Problem set 1

Due on Thursday September 15, 2022 (11:59 PM EST)

Note: No credit will be given if you report only the final answers without showing formulas and calculations when appropriate. This applies to both theoretical and empirical questions. For the empirical questions, make sure to submit your R codes and output on Latte. No credit will be given if the R codes and output are missing.

Problem 1

It is now early September, and you are not too far from taking your first Econometrics test (relax, there is still a lot of time!).

A friend bets you $10 that she will outscore you on the exam. Let X be a random variable denoting your winnings. X can take on the following values: 10, -10, 0 (0 if you tie). You know that the outcome depends on whether she studied for the exam. Let Y = 1 if she studied and 0 if she did not. The joint distribution is:

	Y = 0	Y = 1	Pr(X)
X = -10		.175	
X = 0			.05
X = 10	.4875		
Pr(Y)	.75		

a. Fill in the missing elements in the joint and marginal distributions, using the information that $Pr(X = 0 \mid Y = 0) = .05$.

b. Compute E(X), E(Y), Var(X) and Var(Y). Should you take the bet? Why or why not?

c. What is the expected value of your winnings if your friend did study for the exam? If she did not study?

Problem 2

Math and verbal SAT scores are each distributed normally with N (500,10000).

(a) What fraction of students scores above 750? Above 600? Between 420 and 530? Below 480? Above 530?

(b) If the math and verbal scores were independently distributed, which is not the case, then what would be the distribution of the overall SAT score? Find its mean and variance.

(c) Next, assume that the correlation coefficient between the math and verbal scores is 0.75. Find the mean and variance of the resulting distribution.

(d) Finally, assume that you had chosen 25 students at random who had taken the SAT exam. Derive the distribution for their average math SAT score. What is the probability that this average is above 530? Why is this so much smaller than your answer in (a)?

Problem 3

The development office and the registrar have provided you with anonymous matches of starting salaries and GPAs for 108 graduating economics majors. Your sample contains a variety of jobs, from church pastor to stockbroker.

(a) The average starting salary for the 108 students was $38,644.86 with a standard deviation of $7,541.40. Construct a 95% confidence interval for the starting salary of all economics majors at your university/college.

(b) A similar sample for psychology majors indicates a significantly lower starting salary. Given that these students had the same number of years of education, does this indicate discrimination in the job market against psychology majors?

(c) You wonder how much (in monetary terms) good grades are worth by calculating the average salary for economics majors who graduated with a cumulative GPA of B+ or better, and those who had a B or worse. The data is as shown in the accompanying table.

Cumulative GPA	Average Earnings $\overline{Y}$	Standard deviation s_Y	n
B+ or better	$39,915.25	$8,330.21	59
B or worse	$37,083.33	$6,174.86	49

Conduct a t-test for the hypothesis that the two starting salaries are the same in the population. Given that this data was collected in 1999, which was a boom year, do you think that your results will hold for other years, such as 2009 for example?

Problem 4

The new management of a bakery claims that workers are now more productive than they were under the old management, which is why the wages have ”generally increased”. Let W_i^B be Worker i's wage under the old management, and let W_i^a be worker i's wage after the change. The difference is $D_i = W_i^a - W_i^b$. Assume that the D_i are a random sample from a Normal(μ, σ^2) distribution.

a. Formally state the null hypothesis that there has been no change in average wages. In particular, what is $E(D_i)$ under H_0? If you are hired to examine the validity of the new management claim, what is the relevant alternative hypothesis in terms of $\mu = E(D_i)$?

b. Using the following data on 15 workers, test the null hypothesis from part (a) against the stated alternative at the 5% and 1% levels. Note that with n = 15, you are working with a rather small sample, and the central limit theorem does not apply yet

Worker	Wage before	Wage after
1	8.30	9.25
2	9.40	9.00
3	9.00	9.25
4	10.50	10.00
5	11.40	12.00
6	8.75	9.50
7	10.00	10.25
8	9.50	9.50
9	10.80	11.50
10	12.55	13.10
11	12.00	11.50
12	8.65	9.00
13	7.75	7.75
14	11.25	11.50
15	12.65	13.00

c. Obtain the p-value for the test described in part (b) [Hint: you can use Stata to compute this pvalue]

d. Construct an exact 95% confidence interval for μ.

Problem 5

What determines success at the Winter Olympics? Does population matter? Income? Or is it simply a matter of being in a cold place with lots of mountains? The table below describes variables in the **Olympics_HW.csv** dataset related to the winter Olympics games from 1980 to 2014.

Variable name	**Description**
ID	Unique number for each country in the data
Country	Name of country
Year	Year
Medals	Total number of combined medals won
Athletes	Number of athletes in Olympic delegation
GDP	Gross domestic product of country (per capital GDP in 10,000 US dollars)
Temp	Average high temperature (in Fahrenheit) in January if country is in Northern Hemisphere or July if Southern Hemisphere (for largest city)
Population	Population of country (in 100,000)
Host	Equals 1 for host nation and zero otherwise

a) Summarize the medals, athletes, and GDP variables;

b) List the first five observations for the country, year, medals, athletes, and GDP data
c) How many observations are there for each year;

d) Produce a scatterplot of medals and the number of athletes. Describe the relationship depicted;

e) Explain any suspicion you may have that other factors could explain the observed relationship between the number of athletes and medals;

f) Create a scatterplot of medals and GDP. Briefly describe any clear patterns;

g) Create a scatterplot of medals and population. Briefly describe any clear patterns;

h) Create a scatterplot of medals and temperature. Briefly describe any clear patterns;