Comparing two groups

Carol Howald

6/12/2020

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 Confidence and T Little App. (See footnote[1]) You set the app to work with the `NHANES` data, by setting Source package to be Little Apps, and Data set to NHANES2. 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. To set the sample size to 100, click the icon that says n= and then pick 50.

1. You set the response variable in the app to a variable called "income_poverty." Set the explanatory variable to Select explanatory variable. We could see at a glance that this variable ranged from 0 to 5. That didn't make sense to us.

    Documentation on the meaning of the variables is found in an icon called '_____'. Fill in the blank.

    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 an individual person's level for the poverty variable based on their 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 by

---

[1] https://maa-statprep.shinyapps.io/Little__App__T/

clicking on Apps control (three parallel lines) and then clicking on the check box next to show 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, by clicking on Graph in the top tool bar and then clicking on the higher mean and dragging the curser to the lower mean. (Or, you could read it off the scale for the vertical axis.)

What was the distance you measured? . . . .

4. Next you told us to watch the two means while you pressed over and over the New Sample button (the pair of dice icon). 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, look at the difference between the two means. Record "Own" when the mean for "Own" is higher than the mean for "Rent," and vice versa.

   Write down 20 new-sample trials (each with an "Own" or a "Rent") in this space. . . .

   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" 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 20 times and observe on each trial which mean is higher than the other.

   Give a simple description of the pattern you see. . . .

Version 0.3, 2020-08-12