Tutorial #1 - Introduction to Linear Modelling

Guy F. Sutton

05/05/2020

What is a linear regression?

A linear regression model is one of the most basic models used to analyse ecological data. But, what makes something a linear regression model?

1. Linear - Our predictor (independent) variable shows a linear relationship with our response (dependent) variable.
2. Regression - We are measuring the response between one (or more) predictor variables and a response variable.

When we model one numeric predictor and a numeric response variable, we have a *simple linear regression (SLR)*. When we have more than one numeric predictors and a numeric response variable, we have a *multiple regression*. Today, we are going to focus on simple linear regressions in R.

So, when we model Y as a linear function of X, we are performing an SLR.

What are the assumptions of a SLR?

There is a nice acronym which makes this easy to remember: **LINE**.

1. Linear - The relationship between your predictor and response is linear.
2. Independent - The errors (residuals) are independent (no autocorrelation, pseudo-replication, ect...).
3. Normal - The errors (residuals) are normally distributed. *NB! - SLR does not assume your raw data are normally distributed, just the errors*
4. Equality of variance (homogeneity) - At each value of your predictor variable, the variance in your response variable is equal.

How do we run a simple linear regression in R?

R has a built-in function called `lm` that is the workhorse used to fit linear models. You can find more information on this function by typing `?lm` into your console.

The general formula for running a linear regression in R is:

```
model_name <- lm(response ~ predictor, data = data_frame_name)
```

Let's break this down:

1. We are telling `R` to perform a linear regression by using the `lm` function.
2. We are going to store/save our linear regression in a variable called `model_name` using the <- (assign) key.
3. We would like to model our `response` variable as a linear function of the `predictor` variable.
4. The `data = ...` argument tells `R` where to look for your data.