Statistical inference on proportions

Outline for today

Review of hypothesis tests for a proportion (whether Paul is psychic)

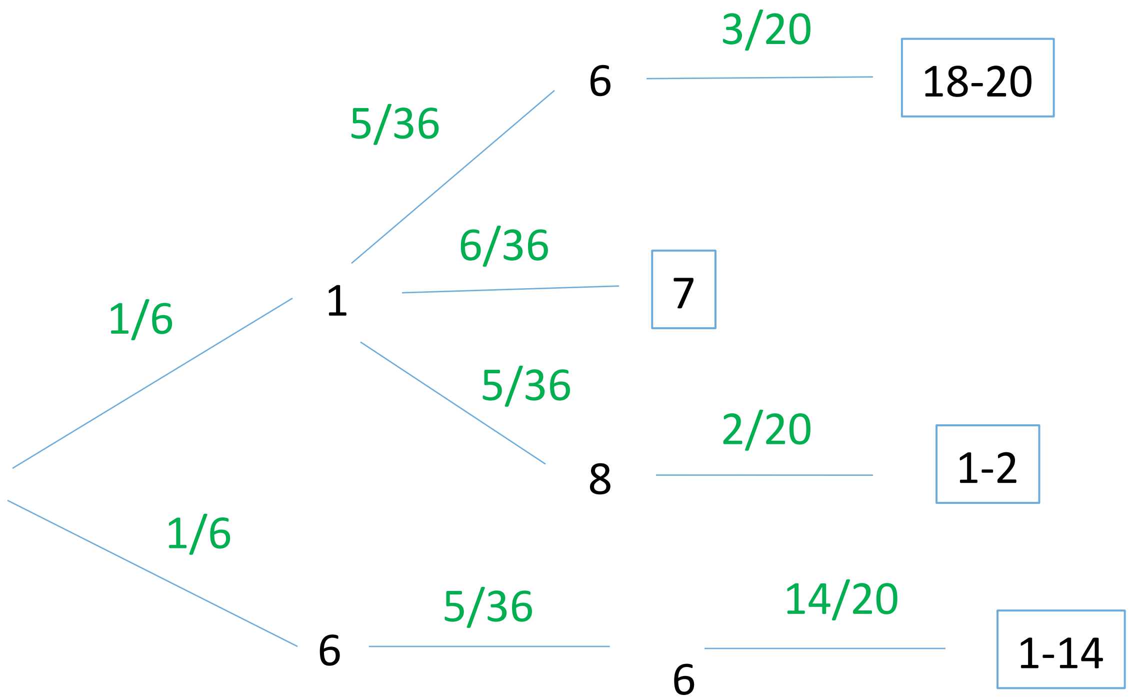
More hypothesis tests for a single proportion

Worksheet 8

Better know a player

Doc Ellis

Tree diagrams: you need to calculate the final probability of the event

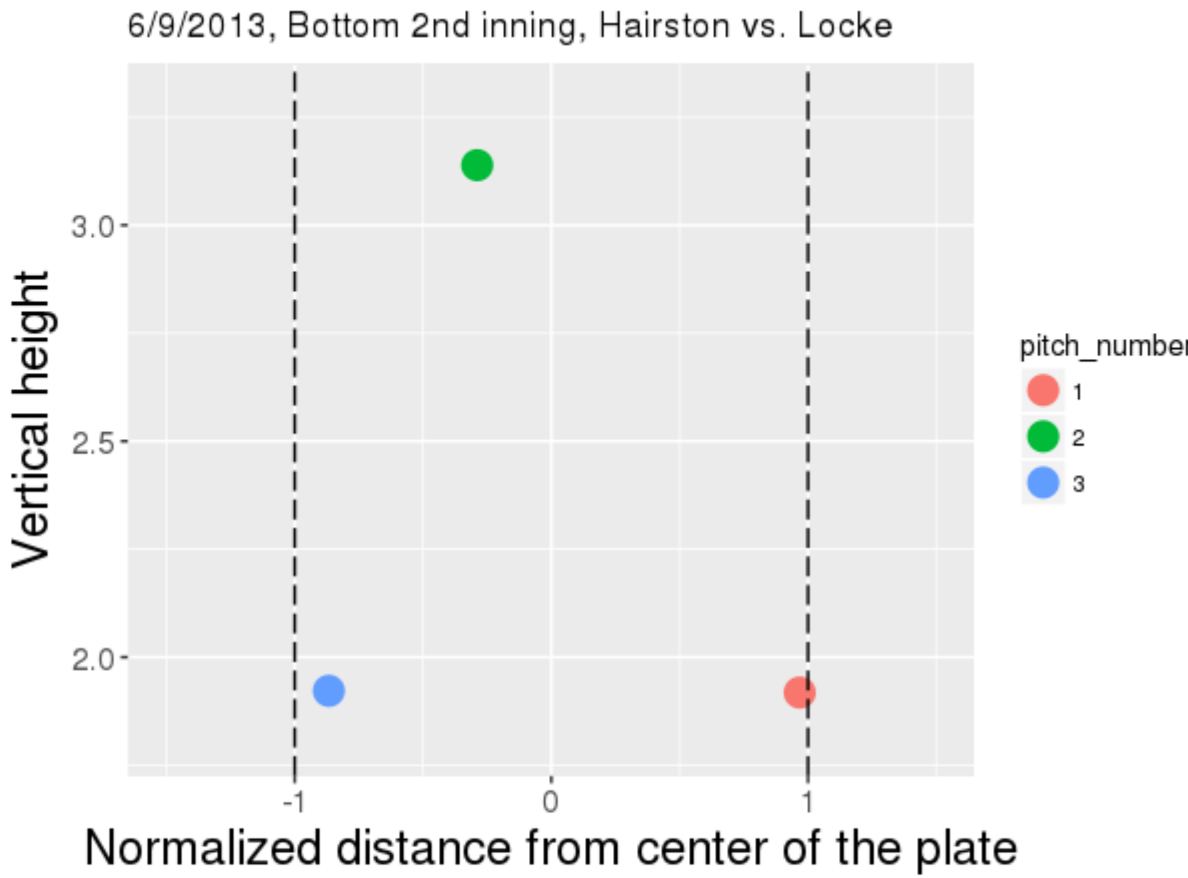


$$1/6 \cdot (5/36 \cdot 3/20 + 6/36 + 5/36 \cdot 2/20) + 1/6 \cdot 5/36 \cdot 14/20 = .04977$$

Big Data Baseball – pg 152

Sawchik claims Locke strikes out Hairston on a pitch that was called a strike even though it was outside the strike zone

Was it really a ball?



Big Data Baseball – pg 152

```
library("pitchRx")

pirates_cubs  <- scrape(game.ids =
        "gid_2013_06_09_pitmlb_chnmlb_1")

pitches <- pirates_cubs$pitch

# px and pz show the location of the pitch
View(pitches)
```

Statistical inference

Statistical inference: use sample of data to deduce properties of an underlying population, or stochastic process

- E.g., looking at a player's **performance** to estimate their **ability**

Hit, Out, Hit, Out, Out, Out, … Estimate π_{hit}

Hypothesis tests: Paul the Octopus

In the 2010 World Cup, Paul the Octopus (in a German aquarium) became famous for correctly predicting 11 out of 13 soccer games

Question: is Paul psychic?

Paul the Octopus: second approach

Question: If Paul was psychic, what proportion of games would he have guessed correctly?

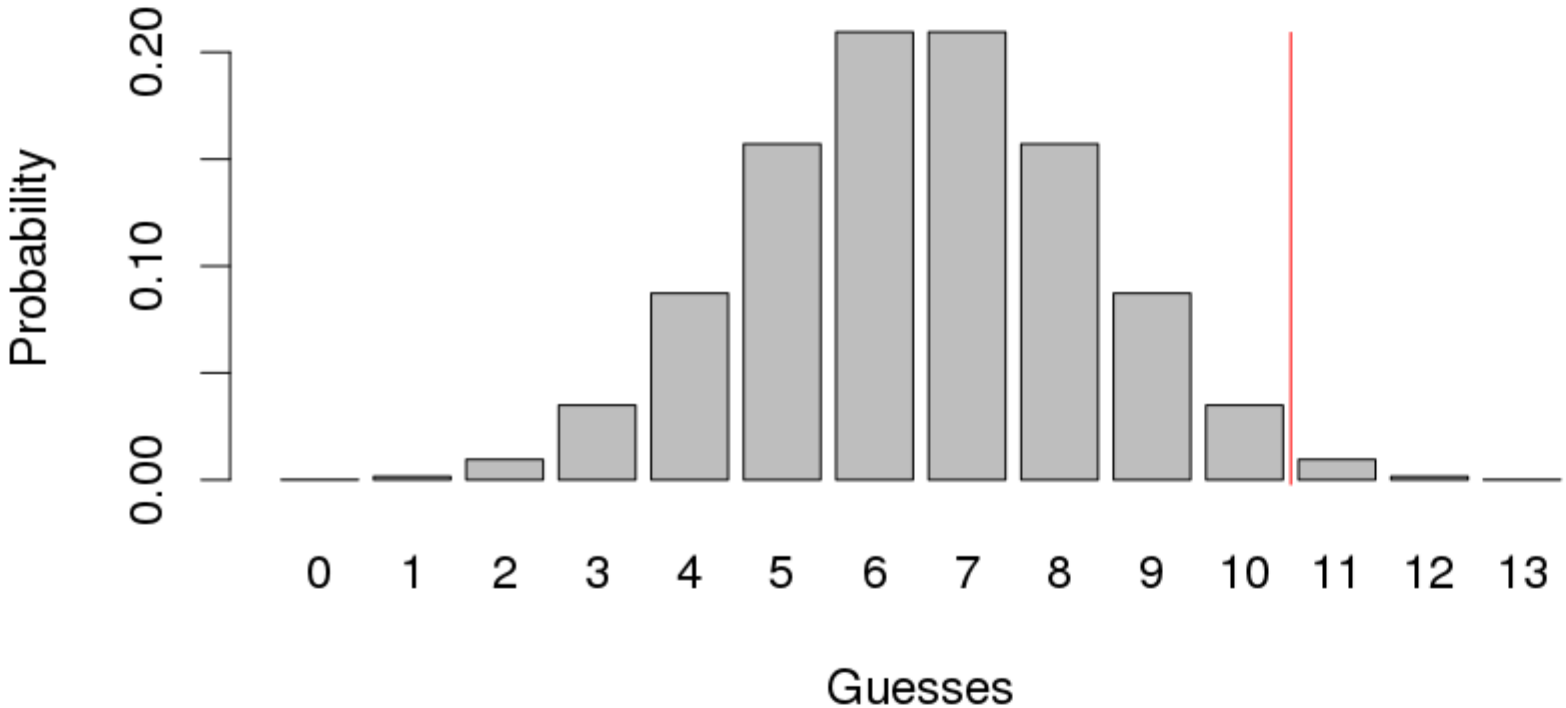
- Answer: $\pi = .5$

Question: How can we calculate the probability he would guess 11 games correctly?

- We could use the binomial distribution to the probability of getting 11 or more heads

- We can do this in R using:
  - `> sum(dbinom(11:13, 13, .5))`   # sum Pr(X = 11) + Pr(X = 12) + Pr(X = 13)
  - `> 1 – pbinom(10, 13, .5)`   # equivalently: 1 – Pr(X ≤ 10)

Paul the Octopus



Using the binomial distribution we have:

$\Pr(X \geq 11;\ n = 13,\ \pi = .5) \quad = \quad .0112$

Paul the Octopus

Question: If Paul was psychic, what proportion of games would we expect him to guess correctly?

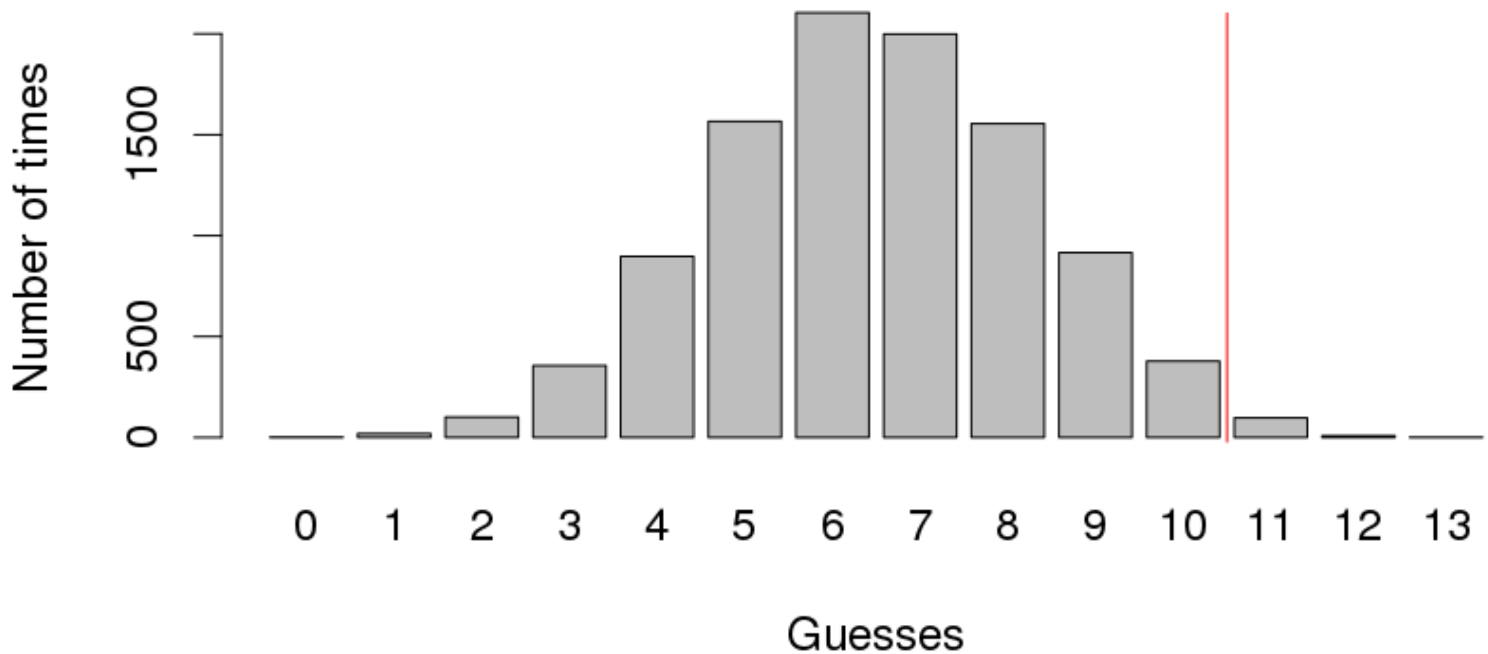
- Answer: $\pi = .5$

Question: How could we calculate the probability Paul would guess 11 or more games correctly?

- Answer: We could flip a fair coin 13 times and see how many times we get 11 or more heads. Then repeat this process 10,000 times.

- We can do this simulation in R using the following commands:
  - `> simulated.correct.guesses <- rbinom(10000, 13, .5)`
  - `> num.sims.as.good.as.paul <- sum(simulated.correct.guesses >= 11)`
  - `> proportion.as.good.as.paul <- num.sims.as.good.as.paul/10000`

Paul the Octopus



I got: 129 of 10,000 simulated trials have 11 or more correct guesses

- If Paul was guessing, he would only get 11 right 129/10,000 = .0129 of the time

Paul the Octopus

Formalizing hypothesis testing

Let's describe what we just did…

First we stated two hypotheses which were:

- Null hypothesis: Paul is guessing H_0: $\pi = .5$
- Alternative hypothesis: Paul is psychic H_A: $\pi > .5$

Next we created a ***null distribution*** that was consistent with what we would expect if the null hypothesis was true by either:

- 1. creating a binomial distribution that is consistent with H_0
- 2. simulating data consistent with the null hypothesis H_0

Finally we examined how likely we were to get the observed statistic, $\hat{p} = 11/13$, from the null distribution

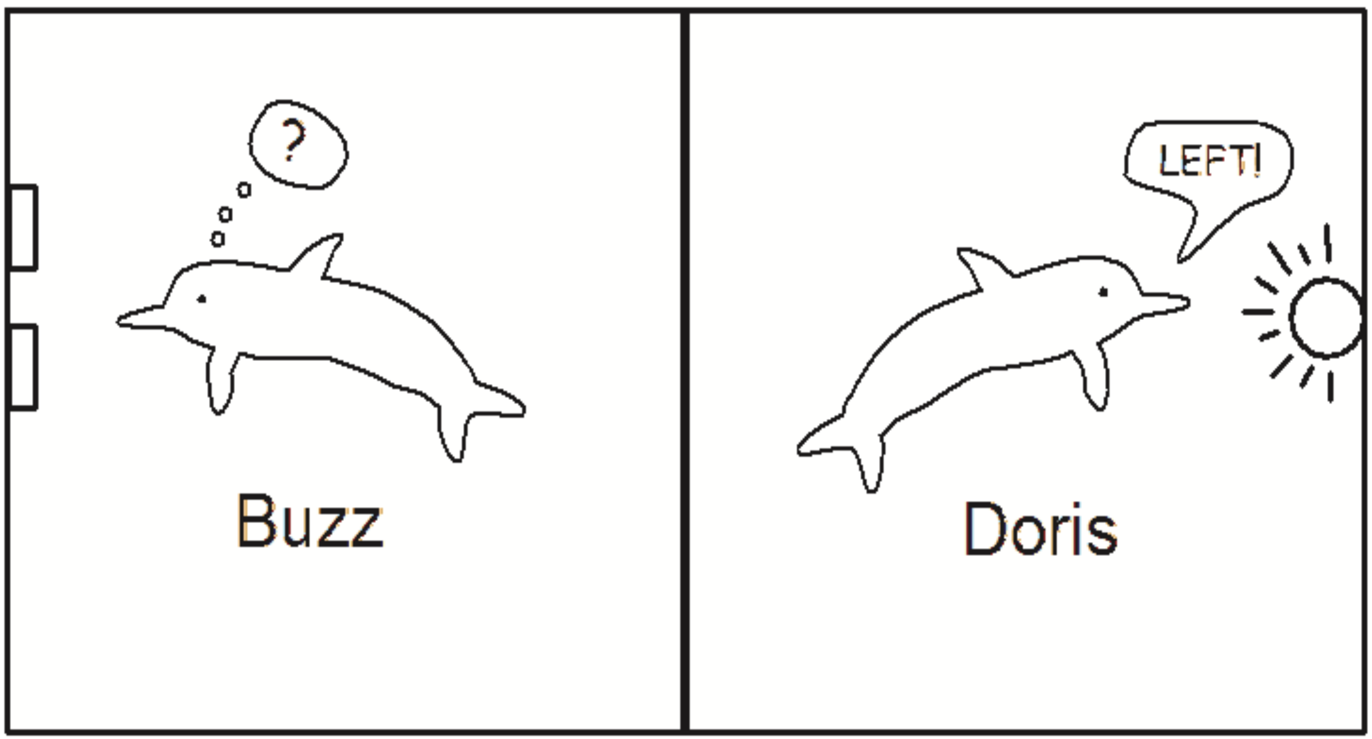
Can dolphins communicate?

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

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.



Got 15 out of 16 right

The dolphin communication study

1. What is the null hypothesis?

2. How would you write it in terms of the population parameter?

H_0: $\pi = 0.5$

3. What is the alternative hypothesis?

H_A: $\pi > 0.5$

Calculate the probability that buzz would have gotten 15 out of 16 correct if he was guessing

Binomial distribution:

```
> sum(dbinom(15:16, 16, .5))   # sum Pr(X = 15) + Pr(X = 16)
> 1 – pbinom(10, 13, .5)      #  equivalently:  1 – Pr(X ≤ 15)
```

Simulation:

```
> simulated.correct.guesses <- rbinom(10000, 16, .5)
> num.sims.as.good.as.buzz <- sum(simulated.correct.guesses >= 15)
> proportion.as.good.as.buz <- num.sims.as.good.as.buz/10000
```

The dolphin communication study

1. Assume nothing interesting is happening
    - Everything is due to chance, dolphins just guessing ($\pi = 0.5$)

2. Create a probability distribution consistent with the null hypothesis (null distribution)

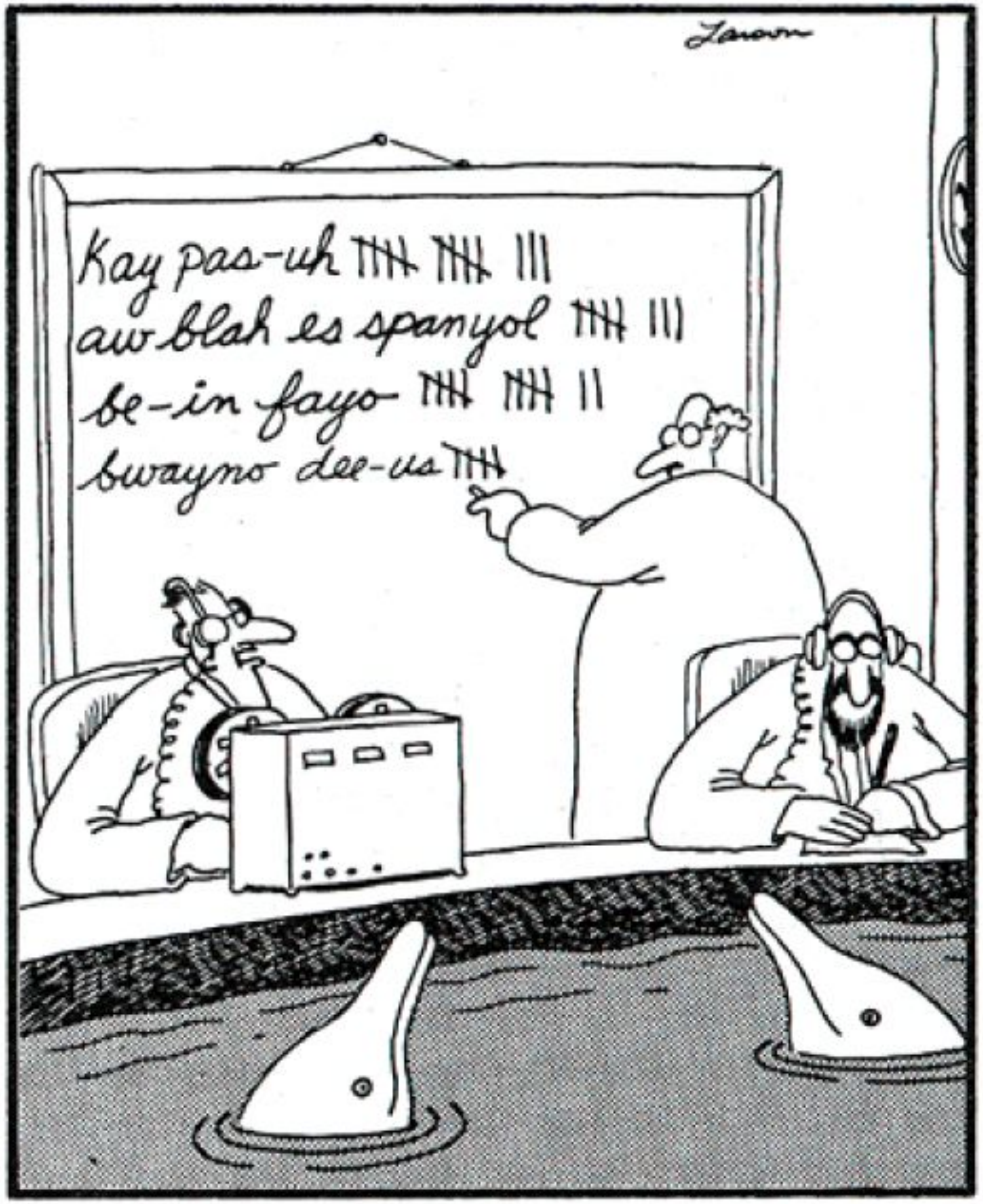
3. Examine how likely evidence is if everything was due to chance
    - What is the probability that the dolphins would guess 15 of 16 correct?

P-values

The **p-value** is the probability, when the null hypothesis is true, of obtaining a statistic as extreme as (or more extreme than) the observed statistic

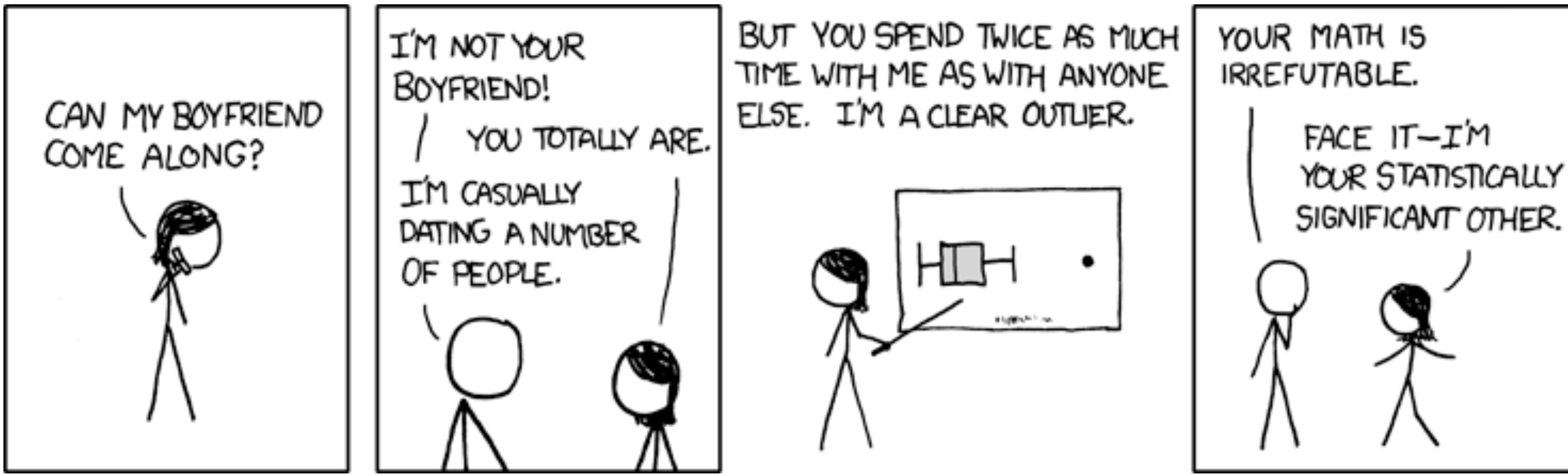
$$Pr(STAT > observed\_statistc \mid H_0 = True)$$

The smaller the p-value, the stronger the statistic evidence is against the null hypotheiss and in favor of the alternative



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

Bad example?



What is the null hypothesis here?

Are we on safe grounds saying the null hypothesis is not likely to be true?

How good is A-Rod's really?

In 2012, Alex Rodriguez had a .353 OBP based on 529 plate appearances

- Let us denote this observed performance with the symbol $\hat{p}$

This observed performance ($\hat{p}$) is due to both:

- a) A-Rod's innate ability or skill (π)
- b) Luck, randomness, chance

So how good is A-Rod really?

- Is it plausible that A-Rod's OBP ability π was really .300 and he just got lucky to get a $\hat{p}$ of .353?

How good is A-Rod's really?

<u>Question</u>: Is it plausible that A-Rod's OBP ability π was really .300 and he was just got lucky to get a $\hat{p}$ of .353?

What are the null and alternative hypotheses?

- H_0: A-Rod's ability is $\pi = .300$
- H_A: A-Rod's ability is $\pi > .300$

How good is A-Rod's really?

Let's model on base events using the binomial distribution:

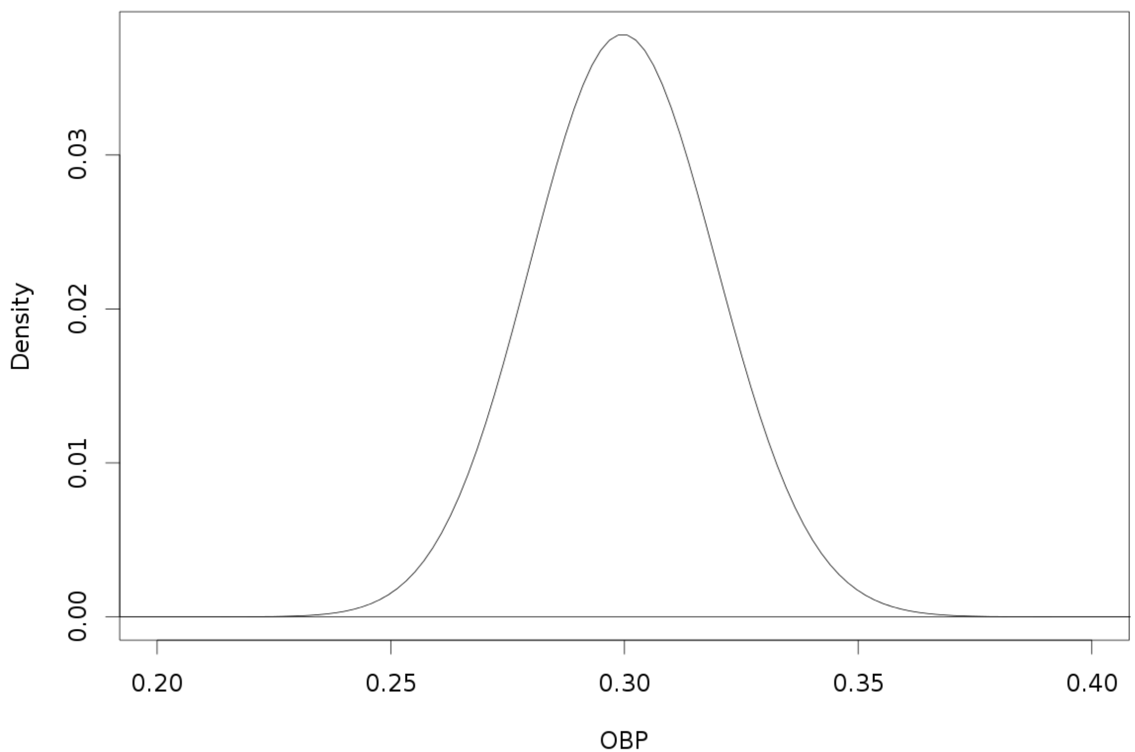
- Pr(k; n, π)
  - Probability of getting on base k times
  - Given n plate appearances, and an OBP ability of π

If A-Rod's real ability was π = .300, then for 2012 (when he had 529 PA) we have:

- Pr(k; 529, .3)

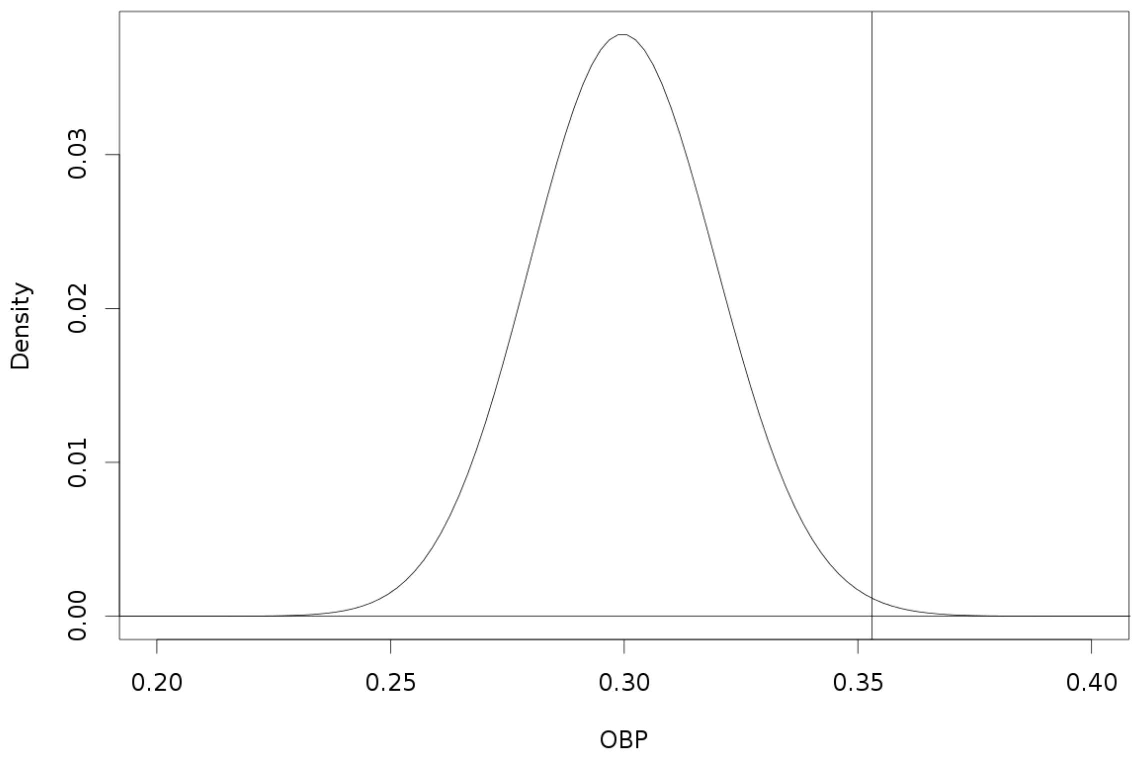
How good is A-Rod's really?

Let's look at the distribution: Pr(k; 529, .3)/PA



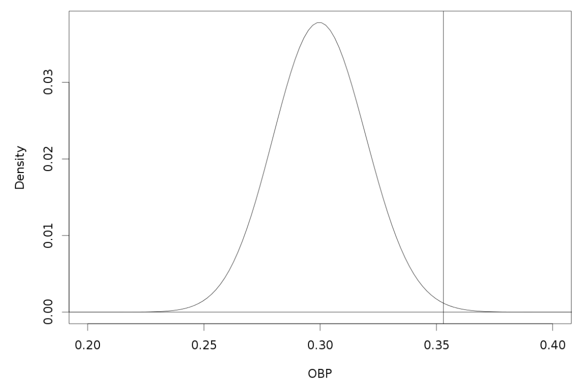
How good is A-Rod's really?

Let's look at the distribution: Pr(k; 529, .3)/PA



How good is A-Rod's really?

Let's look at the distribution: Pr(k; 529, .3)/PA



The probability that A-Rod would have an observed OBP ($\hat{p}$) of .353 or higher if his OBP ability was really .300 ($\pi = .300$) is: .0046

- i.e., p-value = .0046

Typically if a p-value is < 0.05 we "reject the null hypothesis"

- i.e., H_0 is a bad model/explanation for the data we observed

Worksheet 8

Let’s start on worksheet 8 now…

```
>source('/home/shared/baseball_stats/baseball_class_functions.R')
```

```
> get.worksheet(8)
```

Hint: all three questions are very similar!