20180905 - Introductory Lab

Adapted from Mine Cetinkaya-Rundel and Andrew Bray

September 05, 2018

Objectives

This lab will give you a first look at some important ideas that we'll come back to throughout this class:

1. We can use data from a sample to learn about a larger population.
2. We can quantify the *strength of evidence* an experiment provides

Along the way, we'll also get a first look at `R`, the statistical programming language we'll use in this class.

Evaluation

You will not be turning this lab in for a grade. This is just a first look at many of the ideas we'll talk about over the coming weeks and months - we will revisit everything here again in future classes.

Introduction

As a part of a study on gender discrimination, 48 male bank supervisors were each given the same personnel file and asked to judge whether the person should be promoted to a branch manager job that was described as "routine" (B.Rosen and T. Jerdee (1974), "Influence of sex role stereotypes on personnel decisions", J.Applied Psychology, 59:9-14.). The files were identical except that half of the supervisors had files showing the person was male while the other half had files showing the person was female. It was randomly determined which supervisors got "male" applications and which got "female" applications.

Overall, 35 out of 48 of the supervisors decided to promote the candidate. The results of the study are summarized below:

Gender	Promoted	Not Promoted	Total
Male	21	3	24
Female	14	10	24
Total	35	13	48

Our goal is to **quantify** the strength of evidence provided by these data for distinguishing between the following two hypotheses:

1. Null Hypothesis: "There is nothing going on."

- Promotion and gender are independent
- There is no systematic gender discrimination in promotion decisions
- Any observed differences are simply due to chance

2. Alternative Hypothesis: "There is something going on."

- Promotion and gender are dependent
- There is systematic gender discrimination in promotion decisions
- Any observed differences are not due to chance

Section 1: Thinking About the Data

a) Identify the *observational units* and the *variables* in this study.

The **observational unit** is the person or thing whose characteristics we measure in a study. Depending on the context, an observational unit may also be referred to as a *case*, *respondent*, *subject*, or *participant*.

The **variables** are the characteristics recorded about each observational unit.

b) We are often interested in the relationships between multiple variables; in many cases, we can identify one or more variables as *explanatory* variables, that explain something about differences in the *response* variable for different observational units. In this study, what is the explanatory variable? What is the response variable?

Explanatory:

Response:

c) Based on this sample, what is the conditional probability of promotion given that the applicant was "male"? A probability is a number between 0 and 1: in this case, the proportion of "males" who were promoted.

d) Based on this sample, what is the conditional probability of promotion given that the applicant was "female"?

e) One way of comparing these probabilities is by calculating their difference: (probability of promotion in the male group) - (probability of promotion in the female group). Based on this sample, what is the difference in probabilities of promotion for the male group and the female group?

This difference is an example of a **statistic**. A statistic is a numeric summary of the data in a sample. Here, we summarized all of the information we recorded about our sample in a single number that is relevant to the question we want to answer.

Your answers to parts c, d, and e summarized the results for this particular *sample* of 48 promotion decisions. Our goal is to use the data from this sample to try to say something about the relationship between gender and the probability of promotion in the *population* of all promotion decisions.

f) You saw that in this sample, there was a difference in the probability of promotion for the two groups. Is it possible that this difference could have happened in the sample even if there was no association between gender and probability of promotion more generally?

Section 2: A Simulation Study (group activity, groups of 2 or 3)

The key question here is whether results like those obtained in this study (a difference in probability of promotion of 0.292 between the two groups) could reasonably occur if there was no association between gender and probability of promotion in the population of all promotion decisions. We will call this assumption of no association the null model or the null hypothesis.

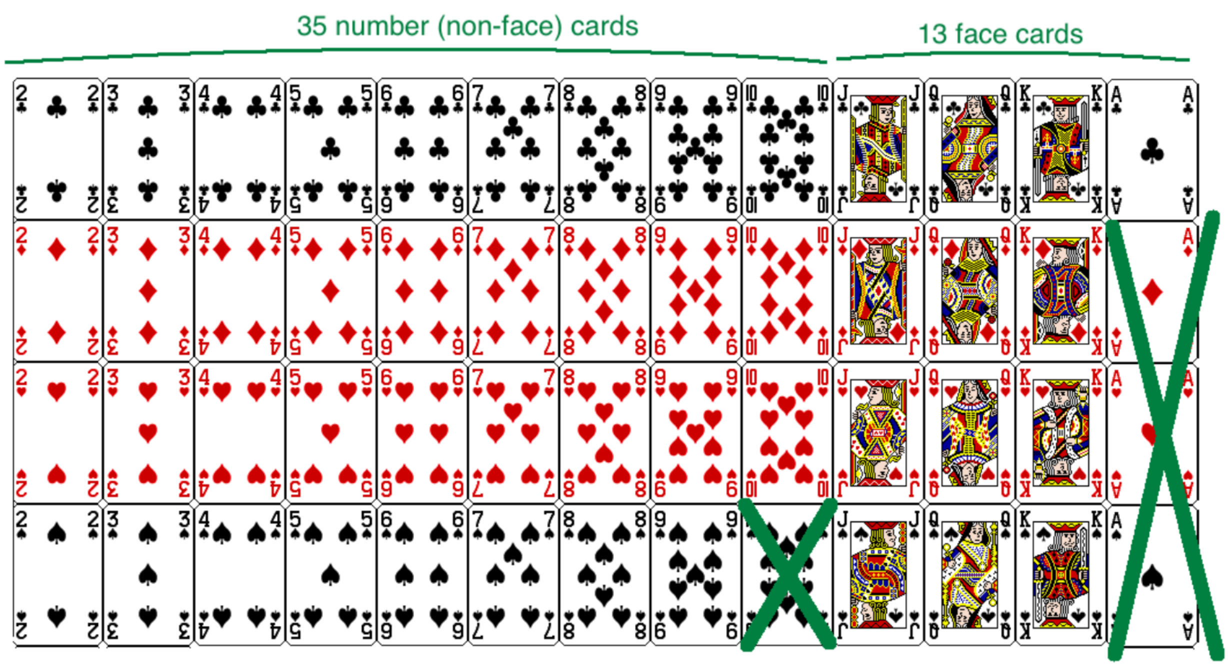
We will answer this question by simulating (artificially re-creating) the assignment of promotion decisions to the male and female groups over and over again, assuming that the cases who were promoted were equally likely to be in either of those groups. This will let us see how the results we actually observed in our sample compare with the kinds of results we might have observed if there was actually no association between gender and promotion decisions.

Throughout this procedure, let's assume that in every replication of the experiment, the researchers assign 24 cases to the "male" group and 24 cases to the "female" group. Let's also assume that overall, the same number of cases are promoted (35). The question, then, is which group those 35 promotions end up in, just by chance.

One Simulation

Do the steps below to sample from this table, assuming that there is no association between gender and probability of promotion.

1. We'll let a face card (J, Q, K, or A) represent "not promoted" and a non-face card (a number 2 through 10) represent "promoted". To match the totals in the data, we want to create a deck with 13 face cards and 35 non-face cards. To do this:
   - Set aside the jokers
   - Take out 3 aces
   - Take out one number card



(continued on next page)

2. Shuffle the cards well
3. Deal out 24 cards, representing “female” candidates; the remaining cards will represent “male” candidates. Fill in the table below based on your simulated results; for example, the top left number will be the number of non-face cards in the stack for the “male” group.

Gender	Promoted	Not Promoted	Total
Male			24
Female			24
Total	35	13	48

4. Based on your simulated data in the table above, find the difference in the probability of promotion for the “male group and the”female" group.

Your result in step 4 is one possible value for a difference in risk of death for the two groups that you might have observed in this study, if there was actually no association between gender and the probability of promotion in the population of all promotion decisions.

Additional Simulations: Repeat steps 2, 3, and 4 above 4 more times, taking turns:

Simulation 2

Gender	Promoted	Not Promoted	Total
Male			24
Female			24
Total	35	13	48

Difference in the probability of promotion for the “male” and “female” groups:

Simulation 3

Gender	Promoted	Not Promoted	Total
Male			24
Female			24
Total	35	13	48

Difference in the probability of promotion for the “male” and “female” groups:

Simulation 4

Gender	Promoted	Not Promoted	Total
Male			24
Female			24
Total	35	13	48

Difference in the probability of promotion for the “male” and “female” groups:

Simulation 5

Gender	Promoted	Not Promoted	Total
Male			24
Female			24
Total	35	13	48

Difference in the probability of promotion for the “male” and “female” groups:

When your group is done, add your results to the list on the board.

Examining the Results

h) On average, what would we expect the difference in probability of promotion to be between the “male” and “female” groups, if there is no effect of gender on promotion?

i) What did we observe in our combined simulation results?

j) Looking at the plot, does it seem that the result obtained by the researchers would have been “unusual” if in fact there was no association between gender and probability of promotion? What does this suggest about whether the researchers’ result provides much evidence that there is actually an association between gender and promotion? Explain.

Section 3: Automating the Simulation using R

We really need to simulate this random selection process hundreds, preferably thousands of times. This would be very tedious and time-consuming with cards, so we'll turn to technology. At this point, you'll need to register for our course on gryd.us. This is a website where you can run R code, and the next few steps of the lab are there. (Don't worry, all the R code is provided for you.) There's a registration link on our course website at http://www.evanlray.com/stat140_f2018/schedule.html.

After you've gone through some of the code in that lab, you'll be directed back here to answer the questions below.

k) Based on the plot of your simulation results, would you say that it would be very surprising, if there is actually no association between banding and risk of death, that we would have observed a difference of 0.292 by chance? Explain.

l) Report how many of your 1000 simulated random assignments produced a difference in risk of death between the two groups at least as extreme as what we saw in the actual study (which, you'll recall, was 0.292). Also determine the proportion of these 1000 repetitions that produced such an extreme result.

This proportion is called an approximate p-value. A p-value is the probability of obtaining a result as extreme as the one observed, assuming that there is no actual relationship between the variables. If the p-value is small, that means that it would be unlikely to get data like what we actually got in our experiment if there was actually no association between the variables. Therefore, a small p-value casts doubt on the null model/hypothesis used to perform the calculation (in this case, that there is no association between gender and promotion). A large p-value indicates that the data are consistent with the null model: we could have observed data like what we got in our sample just by chance, even if there was no association.

- A p-value of .10 or less is generally considered to be some evidence against the null model/hypothesis.
- A p-value of .05 or less is generally considered to be fairly strong evidence against the null model/hypothesis.
- A p-value of .01 or less is generally considered to be very strong evidence against the null model/hypothesis.
- A p-value of .001 or less is generally considered to be extremely strong evidence against the null model/hypothesis.

m) Is the proportion you got in part l) small enough to consider the actual result obtained by the researchers surprising, assuming the null model that there is no association between gender and promotion?

Section 4: Summary

We just covered a lot! We will revisit all of these ideas over the rest of the semester, but here is a summary of the vocabulary and big ideas introduced in this lab:

- An **observational unit** is the person or thing whose characteristics we measure in a study.
- The **variables** are the characteristics recorded about each observational unit.
- An **explanatory** variable explains something about differences in the **response** variable for different observational units.
- A **sample** is the particular set of observational units we included in our study.
- A **population** is the larger set of people or things we are really interested in. Our goal is to use the data in the sample to say something about the whole population.
- A **statistic** is a numeric summary of the data in a sample.
- We saw the structure of a **hypothesis test**, which is a way of answering the question of whether data in a sample provide strong evidence that something interesting is going on in a population.
  - We temporarily assume a **null model** or **null hypothesis** that expresses the idea that in the population, nothing interesting is going on. In this case, the null hypothesis was that there was no association between gender and promotion.
  - Assuming that the null model is correct, we find a distribution of values we might see for a statistic we are interested in.
  - We can then calculate a **p-value**: if the null model were true, how often would we get samples that had a statistic at least as extreme as what we actually observed in our data set?
  - Based on the p-value, we can make a decision about our hypothesis:
    * If the p-value is small, it would be unlikely to observe data as extreme as what we got in our actual data set if the null hypothesis was true. So our data are not consistent with the null hypothesis. That means that our data offer evidence that the null hypothesis is not true.
    * If the p-value is large, it is reasonable that we could have observed data as extreme as what we got in our actual data set if the null hypothesis was true. That means that our data do not offer enough evidence to rule out the null hypothesis.