

Introduction to hypothesis tests

Overview

Questions about worksheet 7

Last quick review of confidence intervals

A brief note on probability

Introduction to hypothesis tests

Questions about worksheet 7?

Do we feel confident about constructing confidence intervals using the bootstrap?

There will be a few more opportunities to practice computing CIs through the rest of the semester 😊



Questions about worksheet 7?

Exercise 2: Comparing Mazda and Toyota prices

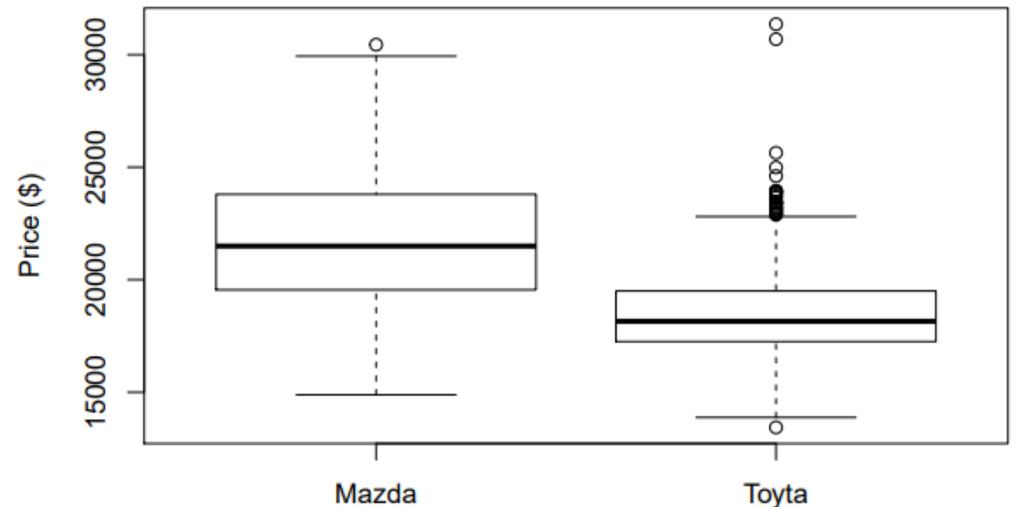
What is a good first thing to do when analyzing data?

Confidence intervals:

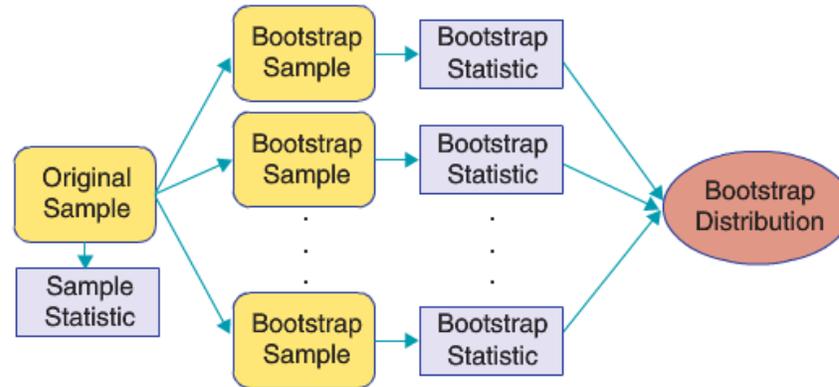
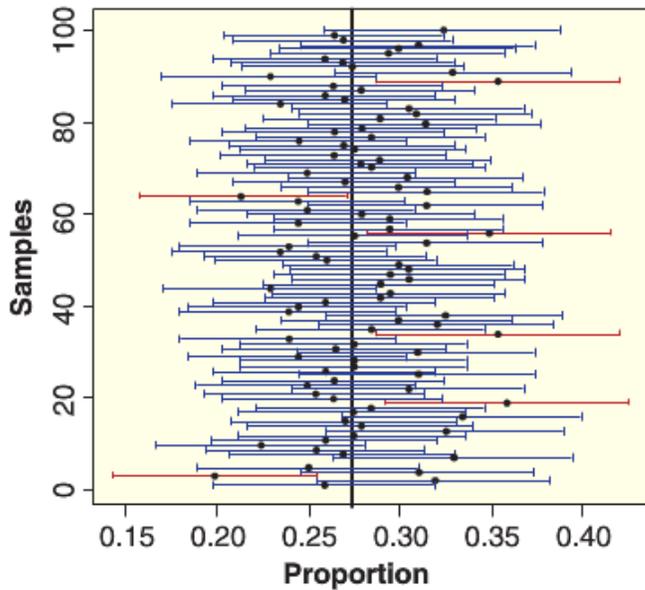
Mazda price CI = [\$21,618 \$21,899]

Toyota price CI = [\$18,415 \$18,610]

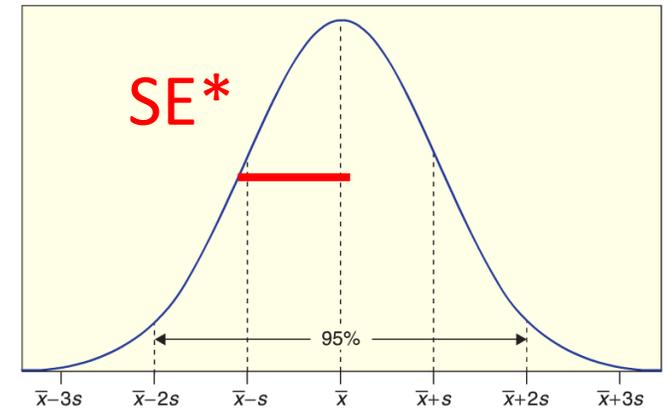
Is it likely that $\mu_{\text{mazda}} = \mu_{\text{Toyota}}$?



Are we feeling confident about confidence intervals?



Bootstrap distribution



95% confidence interval: $\bar{x} \pm 2 \cdot \hat{SE}$

Bootstrap distribution in R for μ

```
my_sample <- c(3, 2, 4, 2, 5, 9, 10, 2, 4, 3) # n = 10 points here  
my_stat <- mean(my_sample)
```

```
bootstrap_dist <- NULL
```

```
for (i in 1:10000) {
```

```
  curr_boot <- sample(my_sample, 10, replace = TRUE)  
  bootstrap_dist[i] <- mean(curr_boot)
```

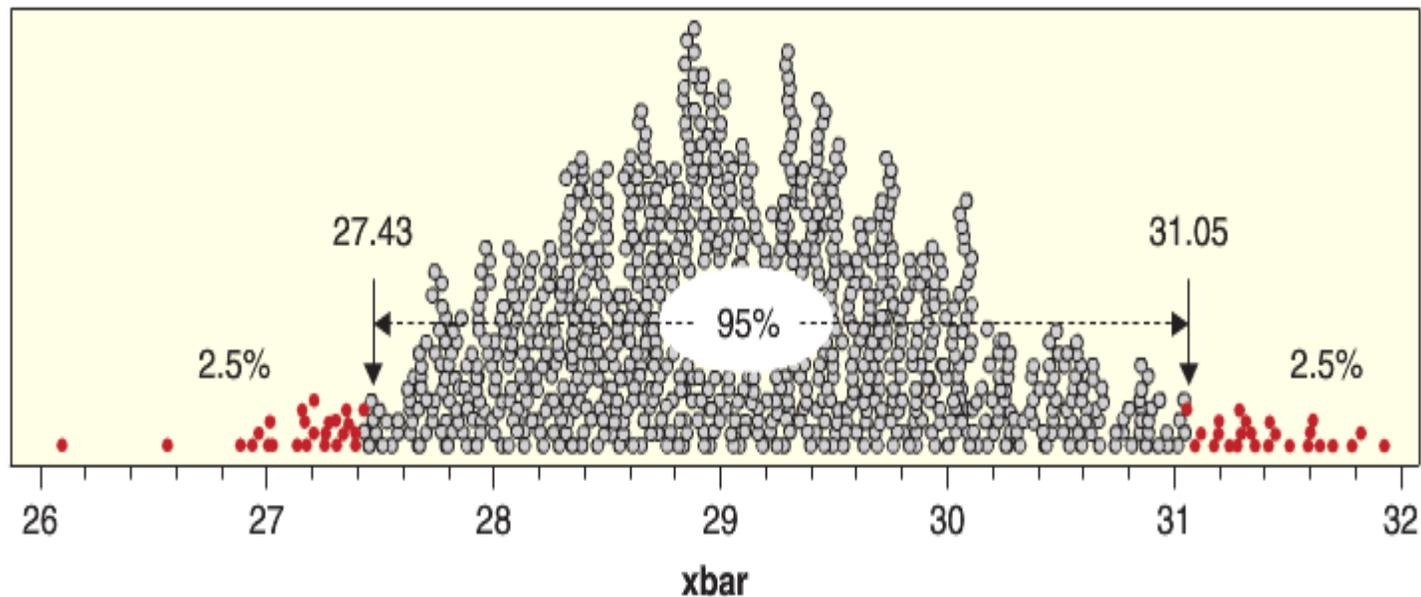
```
}
```

```
SE_boot <- sd(bootstrap_dist)
```

```
CI <- c(my_stat - 2 * SE_boot, my_stat + 2 * SE_boot)
```

What if the bootstrap distribution is not normal?

If the bootstrap distribution is approximately symmetric, we can use percentiles in the bootstrap distribution to an interval that matches the desired confidence level

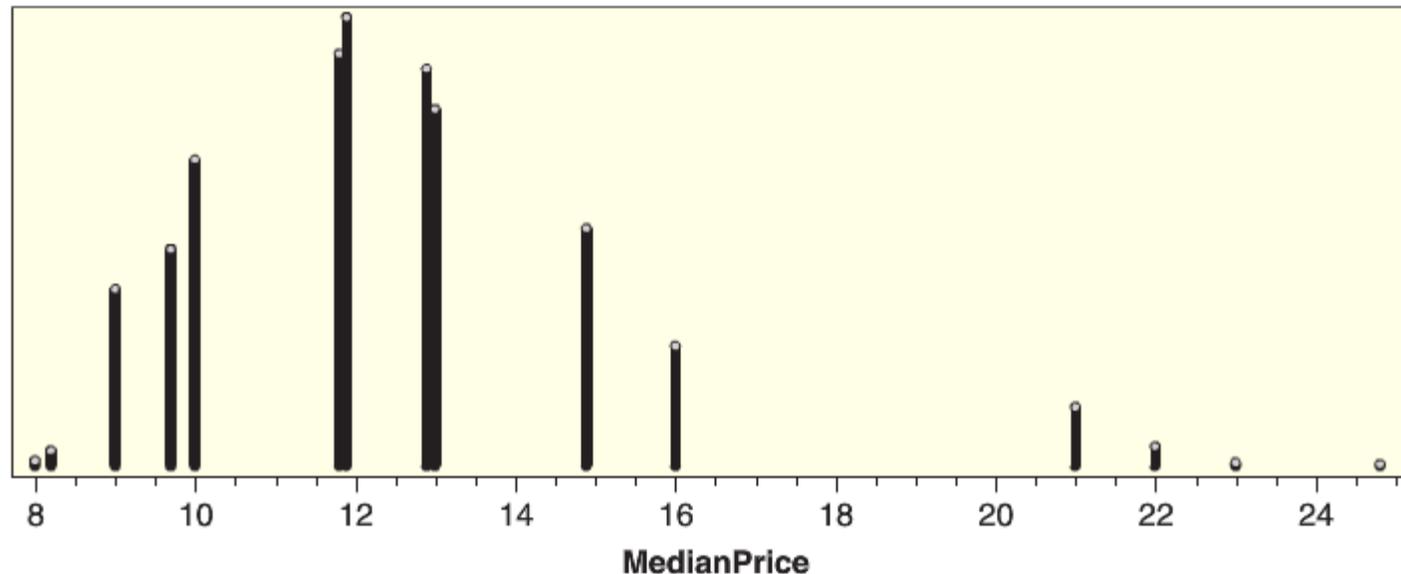


```
quantile(bootstrap_dist, c(.025, .975))
```

Caution: the bootstrap does not always work

Always look at the bootstrap distribution, if it is poorly behaved (e.g., heavily skewed, has isolated clumps of values, etc.), you should not trust the intervals it produces.

Median prices of Mustangs



A quick note on probability

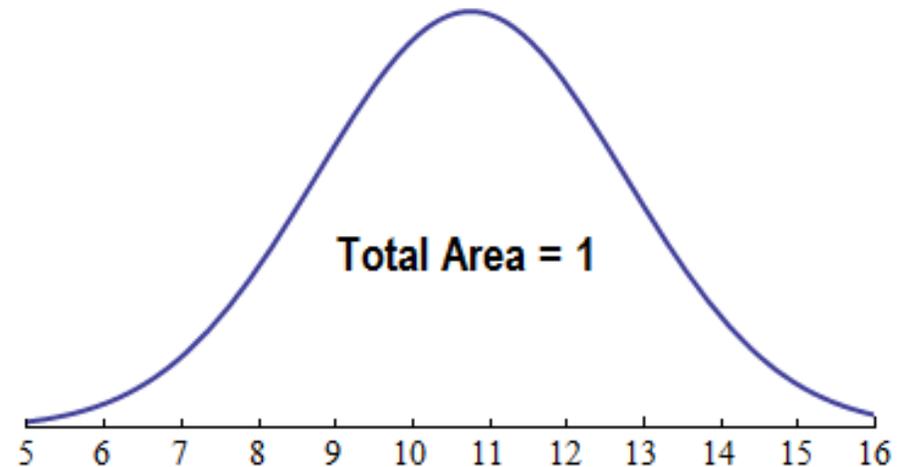
Probability is a way of measuring the likelihood that an event will occur

Probability models assigns a number between 0 and 1 to the outcome of an event occurring

We can use a probability model to calculate the probability of an event

For example:

- $\Pr(X < 11) = 0.55$
- $\Pr(X > 20) = 0$



Statistical tests (hypothesis test)

A **statistical test** uses data from a sample to assess a claim about a population

Example 1: we might make the claim that Trump's approval rating for all US citizens is 45%

How can we write this using symbols?

- $\pi = .45$

Statistical tests (hypothesis test)

A **statistical test** uses data from a sample to assess a claim about a population

Example 2: we might make the claim that the average height of a baseball player is 72 inches

How can we write this using symbols?

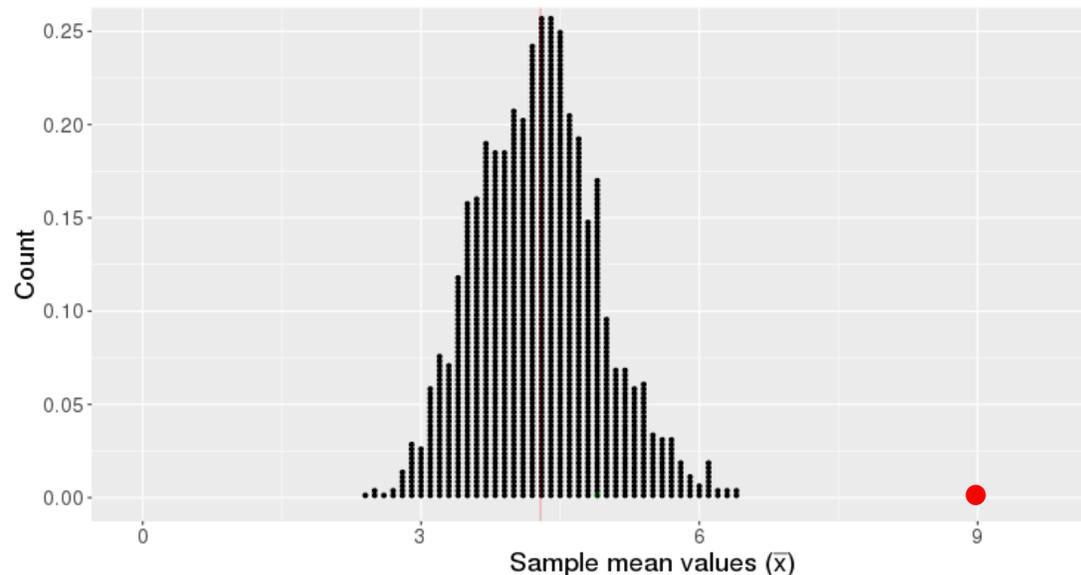
- $\mu = 72$

Basic hypothesis test logic

We start with a claim about a population parameter

- E.g., $\mu = 4$ 

This claim implies we should get a certain distribution of statistics



If our observed statistic is highly unlikely, we reject the claim

Can dolphins communicate?

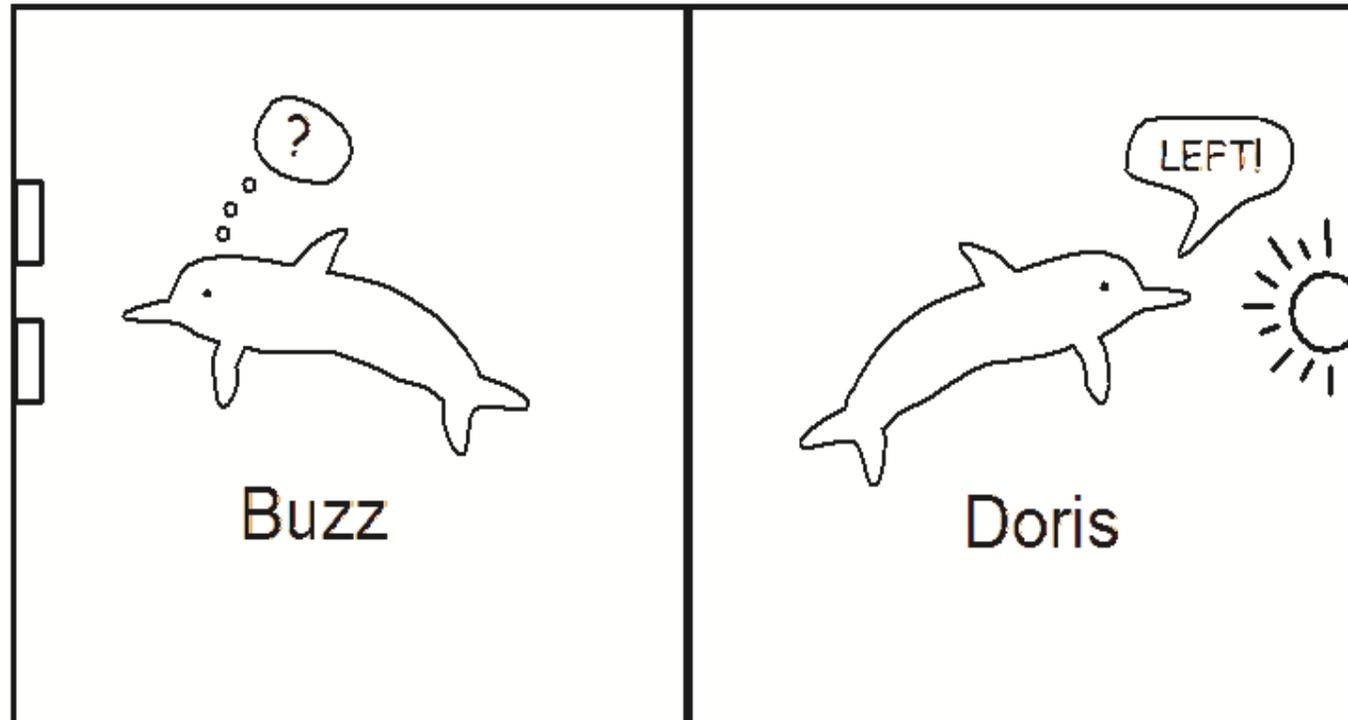
Dr. Jarvis Bastian in the 1960's wanted to know whether dolphins are capable of abstract communication

He used an old headlight to communicate with two dolphins (Doris and Buzz)

- Stead light = push button on right to get food
- Flashing = push button on the left to get food



A canvas was then put in the middle of the pool with Doris on one side and buzz on the other



Buzz got 15 out of 16 trials correct

Write down/discuss answers to the following questions

1. What are the cases here?
2. What is the variable of interest and is it categorical or quantitative?
3. What is the observed statistic - and what symbols should we use to denote it?
4. What is the population parameter we are trying to estimate - and what symbol should we use to denote it?
5. Do you think the results are due to chance?
 - i.e., how many correct answers do you think Buzz would have gotten if he was guessing?
6. Are dolphins capable of abstract communication?

The dolphin communication study

If Buzz was just guessing, what would we expect the value of the parameter to be?

$$\pi = 0.5$$

If Buzz was not guessing, what would we expect the value of the parameter to be?

$$\pi > 0.5$$

Chance models

How can we assess whether 15 out of 16 correct trials ($\hat{p} = .975$) is beyond what we would expect by chance?

If buzz was guessing we can model his guesses as a coin flip:

Heads = correct guess

Tails = incorrect guess

Let's flip a coin 16 times and see how many times we get 15 heads

Chance models

To really be sure, how many repetitions of flipping a coin 16 times should we do?

Any ideas how to do this?



How to estimate p-values in R?

We can simulate coin flipping using the `rbinom()` function

```
flip_simulations <- rbinom(num_sims, size, prob)
```

num_sims: the number of simulations run

- Typically we do around 10,000 repeats

size: the number of trials on each simulation

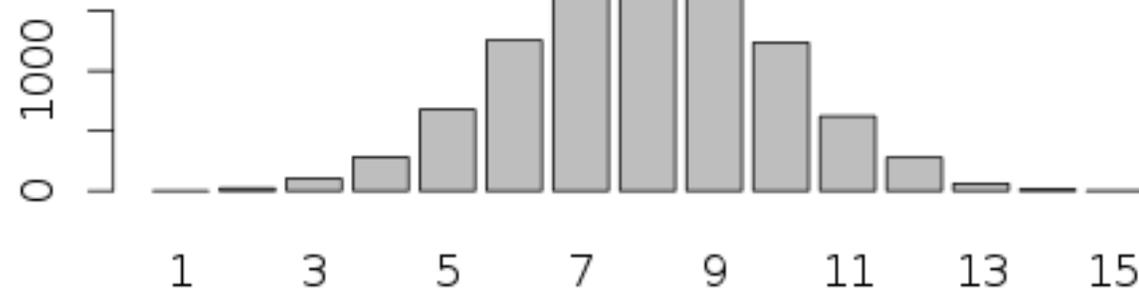
- 16 for Doris/Buzz

prob: the probability of success on each trial

- .5 Doris/Buzz

Simulating Flipping 16 coins 10,000

0	0
1	1
2	22
3	105
4	283
5	679
6	1257
7	1786
8	1920
9	1726
10	1238
11	623
12	279
13	63
14	15
15	3
16	0



Is it likely that Buzz was guessing?

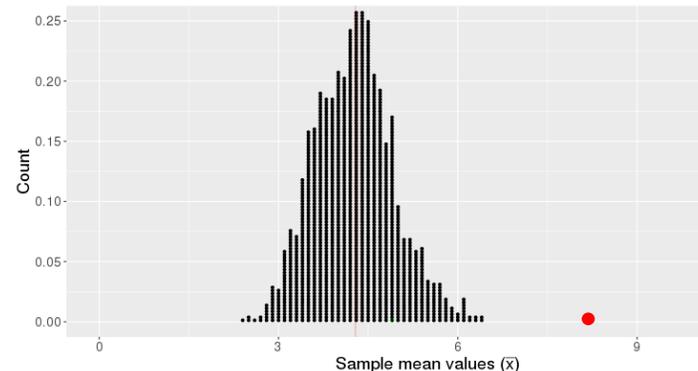
Are dolphins capable of abstract communication?

Terminology

Null Hypothesis (~~H_0~~): Claim that there is no effect or no difference

Alternative Hypothesis (H_a): Claim for which we seek significant evidence

The alternative hypothesis is established by observing evidence that inconsistent with the null hypothesis



The dolphin communication study

1. What is the null hypothesis?

2. We can write this in terms of the population parameter as:

$$H_0: \pi = 0.5$$

3. What is the alternative hypothesis?

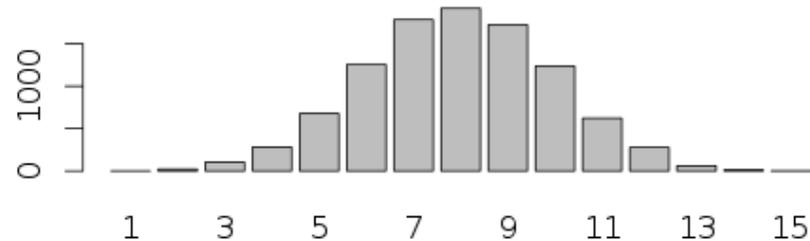
$$H_A: \pi > 0.5$$

Null Distribution

A **null distribution** is the distribution of statistics one would expect if the null hypothesis (H_0) was true

i.e., the null distribution is the statistics one would expect to get if nothing interesting was happening

- note: the book calls this the randomization distribution



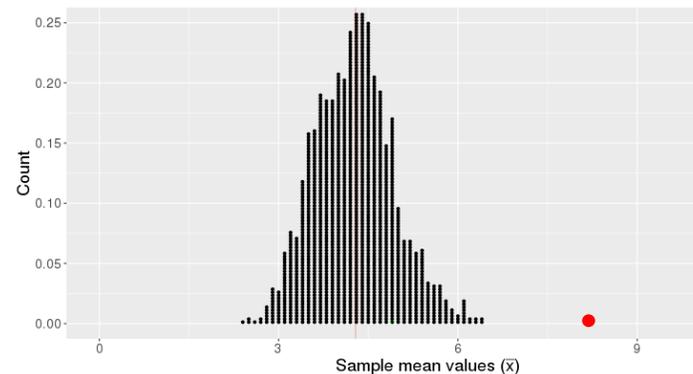
P-values

A **p-value** is the probability, of obtaining a statistic as as (or more) extreme than the observed sample *if the null hypothesis was true*

- i.e., the probability that we would get a statistic as extreme as our observed statistic from the null distribution

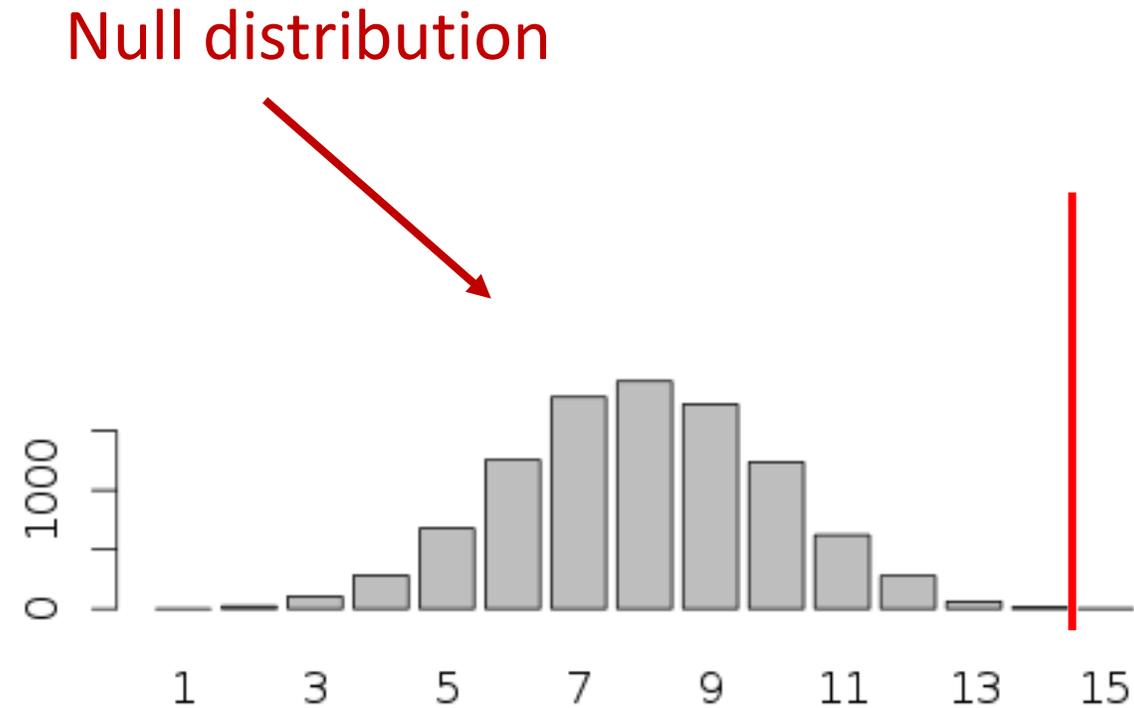
$$\Pr(\text{STAT} \geq \text{observed statistic} \mid H_0 = \text{True})$$

The smaller the p-value, the stronger the statistic evidence is against the null hypothesis



Buzz and Doris example

0	0
1	1
2	22
3	105
4	283
5	679
6	1257
7	1786
8	1920
9	1726
10	1238
11	623
12	279
13	63
14	15
15	3
16	0



$$p\text{-value} = 3/10000 = .0003$$

Statistical significance

When our observed sample statistic is unlikely to come from the null distribution, we say the sample results are **statistically significant**

- i.e., we have a small p-value

‘Statistically significant’ results mean we have strong evidence against H_0 in favor of H_a

Key steps hypothesis testing

1. State the null hypothesis... and the alternative hypothesis

- Buzz is just guessing so the results are due to chance: $H_0: \pi = 0.5$
- Buzz is getting more correct results than expected by chance: $H_A: \pi > 0.5$

2. Calculate the observed statistic

- Buzz got 15 out of 16 guesses correct, or $\hat{p} = .973$

3. Create a null distribution that is consistent with the null hypothesis

- i.e., what statistics would we expect if Buzz was just guessing

4. Examine how likely the observed statistic is to come from the null distribution

- What is the probability that the dolphins would guess 15 or more correct?
- i.e., what is the p-value

5. Make a judgement

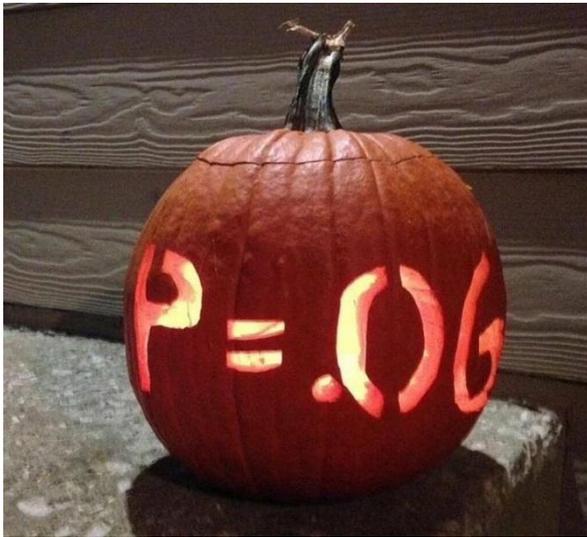
- If we have a small p-value, this means that $\pi = .5$ is unlikely and so $\pi > .5$
- i.e., we say our results are 'statistically significant'



"Matthews ... we're getting another one of those strange 'aw blah es span yol' sounds."

Next class – hypothesis tests continued

Halloween edition...



Read Lock5 chapters 4.1, 4.2

Practice problems Lock5 first edition: 4.30-4.37, 4.45

Practice problems Lock5 second edition: 4.31-4.37, 4.53