DSFA
Spring 2020

Lecture 14

Simulation

Announcements

- HW 4 out, due Friday 3/13, 5:59PM

Distribution

- A distribution is a description of the likelihood of *events or outcomes*

- Probability distribution:
  - Theoretical: made from mathematics
  - Probability of each event

- Empirical distribution:
  - Experimental: made from observations
  - Proportion of each event in sample

Law of Large Numbers

If an experiment is repeated many times,
independently and under the same conditions,
then the proportion of times that an event occurs
gets closer to the theoretical probability of the event

Sometimes called *Law of Averages*

Terminology

Parameter

A number associated with the population

Statistic

A number associated with the sample

A statistic can be used as an **estimate** of a parameter

Simulation Structure

(Demo)

Mendel's Pea Plants

Mendel's Pea Plants

Gregor Mendel (1822-1884)

- Founder of the modern field of genetics
- Theorized that pea plants will bear purple or white flowers at random, in the ratio 3:1
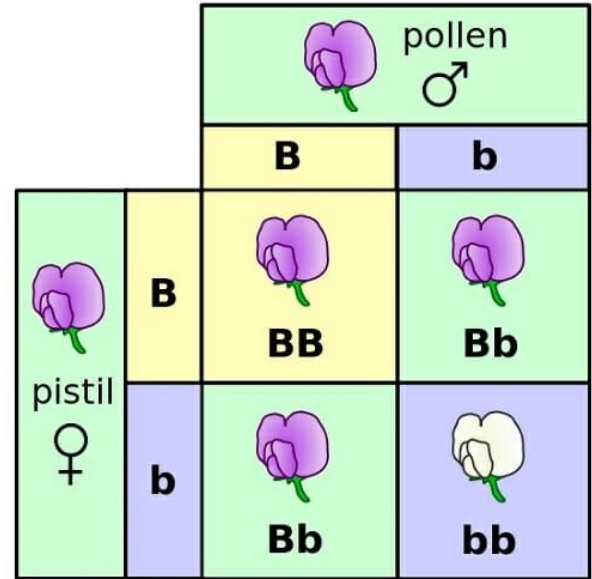- Planted 929 plants and observed their colors
- Let's simulate that...

Image from biologydictionary.net

(Demo)

Mendel's Observations

- Mendel observed 705 of the 929 plants had purple flowers
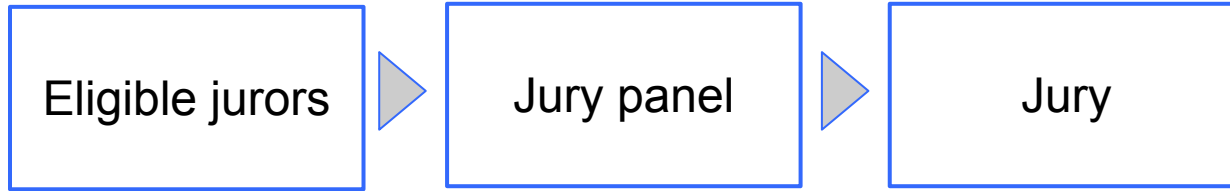- Is that consistent with our simulation?

(Demo)

Different Statistics

The statistic was "proportion of plants with purple flowers". We could have chosen a different statistic, and predicted the values of that statistic.

(Demo)

Swain v. Alabama

Juries

Eligible jurors ▷ Jury panel ▷ Jury

United States Jury Act of 1968: [A]ll litigants in Federal courts entitled to trial by jury shall have the right to...juries selected at random from a fair cross section of the community in the district or division wherein the court convenes.
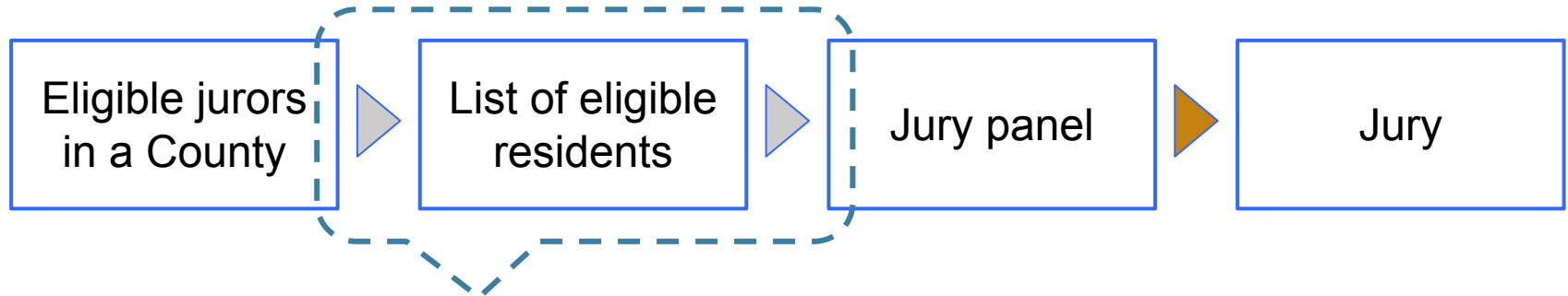
Robert Swain v. Alabama

1965 Supreme Court case about jury selection

- In Talladega, Alabama, 26% of residents were black
- In Swain's jury panel, 8 of 100 panelists were black
- All 8 were struck from the jury by the prosecution (using peremptory challenges)

Ruling: "The overall percentage disparity has been small and reflects no studied attempt to include or exclude a specified number of [black men]."

(Demo)

Jury Panels

Eligible jurors in a County → List of eligible residents → Jury panel → Jury

Section 197 of California's Code of Civil Procedure says, "All persons selected for jury service shall be selected at random, from a source or sources inclusive of a representative cross section of the population of the area served by the court."

(Demo)

Statistics

A statistic is a number calculated from a sample

Total Variation Distance

Every distance has a computational recipe

Total Variation Distance (TVD):

- For each category, compute the difference in proportions between two distributions

- Take the absolute value of each difference

- Sum & divide by 2

(Demo)