ANALYSIS PLAN:

1. Group level – NetworkStats – stores all group level stats (Network\_stats.xlsx).  
   - Check if there are any significant differences in the statistical measurements between the smoking session and the sober session. Using the Wilcoxon signed-rank test (also called the Wilcoxon rank-sum test), which is a non-parametric test for independent data, to compare the two conditions. Analysis class stores statistic value, p-value, effect size for each network statistic in sober\_vs\_cannabis\_group\_level.xlsx.  
   - Check if the task order influences the data collected from the two sessions. This can be done at the group level because we’re comparing the effects of starting with TAP vs. FAST across subjects. Use the Wilcoxon signed-rank test here. Analysis class stores statistic value, p-value, z-value, effect size for each network statistic in TAP\_vs\_FAST\_group\_level.xlsx  
   - Check if smoking in session 1 or session 2 influences the data. This is also done at the group level because we’re comparing the effects of smoking in session 1 vs. session 2 across subjects. Use the Wilcoxon signed-rank test here. Analysis class stores statistic value, p-value, z-value, effect size for each network statistic in smoking-ses-1\_vs\_smoking-ses-2\_group\_level.xlsx  
   - Check if the start task and the session in which the participant was assigned to smoke together do not influence the data collected. This is analysed at group level because we’re looking at the interaction of two variables (start task, and smoking session). Us a non-parametric version of the two-way ANOVA, like the Scheirer-Ray-Hare test. This test allows us to observe the main effects of each independent variable and their interaction. We might be able to remove the second and the third conditions mentioned above since this test might already be analysing the same along with the interaction (working on implementing this today (July 28th)).  
   - Analysis class also calculates the group level descriptive statistics.  
   - Can use each box plots or violin plots to visualize all these conditions. Can also use bar graphs to compare all the test statistic values for each network statistic (x-axis = test statistic (p-value, effect size…))
2. Subject level - NetworkStats – stores all subject level stats (‘sub(sub)\_ses-(ses)\_(smoking/sober).xlsx’) in stats/subject\_level\_analysis folder.   
   - Check if there are any significant differences in the statistical measurements between the smoking session and the sober session. Using the Wilcoxon signed-rank test (since its paired data) which is an alternative to the independent sample t-test. Analysis class stores the U-statistic, p-value, and effect size for each network statistic per subject. This data will be stored in an excel file that has the subject id, smoking session, and test statistics.  
   - We can use box plots here to visualize the smoking and sober session statistics (x-axis = network stat (like degree, centrality…)).

CHANGES IN CODE

* Added functionality to add participant info from participant\_log.xlsx to stats data excel file (to use in group level analysis)- main.py
* Calculate individual subject network statistics (for subject-level analysis) – network\_stats.py
* Save individual subject network statistics in stats/subject\_stats. Added functionality to include participant smoking session info to file name – main.py
* Added functionality to store all estimated network .gexf files in sats/networks folder with all file names containing subject#, session, smoking/sober, and start task info – main.py.
* Added a new module (analyses.py) that contains the Analyses class to implement group-level and subject-level statistical analyses.
* Add new module called plot\_generator.py which will contain the PlotGenerator class with implementation to create all the plots for analyses (Working on this tomorrow (July 29th)).

NETWORK STATISTS:

- Degree centrality - Calculates the number of direct connections a node has. Helps identify important hub nodes in the network and examine the structure of the network to understand how it varies in the two conditions. High degree centrality = central theme

- Betweenness centrality - It is a measure of the number of times a node lies on the shortest path between other nodes. Could help identify thoughts that influence the flow of other thoughts. High betweenness centrality = a thought that connects or mediates other thoughts

- Closeness centrality - Measures the distance of a node from all other nodes in the network. a thought with high closeness centrality measure would be connected to many thoughts suggesting the participant's readiness to transition to other themes.

- Clustering coefficients - It is a measure of the extent to which nodes linked to a node are also interconnected. a high clustering coefficient would indicate that thoughts are tightly linked to one another, indicating a convergent thought pattern.

- Community detection algorithms - can help detect clusters of thought units that co-occur frequently (changes in composition and structure of these clusters could indicate changes in thought patterns).

- Measurements like the diameter or density help to characterize a network structure in comparison with other networks.

- The number of nodes can indicate the qualitative variety of present words in semantic fields and thus a higher variety in language expression.

- Nodes can quantitatively be characterized by degree which indicates the number of adjacent nodes and measures a node’s connectedness. As nodes represent words, the number of adjacent nodes also provides qualitative insights in the diversity of a word’s usage referring to the position in the underlying semantic network model. Words with a high degree imply a highly differentiated relatedness and typically indicate local hubs in a semantic network.

- Paths can be used to determine possible (shortest) connections between two nodes which help to find out how nodes are connected to each other. If a path exists between two nodes, we can state for a semantic network that there is either a direct or an indirect semantic relation on word basis.

ATATISTICAL ANALYSIS THEROY:

Why do we need statistics?

- Inferential statistics - Refers to the kinds of statistical tests that are used to determine if a particular finding is statistically significant (p-value, F, t, etc.) - helps us to know which results to interpret

- Descriptive statistics - It is the characteristics of sets of numbers (mean, variance, spectrum)

- The basis of all inferential statistics boils down to the same fundamental quantity which is a ratio.

- If you have two groups - the inferential statistics is the difference of central tendencies/ widths of distributions -> signal/noise ratio.

- So, if you want statistically significant effects, you need to maximize the signal:noise ratio (get the signal to be big or the noise to be very small or both).

- For example, if you want to study people who have Parkinson's disease who are also diagnosed with schizophrenia, you'll end up with a fairly small sample size. So, you can't acquire a huge amount of data and there might also be a lot of variability among these patients which you won’t have any control over. In other words, the noise will be very large, and you won’t have a lot of control over it. Therefore, you need to design the experiment knowing that the noise is going the be very large and so, your experiment needs to be able to give you really big signals in each trial.

- In contrast, imagine you are doing in-vitro experiments studying slices of brain tissue. That way it is set up, you can get as much data as you want and since it’s an in-vitro experiment, you'll have a lot of control over the various environmental variables. So, you can control the noise to be really really small and so, you can afford to design an experiment that yields a very small effects size because the noise is going to be even smaller.

How do we know if the signal-to-noise ratio is big enough to interpret?

- We need to first figure out a way of defining a distribution of signal-to-noise values or test statistic values that we expect under the Null Hypothesis (H0 - the hypothesis that there is no effect).

- There are different ways to estimate the distribution of the signal-to-noise ratios would be under H0 which can be done parametric, non-parametric methods

- The idea is that you have your observed signal-to-noise value, and you compare it to the null hypothesis SNR values, and you come up with some threshold. Now if the observed SNR value is greater than the threshold, then you would call this statistically significant.

"Levels" in statistical analyses

- A level in a statistical analysis is referred to as the grouping at which you are collecting data

- For example:

- Level 1: Trial (smallest unit of meaningful data collected)

the lowest level associated with the most fundamental aspect of data acquisition.

A lot of within subject analyses (statistical analyses done in each individual data set) is called level 1 analyses.

- Level 2: Subjects [Most of the inferential statistics are done at this level]

500 trials and 20 subjects each doing 500 trials.

At this level, we usually want to test whether the findings that we observe are consistent across a group of individuals. The idea is that, if the findings are consistent across our sample, then maybe we the findings are also consistent over a larger population.

- Level 3: Groups [SOL here TAP vs. FAST]

If you have multiple groups. 40 subjects in total. 20 in group A and group B

- Level 4: Cultural

You might want to see the effect of being in group A or group B on how the individuals are doing the task depends on say whether they are from an Asian or Western culture.

- The recommended sample size N is larger at lower levels.

- If there is some kind of behavioural or cognitive component, its reasonable to expect that there is going to be a lot of variability at the level of individual trials. Therefore, we want to have a lot of data at level 1 because the variability will be very high. And then we expect less variability at level 2 and even less variability at level 3.

Parametric Statistics and Non-Parametric Statistics

- You have your observed SNR value. How do you know if this value is big enough?

- The observed SNR value is big enough if an effect size of that size is unlikely to occur given a distribution of null hypothesis values

- Now the question is, where does this distribution come from or how do we get this distribution to compare against our observed value?

- Essentially, we have two was of generating this distribution:

- Parametric: We can generate it based on a mathematical formula which would be an analytic solution where you rely on several assumptions about the underlying population. This is called parametric because we are computing parameters based on analytic formulas that we have.

- Nonparametric: Or we can generate the distribution based on null hypothesis empirical distributions and this generally comes from permutation-based approaches. This is referred to as nonparametric because we do not have specific parameters that we are evaluating based on a formula. Instead, we are generating these distributions empirically based on the actual data that we have and simulating situations that could arise under the null hypothesis.

What are the assumptions underlying parametric statistics?

- Assumption that allows to validly use a particular formula to evaluate against an observed test statistic

1. Data is sampled randomly from a population

- to avoid bias in data sampling measure

- this assumption is sometimes met and sometimes not

2. That population has known parameter distributions ()

- Here the parameters refer to the key descriptive statistics like the mean, variance, shape of spectrum etc.

- This assumption is sometimes met but sometimes not, it depends on the analyses

- Some distributions can be transformed into known distributions

3. Variance across repeated measurements is homogeneous

- This means that if you take repeated measurements of some system like the brain for example, that the variance across different measurements should be the same or homogeneous

- This assumption is sometimes met but sometimes not met.

4. Measurement error and sample variability are independent across different measurements

- This means that if you measure two different individuals from the population, they should be independent of each other, their measurement error, their sampling variability should be independent.

- This assumption is often not met when doing large-scale recordings

- Do spectral data meet the assumptions that are required by parametric statistics?

- Sometimes yes, sometimes no

What are the assumptions underlying nonparametric statistics?

- It turns out that nonparametric statistics arguably do not have very strong assumptions

1. There exists a null hypothesis

- Sometimes it is a little bit difficult to think about what exactly the null hypothesis is and how to translate that null hypothesis into a concrete situation that we can construct to get an empirical null hypothesis distribution

2. You have sufficient data from which to create a situation that would arise under the null hypothesis

- For example, if you only collect data from only 3 trials, you won’t be able to create a situation that could arise under the null hypothesis, and you won’t be able to do permutation testing with 3 trials

Permutation-based statistics

- Advantages:

- Don't need to check assumptions for parametric statistics

- Ability to deal appropriately with multiple comparisons testing

- Mechanism of permutation testing

- 7 measured data sets and 2 conditions.

- Group A: 1, 2, 3, 4

- Group B: 5, 6 ,7

- first step is to pool all the data together and pool them all into one condition

- Nest step is to randomly reassign these conditions to different data points. The number of conditions and the number of elements in each condition does not change. What does change is that we are randomly remapping the condition label to the data point. Now the shuffled data might look like this:

- Group A: 1, 4, 5, 6

- Group B: 2, 3, 7

- Next, you take your shuffled data and compute your test statistic (t-test, correlation coefficient...). You generate a statistical value. For example, if you used a t-test, your statistical value would look something like this $$Statistic Value = {mean\_A - mean\_B \over sd\_A sd\_B}$$

- What would you expect the value of this test statistic to be?

You expect that this value will be zero because you've randomly assigned each data point to be in condition A or condition B. So, the means for each condition should be very similar since its just total random shuffling.

In reality, the test statistic is probably not going to be exactly zero because we are randomly shuffling so we can get larger values in A and smaller ones in B making the numerator bigger than zero just by chance, and that ok.

- We now take the test statistic value, and we build up a distribution. This is a test statistic value that I can expect under a null hypothesis.

- If the null hypothesis is true, then we can expect that the conditions A and B are meaningless, so they are randomly assigned to the different data points

- After this, we go back to the shuffled data and we re-shuffle the labels again. We come up with a new random assignment of labels to data points which will give us a new condition mapping. Then we generate a new test statistic value and add it to the distribution.

- We repeat this many many times (100s or even 1000s of shuffles) to build up a distribution.

- Then, we will have out empirical null hypothesis distribution.

- Finally, we go back to the original data set and calculate the observed test statistic. Then you check where this observed test statistic lies in our empirical null hypothesis distribution. If the observed test statistic is far enough away from the centre of the distribution, then it is statistically significant, and we can safely interpret this difference between the conditions.

- Alternatively, if the observed test statistic ends up being close to the centre of the distribution, then we would say that we are likely to observe a test statistic value like this in the null hypothesis data where we did random shuffling. So therefore, we cannot interpret the observed test statistic as being statistically significant.

- How to find out if you are far enough away from this null hypothesis distribution?

- Generate p-values

- One way to compute a p-value in empirical null hypothesis testing is to create a Z-value (give us a normalized distance in standard deviation units of the observed statistic value away from the centre of the null hypothesis distribution). Then we convert the Z-value to a p-value. Larger the magnitude of the Z-value, smaller the p-value.

- This Z-value formula is appropriate only for approximately Gaussian null hypothesis distributions because the mean (E[H0]) of a distribution and the standard deviation (std[H0]) are sensible metrics only for a roughly Gaussian distributions. $$ Z = {ObsStat - E[H\_0] \over std[H\_0]} $$

- If our data does not conform to this rough Gaussian distribution or if you don't prefer the Z-value method, a second way to get a p-value is to calculate P\_c (the c here is for count).

- In this method, you are counting the number of null hypothesis tests statistics that were larger (or smaller) and more extreme than the observed test statistic value. $$ p\_c = {\sum(H\_0 > ObsStat) \over N\_{H\_0}} $$

- P\_c is the sum of all the null hypothesis test statistics that are greater (or lesser) than the observed test statistic over the total number of null hypothesis test statistics. This formula is generally appropriate (be careful of which tail).

- After every iteration, you will end up with a different z-value. And there isn't really a way to determine how many iterations is enough. This leads to the idea of doing a meta permutation test according to which you repeat your permutation tests multiple times, and you average the x-value over all of those individual permutation tests which will quickly converge to a central which presumably reflect the true underlying z-value.

- Types of Nonparametric tests

- Sign test - One sample (numerical variable) or Matched pairs (sample the same subjects before and after some kind of intervention

- Wilcoxon signed rank test - One sample or Matched pairs

- Mann-Witney (Wilcoxon rank sum) - Two independent samples (male vs female on some measure or sober vs cannabis)

Subject- vs. group-level analyses

- Clarify the distinction between single-subject and group-level effects

- level 1 statistical analyses - subject level

- level 2 statistical analyses - group level

- Situation 1: Subject effects and group effects

- We hope for this

- Situation 2: Subject effects and no group effect

- Each individual has an effects parameter with error bars that are non-overlapping with zero. So, there is statistically significant effect in the subject level.

- No group level significant effect if individual data sets do not have a consistency in the direction of the effects across the population

- Situation 3: No subject effects and group effects

- Direction of effects across individuals is very consistent so significant group level

- No significant subject level effect

- Single-subject statistics are looking at the variance of data over trials. So, if you have an effect that has a lot of variances, it might be non-statistically significant. Does no allow for population generalization. Generally, requires a large effects size or a very small variance.

- Group-level statistics is looking for consistency of effects across the population (average over different subjects). Hence, you can have a lot of variances within individuals but if the means are all on the same side of zero then the group level statistics are significant. It does allow for population generalization. Tends to be more sensitive to small effects sizes.

- linear mixed effects models or hierarchical linear models are statistical methods that combine the within subject variance and the across subject variance (they take the variance of the individual trials into account at group level inferences). however, it’s not so frequently used.

Variables:

- Categorical variables: you can count the individual entities in each category (male, female) and represent the sum and % composition in a table or visualize it in a bar graph

- Numerical variables: can be represented on a number line, their descriptive statistics can be represented in a table (Range, IQR, mean, median, standard deviation) and can be visualized in a box plot.

- One categorical

- What we observe in our sample give this combination of data types: Gender

- What statistical tests can be applies to find out significance: 1 sample proportion test

- Two categorical

- What we observe in our sample give this combination of data types: Gender and age

- What statistical tests can be applies to find out significance: Chi squared test

- One numeric

- What we observe in our sample give this combination of data types: Height

- What statistical tests can be applies to find out significance: t-test

- One numeric and one categorical

- What we observe in our sample give this combination of data types: Height and Gender

- What statistical tests can be applies to find out significance: t-test or ANOVA (if more than 2 categorical variables)

- Two numeric

- What we observe in our sample give this combination of data types: Weight and height

- What statistical tests can be applies to find out significance: correlation test

Choosing a statistical test:

1. Your question - Compare a difference or check a relationship

2. The scale of your data - continuous (numeric variable), ordinal (categories), or nominal

3. Assumptions for parametric tests - normality (use nonparametric tests is n is small <30)

4. The experimental design - e.g., paired, or unpaired design

5. Number of groups - one, two, or more groups