



DATA 8
Fall 2016

Lecture 18, October 5

Testing Hypotheses

Slides created by Ani Adhikari and John DeNero

Announcements

- Nice work on Project 1!
 - Homework: early deadline today at **7 pm**, regular deadline tomorrow 10/6 at **7 pm**.
 - All regrade requests for Homework 1-4 are due by Friday 10/7 at 7PM.
 - Midterm is on Friday Oct 14, less than two weeks away.
 - Talk to your GSI if you are interested in few-on-one tutoring.
-

“Goodness of Fit”

Assessing if a categorical sample was drawn randomly from a known population:

- Decide on a statistic that measures the distance between distributions
 - Compute the statistic from the sample; that is, the distance between the distributions of sample and the known population
 - Sample at random from the population and compute the statistic from the random sample; repeat numerous times
 - Compare the empirical distribution with the value that was calculated from the sample
-

Testing a Hypothesis

Step 1: The Hypotheses

- A test chooses between two views of how data were generated
- **Null hypothesis:** Data are generated by a specified chance process
- **Alternative hypothesis:** There is some effect other than chance

Step 2: The Test Statistic

- A statistic that can be used to help choose between the hypotheses

Step 3: The Probability Distribution of the Test Statistic

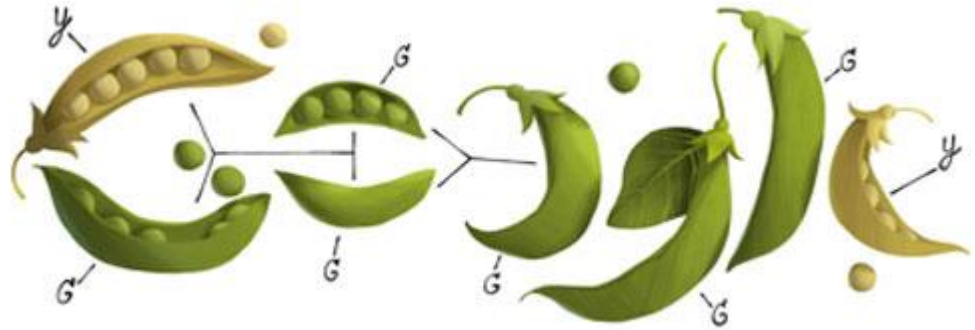
- What the test statistic might be if the null hypothesis were true
 - Approximate the probability distribution by an empirical distribution
-

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 statistic is consistent with the null distribution**, then the data support the null better than they support the alternative
 - If the **observed statistic is not consistent with the null distribution**, then the data support the alternative better than they support the null
-

Gregor Mendel, 1822-1884



(Demo)
