

Week 1: Welcome to statistics and data

3. Statistics to illuminate

Stat 140 - 04

Mount Holyoke College

Dr. Shan Shan

Slides posted at <http://sshanshans.github.io/stat140>

Rename yourself on Zoom with the following format:

[Group number]_[Your preferred name] (Your pronoun)

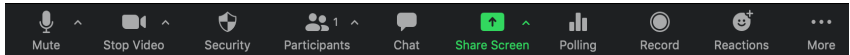
E.g.,

- ▶ 1_Shan Shan (she/her) (if I'm in group 1)
- ▶ 2_Shan Shan (she/her) (if I'm in group 2)

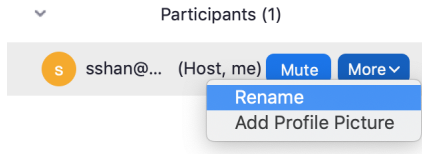
Find your group information here:

<https://docs.google.com/spreadsheets/d/17Q8S2iAjZJ81PBterieARC-InLgz3MrY-nHEyAE2vEw/edit?usp=sharing>

Step 1: Find 'Participants' tab



Step 2: Find yourself → More → Rename



1. Summary to the admission example

A table of counts, based on data from UC Berkely's graduate admission process in 1973.

	Men	Women
Accepted	533	113
Denied	663	336
Total	1198	449

A table of counts, based on data from UC Berkely's graduate admission process in 1973.

	Men	Women
Accepted	533	113
Denied	663	336
Total	1198	449

Proportion of male applicants who were accepted out of the total male applicants is $533/(533 + 663) \approx .445 \approx 45\%$.

Proportion of female applicants who were accepted out of the total female applicants is $113/(113 + 336) \approx .25 \approx 25\%$.

A table of counts, based on data from UC Berkely's graduate admission process in 1973.

	Men	Women
Accepted	533	113
Denied	663	336
Total	1198	449

Proportion of male applicants who were accepted out of the total male applicants is $533/(533 + 663) \approx .445 \approx 45\%$.

Proportion of female applicants who were accepted out of the total female applicants is $113/(113 + 336) \approx .25 \approx 25\%$.

We observed that the acceptance rate for men is almost 20 percentage points higher than the acceptance rate for women (44.5% vs. 25.2%).

Let's proceed to dig a little deeper. The data actually came from two different programs, let's call them A and F. This table show the counts for these two programs separately.

	M(accpt)	M(denied)	W(accpt)	W(denied)
ProgA	511	314	89	19
ProgF	22	351	24	317
Total	533	665	113	336

Let's proceed to dig a little deeper. The data actually came from two different programs, let's call them A and F. This table show the counts for these two programs separately.

	M(accpt)	M(denied)	W(accpt)	W(denied)
ProgA	511	314	89	19
ProgF	22	351	24	317
Total	533	665	113	336

Within A, the proportion of male applicants who were accepted out of the total male applicants who applied for A is $511/(511 + 314) = 511/825 \approx .619 \approx 62\%$

Let's proceed to dig a little deeper. The data actually came from two different programs, let's call them A and F. This table show the counts for these two programs separately.

	M(accpt)	M(denied)	W(accpt)	W(denied)
ProgA	511	314	89	19
ProgF	22	351	24	317
Total	533	665	113	336

In summary,

Program A, men: $511/(511+314) = 511/825 \approx .619 \approx 62\%$

Program A, women: $89/(89+19) = 89/108 \approx .824 \approx 82\%$

Program F, men: $22/(22+351) = 22/373 \approx .059 \approx 6\%$

Program F, women: $24/(24+317) = 24/341 \approx .070 \approx 7\%$

This is very strange!

Hint: think about what's the difference between program A and program F?

Hint: think about what's the difference between program A and program F?

Let' look at the data table more carefully.

	M(accpt)	M(denied)	W(accpt)	W(denied)
ProgA	511	314	89	19
ProgF	22	351	24	317
Total	533	665	113	336

What's the acceptance rate for program A:

$$(511 + 89)/(511 + 314 + 89 + 19) \approx 64\%$$

What's the acceptance rate for program F:

$$(22 + 24)/(22 + 351 + 24 + 317) \approx 5\%$$

Hint: think about what's the difference between program A and program F?

Let' look at the data table more carefully.

	M(accpt)	M(denied)	W(accpt)	W(denied)
ProgA	511	314	89	19
ProgF	22	351	24	317
Total	533	665	113	336

How many men applied for program A? $511 + 314 = 825$

How many women applied for program A? $89 + 19 = 108$

How many men applied for program F? $22 + 351 = 373$

How man women applied for program F? $24 + 317 = 341$

The odds is because more women applied for the program that is difficult to get into.

This explains how it happens that women have a higher acceptance rate than men in both programs but a lower acceptance rate than men when the programs are combined.

A phenomenon where a trend appears in several different groups of data, but disappears or reverses when they are combined.

HW01: find another example of the Simpson's paradox.

Tutorial exercise: For the rest of the class

Work on mini-project 1

Goal: practice using R markdown to produce report

Let me know if you have any questions

You are allowed to leave once you are done.

Tips:

- ▶ Each one of you work on your own markdown file first. Then copy your write up to a shared google doc with your group and collaboratively edit the google doc together.
- ▶ Figure out a time to meet outside of class first.
- ▶ Read the grading guidelines carefully, and let me know if I can help clarify.