190F Foundations of Data Science

Spring 2020

Lecture 10

Hypothesis Testing

Announcements

Science

- **Hypothesis:** The sun always rises in the east.
- **Prediction:** If you observe a bunch of sunrises, all of them will be in the east.
- **Test:** Go watch some sunsets.

Science

- **Hypothesis:** All swans are white.
- **Prediction:** If you observe a bunch of swans, all of them will be white.
- **Test:** Go find swans.

Where it gets hairy

- **Hypothesis:** Most swans are white.
- **Prediction:** ???
- **Test:** ???

Dealing with randomness

- **Hypothesis:** At least 99% of swans are white.
- **Prediction:** ???
- **Test:** ???

Dealing with randomness

- **Hypothesis:** At least 99% of swans are white.
- **Prediction:** If you observe a random sample of 100 swans, at least 99 will be white.

- **Test:** ???

Dealing with randomness

- **Hypothesis:** At least 99% of swans are white.
- **Prediction:** If you observe a random sample of 100 swans, more likely than not, at least 99 will be white.
- **Test:** ???

Dealing with randomness

- **Hypothesis:** At least 99% of swans are white.
- **Prediction:** If you observe a random sample of 100 swans, almost certainly, at least 90 will be white.
- **Test:** ???

Dealing with randomness

- **Hypothesis:** At least 99% of swans are white.
- **Prediction:** If you observe a random sample of 100 swans, almost certainly, at least 90 will be white.
- **Test:** Randomly sample 100 swans. Observe their color.

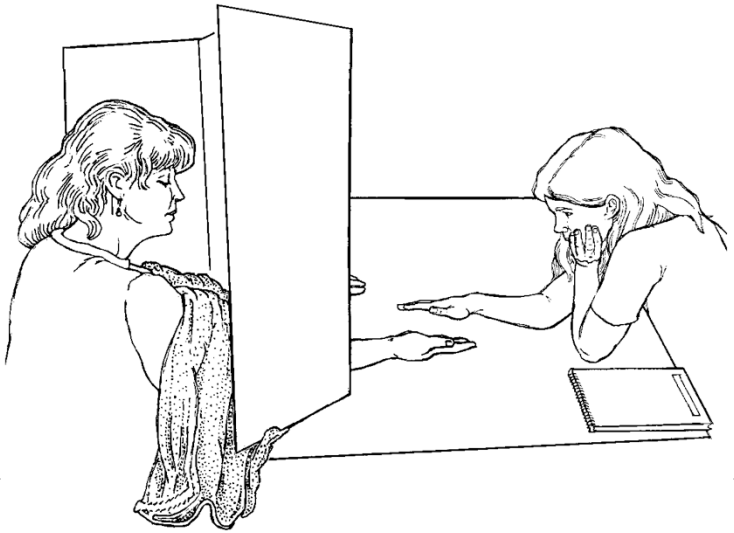
Interpreting the result of a test

- **Test:** Randomly sample 100 swans. Observe their color.
- If at least 90 are white: hypothesis could be true.
- If fewer than 80 are white: either the hypothesis is false, or a ginormous coincidence just happened.

Emily Rosa and the healing touch

Emily Rosa's 4th grade science fair project:

Touch healers say they detect and use the healing energy field (HEF) to cure patients. Can they detect HEF?

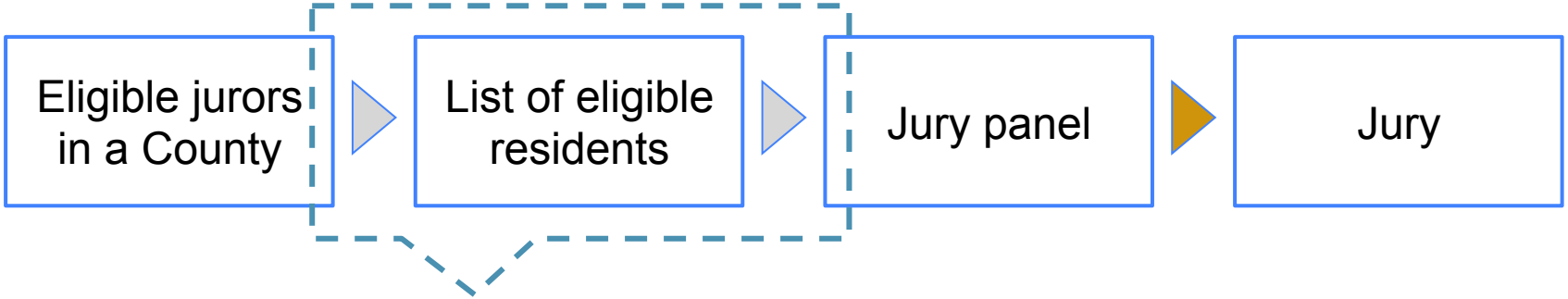


Science

- **Hypothesis:** Healers cannot detect HEF better than random guessing.
- **Prediction:** If you test healers 150 times, almost certainly, they'll be right at most 60% of the time.
- **Test:** Go find healers and test them.

Jury Panels

Jury Panels



"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)

Distances

Computing 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

Chi Squared (χ^2 *Optional*):

- For each category, compute the difference in proportions between two distributions
- Square each difference and divide by the first proportion
- Sum & multiply by sample size

(Demo)

Empirical Distributions

Distribution of a Statistic

Statistic: A quantity computed for a particular sample

Distribution: The chance of each outcome of sampling

Sampling distribution: Chance of each value of a statistic (computed from all possible samples)

Also known as the *probability distribution of the statistic*

Empirical distribution: Observations of a statistic (computed from some samples drawn at random)

(Demo)

Hypothesis Testing

Testing a Hypothesis

Step 1: Select Two Hypotheses

- A test chooses between two views of how data were generated: *Null hypothesis* proposes that data were generated at random; *Alternative hypothesis* proposes some effect other than chance

Step 2: Choose a Test Statistic

- A value that can be computed from the data

Step 3: Compute What The Null Hypothesis Predicts

- Compute the distribution of the test statistic: what the test statistic might be if the null hypothesis were true.

Step 4: Compare the Prediction to the Observed Data

Hypothesis Testing Logic

Define 2 mutually exclusive descriptions: either this or that.

- One of them can be evaluated using probability (the null hypothesis)
- You can "reject the null," so then you “accept” the alternative.
- Otherwise: you're still not sure, but null looks plausible.

Step 1:
Select Two Hypotheses

Example: The Two Hypotheses

Gregor Mendel (1822-1884) was an Austrian monk and founder of the modern field of genetics. Among many experiments, he tested the hypothesis that pea plants will bear purple or white flowers at random, in the ratio 3:1.

- **Mendel's model describes the world.** If the distribution of the observed plants is different from the distribution in the model, it's just chance variation.

- **Mendel's model doesn't.**

Null

Alternative

(Demo)

Example: Smoking and Babies

Researchers are interested in whether there is an association between smoking mothers and the health of their babies. For each birth, they record the baby's birth weight and whether the mother smokes or not.

- **Birth weights aren't affected by maternal smoking.** The birthweight distribution for babies of smokers is same as that of babies of non-smokers.

Null

- **They are affected.**

Alternative

Example: Smoking and Babies

Researchers are interested in whether there is an association between smoking mothers and the health of their babies. For each birth, they record the baby's birth weight and whether the mother smokes or not.

- **Birth weights aren't affected by maternal smoking.** The birthweight distribution for babies of smokers is same as that of babies of non-smokers.
- **They are lower.** Birthweight of babies of smokers are lower than birthweights of babies of non-smokers.

Step 2:
Choose a Test Statistic

Choosing a Test Statistic

Test statistic: The statistic that you have chosen to calculate, to help you decide between the two hypotheses

Goal: If the null hypothesis is false, then you expect that measuring the test statistic will allow you to reject the null

Choosing a Test Statistic

For Mendel's pea flower
would be reasonable tes
(Choose all that are OK.)

If the alternative hypothesis is true, will test statistic be *larger* than prediction, *smaller*, or *could be either way*?

1. The proportion of plants with purple flowers.
2. The proportion of plants with white flowers.
3. abs(p - 0.75), where p is the proportion of plants with purple flowers.
4. The number of different colors in the plants flowers.
5. The total variation distance between the distribution in the observed data, vs the model distribution (0.75, 0.25)

Choosing a T

If the alternative hypothesis is true, will test statistic be *larger* than prediction, *smaller*, or *could be either way*?

For the baby birth we … following would be re… (Choose all that are OK.)

1. The average birth weight of all the babies.
2. The proportion of babies whose mother smoked.
3. The average birth weight of babies of smokers, minus the average birth weight of babies of non-smokers.
4. The absolute value of the previous difference.

Absolute Values & Alternatives

- Choose a test statistic where alternative hypothesis suggests which direction statistic will go.
- **Alternative: Smoking causes poor health.**
  - Test statistic: Average birth weight for smokers, minus average for non-smokers.
- **Alternative: Smoking has some relation to health.**
  - Test statistic: Absolute value of that difference.

Step 3:
Compute the Distribution of the Test Statistic under the Null Hypothesis

Step 4:
Compare the Prediction to the Observed Data

Conclusion of a Test

Resolve choice between null and alternative hypotheses

- Compare observed test statistic to its empirical distribution under the null hypothesis
- If the observed value is **consistent** with the distribution, then the test *does not* reject the null hypothesis

Whether a value is consistent with a distribution:

- A visualization may be sufficient
- Convention: The observed significance level (P-value)

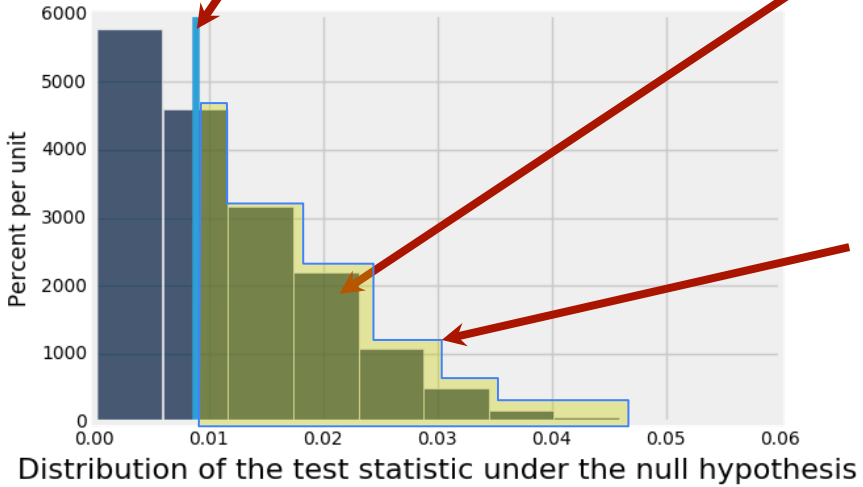
Definition of *P*-value

The P-value is the chance,

- under the null hypothesis,
- that the test statistic
- is equal to the value that was observed in the data or is even further in the direction of the alternative.

Quantifying Conclusions

P(the test statistic would be equal to or more extreme than the observed test statistic under the null hypothesis)



Evaluating Mendel's pea flower hypothesis

This area is the P-value (approximately)

Conventions of Consistency

- **"Inconsistent":** The test statistic is in the tail of the null distribution.
- **"In the tail," first convention:**
  - The area in the tail is less than 5%.
  - The result is "statistically significant."
- **"In the tail," second convention:**
  - The area in the tail is less than 1%.
  - The result is "highly statistically significant."