**Terms, topics, or concepts you should be familiar with:**

p-value

reproducibility crisis

statistic

parameter

continuous variable

discrete variable

nominal variable

ordinal variable

experimental vs observational studies

blinding

pseudoreplication

biological vs technical replicates

confounding variables

common faults in plots

sample

population

transformation

parametric

non-parametric

**R skills you should have**

Create matrices and vectors

Subset a vector, dataframe, or matrix to select only specific elements

Read a csv file to import data

Make a publication qulity plot of 1, 2, or 3 variables that have a mix of continuous and discrete values

Perform the statistical tests mentioned below

**Tests you should be able to run**

Binomial

Chi-square

T-test (single sample, two sample, paired)

Anova and posthoc test

Permutation

Correlation test

General linear model

**Example Problems**

You administer a vaccine candidate for covid-19 and a placebo each to 1000 individuals. You find that after 2 months there have been covid-19 cases in 32 placebo group and 6 vaccine group participants.

Does this vaccine work? \_\_\_\_\_\_\_\_\_\_\_

What test did you use? \_\_\_\_\_\_\_\_\_\_\_

What p-value was associated with this test? \_\_\_\_\_\_\_\_\_\_\_

You measure height of students at the MSC and the gym. Are the heights you measured significantly different

MSC: 126, 164, 148, 120, 178, 183

Gym: 151, 109, 151, 174, 118, 136

What test did you use for this question? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What p-value was associated with this test? \_\_\_\_\_\_\_\_\_\_\_\_\_

What do you infer from your test? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

You grow plants with three different potting soils and measure height at 21 days. Describe the results of your study.

Soil1: 23, 12, 45, 23, 21, 45, 21

Soil2: 35, 45, 21, 34, 67, 23, 16

Soil3: 16, 21, 18, 33, 16, 21, 19

Stickleback fish occur in deep water and shallow water populations. These populations rarely interbreed. It has been hypothesized that these fish have genetic adaptations to their habitat. To test this, you grow fish from both strains in both deep and shallow water. Does the data below support the hypothesis that these fish are adapted to their natural habitat? The values in the table are fitnesses for fish in your experiment

|  |  |  |
| --- | --- | --- |
|  | Deep water habitat | Shallow water habitat |
| Deep water fish | .97, .78, .99, .87, .91, .89 | .61, .87,.88, .78, .80, .37 |
| Shallow water fish | .56, .95,.73, .81, .89, .64 | .77, .95,.93, .95, .89, .94 |

You measure reproductive success of fish in your study you use fat body measure as a correlate of health and record sex and number of offspring that survive to adulthood for each fish in the study. Interpret the results of your analysis of this data.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fat body measure | 0.65 | 0.35 | 0.65 | 0.41 | 0.62 | 0.42 | 0.64 | 0.51 | 0.49 | 0.64 | 0.25 | 0.68 | 0.8 | 0.23 | 0.63 | 0.48 | 0.28 | 0.71 | 0.61 | 0.36 |
| Sex | f | m | f | f | m | f | f | f | m | m | m | m | m | f | m | m | f | m | f | f |
| Total reproduction | 10 | 16 | 9 | 8 | 0 | 9 | 9 | 9 | 8 | 3 | 17 | 14 | 17 | 8 | 9 | 18 | 8 | 7 | 7 | 8 |