Tim Vigers

Consulting 1

Homework 9

11/8/2018

Exercise 2

I think the best way to approach this is to go through each of the variables in order and come up with a few different ways of looking at each one. There are only two study visits, so most of the questions will be about change between baseline and one year. However, it may also be worth looking at the baseline cohort overall first, to see if they are a particularly sick (or healthy) group. For instance, are there more patients with x-rays positive for other lung diseases than a clinician might expect? Or is the presence of other diseases pretty consistent with expectations? You could do the same kind of thing with total lung capacity and FVC as well. If the group starts out sicker or healthier than average, it will influence the interpretation of the findings. The next variable to look at is probably the proportion of patients who died during the year, which is likely related to the health of the cohort at the start. None of these are really statistical questions in the sense that you’re looking for a difference or a trend, but they’re worth thinking about as research questions anyway, and the info can go in table 1 to inform readers.

Once you’ve looked at these questions, I think the first thing to ask is whether there is a difference between baseline and one year. Since it’s the same people being measured longitudinally, you’ll have to use paired tests for everything. As far as we know, there isn’t any sort of intervention in this study (i.e. patients aren’t receiving any sort of medication), so we’re really just looking at the progression of disease. Still, we can probably look at some interaction effects between lung disease, total lung capacity, and FVC. I just noticed that this answer is getting really long, so here are some questions in list form:

1. For each patient, count the total number of diseases diagnosed through x-ray at baseline and 1 year. Was there more disease at 1 year than at baseline? You could define this as “did more people have at least one disease?” or “was the overall number of diseases larger?”
2. Group the lung diseases into categories, like mild, moderate, severe, etc. and test if severe diseases were more prevalent at 1 year compared to baseline.
3. On average, does the ratio of black to white pixels change from baseline?
4. Is the ratio black to white pixels a good measure of lung disease in this cohort? For example, is the ratio higher among those with more severe disease? Does it increase as you would expect for those participants whose disease got worse over the year?
5. Similarly, does the ratio correlate with lung capacity measures? If a participant’s lung capacity goes down over the year, does that predict the change in x-ray ratio? If so, how well?
6. Did average lung function decrease from baseline to 1 year?
   1. One issue with this is how do you deal with the participants who died? They aren’t included in the 1 year data, but you’d assume that they probably had lower lung function before they died. Probably best to exclude them completely.
7. If lung function does indeed decrease on average, which other variable best predicts the decline? Similar to some of the above questions, but you could look at e.g. whether pixel ratio predicts decline better than other disease status.
   1. You can also dichotomize these variables if there are clinical cutoffs (e.g. if FVC > 80% is considered “good”).

This certainly isn’t an exhaustive list of questions, but I think it’s a good place to start with the investigator. They can then help weed out the questions that might have been answered in previous papers, or just aren’t clinically interesting, and they can assist with dichotomizing certain variables if it’s more clinically meaningful.

Exercise 3

Exercise 4

1. A mediator is a variable that explains the relationship between other variables. For example, if they found that African American people have worse outcomes than whites, level of care would be a mediator if it explained the overall negative relationship. In other words, race impacts the level of care, which in turn results in worse outcomes.
2. An interaction means that a second variable modifies the relationship of other variables. For example, if there is a relationship between race and outcomes, but the relationship is different among men and women (e.g. African American women have significantly worse outcomes than African American men).