

Is A Person's Level of Education Associated With Their Beliefs Regarding the Cause of Global Warming?

Part 1: Introduction:

In this research project, I will investigate the relationship between a person's highest level of education and whether they believe global warming is caused by human activity, natural causes, or both causes equally. According to many climate change research articles published in highly cited scientific journals, there *is an association between global warming and human activity*. However, based on my preliminary data analysis, a large proportion of Americans polled believe that global warming is caused by natural mechanisms or caused equally by natural and human activity.

Therefore, this question is very important to me because I am interested in discovering potential reasons for the discrepancy between the broad consensus among climate change scientists and the general population. One possible reason may be related to person's level of education – where more education could be associated with believing that global warming is caused by human activities. Thus, it is important for the general USA public to determine why some American's do not believe the climate change research scientists and determine whether additional formal education might be associated with believing that humans are responsible for global warming.

In this research project, I will use polling data from an observational study, and therefore, will only be able to test for an association between a person's education level and their beliefs regarding global warming. I will be unable to make causal conclusions to suggest that additional education would cause a person to belief that global warming is caused by human activity.

However, the first step is to determine whether these two variable are independent or dependent, which will be investigated below.

Part 2: Data:

To investigate this research question, I will use polling data from the American National Election Studies (ANES). To collect the polling data, an interview was completed in early September 2012 to just before the 2012 presidential election. The data for the two variables of interest (education and global warming) came from the Pre-election or “Pre” interview group in the ANES data set and were not asked again after the election (“post” interview group). Face-to-face (FTF) interviewing data was supplemented with data collected from the internet. The persons were given interviews over the internet were a representative sample of the FTF sample. The FTF sample contained oversampling of blacks and Hispanics. For the first time in ANES testing, FTF interviews were given with CASI (Computer Assisted Self-Interviewing) tools (such as tablet computers).

Each respondent was asked the following two questions: (1) what is your highest education level, and (2) Do you think a rise in the world's temperatures would be caused mostly by human activity, mostly by natural causes, or about equally by human activity and by natural causes?

There are two categorical variables in the data set (1) Education, with 5 levels, and (2) Global Warming Cause, with 3 levels.

Education levels:

1. Less than high school credential
2. High school credential
3. Some post-high-school, no bachelor's degree
4. Bachelor's degree
5. Graduate degree

Global Warming Cause:

1. Mostly by human activity
2. Mostly by natural causes
3. About equally by human activity and natural causes

Below are the frequency counts for the two variables:

```
data_freq <- table(data)
data_freq
```

```
##           envir
## edu      Human Natural Equal
## Less HS    172    139   295
## HS        424    274   730
## Post-HS   621    383   947
## Bach      426    206   484
## Grad Deg  318    118   265
```

```
dim(data)
```

```
## [1] 5914    2
```

There are 5914 cases in the data set (indicated by the number of rows in the data set). Thus, 5914 persons were asked these two interview questions (indicated by the number of columns).

This is an observational study. The sample groups were not randomly assigned to different groups (e.g. experiment and control). The sample was randomly selected, but the sample was not randomly assigned to any groups before the interview – making this an observational study.

The population of interest is derived from the characteristics of the sampled population. For this study, both the FTF and internet sample groups included persons that were “U.S. eligible voters” and were all fresh (new) cross-sectional cases (with 2 over-sampled variables for ethnicity: blacks and Hispanics). The sample included persons from every state, a wide range of income levels, and genders. This suggests that these results may be generalized to the entire USA population of eligible voters. Biases may include the oversampling of black and Hispanic voters, or that these conclusions only apply to eligible USA voters, not the general opinions of ALL persons in the USA.

Because these data are observational (no random assignment of subjects to treatment conditions), therefore, these data cannot provide causal conclusions. Thus, we can only establish an association between the explanatory and response variables.

Part 3: Exploratory data analysis:

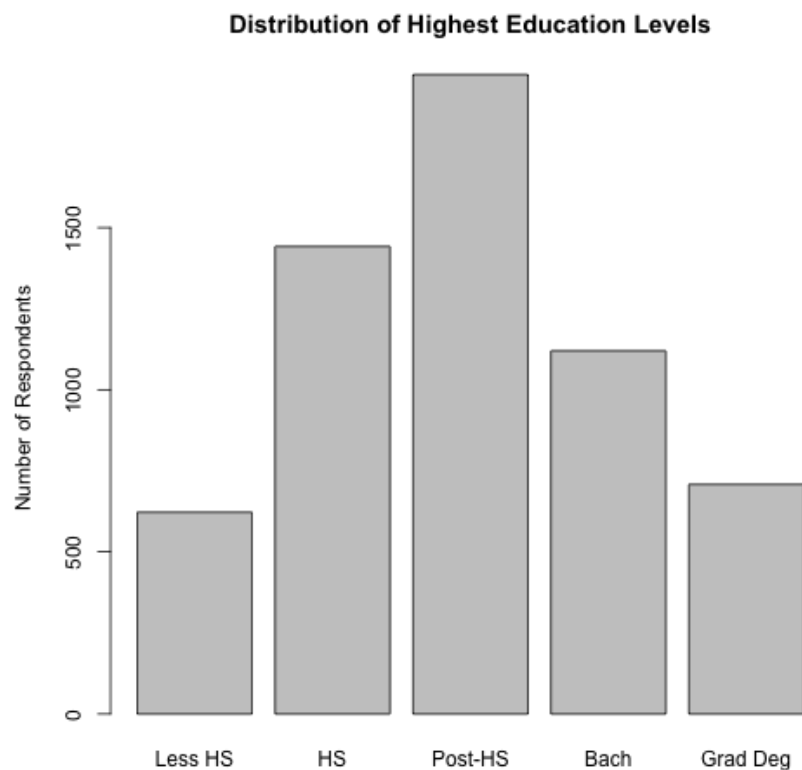
To begin this analysis, I will perform an exploratory data analysis using summary statistics (i.e. investigate the frequencies of each level for each variable) using tables and bar plots (see below).

Education Variable

```
summary(data$edu)
```

```
## Less HS      HS Post-HS      Bach Grad Deg      NA's
##      622    1442    1972    1120      708      50
```

```
plot(data$edu, main = "Distribution of Highest Education Levels", ylab = "Number of Respondents")
```



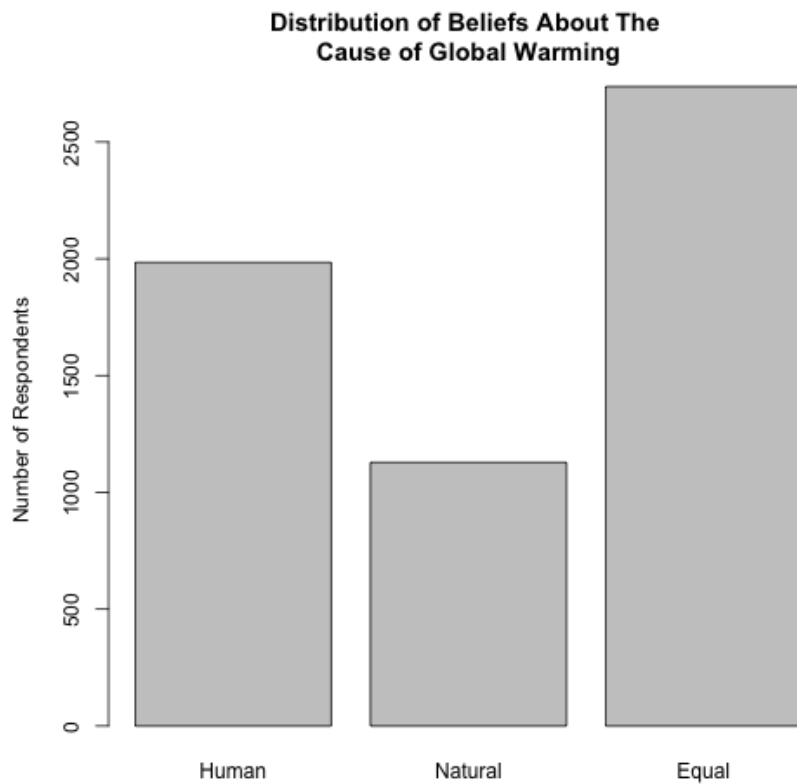
Looking at this plot, it appears education levels are fairly balanced among the polled respondents, where the greatest number of persons have the middle level: “Some Post-High-School, No Bachelor'S Degree.”

Global Warming Variable

```
summary(data$envir)
```

##	Human	Natural	Equal	NA's
##	1984	1128	2737	65

```
plot(data$envir, main = "Distribution of Beliefs About The \nCause of Global warming",  
      ylab = "Number of Respondents")
```



In addition, looking at the global warming data, these levels are about equally represented among the polled respondents, where most persons believe that global warming is caused “About Equally By Human Activity And Natural Causes.”

Another convenient way to look at categorical data is to use a mosaic plot. Below, you can see that as a person's education level increases, so does their belief that global warming is caused by human activity.

```
library(vcd)
mosaic(data_freq, legend = TRUE)
```

		envir		
		Human	Natural	Equal
edu	Less HS			
	HS			
	Post-HS			
	Bach			
	Grad Deg			

Therefore, with these preliminary summary statistics, I believe it will be important to use a statistical test to determine whether a person's education level and their beliefs about the causes of global warming are associated.

Part 4: Inference:

State Hypotheses:

Null Hypothesis: A person's highest education level and their beliefs about the causes of global warming are independent. A person's education level *does not vary* by their beliefs about the causes of global warming.

Alternative Hypothesis: A person's highest education level and their beliefs about the causes of global warming are dependent. A person's education level *does vary* by their beliefs about the causes of global warming.

Choose Appropriate Test:

We have a 2-way table of categorical data (with multiple levels for each variable), therefore we will use the *Chi-Squared Test of Independence* to investigate our hypotheses at the $\alpha = 0.05$ significance level.

Check Conditions for Test:

- Independence:* The sampled observations in this study were randomly sampled, without replacement, from the USA voting population. In addition, the n (number of cases in this analysis, 5914) are $< 10\%$ of the total population. In fact, after subtracting the "NA's" from the data, we have an $n = 5802$ which is still $< 10\%$ of the population. Each case only contributes to one cell in the table.
- Sample size:* Each cell in the table has at least 5 expected cases (see table below).

STEP 1. Find the expected counts for a 2-way table. Use the formula:

Expected = (row total * column total) / table total

The expected values are presented inside the parentheses in the table below (raw values were rounded so that the totals for all columns or rows add up properly):

Human Natural Equal Total

Less HS	172 (205)	139 (117)	295 (284)	606
HS	424 (483)	274 (275)	730 (670)	1428
Post-HS	621 (659)	383 (377)	947 (915)	1951
Bach	426 (377)	206 (216)	484 (523)	1116
Grad Deg	318 (237)	118 (135)	265 (329)	701
Total	1961	1120	2721	5802

STEP 2: Test the hypothesis that education level and beliefs about global warming are associated at the 5% level.
Find the Chi-Squared Statistic:

Chi-Squared = SUM ((observed - expected)²) / expected
(summation for every cell in the contingency table)

```
chisq.test(data_freq)
```

```
##
##  Pearson's Chi-squared test
##
## data:  data_freq
## X-squared = 77.86, df = 8, p-value = 1.318e-13
```

STEP 3: Interpret the results:

The Chi-Squared Test of Independence provided a $P\text{-value} = 1.318e-13$ which is much smaller than our alpha threshold (5%) and therefore, we can reject the null hypothesis and accept the alternative hypothesis: “A person's highest education level and their beliefs about the causes of global warming are dependent.”

We cannot report a confidence interval for this hypothesis test because they do not exist for Chi-Squared tests.

Part 5: Conclusion:

In this research project, I have investigated the relationship between a person's education level and their beliefs about the causes of global warming. I used a Chi-Squared Test of Independence to determine that these two variables *are indeed associated with each other with high statistical significance*. However, because these data were derived from an observational study (with random sampling, but not random assignment), *I cannot make any casual conclusions between these two variables* (i.e. I cannot conclude that more education causes a person to believe that humans are responsible for global warming).

Besides the conclusion that these two variables are associated, I was surprised to learn that more people believe that global warming is caused by BOTH human and natural causes (2721 persons), compared to these two levels alone (1961 and 1120 persons, respectively). In the future, I would like to extend this research question by doing an experiment that would test the causality of a person's education level and their beliefs regarding global warming. Only an experiment would allow me to test whether more education causes a person to believe that global warming is caused by human activity.

Reference:

The American National Election Studies (ANES; www.electionstudies.org). The ANES 2012 Time Series Study (data set). Stanford University and the University of Michigan (producers).

Appendix (one page of data):

	edu	envir
1	Less HS	Equal
2	Less HS	Human
3	Post-HS	Equal
4	Post-HS	Natural
5	HS	Equal
6	Less HS	NA
7	Less HS	Natural
8	Less HS	NA
9	Bach	Equal
10	Grad Deg	Equal
11	Bach	Human
12	Grad Deg	Equal
13	Post-HS	Equal
14	Post-HS	Natural
15	Bach	Equal
16	Grad Deg	Equal
17	Grad Deg	Natural
18	HS	Equal
19	Less HS	Equal
20	Less HS	Equal
21	Post-HS	Equal
22	HS	Equal
23	HS	Human
24	Less HS	NA
25	Less HS	Equal
26	Post-HS	Human
27	Post-HS	Natural
28	Post-HS	Equal
29	Less HS	Human
30	Post-HS	Equal
31	Bach	Human
32	Post-HS	Equal
33	Post-HS	Equal
34	HS	Equal
35	HS	Human
36	Post-HS	Human
37	Post-HS	Human