***Statistics for Humanities Researchers***

**Lab 4: the general linear model**

**FILL OUT THE FOLLOWING EXERCISES. CUT AND PASTE ANY PLOTS YOU’RE ASKED TO PRODUCE. FOR THE PROGRAMMING, MAKE A NEW *R* SCRIPT TO TRACK YOUR WORK.**

**Exercise 1) – One Way ANOVA**

1. In your own words, explain what family wise error is?
2. How does ANOVA avoid having to deal with family wise error?
3. In your own words, explain what an *F* test tells us? What are the differences between an *F* test and a *t* test?

*For the following exercises, run your script from Monday, so that you again have the fuel economy data in your workspace. For the following questions, you will use the data frame you used to prepare the barplots on Monday.*

1. Is there a significant overall effect of region on ***city fuel economy***? Provide the relevant statistics in the correct form for a lab report. (don’t worry about contrasts yet, that’s for the next set of exercises)
2. On Monday you created a barplot that shows the mean ***city fuel economy*** for each region. Reproduce the figure from your code from Monday. Is there any region that clearly stands out from the others (i.e. which group has a mean that looks much higher or lower than the other groups)?
3. Is there a significant overall effect of region on ***how many cylinders are in a car***? Provide the relevant statistics in the correct form for a lab report.
4. Using the same method that you used on Monday, make a new barplot that shows the mean ***number of cylinders*** for each ***region***. What can you say about the relationship between these two variables?
5. Is there a significant overall effect of region on ***engine displacement***? Provide the relevant statistics in the correct form for a lab report.
6. Using the same method that you used on Monday, make a new barplot that shows the mean ***engine displacement*** for each ***region***. What can you say about the relationship between these two variables?

**Exercise 2) – Planned contrasts**

1. In your own words, explain the idea behind planned contrasts? When is it appropriate to run planned contrasts?
2. In your own words, explain the idea behind post-hoc comparisons? When it is appropriate to run post-hoc comparisons?
3. In your own words, explain orthogonality? Why is it important that contrasts are orthogonal?
4. Re-run the ANOVA required to answer 1d), but include contrasts for simultaneously asking i) whether Asian cars are more economical for city driving than American cars and ii) whether Asian cars are more economical for city driving than European cars? Answer i) and ii) separately, and provide the contrast vectors and the necessary statistics for proper reporting. How do these statistics relate to the figure you produced in question 1e?
5. Re-run the ANOVA required to answer 1f), but include contrasts for simultaneously asking i) whether Asian cars have few cylinders than American cars and ii) whether Asian cars have fewer cylinders than European cars? Answer i) and ii) separately, and provide the contrast vectors and the necessary statistics for proper reporting. How do these statistics relate to the figure you produced in question 1g?
6. Re-run the ANOVA required to answer 1h), but include contrasts for simultaneously asking i) whether American cars have greater engine displacement than cars from other regions? and ii) whether Asian and European cars have different engine displacement? Answer i) and ii) separately, and provide the contrast vectors and the necessary statistics for proper reporting. How do these statistics relate to the figure you produced in question 1i?

**Exercise 3) – Factorial ANOVA**

1. In your own words, explain what is meant by “factorial design”?
2. In your own words, explain what is meant by “main effect”?
3. In your own words, explain what is meant by an “interaction”?
4. What are effect sizes? How are effect sizes different to *p* values? Why is it important to report effect sizes?
5. Using the same method you used for question 6 on Monday, make a new variable which codes cars by number of cylinders. Specifically, your new vector should code for “small engines” which are less than 6 cylinders, “medium engines” which are exactly 6 cylinders, and “large engines” which are greater than 6 cylinders.

*For the following 3 questions, run ONE 3 x 3 factorial analysis of variance*

1. Is there a main effect of region on ***city fuel economy***? Provide the relevant statistics for reporting in a lab report. Include the effect size (for effect sizes you can just use the **etasq** function in the **heplots** package
2. Is there a main effect of engine size (your new variable) on ***city fuel economy***? Provide the relevant statistics for reporting in a lab report. Include the effect size
3. Is there an interaction between region and size for ***city fuel economy***? Provide the relevant statistics for reporting in a lab report. Include the effect size
4. Use ggplot to make a bar graph that represents average city fuel economy for each of the three regions across the three engine sizes. From this graph, how would you explain any interactions in 3g)? Report the results of relevant post-hoc Bonferroni corrected t-tests.

*For the following 3 questions, run ONE 3 x 3 factorial analysis of variance*

1. Is there a main effect of region on ***engine displacement***? Provide the relevant statistics for reporting in a lab report. Include the effect size
2. Is there a main effect of size on ***engine displacement***? Provide the relevant statistics for reporting in a lab report. Include the effect size
3. Is there an interaction between region and size for ***engine displacement***? Provide the relevant statistics for reporting in a lab report. Include the effect size
4. Use ggplot to make a bar graph that represents average ***engine displacement*** for each of the three regions across the three engine sizes. From this graph, how would you explain any interactions in 3g)? Report the results of relevant post-hoc Bonferroni corrected t-tests.