DSFA

Spring 2019

Lecture 18

Percentiles and the Bootstrap

Announcements

- Project 2: Due Tuesday, April 9 and April 16
- Prelim 2: In-class. Tuesday, April 16
  (Not Tuesday after spring break)

One- versus Two-sided Tests

- **Mendel's Peas**
  Null Hypothesis: Probability of purple flower is 0.75 ($p = 0.75$)
  Alternative Hypothesis: Probability is not 0.75 ($p != 0.75$)
  Test Statistic: | p_hat - 0.75 |

  Two-sided test: 'Large' deviation from the null in either direction leads to rejection (of the null hypothesis)

- **Jelly Beans**
  Null Hypothesis: No effect on the probability of acne ($p = 0.2$)
  Alternative Hypothesis: Increase the probability of acne ($p > 0.2$)
  Test Statistic: p_hat - 0.2

  One-sided test: 'Large' positive deviation from the null leads to rejection

Conclusions From a Test

Hypothesis test

Fail to reject the null hypothesis (data is not inconsistent with the null hypothesis - inconclusive)

Reject the null hypothesis (data is inconsistent with the null hypothesis - accept the alternative)

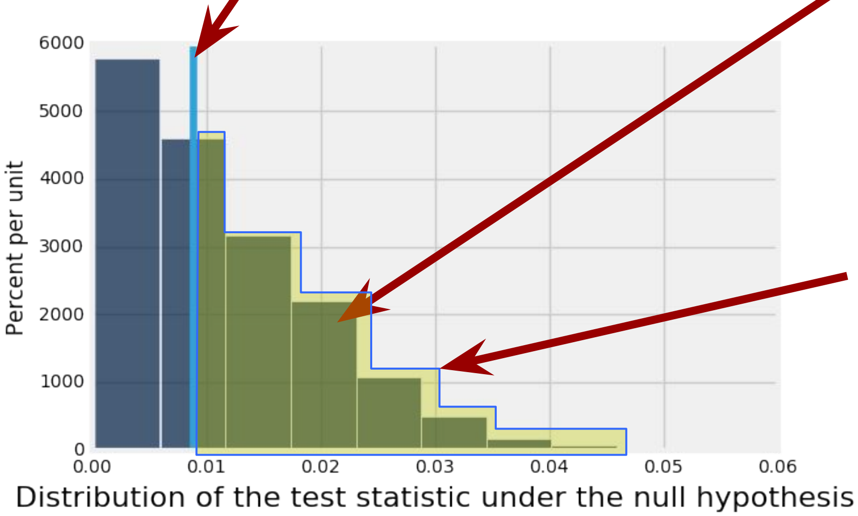
Definition of *P*-value

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."

Can the Conclusion be Wrong?

Yes.

	Null is true	Alternative is true
Test rejects the null	✗	✓
Test doesn't reject the null	✓	✗

(Demo)

An Error Probability

- The cutoff for the P-value is an error probability.
- If:
  - your **cutoff is 5%** (your significance level)
  - and the **null hypothesis happens to be true**
  - (but you don’t know that)
- then there is about a **5% chance** that **your test will reject the null hypothesis anyway**.

Type I and Type II errors

- The significance level (or p-value cutoff) is the probability of a Type I error

  Type I error = Reject null when it is true

- What if the alternative is true?

  Type II error = Fail to reject null when it is false

More on P-Hacking

Suppose you conduct 10 independent hypothesis test, each at a 5% significance level; i.e. the null hypothesis is rejected if $p < 0.05$.

The probability that at least one null hypothesis is rejected is

A. 0.05 B. <0.4 C. >0.4 D. >0.5 E. 0.95

Mendel versus Fisher

Ronald Fisher (1936), Commenting on the fact that Mendel's results were too good to be true:

"Mendel was deceived by some assistant who knew too well what was expected"

Pea - hacking !!

Percentiles

Computing Percentiles

The 80th percentile of a set of numbers is the smallest value in the sample that is at least as large as 80% of the sample

For `s = [1, 7, 3, 9, 5]`, `percentile(80, s)` is 7

The 80th percentile is ordered element 4: `(80/100) * 5`

Percentile | Size of set

For a percentile that does not exactly correspond to an element, take the next greater element instead

The `percentile` Function

- The *p*th percentile is the smallest value at least as large as *p*% of the values in the sample
- Function in the `datascience` module:

  `percentile(p, values)`

- **`p`** is between 0 and 100
- Returns the *p*th percentile of the array

Discussion Question

Which are `True`, when `s = [1, 7, 3, 9, 5]`?

```
percentile(10, s) == 0

percentile(39, s) == percentile(40, s)

percentile(40, s) == percentile(41, s)

percentile(50, s) == 5
```

(Demo)

Estimation (Review)

Inference: Estimation

- What is the value of a population parameter?

- If you have a census (that is, the whole population):
  - Just calculate the parameter and you're done

- If you don't have a census:
  - Take a random sample from the population
  - Use a statistic as an **estimate** of the parameter

(Demo)

Variability of the Estimate

- One sample ➜ One estimate
- But the random sample could have come out differently
- And so the estimate could have been different
- Main question:
  - **How different could the estimate have been?**
- The variability of the estimate tells us something about how accurate the estimate is:

  estimate = parameter + error

(Demo)

Where to Get Another Sample?

- One sample ➜ One estimate
- To get many values of the estimate, we needed many random samples
- Can’t go back and sample again from the population:
  - No time, no money
- Stuck?

The Bootstrap

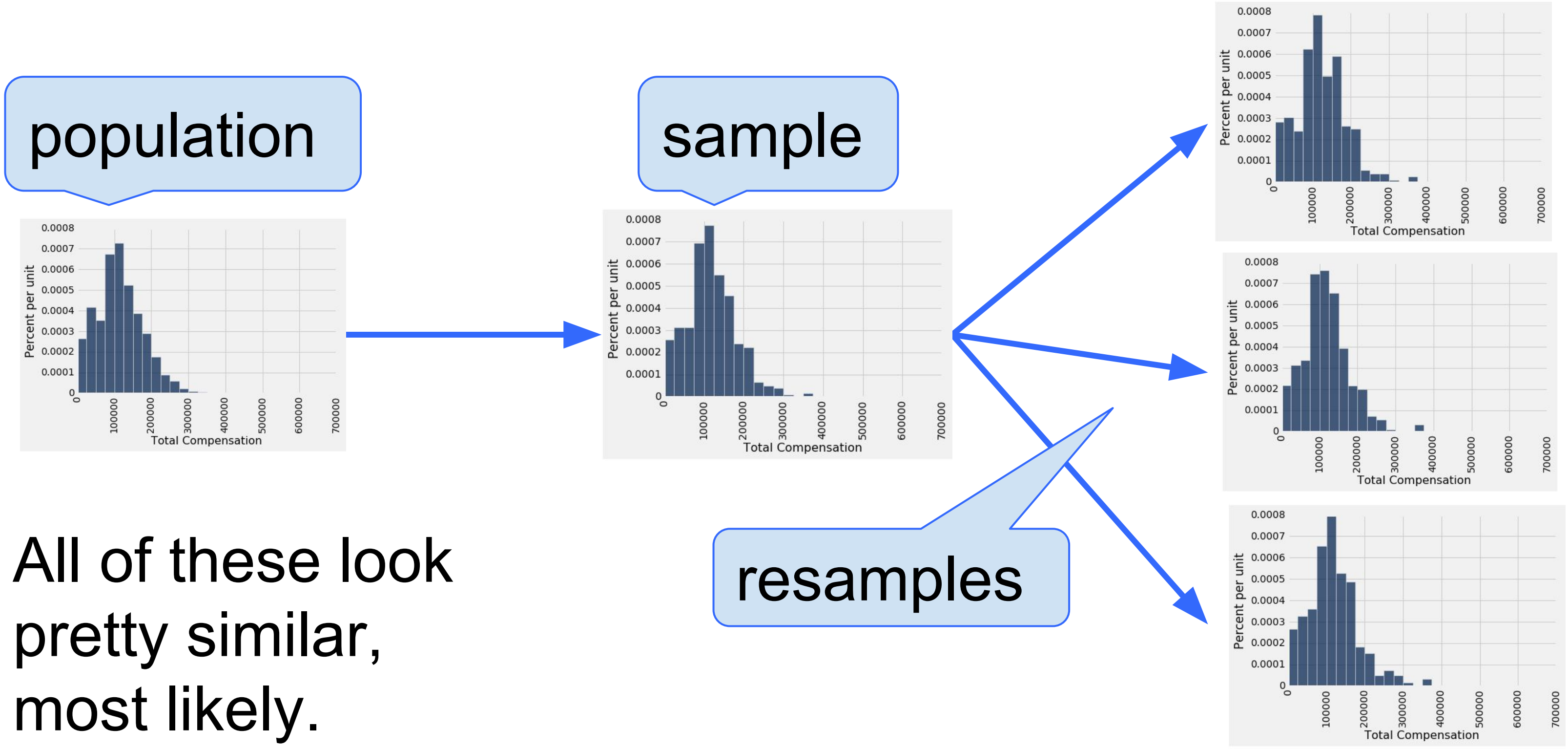
The Bootstrap

- A technique for simulating repeated random sampling

- All that we have is the original sample
  - … which is large and random
  - Therefore, it probably resembles the population

- So we sample at random from the original sample!

Why the Bootstrap Works



All of these look pretty similar, most likely.

Key to Resampling

- From the original sample,
  - draw at random
  - with replacement
  - as many values as the original sample contained

- The size of the new sample has to be the same as the original one, so that the two estimates are comparable

(Demo)