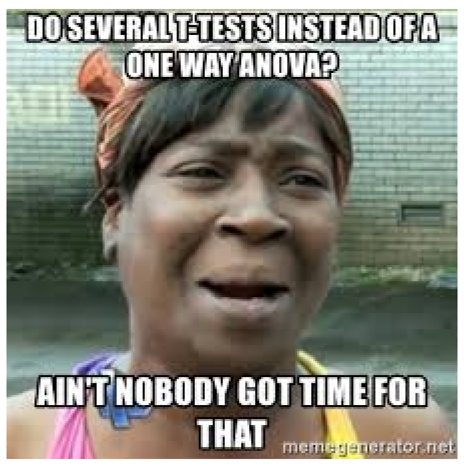
Welcome to Week 4!

Welcome to Week 4!

Last week we learnt all about the independent t-test. We remember that it was a statistical test that was used to compare the means of two groups that have continuous data. SPSS however, has the ability to analyse more than two groups using the independent samples t-test but what it actually does is multiple t-tests and comparisons for each pair of groups (aka a pairwise comparison). In practice, this is not only cumbersome and time-consuming, it also involves a higher chance of making a type I error (rejecting the null hypothesis when it is true). So how do we compare the means of 3 or more groups in an ideal way? We use what's called a **One-Way ANOVA**. ANOVA is short for "Analysis of Variance" and it allows us to compare the means of multiple groups using only one test at a 5% (alpha = 0.05) error rate.

Let me explain using some basic probability: Suppose you have a 20 cent coin, well you know the probability of getting a heads or tails in one coin flip is 50% or 0.5. What about the probability of getting two tails (no heads) in a row for two coin flips? This would be 0.5 x 0.5 = 0.25. What about the probability of getting 1 or more heads in two coin flips? This would be the total probability - probability of no heads, so 1 - 0.25 = 0.75. Suppose now that we wanted to do 4 coin flips then the probability for getting at least one or more heads would be 1 - (0.5 x 0.5 x 0.5 x 0.5) or 1 - (0.5)^4 = 0.94 ie. 94% of the time you will likely get at least one heads in 4 coin flips.

Let's now translate this probability example into a statistical t-test. We should now understand that by convention an alpha = 0.05 confidence level is usually chosen. Here we are saying that if the null hypothesis is true, the probability of not getting a significant result is 0.95 (ie. p > 0.05). So this value of 0.95 needs to be multiplied by the number of tests we conduct using a t-test alone. For example: if we had 4 groups, Group A, Group B, Group C and Group D then we would need to perform 6 t-tests (A vs B, A vs C, A vs D, B vs C, B vs D and C vs D), this means we have 0.95^6 = 0.735, which is then subtracted from 1 giving us 0.265 or 26.5% error rate. Compare this to doing one One-Way ANOVA test at a 5% error rate and we see that using ANOVA is a much more efficient method for the comparison of means for 3 or more groups. In many ways, I feel this meme sums up what I am trying to say!



With ANOVA comes a new set of terminology namely **"factors" and "levels"**. Factors are independent variables that change the outcome of the study. Levels are simply the values that factors can have. A classic example is the classification of human age groups. The factor here is age and the levels are Child (0-12 years), Adolescence (13-18 years), Adult (19-59 years) and Senior Adult (60 years and over).

There are many variants of the ANOVA test but in this unit we will just focus on the most important one, which is the **One-Way Between Groups ANOVA**. Other variants include Two-Way ANOVA, Repeated Measures ANOVA, ANCOVA and MANOVA however, these are perhaps a bit more advanced.

For an independent t-test, if we find that after conducting the test we have statistical significance, it is always important to state in the reporting which of the two means is higher. How would you do this if there were multiple means? You would need to use what's called a **Post-Hoc test**. There are two types of post-hoc tests that we will cover in this unit, these are: **Bonferroni's test and Tukey's test.** We will expand briefly on these in this week's lecture.