Testing Hypotheses


Are these fingernail lengths extreme?

Testing Hypotheses

1. Hypotheses and Testing Philosophies
2. Introducing the Frequentist P-Value
3. Test Statistics and All That
4. You're Doing it Wrong (or are you)

What is a Hypothesis? (in a statistical sense)

- **Point**: Mean Fingernail Length == 5cm
- **Model**: If I increase Kelp, Biodiversity increases
- **Explanatory**: Fertilizers explain more variation in the data than Pesticides

Three Flavors of Statistical Inference

1. **Frequentist:** What's the chance we observed something like the data given our hypothesis?

2. **Likelihoodist:** What's the likelihood of our hypothesis relative to others?

3. **Bayesian:** How much do we believe our hypothesis given our data?

Frequentist Inference: the probability of observing this data, or more extreme data, given that a hypothesis is true

P-value
or
Confidence Interval

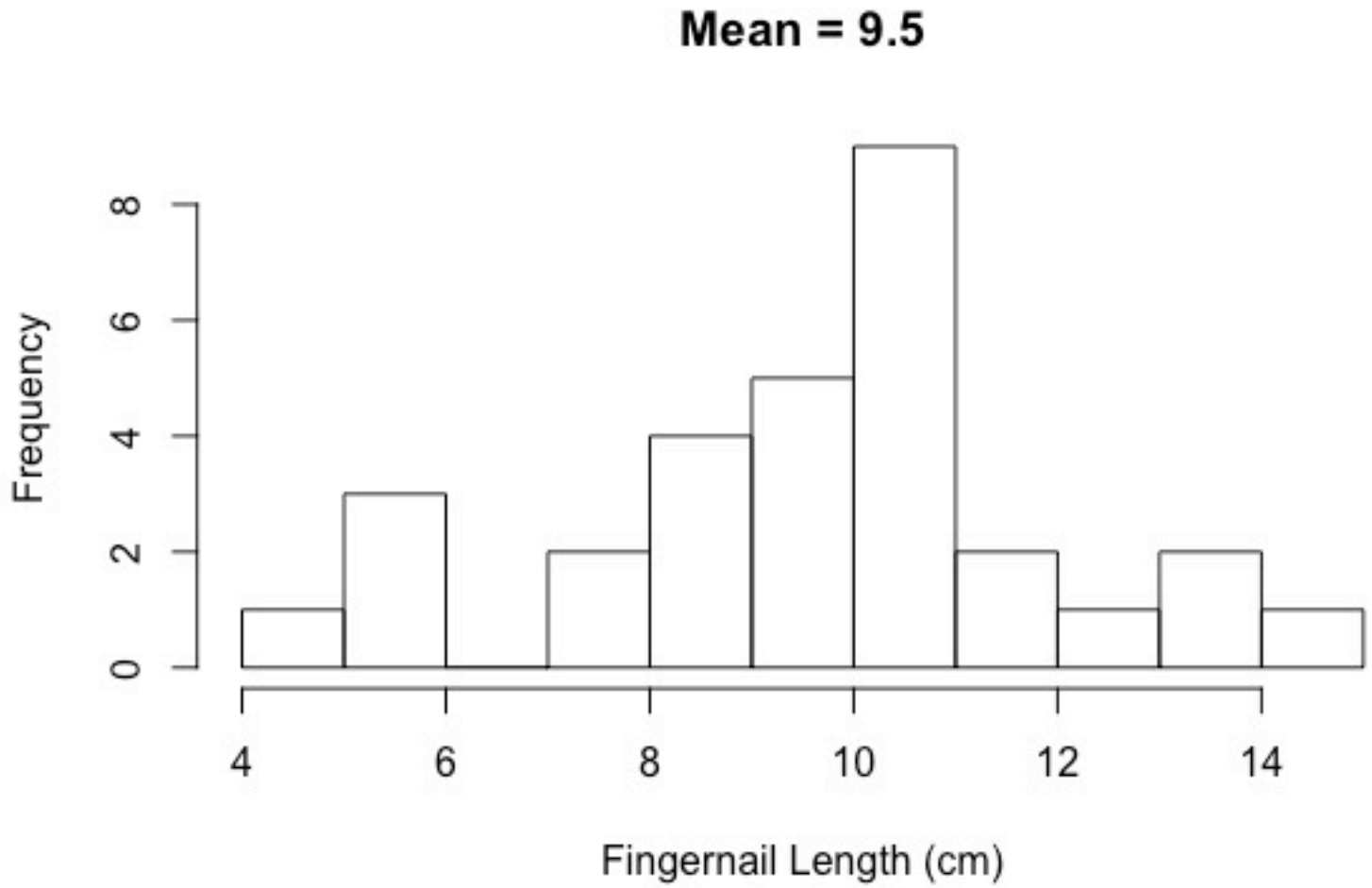
Our Hypothesis

‘Extreme’ fingernail lengths are not different from standard fingernail lengths

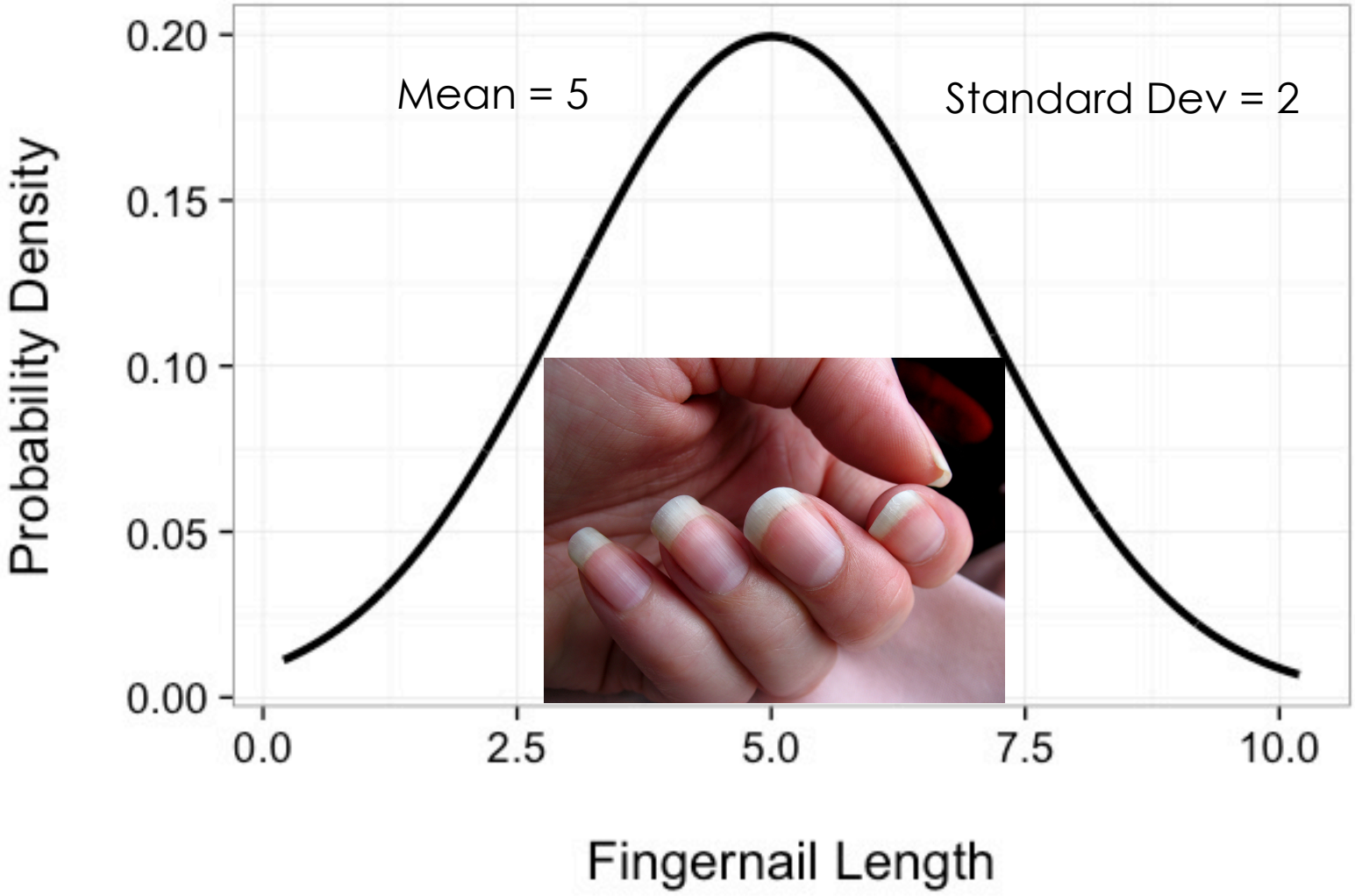
Testing Hypotheses

1. Hypotheses and Testing Philosophies
2. Introducing the Frequentist P-Value
3. Test Statistics and All That
4. You're Doing it Wrong (or are you)

Our Sample



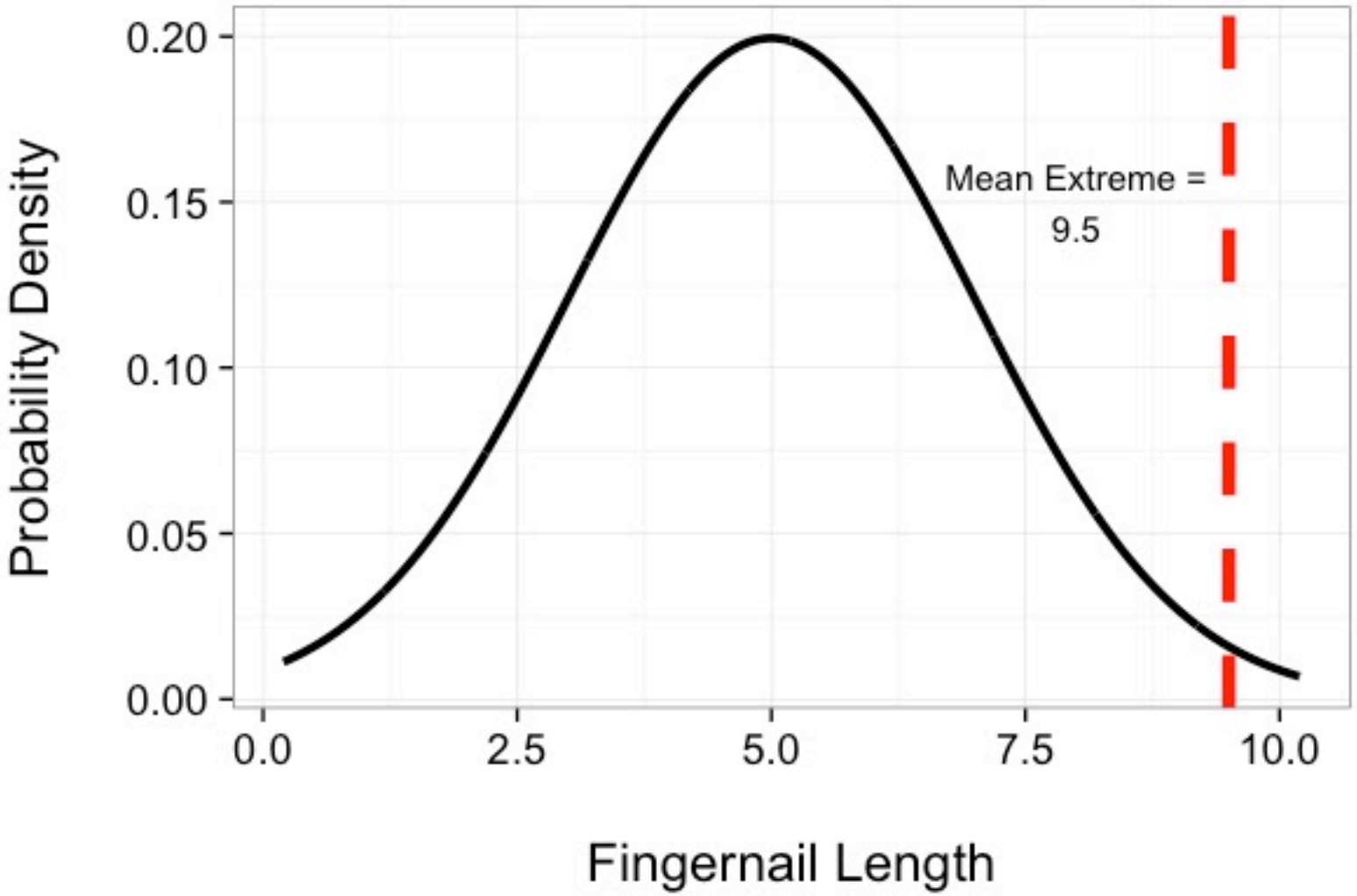
Assume a Distribution of Fingernail Sizes



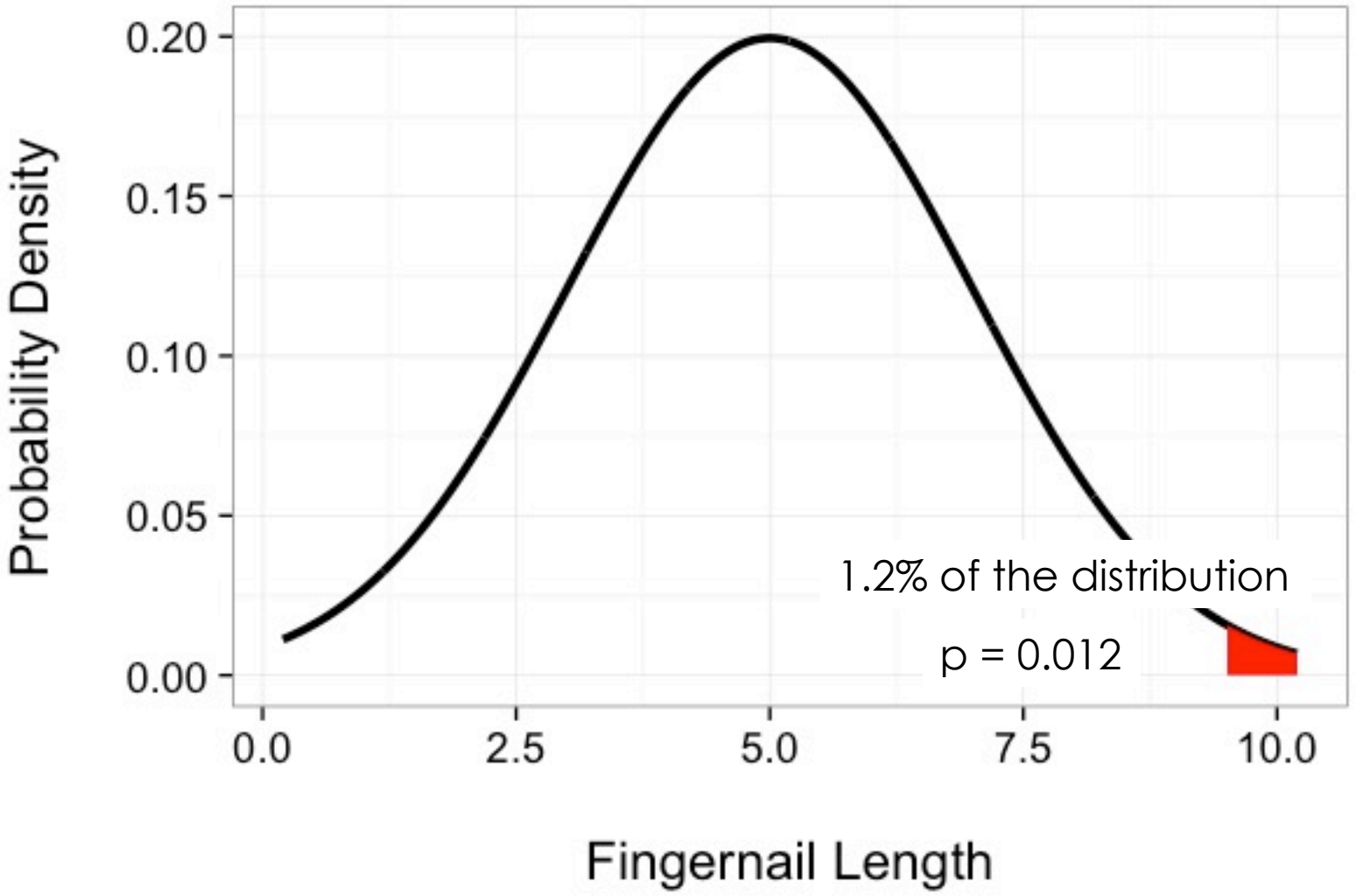
Our Hypothesis

‘Extreme’ fingernail lengths are from a distribution with mean of 5 and SD of 2.

Is our size extreme?



What about even more extreme values?




Extreeeeeeme!

P-Values: Fisher

R.A. Fisher

The probability of observing a value or more extreme value given a specified hypothesis.

Testing Hypotheses

1. Hypotheses and Testing Philosophies
2. Introducing the Frequentist P-Value
3. Test Statistics and All That
4. You're Doing it Wrong (or are you)

Null Hypothesis Testing

- Frequentist testing of whether ***something*** is different from a null expectation

- Example uses:
  - An estimate is not different from 0
  - The difference between two groups is not different from zero
  - A predictor provides no additional explanation of patterns in the data

Test Statistics: Making the World Sensible (and Null)

1. Create a null distribution

2. Use your data to calculate a test score

3. Calculate the p-value for your data in the context of that null distribution
   - $P(D|H_0)$:

Testing Fingernails Against a Normal

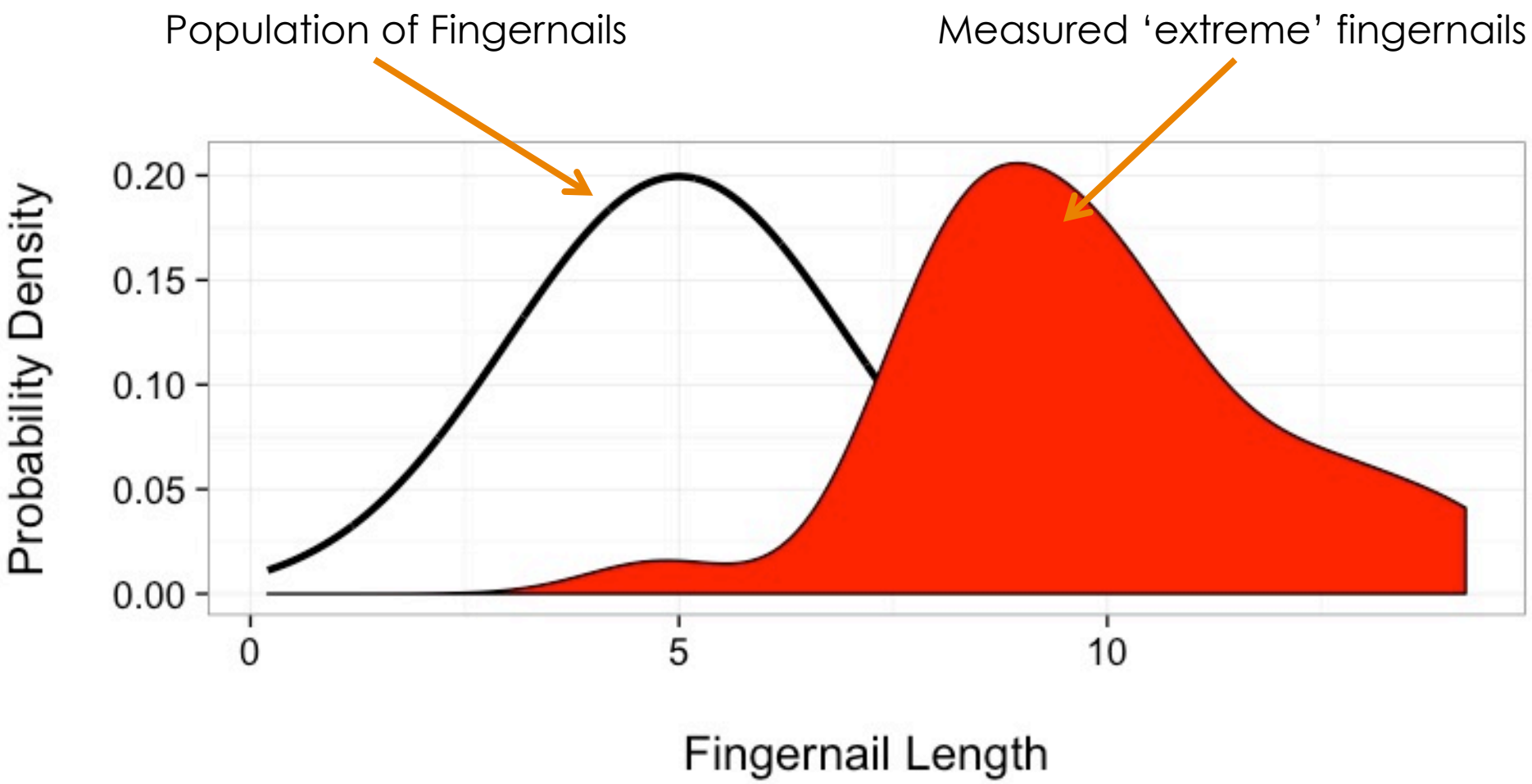
H_0 = Mean of sample is not different from the rest of the population

1. Assume a normal distribution with an SD of 2

2. Calculate the difference between the mean of our sample and 5 = 4.5

3. Assess the p-value of 4.5 against the normal distribution with mean 0

But…this is my data



The Arrival of a Test Statistic

H_0 = Mean of sample is not different from the rest of the population

1. Assume a normal distribution with an SD of 1 (standard normal curve)

2. Calculate the difference between the mean of our sample and 5 = 4.5

3. Divide by the Standard Deviation of the population and the square root of the sample size (assumed SE of a population Sample) – z score!

- Z-Score

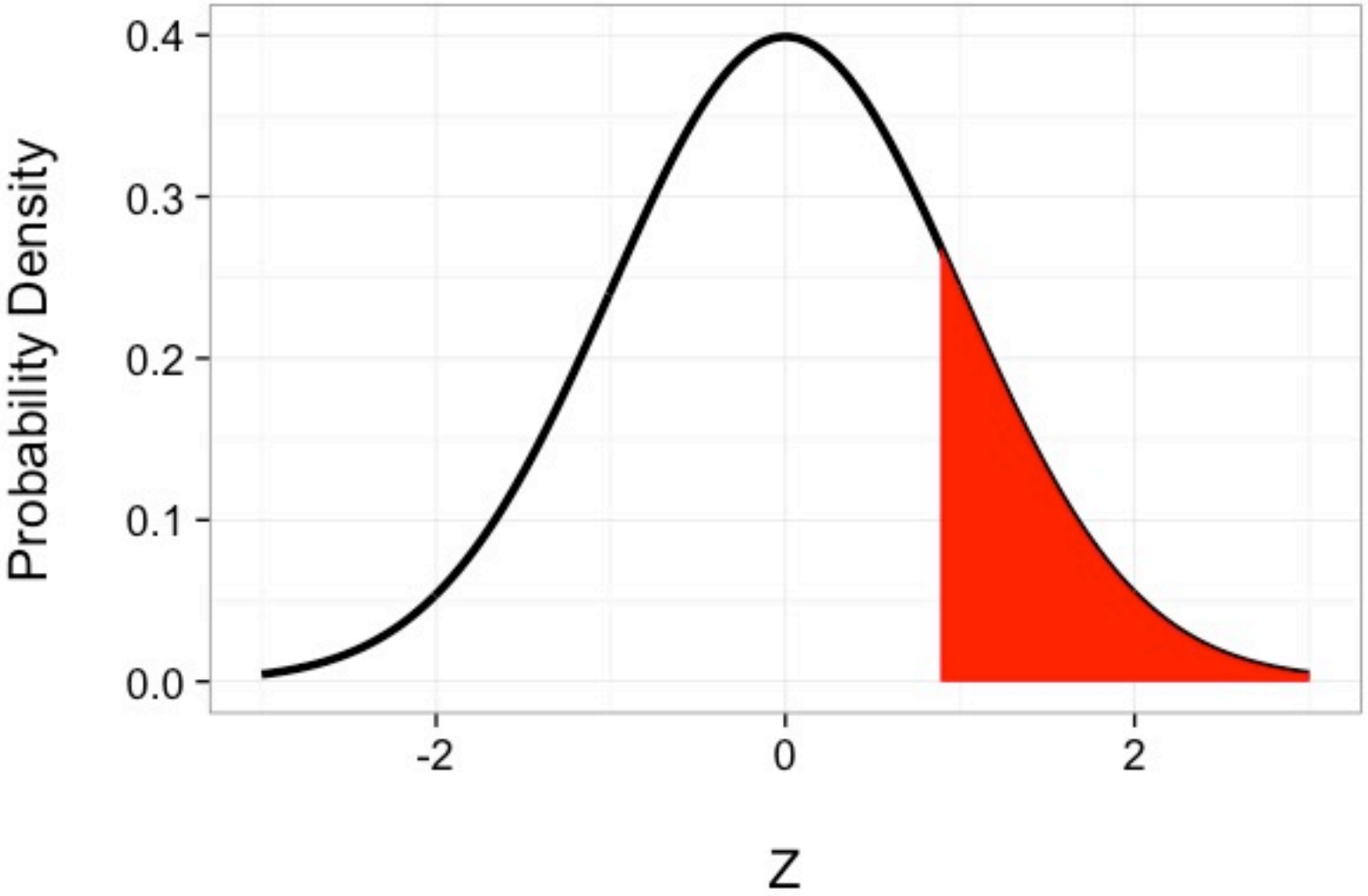
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

z = 4.5 / (2/sqrt(30))

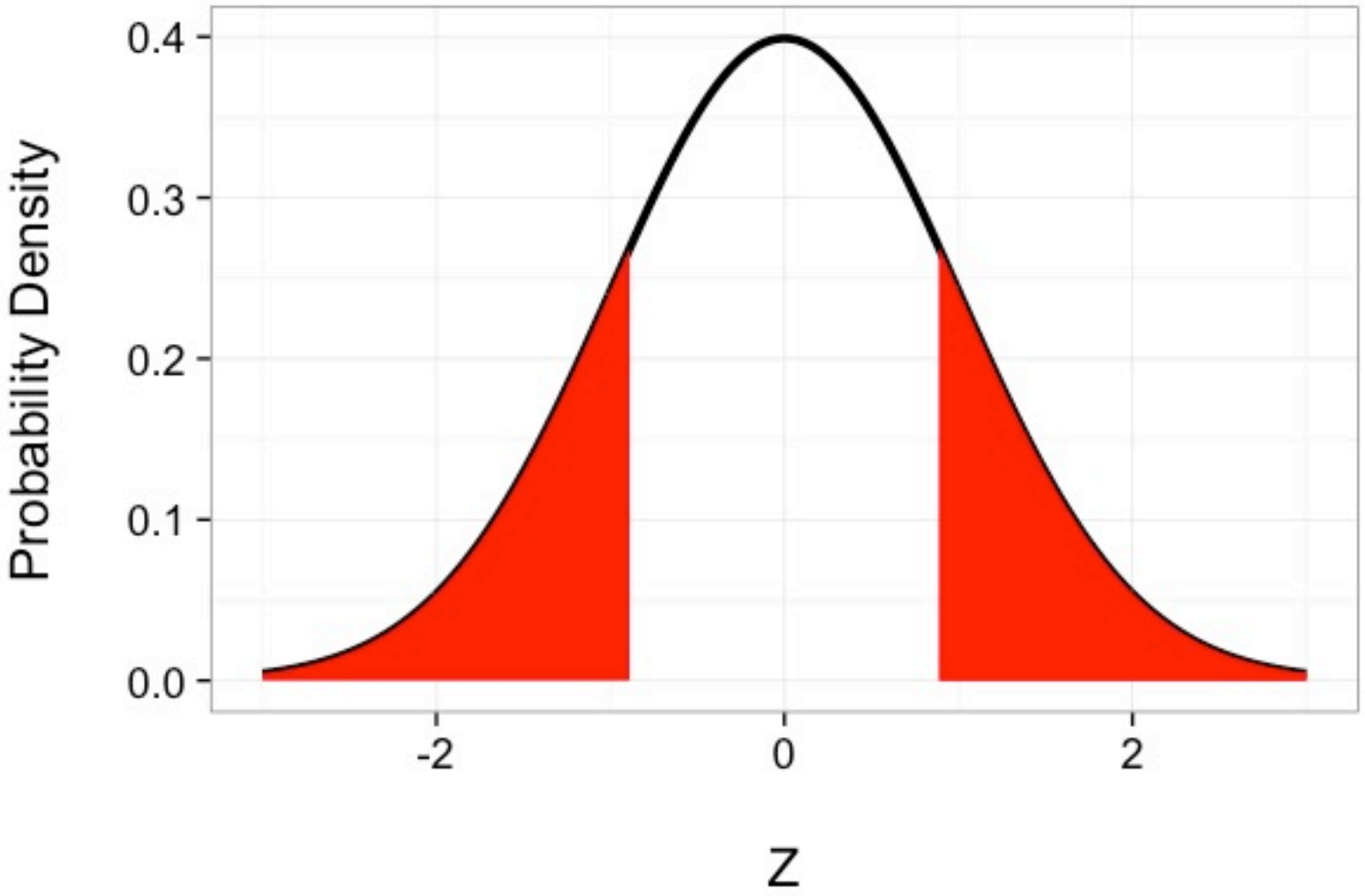
= 12.32

p(12.32|μ=0) < 0.0001

Z-Test



There Are Two Tails




Yes. These are super odd.

What do you do with a p-value?

- P-values give the evidence for the probability of your data, or more extreme data, given a hypothesis

- As the scientist, ***you*** decide whether it is grounds for rejecting a hypothesis

- In a frequentist framework, you can only reject a hypothesis – never `accept`

Testing Hypotheses

1. Hypotheses and Testing Philosophies
2. Introducing the Frequentist P-Value
3. Test Statistics and All That
4. You're Doing it Wrong (or are you)

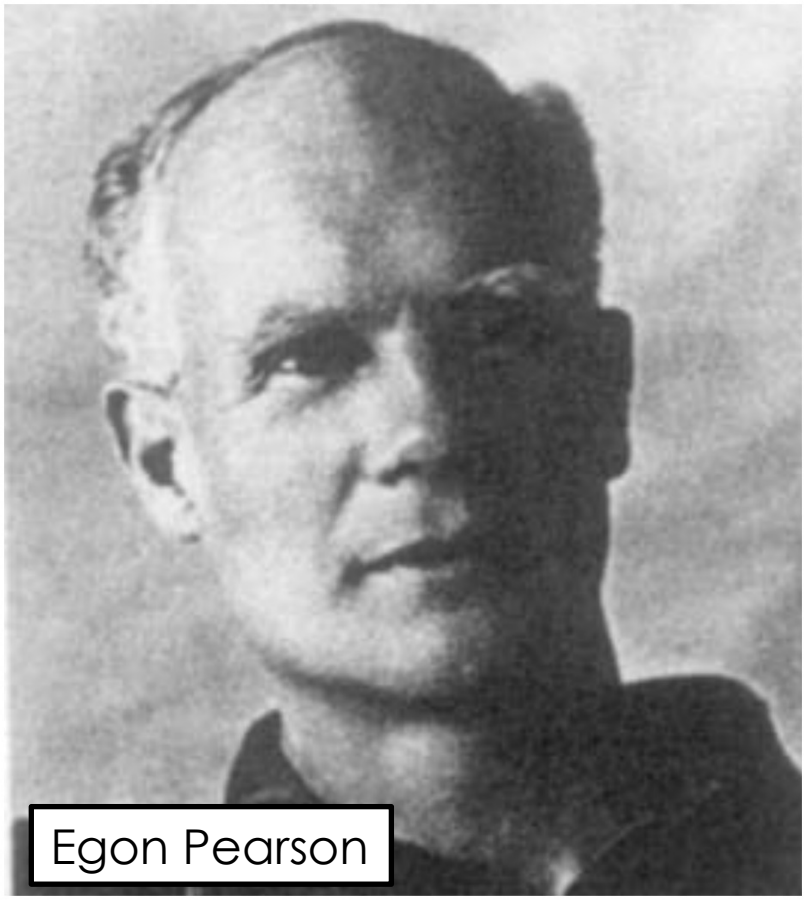
How We Can Screw This Up

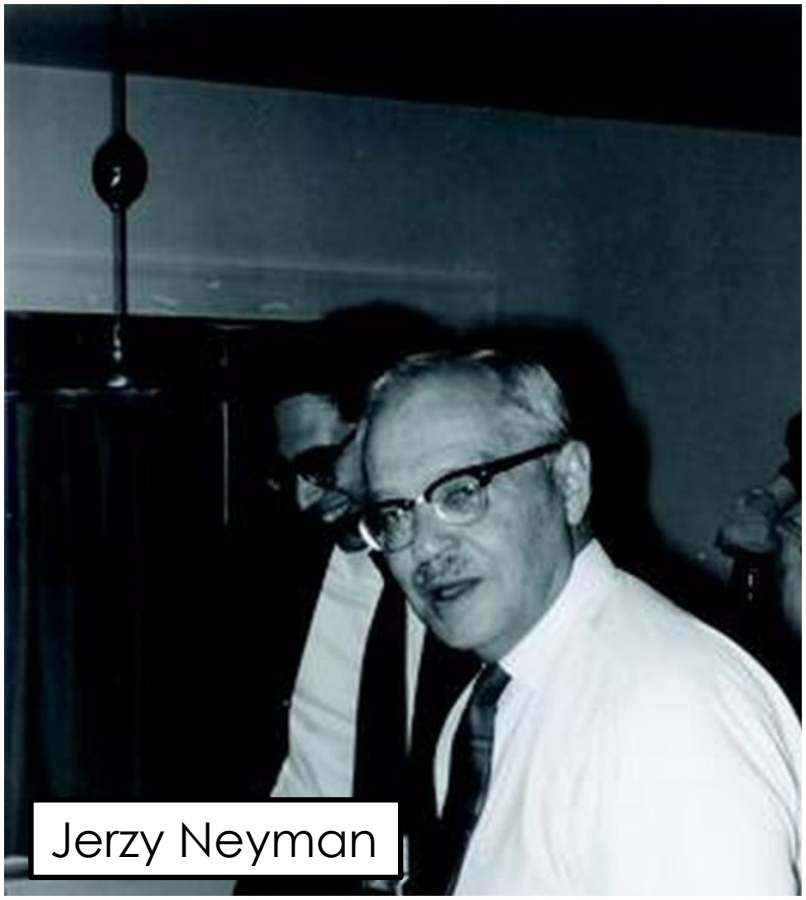
		Is the Hypothesis True or False?	
		True	False
Test Result Against Hypothesis	Hypothesis Not Rejected	☺	Type I error
	Hypothesis Rejected	Type II error	☺

Probability of a type I error = α

Probability of a type II error = β

Null Hypothesis Significance Testing


Egon Pearson


Jerzy Neyman

Problem: What is an acceptable α?

Answer: 0.05. You have a 1 in 20 chance of committing a type I error

Problems with NHST

- A realistic alpha can depend on your study design
  - E.g., Large sample size = lower p value
- Ignores β
  - Tradeoff between α and β
- Conflation of scientific significance and statistical significance
  - File-drawer effect
- We are human
  - If $p \leq 0.05$ makes your career, you will do a lot to obtain it!