**Vigilance in geese: Instructions for Part 2**

If you haven’t already done so, you first need to enter the data from the video that you watched after Part 1 of the practical into an Excel spreadsheet. You should enter data on group size, the two measures you took from the video (i.e., ‘Frequency’ and ‘Duration’), as well as a derived variable calculated as ‘Duration/Frequency’. This derived variable is useful as it provides a measure of the average duration of each vigilance scan. You now have one explanatory variable – group size – and three response variables: the frequency of vigilance scans, the total duration of time spent being vigilant, and the average duration of each vigilance scan.

Save your spreadsheet as a Comma-separated Values file (.csv file). Go to ‘Save As’ in Excel and choose ‘Comma-separate Values (.csv)’ from the dropdown list. Once you’ve done so, open RStudio and import your data by running this code: goosedata <- read.csv(file.choose(), header = T).

You should now conduct three separate linear regressions, one for each of your three response variables. In each regression, you will test for an effect of group size on the vigilance behaviour in question. Keep in mind that already you did a regression as part of the introduction to R earlier in the course. You are now doing the same test but on a dataset that you collected yourself. If you need a reminder on how to do a regression in RStudio, please go back the instructions from the introduction to R.

Once you have run your regressions, you should get an output that looks something like this for each regression:

A picture containing text, receipt, screenshot

Description automatically generated

This is an output from an analysis on a dataset that I named ‘GoosePractical’, where I named the response variable ‘Dur’ (total duration of time spent being vigilant) and the explanatory variable ‘GrSize’ (group size). You can obviously choose your own names. Once you have done you own analyses, you need to extract the relevant information for your report. In this case, this is the test statistic (F-statistic), the degrees of freedom (DF), the significance level (p-value), the R-squared, and the regression equation. By convention, we always report the test statistic, the degrees of freedom and the significance (note that there is an exception for correlation and some other tests where the convention is to report the sample size (n) rather than the degrees of freedom). In the case above, the F-statistic is 15.47, the degrees of freedom are 1,14, and the p-value is 0.0015. Make sure that you can identify these in your own output.

For regression, we also report two additional bits of information: the R-squared and the regression equation. The R-squared provides information on how much of the variation in your response variable (here: vigilance behaviour) is explained by your explanatory variable (here: group size). For example, even though group size has a significant effect on vigilance behaviour, it may explain only a relatively small amount of the variation in this behaviour. This is useful to know as it suggests that there are other sources of variation in the behaviour. In this case, the Adjusted R-squared is 0.4911, which means that group size explains 49.11% of the variation in vigilance behaviour. This may not seem like much, but this is a high R-squared for a behaviour. You will see two R-squared values in your output: Multiple R-squared and Adjusted R-squared. The Adjusted R-squared adjusts for improvements in your model should you add additional explanatory variables. This is not a problem in this case because there is only one explanatory variable in your model. This is also why these two values are similar. Because you have only one explanatory variable, it doesn’t matter which of the two you report in this particular case. Pick one and report that one only.

The regression equation describes the relationship between your response variable and your explanatory variable. To find the regression equation, find ‘Coefficients’ in the output, and then find ‘Estimate’ for Coefficient and GrSize. In this case, the values are 35.25 and –1.80, which means that the regression equation is as follows: total duration of time spent being vigilant = 35.25 – 1.80 ✕ group size. ‘Intercept’ describes where the straight line crosses the Y-axis. GrSize describes the regression slope, which is the extent to which your vigilance behaviour changes with a given increase in group size. In this case, the total duration of time spent being vigilant decreases by 1.8 s for each additional goose in the flock.

Once you have extracted the required information, you need to ensure that you report this information professionally in your written report. For further advice on how to do this, see the information provided in the lecture slides from Part 1. Furthermore, you may read the information provided in the document entitled ‘Instructions to Authors’. Finally, keep in mind that you should model your report on scientific journal articles. You can learn a lot about how to write professionally by reading some published journal article. One such article is available on Learn, but you can use your favourite article instead if you prefer to do that.

Finally, you should prepare at least one figure to illustrate your main results. You should prepare your figure(s) using the package ggplot2 in RStudio. The figure(s) should be a simple scatterplot showing the relationship between group size and vigilance behaviour. The figure should include the regression slope as shown below. The figure(s) should be prepared in a professional manner with an informative figure legend placed under the figure, as explained in the lecture slides from Part 1. You should prepare the figure legend in Word afterwards. For further advice, please take a look at how figures are prepared in papers published in a scientific journal.

Below is a scatterplot based on the same dataset that I described above. Please note that I have made some changes to code that you used in the introduction to R earlier in the course. These changes allow me to use the full names of the variables on the X and Y axes and include the unit of measurement on the Y axis. Below is the code I used for the figure. You should copy this code when making your own, but obviously make sure that you rename your code to match the name of your data set and the names of the variables in your data set.

Chart, scatter chart

Description automatically generated

> ggplot(GoosePractical, aes(x = GrSize, y = Dur)) + geom\_point() + theme\_classic() + xlab("Group size") + ylab("Time spent being vigilant (s)") + stat\_smooth(method=lm, se=FALSE)