

**DSFA**  
Spring 2021

# Lecture 19

---

Error Probabilities

# Announcements

---

- Project 1, part 2 due today, 5:59PM
  - HW 4 out today, due next Friday
  - Prelim 1 regrade requests due Monday at 5PM
-

# 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 two mutually exclusive descriptions: either this or that.  
One of them can be evaluated using probability (the null hypo.)  
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.**

Alternative

Null

(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

For Mendel's pea flower color data, which test would be reasonable to use?  
(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.  $\text{abs}(p - 0.75)$ , 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  $(0.75, 0.25)$

When poll is active, respond at [pollev.com/dsfa](https://pollev.com/dsfa)

Text **DSFA** to **22333** once to join

# For Mendel's pea plant experiment, which of the following would not be a reasonable test statistic?

The proportion of plants with purple flowers

The proportion of plants with white flowers

$\text{abs}(p - 0.75)$ , where  $p$  is the proportion of plants with purple flowers

The number of different colors in the plants' flowers

The total variation distance between the distribution in the observed data, vs the model distribution (0.75, 0.25)



When poll is active, respond at [pollev.com/dsfa](https://pollev.com/dsfa)

Text **DSFA** to **22333** once to join

# If we choose the proportion of purple flowers, and if the alternative hypothesis is true, what will happen?

The test statistic will be larger than the predicted value

The test statistic will be smaller than the test statistic

It could be either smaller or larger



# Choosing a T

---

For the baby birth weight, the following would be reasonable test statistics (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 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.
-

When poll is active, respond at [pollev.com/dsfa](https://pollev.com/dsfa)

Text **DSFA** to **22333** once to join

# For the baby birth weight experiment, which of the following would be a reasonable test statistic?

Average birth weight of all babies

The proportion of babies whose mother smoked

The average birth weight of babies of smokers minus the average birth weight of babies of non-smokers

The absolute value of the previous difference

None of the above





# 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

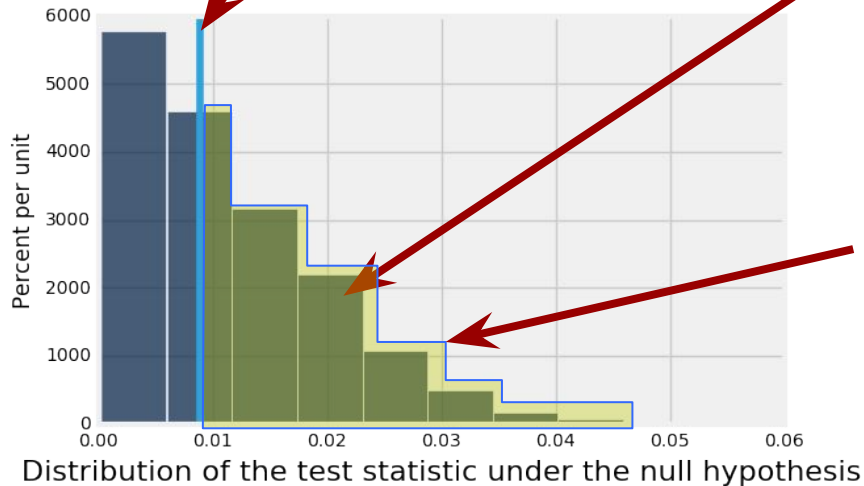
---

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.”
-

# Sir Ronald Fisher, 1890-1962

---



*"We have the duty of formulating, of summarizing, and of communicating our conclusions, in intelligible form, in recognition of the right of other free minds to utilize them in making their own decisions."*

*Ronald Fisher*

---



# 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 ...”

---

# Can the Conclusion be Wrong?

---

**Yes.**

	<b>Null is true</b>	<b>Alternative is true</b>
<b>Test rejects the null</b>		
<b>Test doesn't reject the null</b>		

(Demo)

---

# An Error Probability

---

- The cutoff for the P-value is an error probability.
  - If:
    - your **cutoff is 5%**
    - 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**.
-

# Assess this:

---

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

— Letter to *Nature*, 1994

---