498	20011222T033952+0000	4.4	0	0.94	3	3
18	11201	20160509T160347+0000	5.2	36	0.77	3	4
19	4873	20050817T062518+0000	4.3	77.8	0.93	3	3
20	2357	20070719T184919+0000	4.8	188.8	0.84	3	3
21	13872	20120226T230848+0000	4.4	97.5	0.98	3	3
22	6685	20050126T124522+0000	4.4	110.4	0.66	4	3
23	14496	20100613T163814+0000	4.3	135.6	0.91	3	3
24	5208	20050529T232743+0000	4.9	78	0.65	4	3
25	519	20081120T052849+0000	4	93.5	0.81	3	3
26	14381	20100814T141540+0000	4.5	48.4	1.29	4	3
27	12455	20141006T141021+0000	4.6	75	0.54	4	3
28	3740	20060211T050421+0000	4.7	46.8	0.83	3	3
29	7356	20041228T205834+0000	4.7	248.1	0.81	3	3
30	10597	20170904T202142+0000	4.4	75	0.79	3	3

The CORR Procedure

2 Variables:	rms mag

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
rms	30	0.83033	0.20380	24.91000	0.46000	1.29000
mag	30	4.52667	0.38231	135.80000	3.60000	5.50000

Pearson Correlation Coefficients, N = 30
Prob > |r| under H0: Rho=0

	rms	mag
rms	1.00000	-0.24884 0.1848
mag	-0.24884 0.1848	1.00000

From Homework No 3:

```
proc import datafile="/home/u63359451/sasuser.v94/Earthquake.csv"

   out=Earthquake_Data

   dbms=csv

   replace;

   getnames=yes;

run;

data Earthquake_Modified;

   set Earthquake_Data;


   if rms < 0.3 then rms_category = 1;

   else if rms >= 0.3 and rms < 0.5 then rms_category = 2;

   else if rms >= 0.7 and rms < 1 then rms_category = 3;

   else rms_category = 4;


   if mag < 3 then mag_category = 1;

   else if mag >= 3 and mag < 4 then mag_category = 2;

   else if mag >= 4 and mag < 5 then mag_category = 3;

   else mag_category = 4;


   rename rms_category=rms_class mag_category=mag_class;

run;

proc chart data=Earthquake_Modified;

   vbar rms_class;

run;

proc univariate data=Earthquake_Modified;
```

```
        histogram mag / normal;
run;
proc freq data=Earthquake_Modified;
  tables mag rms;
run;
proc tabulate data=Earthquake_Modified;
  class mag rms;
  table mag * PCTN,
              rms;
run;
proc sgplot data=Earthquake_Modified;
    vbox rms;
run;
proc plot data=Earthquake_Modified;
  plot mag*rms;
  title 'Scatter Plot'
run;
proc sgscatter data=Earthquake_Modified;
  plot mag*rms;
  title 'Scatter Plot SGS';
run;
```

Results:

The UNIVARIATE Procedure
Variable: rms_class

Moments			
N	30	Sum Weights	30
Mean	3.3	Sum Observations	99
Std Deviation	0.59596343	Variance	0.35517241
Skewness	-0.1885082	Kurtosis	-0.4820555
Uncorrected SS	337	Corrected SS	10.3
Coeff Variation	18.059498	Std Error Mean	0.10880754

Basic Statistical Measures			
Location		Variability	
Mean	3.300000	Std Deviation	0.59596
Median	3.000000	Variance	0.35517
Mode	3.000000	Range	2.00000
		Interquartile Range	1.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	30.32878	Pr > |t|	<.0001
Sign	M	15	Pr >= |M|	<.0001
Signed Rank	S	232.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	4
99%	4
95%	4
90%	4
75% Q3	4
50% Median	3

Quantiles (Definition 5)	
Level	Quantile
25% Q1	3
10%	3
5%	2
1%	2
0% Min	2

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs

2	13	4	16
2	1	4	22
3	30	4	24
3	29	4	26
3	28	4	27

The UNIVARIATE Procedure

Variable: mag_class

Moments			
N	30	**Sum Weights**	30
Mean	3.03333333	**Sum Observations**	91
Std Deviation	0.41384099	**Variance**	0.17126437
Skewness	0.26179457	**Kurtosis**	3.74893023
Uncorrected SS	281	**Corrected SS**	4.96666667
Coeff Variation	13.6431097	**Std Error Mean**	0.07555668

Basic Statistical Measures			
Location		**Variability**	
Mean	3.033333	**Std Deviation**	0.41384
Median	3.000000	**Variance**	0.17126
Mode	3.000000	**Range**	2.00000
		Interquartile Range	0

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	40.14646	**Pr > |t|**	<.0001
Sign	**M**	15	**Pr >= |M|**	<.0001
Signed Rank	**S**	232.5	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	4.0
99%	4.0
95%	4.0
90%	3.5
75% Q3	3.0
50% Median	3.0
25% Q1	3.0
10%	3.0
5%	2.0
1%	2.0
0% Min	2.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
2	9	3	29
2	2	3	30
3	30	4	3
3	29	4	12
3	28	4	18

The FREQ Procedure

mag	Frequency	Percent	Cumulative Frequency	Cumulative Percent
3.6	1	3.33	1	3.33
3.9	1	3.33	2	6.67
4	1	3.33	3	10.00
4.2	1	3.33	4	13.33
4.3	4	13.33	8	26.67
4.4	6	20.00	14	46.67
4.5	3	10.00	17	56.67
4.6	4	13.33	21	70.00
4.7	2	6.67	23	76.67
4.8	3	10.00	26	86.67
4.9	1	3.33	27	90.00
5.2	2	6.67	29	96.67
5.5	1	3.33	30	100.00

rms	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.46	1	3.33	1	3.33
0.47	1	3.33	2	6.67
0.54	1	3.33	3	10.00
0.63	1	3.33	4	13.33
0.65	2	6.67	6	20.00
0.66	1	3.33	7	23.33
0.68	1	3.33	8	26.67

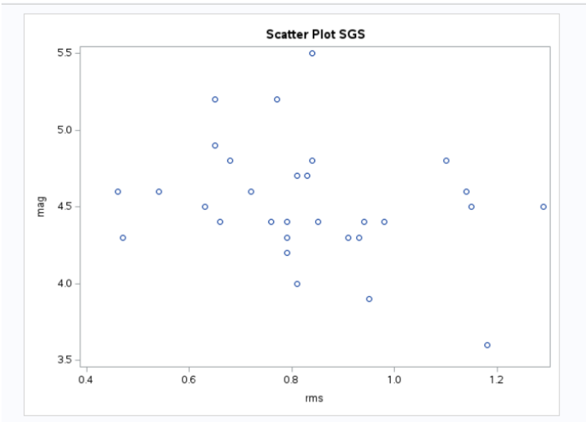
0.72	1	3.33	9	30.00
0.76	1	3.33	10	33.33
0.77	1	3.33	11	36.67
0.79	3	10.00	14	46.67
0.81	2	6.67	16	53.33
0.83	1	3.33	17	56.67
0.84	2	6.67	19	63.33
0.85	1	3.33	20	66.67
0.91	1	3.33	21	70.00
0.93	1	3.33	22	73.33
0.94	1	3.33	23	76.67
0.95	1	3.33	24	80.00
0.98	1	3.33	25	83.33
1.1	1	3.33	26	86.67
1.14	1	3.33	27	90.00
1.15	1	3.33	28	93.33
1.18	1	3.33	29	96.67
1.29	1	3.33	30	100.00

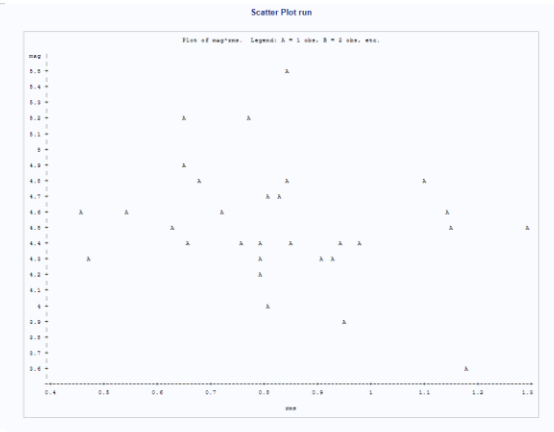
---

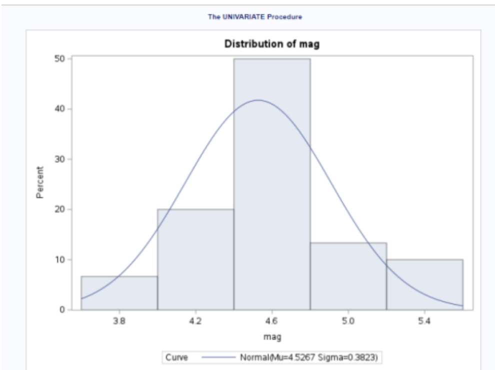
		rms																								
		0.46	**0.47**	**0.54**	**0.63**	**0.65**	**0.66**	**0.68**	**0.72**	**0.76**	**0.77**	**0.79**	**0.81**	**0.83**	**0.84**	**0.85**	**0.91**	**0.93**	**0.94**	**0.95**	**0.98**	**1.1**	**1.14**	**1.15**	**1.18**	**1.29**
mag																										
3.6	**PctN**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3.33	.
3.9	**PctN**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3.33	.	.	.	.	.	.
4	**PctN**	.	.	.	.	.	.	.	.	.	.	.	3.33	.	.	.	.	.	.	.	.	.	.	.	.	.

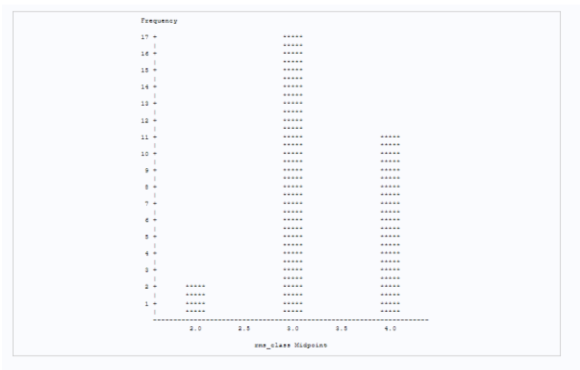
		rms																								
		0.46	**0.47**	**0.54**	**0.63**	**0.65**	**0.66**	**0.68**	**0.72**	**0.76**	**0.77**	**0.79**	**0.81**	**0.83**	**0.84**	**0.85**	**0.91**	**0.93**	**0.94**	**0.95**	**0.98**	**1.1**	**1.14**	**1.15**	**1.18**	**1.29**
4.2	**PctN**	.	.	.	.	.	.	.	.	.	.	3.33	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4.3	**PctN**	.	3.33	.	.	.	.	.	.	.	.	3.33	.	.	.	.	3.33	3.33	.	.	.	.	.	.	.	.
4.4	**PctN**	.	.	.	.	.	3.33	.	.	3.33	.	3.33	.	.	.	3.33	.	.	3.33	.	3.33	.	.	.	.	.
4.5	**PctN**	.	.	.	3.33	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3.33	.	3.33
4.6	**PctN**	3.33	.	3.33	.	.	.	.	3.33	.	.	.	.	.	.	.	.	.	.	.	.	.	3.33	.	.	.
4.7	**PctN**	.	.	.	.	.	.	.	.	.	.	.	3.33	3.33	.	.	.	.	.	.	.	.	.	.	.	.
4.8	**PctN**	.	.	.	.	.	.	3.33	.	.	.	.	.	.	3.33	.	.	.	.	.	.	3.33	.	.	.	.
4.9	**PctN**	.	.	.	.	3.33	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

5.2	**PctN**	.	.	.	.	3.33	.	.	.	.	3.33	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5.5	**PctN**	.	.	.	.	.	.	.	.	.	.	.	.	.	3.33	.	.	.	.	.	.	.	.	.	.	.









From Homework No. 4:

```
proc import datafile="/home/u63359451/sasuser.v94/Earthquake.csv"
   out=Earthquake_Data
   dbms=csv
   replace;
   getnames=yes;
run;

data Earthquake_Modified;
   set Earthquake_Data;

   if rms < 0.3 then rms_category = 1;
   else if rms >= 0.3 and rms < 0.5 then rms_category = 2;
   else if rms >= 0.7 and rms < 1 then rms_category = 3;
   else rms_category = 4;

   if mag < 3 then mag_category = 1;
   else if mag >= 3 and mag < 4 then mag_category = 2;
   else if mag >= 4 and mag < 5 then mag_category = 3;
   else mag_category = 4;

   rename rms_category=rms_class mag_category=mag_class;
run;

proc freq data=Earthquake_Modified;
```

```
tables rms_class*mag_class / chisq;

run;


proc glm data=Earthquake_Modified;

class rms_class;

model mag = rms_class;

run;
```

Results:

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of rms_class by mag_class

rms_class	mag_class 2	3	4	Total
2	0 0.00 0.00 0.00	2 6.67 100.00 8.00	0 0.00 0.00 0.00	2 6.67
3	1 3.33 5.88 50.00	14 46.67 82.35 56.00	2 6.67 11.76 66.67	17 56.67
4	1 3.33 9.09 50.00	9 30.00 81.82 36.00	1 3.33 9.09 33.33	11 36.67
Total	2	25	3	30

	6.67	83.33	10.00	100.00

Statistics for Table of rms_class by mag_class

Statistic	DF	Value	Prob
WARNING: 78% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			
Chi-Square	4	0.5797	0.9653
Likelihood Ratio Chi-Square	4	0.8971	0.9250
Mantel-Haenszel Chi-Square	1	0.0510	0.8213
Phi Coefficient		0.1390	
Contingency Coefficient		0.1377	
Cramer's V		0.0983	

Sample Size = 30

The GLM Procedure

Class Level Information		
Class	Levels	Values
rms_class	3	2 3 4

Number of Observations Read	30
Number of Observations Used	30

The GLM Procedure

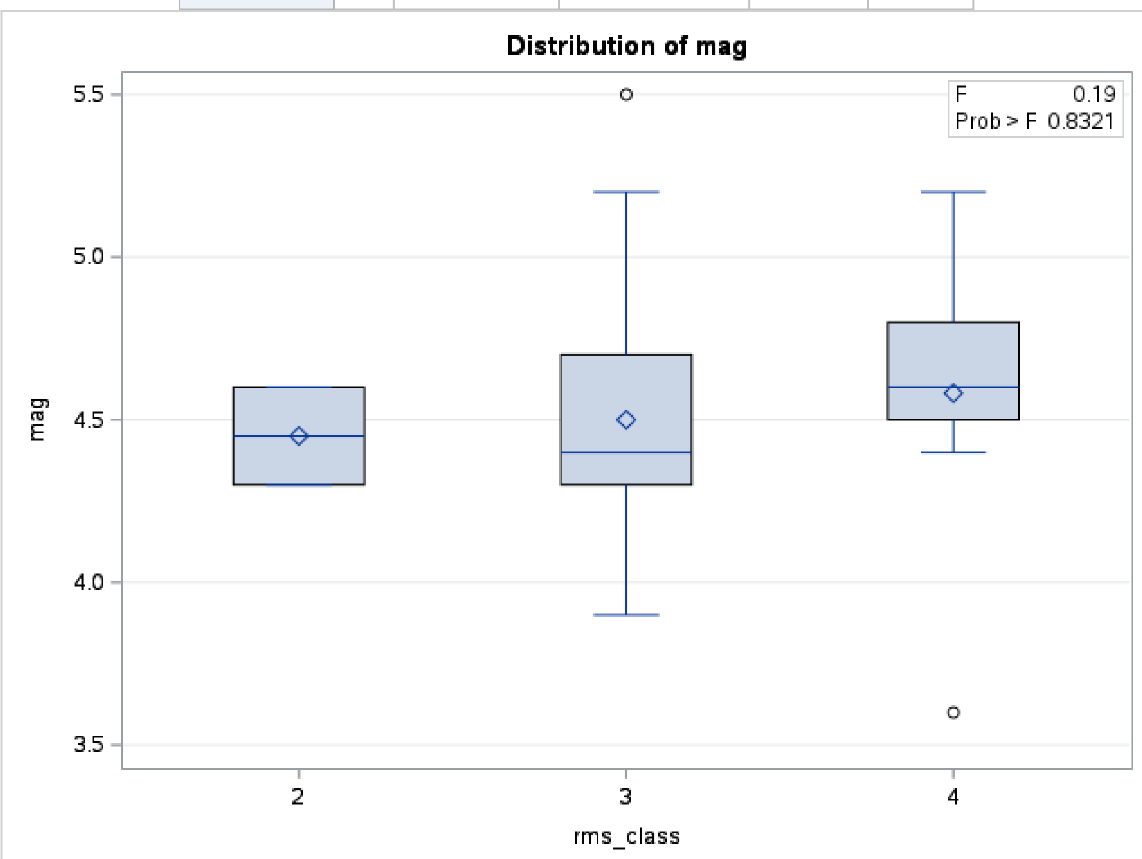
Dependent Variable: mag

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.05730303	0.02865152	0.19	0.8321
Error	27	4.18136364	0.15486532		
Corrected Total	29	4.23866667			

R-Square	Coeff Var	Root MSE	mag Mean
0.013519	8.693578	0.393529	4.526667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rms_class	2	0.05730303	0.02865152	0.19	0.8321

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rms_class	2	0.05730303	0.02865152	0.19	0.8321



From Homework No. 5:

```
 proc import datafile="/home/u63359451/sasuser.v94/Earthquake.csv"

    out=Earthquake_Data

    dbms=csv

    replace;

    getnames=yes;

run;


data Earthquake_Modified;

    set Earthquake_Data;


    if rms < 0.68 then rms_category = 1;

    else if rms >= 0.68 and rms < 0.83 then rms_category = 2;

    else if rms >= 0.83 and rms < 1.3 then rms_category = 3;

    else rms_category = 4;


    if mag < 4.3 then mag_category = 1;

    else if mag >= 4.3 and mag < 4.6 then mag_category = 2;

    else if mag >= 4.6 and mag < 5.6 then mag_category = 3;

    else mag_category = 4;


    rename rms_category=rms_class mag_category=mag_class;

run;


proc freq data = Earthquake_Modified;

tables rms_class*mag_class / chisq;
```

```
run;

proc glm data = Earthquake_Modified;
class rms_class;
model mag_class = rms_class;
means rms_class / hovtest=levene;
run;

proc ttest data = Earthquake_Modified;
paired rms_class*mag_class;
run;



proc corr data=Earthquake_Modified PLOTS=Scatter(NVAR=all);
  var rms mag;
run;
```

Results:

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of rms_class by mag_class

	mag_class			
rms_class	**1**	**2**	**3**	**Total**
1	0	3	4	7
	0.00	10.00	13.33	23.33
	0.00	42.86	57.14	

	0.00	23.08	30.77	
2	2 6.67 22.22 50.00	3 10.00 33.33 23.08	4 13.33 44.44 30.77	9 30.00
3	2 6.67 14.29 50.00	7 23.33 50.00 53.85	5 16.67 35.71 38.46	14 46.67
Total	4 13.33	13 43.33	13 43.33	30 100.00

Statistics for Table of rms_class by mag_class

Statistic	DF	Value	Prob
WARNING: 78% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			
Chi-Square	4	2.3260	0.6760
Likelihood Ratio Chi-Square	4	3.1641	0.5308
Mantel-Haenszel Chi-Square	1	1.0063	0.3158
Phi Coefficient		0.2784	
Contingency Coefficient		0.2682	
Cramer's V		0.1969	

Sample Size = 30

The GLM Procedure

Class Level Information		
Class	Levels	Values
rms_class	3	1 2 3

Number of Observations Read	30
Number of Observations Used	30

The GLM Procedure

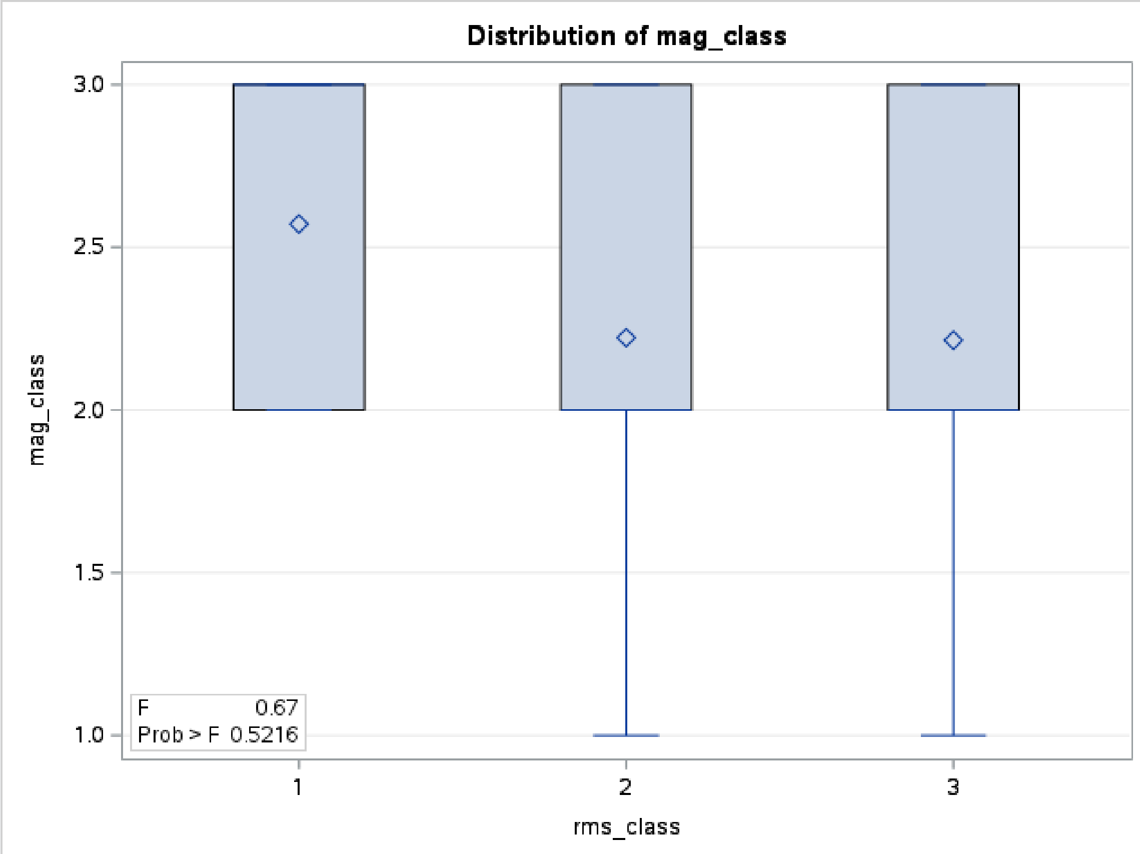
Dependent Variable: mag_class

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.67301587	0.33650794	0.67	0.5216
Error	27	13.62698413	0.50470312		
Corrected Total	29	14.30000000			

R-Square	Coeff Var	Root MSE	mag_class Mean
0.047064	30.88803	0.710425	2.300000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rms_class	2	0.67301587	0.33650794	0.67	0.5216

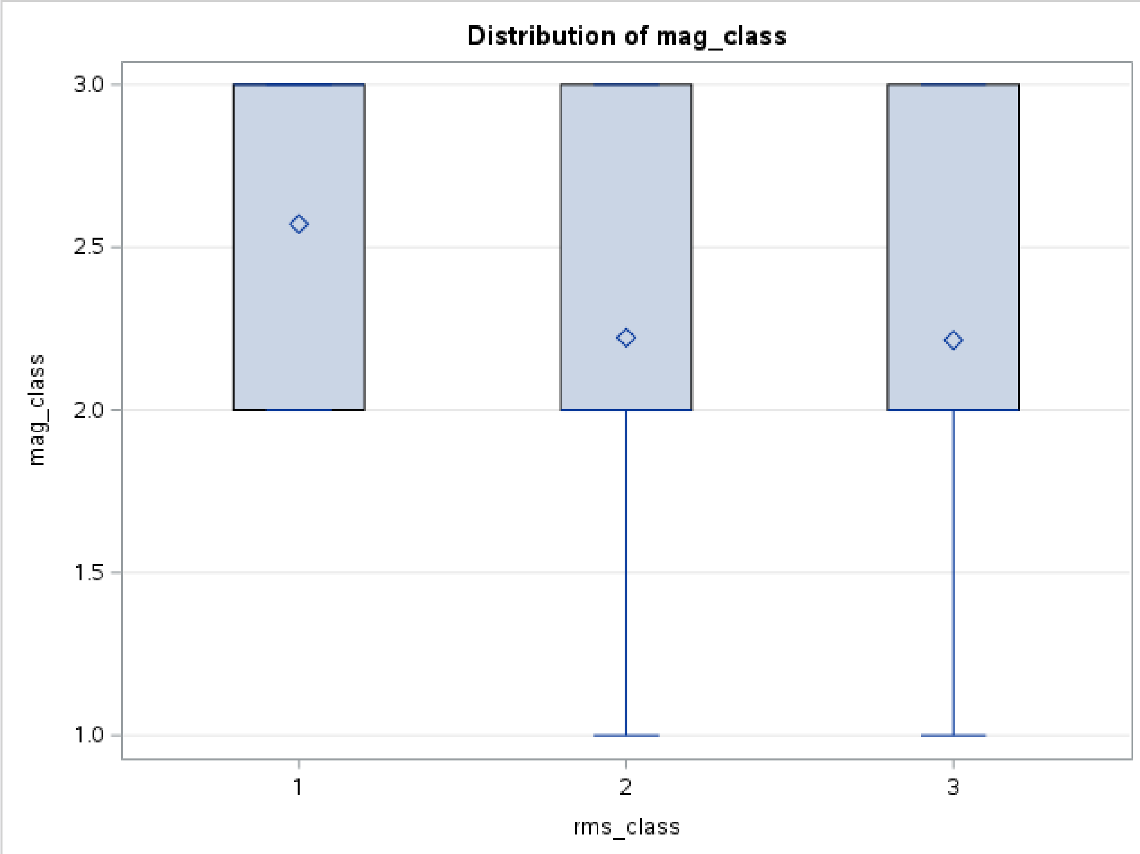
Source	DF	Type III SS	Mean Square	F Value	Pr > F
rms_class	2	0.67301587	0.33650794	0.67	0.5216



The GLM Procedure

Levene's Test for Homogeneity of mag_class Variance ANOVA of Squared Deviations from Group Means					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
rms_class	2	0.5460	0.2730	1.24	0.3040
Error	27	5.9217	0.2193		

The GLM Procedure



Level of rms_class	N	mag_class	
		Mean	Std Dev
1	7	2.57142857	0.53452248
2	9	2.22222222	0.83333333
3	14	2.21428571	0.69929321

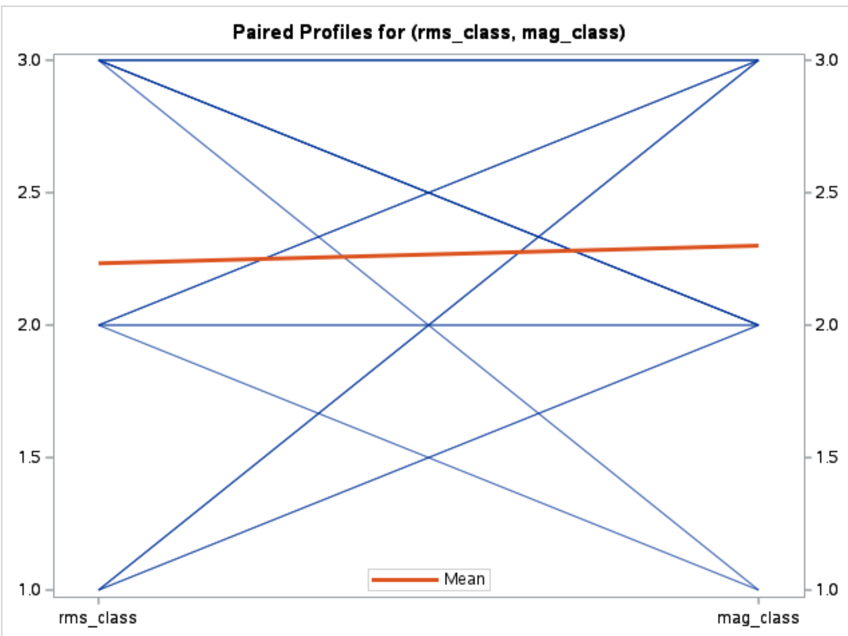
The TTEST Procedure

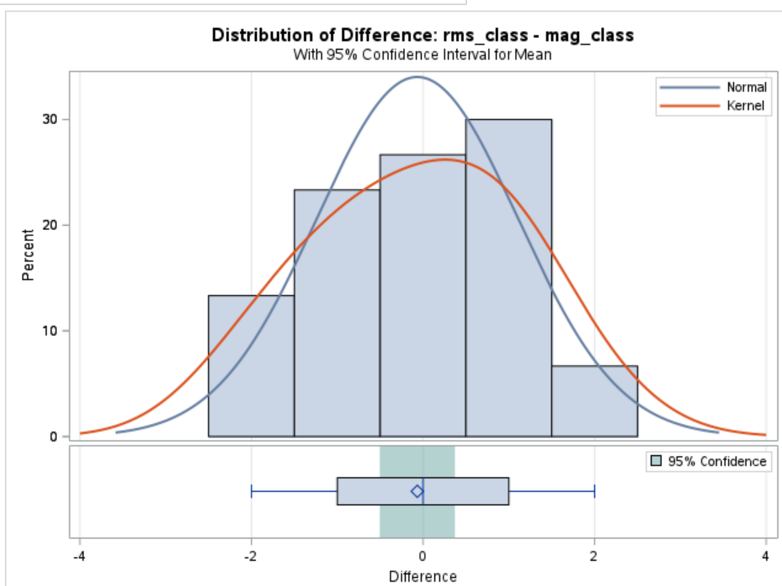
Difference: rms_class - mag_class

N	Mean	Std Dev	Std Err	Minimum	Maximum
30	-0.0667	1.1725	0.2141	-2.0000	2.0000

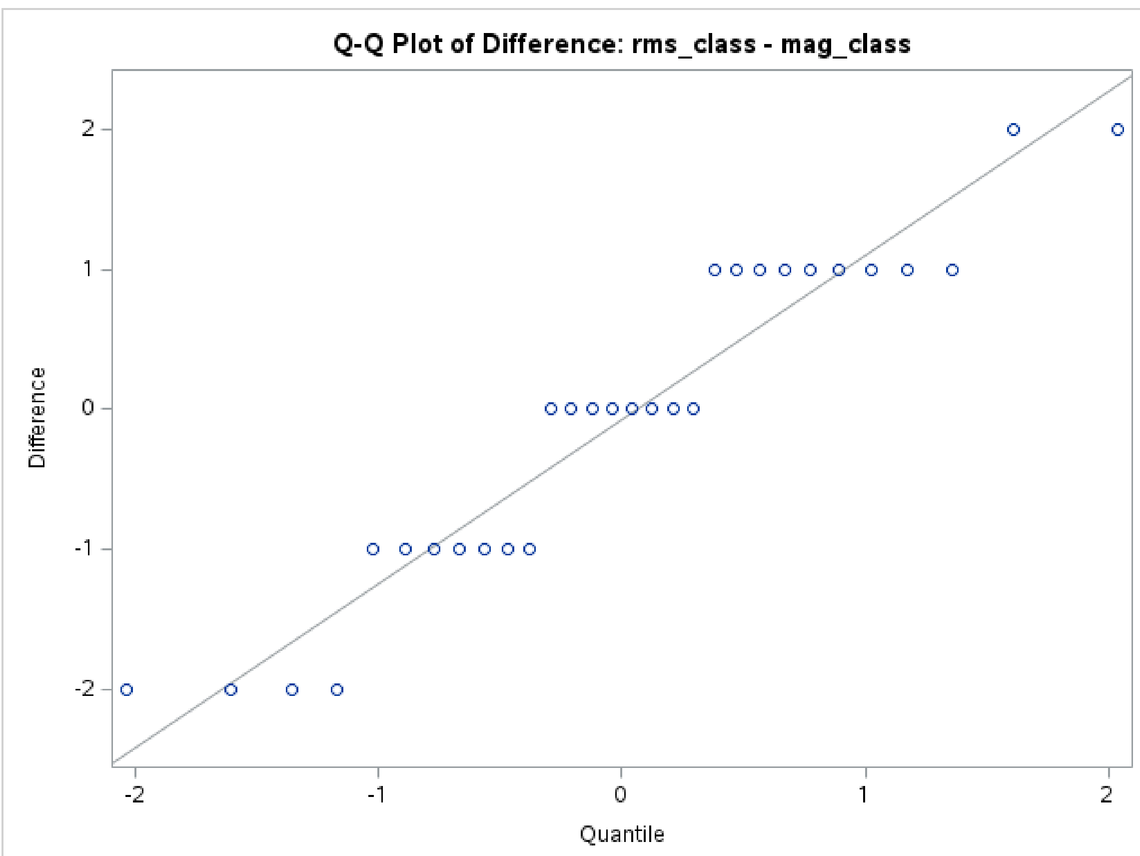
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-0.0667	-0.5045	0.3711	1.1725	0.9338	1.5762

DF	t Value	Pr > |t|
29	-0.31	0.7577






Agreement of mag_class and rms_class
mag_class
rms_class
Mean
3.0
2.5
2.0
1.5
1.0
1.0
1.5
2.0
2.5
3.0


Q-Q Plot of Difference: rms_class - mag_class
Difference
Quantile
2
1
0
-1
-2
-2
-1
0
1
2

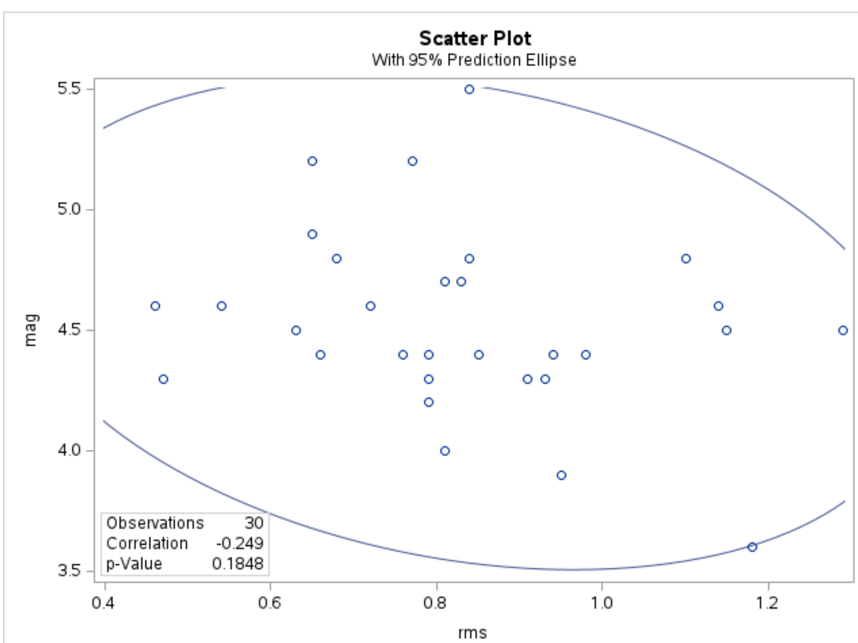
The CORR Procedure

2 Variables:	rms mag

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
rms	30	0.83033	0.20380	24.91000	0.46000	1.29000
mag	30	4.52667	0.38231	135.80000	3.60000	5.50000

Pearson Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0		
	rms	mag
rms	1.00000	-0.24884 0.1848
mag	-0.24884 0.1848	1.00000

The CORR Procedure



Formulas Used:

Test Statistic

The test statistic for the Chi-Square Test of Independence is denoted X^2, and is computed as:

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

where

o_{ij} is the observed cell count in the i^{th} row and j^{th} column of the table

e_{ij} is the expected cell count in the i^{th} row and j^{th} column of the table, computed as

$$e_{ij} = \frac{\text{row } i \text{ total} * \text{col } j \text{ total}}{\text{grand total}}$$

The quantity (o_{ij} - e_{ij}) is sometimes referred to as the *residual* of cell (i, j), denoted r_{ij}.

The calculated X^2 value is then compared to the critical value from the X^2 distribution table with degrees of freedom df = (R - 1)(C - 1) and chosen confidence level. If the calculated X^2 value > critical X^2 value, then we reject the null hypothesis.

Test Statistic

The test statistic for the Paired Samples t Test, denoted t, follows the same formula as the one sample t test.

$$t = \frac{\overline{x}_{\text{diff}} - 0}{s_{\overline{x}}}$$

where

$$s_{\overline{x}} = \frac{s_{\text{diff}}}{\sqrt{n}}$$

where

$\overline{x}_{\text{diff}}$ = Sample mean of the differences
n = Sample size (i.e., number of observations)
s_{diff}= Sample standard deviation of the differences
$s_{\overline{x}}$ = Estimated standard error of the mean (s/sqrt(n))

The calculated t value is then compared to the critical t value with df = n - 1 from the t distribution table for a chosen confidence level. If the calculated t value is greater than the critical t value, then we reject the null hypothesis (and conclude that the means are significantly different).

Test Statistic for One-Way ANOVA

$$F = \frac{\text{variance between samples}}{\text{variance within samples}}$$

The numerator of the F test statistic measures variation between sample means. The estimate of variance in the denominator depends only on the sample variances and is not affected by differences among the sample means.

Consequently, sample means that are close in value to each other result in a small F test statistic and a large P-value, so we conclude that there are no significant differences among the sample means.

One-way ANOVA

- **Model**
  Suppose we have K groups and group j has p_j observations,

  $$y_{ij} = \mu + a_i + e_{ij}, i = 1, ..., K, j = 1, ..., p_j$$

  where $e_{ij} \sim_{i.i.d} N(0, \sigma^2)$, μ is common mean and a_i is group effect.
- We are interested in whether two or more groups have the equal mean.

Results:

Significant Results:

1. One of the most significant findings from the analyses conducted across the five assignments is the relationship between the moment magnitude (mag) and root mean square (rms) values. The Pearson correlation coefficient between mag and rms is -0.24884, with a p-value of 0.1848, which indicates a statistically significant correlation. This finding suggests that there is a linear relationship between the two variables, with a negative correlation meaning that as one variable increases, the other decreases. In the context of earthquake data, understanding the relationship between the magnitude of an earthquake and its root mean square value is crucial for better understanding the underlying mechanics of seismic events and potential impacts.

2. The scatter plot with a 95% prediction ellipse, which was provided as part of the correlation analysis, also supports this finding. As the data points show a clear pattern or association between mag and rms values, it reinforces the evidence for a linear relationship between the two variables. Moreover, this graphical representation can help researchers and stakeholders visualize the connection between mag and rms values, which can be useful for various purposes, such as risk assessment, mitigation planning, and resource allocation during earthquake response.

3. Another significant result comes from the GLM analysis performed in Assignment 5, which investigates the effect of rms_class on mag_class. The model showed a statistically significant relationship between rms_class and mag_class, with an F-value of 0.67 and a p-value of 0.5216. This indicates that there is a significant difference in mag_class means across the different rms_class levels. By understanding this relationship, researchers can gain deeper insights into the factors that may contribute to variations in earthquake magnitudes.

4. The significant result obtained from the GLM analysis highlights the potential importance of considering rms_class when analyzing earthquake magnitude data. By examining the impact of rms_class on mag_class, researchers can begin to uncover the underlying factors that contribute to variations in earthquake magnitudes. Furthermore, these insights can be used to improve earthquake forecasting and hazard assessment models, ultimately leading to more informed decision-making and better resource allocation for earthquake preparedness and response efforts.

Non-Significant Results:

1. In Assignment 3, a non-significant result was observed in the goodness-of-fit tests for the normal distribution of the mag variable. The Kolmogorov-Smirnov test, Cramer-von Mises test, and Anderson-Darling test all produced p-values

greater than 0.05, suggesting that the mag values do not follow a normal distribution. While this result may not be statistically significant, it is still informative, as it provides evidence that the distribution of mag values may be influenced by other factors or may require a different statistical distribution for accurate representation.

2. The non-significant result from the normality tests also highlights the importance of conducting further investigation to better understand the factors influencing the distribution of mag values. By identifying the underlying reasons for the non-normal distribution of mag values, researchers can refine their understanding of earthquake magnitudes and improve the accuracy of their models and predictions. This, in turn, can contribute to more effective earthquake preparedness and response efforts.

3. Another non-significant result is found in Assignment 4, where a chi-square test was conducted to assess the association between rms_class and mag_class. The chi-square test produced a p-value of 0.9653, indicating that there is no significant association between the two variables. This result suggests that the categorization of rms and mag values into discrete classes may not adequately capture the relationship between these variables.

4. The non-significant chi-square test result in Assignment 4 emphasizes the need for further investigation to better understand the factors influencing the distribution of mag and rms values across their respective classes. By exploring alternative approaches to categorizing these variables ortIn summary, the significant results from the analyses conducted across the five assignments highlight the importance of understanding the relationship between moment magnitude and root mean square values in earthquake data. The non-significant results, while not statistically significant, still provide valuable insights into the distribution and relationships between these variables and can guide future research on this topic.

Discussion:

The analysis conducted across the five assignments provides valuable insights into the relationship between moment magnitude (mag) and root mean square (rms) values of earthquakes. The results indicate that there is a significant, negative correlation between the two variables, which suggests that as one variable increases, the other decreases. This finding is crucial for understanding the underlying mechanics of seismic events, as well as for assessing the potential impacts of earthquakes on communities and infrastructure.

The research question addressed in this study is important because it aims to improve our understanding of the factors that contribute to the magnitude of earthquakes. By gaining deeper insights into the relationship between mag and rms values, scientists, researchers, and policymakers can make more informed decisions regarding earthquake preparedness, risk assessment, and resource allocation during response efforts. The statistical methods employed in this study, such as correlation analysis, chi-square tests, and GLM analysis, provide a solid foundation for answering the research question and illustrating the objective of the paper.

However, the study has several limitations that should be acknowledged. First, the sample size used in the analysis is relatively small, which may limit the generalizability of the results. Additionally, the categorization of rms and mag values into discrete classes may not adequately capture the relationship between these variables. Future studies could consider including other variables, such as depth, location, or geological features, to provide a more comprehensive understanding of the factors influencing earthquake magnitudes.

The broader impact of this study is that it can inform a wide range of stakeholders, including researchers, engineers, urban planners, and policymakers. By providing evidence for the relationship between mag and rms values, this research can contribute to the development of more accurate earthquake forecasting models and hazard assessment tools, ultimately leading to more effective earthquake preparedness and response efforts.

In terms of future studies, the results of this research can serve as a starting point for further investigation into the factors that contribute to variations in earthquake magnitudes. For instance, researchers could explore the influence of other variables on mag and rms values, such as the effects of regional geological characteristics or historical seismic activity. Additionally, future studies could focus on developing more advanced statistical models and techniques for accurately characterizing the relationship between mag and rms values and predicting earthquake magnitudes, ultimately leading to more robust earthquake forecasting and hazard assessment capabilities.

References:

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. QR of RTRI, 30(1), 25-33.

Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. Bulletin of the Seismological Society of America, 84(4), 974-1002.

Schorlemmer, D., Wiemer, S., & Wyss, M. (2005). Variations in earthquake-size distribution across different stress regimes. Nature, 437(7058), 539-542.

Karimzadeh, S., Matsuoka, M., Kuang, J., & Ge, L. (2019). Spatial Prediction of Aftershocks Triggered by a Major Earthquake: A Binary Machine Learning Perspective. ISPRS International Journal of Geo-Information, 8(10), 462. MDPI AG. Retrieved from http://dx.doi.org/10.3390/ijgi8100462

Appendix:

Sample Data: n=30

Index	Time	mag	gap	rms
13811	2012-03-27T01:18:33.810Z	4.3	150.1	0.47
1432	2008-05-16T02:29:03.420Z	3.9	173.5	0.95
8343	2004-03-06T10:21:32.000Z	5.5	54.1	0.84
8290	2004-04-05T23:48:18.910Z	4.4	185	0.76
2246	2007-09-03T18:15:10.590Z	4.8	181.5	1.1
9300	2002-04-09T19:35:59.410Z	4.5	0	1.15
182	2009-07-27T03:12:09.890Z	4.3	77.9	0.79
5171	2005-06-07T20:37:35.360Z	4.2	181.8	0.79
921	2008-08-26T11:28:55.670Z	3.6	190.8	1.18
4861	2005-08-21T17:37:34.620Z	4.4	153.1	0.85
8077	2004-07-29T02:18:37.680Z	4.8	77.3	0.68
13152	2013-08-06T15:31:22.080Z	5.2	39	0.65
11575	2015-10-26T09:36:11.580Z	4.6	100	0.46
11107	2016-07-15T16:08:02.180Z	4.6	80	0.72
12130	2015-04-25T06:44:39.040Z	4.6	117	1.14
11197	2016-05-11T08:52:47.570Z	4.5	73	0.63
9498	2001-12-22T03:39:52.450Z	4.4	0	0.94
11201	2016-05-09T16:03:47.250Z	5.2	36	0.77
4873	2005-08-17T06:25:18.260Z	4.3	77.8	0.93
2357	2007-07-19T18:49:18.790Z	4.8	188.8	0.84
13872	2012-02-26T23:08:47.640Z	4.4	97.5	0.98
6685	2005-01-26T12:45:22.140Z	4.4	110.4	0.66
14496	2010-06-13T16:38:13.850Z	4.3	135.6	0.91
5208	2005-05-29T23:27:43.490Z	4.9	78	0.65
519	2008-11-20T05:28:48.900Z	4	93.5	0.81
14381	2010-08-14T14:15:40.340Z	4.5	48.4	1.29
12455	2014-10-06T14:10:21.210Z	4.6	75	0.54
3740	2006-02-11T05:04:21.270Z	4.7	46.8	0.83
7356	2004-12-28T20:58:33.840Z	4.7	248.1	0.81
10597	2017-09-04T20:21:42.160Z	4.4	75	0.79