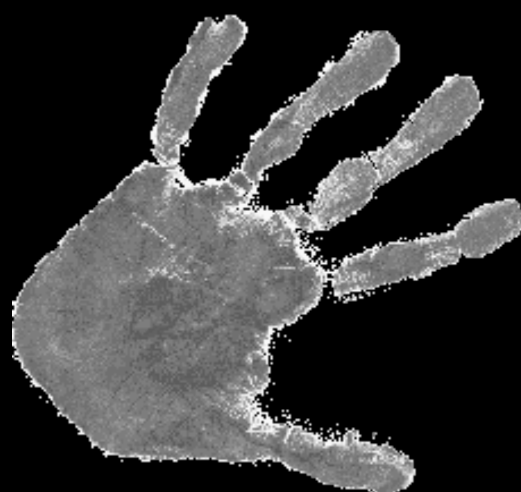
[redacted]

hypothesis testing

Bisazza et al. (1996) tested the possibility of handedness in European toads, *Bufo bufo*, by sampling and measuring 18 toads from the wild.

Of the 18 toads tested, 14 were right-handed and 4 were left-handed. Are these results evidence of a predominance of one type of handedness in toads?



stating the hypothesis

number of interest in population is the proportion that are right-handed

our null hypothesis should be that the two handedness types are equally frequent in the population

H_o: p=0.5

stating the hypothesis

our alternative hypothesis should be that left- and right-handed toads are not equally frequent in the population

H_A: p is not equal to 0.5

this is a two-sided hypothesis because the alternative hypothesis includes values on both sides of the value specified by the null hypothesis

test statistic

quantity calculated from the data that is used to evaluate how compatible the data are with the result expected under the null hypothesis

test statistic

on average, if the null hypothesis were correct, we would expect to observe nine right-handed toads (out of the 18)

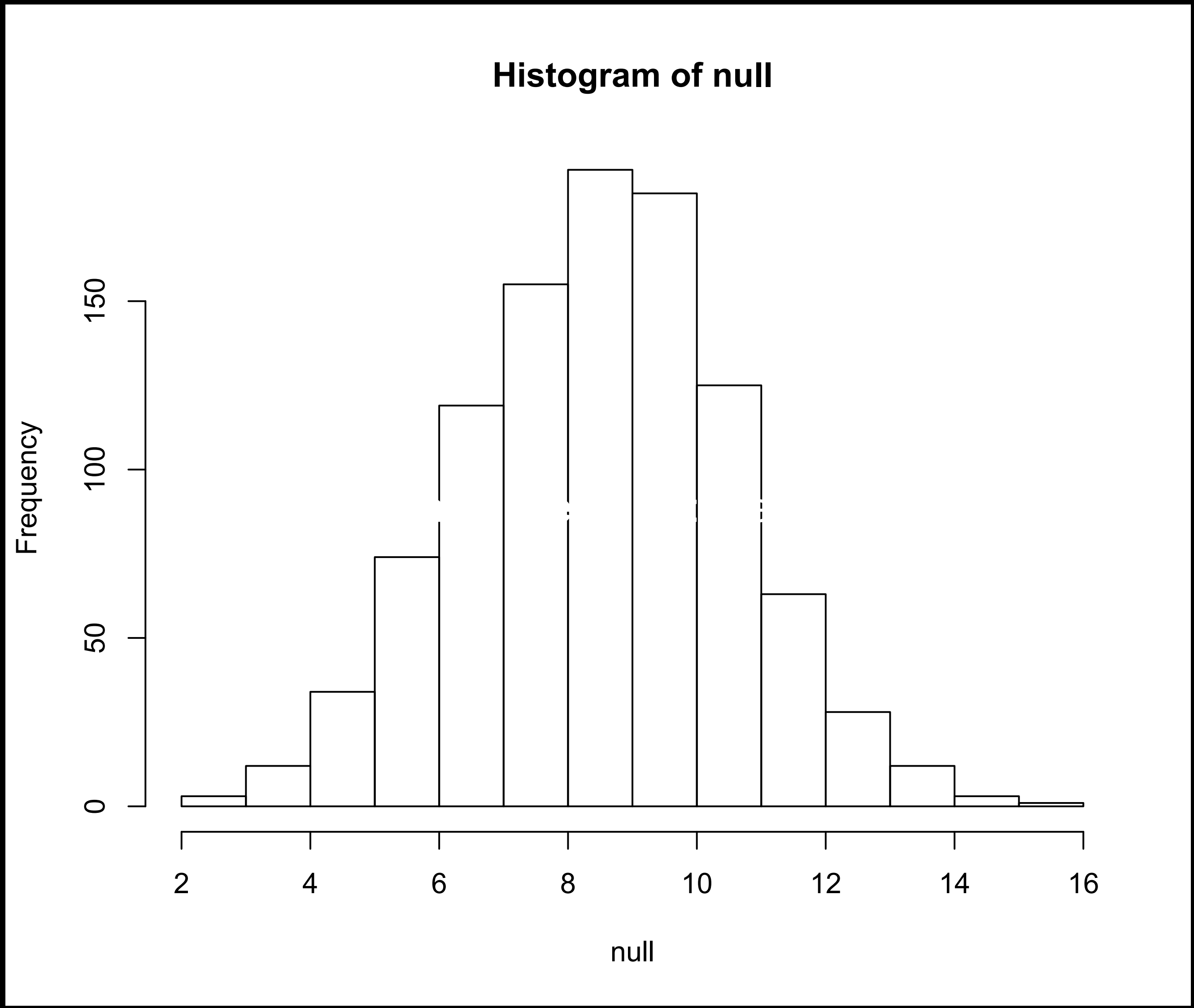
instead, we observed 14 right-handed toads out of the 18 sampled

null distribution

sampling distribution of outcomes for a test statistic under the assumption that the null hypothesis is true

sampling 18 toads under the null hypothesis is like tossing a coin in the air 18 times and counting the number of “heads” that come up (heads=right-handed)

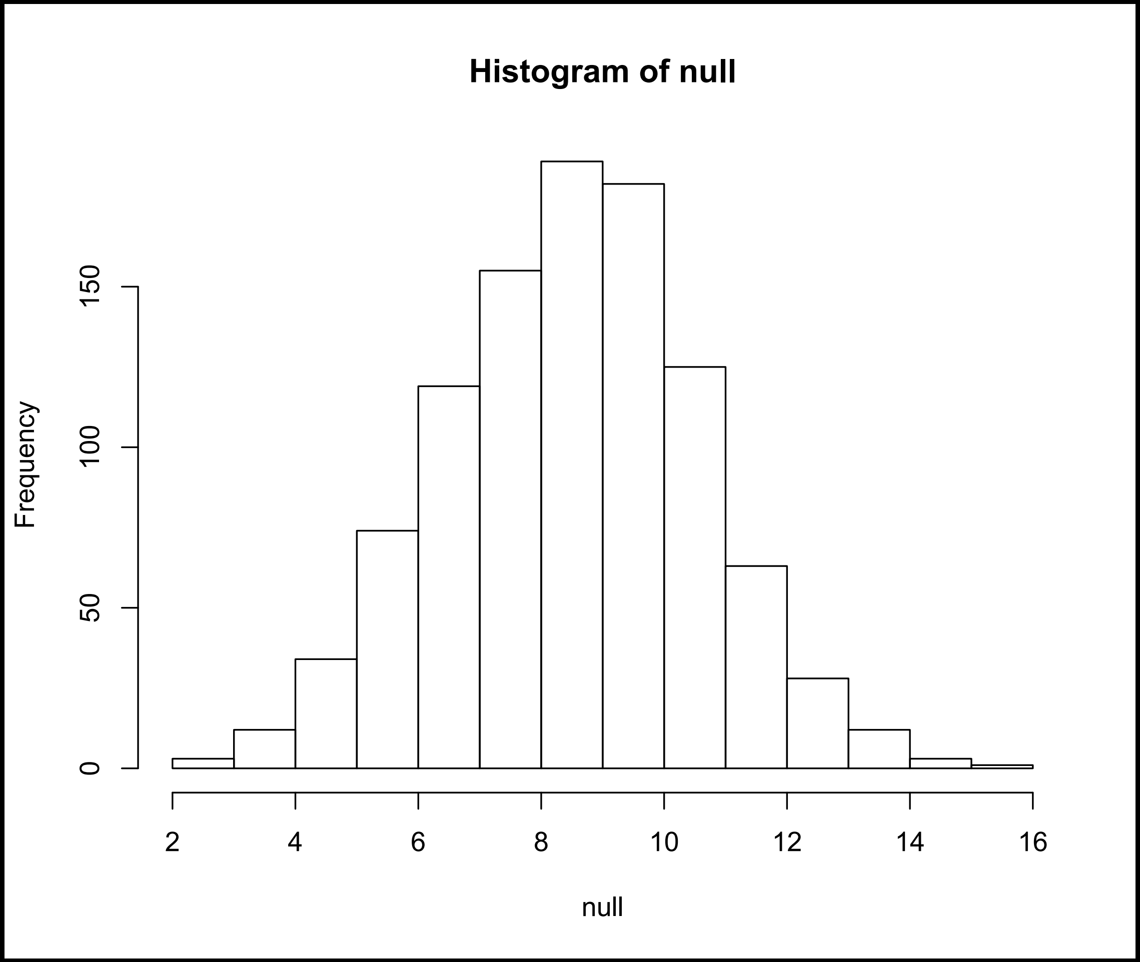
R Exercise:
Generate Null Distribution


Histogram of null
Frequency
0
50
100
150
2
4
6
8
10
12
14
16
null

quantifying uncertainty: the *P*-value

the *P*-value is the probability of obtaining the data (or data showing as great or greater difference from the null hypothesis) if the null hypothesis were true

quantifying uncertainty: the *P*-value



quantifying uncertainty: the *P*-value

don't worry about the calculation of the *P*-value at the moment—we'll get to that next week

our *P*-value is around 0.031

statistical significance

significance level (α): probability used as a criterion for rejecting the null hypothesis. If the *P*-value for a test is less than or equal to α, then the null hypothesis is not rejected

a widely used significance level is α=0.05

interpreting non-significant results

can never conclude that the null hypothesis is true

always possible

- true value differs from the null hypothesis by a small amount
- null was not rejected because of chance
- power of the test was limited by sample size

We interpret our results as the data are *compatible* or *consistent* with the null hypothesis.

reporting the results

include the following information in the summery of the results of a statistical test

- value of the test statistic
- the sample size
- the *P*-value

It's also useful to provide confidence intervals, or at least the standard errors, for the parameters of interest

errors in hypothesis testing

	Reality	
Decision	**H_0 True**	**H_0 False**
Reject H_0	Type I error	Correct
Do Not Reject H_0	Correct	Type II Error

Type I error is rejecting a true null hypothesis. The significance level α sets the probability of committing a Type I error.

Type II error is failing to reject a false null hypothesis.

The power of a test is the probability that a random sample will lead to rejection of a false null hypothesis

one-sided tests

alternative hypothesis includes parameter values on only one side of the value specified by the null hypothesis

H_o is rejected only if the data depart from it in the direction stated by the H_A