**Return to Education**

The data includes 3,010 men aged between 24 and 34 with 34 variables that is collected from the Young Men Cohort of the National Longitudinal Survey.

Firstly, you should read description data on second spreadsheet. Then, you answer the following questions.

Use the following R-code to take the subsample (You can use Python:

id<- 459 # Your ID: SE140459 🡺 id <- 459

set.seed(id)

yourdata<-wage[sample(1:nrow(wage),30),]

# generate a random sample with size of 30 from the data.

Topic: Probability

1. Choose a simple random sample with size of 30 from the population. What is the probability that there are six black ones.
2. Choose 5 people in your subsample randomly. Find the probability that there is at least one black.

Topic: Discrete Random Variables and Probability Distribution

1. Suppose that the proportion of blacks in the population equals the proportion of backs in your subsample.
2. How many blacks do you expect in your subsample?
3. How many blacks are in the data?
4. Suppose that number of education (educ) is a discrete random variable that has a Poisson distribution such that its mean equals to the sample mean in your subsample. If we randomly choose a person, what is the probability that he/she has 6 years in education?
5. Assume age is an integer-valued discrete uniform random variable, 24 ≤ age ≤ 34. Find the mean and the variance of age. Compare these corresponding values to the values in your subsample. Create the bar chart for age. Is age a uniform distribution?

Topic: Continuous Random Variables and Probability Distribution

1. Suppose that lwage is a normally distributed random variable with mean μ = 6.26 and standard deviation σ = 0.44. Find the probability P(5 < lwage < 6).
2. Assume , where and .
3. For your random sample of n = 20 observations. Find the probability that the sample mean of IQ lies in [100, 110].
4. How large must the random sample be if we want the standard error of the sample mean to be 1?

Topic: Descriptive Statistics

1. Calculate the statistics for wage, IQ, educ and exper.

Topic: Sampling Distributions and Point Estimation of Paramaters

1. Find the point estimation of mean and variance of IQ.

Topic: Statistical Intervals for a Single Sample

1. Suppose that lwage is a normally distributed random variable with standard deviation σ = 0.44. Construct a 95% confidence interval on the true mean lwage using data in your subsample.
2. Construct a 96% confidence interval on the true mean IQ using data in the subsample. Assume that IQ is a normally distributed random variable.
3. If we want the error in estimating the mean lwage from the two-size confidence interval to be 0.2 at 95% confidence. What sample size should be used? Assume that lwage is a normally distributed random variable with standard deviation σ = 0.44.
4. Calculate a 99% confidence interval on the true proportion of all people who near 4-year college (nearc4).
5. If we want the error in estimating the proportion of blacks from two-size confidence interval to be 0.01 at 99% confidence. What sample size should be used?

Topic: Test of Hypotheses for a Single Sample

1. Use your subsample to test the hypothesis H0: mean(IQ) = 100 against H1: mean(IQ) ≠ 100 at . Assume that IQ is a normally distributed random variable with standard deviation σ = 15.
2. Use your subsample to test the hypothesis H0: mean(lwage) = 6 against H1: mean(lwage) > 6 at . Assume that lwage is a normally distributed random variable.
3. It is argued that the proportion of people with less than 10 years of work experience is not more than 0.07. Use the given data to test this argument at .
4. Assume that the variables are normally distributed and the variance within each population be equal for all populations.
5. Test the difference between the averages wage by variable black at 1% significance level based on the subsample.
6. Test the difference between two independent population proportions that people with low IQ (IQ < 90) by variable black at 5% significance level based on the data.
7. Test the difference between the variances lwage by variable black at 1% significance level based on the subsample.

Topic: Simple Linear Regression and Correlation

1. Suppose you are interested in studying the effect of education on wages. Use your subsample to calculate the regression equation of educ on wage.
2. Interpret the slope and the intercept for the output.
3. Test the significance of the regression () at 10% significance level.
4. Find the coefficient of determination and explain its meaning.
5. Test the significance of the regression () at 10% significance level.
6. Use your subsample to find the regression line of educ on lwage. Which model is better? Why?