

# Analysis of Ecosystems homework week 5

Intro to statistical inference

*The assignment*

*05 February 2019*

## Data preparation

This assignment will use data in the `mpg` data from `ggplot2`—`cty` and `displ` are provided, and script to add the predictor variable, `origin`, is already in Code Chunk 3 for you. Note that `echo=FALSE` so you should not see this script in the Word file, but `results='verbatim'` so we should see the results of the call to `str()`.

## The t test

### Assumptions

Briefly describe three assumptions of the t test. Recall that two of them relate to the data themselves and a third is more about experimental design.

### Distribution

#### Graphing

Make one or more graphs that visualizes the distribution of the response variable `mpg$cty` with respect to the assumptions of the t test model. Use and copy Code Chunk 4 as necessary. Relevant items in your toolbox include histograms and density plots, faceting and colour aesthetics, Q-Q plots, and transformations.

### Interpretation

Below the code chunk, respond to the following:

- What do you conclude about the distribution of these data with respect to the assumptions of the t test?
- Describe what modifications, if any, you applied to the data to better fit model assumptions.

### Variance

#### Test

In Code Chunk 5, conduct one or more relevant statistical tests to assess whether variance in the response variable `mpg$cty` meets the assumptions of the t test model. Manipulate the data as necessary.

## Interpretation

Below the code chunk, respond to the following:

- What do you conclude about the variance of these data with respect to the assumptions of the t test?
- Describe the modification you applied to the data to better fit model assumptions.

## The linear model

### Assumptions

Briefly describe four assumptions of the linear regression model.

### Distribution

### Graphing

Make one or more graphs that visualize the distribution of the response variable `mpg$cty` with respect to the assumptions of the linear model. Use and copy Code Chunk 6 as necessary. Relevant items in your toolbox include histograms and density plots, Q-Q plots, and transformations.

## Interpretation

Below the code chunk, respond to the following:

- What do you conclude about the distribution of these data with respect to the assumptions of the linear model?
- Describe the modification you applied to the data to better fit model assumptions.

## Fit a linear model

### Test

In Code Chunk 7, use `cty` and `displ` from `mpg` as response and predictor variables, respectively, to:

- Fit a linear model using formula notation
- Provide summary statistics of the model

### Plot

In Code Chunk 8, produce an appropriate graph of the linear model fit above using `ggplot`. Include the following:

- Data
- Trendline
- Informative axis labels

## Evaluate the linear model

### Model fit

Use Code Chunk 9 to evaluate the model fit with respect to assumptions of the linear model. Below the Code Chunk:

- Give an interpretation of model fit
- Describe what you base this assessment on

### Model results

Report the following from the linear model (enter as text, not R output). When math is required on your part using information from R output, show your work. Provide both the letter that designates the test statistic/parameter as well as the value, separated by an = sign :

- Degrees of freedom
- Total Sum of Squares (show your math)
- Test statistic for the overall model
- Test statistic for the dependent variable
- Results of significance test
- How much variation is explained by the linear model?

### Interpretation

Below the code chunk, give your conclusion about the relationship between these variables.