Backpack Weights

Carrying a heavy backpack can be a source of "chronic, low-level trauma," and can cause long-lasting shoulder, neck, and back pain. The American Academy of Orthopedic Surgeons recommends that a backpack not weigh more than 10 to 15% of the wearer's body weight. As the problem of overheavy backpacks has received more attention recently among primary and secondary school students, a student group decided to investigate the issue with students at their own university (approximately 20,000 enrolled students). Twenty-five students were sampled from each of four locations across the campus, during the course of four days, and four different times of the day (early morning, lunch time, evening, late afternoon). Scales were used to measure backpack weight to the nearest pound, and students were asked to report their body weight on a survey. One researcher was the designated talker and introduced the survey to the participants (survey questions included major, year in school, whether have back problems, and gender). Another person in the group was in charge of reading the scale, and another in charge of gathering the surveys.

STEP 1: Ask a research question.

1. Based on the study description, phrase a research question that you would be interested in testing with these students' data. Be sure to identify the population and parameter of interest. What symbol could you use for this parameter?

STEP 2: Design a study and collect data.

These data have already been collected, so now let us think about the students' data collection process.

2. Was this a random sample? Is it likely to be representative of the population identified in the previous question?

3. What other steps did the students take to minimize non-sampling concerns?

The students want to test whether students at their university were adhering to these backpack weight guidelines.

Specifically, they wondered if students at their university had a backpack weight to body weight ratio that was different from 0.10 (10%).

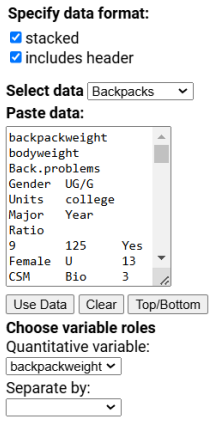
4. State the null and alternative hypotheses for investigating this question.

Null hypothesis:

Alternative hypothesis:

STEP 3: Explore the data.

After you collect data, the next step is to explore the data. In the case of a quantitative variable (like we have here), exploring the data is a bit more involved than exploring a categorical variable. We can summarize a quantitative variable using a dotplot and then describe the shape, center, variability, and unusual observations. Use the **Descriptive Statistics** applet to examine a dotplot of the sample data. To do this, click on the **Select data** drop down box and choose the *Backpacks* data, and press **Use Data**. Under **Choose variable roles**, the default quantitative variable to explore is the *backpackweights*. We want to change this to the *ratio* of backpack weight to body weight in the drop down menu.



These materials were developed for the EAPOST Scholar Program and supported by the National Science Foundation under Grant NSF DUE2235355. They are covered under the Creative Commons license BY-NC which allows users to distribute, adapt, and build upon the materials for noncommercial purposes only, and only so long as attribution is given to the EAPOST Program.

5. Describe the distribution of the percentages for this sample of 100 students. Remember to focus on shape, center, and variability, and any unusual observations or outliers.
6. Do these data provide any preliminary evidence that students at this California university *tend to* have a backpack weight to body weight ratio different from 0.10% (10%)?

7. Predict how the *median* percentage will compare to the mean ratio, keeping in mind that the median will divide the data set in half, 50 students below the median and 50 students above the median.

8. The largest ratio of backpack weight to body weight in this data set is 0.181 (18.1%). Suppose this had been incorrectly recorded as 1.81 (181%)! How will the mean change? How will the median change?

STEP 4. Draw inferences beyond the data. Is it plausible that this sample came from a population with a population mean of 0.10 (10%) or is the population mean different from 0.10 (10%)?
9. State appropriate null and alternative hypotheses, in words, about the population mean for this research question.

10. Outline a simulation strategy for this null hypothesis:
Population
Parameter
Sample
Statistic

In previous classes, we created hypothetical populations from which to sample. We wanted the population to have similar characteristics as the observed sample, but also assuming the null hypothesis to be true. We could generate a population of 20,000 ratios, with a population mean of 0.10? One method for generating a hypothetical population is to make many copies of the observed sample. As long as the population is more than 20 times the size of the sample, then the actual population size does not matter.
11. In the applet, choose the **x40** ratio button. This creates 40 copies of the observed sample and updates the "population distribution" in the first graph. How would you characterize the shape, center, and variability of this population compared to the actual sample?

12. In the **Sample size** box, use a value of 100 (the number of students actually measured), and press **Draw Samples**. If you scroll through the window showing the sample data, is it possible that you will see the same ID number more than once?

Now change the **Number of samples** to 999 (for 1,000 total). Choose **Mean** as the statistic and press **Draw Samples.** This process, sampling from a large number of copies of the sample distribution, is equivalent to sampling from the

DEFINITION
Bootstrapping randomly samples n observations from the original sample of n observations with replacement to estimate the sample-to-sample variation in a statistic. This is equivalent to treating the "population" from which you are sampling to be a very large number of copies of the sample.

dation
users to
s given to

CC BY NC

original sample *with replacement*, a technique known as ***bootstrapping***. The distributions you just generated are often called "bootstrap distributions."

Null Distribution? We cannot use the bootstrap distribution directly to estimate a p-value because it will be

> **KEY IDEA**
> Bootstrapping allows you to estimate the sample to sample variation in any statistic.

centered at the observed sample statistic rather than at the hypothesized value for the parameter. However, the bootstrap distribution does tell us about the sample- to-sample variation in the statistic. In fact, with quantitative data, we can simply shift the "population distribution" over to center at the hypothesized parameter value without changing
the shape or variability of the distribution.

13. In the applet, use the **Shift mean/median** slider to the right so that the "population distribution" centers at (approximately) 0.10. Now specify the observed sample mean ratio of backpack weight to body weight (0.078) in the **Count Samples** box. Since we have a 2-sided alternative, you will need to select *beyond* in front of the **Count Samples** box. Is it plausible that the observed sample of 100 ratios of backpack weight to body weight came from a population where the mean ratio was 0.10? Explain your reasoning. (Use the p-value from the applet to help.)

14. We have already seen that the variability in the sample mean can be calculated using the theory-based formula of $(\sigma/\sqrt{n})$. In reality, we don't have measurements on everyone in the population, so we can't know σ. We can, however estimate it with our sample SD, s. Let's now look at using the sample standard deviation (s) instead of σ in our theory-based formula to predict the variability in the sample means. Calculate *and interpret* $s/\sqrt{n}$, called the standard error of the mean.

15. Calculate a standardized statistic using the shifted null distribution. Is your sample mean ratio of the backpack weight to the body weight surprising if in the population the average ratio is 0.10 (10%)? Use your standardized statistic to help you explain.

> **KEY IDEA:** We can calculate a *standardized statistic* based on the observed sample data by computing
>
> *standardized statistic* =
>
> (where is a symbol used to represent the hypothesized value of the population mean).

STEP 5: Formula conclusions.

16. Write a paragraph summarizing your conclusions from this study. Be sure to discuss both what you learned from the sample and what you believe to be plausible about the population.

These materials were developed for the EAPOST Scholar Program and supported by the National Science Foundation under Grant NSF DUE2235355. They are covered under the Creative Commons license BY-NC which allows users to distribute, adapt, and build upon the materials for noncommercial purposes only, and only so long as attribution is given to the EAPOST Program.

STEP 6: Look back and ahead.

17. Return to your critique of the study design. If you were to carry out such a study on your own campus, discuss what you would do differently and why.

These materials were developed for the EAPOST Scholar Program and supported by the National Science Foundation under Grant NSF DUE2235355. They are covered under the Creative Commons license BY-NC which allows users to distribute, adapt, and build upon the materials for noncommercial purposes only, and only so long as attribution is given to the EAPOST Program.