

Comparing two groups

StatPREP Class Lesson

Orientation

A friend and I were discussing ways to measure poverty. Of course, there's the so-called "poverty level" of income defined by the US government. But we were interested in whether it might be worthwhile to do a detailed interview with, say, 100 families, quantifying how much trouble they have paying for different things: food, housing, healthcare, transportation, emergencies etc. Then we'd look at easy things to measure so that we could have an index for these different things, food stress, housing stress, and so on.

To get started, we decided to look at some already available data to generate ideas about what things might be related to poverty. This could help us design our interview.

We know that you're taking a statistics course, so we arranged to visit with you so that you could help us handle the data. You told us that you already have some data about income, health, socio-economic factors and such. It's called the NHANES data.

We arranged to meet you in a local coffee shop, which has free internet.

Activity

When we got to the coffee shop, you opened up a statistics app, called the [t-test Little App](#). You set the app to work with the NHANES2 data. You explained that there is a lot of data, but maybe it would be good to start with a sample of 100, so that we would get an idea of what things might work in our planned interviews of 100 families.

1. You set the response variable in the app to a variable called "income_poverty." We could see at a glance that this variable ranged from 0 to 5. That didn't make sense to us.
 - Explain what the `income_poverty` variable measures and how it's related to income.
2. We decided to start with home ownership, so you set the explanatory variable to "home_type".
 - Describe the pattern shown by the data in the graph. Do you think that you can say much about the poverty variable based on home ownership?
3. The dots were all over the place. You told us that it can be helpful to look at the mean poverty level separately for people who own and who rent their home. You turned on a display of the mean. Exciting! The mean was different for the two groups.

- “Slow down,” you said to your friends. Explain to them why any meaningful comparison of the means has to take into account the random nature of sampling.
 - Use the app’s measuring stick to quantify the difference “Own” mean versus “Rent” mean. (Or, you could read it off the scale for the vertical axis.) Write down the distance you measured.
4. Next you told us to watch the two means while you pressed over and over the New Sample button. You told us to watch whether there was a consistent pattern, from sample to sample, of one of the means being higher than the others.
- Each time you press New Sample, measure the difference between the two means. Take care to record which mean is higher than which.
 - Is there a consistent pattern or is the mean for “Own” sometimes above and sometimes below the mean for “Rent?”
5. To show what pattern appears in the difference between the two means *when there is no meaningful difference*, check the “Shuffle groups” box. This will show the same data, but the explanatory variable will be randomly shuffled. This eliminates any possible systematic relationship between the response and explanatory variables.
- Press the New Sample button many times and observe which mean is higher than the other. Give a simple description of the pattern you see.