IS THERE A RELATIONSHIP BETWEEN A PERSON'S AFFLUENCE AND THEIR OPINION ON GOVERNMENT SPENDING FOR HEATLHCARE?

Introduction:

For many in the United States (U.S.), the subject of Government spending has long been a polarizing topic. Often, polarization occurs around the decision of whether or not this spending is too much or too little, and if that spending (regardless of level) goes towards worthwhile efforts. With the passage of the Affordable Care Act in 2010, the level of Government attention and spending on the healthcare of US citizens became one of the most hotly debated issues in the country, which continues to this day.

This study seeks to determine what relationship, if any, exists between the affluence of a person and their opinion on Government involvement in healthcare as a means to understand what may be driving some of the polarity on this subject. With a better understanding of some of the reasons for disagreement, lawmakers and other leaders may begin to address the rift in opinion and find a position of compromise on this subject with the American public.

Data:

To test for a relationship between a person's affluence and their opinion on Government spending on healthcare, this study will utilize results from the General Social Survey (GSS). These data were collected as part of a comprehensive survey of United States residents by the National Opinion Research Center between 1972 and 2012 to provide a measure of Americans' opinion on a variety of issues, as well as characteristics about the respondents' lives (such as employment status, level of education, marital status, etc.). To ensure a representative random sample, survey takers utilized a modified probability approach and quota elements at the block level. Both of these methods are in accordance with National Science Foundation standards for random representative samples.

For this analysis, the results of interest from the GSS are cases in which a respondent has provided both their family's level of income in constant dollars (the variable 'coninc') and their opinion on the amount of Government spending aimed at improving and protecting the nation's health (the variable 'natheal'). There are 29,094 cases that contain values for both variables in the dataset.

The level of income is a numerical variable measuring the affluence of the respondent's family with inflation effects normalized. The opinion on Government spending on the Nation's health is an ordinal categorical variable of three categories.

This is an observational study taking advantage of the GSS longitudinal data profiling United States residents' (the population of interest) opinions on a variety of questions. This is not an experimental study as the survey takers did not institute controls over variables.

Since the methodology of the survey was consistent with random sampling techniques accounting for proportions of all different segments of population to ensure accurate representation of U.S. residents, the results of this study may be generalized to the overall population of U.S. residents.

The survey respondents accounted for several possibilities of bias:

- disproportionate representation of groups of U.S. residents: the researchers used a modified probability approach and quota elements at the block level to survey a true cross-section of residents from various societal groupings
- non-responses: to mitigate the bias of non-response, survey takers only sought responses on weekends, holidays, or after 3:00PM on workdays
- non-English speaking respondents: researchers conducted the survey in English and Spanish for the significant amount of Spanish-speaking populations within the U.S.

Finally, since this study only takes the relationship between two variables (family income in constant dollars and opinion on government spending to improve or protect health) into account, and there are most certainly confounding variables affecting this relationship, we cannot rely upon the outcome of this study to establish a causal link between the variables. For instance, the political views variable may affect the survey respondents' opinion on level of government spending to protect or improve health in addition to their family income. With the presences of such confounders, any relationship between the two variables of the study may be considered correlation but not causation. One needs experimental data, with external/confounding variable control, to determine such a relationship.

```
# subset GSS data set to only variables of interest
StudyVars <- c("coninc", "natheal")
StudyGSS <- gss[StudyVars]
# subset data further to only valid cases (i.e., where both coninc and
# natheal have values other than NA)
SubStudyGSS <- subset(StudyGSS, !is.na(coninc) & !is.na(natheal))
```

Exploratory data analysis:

To begin working with this data, we take a look at some basic information on the data set:

```
# summarize the data
summary(SubStudyGSS)
```

```
##       coninc              natheal
##   Min.   :    383    Too Little :19396
##   1st Qu.: 18445    About Right: 7987
##   Median : 35722    Too Much   : 1711
##   Mean   : 43945
##   3rd Qu.: 58849
##   Max.   :180386
```

Immediately, we note that the vast majority of respondents believe the Government either spends too little or just the right amount on improving and protecting the Nation's health, with only a small minority thinking the Government spends too much. The income data demonstrates a general right-skew with the median less than the mean value and the large range between 75th and 100th percentile of data. For a closer look at this by category, we present a boxplot of the data:

```
ggplot(SubStudyGSS, aes(natheal, coninc)) +
geom_boxplot(aes(fill = natheal)) +
    scale_color_discrete(name = "Opinion") + labs(title =
"Opinion on Government Spending on Health v. Income",
    x = "Opinion on Amount of Gov't Spending on Health", y
= "Family Income [constant USD]")
```
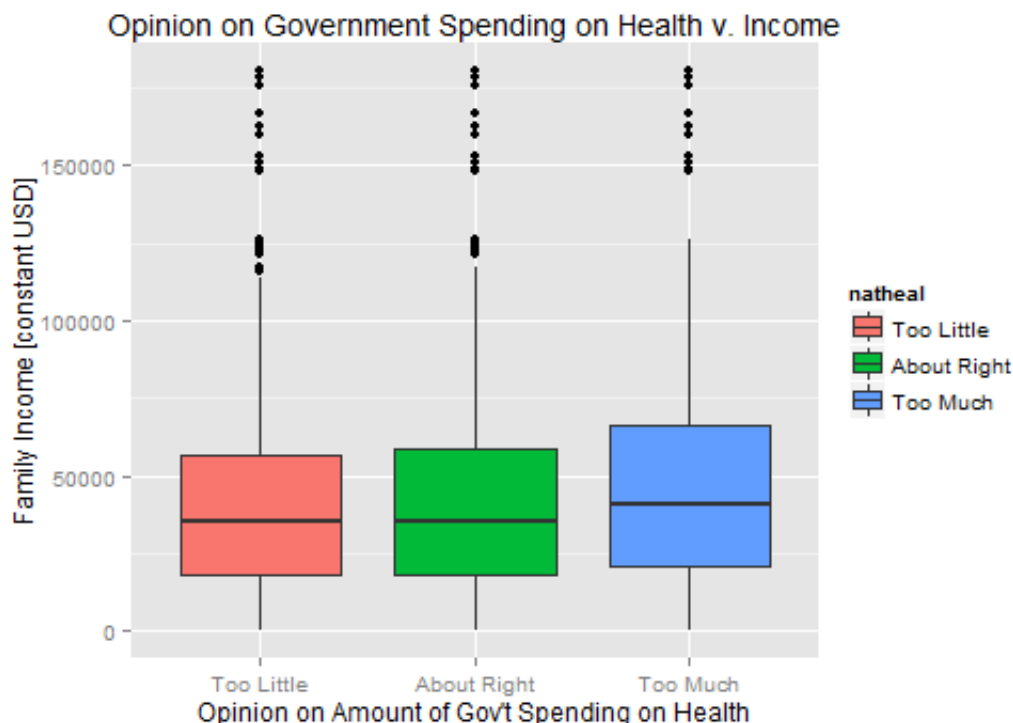

The boxplot allows us to expand on our earlier right-skewed conclusion to note that all opinion categories' income data is right skewed. With this income data now separated by category, we also note that the median income in constant dollars for survey respondents who believe

the Government spends too much improving or protecting the Nation's health is higher than income data in other categories. This may suggest a relationship a statistically significant difference in the income of respondents and their opinion on this healthcare spending.

Inference:

To determine if there is a relationship between income and opinion on Government healthcare spending, we will run a hypothesis test comparing the mean income of respondents based upon their income (ANOVA test), then pairwise comparisons of each opinion group's mean income if the hypothesis test returns a statistically significant difference in the means to determine which means are different (T-tests).

The hypotheses for this test are then: H0: the mean income is equal across all opinion groups (meanTL = meanAR = meanTM) HA: the mean income differs for at least one opinion group

Before testing these hypotheses, we must check the conditions of use for ANOVA:

1) Independence

- Within Groups: observations are independent per the random sampling methodology of the survey
- Between Groups: observations are independent per the random sampling methodology of the survey

2) Approximate Normality

- Each group's distribution appears to be nearly normal, with the large sample sizes sufficient enough to overcome any skew

3) Equal Variance

- The variances for each group appear to be roughly equal from the earlier boxplot; this is especially important given the disparity in sample size between the groups

As earlier stated, we will use an ANOVA test to determine if the mean income level varies across groups expressing different opinions on the amount of Government spending on healthcare. We use ANOVA because we are comparing one numerical variable (income) with one categorical variable (opinion) with more than two groupings. The ANOVA test will generate a measure of the variability between groups over the variability within groups as a way to determine if the means are significantly different and if the groupings offer a statisically valid relationship for that variability. If this test results in evidence of a statistically significant variation in the mean income between one or more groups, we will perform a T-test for the mean income for each possible pairwise comparison of two groups to determine which groups have the different means.

```
# perform ANOVA and summarize results
IncomeANOVA <- aov(coninc ~ natheal, data = SubStudyGSS)
ANOVATable <- anova(IncomeANOVA)
ANOVATable
```

```
## Analysis of Variance Table
##
## Response: coninc
##              Df    Sum Sq   Mean Sq F value   Pr(>F)
## natheal       2  7.28e+10  3.64e+10    29.8  1.1e-13 ***
## Residuals 29091  3.55e+13  1.22e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ' ' 1
```

With such a large F leading to a very small p-value (< alpha = 0.05 significance), we can reject the H0 and conclude that at least one mean income is statistically different across opinion categories. With that conclusion, we set about to compare the means among groups for statistically significant difference with T-tests. For each of these tests, we will judge significance using the adjusted significance alpha-star (Bonferroni Correction) to account error. For this study alpha-star is alpha=0.05/3 or 0.0167.

```
# Subset income data by opinion category
SubStudyIncTL <- subset(SubStudyGSS, natheal == "Too
Little")
SubStudyIncAR <- subset(SubStudyGSS, natheal == "About
Right")
SubStudyIncTM <- subset(SubStudyGSS, natheal == "Too Much")
# record 'n' for each category
nTL <- sum(SubStudyIncTL$natheal == "Too Little")
nAR <- sum(SubStudyIncAR$natheal == "About Right")
nTM <- sum(SubStudyIncTM$natheal == "Too Much")
# extract MSE value from table for SE calculations and dfE
value for
# T-test
IncMSE <- ANOVATable$"Mean Sq"[length(ANOVATable$"Mean
Sq")]
dfE <- ANOVATable$Df[length(ANOVATable$Df)]
# set Null difference in mean income
H0_Inc <- 0
```

Compare income means for the "Too Little" (TL) category vs. the "About Right" (AR) category

```
# this comparison has full commenting; subsequent
comparisons will not
# calculate point estimate for the difference in means
PointTLvAR <- mean(SubStudyIncTL$coninc) -
mean(SubStudyIncAR$coninc)
# calculate the Standard Error based on average variability
between groups
# (MSE) and n for each group
SE_TLvAR <- sqrt((IncMSE/nTL) + (IncMSE/nAR))
# calculate the T statistic for comparison on TL and AR
income means
T_TLvAR <- (PointTLvAR - H0_Inc)/SE_TLvAR
# Calculate two-tail p-value for this comparison
P_TLvAR <- 2 * pt(abs(T_TLvAR), df = dfE, lower.tail =
FALSE)
# report result
P_TLvAR
```

```
## [1] 0.0002315
```

This low of a p-value for the comparison of mean income between those that think the Government spends "Too Little" on healthcare and those that believe the Government's level of spending is "About Right" allows us to conlude there is a significantly significant difference in the means between the two opinion groups' incomes (p-value = 0.0002 < alpha-star of 0.0167).

Compare income means for the "Too Little" (TL) category vs. the "Too Much" ™ category

```
PointTLvTM <- mean(SubStudyIncTL$coninc) -
mean(SubStudyIncTM$coninc)
SE_TLvTM <- sqrt((IncMSE/nTL) + (IncMSE/nTM))
T_TLvTM <- (PointTLvTM - H0_Inc)/SE_TLvTM
P_TLvTM <- 2 * pt(abs(T_TLvTM), df = dfE, lower.tail =
FALSE)
P_TLvTM
```

```
## [1] 3.546e-13
```

With this p-value even smaller, we note that there is strong statistical evidence between the mean income of those who think the Government spends "Too Little" and those who think the Government spends "Too Much" on healthcare.

Compare income means for the "About Right" (AR) category vs. the "Too Much" ™ category

```
PointARvTM <- mean(SubStudyIncAR$coninc) -
mean(SubStudyIncTM$coninc)
SE_ARvTM <- sqrt((IncMSE/nAR) + (IncMSE/nTM))
T_ARvTM <- (PointARvTM - H0_Inc)/SE_ARvTM
P_ARvTM <- 2 * pt(abs(T_ARvTM), df = dfE, lower.tail =
FALSE)
P_ARvTM
```

```
## [1] 4.455e-07
```

Just as with the other two comparisons, there is a statistically significant difference in the mean income groups for these two opinion categories.

Conclusion:

In comparing the mean incomes of GSS survey respondents of varying opinion on the amount of Government spending to increase and protect the Nation's health, we determined (through an ANOVA test at the 0.05 significance level) that not only did one mean differ from others, but all means statistically varied from one another when following up the ANOVA test with a series of pairwise T-tests. This result leads us to conclude that there is correlation with U.S. residents' income level and their opinion on Government spending in this manner. For a visual depiction of the difference in mean income at the 95% confidence level, see the graph below:

```
ggplot(MeanInc_Category, aes(x = Inc.mean, y = natheal)) +
geom_point() + geom_errorbarh(aes(xmin = Lower,
    xmax = Upper), height = 0.3) + labs(title = "Opinion on
Government Spending on Health v. Mean Income",
    x = "Mean Family Income C.I. [constant USD] @ 95%
Confidence", y = "Opinion on Amount of Gov't Spending on
Health")
```

Opinion on Government Spending on Health v. Mean Income
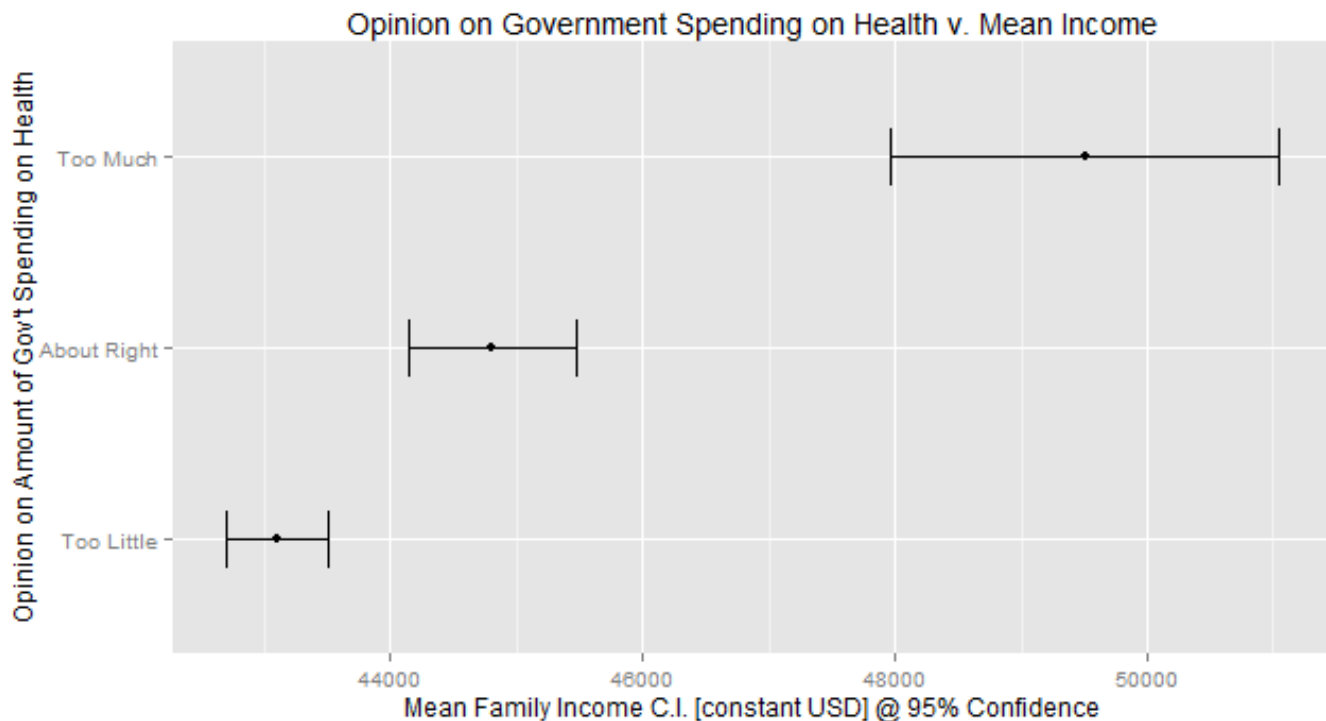
With this information, those seeking to justify or publicize Government spending on healtcare can target messages to U.S. residents of varying income levels to either win support (in the case of those with higher incomes) or advertise increased investment (in the case of those with lower incomes) to possibly bring the two opinions closer together and reduce the issue's polarity.

References:

Smith, Tom W., Michael Hout, and Peter V. Marsden. General Social Survey, 1972-2012 [Cumulative File]. ICPSR34802-v1. Storrs, CT: Roper Center for Public Opinion Research, University of Connecticut /Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributors], 2013-09-11. doi:10.3886/ICPSR34802.v1

Persistent URL: http://doi.org/10.3886/ICPSR34802.v1

Dataset URL: http://bit.ly/dasi_gss_data

Appendix - Data Sample:

```
# the first 25 cases in the study are presented here for
reference
SubStudyGSS[1:25, ]
```

```
##       coninc      natheal
## 1614   44643   Too Little
## 1615   26786   Too Little
## 1616   44643   Too Little
## 1617   44643  About Right
## 1618   44643  About Right
## 1619   44643  About Right
## 1620   16071  About Right
## 1621   26786   Too Little
## 1622    7143   Too Little
## 1623   62500   Too Little
## 1624   19643  About Right
## 1625   80357  About Right
## 1626   12500     Too Much
## 1627   62500   Too Little
## 1628   44643   Too Little
## 1629   19643   Too Little
## 1630   62500   Too Little
## 1631   16071   Too Little
## 1632   12500  About Right
## 1634   12500     Too Much
## 1635    1786   Too Little
## 1636   62500   Too Little
## 1637   12500   Too Little
## 1638   19643  About Right
## 1639   32143   Too Little
```