NYU Center for Data Science

DS-UA 111
Data Science for Everyone

Week 11: Lecture 1
Comparing Two Samples

NYU Center for Data Science

Could we use hypothesis testing to assess the relationship between two samples?

DS-UA 111
Data Science for Everyone

Week 11: Lecture 1
Comparing Two Samples

Adapted from Adhikari, DeNero, Wagner, Milner, Dudoit

Announcements

- Please check Week 11 agenda on NYU Classes
  - Homework 3/4
  - Lab 6
  - Project Milestone
- Refer to the Calendar linked to NYU Classes

Review

- Hypothesis Testing
  - Null Hypothesis
  - Alternative Hypothesis
  - Test Statistics
  - Empirical Distribution
  - P-Values

References

- Hypothesis Testing:
  - Chapter 11.3, 11.4

Review

- Hypothesis Testing
  - We use tests to connect questions and answers about the data generating processes in a population
  - With hypothesis testing, we have two possibilities summarized by null hypothesis and alternative hypothesis

False Reject

False Accept

	Null is True	**Alternative is True**
Test Favors the Null	Correct result	Error
Test Favors the Alternative	Error	Correct result

- The null hypothesis asserts that the processes follow a pattern. However, the alternative hypothesis asserts the processes do not follow the pattern.
- While we accept or reject the null hypothesis, we should think in terms of rejecting or failing to reject the null hypothesis. We must remember that hypothesis tests provide evidence not proof for the null hypothesis or alternative hypothesis

Review

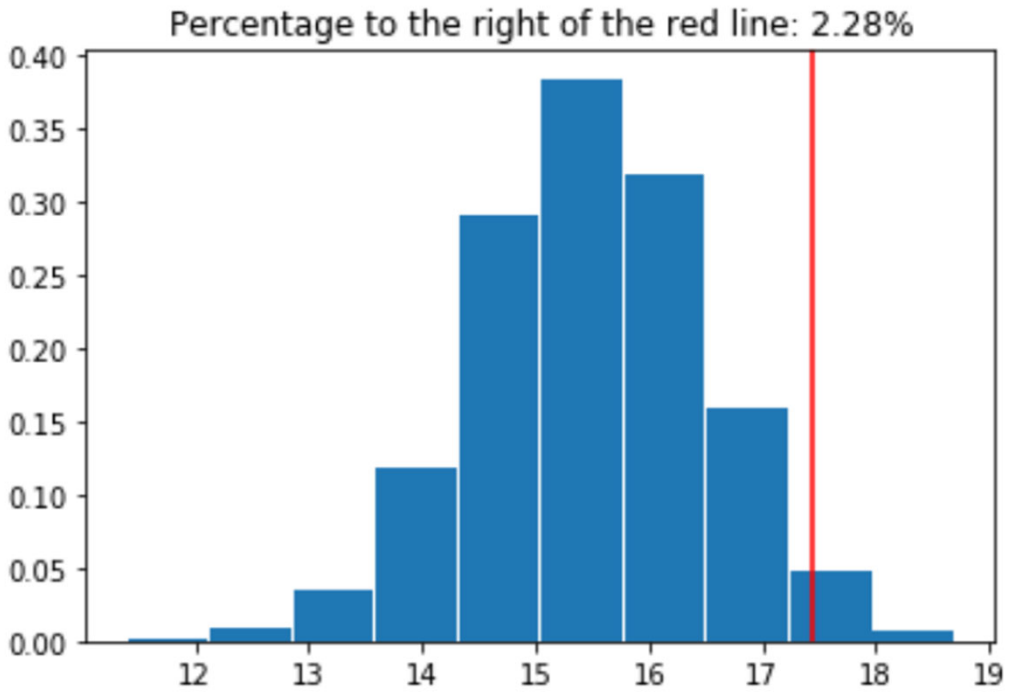
- Hypothesis Testing
  - If the observed test statistic is consistent with the simulated test statistics in the empirical distribution then we lack evidence to reject the null hypothesis
  - However, the randomness in the testing procedure allows for the possibility of false accept and false reject of the null hypothesis

	Null is True	**Alternative is True**
Test Favors the Null	Correct result	Error
Test Favors the Alternative	Error	Correct result

- If the alternative hypothesis is true, then we expect the observed test statistic to be an outlier. So we would be less worried about false accept.
- However, if the null hypothesis is true, then random sampling will lead to false reject with probability related to the threshold for p-values
- We state in the null hypothesis that any deviation in outcomes owes to chance because it reminds us that some test statistics must be outliers

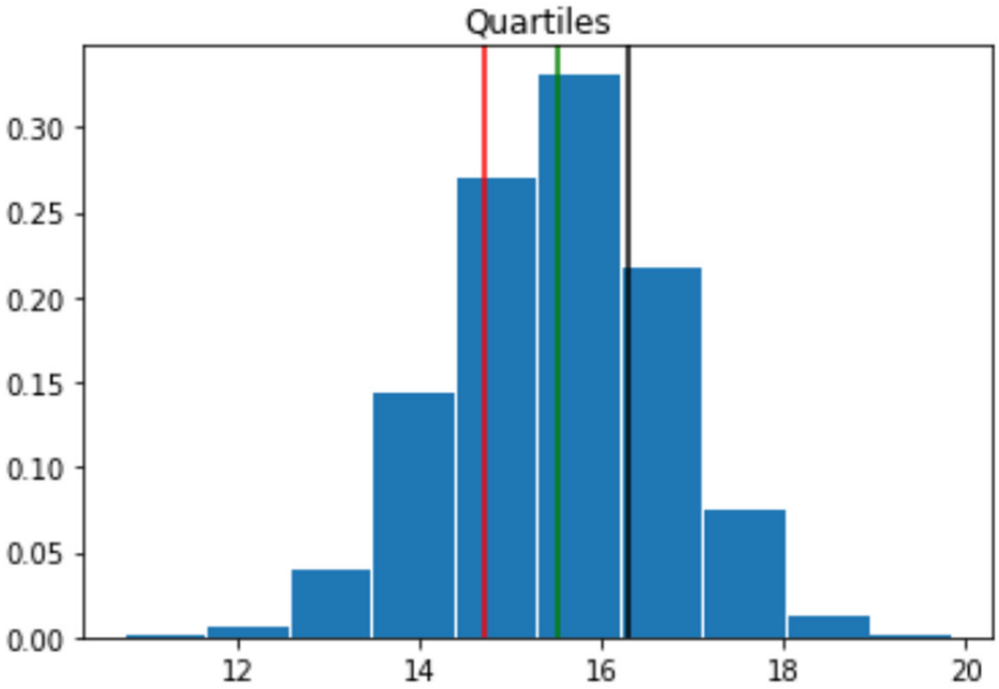
Review

- For hypothesis testing we need to compare the observed test statistic to the simulated test statistic under the null hypothesis. We check the left tail and right tail for outliers.
- We want to estimate the probability of a test statistic obtaining a value farther to the left in the left tail or farther to the right in the right tail.
- The number of simulated test statistics provides an estimate of the probability nicknamed p-value.



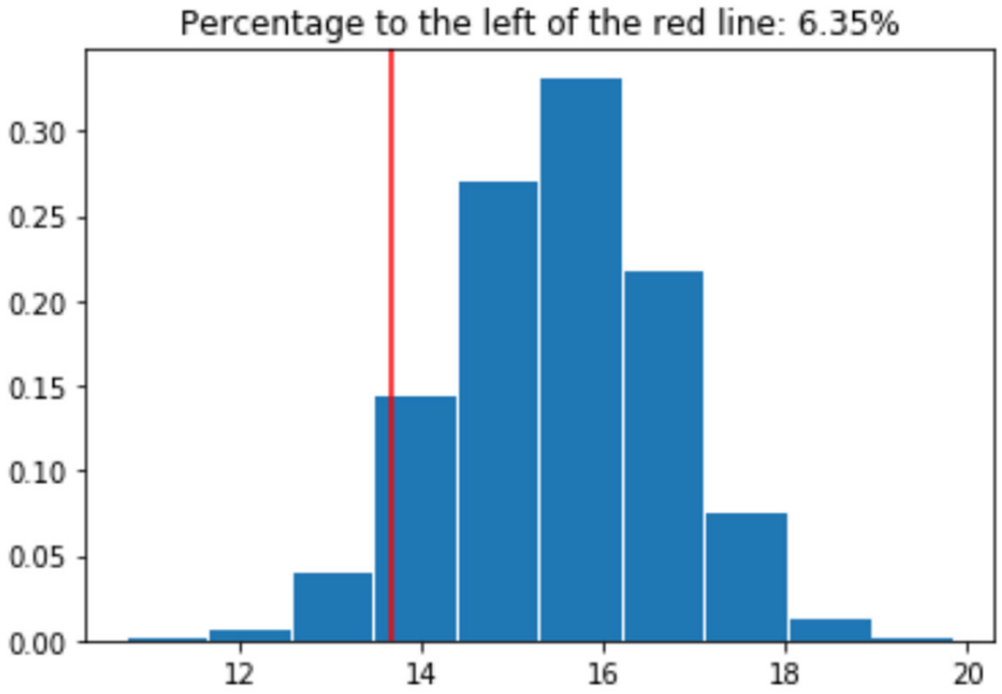
Review

- We can set thresholds for p-values with quantiles. Remember that quantiles are cut-points. These cut-points divide the data into subsets of equal size.
- We need to sort the data in increasing order to determine the cut-points. Commonly we use percentiles which divide the data into subsets of size 1/100.
- Just like we can compare a p-value to a level that sets a threshold for accepting or rejecting



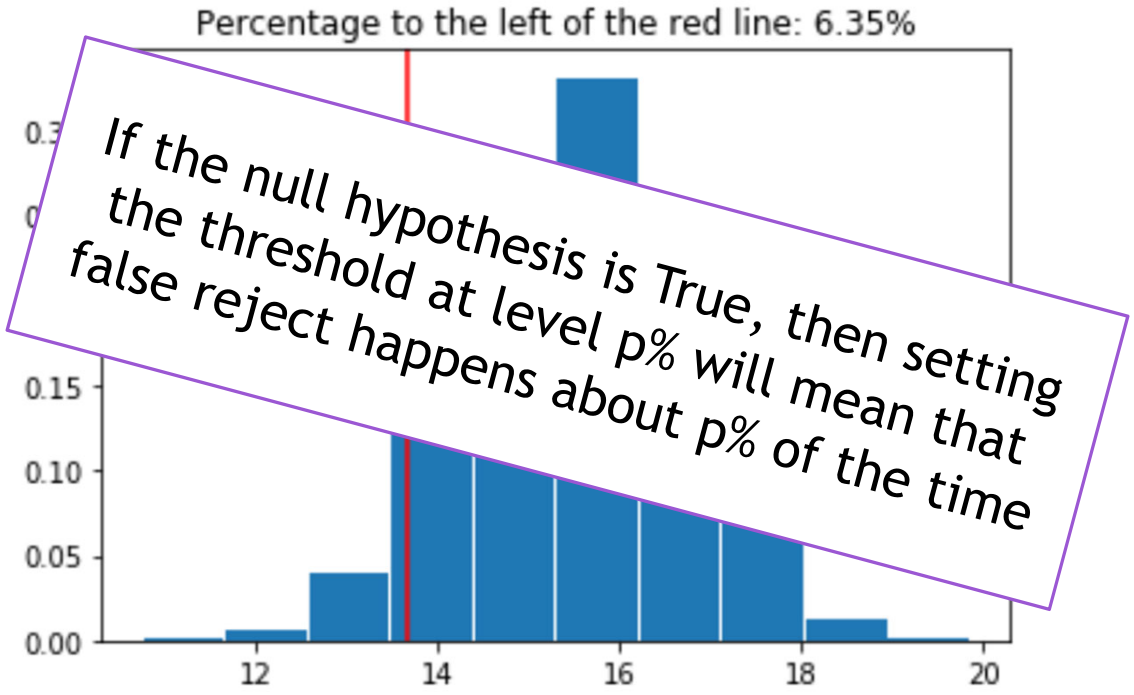
Review

- We can compare a p-value to an observed significance level that sets a threshold for accepting or rejecting the null hypothesis. Common levels are 1% or 5%.
- Instead of comparing the p-value to the level, we can compare the corresponding percentile to the observed test statistic.
- So the level sets a threshold for a p-value and the corresponding percentile determines rejection region in the left and / or right tails



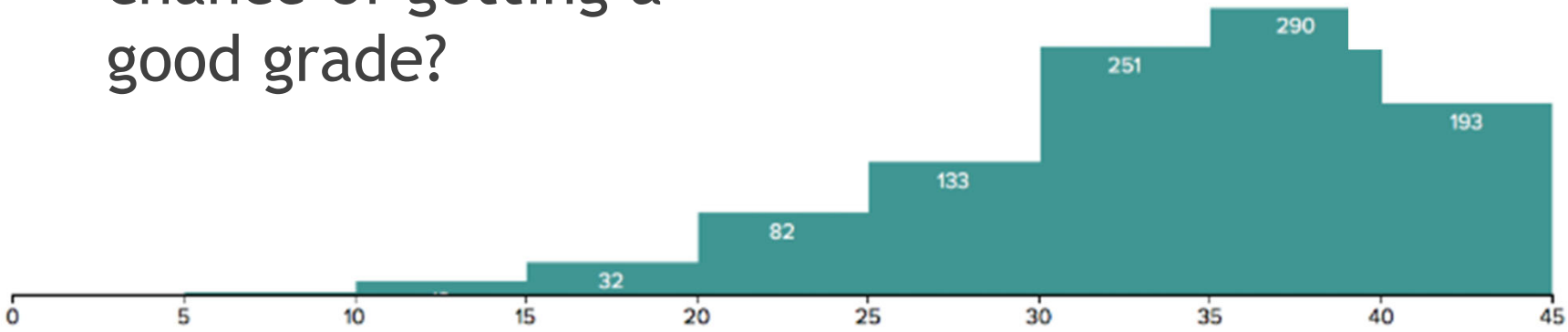
Review

- We can compare a p-value to an observed significance level that sets a threshold for accepting or rejecting the null hypothesis. Common levels are 1% or 5%.
- Instead of comparing the p-value to the level, we can compare the corresponding percentile to the observed test statistic.
- So the level sets a threshold for a p-value and the corresponding percentile determines rejection region in the left and / or right tails



Exercise

- How can we use hypothesis testing to validate the assumption that each section has the same chance of getting a good grade?
- Suppose we want to assess the grades on assignments across the different sections of class



Agenda

- Studies
  - Observational
  - Experimental
- Permutation Testing
  - Do two samples come from the same population?

References

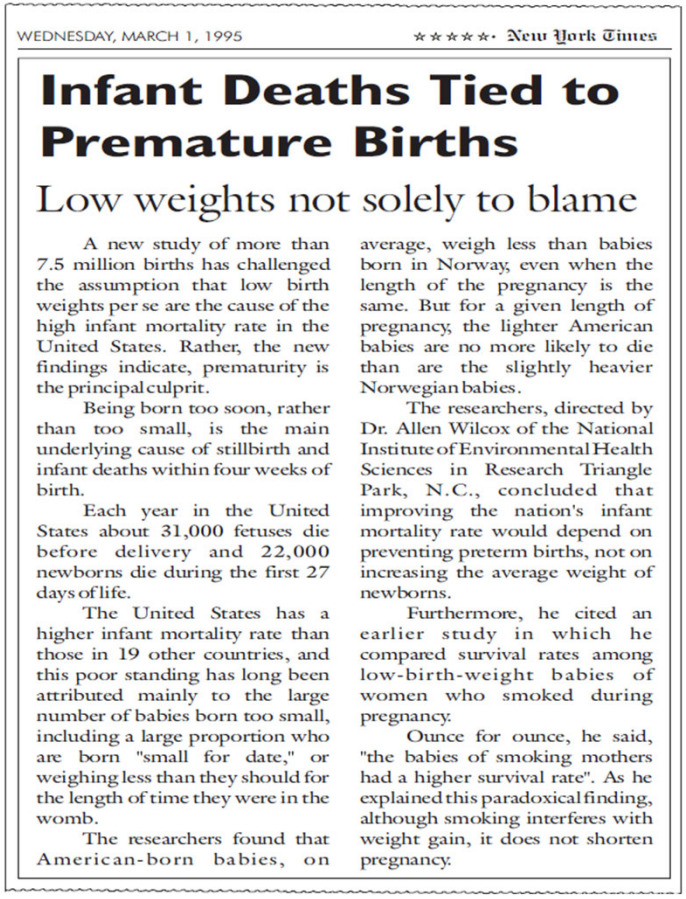
- Comparing Samples
  - Chapters 12.1

Observations and Experiments

- Researchers use experiments to study the association or causation among variables in situations where they control on values of these variables.
- Search engines and social media use different versions of their websites to experiment with user experience

Observations and Experiments

- Researchers use observations to study association among variables in situations where they cannot control the values of these variables.
- Physician gathered data in the Child Health and Development Studies on links between maternal smoking and infant health



WEDNESDAY, MARCH 1, 1995 ★★★★★ New York Times

Infant Deaths Tied to Premature Births

Low weights not solely to blame

A new study of more than 7.5 million births has challenged the assumption that low birth weights per se are the cause of the high infant mortality rate in the United States. Rather, the new findings indicate, prematurity is the principal culprit.

Being born too soon, rather than too small, is the main underlying cause of stillbirth and infant deaths within four weeks of birth.

Each year in the United States about 31,000 fetuses die before delivery and 22,000 newborns die during the first 27 days of life.

The United States has a higher infant mortality rate than those in 19 other countries, and this poor standing has long been attributed mainly to the large number of babies born too small, including a large proportion who are born "small for date," or weighing less than they should for the length of time they were in the womb.

The researchers found that American-born babies, on average, weigh less than babies born in Norway, even when the length of the pregnancy is the same. But for a given length of pregnancy, the lighter American babies are no more likely to die than are the slightly heavier Norwegian babies.

The researchers, directed by Dr. Allen Wilcox of the National Institute of Environmental Health Sciences in Research Triangle Park, N.C., concluded that improving the nation's infant mortality rate would depend on preventing preterm births, not on increasing the average weight of newborns.

Furthermore, he cited an earlier study in which he compared survival rates among low-birth-weight babies of women who smoked during pregnancy.

Ounce for ounce, he said, "the babies of smoking mothers had a higher survival rate". As he explained this paradoxical finding, although smoking interferes with weight gain, it does not shorten pregnancy.

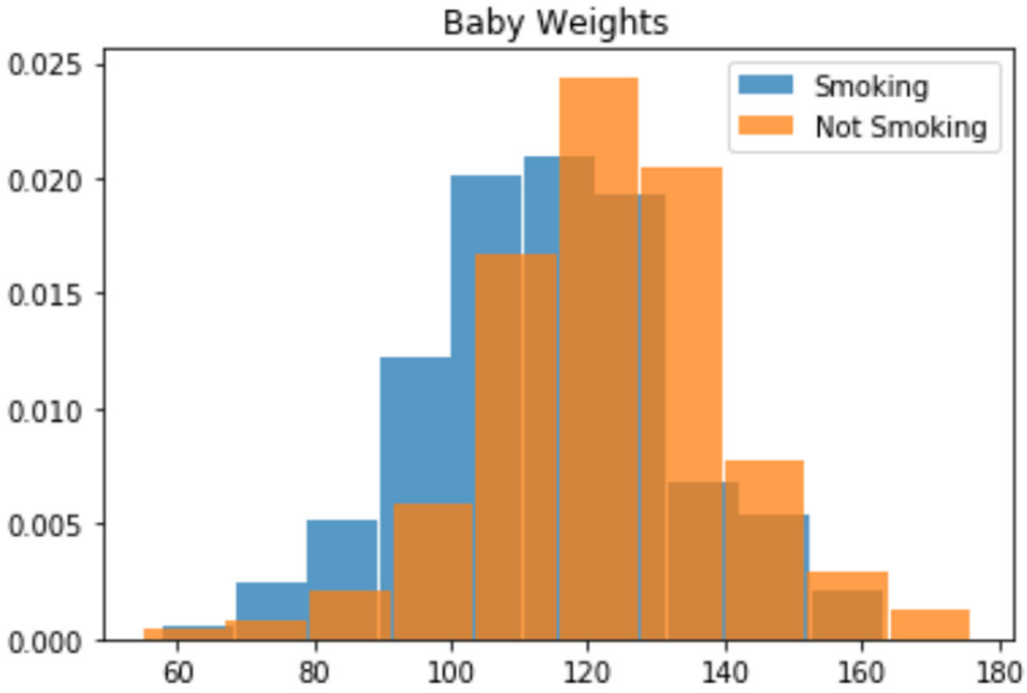
Permutation Testing

- We can use permutation testing in both experimental studies and observational studies to compare samples
  - Do two samples come from the same population?
  - Does the distribution of a property in sample A match the distribution of a property in sample B
- Permutation means shuffling. We will use hypothesis testing with data generated from shuffling the records between the two group.

	Maternal Smoker	Birth Weight	Shuffled Labels
0	False	120	False
1	False	113	False
2	True	128	False
3	True	108	False
4	False	136	False

Permutation Testing

- Commonly permutation testing will compare the means of the groups. We can use the difference in means for the test statistic.
- We use an hypothesis test of the form
  - Null Hypothesis : The samples come from the same population. Any difference in the properties of the samples owes to randomness.
  - Alternative Hypothesis : The samples come from different populations



Summary

- Studies
  - Observational
  - Experimental
- Permutation Testing
  - Do two samples come from the same population?

Goals

- Understand the probability of false reject under the null hypothesis based on level
- Use random sampling with replacement to shuffle the labels between two groups