Review of Statistical Inference

Submit your answers in hard copy only. (no rough draft)

1. *Cross Section Data*

For this question, use the 1993 data from the Survey of Income Dynamics (PSID) on annual earnings (measured in dollars) and years of completed schooling for 21 year-old males and females. The data are available on the course web site.

"Clean" the sample on females by eliminating missing values "99" and those reporting 0 income. Save this as a new file.

(a) Construct a table with the following information for earnings and education for men and women:

- Mean
- Median
- Standard Deviation

You can organize your table however you like, but make sure the information is clear and easy to read. You could to organize your table as follows (for example):

Table: Statistics on earnings and education for 21 year-old men and women from PSID

	Men		*Women*	
	Earnings	Education	Earnings	Education
Mean				
Median				
Standard Deviation				

(b) Conduct a hypothesis test that women earn the same income as men. You can assume that the average men's income is known (i.e., you can treat this as a hypothesized value, rather than a random variable). Do you reject the null? What does this imply about earnings across genders? Report your t-statistic.

(c) Conduct a hypothesis test that women have the same level of education as men. You can assume that the average men's wage is known (i.e., you can treat this as a hypothesized value, rather than a random variable). Do you reject the null? What does this imply about earnings across genders? Report your t-statistic.

Create one data file that combines data for men and women.

(d) Compute the correlation between earnings and education. Is the correlation positive or negative?

(e) Construct a scatterplot of earnings and education (for both men and women combined).

(f) Construct a scatterplot of earnings and education for men, and then construct one for women.

(g) Taking what you have done above, comment on any important differences/similarities in the earnings of men versus women. You should make use of all of your statistics and graphs from above to support your statements. Are there any variables that could help explain this gap between the earnings of 21 year-old men and women? Identify at least two other explanatory variables that might explain why there appears to be a wage gap. This explanation should be about 5-6 sentences long.

2. Time Series Data

This section makes use of time series data. While cross section data often need to be "cleaned" and organized, time series data often need to be transformed.

Go to the Federal Reserve Economic Database (FRED) and obtain the following seasonally-adjusted monthly data:

- Consumer Price Index (CPI) – All items
- Consumer Price Index – Less food and energy ("Core CPI")
- Unemployment rate

Using each of the CPI measures, compute the ***annualized*** inflation rate. To compute this, you will calculate the growth in prices from the previous year. For example:

$$\text{Inflation (March 2001)} = [\text{CPI in March 2001} - \text{CPI in March 2000}]/\text{CPI in March 2000}$$
$$= 0.0292 = 2.92\%$$

This will give you two inflation series (1948:1-present)

(a) Create a time series plot of the two inflation series. Comment on any differences you notice and compute average and standard deviation for each inflation series. Explain why these differences may exist.

Go to the National Bureau of Economics web page and locate the business cycle dates.

(b) Using your graph from (a) and the NBER business cycle dates, identify how many recessions the U.S. has experienced since 1948:1. Indicate the recession dates by shading the dates on a time series plot of the inflation rate. Describe the behavior of inflation during recessions – are there any general conclusions you can draw based on your graph and knowledge of macroeconomics?

Combine the unemployment rate data, and the two inflation time series into one data file.

(c) The Phillips Curve plots the relationship between inflation (vertical axis) and unemployment (horizontal axis). This curve actually began as a simple statistical observation: the negative relationship between wage inflation and unemployment for the U.K. through the 1950s. Create a Phillips Curve using the U.S. data from 1948:1 – 1968:12, using the CPI-All items inflation rate. Does your Phillips Curve show a negative relationship? Compute the correlation between inflation and unemployment.

(d) Now, construct a Phillips Curve using data 1948:1-present. How does your answer differ to the one from the 1948-1968? Thinking back to intermediate macroeconomics, why might this be the case? Use an economic model in your justification. [HINT: think about shocks that affected the economy in the 1970s].

3. *Data for your project*

Create a table of summary statistics for the data you've collected for your project thus far. Your table should include the following for each:

- Average (Mean), Median, and Standard deviation
- Number of observations (the more the better, but you'll need 35 observations at the bare minimum)
- A description of either:
  - Sample period and frequency of data (for time series) OR
  - Point in time data were collected and unit of analysis (for cross section)
- Identify the following in your dataset:
  - which variable in your data set is the dependent variable (it is ok if you plan to have more than one),
  - choose one of your explanatory/independent variables and describe the correlation you expect to find, based on economic theory,
  - compute the correlation between the two variables you described above.