A Cross-country Analysis of Vulnerability to Natural Hazards Using Principal Component Regression[1]

Edward Nataniel C. Apostol

School of Statistics

University of the Philippines

December 2019

[1]A term paper submitted in partial fulfillment of the requirements in Statistics 226: Applied Multivariate Analysis under Prof. Francisco N. de los Reyes, 1st Semester A.Y 2019-2020.

ABSTRACT

This research aims to build a parsimonious model that explains a country's vulnerability to extreme natural events using socio-economic indicators. The indicators used in this study are the following: exposure index, GDP per capita, health expenditure per capita, percentage of internet users, literacy rate, mortality rate from common diseases, percentage of females in the population, percentage of population under poverty, percentage of tertiary school enrollment, unemployment rate, urban population, percentage of population using at least basic drinking water services, and percentage of working age population. Due to a high degree of association among the indicators mentioned, principal components were derived from the original set of variables and then regressed to the vulnerability scores. Three of the five principal components were found to be significant, namely, quality of life, percentage of women in the population, and urban population. The findings of this study may help the ongoing disaster risk reduction efforts of national and local communities.

INTRODUCTION

Extreme natural events such as earthquakes, floods, tropical cyclones, heatwaves, landslides and tsunamis are claiming the lives of thousands, and destroying billions worth of properties every year. Due to the scientific and technological advancements in the recent years, the capacity to predict incoming natural hazards have improved to some extent. Some countries are more exposed to certain natural hazards due to their geographical location. However, exposure, intensity, and duration of natural hazards alone will not determine its overall impact to a certain country or community. Some communities may be able to recover relatively quickly than others, while some may only have limited capabilities to cope with disasters. Exposure must be examined in conjunction with vulnerability, a multifaceted concept related to the societal factors and economic structures within a community that makes it susceptible to the negative impacts of natural hazards. In this regard, a research in disaster risk management should not only include the scientific aspects of natural disasters but also the interplay of social, political, and economic dimensions.

Several studies were already done in the past to uncover the indicators of vulnerability to natural disasters. This study aims to build the simplest possible intuitive model that will explain the vulnerability of a country to natural disasters using socio-economic indicators. This study aims to build a linear regression model that is simple, and can easily be presented to communities in line with the efforts for disaster awareness and preparedness. The final model should still be able to maximally explain the variability of vulnerabilities across different countries.

REVIEW OF RELATED LITERATURE

The United Nations Office for Disaster Risk Reduction (UNDRR), disaster risk is defined as "the potential loss of life, injury, or destroyed or damaged assets which could occur to a system, society or a community in a specific period of time, determined probabilistically as a function of hazard, exposure, and capacity". Risk can be thought of as how prone a certain community is for a natural hazard to turn into a disaster.

The WorldRiskReport, developed by the Development Helps Alliance, aims to evaluate the exposure of 180 selected countries to natural hazards and to assess the vulnerabilities of these countries to the adverse impacts when facing these hazards. In accordance to the definition of UNDRR, the World Risk Index is calculated by multiplying the score for exposure to the score for vulnerability. The exposure score is a function of the proportion of population exposed to natural hazards. While the vulnerability score is determined by susceptibility, lack of coping capacity, and lack of adaptive capacity. Susceptibility suggests how likely a community will adversely be affected by extreme natural events. Susceptibility depends on infrastructure, food supply, and economic framework conditions. Lack of coping capacity means having a limited capability to mitigate the impact of a natural event. Coping capacities depend on governance, healthcare, social and material security. Lack of adaptive capacities means lack of long-term strategies for societal change. Adaptive capacities depend on education, literacy rate, gender equality, agricultural management, and investments to public and private health.

Japan, a country with thousands of islands extending along the Pacific coast of East Asia, is located in a volcanic zone in the Pacific Ring of Fire, making it one of the most exposed countries to natural hazards. Regardless of this, Japan received a moderate score in the World Risk Index due to its very low vulnerability score. Japan has taken huge steps in disaster preparedness for the past few years. Their government was able to manage disaster risks through continuous revision of their basic disaster prevention law and through Public-private partnerships (PPPs) where the government attracts the private sector to invest in infrastractures and skills to address the challenges of natural disasters.

De Silva et al. (2018), confirmed in their study a positive correlation between disaster damage, poverty, and vulnerability. Poor families usually depend on agriculture to make a living. Crop destruction may result from natural events such as tropical storms or droughts that will really impose a challenge for the survival of these poor families.

Frankenberg et al. (2014) investigated on the extent on which one's education can provide protection during large-scale natural disasters. They concluded that education is associated with higher levels of resilience in the longer term. According to their findings, the better educated people have more capabilities to minimize dips or declines in their spending levels than their less educated counterparts. Also, the better educated people were also less likely to live in a camp or other temporary housing. This reflects the greater availability of financial and social resources of those who receive better education.

Kaigo (2012) investigated the role of social media during disasters. Even though social media has the potential to transmit false information quickly, the study provided evidence that social media, especially Twitter, is beneficial during disasters. Twitter became the main communication channel among the citizens of Tsukuba and their disaster countermeasures headquarters during the Great East Japan Earthquake. Twitter played a significant role in disseminating vital information during the disaster.

Due to its geographical location, the Philippines is prone to tropical cyclones. Yonson et al. (2017), developed a tool that estimates the tropical cyclone-induced fatalities in the provinces of the Philippines. Their study also tried to explain the variability of fatalities across provinces using an evidence-based approach. The study found a strong evidence that socioeconomic development and good local governance help reduce fatalities during disasters while unplanned urbanization is associated with more deaths.

METHODOLOGY

The data for this study were primarily obtained from the World Bank Open Data (https://data.worldbank.org). The percentages of the population living under the poverty line were obtained from the 2017 Central Intelligence Agency (CIA) World Factbook. The vulnerability and exposure indices were obtained from the 2019 WorldRiskReport (WRR). The data per indicator were downloaded separately and then combined into a single dataset. Countries with missing values for a specific variable were excluded from the analysis. A total of 168 countries were included in this study.

A. Definition of Variables.

The following variables will be considered for this study:

a. **Vulnerability index (%) (VUL)** - This is the dependent variable. Vulnerability is calculated as the mean of susceptibility (S), lack of coping capacity (LCC), and lack of adaptive capacity (LAC).

b. **Exposure (%) (EXPOSURE)** - Defined as the proportion of population exposed to natural disasters.

c. **Urban Population** (**URBAN_POP**) - Refers to people living in urban areas as defined by national statistical offices.

d. **Mortality from CVD, cancer, diabetes or CRD between exact ages 30 and 70 (%) (MORTALITY_DISEASES)** - Refers to the percentage of 30-year-old people who would die before their 70th birthday from any of cardiovascular disease, cancer, diabetes, or chronic respiratory disease.

e. **Population, female (%) (PERCENT_FEMALE)** - The percentage of the population that is female.

f. **Population living under the Poverty Line (%) (POVERTY)** - The percentage of the population living under the poverty line. Definition of poverty line varies per country.

g. **Unemployment, total (%) (UNEMPLOYMENT)** - Refers to the share of the labor force that is without work but available for and seeking employment.

h. **Literacy rate, adult total (%) (LITERACY)** -The percentage of people ages 15 and above who can both read and write with understanding a short simple statement about their everyday life.

i. **School enrollment, tertiary (%) (TERTIARY_ENROLLMENT)** - Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to tertiary level.

j. **GDP per capita (current US$) (GDPPC) -** The gross domestic product divided by midyear population.

k. **Population ages 15-64 (%) (WORKING_AGE)** - Total population between the ages 15 to 64 as a percentage of the total population.

l. **Current health expenditure per capita (current US$) (HEALTH_EXPENDITURE)** - Current expenditures on health per capita in current US dollars.

m. **Individuals using the Internet (%) (INTERNET_ACCESS)** - Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.

n. **People using at least basic drinking water services (% of population) (%)** - The percentage of people using at least basic water services. This indicator encompasses both people using basic water services as well as those using safely managed water services.

B. Principal Component Regression

Since many of the independent variables above are highly correlated in nature, principal components will be derived from the original set of predictors. A subset of the principal components will be regressed with the dependent variable in order to obtain a final linear model. The following are the set of assumptions used in the analysis:

a. All tests and variable selection processes uses 0.05 level of significance.

b. The relationship between vulnerability and the predictor variables are linear and subject to random error.

c. Error terms are assumed to be have a mean of 0, normal, homoscedastic, and uncorrelated unless there is a significant evidence to conclude otherwise.

RESULTS AND DISCUSSION

The correlation matrix (see Appendix) shows that several independent variables have a very strong correlation. Some of these are GDPPC and HEALTH_EXPENDITURE (r=0.93), WATER_ACCESS and WORKING_AGE (r=0.76) and TERTIARY_ENROLLMENT and INTERNET ACCESS (r=0.74). This justifies the use of Principal Component Regression to avoid the complications of multicollinearity.

Principal Component Analysis were performed on all of the 13 independent variables. The correlation matrix instead of the covariance matrix was used to obtain the principal components. Using a cutoff of 80% in the cumulative percentage variance criterion, the first 5 PCs were retained. The eigenvalues of the correlation matrix table is shown below in Table 1. Four of the five PCs that were retained have information explained larger than any of the original variables. The fifth principal component was included to reach the 80% cumulative variance target.

Eigenvalues of the Correlation Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	5.83089902	4.25523704	0.4485	0.4485
2	1.57566198	0.39252895	0.1212	0.5697
3	1.18313303	0.1064696	0.091	0.6607
4	1.07666343	0.13410906	0.0828	0.7436
5	0.94255437	0.22545118	0.0725	0.8161
6	0.71710319	0.30941884	0.0552	0.8712
7	0.40768435	0.05855621	0.0314	0.9026
8	0.34912814	0.0713773	0.0269	0.9294
9	0.27775083	0.02866546	0.0214	0.9508
10	0.24908537	0.05707545	0.0192	0.97
11	0.19200993	0.04353797	0.0148	0.9847
12	0.14847195	0.09861754	0.0114	0.9962
13	0.04985441		0.0038	1

Table 1: Eigenvalues of the Correlation Matrix

The loadings of each PC with the original variables were computed. The second to fifth PCs are strongly loaded to only 1 variable each. While the first PC is strongly correlated with 7 variables. The loadings for each PC are shown below in Table 2. Highlighted in bold are the PC where each variable is strongly loaded.

The first PC is strongly correlated with the indicators of the quality of life. PC1 takes high scores for observations with low mortality due to common diseases, low percentage of the population living under the poverty line, high literacy rate, high tertiary enrollment, high GDP per capita, high health expenditure per capita, high percentage of internet users, high percentage of population ages 15-64, and high percentage of people with access to at least basic drinking water services.

The second PC is strongly loaded with unemployment rate. The third PC is strongly correlated with the percentage of females in the population. The fourth PC is strongly loaded with the exposure index. And lastly, the fifth PC is strongly loaded with urban population. The retained PCs were labeled as follows:

a. PC1 - Quality of Life (QUALITY_LIFE_PC)
b. PC2 - Unemployment Rate (UNEMP_PC)
c. PC3 - Percentage of female in the population (PERCENT_FEMALE_PC)
d. PC4 - Exposure Index (EXPOSURE_PC)
e. PC5 - Urban Population (URBAN_PC)

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
EXPOSURE	-0.1034	0.1488	-0.4419	**0.7637**	-0.1613
URBAN_POP	0.2038	-0.3360	0.0582	0.2297	**0.8798**
MORTALITY_DISEASES	**-0.8158**	0.1741	-0.0981	-0.0031	0.1875
PERCENT_FEMALE	-0.0023	-0.1037	**0.7837**	0.5345	-0.1597
POVERTY	**-0.7851**	-0.1414	0.1100	-0.0994	-0.0268
UNEMPLOYMENT	0.0694	**0.5869**	0.4585	-0.2801	0.0921
LITERACY	**0.7959**	0.3425	0.0537	0.1072	-0.0384
TERTIARY_ENROLLMENT	**0.8256**	-0.0721	0.2427	0.0270	0.0100
GDP_PER_CAPITA	**0.7565**	-0.4896	-0.1075	-0.1573	-0.1955
HEALTH_EXPENDITURE	**0.7041**	-0.5960	-0.0104	-0.0896	-0.0664
INTERNET_ACCESS	**0.9263**	0.0474	-0.0207	-0.0054	-0.0248
WORKING_AGE	**0.7368**	0.4660	-0.2510	-0.0747	0.1527
WATER_ACCESS	**0.8413**	0.3134	-0.0375	0.1268	0.0672

Table 2: PC Loadings

The new set of independent variables (retained PCs) were regressed with the dependent variable VULNERABILITY. Stepwise variable selection procedure was used.

Summary of Stepwise Selection								
Step	**Variable Entered**	**Variable Removed**	**Number Vars In**	**Partial R-Square**	**Model R-Square**	**C(p)**	**F Value**	**Pr > F**
1	quality_life_pc		1	0.9479	0.9479	21.7208	3022.36	<.0001
2	urban_pc		2	0.0043	0.9522	8.5008	14.73	0.0002
3	percent_female_pc		3	0.0018	0.954	4.0881	6.41	0.0123

Table 2: Summary of Stepwise Variable Selection

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	32113	10704	1133.74	<.0001
Error	164	1548.44607	9.44174		
Corrected Total	167	33662			

Root MSE	3.07274	**R-Square**	0.954
Dependent Mean	46.40304	**Adj R-Sq**	0.9532
Coeff Var	6.62186		

Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	Intercept	**1**	46.40304	0.23707	195.74	<.0001	0
quality_life_pc		**1**	-5.72443	0.09847	-58.13	<.0001	1
percent_female_pc		**1**	-0.55342	0.2186	-2.53	0.0123	1
urban_pc		**1**	0.95523	0.24491	3.9	0.0001	1

Table 3: Result of Multiple Linear Regression

Three significant PCs out of the five PCs were retained in the model. The model has an R^2 value of 0.954, indicating that the three remaining PCs can explain 95.4% of the variability in the vulnerability scores. The reason for this high number is that even though we only have 3 predictors left, the principal components still contain most of the information from the original variables since each PC is a linear combination of all the independent variables.

The assumptions of multiple linear regressions were checked. Figure 1 displays the residual plots of the regressors for VULNERABILITY. Based from the residual plots, there are no signs of nonlinearity and other irregularities aside from possible outlying observations.

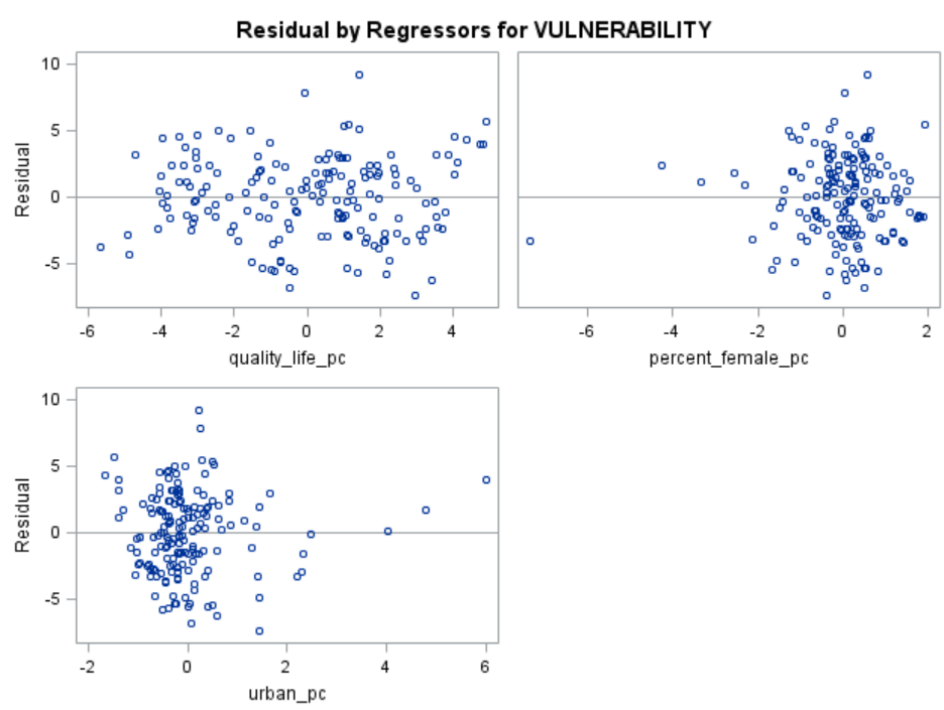

Figure 1: Residual plots of regressors for VULNERABILITY

There are also no signs of departures from normality since all the tests for normality fail to reject the null hypothesis of normality of the error terms.

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.993257	**Pr < W**	0.6297
Kolmogorov-Smirnov	**D**	0.044665	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.048503	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.321712	**Pr > A-Sq**	>0.2500

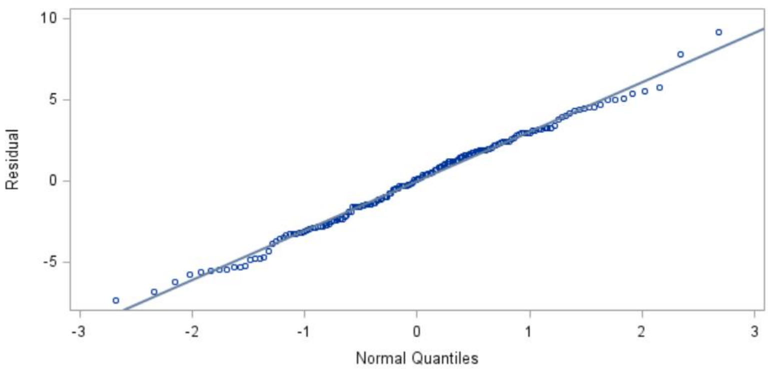

Table 4: Tests for Normality and Figure 2: Normal Quantiles Plot

The assumption of heteroscedasticity was also not rejected.

Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
9	6.06	0.7341

Table 5: Test for Heteroscedasticity

The VIF values and Condition Indices are all 1. This is expected since correlation matrix was used to obtain the principal components and since all the principal components are uncorrelated with each other.

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	**1**	46.40304	0.23707	195.74	<.0001	0
quality_life_pc		**1**	-5.72443	0.09847	-58.13	<.0001	1
percent_female_pc		**1**	-0.55342	0.2186	-2.53	0.0123	1
urban_pc		**1**	0.95523	0.24491	3.9	0.0001	1

Collinearity Diagnostics							
Number	Eigenvalue	Condition Index	Proportion of Variation				
			Intercept	quality_life_pc	percent_female_pc	urban_pc	
1	1	1	0	0	1	0	
2	1	1	0	0	0	1	
3	1	1	1	0	0	0	
4	1	1	0	1	0	0	

Table 5: VIF and Collinearity Diagnostics

Finally, five observations were tagged as possible outliers using the cook's distance cutoff of 4/n = 4/168 = 0.024. But since refitting the model without these observations will yield almost the same results, the observations were not dropped.

The final model obtained is given by:

VULNERABILITY = 46.40304 - 5.7244*QUALITY_LIFE_PC - 0.55342*PERCENT_FEMALE_PC + 0.95523*URBAN_PC

The quality_life_pc predictor was found to be significant in the final model. Note that this PC alone can already explain 94.79% of the variation in vulnerability scores. This makes sense because this PC is strongly correlated to 7 of the variables that are usually associated with vulnerability in various studies.

The percent_female_pc was also retained. Note that the parameter estimate for this predictor has a negative sign. This means that as the percent_female_pc score increases, the vulnerability decreases holding all the other predictors constant. This is an interesting result since females are stereotyped as physically weaker than males. But in this study, a significant evidence is seen that having a higher percentage of females can decrease the vulnerability score. One possible explanation is there are more college-educated women nowadays than men in some countries. Another possible reason is that women are seen as more altruistic than men. There have been some studies that show that women tend to be more charitable than men. In a disaster-stricken area, a generous helping hand can help people in coping with adversities.

The third PC retained is the urban population. As mentioned by Yonson et al. (2017), unplanned urbanization can increase vulnerabilities to natural disasters. Weak infrastructures, lack of rescuers and rescue equipment, and limited space for evacuation, are just some of the problems that might be encountered in a densely populated urban area.

Conclusion and Recommendations

The final model given above is a parsimonious model that can be used to explain the vulnerability score of a country. Three principal components were retained in the model. They are all significant and has an R^2 value of 0.954.

The first significant principal component is labeled as the quality of living in the country. It is correlated with 7 of the original indicators and will possibly decrease the vulnerability score when there is low mortality due to common diseases, low percentage of the population living under the poverty line, high literacy rate, high tertiary enrollment, high GDP per capita, high health expenditure per capita, high percentage of internet users, high percentage of population ages 15-64, and high percentage of people with access to at least basic drinking water services.

The second significant principal component is the percentage of females in the population. Females tend to be more educated, altruistic, and charitable than men so a higher female population can boost the morale of the disaster-stricken communities.

The third significant principal component is the urban population. Weak infrastractures, lack of rescuers and rescue equipment, and limited space for evacuation are some of the possible reasons why an increase in urban population tends to also increase the vulnerability score.

For future work, other predictors that are not included in the scope of this study such as the World Giving Index (WGI) and Technological Readiness Level (TRL) may be included in the analysis.

References

2019 WorldRisk Report. Retrieved from: https://reliefweb.int/sites/reliefweb.int/files/resources/WorldRiskReport-2019_Online_english.pdf

Frankenberg E., Sikoki B., Sumantri C., Suriastini W., Thomas D. (2013). Education, Vulnerability, and Resilience after a Natural Disaster. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/25170339

IRDR International (2016). World Risk Index. Retrieved from: http://www.irdrinternational.org/2016/03/01/word-risk-index/

Kaigo (2012). Social Media Usage During Disasters and Social Capital. Twitter and the Great East Japan Earthquake. Retrieved from: http://www.mediacom.keio.ac.jp/publication/pdf2012/KCR34_02KAIGO.pdf

National Institute of Disaster Management. Japan Country Profile. Retrieved from: https://nidm.gov.in/easindia2014/err/pdf/country_profile/Japan.pdf

Samori S. & Shibuya, N. (2018). Learning from Japan: PPPs for infrastructure resilience. Retrieved from: http://blogs.worldbank.org/ppps/learning-japan-ppps-infrastructure-resilience

United Nations Office of Outer Space Affairs. Disaster Risk Management. Retrieved from: http://www.un-spider.org/risks-and-disasters/disaster-risk-management

Yonson R., Noy I., Gaillard J. (2017). The measurement of disaster risk: An example from tropical cyclones in the Philippines. Retrieved from: https://onlinelibrary.wiley.com/doi/abs/10.1111/rode.12365

APPENDIX

Pearson Correlation Coefficients, N = 168							
	VULNERABILITY	**EXPOSURE**	**URBAN_POP**	**MORTALITY_DISEASES**	**PERCENT_FEMALE**	**POVERTY**	**UNEMPLOYMENT**
VULNERABILITY	1	0.10909	-0.1338	0.77922	-0.04902	0.75732	-0.07188
EXPOSURE	0.10909	1	-0.03024	0.11692	0.00966	0.01838	-0.12112
URBAN_POP	-0.1338	-0.03024	1	-0.09854	0.04579	-0.12329	-0.09093
MORTALITY_DISEASES	0.77922	0.11692	-0.09854	1	-0.07575	0.55196	0.02953
PERCENT_FEMALE	-0.04902	0.00966	0.04579	-0.07575	1	0.01732	0.05975
POVERTY	0.75732	0.01838	-0.12329	0.55196	0.01732	1	-0.00568
UNEMPLOYMENT	-0.07188	-0.12112	-0.09093	0.02953	0.05975	-0.00568	1
LITERACY	-0.79817	0.00176	0.05786	-0.59897	0.05393	-0.59554	0.19282
TERTIARY_ENROLLMENT	-0.81605	-0.17383	0.19914	-0.69561	0.1524	-0.59555	0.06854
GDP_PER_CAPITA	-0.73208	-0.11211	0.1214	-0.65266	-0.08135	-0.47744	-0.12352
WORKING_AGE	-0.69735	-0.02554	0.08066	-0.44465	-0.2385	-0.63447	0.16574
HEALTH_EXPENDITURE	-0.67516	-0.1183	0.2762	-0.61261	0.00611	-0.42683	-0.12607
INTERNET_ACCESS	-0.90985	-0.06814	0.15092	-0.73051	-0.01691	-0.69422	0.09102
WATER_ACCESS	-0.84219	0.02051	0.13648	-0.59134	0.00981	-0.68612	0.15095

Pearson Correlation Coefficients, N = 168							
	LITERACY	**TERTIARY_ENROLLMENT**	**GDP_PER_CAPITA**	**WORKING_AGE**	**HEALTH_EXPENDITURE**	**INTERNET_ACCESS**	**WATER_ACCESS**
VULNERABILITY	-0.79817	-0.81605	-0.73208	-0.69735	-0.67516	-0.90985	-0.84219
EXPOSURE	0.00176	-0.17383	-0.11211	-0.02554	-0.1183	-0.06814	0.02051
URBAN_POP	0.05786	0.19914	0.1214	0.08066	0.2762	0.15092	0.13648
MORTALITY_DISEASES	-0.59897	-0.69561	-0.65266	-0.44465	-0.61261	-0.73051	-0.59134
PERCENT_FEMALE	0.05393	0.1524	-0.08135	-0.2385	0.00611	-0.01691	0.00981
POVERTY	-0.59554	-0.59555	-0.47744	-0.63447	-0.42683	-0.69422	-0.68612
UNEMPLOYMENT	0.19282	0.06854	-0.12352	0.16574	-0.12607	0.09102	0.15095
LITERACY	1	0.61036	0.41803	0.69152	0.3546	0.71166	0.74323
TERTIARY_ENROLLMENT	0.61036	1	0.56664	0.49617	0.55741	0.74069	0.63202
GDP_PER_CAPITA	0.41803	0.56664	1	0.34501	0.92504	0.68344	0.45196

WORKING_AGE	0.69152	0.49617	0.34501	1	0.23058	0.69943	0.75597
HEALTH_EXPENDITURE	0.3546	0.55741	0.92504	0.23058	1	0.60792	0.39436
INTERNET_ACCESS	0.71166	0.74069	0.68344	0.69943	0.60792	1	0.77407
WATER_ACCESS	0.74323	0.63202	0.45196	0.75597	0.39436	0.77407	1

A1. Correlation Matrix

Simple Statistics							
	EXPOSURE	**URBAN_POP**	**MORTALITY_DISEASES**	**PERCENT_FEMALE**	**POVERTY**	**UNEMPLOYMENT**	**LITERACY**
Mean	16.64958333	16509832.48	15.93928571	50.1932836	0.2946785714	7.420416654	85.50753893
StD	12.21303173	33140983.35	5.59164299	2.30869494	0.1872911954	5.509755379	18.36537247

Simple Statistics						
	ARY_ENROLL	**GDP_PER_CAPITA**	**HEALTH_EXPENDITURE**	**INTERNET_ACCESS**	**WORKING_AGE**	**WATER_ACCESS**
Mean	37.56731102	13882.89507	1014.862339	53.73986283	63.1789779	87.52277435
StD	29.22977608	19747.52802	1766.249822	27.7634882	5.96543501	16.0209021

A2. Simple Statistics

SAS code used in the analysis:

```
proc import
        datafile = 'vul.xlsx'
        out = vulnerability
        dbms = xlsx
        replace;
run;

proc corr data = vulnerability noprob;
run;

proc princomp data = vulnerability out = vul_pc;
var exposure urban_pop mortality_diseases percent_female poverty
        unemployment literacy tertiary_enrollment gdp_per_capita health_expenditure
        internet_access working_age water_access;
run;

data vul_pc;
set vul_pc;
quality_life_pc = prin1;
unemp_pc = prin2;
percent_female_pc = prin3;
exposure_pc = prin4;
urban_pc = prin5;
run;
```

```
proc reg data = vul_pc;
model vulnerability = quality_life_pc unemp_pc percent_female_pc exposure_pc urban_pc / collin
vif dw spec selection = stepwise;
output out = res p = yhat r = resid student = s_resid cookd = cookd h = lev;
run;

proc print data = res;
run;

proc univariate data = res plot normal;
var resid;
run;
```