

## Worksheet on CLT & inferences

1. According to a study by the U.S. Public Interest Research Group (USPIRG), a typical student spends an average of about \$898 per year on textbooks, with a standard deviation of \$268. A smart statistics student at Earlham College is interested in studying the implications of this for her class of 30 students.
  - (a) Describe (with numbers, sketch and a few words) the sampling distribution of mean textbook costs per year for samples of 30 students.
  - (b) Do you think the necessary conditions/ assumptions are met in this situation? Explain.
  - (c) Assume the conditions are met, and find the probability that the total textbook expenditure for that class will be more than \$30,000 this year. Show all steps in your solution.
  - (d) There is only a 5% chance that the total textbook costs for the class will be less than what value?
2. In order to gauge public opinion about a referendum on the ballot in the upcoming elections in Indiana, a news organization conducted telephone interviews with 602 voters registered in the state. They found 74% of those interviewed supported the referendum.
  - (a) Create a 90% confidence interval to estimate the true proportion of supporters. For full credit, your solution must show all key steps, including an appropriate conclusion at the end.
  - (b) Explain what 90% confidence means.
  - (c) Suppose we want to cut the margin of error to 2%. Find two different ways to do that, and discuss the pros and cons of each method.
3. A public-service organization carried out a survey of the tuition and fees for the 2017-18 school year at a sample of 44 private colleges in the midwestern U.S. The mean and standard deviation (in thousands of dollars) were found to be 32.4 and 7.2, respectively.
  - a. Carry out a hypothesis test to determine whether these data suggest the mean costs in the midwest differ significantly from the national average of \$34,800 published by the U.S. Department of Education.
  - b. Infer an appropriate conclusion to the same question by constructing and interpreting a matching confidence interval.
  - c. What sample size should we use if we want to estimate the true mean costs at midwestern colleges to within a \$1000 margin of error?
4. The distribution of annual income in a geographic region is skewed right, and has two modes. The overall mean and standard deviation (in thousands of dollars) are 45.8 and 36.2, respectively. What shape and summary statistics (i.e., mean and standard deviation) would we expect for the following distributions:
  - a. Incomes in a single random sample of size 1025 drawn from this population.
  - b. The sampling distribution of mean incomes in random samples of size 5.
  - c. Same as previous question, except random samples are of size 1025.

5. The normal resting pulse rate for adults is considered to be about 72.5 (beats per minute). A medical researcher measured the resting pulse rates of a sample of 20 adult patients at a clinic and found the mean to be 74.2, with standard deviation of 4.8.
  - a. Carry out a hypothesis test to determine whether the pulse rates for this group differ significantly from the normal.
  - b. Using an appropriate confidence interval, verify whether it supports the inference from your hypothesis test.

Think carefully about how to pick consistent yardsticks for comparing inferences from a hypothesis test with those from a confidence interval.
6. It is known from past meteorological data that the average daily snowfall for the month of December in Richmond, IN, is 0.089 inches, with a standard deviation of 0.025 inches. We are interested in statistically predicting the implications of these data for snowfall this December.
  - a. Describe (with numbers, sketch and a few words) the sampling distribution of mean daily snowfall for random samples of 31 December days. Clearly state any assumptions you make.
  - b. Find the probability that the total snowfall this December will be in excess of 3 inches.