Directions

Scan and upload your *handwritten* solutions to eLearning by the end of the day on Monday, April 10. STAT functions are permitted for any computations where possible. If a calculator function is not available, then provide a formula instead. For the multiple choice questions, simply provide the letter of your choice.

Problem 1 (1.5 points) Multiple Choice

In a May 1999 national poll on CNN, 43% of high school students reported fear of going back to school (History: Columbine was April 20, 1999). Which of the following best describes what is meant by the poll having a margin of error of 5%?

(A) It is likely that the true proportion of high school students afraid to go back to school is between 38% and 48%.
(B) 5% of students refused to participate in the poll.
(C) Between 38% and 48% of those surveyed expressed fear about going back to school.
(D) There is a 5% probability that the 43% result is in error.
(E) If similar polls were repeatedly taken, 5% of the confidence intervals would fail to capture the true proportion.

Problem 2 (1.5 points) Multiple Choice

A random sample of size 15 is taken from a population, and a 95% confidence interval for the population mean is calculated from the sample data to be $(64.06, 66, 96)$. Which choice below gives the best interpretation of this interval?

(A) 95% of the population measurements lie within the interval.
(B) If many random samples of size 15 are taken from this population, then 95% of the time the population mean will lie within the interval.
(C) If many random samples of size 15 are taken from this population, then 95% of the time the sample mean will lie within the interval.
(D) If many random samples of size 15 are taken from this population, then 95% of the confidence intervals will contain the population mean.
(E) If many random samples of size 15 are taken from this population, then 95% of the confidence intervals will contain the sample mean.

Problem 3 (3 points)

How much time do Americans spend eating or drinking? Suppose for a random sample of 101 Americans, the mean time eating or drinking per day is 73.2 minutes with a sample standard deviation of 39 minutes. Assume the time is a normal random variable.

(a) What is the value of t_c for calculating a 98% confidence interval for the mean time?

Since σ is unknown, we should use the t-distribution. (from the flowchart!)

$t_c = \text{InvT}(1 - 0.02/2, 101 - 1) = 2.364.$

(b) Construct and interpret a 98% confidence interval for the mean amount of time Americans spend eating or drinking per day.

The population standard deviation σ is unknown, so we use the t-distribution with $\alpha = 0.02$ and $t_c = 2.364$.

The 98% confidence interval for μ is

$$73.2 \pm 2.364 \times \frac{39}{\sqrt{101}} = [64.025, 82.375]$$

and we are 98% confident that the true mean time Americans spend eating or drinking per day is between 64.025 minutes and 82.375 minutes.

(c) Suppose you want to conduct your own survey on the same topic. How large of a sample is required to estimate the mean time Americans spend eating or drinking per day so that the estimate is within 15 minutes of the true mean with 92% confidence?

Use the z-distribution at $\alpha = 0.08$ since t can't be used since it depends on knowing n for the degrees for freedom. Then,

$z_c = \text{InvNorm}(1 - 0.08/2, 0, 1) = 1.751$

Then, the minimum sample size required is

$$n \geq \left(\frac{1.751 \times 39}{15}\right)^2$$
$$n \geq 20.7$$

$\boxed{\text{Therefore, we need at least 21 Americans.}}$

Problem 4 (4 points)

From last semester's class survey, consider their total number of computing languages which follows an approximate normal distribution. Treat this variable as continuous in this problem. I randomly selected 10 students from each section and recorded the number of computing languages they know in the table below. One student from the 11:30 am class did not give a response and so their recorded value is not available (NA). Thus, the sample size for the 11:30 am class is reduced from 10 to 9. Enter the data into two lists on your calculator and then do the following:

Class	Number of Languages
11:30 am	2 2 4 1 4 3 3 3 4 NA
1:00 pm	6 2 5 2 5 5 5 3 4 5

(a) Construct a 95% confidence interval for the difference in mean computing languages between the two classes. Assume that the population variances are equal. Is there evidence that one group knows more computing languages than the other? Explain. If so, which group?

Using `2-SampTInt` STAT function with `Pooled=Yes` we get the interval:

$$[-2.521, -0.101]$$

The interval is strictly negative implying that $\mu_1 - \mu_2 < 0 \implies \mu_1 < \mu_2$. There is evidence that one group knows more computing languages than the other. Since μ_2 is greater, than there is evidence that the 1:00 pm class knows more computing languages than the 11:30 am group.

(b) Redo (a) but assume the population variances are not equal.

Using `2-SampTInt` STAT function with `Pooled=No` we get the interval:

$$[-2.505, -0.1169]$$

Same conclusion as (a).

(c) Show how the degrees of freedom ν is computed by hand in (b).

Using the sample standard deviations and sample sizes in Satterthwaite's approximation we get

$$\nu = \frac{\left(\frac{1.054^2}{9} + \frac{1.398^2}{10}\right)^2}{\frac{1.054^4}{9^2(9-1)} + \frac{1.398^4}{10^2(10-1)}} \approx 17$$

DO NOT WRITE ON THIS FORM