Data 8

Fall 2016

Lecture 19, October 7

Error Probabilities

Slides created by Ani Adhikari and John DeNero

Announcements

- New Homework today! Due Wed/Thu of next week. We will post solutions promptly.
- All regrade requests for Homework 1-4 are due today by 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.

Statistical tests of hypotheses

- You have two hypotheses about the world. For example:
  - “The world behaves in this way.”
  - “The world doesn’t behave in this way.”

- You have a sample of data.

- You try to see which of the two hypotheses is better supported by the data.

The two hypotheses

- **Mendel's model is good.** The distribution of the plants is different from the distribution in the model due to chance variation.

Null

Alternative

- **Mendel's model is not good.**

Common informal statements

- **Null**:
  - Nothing is going on, except chance variation.
  - The difference between the two distributions “isn’t real”.
- **Alternative:**
  - Something other than chance is going on (but we’re not hypothesizing what that “something” is).
  - The difference “is real”.

The evidence

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

In our example, the test statistic is the **absolute distance between the percent purple in Mendel's model and the observed percent purple.**

The null hypothesis, in general

The sample has been generated at random

under precise assumptions about the randomness

that make it possible for us to calculate chances involving the test statistic.

The null again, informally

“The difference is due to chance.”

1. Which difference?

   The difference between:

- the observed value of the test statistic
- the values of the test statistic as predicted by the null hypothesis

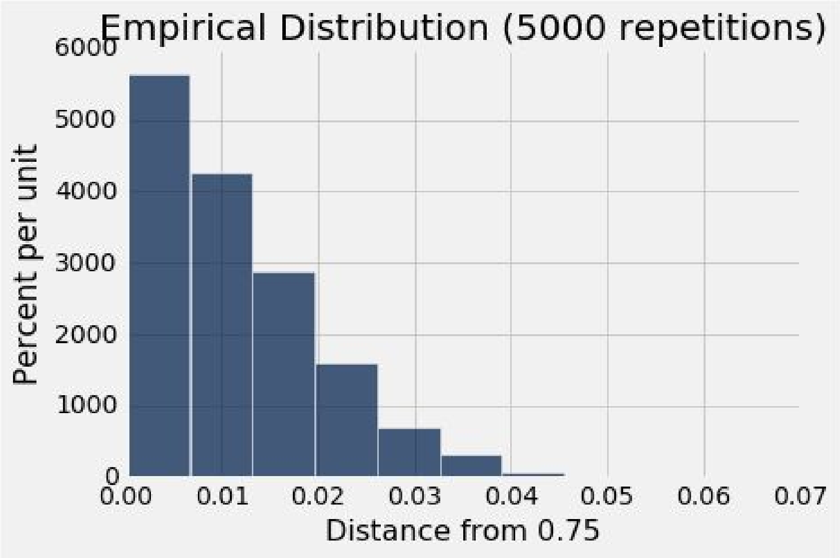
2. Where does chance come in?

   The assumptions of randomness made in the null hypothesis

The statistic under the null

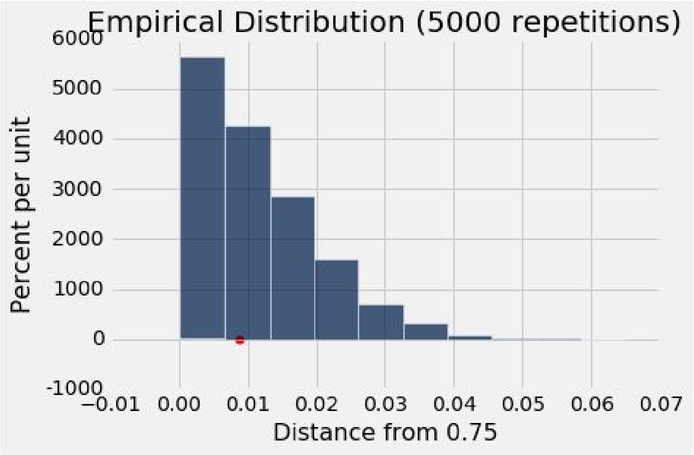
Assume the null hypothesis is true.

Under this assumption, calculate (at least roughly) the probability distribution of the test statistic.



Compare with the observed statistic

Compare this “null distribution” with the observed value of the test statistic in your sample.



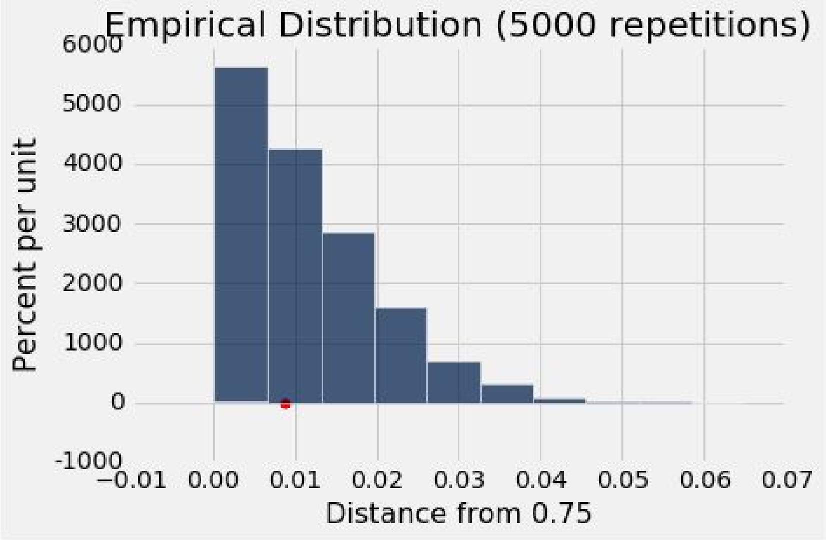
The observed statistic: 0.00888

The conclusion

- If your observed test statistic is **inconsistent** with what the null predicts, then the data are inconsistent with the null. So you **reject the null hypothesis.**

- If your observed test statistic is **consistent** with what the null predicts, you **can’t reject the null hypothesis.** The data support the null hypothesis better than they support the alternative.

The conclusion in our example



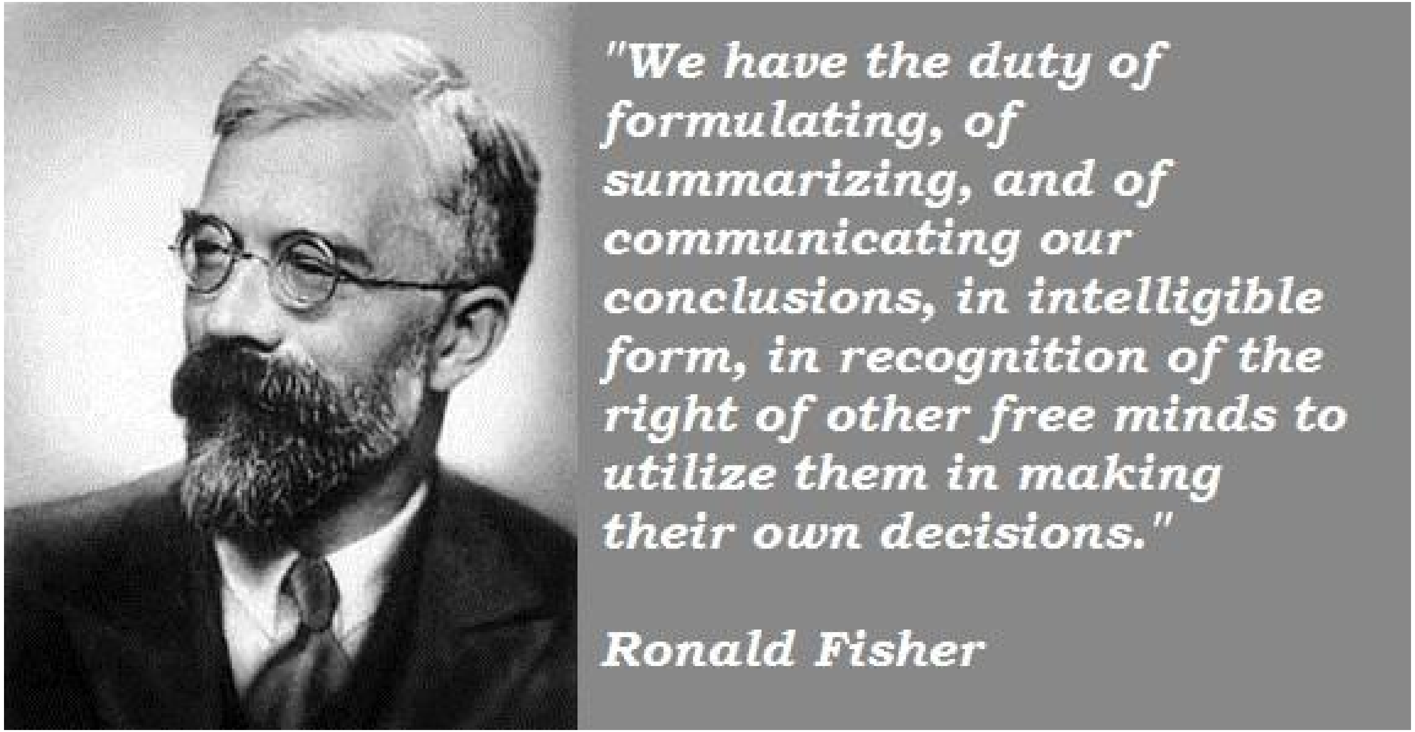
By any reasonable definition of "consistent," the observed value is consistent with the values predicted by the null.

Conclusion: The data support the null better than the alternative. Mendel's model looks good by this test.

“Inconsistent”: the conventions

- **“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.”

Sir Ronald Fisher, 1890-1962



Sir Ronald Fisher, 1925

"It is convenient to take this point [5%] as a limit in judging whether a deviation is to be considered significant or not."

— *Statistical Methods for Research Workers*

Sir Ronald Fisher, 1926

"If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty (the 2 percent point), or one in a hundred (the 1 percent point). Personally, the author prefers to set a low standard of significance at the 5 percent point …"

P-value

Mark your test statistic on the horizontal axis of the histogram of the null distribution. The *P*-value is:

- the area in the tail
- the “**observed significance level**”

In **Methods** sections of papers:

$$s = 13.82,\ P = 0.0071^{**}$$

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.

(Demo)

Can a test's conclusion be wrong?

Yes.

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

An error probability

- The cutoff for the P-value is an error probability.
- If:
  - your **cutoff is 6%**
  - and the **null hypothesis happens to be true**
  - (but you don't know that)
- then there is about a **6% chance** that **your test will reject the null hypothesis**.

Assess this:

"Statistical significance is an objective, unambiguous, universally accepted standard of scientific proof.

— Letter to *Nature*, 1994