III. Discussion of statistical tests or methodology to be used and why, including discussion of assumptions if they are required to perform any mentioned test.

1. How do education and capital gains affect the hours per week for working?

We plan to use the regression to address this problem. Regressions are used when we want to find the scale effect of how variables can affect a dependent variable. In this case, we hope to see answers to two basic questions: (1) if people with higher education tend to work shorter and (2) if people with extra capital gains tend to affect their decisions of how long to work.

To be more specifically, we plan to use the Bootstrapping Regression Method to address the problem. Bootstrapping is a nonparametric approach to statistical inference that substitutes computation for more traditional distributional assumptions and asymptotic results. As the simple multiple linear regression requires normal distribution for the data, we can not simply apply that to our analysis as the variables Hoursperweek, EducationNum and CapitalGain are all skewed and not normally distributed due to our small data size and the nature of those data. The bootstrap can provide more accurate inferences when the data are not well behaved or when the sample size is small. It is also possible to apply the bootstrap to statistics with sampling distributions that are difficult to derive, even asymptotically. Further, the bootstrap regression will provide us with a narrower confidence interval.

We are going to construct a bootstrap interval for the slope of the line for predicting Hoursperweek from numbers of education and extra capital gains. We are performing the standard test of slope:

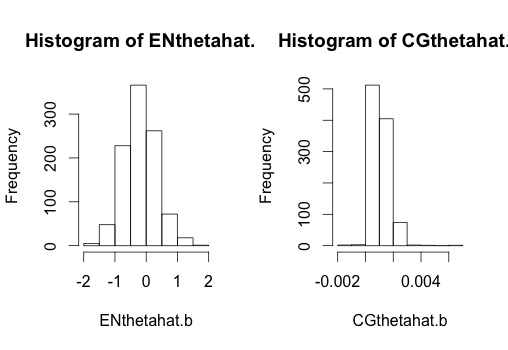
*H0 : β1 = β2 = 0* versus *H1 : β1 ≠0, β2 ≠ 0.*

IV. Results of applications of statistical tests to respective data

1. How do education and capital gains affect the hours per week for working?

|  |  |  |  |
| --- | --- | --- | --- |
| Independent Variable | 95% Lower Bootstrap Interval | 95% Upper Bootstrap Interval | P-value |
| EducationNum | -1.141598 | 0.9127412 | Greater than 0.05 |
| CapitalGain | 0.0002935431 | 0.002403871 | Smaller than 0.05 |

Table. Histogram of bootstrap distributions



V. Discussion of results

1. How do education and capital gains affect the hours per week for working?

The 95% bootstrap confidence interval for the slope of Education Numbers is (-1.141598, 0.9127412). The p − value of SAT will be greater than 0.05 because 0 is in the confidence interval. We can make a conclusion that we have strong evidence that the slope may be equal to 0. Education may not affect working hours per week.

The 95% bootstrap confidence interval for the slope of Extra Capital Gains is (0.0002935431, 0.002403871). The p − value of SAT will be less than 0.05 because 0 is not in the confidence interval. Since the bootstrap interval is so close to 0, we can make a conclusion that we have weak evidence that the slope may not be equal to 0. Capital Gains may affect working hours per week.

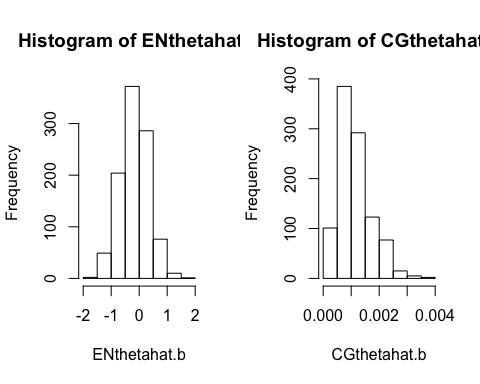
VII. Software/ Code Appendix

adult=read.csv("~/Desktop/Fourth-Year Second Semester/STAT 3480/Final/Adult.csv")  
attach(adult)  
  
### create our data  
oursample = adult  
ENthetahat = lm(HoursPerWeek~EducationNum+CapitalGain,data=oursample)$coeff[2]  
CGthetahat = lm(HoursPerWeek~EducationNum+CapitalGain,data=oursample)$coeff[3]  
ENthetahat; CGthetahat;

## EducationNum   
## -0.1740954

## CapitalGain   
## 0.0009557792

ENthetahat.b = rep(NA,1000); CGthetahat.b = rep(NA,1000);   
for (i in 1:1000) {  
 ### draw the bootstrap sample and calculate thetahat.b  
 index = 1:89  
 bootindex = sample(index, 89, replace=T)  
 bootsample = oursample[bootindex,]  
 ENthetahat.b[i] = lm(HoursPerWeek~EducationNum+CapitalGain,data=bootsample)$coeff[2]  
 CGthetahat.b[i] = lm(HoursPerWeek~EducationNum+CapitalGain,data=bootsample)$coeff[3]  
}  
par(mfrow=c(1,2))  
hist(ENthetahat.b); hist(CGthetahat.b);



quantile(ENthetahat.b, .025); quantile(ENthetahat.b, .975)

## 2.5%   
## -1.187911

## 97.5%   
## 0.8282693

quantile(CGthetahat.b, .025); quantile(CGthetahat.b, .975)

## 2.5%   
## 0.0003073671

## 97.5%   
## 0.002459157