Math 107

Hypothesis Tests: p-values
(Sections 4.2 & 4.4)

Paul the Octopus

http://www.youtube.com/watch?v=3ESGpRUMj9E

Hypotheses

- In 2008, Paul the Octopus predicted 8 World Cup games, and predicted them all correctly.
- Is this evidence that Paul's chance of guessing correctly, *p*, is really greater than 50%?
- What are the appropriate hypotheses?

1. $\mathbf{H_0}: \ \mathbf{p \neq 0.5}$ vs. $\mathbf{H_a}: \ \mathbf{p = 0.5}$
2. $\mathbf{H_0}: \ \mathbf{p = 0.5}$ vs. $\mathbf{H_a}: \ \mathbf{p \neq 0.5}$
3. $\mathbf{H_0}: \ \mathbf{p = 0.5}$ vs. $\mathbf{H_a}: \ \mathbf{p > 0.5}$
4. $\mathbf{H_0}: \ \mathbf{p > 0.5}$ vs. $\mathbf{H_a}: \ \mathbf{p = 0.5}$

How do we measure how unusual a sample statistic is, if the null is true?

Measure Evidence Against the Null

To see if a statistic provides evidence against the null hypothesis, we need to understand what values of the sample statistic we would observe just by random chance, **assuming that the null hypothesis is true**.

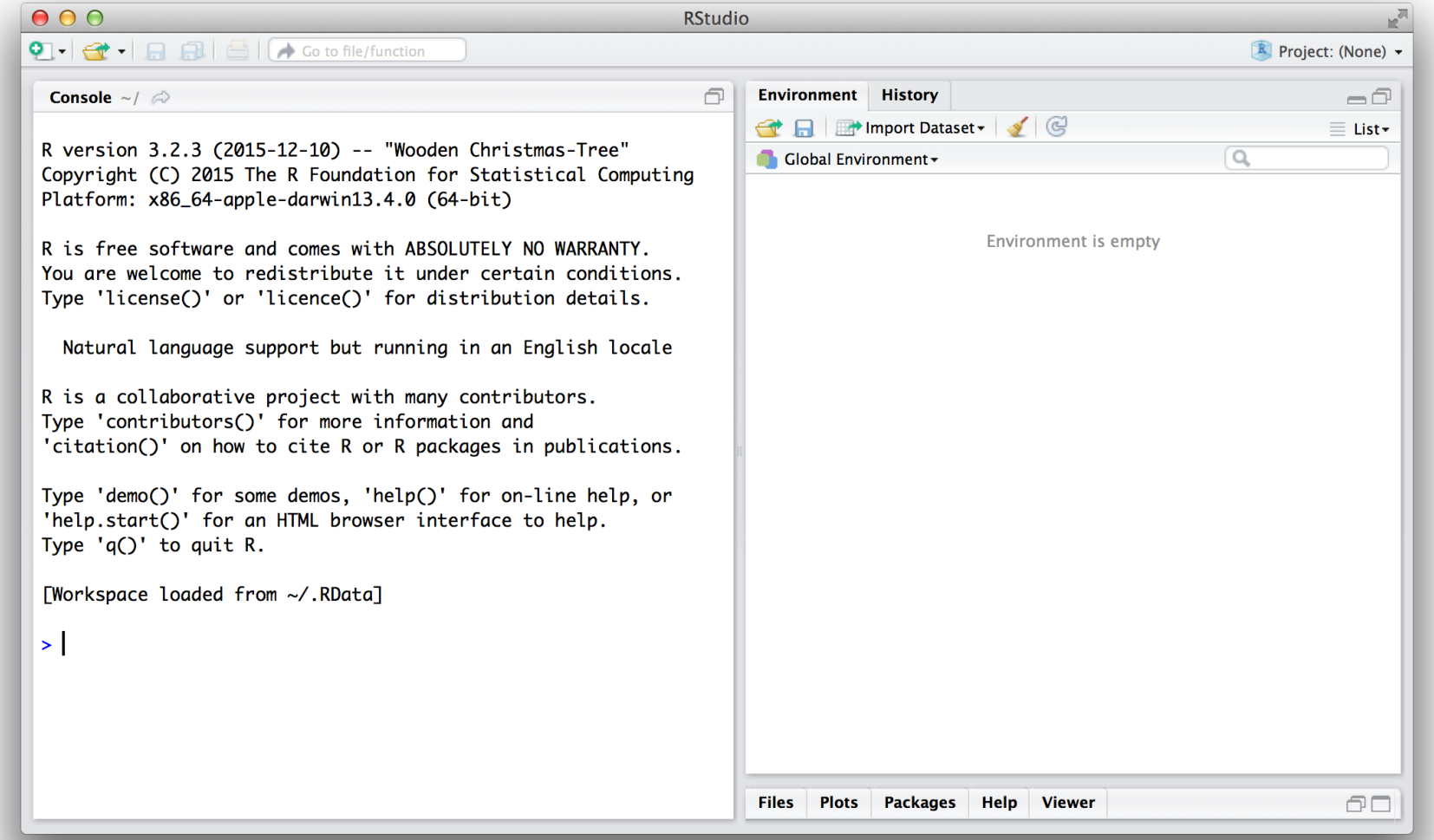
Paul the Octopus

- We need to know what values of the sample statistics we would observe just by random chance, if the null hypothesis were true.
- How could we figure this out?

Simulation

- We can simulate this choice with a fair coin.
- Heads = correct guess
- Tails = incorrect guess
- Chance of heads = 1/2 = probability Paul will guess correctly
- One repetition = one set of 8 coins flips

R Demo


RStudio
Go to file/function
Project: (None)
Console ~/
R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
>
Environment
History
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer

Randomization Distribution

A **randomization distribution** is a collection of statistics from samples simulated by assuming that the null hypothesis is true.

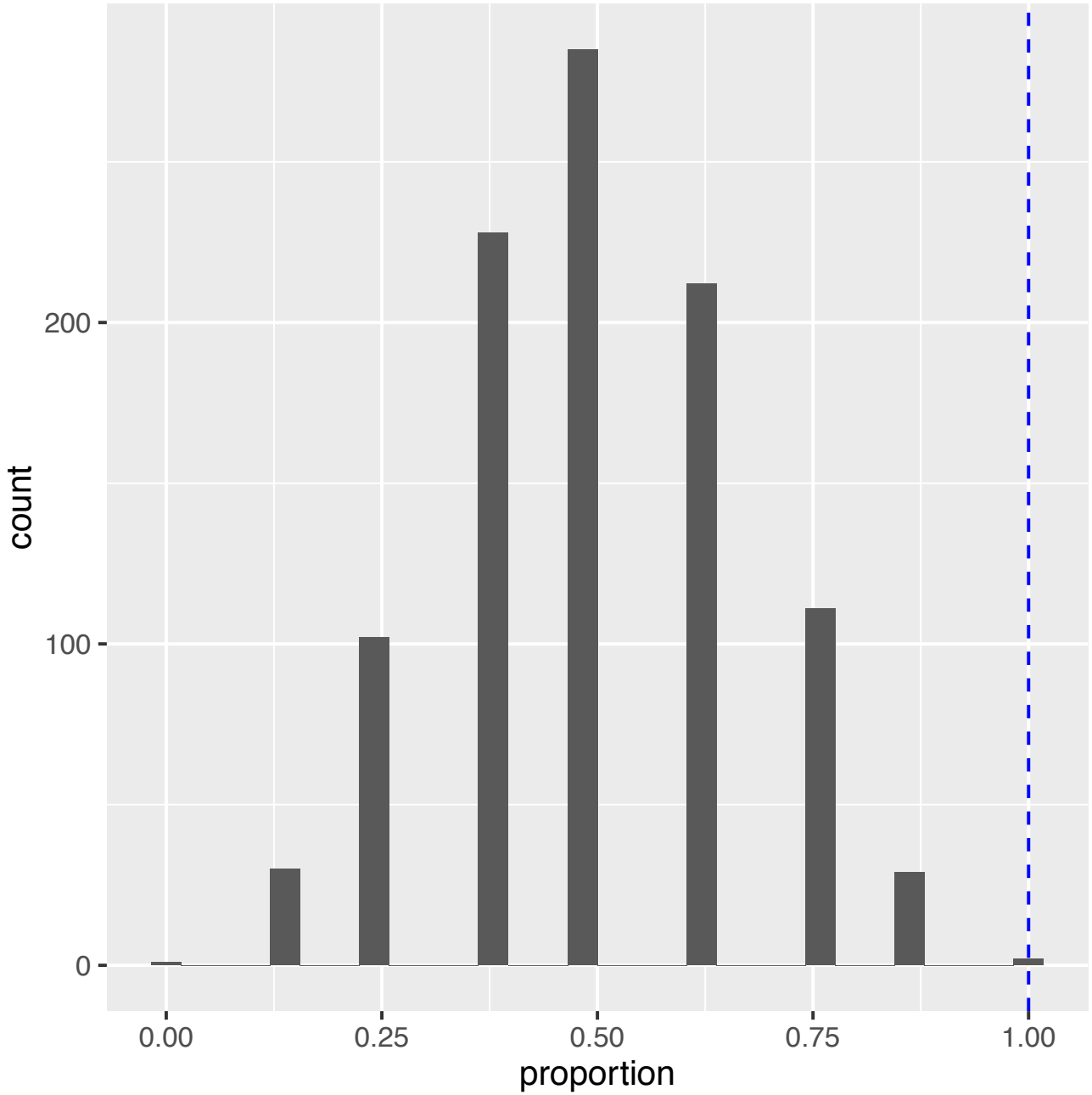
i.e. a distribution of what we would expect to observe by random chance, if the null hypothesis were true.

Quantifying Evidence

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

Calculation: the proportion of statistics in a randomization distribution that are as extreme as (or more extreme than) the observed sample statistic

1000 Simulations



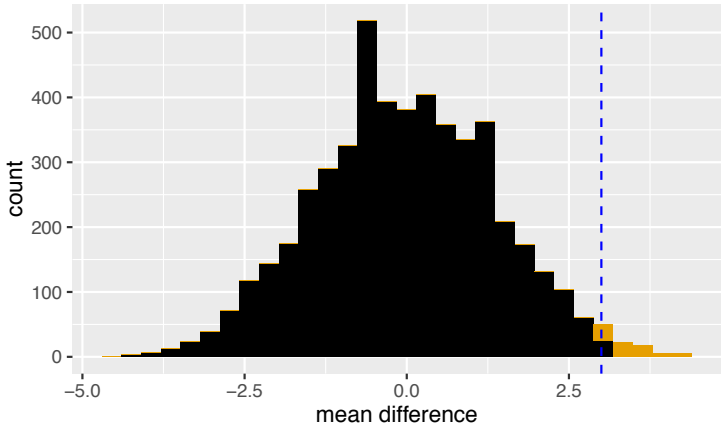
Alternative Hypothesis

- A **one-sided** alternative contains either $>$ or $<$
- A **two-sided** alternative contains $\neq$
- The p-value is the proportion in the tail in the direction specified by H_a
- For a two-sided alternative, the p-value is twice the proportion in the smallest tail

p-value and H_a

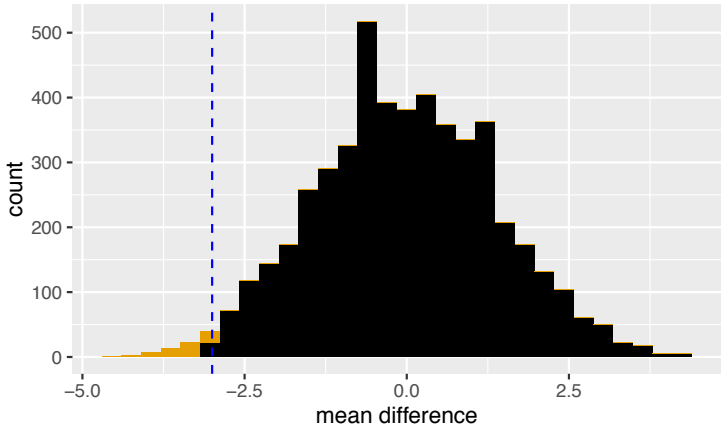
Upper-tail
(Right Tail)

$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 > \mu_2$
Observe: 3



Lower-tail
(Left Tail)

$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 < \mu_2$
Observe: -3



Two-tailed

$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 \neq \mu_2$
Observe: 3