

8.1: Practical steps to Statistical Modelling

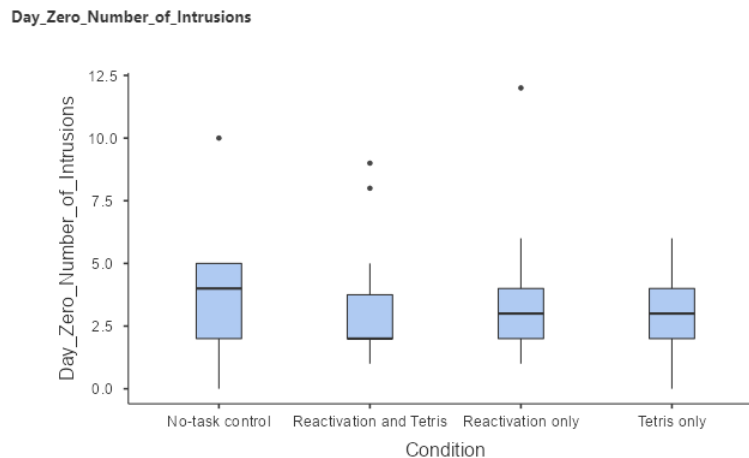


Figure 8.1.1. Box plot of a number of intrusive memories on Day 0 before the experimental task by the different conditions

The boxplot above shows that the intrusive thoughts are relatively similar for all conditions. Remember, we expect that all groups should have the same amount of bothersome memories during the first 24 hours since this is before any changes (Day 0). This is just to ensure that all groups started with a relatively similar baseline.

However, what we really want to test is the effect of the experimental manipulation. In particular, we want to examine whether there is a significant difference between the conditions on the number of memory intrusions in the seven days following the experimental task. We will use the variable named *Day_One_to_Seven_Number_of_Intrusions* and visualize them to see if there are any outliers:

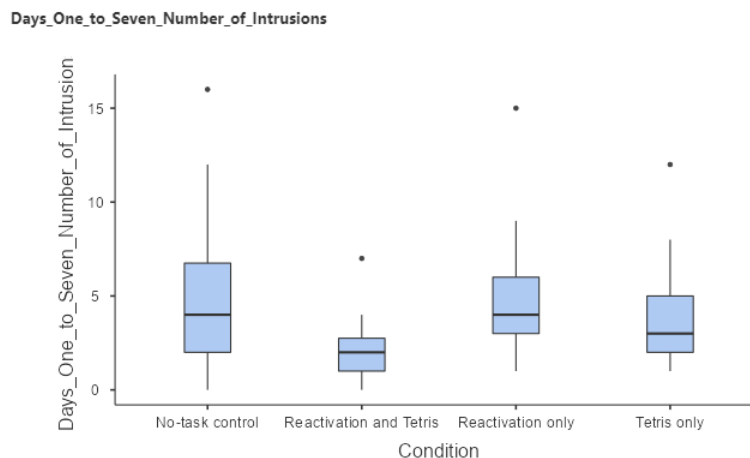


Figure 8.1.2. Box plot of a number of intrusive memories on Day 1 to 7 after the experimental task by the different conditions

Box plots are useful to see the shape of the distributions, as shown in Figure 8.1.2. Those data look fairly reasonable – there are a couple of outliers (indicated by the dots outside of the box plots), but they don't seem to be extreme. We can also see that the distributions seem to differ a bit in their variance, with the reactivation and Tetris showing somewhat less variability than the other groups, while the no-task control has the most variability. This means that any analyses that assume the variances are equal across groups might be inappropriate. Fortunately, the statistical model that we plan to use is fairly robust to this.

Step 4. Determine the Appropriate Model

There are several questions that we need to ask in order to determine the appropriate statistical model for our analysis.

Figure 8.1.3. Results of the General Linear Model (jamovi screenshot)

Note that the software automatically generated dummy variables that correspond to three of the four conditions, leaving the no-task control without a dummy variable. This means that the intercept represents the mean of the no-task control condition, and the other

three variables model the difference between the means for each of those conditions and the mean for the no-task control condition. No-task control condition was chosen as the unmodeled baseline variable simply because it is first in alphabetical order.

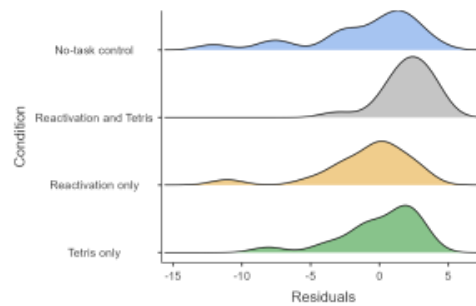


Figure 8.1.4. Distribution of residuals for each condition

Another important assumption of the statistical tests that we apply to linear models is that the residuals from the model are normally distributed. It is a common misconception that linear models require that the *data* are normally distributed, but this is not the case; the only requirement for the statistics to be correct is that the residual errors are normally distributed. The right panel of Figure 8.1.5 shows a Q-Q (quantile-quantile) plot, which plots the residuals against their expected values based on their quantiles in the normal distribution. If the residuals are normally distributed then the data points should fall along the dashed line – in this case, the plot doesn't look the best. What we want is for the residuals (denoted by the dots) to be tightly packed around a line (in other words, linear). However, given that this model is also relatively robust to violations of normality, we will go ahead and continue with our analysis.^[3]

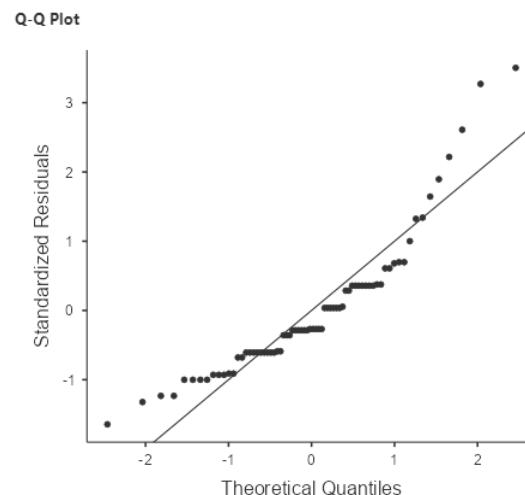


Figure 8.1.5. Q-Q plot of actual residual values against theoretical residual values

Figure 8.1.6. Fixed effects parameters estimates table for the dummy coded variables (jamovi screenshot)

From the table above, we can see that the frequency of intrusive memories for participants under the no-task control and reactivation-only conditions was significantly different from the reactivation task with the Tetris condition.

Post-Hoc Comparisons

For the following analysis, we will differ from the original paper to show you how you would conduct the analysis if they did not provide a specific hypothesis.

Because we are doing several comparisons, we also need to correct those comparisons, which is accomplished using a procedure known as the Tukey method, which can be requested by going into the Post Hoc Tests, putting the condition into the variable window and checking Tukey under correction.

Post Hoc Tests

Post Hoc Comparisons - Condition						
Comparison		Difference	SE	t	df	Ptukey
Condition	Condition					
No-task control	- Reactivation and Tetris	3.22	1.06	3.04	68.00	0.017
No-task control	- Reactivation only	0.28	1.06	0.26	68.00	0.994
No-task control	- Tetris only	1.22	1.06	1.15	68.00	0.657
Reactivation and Tetris	- Reactivation only	-2.94	1.06	-2.78	68.00	0.034
Reactivation and Tetris	- Tetris only	-2.00	1.06	-1.89	68.00	0.242
Reactivation only	- Tetris only	0.94	1.06	0.89	68.00	0.809

Figure 8.1.7. Post-Hoc Tests between the different conditions (jamovi screenshot)

The column titled Ptukey in the rightmost column shows us which of the groups differ from one another, using a method that adjusts for the number of comparisons being performed. Anything below the p-value of .05 is significantly different from one another. This shows that the pairing of no-task control and reactivation and Tetris as well as reactivation and Tetris and reactivation only were the only pairs that significantly differ from one another.

What about Possible Confounds?

If we look more closely at the James et al. paper, we will see that they also collected data on attention paid to the film. Let's plot this data on a bar plot for each condition.

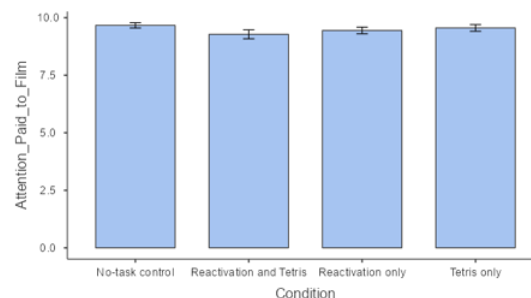


Figure 8.1.8. Barplot of attention given to the film per condition

Looking at the data it seems that the rates were consistent across the conditions. If the data is quite different across groups, then we may be concerned that these differences could have affected the results of the intrusive memory outcomes. In our case, this is not an issue. However, it is also good to check potential confounding variables that may be affecting your data.

Getting Help

Whenever one is analysing real data, it's useful to check your analysis plan with a trained statistician, as there are many potential problems that could arise in real data. In fact, it's best to speak to a statistician before you even start the project, as their advice regarding the design or implementation of the study could save you major headaches down the road. Most universities have statistical consulting offices that offer free assistance to members of the university community. Understanding the content of this book won't prevent you from needing their help at some point, but it will help you have a more informed conversation with them and better understand the advice that they offer.

Chapter attribution

This chapter contains material taken and adapted from *Statistical thinking for the 21st Century* by Russell A. Poldrack, used under a CC BY-NC 4.0 licence.

Screenshots from the jamovi program. *The jamovi project* (V 2.2.5) is used under the [AGPL3](https://www.gnu.org/licenses/agpl-3.0.html) licence.

1. James, E. L., Lau-Zhu, A., Tickle, H., Horsch, A., & Holmes, E. A. (2015). Playing the computer game Tetris prior to viewing traumatic film material and subsequent intrusive memories: Examining proactive interference. *Journal of Behavior Therapy and*

Experimental Psychiatry, 53, 25-33. <https://doi.org/10.1016/j.jbtep.2015.11.004> ↩

2. This example came from OpenStatsLab. For more practical exercises such as this one, visit:

<https://sites.google.com/view/openstatslab/about> ↩

3. Some may argue that these violations suggest that we should not fit the GLM in our data. This is fine – we can instead conduct a Generalised Linear Model if you are concerned about these violations. Another option is to conduct the non-parametric equivalent of ANOVA, which is the Kruskal-Wallis test. ↩

This page titled [8.1: Practical steps to Statistical Modelling](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Klaire Somoray](#) (Council of Australian University Librarians Initiative) .