Revisoin Week 3

## Familywise Error Increases with Number of Categories

A graph with a line

Description automatically generated with medium confidence

That results in a familywise error of .659

* So if we went with all possible contrasts, we’d have a 66% of getting at least one false positive!

## Methods for correcting for multiple comparisons

* Anytime you subject the same data to multiple tests, you should correct for multiple comparisons
* There are several methods available:
  + **Bonferroni**
  + **Holm-Bonferroni**
  + **Tukey’s HSD (Honestly Significant Differences)**
  + False Discovery Rate
  + Cluster-based correction
  + Threshold-Free Cluster Enhancement (TFCE)
  + And so many more…
* We will discuss the most common

### Bonferroni

A screenshot of a math method

Description automatically generated

The Bonferroni method simply scales your critical value by the number of contrasts.

So let’s say your initial alpha, that is the value of p that you will deem statistically significant is 0.05 and you are running 10 post-hoc comparisons, your NEW CORRECTED alpha will be: 0.05 / 10 or 0.005.

So for each pairwise comparison, the p-value must now be **below** 0.005.

* The Bonferroni method is very popular because it is easy to implement.
* However, it is very conservative, especially when correcting for 1000s if not 1,000,000s of contrasts like you would be when analysing FMRI data.

### Holm-Bonferroni

A screenshot of a math method

Description automatically generated

Holm introduced a modification whereby the critical p-value for each contrast was determined separately. This was done by dividing the Bonferroni-corrected alpha by the ranking of the current contrast.

So what this means is that each contrast is “penalized” differently, with the contrast with the smallest p-value – that is the one that is “most” significant - being penalized most harshly.

For example, let’s say we have 5 contrasts that give us five different p values. We rank those p values from lowest to highest. Then, we calculate a different p-critical value for each using this formula. In the case of the first contrast, we divide 0.05 by k, which is the total number of contrasts (which in our hypothetical example is 5) minus j (which is the rank of the current p-value, so for our first example that is 1 and then add back 1. So that is 0.05 / 5. The next p-value would have a critical value of 0.05 / 4, and so on.

What this means is the critical p’s will get closer and closer to 0.05 as our p-values get larger and larger.

### Tukey’s Test

The last example that we will look at is Tukey’s HSD or Tukey’s Test or Tukey’s Range test.

In essence, Tukey’s runs pairwise t-tests for all possible pairings BUT it uses a special t-distribution when determining the p-value.

Briefly, this is a modified version of the Student t-distribution (the one that is used for standard t-tests) but constrained by the range of your different group means. So instead of a t-distribution that would extend well beyond the useful range of values, this “Studentized distribution” is specific to your dataset. In effect, it means it’s “easier” to get a statistically significant result.

This is a true post-hoc test as it is the one method where you must use all pairwise comparisons. Bonferroni and Holm methods can theoretically be applied to a subset of pairwise comparisons but Tukey’s must include all comparisons to work properly.

Also – it is more sensitive to violations of the equal variances assumption of ANOVA.

So if your Levene’s test is significant, do not use Tukey’s HSD.

A screenshot of a test

Description automatically generated

It is unfortunately not possible to use pairwise.t.test to run Tukey’s so we need to introduce some new code and syntax.

Here we will use glht() (which stands for general linear hypothesis testing) which is part of the multcomp library to run our correction.

We need to feed this function our ANOVA object (or LM object) and set the linear function – or rather the “hypothesis to be tested”, which in this case is the multicomparison across decade using Tukey’s.

Conveniently or inconveniently, depending on your perspective, this method spits out the t-stat and p-value for each possible pairwise contrast, along with significance codes.

And that is Tukey’s.

### Importance of Orthogonal Planned Contrasts

A table with blue and white text

Description automatically generated

* In the previous examples, you may have noticed we had an additional column called “Product”, which contained the product of all weights for a given group
* When the **TOTAL** of this product column (marked in GREEN above) is zero, the contrasts are said to be ***orthogonal***to one another – meaning they are essentially **independent** from one another.
* This is desirable because it justifies our NOT correcting for multiple comparisons. It basically means that the probability of getting a significant effect in one contrast *has no impact* on the probability of getting a significant effect in another contrast.
* **If the contrasts are not orthogonal, they are not independent and therefore you should somehow correct for multiple comparisons**

Parametric VS Non-Parametric Tests

* A **parametric test** (e.g., t-test) assumes the data are distributed in a certain way. In most “standard” cases, it assumes the data follow a normal distribution with two parameters: mean, 𝜇 and standard deviation, 𝜎.
* A **non-parametric test** makes no assumptions about the underlying distribution of the data (that is, it is *parameter-free*). It often uses *relative rankings*

A screenshot of a test

Description automatically generated

# TODOs

./