

Week 5 Homework

Aaron R. Caldwell

Reading and Resources

So, this week is going to be a little different. Instead of having step-by-step instructions I am going to give you some questions to answer and some general links for how to answer those questions in R. If you ever get stuck you can check the answers document.

Data

For this activity you will need three datasets that are freely available in R.

```
data("sleep")
data("ChickWeight")
data("iris")
```

Correlation

Let's assume you are a botanist and are interested in the relationship between different measures of flower size. You are going to use the `iris` dataset to determine the correlation between 4 different measures: sepal length, sepal width, petal length, and petal width.

Using the `iris` dataset, measure the correlations between the 4 variables `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`.

Correlation Questions

1. What is the strongest correlation?
2. Which variables were “significantly” correlated?

Help on Correlations

1. Check out the `cor.test` function in R with `?cor.test`
2. You could possibly use the `correlation` R package
3. How to measure correlations in R
4. How to create correlation matrices in R

t-test

Let's compare the groups (i.e., drug) in the “sleep” dataset and see if there is a difference in the hours of sleep each night

Now we can use the `sleep` dataset to test whether the `group` variable causes a difference in `extra`

t-test Question

1. Does `group` result in differences in `extra`?
2. Are the differences in group practically equivalent (equivalence being any difference between -1 and 1)?
3. Does a non-parametric test result in different conclusions (i.e., Wilcoxon signed-rank test)?

Help on t-tests

1. Please read the basic instructions on paired t-tests and non-parametric tests
2. For equivalence tests check out the TOSTER package.

ANOVA

Suppose you are an animal scientist and you have data on different diets fed to chickens over 21 weeks. Let's assume the scientist wants to test if the weight of the chicks is moderated by diet over time

The `ChickWeight` dataset can be treated as a factorial ANOVA with a within (time) and between subjects factor.

ANOVA questions

1. Is there a significant interaction between Time and Diet? I.e., is the effect of time moderated by diet?
2. *Bonus*: are the assumptions of ANOVA tenable?
3. *Bonus*: would a mixed model (HLM/multilevel model) be more appropriate?

Help on ANOVAs

1. Lecture materials on “mixed” ANOVAs
2. Blog post on the basics of ANOVA with `afex`
3. `afex` documentation
4. Check out `?aov_car` in R

Additional Help

1. `emmeans` R package
2. `interactions`
3. Introduction to Mixed Models in R
4. `easystats` also has ways of checking the model fit.