# Project Proposal

Analysis 1: Estimating the Average Anxiety Level in the Population (with p-value)

Research Question: What is the average anxiety level among individuals in the dataset, and is it significantly different from a known population anxiety level

Rationale: Estimating and testing the average anxiety level provides insights into the mental wellness of the population, which could inform health interventions. If a typical population anxiety level is known, we can compare it to our sample's mean to assess if there’s a significant difference.

Variables and Exploration Methods:

Variable: “anxiety\_level” (measured on a scale from 1 to 10).

Exploration: A histogram will display the distribution of anxiety levels, and a box plot will identify any outliers that might influence the mean.

Justification for Visualizations: Histograms provide a clear picture of the distribution of anxiety levels, and box plots highlight any extreme values that may impact the mean.

Analysis Plan:

Step 1: Calculate the sample mean of anxiety levels to obtain a central estimate.

Step 2: If we have a hypothesized population mean (e.g., an average anxiety level from prior studies or national data), I will perform a one-sample t-test to see if the sample mean is significantly different from this hypothesized mean.

The result will include a p-value, which will help us determine whether to reject or fail to reject the null hypothesis.

Null Hypothesis (H0): The average anxiety level in the population is equal to the hypothesized mean (e.g., no difference).

Alternative Hypothesis (H1): The average anxiety level in the population differs from the hypothesized mean.

Interpretation of p-value:

A p-value less than 0.05 would indicate a statistically significant difference, suggesting that our sample's anxiety level is meaningfully different from the population average.

If the p-value is greater than 0.05, we would not reject the null hypothesis, implying that any difference in anxiety levels might be due to random variation.

Assumptions:

The sample is representative of the population we’re comparing against.

The data for anxiety level is approximately normally distributed for the one-sample t-test to be valid.

Hypothesis and Expected Results:

Hypothesis: If the hypothesized mean is known, I expect the average anxiety level to be close to it, with minor random variation.

Relevance: This p-value-based analysis provides a statistical test to see if the sample’s anxiety differs from a known benchmark, answering the research question by confirming or refuting significant deviation from the population average.

Analysis 2: Investigating the Association Between Mask-Wearing Frequency and Perceived Wellness

Research Question: Is there an association between the frequency of mask-wearing and individuals’ perceived wellness?

Rationale: Understanding the link between mask-wearing behavior and wellness can provide insights into the psychological effects of health behaviors.

Variables and Exploration Methods:

Variables: “mask\_wearing\_frequency” (categorical: Never, Sometimes, Often, Always) and “perceived\_wellness” (scale 1–10).

Exploration: I’ll use a bar plot to compare average wellness scores across mask-wearing categories and a histogram to check the distribution of wellness scores.

Justification for Visualizations: Bar plots help visualize differences in wellness across mask-wearing frequencies, and histograms reveal the distribution of wellness scores, allowing us to assess normality.

Analysis Plan: After calculating the mean wellness score for each category, I’ll conduct a hypothesis test to determine if there’s a significant association between mask-wearing frequency and wellness. Using a p-value from a chi-squared test (or ANOVA if appropriate), I can test the null hypothesis that mask-wearing frequency and wellness are independent of each other.

Null Hypothesis (H0): There is no association between mask-wearing frequency and perceived wellness; any differences in wellness across categories are due to chance.

Alternative Hypothesis (H1): There is an association between mask-wearing frequency and perceived wellness.

Interpretation of p-value:

A p-value less than 0.05 would suggest rejecting the null hypothesis, indicating a statistically significant association between mask-wearing and wellness.

If the p-value is greater than 0.05, we would not reject the null hypothesis, suggesting that any observed differences in wellness scores could be due to random variation.

Assumptions:

Sample sizes within each mask-wearing category are sufficient to ensure reliable p-value calculations.

The wellness scores are accurately reported without significant biases.

Analysis 3: Comparing Stress Levels Across Education Levels

Research Question: Is there a statistically significant difference in stress levels across different education levels?

Rationale: Identifying stress variations by education level could help tailor mental health programs to those experiencing higher stress.

Variables and Exploration Methods:

Variables: “stress\_level” (scale 1–10) and “education\_level” (categorical: High School, College, Bachelor’s, Master’s, Doctorate).

Exploration: I’ll use side-by-side box plots to compare stress distributions for each education level, showing the spread and central tendency within each group.

Justification for Visualizations: Box plots make it easy to compare distributions across groups, providing insights into both the central values and variability within each education level.

Analysis Plan: I will perform a one-way test to determine if mean stress levels differ significantly across education levels. The p-value from this test will allow us to decide whether to reject or accept the null hypothesis.

Null Hypothesis (H0): There is no difference in mean stress levels across different education levels; any observed differences are due to chance.

Alternative Hypothesis (H1): There is a difference in mean stress levels across different education levels.

Interpretation of p-value:

If the p-value is less than 0.05, we will reject the null hypothesis, suggesting that there is a statistically significant difference in stress levels across education levels.

If the p-value is greater than 0.05, we will not reject the null hypothesis, indicating that any observed differences could be attributed to random variation.

Assumptions:

Stress level data within each education level is approximately normally distributed, and variances are similar across groups, which are assumptions of test.

If variances differ or data deviates significantly from normality, bootstrapping methods may be applied as an alternative.