# POLS 418/GEOG 418: Quantitative Methods

Problem Set 4

Due June 12th, 2020

***Problem Set #4:***

***Hypothesis Testing using Cross-tabulation and Means Comparisons***

**Instructions:** Please complete each task listed below. You should type your problem set, and turn in a hard copy to me. You do not need to include a cover page. I expect, however, that your first page will include your full name, the problem set number, the due date, and the date you submit the work. The final page(s) of your assignment ***must*** include the STATA code you used, otherwise it is incomplete and it will receive the grade of “I”. Likewise, I will grade your assignment “I” if you use screenshots from STATA instead of making your own tables in MS Word.

In this problem set you will apply the techniques of crosstabulation and means comparisons to test some basic hypotheses. We will use data from the 2016 American National Election Study, which is posted on Blackboard. Below you will find four hypotheses**:**

This assignment is designed to familiarize you with the basics of hypothesis testing. For each hypothesis I will grade you on your use of the appropriate tabular hypothesis test; the correct usage of graphical techniques; the proper labeling of tables and graphs; and your associated written analysis of the hypothesis and data. That is to say, for each hypothesis I expect you to draw a substantive conclusion about the hypothesis from the data using crosstab or means comparisons and an associated graph. I expect all of your output to be labeled properly. You will have to do some work to clean the data before you can test your hypotheses.

For each, please provide each of the following:

1. Clean the data. Make a list of all the variables you will use to complete the assignment. In the ANES data missing values are often coded as negative numbers opposed to a “.”. STATA will not recognize these values are supposed to be interpreted as missing unless you replace these values to “.”. Replace these values as “.” as necessary for the variables you will be working with.
2. Set up the data to construct the necessary test, either a difference of means T-Test (Statistics>Summaries, tables, and tests>Classical tests of hypotheses>t test> Two-sample using groups) or some form of tabular analysis (Statistics>Summaries, tables, and tests>Frequency tables>Two-way tables with measures of association). Setting up these tests will likely require you to generate a new variable or replace some of the values in an existing variable.
3. Use the correct tabular analytic technique: crosstabulation or means comparisons. Include your table with appropriate titles, labels and other relevant details. Make sure you report either the p or chi2 value and tell readers whether the difference of means or association is statistically significant.
4. Produce a bar chart (Graphics>Bar chart) or line graph, as appropriate to the hypothesis and data. You will need to show the difference of group means if you are using a difference of means test. You will need to show the frequency of observations within categories if you are using a cross tabs.
5. Label all tables, charts and graphs appropriately, so that an unfamiliar reader can interpret you variables and quantities or percentages you report in tabular and graphical form. The best tables and graphs will be completely self-explanatory.
6. Make a conclusion about the hypothesis. Does your analysis support the hypothesis? Does it refute it? Or is it inconclusive? Make sure to report the p-value or the chi-squared value.

Here are the four hypotheses you will test:

*Hypothesis 1 (tabular):*

“In comparing individuals, those who hold an advanced educational (variable: education) degree be more likely to believe humans are causing climate change (variable: Climate\_Change) than those without an advanced degree.”

*Hypothesis 2 (difference of means T-Test):*

“In comparing individuals, women (variable: gender) will have warmer feelings towards feminists (variable: V162096) than men.”

Table 2. Difference of means T-Test comparing the means of men and women on the feminist thermometer

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|  |  |  |  |  |  |  |
| 1. Male | 1,664 | 51.69832 | .6234168 | 25.43052 | 50.47555 | 52.92108 |
| 2. Female | 1,874 | 59.96371 | .5978191 | 25.87942 | 58.79125 | 61.13618 |
|  |  |  |  |  |  |  |
| combined | 3,538 | 56.07631 | .4370327 | 25.99518 | 55.21945 | 56.93318 |
|  |  |  |  |  |  |  |
| diff |  | -8.265397 | .8646316 |  | -9.960624 | -6.57017 |

p-value=0.0000

t -statistic= -9.5594

Based on the p value and the t-statistic we reject the null hypothesis that men and women have the same mean in the population. We accept the alternative hypothesis that women will have warmer feelings towards feminists than men. The difference between the two groups is 8.3 points on a 100.0 scale. The difference is measurable but not huge.

A screenshot of a cell phone

Description automatically generated

*Hypothesis 3 (tabular):*

“In comparing individuals, those who attend church (variable: attend\_church) will be more likely to favor using the death penalty (variable: death\_penalty) than those who do not.”

*Hypothesis 4 (difference of means T-Test):*

“In comparing individuals, those who are identity as white (use variable: race) will be more likely to favor building a wall on the Mexican border (variable: build\_wall) than those who are do not identity as white.”

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|  |  |  |  |  |  |  |
| 1 | 3,027 | 4.162537 | .0437722 | 2.408266 | 4.076711 | 4.248363 |
| 2 | 1,191 | 5.065491 | .062785 | 2.166764 | 4.94231 | 5.188673 |
|  |  |  |  |  |  |  |
| combined | 4,218 | 4.417496 | .0366051 | 2.377359 | 4.345731 | 4.489262 |
|  |  |  |  |  |  |  |
| diff |  | -.902954 | .0801297 |  | -1.060051 | -.7458575 |

t = -11.2687

degrees of freedom = 4216

Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(­­­­T > t) = 1.0000