

Module 3: Introduction to hypothesis testing

Jake Ferguson (jakeferg@hawaii.edu)

Module goals

1. Understand basic principles of hypothesis testing
2. Introduce programming control-flow structures in R
3. Introduce sampling with and without replacement

Monte Carlo methods

Any method that relies on repeated, **random** sampling with a computer to obtain a result.



Why are we learning this?

- Improves intuition about NHT
- Flexible approach that does not rely on any distributional assumptions (i.e., a nonparametric method)
- Develop new programming skills

Permutation tests

Question: are two groups the same?

Solution: What would the data look like if there were no differences between the groups.

How: Randomly assign observations to each treatment

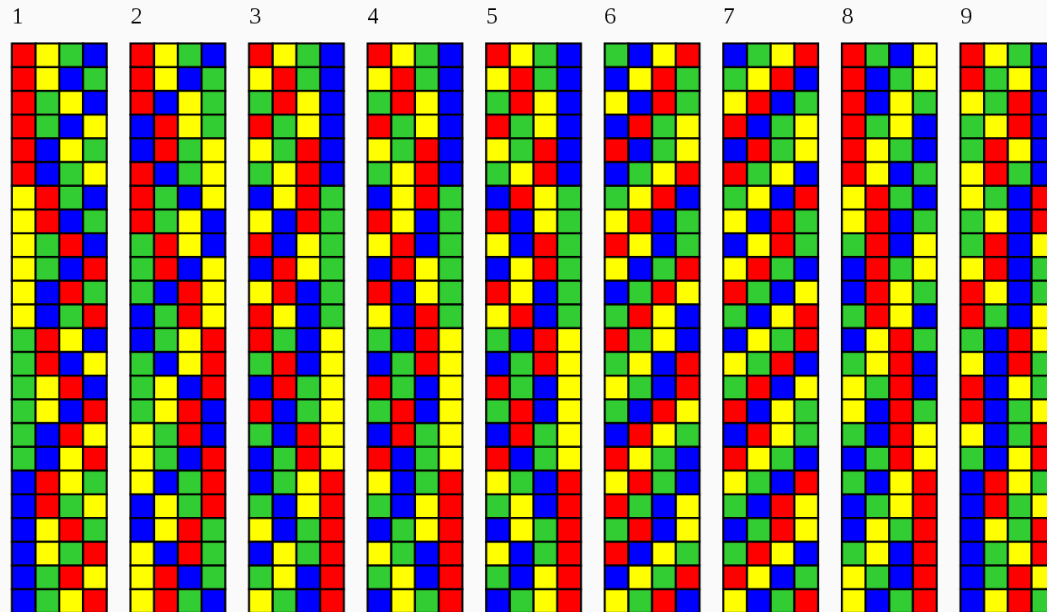


image source: www.wikimedia.org/

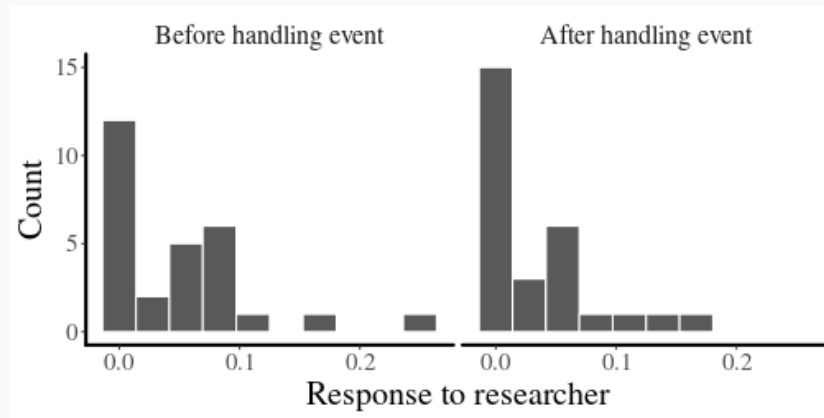
Example

Are Hawai'ian monk seals more responsive to researcher presence after a handling event?

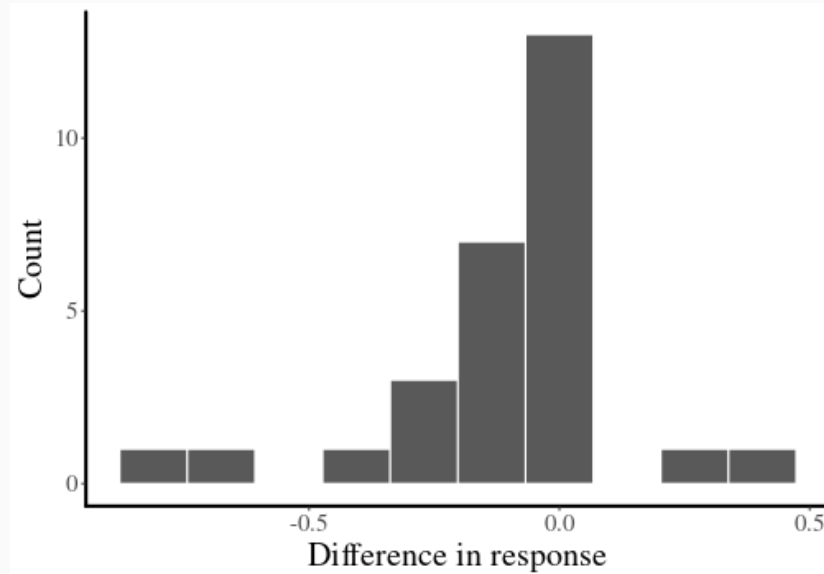


image source: [fisheries.noaa.gov](https://www.fisheries.noaa.gov)

The data



Since this is a **paired** study design, we can study the difference in each animals behavior



The null hypothesis

What does the world look like if there is nothing interesting going on?

H_0 : no difference between the groups.

How can we use our computer to create new datasets that we know are consistent with H_0 , with the same statistical properties as the original dataset?

Permute the data!

Permutation

Randomly assign observations to before or after treatment

##		Before	After	Difference
## 1	0.05263158	0.15151515	0.09888357	
## 2	0.04545455	0.00000000	-0.04545455	
## 3	0.08333333	0.05263158	-0.03070175	
## 4	0.00000000	0.00000000	0.00000000	
## 5	0.08000000	0.04000000	-0.04000000	
## 6	0.11111111	0.16666667	0.05555556	

This is equivalent to randomly changing the sign of the difference (+/-)

Randomization tests for paired data

- Calculate your statistic of interest (e.g., $\overline{x_{After} - x_{Before}}$)
 - 1. Randomly assign each observation to one of the treatments (Before/After) (for a paired t-test this equivalent to multiplying by +1 or -1 with equal probability)
 - 2. Calculate your new statistic of interest (e.g., $\overline{x_{After} - x_{Before}}$) under H_0 . Save result.
 - 3. Go back to 1.
-
- After doing 1-3 a bunch, use to construct H_0 . Compare H_0 to the original statistic.

Generating random numbers

How to randomly switch the sign of the observed difference?

The `sample` function randomly draws values from a set of possibilities:

```
sample(x=c(-1,1), size=10, replace=T) #randomly sample 10 values from -1 and 1
```

```
## [1]  1 -1  1  1 -1 -1 -1 -1 -1  1
```

Sampling with replacement means that once a value is drawn from the original set, it can be drawn again.

Exercise 3A

Using the sample function

The `for` loop

Loops are how we run a chunk of code a bunch of times.

syntax:

```
for(val in sequence)
{
  statement to do stuff.
}
```

A simple example:

```
x ← numeric(10) #an empty vector
for(i in 1:5) {
  print(i)
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
```

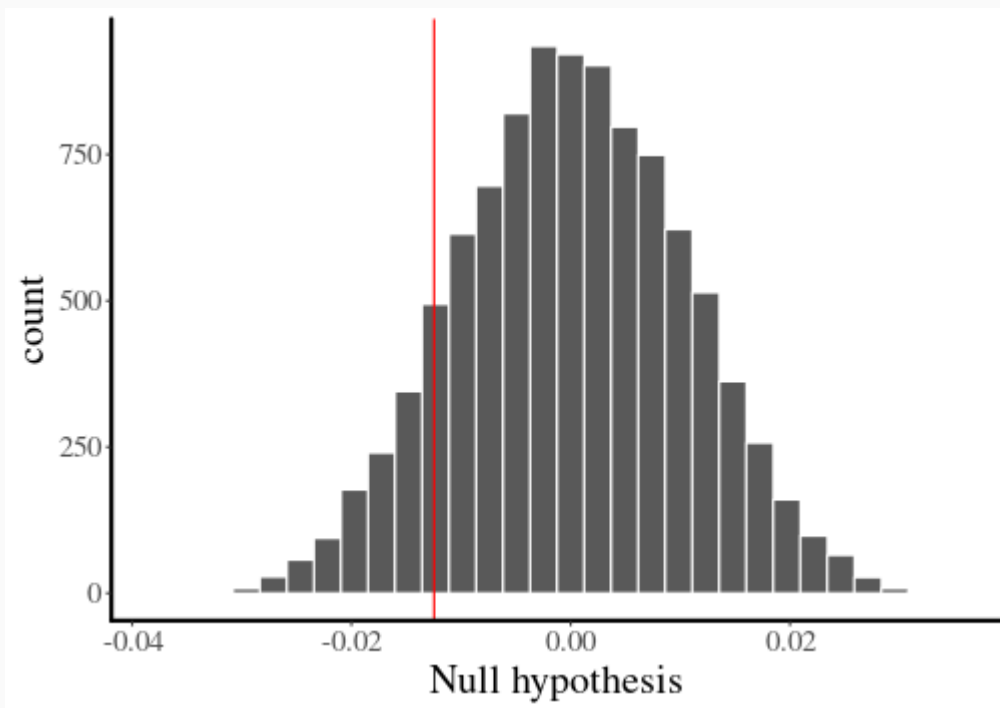
Exercise 3B

Your first for loop

The null distribution

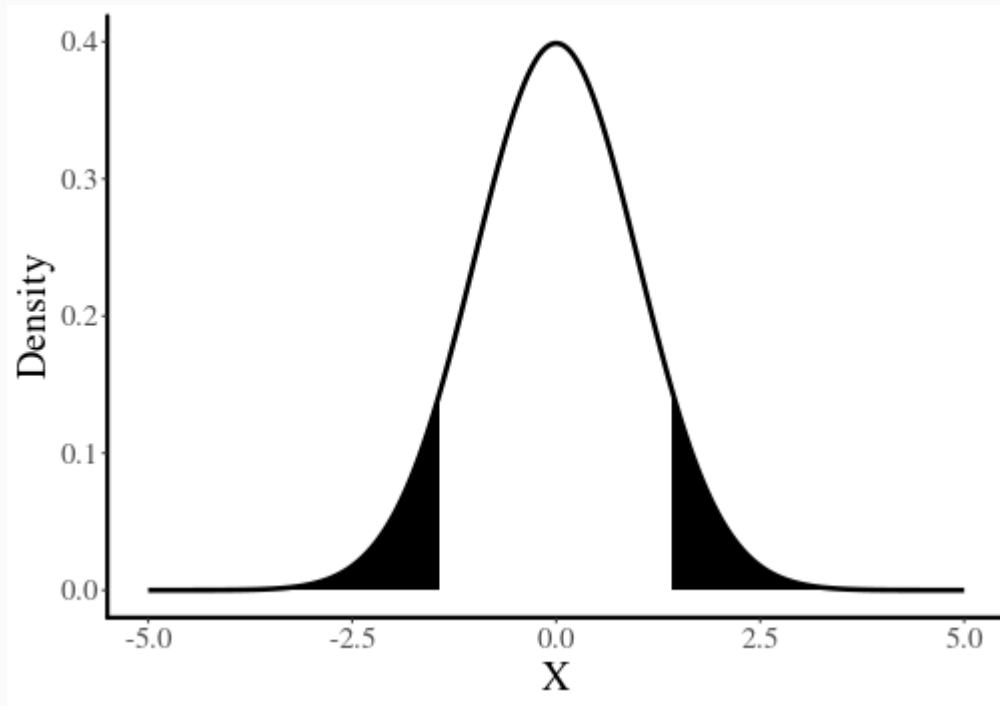
```
obs.stat ← mean(diff.dat$Difference)

null.vec ← numeric(1e4)
for(i in 1:length(null.vec)) {
  multiplier ← sample(x=c(-1,1), size=length(diff.dat$Difference), replace=TRUE)
  null.vec[i] ← mean(diff.dat$Difference*multiplier)
}
```



The p-value

A hypothesis test asks: what is the probability in future studies of observing a result as extreme or more, if the null hypothesis is true? ¹

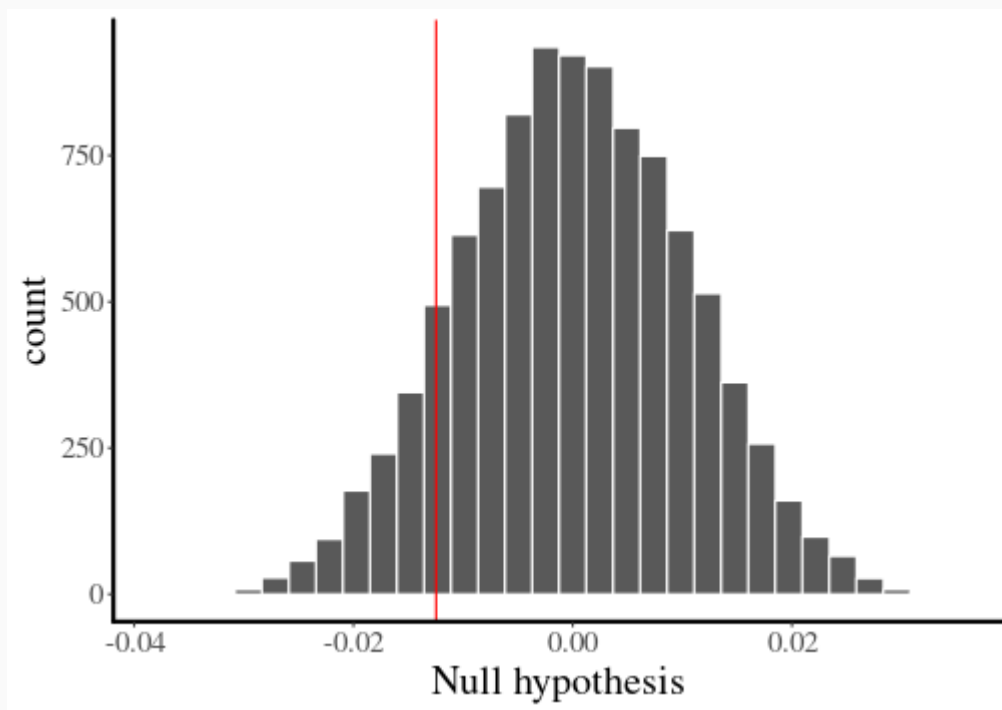


[1] Your life will be improved if you memorize this definition

Getting the p-value

```
p.val ← sum(null.vec < obs.stat)/length(null.vec) #proportion of values that are more  
print(2*p.val) #double this for a two-sided test
```

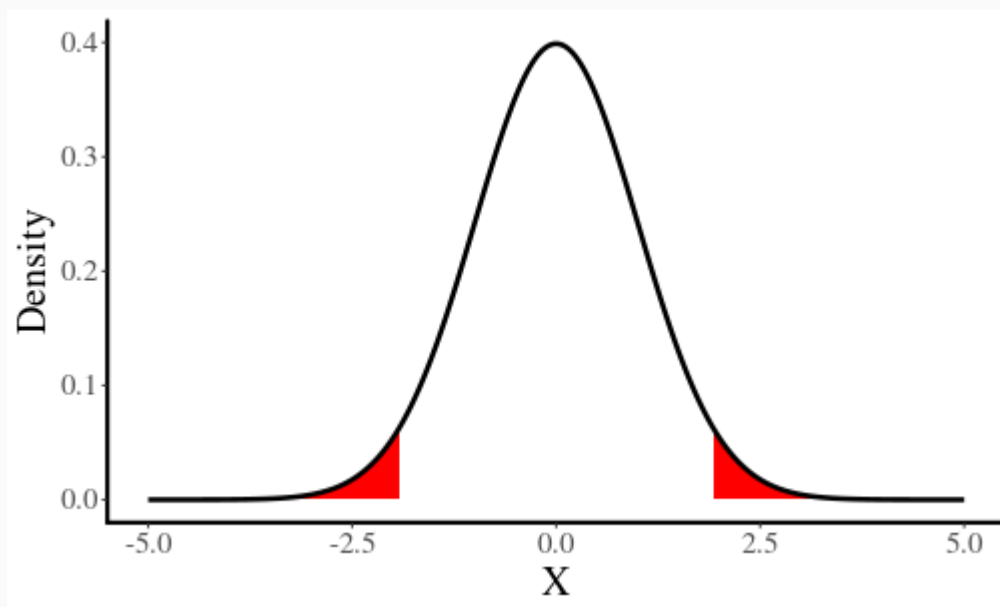
```
## [1] 0.2288
```



The α -level

The p-value measures the consistency of your data with the null hypothesis. So at what point a p-value denote inconsistency with the null hypothesis?

The α -level is a set value that denotes the researchers comfort with rejecting a difference that is true but rare. The standard level is $\alpha = 0.05$. It is completely reasonable to use a different value.



Excercise 3C

Paired-sample randomization test

Two-sample randomization test

If we have samples from two populations and want to compare them the recipe is a bit different

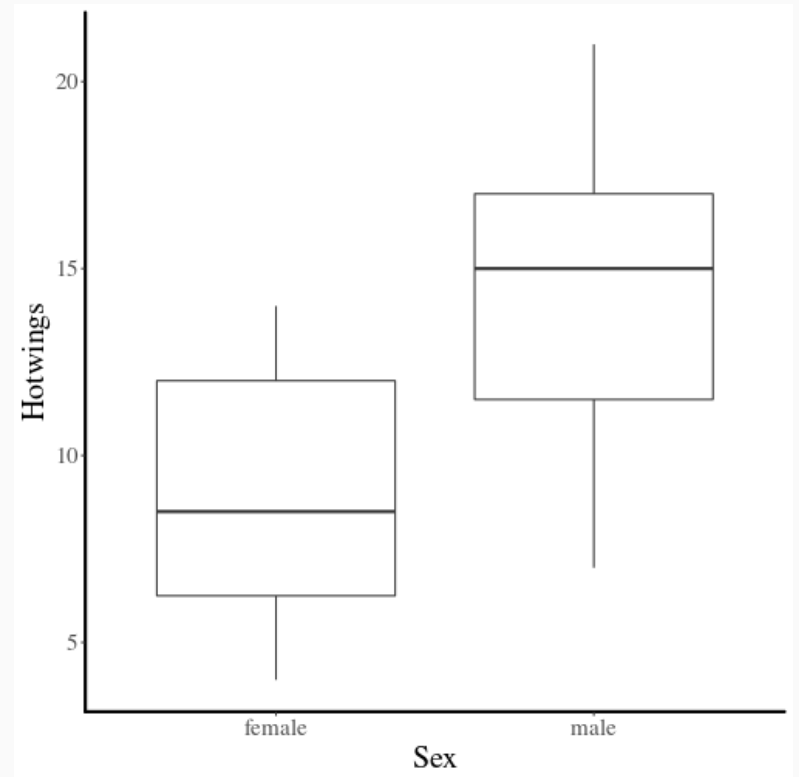
- Calculate your statistic of interest
1. Randomly assign each observation to one of the populations keeping the sample sizes of each population the same
 2. Calculate your new statistic of interest under H_0 . Save result.
 3. Go back to 1.
- After doing 1-3 a bunch, use to construct H_0 and compare to the original statistic.

Hotwings



Image: <https://cdn.pixabay.com>

```
hotwings.dat <- read.csv(file=" ../Data/h  
ggplot(data=hotwings.dat) +  
  geom_boxplot(aes(x=Sex, y=Hotwings))
```



Get the randomization distribution

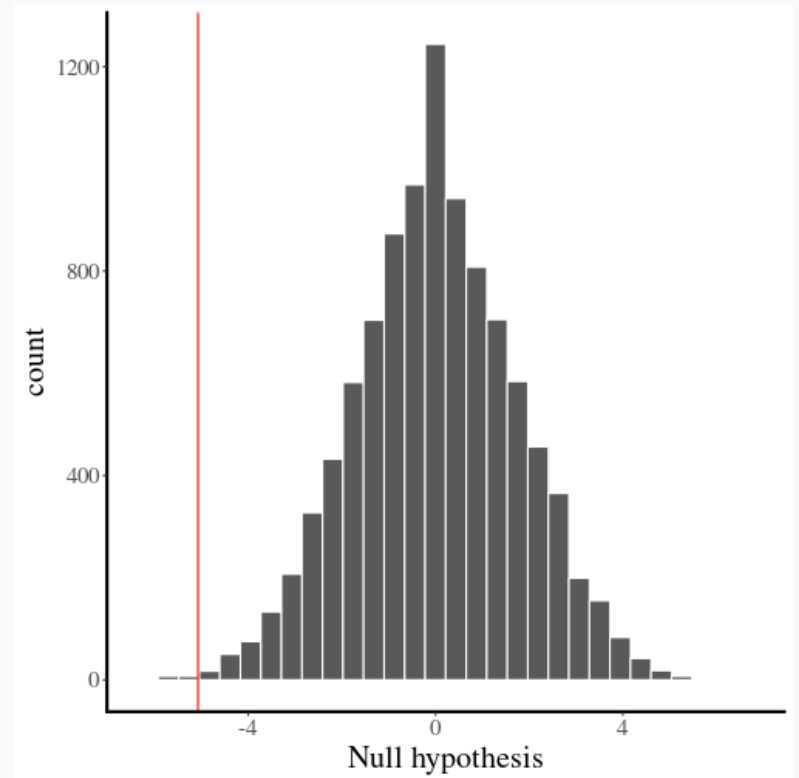
```
obs.val ← mean(hotwings.dat$Hotwings[1:
print(obs.val)
```

```
## [1] -5.071429
```

```
null.vec ← numeric(1e4)
for(i in 1:1e4) {
  index      ← sample(1:28, 14, replac
  null.vec[i] ← mean(hotwings.dat$Hotwi
}

p.val ← 2*sum(null.vec < obs.val)/lengt
print(p.val)
```

```
## [1] 0.0026
```



What do we conclude?

Excercise 3D

Two-sample randomization test

Summary: Permutation tests

- Resamples data without replacement.
- Analogs of two-sample t-test, ANOVA that will be discussed later. Can create new tests as well
- Doesn't make distribution assumptions, doesn't require balanced designs
- Used to perform hypothesis test (calculate p-values).