Project Part 3

Mai Luu

Question

Determine whether the mean of unemployment in 2017 is greater than the mean of unemployment in 2018 conducted on a random sample of 10 pairs

Brief Data description

The data contains a record of Employment and Unemployment in 4 States: Alabama, Alaska, Arizona, and Arkansas surrounding in 2017 and 2018. The purpose of the data is to indicate the correlation between the number of unemployment and employment between the four States. The data was collected by USDA ERS- USDA's Economic Research Service to anticipate trends and emerging issues in population, environment, and employment rate (Reference 1). The type of data is a sample and not the whole population. Therefore, it does not conclude the data in the past 10 years. Citation: Pender, John. "Download Data." USDA ERS - Download Data, USDA ERS, 5 Feb. 2020, <www.ers.usda.gov/data-products/county-level-data-sets/download-data/>

The dataset details rural or urban code. If the continuum code is from 1-3, which means it is in Rural. Otherwise, the area is in the Urban area (Reference 2). The data has 185 observations which are in order from State AL to AR in alphabetical order. The data also provides the number of Employment, Unemployment and Median Household Income so the audience can easily compare the employment rate of each state in two consecutive years

Test selection and Generalization

Test selection: I will be using Paired T-Test to solve the question that I conducted. I chose the Paired T-Test because it is commonly used to compare two population means of the same

group of items. Moreover, the trend of type I error of the Paired T-Test is more consistent than the other T-Test (two samples T-Test). My sampling dataset does not include the data throughout at least five consecutive years. The data only contains information from 2017 to 2018. Thus, I did not know the normal distribution of the mean and the standard deviation of the unemployment rate of the entire population through each year. Therefore, I believe using the Pair T-test will be more reasonable than two samples T-test. The sample size is 10 pairs, which meet the requirement according to the general rule for the Central Limit Theorem (Reference 3). Paired T-test is conducted when the sample size is 15 or less. The Paired T-Test assumes that the observations are independent of one another. Since we have met all the requirements, I believe the Paired T-Test is reasonable in this case. The reason I conducted the test with Monte Carlo Simulation is to model the probability of different outcomes that cannot be easily predicted.

Generalization: The paired T-Test helps to compare the difference between the unemployment rate in 2018 and 2017. If the p-value is smaller than 0.05, it means we reject the null hypothesis and accept the alternative hypothesis. In other words, the average of unemployment in 2017 is greater than the average of unemployment in 2018. If the p-value is greater than 0.05, we fail to reject the null hypothesis and there is no evidence to show that the first population is greater than the second one. The results will partly contribute in further analysis and recommendation whether the average of unemployment in 2018 and the previous year is higher than 0.

Code

```
set.seed(12201996)
m0<-0
pvalue<-0.05
k<-10000
size<-10
do.test<-function(x){
value<-Data_description[sample(nrow(Data_description),10),]

test<- t.test(value$Unemployed_2017,value$Unemployed_2018,mu=m0,
alternative="greater",paired=TRUE)
test$p.value<=pvalue
}
rep.test<-function(x){
```

```
res<-replicate(10000,do.test(x))
sum(res)/10000
}
p_test<-sapply(size,rep.test)
p_test
```

```
## [1] 0.0619
```

Conclusion:

The p-value is high, which is greater than the significant level of 0.05. Thus, we fail to reject the null hypothesis with strong evidence that the unemployment rate in 2017 is not greater than the unemployment in 2018 with 10 pairs randomly selected. This is a sign that the unemployment rate is not decreased from 2017 to 2018. The implications are very important and helpful to the analyst since the data will indicate whether the economy of the four States is stable and prosperous or not. Since the 10 pairs of random selection in 2017 is not greater than 2018, this will be a time for the Government to take action because the residents would reasonably want more job opportunities available and a decrease in the unemployment rate.

References

1. https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/.
2. https://seer.cancer.gov/seerstat/variables/countyattribs/ruralurban.html
3. https://www.statisticshowto.com/large-enough-sample-condition/