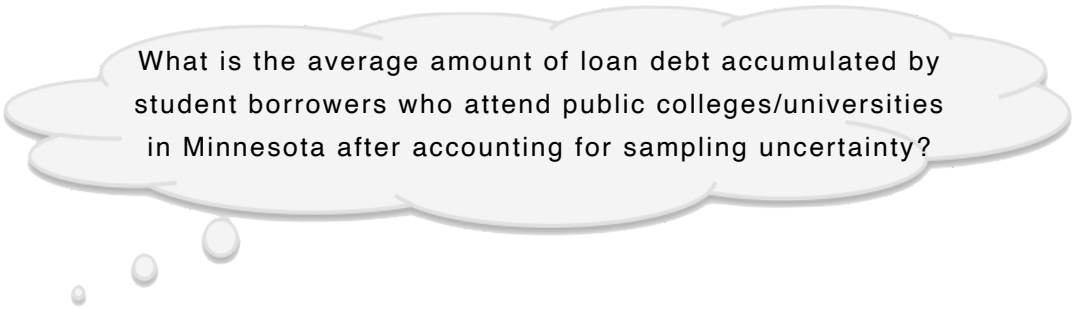
Minnesota College Debt

The Department of Education recently released its College Scorecard data. One of the variables in this data is the typical amount of loan debt accumulated at the institution by student borrowers. In this activity, you will be exploring the following research question:



To answer this question, you will use the data in the file *mn-colleges.tp3.* This data set contains a sample of 25 colleges/universities randomly selected from the larger population of all public colleges/universities in Minnesota. The variable *debt* provides the average student loan debt for students who attend that college/university.

Examine the Observed Data

1. Plot the sample debt data using TinkerPlots™. Describe the sample distribution. Be sure to describe the shape of the distribution and give a measure of center and variability.

Bootstrapping a Compatibility Interval

You can also carry out a bootstrap simulation to estimate the standard error when you have quantitative data.

2. Carry out 500 bootstrap trials. Plot the results from the 500 trials and sketch the plot below. Make sure to label the axis.

Evaluating the Bootstrap Distribution

3. Compute the standard error (use the stdDev() function) based on this simulation.

4. Using the standard error, compute the margin of error.

5. Compute the compatibility interval for the average amount of loan debt for students who attend public college/university in Minnesota.

Design and Inference

6. Based on the validity evidence for this study, is the compatibility interval an unbiased estimate for the average amount of loan debt for ALL students who attend college/university in Minnesota? Explain.

Effect of Sample Size

7. Consider a second sample of Minnesota colleges/universities that had the same sample mean debt as the data in the *mn-colleges.tp3* file. However, this second sample is twice as large; it includes 50 observations. How would the uncertainty in the compatibility interval from this second sample compare to the uncertainty in the compatibility interval you computed in Question #5? Explain.

8. The *mn-colleges-02.tp3* file contains data from 50 public colleges/universities in Minnesota. Use these data to compute a compatibility interval for the average amount of loan debt for students who attend public college/university in Minnesota.

9. Compare and contrast the uncertainty in the compatibility interval you computed in Question #5 with the uncertainty in the compatibility interval you computed in Question #8.

Compatibility Interval for the Median

The distribution of college debt is skewed to the right. This indicates that while the typical amount of debt at most schools in MN is less than $11k, there are a few schools where the typical amount of student debt is much higher. In skewed distributions, the **median** is often a better indication of a "typical" value than the mean.

10. Use TinkerPlots™ to find the median of the observed data in the *mn-colleges.tp3* data. To do this highlight the plot, and then click on the upside-down T button in the toolbar. (This is next to the mean button.) Record the value of the median.

11. How does the value of the median compare to the mean value in this distribution?

You can also carry out a bootstrap simulation to estimate the standard error for the median. To do this we collect the median (similar to how we collect the mean).

12. Carry out 500 bootstrap trials. Plot the results from the 500 trials and sketch the plot below. Make sure to label the axis.

13. Compute the standard error (use the stdDev() function) based on this simulation.

14. Using the standard error, compute the margin of error.

15. Compute the compatibility interval for the **median** amount of loan debt for students who attend public college/university in Minnesota. When you do this, be sure to use the value of the observed median rather than the observed mean in your formula.

16. Compare and contrast the uncertainty in the compatibility interval for the mean (Question #8) with the uncertainty in the compatibility interval for the median.Which has more uncertainty?