Public Policy 240B
Problem Set # 1

1. Many law schools use an 'admission index,' which is a weighted average of college GPA and the Law School Admission Test (LSAT) score, to screen applicants. Applicants whose admission index is above a certain threshold are considered for admission and those with scores below a threshold are rejected immediately. For example, the University of Oklahoma uses an admission index of I = 15*GPA + LSAT. Among the pool of applications for this law school in 2017, the average GPA was 3.2 with a standard deviation of 1.2 the average LSAT score was 150 with a standard deviation of 5, and the **correlation coefficient** between the LSAT and GPA is 0.50. What is the expected value and variance for the admission index at the Oklahoma Law School for their pool of applicants?

2. In this problem, we examine whether wearing a helmet reduces the chance of mortality in a motor cycle crash. The data below is taken from all motorcycle crashes nationwide from 1988-2005 that had a) two people on the bike, and b) the passenger on the bike was not wearing a helmet and died which means the accident was severe enough to generate a fatality. Below is a 2 x 2 table that lists the probabilities of whether the driver on the bike in the accident was wearing a helmet and died in the crash. For example, 28.3% of drivers were wearing a helmet and did not die in a crash, Pr(Wearing a helmet ∩ didn't die) = 0.283

		Did the driver die?		
		No	Yes	Total
Was driver wearing a helmet?	No	0.161	0.343	0.504
	Yes	0.283	0.213	0.496
	Total	0.444	0.556	1.000

Using the table above, answer the following questions:

a) What is the Pr(Died?)
b) What is the Pr(Was wearing a helmet)?
c) What is the Pr(Died | Was wearing a helmet)?
d) What is the Pr(Died | Was not wearing a helmet)?
e) By how much does the probability of dying in a serious crash go down if the driver is wearing a helmet?

To check your answer to part e) please see the abstract on page 5 of the following report http://www-nrd.nhtsa.dot.gov/Pubs/809715.PDF

3. The *Tornado Fuel Saver* is a device that can be installed in the air-intake of your vehicle's engine that is advertised to improve gas mileage and horsepower. (You may have seen an infomercial for the Tornado Fuel Saver on late-night TV! This product is real.). The manufacturer advertises that the Tornado Fuel Saver can boost you cars' miles per gallon by 21 percent. A car magazine recently conducted a test of the Tornado Fuel Saver and installed the device on 25 cars. The magazine found that miles per gallon improved by an average of 18 percent (x) in this sample. The standard deviation (s) of this estimate is an 8 percent improvement. From this sample of 25, construct a 95% confidence interval for the expected percent improvement in miles per gallon (μ) generated by installing the fuel saver. Can you reject or not reject the null that fuel saver boosts MPG by 21%?

4. A popular weight loss program these days is the 'Atkins Diet' that stresses a high protein/low carbohydrate diet. The Atkins diet is in stark contrast to many diets that stress low fats and /high carbohydrates. To examine which diet produces greater weight loss, a group of researchers randomly assigned 40 overweight patients to either an Atkins type diet (n_A=20) or a Low Fat/High Carb diet (n_L=20). Participants were given detailed guidelines about the diets and a phone number for a 'diet hot line' to call with any questions about the program they were assigned. The participants were weighed at the beginning of the diet and again in 6 months. After 6 months, dieters in the Atkins plan lost an average of 16 pounds whereas the participants in the Low Fat/High Carb diet lost only an average of 7 pounds. A summary of results from the experiment are listed below.

	Atkins Diet	Low Fat/High Carb Diet
n (sample sizes)	20	20
x (average change in weight, in pounds)	-16	-7
s (standard deviation)	12	8

a. Construct a **95% confidence interval** for d =u_A - u_L-- the difference in weight loss generated by the Atkins and Low Fat diets. What are the appropriate degrees of freedom and t-value you should use in this context?

b. Using the confidence interval from a), test the null hypothesis that H_o: d=0 -- there is no difference in the weight loss generated by the Atkins versus the Low Fat diets. Can you reject or not reject the null hypothesis?

c. Using a **t-test** and a **99%** confidence level, test the null hypothesis that H_o: d=0 -- there is no difference in the weight loss generated by the Atkins versus the Low Fat diets. What are the appropriate degrees of freedom of the t-statistic you use in this context? Can you reject or not reject the null hypothesis? Explain your answer.

Question 5: In class, we discussed several potential problems associated with inferring the direction and existence of causal relationships from analysis of non-experimental or observational data. Specifically, we discussed how omitted-variables bias, selection bias, and reverse causality (or simultaneous determination of the hypothesized dependent and explanatory variables) plague non-experimental data analysis. The following empirical patterns have been cited in press reports as potential evidence of causal relationships. All three empirical patterns cited are based on observational or non-experimental data. Can you apply any of the criticisms of non-experimental empirical result that we discussed in class to these three examples? If these criticisms were true, how do they alter the qualitative interpretations of these empirical patterns?

· A recent New York Times op ed argued that frequent marijuana use increases the likelihood of the onset of schizophrenia. The author was essentially arguing that the potential medicinal benefits of marijuana are overblown and the risks under-reported and generally under-appreciated. Presume that the the relationship between frequent marijuana use and schizophrenia is based on an observed correlation rather than the finding from a randomized control trial.

· A recent New York Times article reported that adolescents that take a pledge to postpone sex until marriage become sexually active about a year and half later than youths that do not take such a pledge. The implication of the study (and the interpretation of the authors cited in the paper) is that taking such a pledge in and of itself causes teenagers to postpone sex.

· In an op-ed piece arguing that the war on drugs has been a dismal failure, one of the New York Times regulars (who exactly, I can't recall) noted that in states where prison populations have increased the most, drug use has actually increased the most. To the extent that incarceration was intended to reduce drug use and abuse through deterrence and

the incapacitation of dealers, the mystery author offered this empirical pattern as evidence that the strategy of incarcerating drug offenders is not working.

Question 6: Students who graduate in the top 10 percent of their high school class from an accredited Texas school are guaranteed admission to the University of Texas at Austin. Class rank is typically determined by cumulative grade point average (GPA). Assume that the cumulative GPA for the class of 2018 at Tom Landry High School from Arlen, Texas is normally distributed with a mean of 2.9 and a standard deviation of 0.4. What GPA must a student from the Tom Landry class of 2018 receive in order to be admitted to the University of Texas?

Question 7: You are hired by the governor to study whether a tax on liquor has decreased average liquor consumption in your state. You are able to obtain, for a sample of individuals selected at random, the difference in liquor consumption (in ounces) for the years before and after the tax. For person *i* who is samples randomly from the population, Y_i denotes the change in liquor consumption. Treat these as a random sample from a Normal(μ, σ^2) distribution.

(i) The null hypothesis is that there is no change in average liquor consumption. State this formally

(ii) The alternative is that there is a decline in liquor consumption. State the alternative

(iii) Now suppose your sample size is n=900 and you obtain the estimates $\bar{y} = -32.8, \quad s = 466.4$. Calculate the t-statistic for testing the null against the alternative. Obtain the p-value of the test. Do you reject the null at the 5% level? How about the 1% level?

(iv) Would you say that the estimated fall in consumption is large in magnitude? Comment on the practical versus statistical significance of this estimate.

(v) What has been implicitly assumed in your analysis about other determinants of liquor consumption over the two year period in order to infer causality from the tax change to liquor consumption?

Question 8: This question asks you to use the data from the Stanford Education Data Archive (SEDA) that I have provided, which contains for every US district in the country the average standardized academic achievement test score between 2009-2016 for 3rd-8th graders (compiled by Sean Reardon and colleagues). I have also included in the data file a district-level measure of average socioeconomic status of the families of the children.

(i) How do you think the average student in your elementary school or district compared to the national average on math and reading tests? Were they 1 grade level above average? 1 grade level below average? Find out specifically by looking for your district ("leaname") and its achievement in comparison to the national average—note, the standardized outcome measure ("meangs") has already been converted to grade-level equivalents (e.g., 0 represents the national average; 1 represents 1 grade level above average; -1 represents 1 grade level below average).

(ii) What are the mean, median, 10th, 25th, 75th, and 90th percentile of academic achievement? [Hint: use the **sum** command with the **detail** option (options come at the end of a Stata command, following a comma). For help, search on "summarize" in the Stata help menus.] How many public school districts are there in the US? Did you take into account the district's student enrollment size ("mnenrol") in the calculation of the distribution of academic achievement above? [hint: if not, go back & do so]

(iii) What is the sample correlation coefficient between average academic achievement and district-level socioeconomic status? [Hint: use the **corr** command as follows: **corr meangs ses**. For additional help, search on "correlate" in the Stata help menus]

(iv) Create a scatter plot of the relationship between average academic achievement and district socioeconomic status. (Hint: be sure your figure visually distinguishes districts of different size in its display; use the **scatter** command)