Using Standard Error Formulas

Stat 145

The SE formulas

Here are our four SE formulas arranged in a table based on the parameter of interest and the number of groups. When there are two groups, we are interested in the difference in proportions or difference in means.

parameter type	one group	two groups
proportion	$SE = \sqrt{\frac{p(1-p)}{n}}$	$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
mean	$SE = \frac{\sigma}{\sqrt{n}}$	$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

A note on paired designs

When we have a paired design, the first step is to convert two variables into a single variable (usually by substraction, but sometimes the ratio is used instead). After we do that conversion, we are left with a single quantitative variable and we are interested in the mean of that variable. So the paired situation is just a special case (with an extra step) of the situation for one mean.

Design a study

In each scenario below,

- determine what variables you would need to collect and whether each is categorical or quantitative,
- determine the paramater of interest, and state the null and alternative hypotheses if you are going to do a hypothesis test,
- determine whether you are using a paired design.

Sometimes there may be more than one way to design the study, but don't design a poor study when a better option is available.

1. You want to know what proportion of Calvin students got a flu shot this year.
2. You want to know whether male students or female students were more likely to get a flu shot this year.
3. You want to know whether people who got a flu shot were more likely or less likely to get the flu.
4. You want to know whether a new drug works better than an old drug at reducing cholesterol.
5. You want to know whether rhubarb grows faster or slower if you cover it with a bucket for 3 weeks.
6. You want to know whether people can swim faster if they wear wetsuits.

Practice with the formulas

In each of the situations below, compute a p-value or confidence interval (or both).

7. In a study to compare the endurance of male and female mice, mice were made to swim in a bucket with a with attached to their tail and rescued when they became exhausted. The table below gives the some information about the distribution of these "times to exhaustion" (in minutes).

sex	n	mean	sd
female	162	11.4	26.09
male	135	6.7	6.69

   a. Both distributions (for females and males) were skewed. Which direction do you think they were skewed? Why?
   b. Why is OK to use our SE method for this situation even though the sample distributions are skewed?
   c. Give an 98% confidence interval for the mean endurance for female mice.
   d. Give an 98% confidence interval for the mean endurance for male mice.
   e. Give an 98% confidence interval for the difference in the mean endurance for female and male mice.
   f. Is there evidence that endurance varies by sex? How strong is the evidence?

8. Use the data in `StudentSurvey` (from `Lock5withR`) to answer the following questions.
   (This is a sample of students from one particular university, so we can only generalize results to students at that university, and then only if the sample was reasonably representative.)

   a. Do men exercise more than women? How much more?
   b. What is the averate MathSAT for students at this university?
   c. What is the average number of piercings for women at this university?
   d. What proprotion of students have a higher Math SAT than verbal SAT? Does this proportion differ for men and women?