LSN1-Grad School Admissions

MA376 - LTC Starling

Exploring the Berkeley Grad School Dataset

Explore the Berkeley Data set

Let's take a look at the data:

Observational units:

Variables

- **Response variable:**

- **Explanatory variables:**

Let's change the variables to be categorical:

```
grad2 <- grad %>% mutate(Program = as.factor(Program)) %>%
         mutate(Sex = as.factor(Sex)) %>%
         mutate(Accepted = as.factor(Accepted))
```

Create a bar plot of accepted vs rejected.

```
acc2 <- grad2 %>% pull(Accepted) %>%
  fct_count() %>%
  mutate(perc = n / nrow(grad2))
```

What do we see here?

Create a Contingency Table

We can ask the question, "Is the admissions process at Berkeley fair?". Or a more focused question, "Does the admissions process at Berkeley disproportionately affect women applicants?"

What do we see here?

Create Mosaic plot

Create a mosaic plot to see the data related to sex and acceptance:

What does this plot tell us? Are sex and acceptance *associated*?

Does the program a person applied to make a difference? To explore this, we will condition on the program applied to:

What does this plot tell us? Is program associated with acceptance? How is this different than the previous plot?

Here we look at the association between acceptance and program, conditioned on sex:

What do we notice here compared to our first mosaic plot above?

What can we say about our original question, "Does the admissions process at Berkeley disproportionately affect women applicants?"? What do we call the variable *program*?

Sources of Variation Diagram

On p. 8, the author provides the Sources of Variation diagram shown here.

Observed Variation in: Acceptance (Yes or No)	**Sources of explained variation**	**Sources of unexplained variation**
Inclusion criteria • Class (graduate students) • School (Berkeley)	• Sex (male or female) • Program	• Quality of application • Numerical data (e.g., test scores, grade point average) • Unknown ...

Are there other sources to consider?

Six Steps of a statistical investigation:

- Ask a Research Question
- Design a study and collect data
- Explore the data
- Draw inferences beyond the data
- Formulate conclusions
- Look back and ahead

Did we do this with the Berkeley Data? What would you tell the college administration?

Can we create a statistical model for the Berkeley data?