

Univariate Descriptive Statistics

Level	Introductory
Theme	Univariate Descriptive Statistics
Subthemes	Mean, Median, Standard Deviation, Minimum, Maximum, Histogram, Boxplot

REPLICATED ARTICLE

Neumayer, Eric. 2002. Do Democracies Exhibit Stronger International Environmental Commitments? A Cross-country Analysis. Journal of Peace Research, 39(2).

SECTION 1

Research Design

- 1 In a table format, identify all the dependent and independent variables used in the analyses presented in the article.
- 2 (If available) summarize the main theoretical argument(s) and write down all the corresponding hypotheses.
- 3 Discuss the measurement of each of the variable. In your discussion make sure that you identify the type of variable: discrete (ordinal, nominal, dummy), or continuous (interval or ratio) and their theoretical measurement.

SECTION 2

Mean, Median, Standard Deviation, Range

- 1 Identify which type of data are used in the study: survey, experimental, or observational. If multiple are used, identify all.
- 2 Identify what is the level of analysis (level at which variables are measured) in the dataset. If multiple, identify all.
- 3 Replicate the original table of descriptive statistics (Appendix 3) using your statistical software of choice. Present the replication in a neat table highlighting differences (if available) in comparison to the original table.
- 4 Choose three variables and calculate all descriptive statistics manually. Compare your results with software calculation. Note any differences (if applicable).

- 5 **BONUS** — Discuss what could explain the differences between your results and the original findings.

6 BONUS — Discuss the sample differences across variables. Identify which countries are missing and discuss the implications of missing countries for the sample representativeness. Present the missing countries in a neat table. Discuss whether the countries are likely to be randomly or systematically missing from the sample.

7 BONUS — Run the replication of the original table in two different statistical software. Identify any differences in the results and discuss why.

APPENDIX 3

Summary Descriptive Statistics for Variables

	N	Mean	Std. Dev.	Min	Max
<i>Dependent variables</i>					
Sum of MEAs	205	1.76	1.37	0	4
Number of EIOs	121	13.64	6.67	2	35
% of CITES requirements met	121	69.61	30.51	0	100
% land area under protection	154	7.43	7.67	0	42.6
Number of ESI variables available	122	54.48	5.03	47	65
<i>Independent variables</i>					
FREE	187	7.09	3.97	2	14
if FREE-low = 1	48	12.42	1.23	11	14
if FREE-mid = 1	55	8.23	1.45	6	11
if FREE-high = 1	84	3.31	1.15	2	6
POLIT	159	2.78	6.77	-10	10
if POLIT-low = 1	32	-7.47	1.32	-10	-6
if POLIT-mid = 1	46	-.48	2.87	-5	5
if POLIT-high = 1	81	8.68	1.34	6	10
GOV	171	0	.96	-1.79	1.69
if GOV-low = 1	40	-1.24	.29	-1.79	-.854
if GOV-mid = 1	53	-.39	.25	-.778	0
if GOV-high = 1	78	.89	.52	.013	1.69
VAN	183	15.74	12.69	0	43.54

	N	Mean	Std. Dev.	Min	Max
if VAN-autoc = 1	60	1.96	2.85	0	11.73
if VAN-demo = 1	123	22.46	9.88	5.85	43.54
lnGDP	175	8.33	1.10	6.13	10.42
lnPOP	205	15.13	2.26	9.83	20.94
POPdens	183	113.33	153.73	2	965

Source: Neumayer, 2002, pg. 161.

SECTION 3

Boxplot and Histogram

- 1 Plot boxplot for each of the continuous variables in Appendix 3.
- 2 Plot histograms for each of the continuous variables displayed in Appendix 3. Label carefully the x-axis and y-axis.
- 3 Compare the distributions for each continuous variable. Describe the shape of the distributions: normal, bimodal, skewed to the right, skewed to the left, leptokurtic, platykurtic.
- 4 Discuss the added benefit of displaying boxplot and histogram compared to the summary statistics in Appendix 3.

SECTION 4

Final Reflection

Reflect on what you learned through the process of trying to replicate the findings in Appendix 3. In your discussion, focus on issues of clarity and transparency of the research design, insights into measurement challenges, scientific process, sample bias etc...

INSTRUCTOR'S SUPPLEMENT

Before starting with a replication, students and teachers discuss the difference between exact replication and robustness. The students should try to replicate the exact findings in the Appendix 3. However, they may not entirely succeed in doing so. In case of discrepancies, students should systematically identify the potential sources of the differences – sample and measurement issues, transparency of the research design, and plausible own mistakes. Specificity is key, vague descriptions such as: “there is a bias in the research” or “data are not representative” is not sufficient. Students should point out what specifically is meant by “bias”. For example, “data are systematically missing for countries below UN poverty line, such as”

Learning Outcomes

- Identifying dependent and independent variables in a scientific study.
- Identifying and assessing theoretical argument and hypotheses in a scientific study.
- Learning to recognize the difference between theoretical and sample measurement of a variable.
- Calculating means, standard deviations, maximum and minimum, manually and by using a preferred statistical software.
- Introduction to missing data.
- Identification of types of data and the concept of the unit of analysis.
- Working with multiple types of statistical software.
- Thinking through basic aspects of data display and distributions. Reflecting on the added benefit of displaying data distributions compared to the summary statistics.

Section-by-Section Notes

Section 1

The goal is to bring students to critically engage with the research design of the paper. Students evaluate the coherence between the question, theoretical argument, hypotheses and measurement. Is there a logical connection between these research design elements? The focus can be also at a subsection of the argument. The portion on measurement differentiates between the theoretical measurement – i.e. theoretical range of the variable and in-sample measurements – actual range in the sample. If a variable is measured in terms of a percentage, then the theoretical range of that variable is 0–100%. For example, consider the variable in Neumayer’s article “Percent of country’s land under the protection status,” simply put the percentage of protected land in each country. This variable is a ratio variable with a theoretical range 0–100%. Of course, it is safe to assume that no country is composed of only protected land. But as we learn from the dataset, there exist a few countries with protected land close to zero. The theoretical range of a variable gives us the ability to assess descriptive statistics in context and enables students to interpret the effect of a variable in later analyses. In some studies as per different disciplinary standard, the theoretical argument is not articulated or is only implicitly mentioned. Students may then reflect on the disciplinary differences in research design norms.

Section 2

The replication of the Appendix 3 table can be done for the entire table or for a part of the results depending on the level of the students in class. Students learn to calculate basic summary statistics, such as: mean, median, standard deviation, minimum, and maximum. They can calculate the statistics manually or using a statistical software, or both. Students are asked to reflect on which types of summary statistics are useful for different variable types. The idea is to get them to think through the differences between discrete and continuous variables. Students are also asked to reflect on the missing data present in the dataset and point out large discrepancies between the variables. Students should analyze which specific countries are missing and how their “missingness” affects the interpretation of the summary statistics. The goal is to get students to think about non-randomness introduced into data through missing observations.

Section 3

Section 3 builds upon the replication and asks students to compare the advantages and disadvantages of summary statistics and univariate data display – specifically the comparative differences between histograms and boxplots. This is an introduction to data display. Students should focus on the labelling of axes, scaling of axes, and including appropriate titles for the graphs.