# Research Methods and Statistics with R 3

## Week 3 – ANOVA II: Planned vs. *post-hoc* Contrasts

## Introduction to the session

In this session, you will expand on your knowledge of one-way ANOVAs by incorporating contrasts. Contrasts are “follow-up” tests that allow you to further dissect a main effect (or interaction). Contrasts are often necessary when you have a factor (e.g., “condition”) with more than two levels (e.g., “A”, “B”, “C”, “D”). A significant main effect of your factor will **not** tell you which of the levels is greater/less than the other. This necessitates additional tests (i.e., contrasts).

Contrasts come in two flavours: planned/*a priori (Latin: ‘from what is before’*) and *post-hoc.* **Planned contrasts (*a priori*)** are, as the name would imply, planned *before* any analysis has taken place. These would typically be listed in a pre-registration document or analysis plan. **Post-hoc** (Latin: ‘*after this*’) are not planned but are run after the initial ANOVA has been completed and has yielded a significant result. Post-hoc tests are run to further dissect the effect.

Once again, the dataset that we will use for this session is modelled after another study. This time, we will be looking at FMRI data, based on a recent paper published in *Scientific Advances[[1]](#footnote-2)*.

## Background

The **visual word form area** **(VWFA)** is a brain region in the anterior temporal lobe that responses selectively to written words, letter strings, and characters. It is believed to be integral to language acquisition and development.

In 2023, Zhan *and colleagues* published a study reporting that the VWFA in bilingual individuals is actually composed of smaller patches that respond selectively to different aspects of language (reading, writing). Furthermore, while there was considerable overlap between the patches in bilingual English-French speakers, bilingual English-Chinese speakers showed additional non-overlapping small patches selective for Chinese writing.

A close-up of a brain

Description automatically generated

Figure 1- Approximate location of VWFA patches. This figure was taken from Zhan et al., 2023 and actually shows activation to written words in English-French bilingual readers. But it is suitable enough for our purposes, which is to give you an idea of where the VWFA is located (which, incidentally, is very near the fusiform face area).

**In this session, you will use a one-way ANOVA and contrasts to examine whether newly acquired data[[2]](#footnote-3) support these conclusions.** Specifically, you will determine whether characters of different languages (English, French, Chinese, Arabic, and Hebrew) evoke different levels of brain activity in a specific VWFA patch. The data were obtained from bilingual English-Arabic speakers.If a **main effect of** **language is observed**, you will follow up with contrasts to identify which language(s) evokes the strongest response(s) in the brain.

## Learning Outcomes

By the end of this session, you will:

1. Have performed one-way ANOVAs including assumptions checks and **post-hoc contrasts.**
2. Have implemented multiple different methods to correct for multiple comparisons.
3. Be able to describe the benefit of planned vs. post-hoc contrasts.
4. Know how to report the results of an ANOVA including contrasts in APA style.

## Procedure

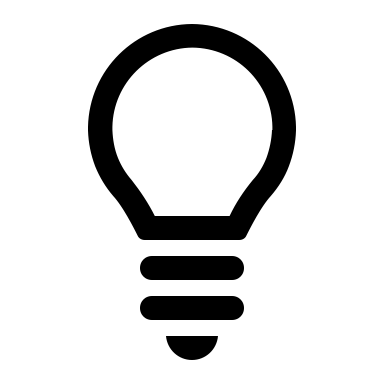
1. Before you do anything, read through the above information and generate some hypotheses for your analysis. For example, what do you believe will be the effect of language? (i.e., will you observe a main effect of language on percent signal change?)

Hypothesis: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

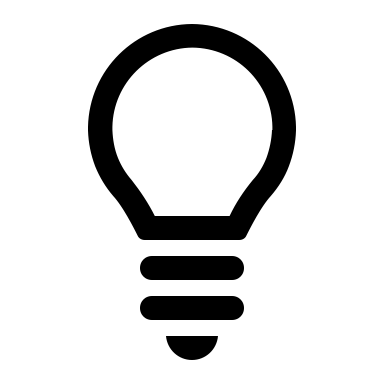
1. Next, generate a set of **orthogonal contrasts**. Remember, the key is to **minimise the number of contrasts** to avoid artificially inflating your family-wise error. Based on the description above, what do you suppose are some reasonable contrasts to conduct? (Note: there are several right answers here). Adapt and complete the following table **before proceeding to the next step**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Level | Contrast 1 | Contrast 2 | Contrast ? | Product |
| Arabic |  |  |  |  |
| Chinese |  |  |  |  |
| English |  |  |  |  |
| French |  |  |  |  |
| Hebrew |  |  |  |  |

1. Now, start a new R script. Be sure to include whatever naming and commenting conventions you feel are appropriate.
2. Download the simulated data from KEATS and load the file into your workspace. You can use **read.csv()** or try **rio::import()[[3]](#footnote-4)**.
3. Inspect your data using an appropriate tool, like **dplyr::glimpse()**, to make sure they have imported appropriately. Pay particular attention to the variable types (i.e., numeric, factor, character, etc.).
4. Compute descriptive statistics for **each language.** Here, you might find it useful to use something like **tapply() , stats::aggregate(),** or **psych::describeBy().** What is your dependent variable? In which column is your independent variable? How many levels does your only factor have?
5. Decide on an appropriate data visualization for such an experimental design and produce it. You will likely find {ggplot2} to be the most useful package for this. There are many suitable options to visualize these data: you can use boxplots, swarm/violin plots, bar graphs, etc. Try out several different types of graphs to see which you feel best informs the reader. See the **KEATS page for Week 2** for links to the R-graph gallery for inspiration and help with the syntax required to generate the various graphs. **You should have successfully produced at least one figure before moving to the next step.**
6. The next step is to test for our assumptions. However, in R - the easiest way to do so is to first generate the **aov() object** so now is a good time to run your aov() command (or other R command of your choice). Consult the lecture material and your script from last week if you are unsure of the proper syntax. Once you have produced your aov() object (by running your ANOVA), you can test for two of the three assumptions for an ANOVA: normal distribution of the residuals and homogeneity of variance. (*Question: what is the third assumption? Why can’t we test for it now?)*

**Tip: Are you getting an error message when you try to run Levene’s test? What does the error message you are getting tell you? Did you perhaps skip an important step? Hint: Check Step #5.**

1. Once you have checked the assumptions and decided about how to proceed based on the results of these checks, you can inspect your ANOVA object.
2. The ANOVA output will tell you if there is a **significant main effect of language**. However, it does not tell you which differences are driving that significant main effect. That is, which pair(s) of conditions are significantly different from each other. To determine this, you will need to run either your **planned contrasts** or **post-hoc comparisons.**

**Recall from the lecture material: You should only do post-hoc comparisons on your data if your main effect/interaction is significant!** If your main effect is not significant, do not conduct any post-hoc pairwise comparisons on those data - Just walk away.

However, for the purposes of this exercise, we will describe how to run both post-hoc and planned contrasts on the same data (something you would NEVER do).

***Post-Hoc Contrasts***

1. To run **post-hoc comparisons**, you have several options:
   1. You could use a function that automatically calculates all pairwise comparisons, such as “**pairwise.t.test()**” or “**lsr::posthocPairwiseT()**” (from Danielle Navarro’s package). These tools have the added benefit of allowing you to specify the correction method for multiple comparisions. For example, ***pairwise.t.test(DV ~ IV, data = dataset, p.adjust.method = “holm”)****.* Check the function documentation to find out how to specify either no correction or Bonferroni correction and compare the results.
   2. You could use individual t.tests() between relevant pairs, being sure to correct your *p*-value manually.
   3. You can use Tukey’s HSD. For example:

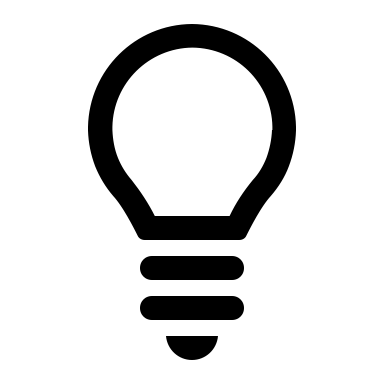
*> postHocs <- glht(anovaObject, linfct = mcp(condition = "Tukey"))*

*> summary(postHocs)*

* 1. Many others…

***Planned Contrasts***

1. To include planned contrasts in your ANOVA call, you must first create a **contrasts** object to run your contrasts (see lecture material).
2. Determine the proper order of the levels (levels(dataset$factor)
3. Assign weights to the levels in R adapting the following syntax:

****contrasts(my\_data$my\_factor) <- cbind(c(x, x, x))

**Tip:** You will need to run aov() again so the weights are saved in the model object.

1. Instead of **summary(my\_model)**, you will need to use **summary.lm(my\_model)** to have an output that takes your planned contrasts into account.

***Effect Sizes***

1. The effect size commonly reported for factorial ANOVAs is ηp2 (partial eta-squared), although η2 (eta-squared) is arguably a more interpretable measure of effect size as it is simply a measure of the variance explained by a factor or an interaction (i.e., the same as *r*2). Thankfully the {*lsr*} package will provide you with both ηp2 and η2 (see Segment lecture material).

## Coding Challenge

Conduct an analysis, including follow-up contrasts, that would be appropriate if the data were not normally distributed. Hint: we discussed one option in the lecture material. What test(s) could you use for the follow-up contrasts?

1. Zhan M, Pallier C, Agrawal A, Dehaene S, Cohen L. Does the visual word form area split in bilingual readers? A millimeter-scale 7-T fMRI study. Sci Adv. 2023 Apr 5;9(14):eadf6140. doi: 10.1126/sciadv.adf6140. Epub 2023 Apr 5. PMID: 37018408; PMCID: PMC10075963. [↑](#footnote-ref-2)
2. Totally fake. [↑](#footnote-ref-3)
3. Q. What are the benefits of rio::import() vs. read.csv()? [↑](#footnote-ref-4)