STA 250

Instructions: You MUST answer question 1 and choose 3 other questions for a total of Four.

1. **Topics: Probability distribution, population moments, sampling distribution of sample mean, standard error of the sample mean, Confidence Interval (CI) for the mean, parts: a - j**

(REQUIRED) Consider the finite population of size N = 3, measurements on the variable *x*: {1, 2, 3}:

a. derive the population probability distribution for the variable *x*.

b. compute the population mean, $\mu_x = E(x) = \sum_{x \in \{1,2,3\}} x\,p(x)$, and standard deviation,

$$\sigma_x = (\mathrm{Var}(x))^{1/2} = \left(\sum_{x \in \{1.2,3\}} x^2\ p(x) - \mu_x^2 \right)^{\frac{1}{2}}.$$

c. Given the above population, sampling with replacement and considering order, how many samples of size n = 2 are possible?

d. derive all samples of n = 2 in the manner as described in *c*.

e. derive the sampling distribution of the mean as per your results obtained in *b*.

f. compute the expectation, E(m), or mean of the sampling distribution of the sample mean, m.

g. use the formula for the standard error of the mean, $\sigma_m = \frac{\sigma_x}{\sqrt{n}}$

h. Using your result obtained in *g*, derive for each sample in your population of samples (cf., result obtained in *d*) a 90% confidence interval.

i. How many of your sample confidence interval cover the true population mean?

j. Does the nominal and actual confidence levels coincide or are they different?

2. Topics: Ratio level data, Correlation, Causality, Control, and Confounding, two parts: a and b

For the below datasets compute the respective simple (zero-order) correlation between the independent variable ***x*** and the dependent variable ***y***. Then use the first-order (partial correlation) correlation to determine whether the relationship is between the variables ***x*** and ***y*** is causal or is due to confounding by the variable ***z*** which precedes both ***x*** and ***y*** in time. Draw the path diagram appropriate to the observed relationship for each dataset.

a. Data set 1

x	***y***	***z***
2	3	1
1	2	1
0	0	0
0	2	0
0	0	0
1	1	0
0	0	1

b. Data set 2

x	***y***	***z***
0	0	-0.03
1	0	0
2	1	0
0	0	-0.03
0	1	0
0	0	-0.03
0	0	-0.03

3. **Topics: Mean, Median, Empirical Influence Function, Influence curve, breakdown point, two parts: a and b**

Given the sample dataset of size n = 4: {1, -3, 1, ***x***}. Note that ***x*** is a real number.

a. Obtain the **empirical influence function** for the mean and sketch its **influence curve**. What does this influence function tell you about the mean (*hint*: try ***x*** = -4 or ***x*** = -100,000)? What is the **breakdown point**?

b. Now suppose that ***x*** < -3, compute the median. If ***x*** = -4, what is the median? If ***x*** = -100,000, what is the median? What do you observe regarding the stability of the median for this dataset?

4. **Topics: Ratio level data, Correlation, Statistical Significance, Effect size, Coefficient of determination**

A random sample of measurements on two quantitative variables *x* and *y* is taken from a population. The data for the variables ***x*** and ***y*** are: {(***x***, ***y***) | (2, 1), (3, 2), (4, 3)}. Assume that ***x*** is an independent variable and ***y*** is a dependent variable. Calculate the Pearson correlation coefficient. What does the correlation mean? Is the correlation statistically significant at the 5% level? What is the effect size? What is the coefficient of determination?

5. **Topics: statistical test, Confidence Interval (CI)**

The average life span of American men is 75.6 years. A random sample of n = 25 male college lecturers is found to have an average life span of 77.6 years, with a sample standard deviation 2.0 years. Would you conclude that college lecturers live longer than average? Use the appropriate two-tailed test with an $\alpha = 0.05$. Also, construct a 95% confidence interval. Do the statistical test and the confidence interval agree regarding whether college lecturers live longer? Write up your results.