Alyssa Jaglowicz, Bradley Calkins, Jimmy Park

STAT:2010 research paper

5/1/17

**Does Education Level Affect Unemployment Rate?**

In recent years, enrolling in college following high school has become increasingly popular, evidenced by the percentage of the population aged 18 to 24 enrolled in a college (40%) being 4.5% higher in 2014 than in 2000 (NCES). A safe assumption would be that the overwhelming majority of students attending college are doing so to earn a degree that can get them a job upon graduation. There are plenty of jobs out there that don't require a college degree though, so what's the incentive behind obtaining one? We determined that pay, security, and mobility are the most important aspects of a job, so being better off in even one of those categories would be an incentive. Since businesses are looking for the most qualified employees, and those with college degrees are generally labeled as such, they will often be willing to compensate at a higher rate and hire college graduates at a higher rate, which many would view as a big benefit in what seems to be an unpredictable United States economy. Since we don't know for sure if earning a college degree does indeed have an impact on job security (unemployment), we decided to test our theory.

To test our theory, we needed to find data on people that fit the description of unemployed (someone who can, and wants to, work and be in full employment, but cannot find a job despite actively looking for one). Each month, the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor finds the total number of unemployed people in the U.S. and announces an unemployment rate. They then release the demographics of the unemployed population and share large amounts of data about those demographics on their website. Out of the datasets on the Bureau of Labor Statistics website, we targeted “people in the United States aged 25 and older with a high school diploma only” and “people in the United States aged 25 and older with a bachelor’s degree.” We converted the given monthly unemployment rates to annual rates and looked at the last 21 years (1996-2016). With these annual rates, we can figure out if there is a difference in unemployment rates between the two populations. Once we record the annual rates, we will compare the sample mean for each education group so that we can make inferences about the entire populations.

Since t procedures can be used, although not entirely accurately in some cases, as long as the sum of the sample sizes are greater than 40 and the samples are both independent, we will do a two-sample t-test in SAS using “proc ttest”. We tested the following hypotheses:

Ho: μ1 = μ2, or μ1-μ2 = 0

Ha: μ1 ≠μ2, or μ1 - μ2≠ 0

α= .05

The “1” represents the people with high school diplomas only and the “2” represents the people with bachelor’s degrees. The first variable is called “education”, which holds values 1 or 2 to differentiate the group’s education level. The second variable “rate”, is the corresponding annual unemployment rate in percentages through the 21 years. Our SAS code and output looks as shown on page 4.

From the output, we found the mean of group one to be 5.75, with a standard deviation of 2.07 and standard error of 0.45. Group two had a mean of 2.83, with a standard deviation of 2.07 and a standard error of 0.21. We would reject the null hypothesis because our p-value is less than our alpha level of 0.05. Our p-value basically says that if the null hypothesis is true and that the means are equal, then there is less than a .01% chance of recording a sample as, or more, extreme than the one recorded. Since that is incredibly unlikely, we should reject the null hypothesis. We would also say that we are 95% confident that the population mean of group one is between 4.8 and 6.7 and that the population mean of group two is between 2.4 and 3.3. However, as you can see in the stem plots above of our data, there are obvious outliers in education group 1 from the recession that the United States is still recovering from. Since t procedures are strongly influenced by outliers, we decided to log-transform the unemployment rates in SAS to get more symmetric distributions of values, hopefully reducing our outliers. Our new SAS code and output is as shown on page 5.

The new mean for group one was 1.69, with a standard deviation of 0.32 and a standard error of 0.07. The new mean for group two was 0.98, with the same standard deviation and standard error of group one. This seemed to fix our problem. There is obviously less variability in the log-transformed data and it reduced our outliers. We are now safe to make conclusions about this data.

The calculated degrees of freedom was 40 and test statistic was t\*=7.04. The probability of recording a sample test statistic more extreme than that (aka the p-value) was <.0001. At a significance level of alpha=0.05, we reject the null hypothesis. Since the p-value was less than our significance level, we have statistically significant evidence to support the alternative hypothesis that the population means are not equal. The test was two-sided, so although the data would suggest that those without a bachelor’s degree would have a higher unemployment rate, we can only conclude that based on the test, the unemployment rates are not equal. Of the log-transformed data, we are 95% confident that group two’s population mean falls within the interval (0.83, 1.13), whereas we are 95% confident that group one’s population means falls within the interval (1.55, 1.84). No overlap between the confidence intervals further supports the fact that the population means are not equal.

To answer our research question: does educational level affect unemployment rate? The two-sample t test for logged unemployment rates produced output that would suggest that education level does affect unemployment rates. A person aged 25 years or older with a bachelor’s degree is not as likely to be unemployed as a person aged 25 years or older with only a high school diploma. If the goal is to simply have a job following high school, pursuing a bachelor’s degree would be best the best option, on average, based on our test.

**SAS Code and Output: t procedure**

**data** project;  **proc** **ttest** data = project;

input education rate;  class education;

datalines;  var rate;

1 4.65  **run**;

1 4.225

1 3.97

1 3.52

1 3.45

1 4.18

1 5.26

1 5.45

1 4.96

1 4.71

1 4.3

1 4.38

1 5.68

1 9.73

1 10.29

1 9.43

1 8.31

1 7.52

1 6

1 5.43

1 5.2

2 2.25

2 2.01

2 1.79

2 1.81

2 1.66

2 2.26

2 2.91

2 3.08

2 2.72

2 2.3

2 2.03

2 2.03

2 2.56

2 4.61

2 4.71

2 4.31

2 4.36

2 3.7

2 3.18

2 2.57

2 2.51

;

**run**;

**SAS Code and Output: Log-form transformation**

**proc** **ttest** data = project;

class education;

var rate1;

**run**;

**data** project;

input education rate;

rate1 = log(rate);

datalines;

1 4.65

1 4.225

1 3.97

1 3.52

1 3.45

1 4.18

1 5.26

1 5.45

1 4.96

1 4.71

1 4.3

1 4.38

1 5.68

1 9.73

1 10.29

1 9.43

1 8.31

1 7.52

1 6

1 5.43

1 5.2

2 2.25

2 2.01

2 1.79

2 1.81

2 1.66

2 2.26

2 2.91

2 3.08

2 2.72

2 2.3

2 2.03

2 2.03

2 2.56

2 4.61

2 4.71

2 4.31

2 4.36

2 3.7

2 3.18

2 2.57

2 2.51

;

**run**;

**References**

*National Center for Education Statistics (NCES)*. U.S. Department of Education. <https://nces.ed.gov/fastfacts/display.asp?id=372> Accessed 30 April 2017.

*Current Population Survey “How the Government Measures Unemployment”.* U.S. Bureau of Labor Statistics. June 2014.

Pettinger, Tejvan. “Definition of Unemployment”. Economics Help, 3 June 2010, <http://www.economicshelp.org/blog/2247/unemployment/definition-of-unemployment/>. Accessed 30 April 2017.

Work:

1. Brainstorming research topics: all 3 members
2. Typing up project proposal: Alyssa
3. Typing up interim report: Alyssa
4. Writing and editing final paper: all 3 members