

# Exercises

## Exercise: Linear model with single continuous explanatory variable

1. As in previous exercises, either create a new R script (perhaps call it `linear_model_1`) or continue with your previous data exploration R script in your RStudio Project. Again, make sure you include any metadata you feel is appropriate (title, description of task, date of creation etc) and don't forget to comment out your metadata with a `#` at the beginning of the line.
2. Import the data file 'loyn.txt' into R and take a look at the summary of the graphical data exploration you performed during the previous exercise. The aim of this exercise is to get familiar with a simple linear model with a continuous response variable, bird abundance (`ABUND`) and a single continuous explanatory variable forest area (`AREA`). Ignore the other explanatory variables for now.
3. Fit an appropriate linear model in R (remember you may have to transform `AREA`, return to your data exploration if in doubt) using the `lm()` function. Try to remember which is your response variable and which is your explanatory variable. Also remember to use the `data =` argument. Assign this linear model to an appropriately names object (`loyn.lm` if you imagination fails you!).
4. Obtain the ANOVA table by using the `anova()` function on your model object. What is the null hypothesis you are testing here? Do you reject or fail to reject this null hypothesis? Explore the ANOVA table and make sure you understand the different components. Refer back to the lectures if you need to remind yourself or ask an instructor to take you through it.
5. Now display the table of parameter estimates using the `summary()` function on your model object. Again, make sure you understand the different components of this output and be sure to ask if in doubt. What is the estimate of the intercept and slope? Write down the word equation of this linear model including your parameter estimates (hint: think  $y = a + bx$ ).
6. What is the null hypothesis associated with the intercept? What is the null hypothesis associated with the slope? Do you reject or fail to reject these hypotheses?

7. Looking again at the output from the `summary()` function how much variation in bird abundance is explained by your log AREA variable?
8. Check the assumptions of your linear model by plotting the residuals (hint: use the `plot()` function on your model object); check all assumptions. Are there patterns in the residuals? Do you have normality and homogeneity of residuals?
9. Using your word equation from Question 5, how many birds do you predict if AREA is 100?
10. Calculate the fitted values from your model using the `predict()` function. Refer back to the model interpretation video if you need a quick reminder of how to do this. Ask an instructor to take you through it if you're in any doubt (they'd be happy to take you through it).
11. Use the `plot()` function to plot the relationship between bird abundance (ABUND) and your log transformed AREA variable. Once you have created the plot then add the fitted values calculated in Question 10 as a line on the plot.
12. OK, this is an optional question. Add the 95 % confidence intervals around the fitted line to your plot. Remember you will need to use the `predict()` function again but this time include `these.fit=TRUE` argument.
13. Ok, and another optional question. This time plot the relationship between bird abundance (ABUND) and the original untransformed AREA variable. Now back-transform your fitted values (remember you got these with the `predict()` function) to the original scale and add these to the plot. Comment on the differences between the transformed and back-transformed fitted values.

End of the linear model with single continuous explanatory variable exercise