

module_4_3

Keaton Wilson

4/22/2020

More hypothesis testing and Extensions of t-tests and permutation tests and visualization

Let's start working with a new data set that complements what we've been working with so far.

New data

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse_1.3.0
```

```
## v ggplot2 3.3.0      v purrr  0.3.3
## v tibble  3.0.1      v dplyr  0.8.5
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
```

```
## -- Conflicts ----- tidyverse_conflicts_0.1.0
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
site_data = read_csv("https://tinyurl.com/yardhofj")
```

```
## Parsed with column specification:
## cols(
##   site_id = col_double(),
##   year = col_double(),
##   tussocks = col_double(),
##   dist_to_water = col_double(),
##   stone_size = col_double(),
##   num_nests = col_double()
## )
```

```
glimpse(site_data)
```

```
## Rows: 200
## Columns: 6
```

```
## $ site_id      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16...
## $ year         <dbl> 1971, 1971, 1971, 1971, 1971, 1971, 1971, 1971, 1971, ...
## $ tussocks     <dbl> 4.219239, 3.932796, 6.263532, 4.116507, 4.118727, 7.1...
## $ dist_to_water <dbl> 27.86114, 25.33264, 28.93208, 28.39797, 28.75344, 28....
## $ stone_size   <dbl> 45.04550, 39.73190, 37.47425, 45.87984, 45.69094, 38....
## $ num_nests    <dbl> 18, 8, 3, 25, 15, 7, 15, 19, 13, 15, 25, 11, 23, 22, ...
```

```
site_data
```

```
## # A tibble: 200 x 6
##   site_id year tussocks dist_to_water stone_size num_nests
##   <dbl> <dbl>   <dbl>       <dbl>       <dbl>   <dbl>
## 1     1     1  1971     4.22         27.9        45.0     18
## 2     2     2  1971     3.93         25.3        39.7      8
## 3     3     3  1971     6.26         28.9        37.5      3
## 4     4     4  1971     4.12         28.4        45.9     25
## 5     5     5  1971     4.12         28.8        45.7     15
## 6     6     6  1971     7.14         28.5        38.9      7
## 7     7     7  1971     4.78         26.4        45.4     15
## 8     8     8  1971     3.42         26.6        41.2     19
## 9     9     9  1971     3.64         25.7        42.6     13
## 10    10    10  1971     6.44         30.7        48.4     15
## # ... with 190 more rows
```

A big one is making comparisons between the 1971 data and the 2011 data. Can you think of some questions we could ask?

Hypothesis brainstorm

Take a few minutes and write down some ideas you have for questions that could be asked and associated hypotheses?

Focusing on one research question

One big question we might ask is whether or not there have been changes in the number of tussocks between the two time periods. Do we have some statistical tools in our toolkit to answer this?

1. Visualization
2. t-test (think about the difference of this dataset though)
3. Permutation test (think about the difference of this dataset though)

Group Discussion and Work

Spend 10 minutes in your groups putting together one (or more) of these tools that demonstrates an answer. You should be able to explain the output and differences in laymans terms.

```
library(tidyverse)
site_data %>%
  ggplot(aes(x = tussocks, fill = as.factor(year))) +
```

```
geom_density(alpha = 0.8) +  
xlab("Number of Tussocks per Site") +  
ylab("Density") +  
scale_fill_discrete(name = "Year") +  
theme_classic()
```

