Individual Project Proposal

Manndon Kahn

Proposal 1:

- **Research Question:** Do those who engage in a religious community have different levels of social support compared to those who don't?
- **Variables:**
  - `PSYCH_zimet_multidimensional_social_support_scale_[need, joys, positive_not_scored, comfort, help, go_wrong, problems_friends, feelings, decisions]:` These are ordinal variables that can take 7 different values ranging from "Very Strongly Disagree" to "Very Strongly Agree". They can be used to get an idea of somebody's social support.
  - `CONNECTION_community_involvement_religious_p3m:` This variable represents the answer to the question: "In the past 3 months, have you participated in a religious-affiliated group (such as a church youth group or choir)?". Although this variable can take 3 different values ("Yes", "No, but I would like to", "No, and I'm not interested in doing so"), for the sake of this question, it can be considered as a binary variable by combining the 2 "No" answers into one. This variable can be used to identify whether the individual is engaged in a religious community.
- **Hypothesis:** H_0*:* There is no difference in social support between those who are engaged in a religious community and those who aren't.
- **Method:**
  - First, we come up with some sort of way to measure somebody's social support (**ex:** ranking each ordinal category of the social support variables from 0 (for "Very Strongly Disagree") to 7 (for "Very Strongly Agree"), and then take the average of those numbers for all the social support variables).
  - Then, we do a **permutation test** by shuffling the binary religious engagement label (`CONNECTION_community_involvement_religious_p3m`). We then infer the p-value, which represents the probability that the dataset matches what we would have expected to see if the null hypothesis is true, and use it to measure the strength of the evidence against the null hypothesis.
- **Discussion of Results:**
  - If the p-value is low, we can say there is strong evidence that there's a difference in levels of social support between those who engage in religious communities and those who don't.

    - Otherwise, we can say that the evidence for there being a difference in social support between those who engage in a religious community and those who don't isn't strong.

Proposal 2:

- **Research Question:** Do more romantically active people have different levels of life satisfaction compared to romantically inactive people?
- **Variables:**
    - `WELLNESS_life_satisfaction:` This discrete variable can take 10 values ranging from 1 to 10, and it represents the answer to the question: "On a scale of 1 to 10, How do you feel about your life as a whole right now?". It can be used to measure how satisfied someone is in life.
    - `CONNECTION_activities_kissed_last` and `CONNECTION_activities_sex_last:` These are variables that can take 8 different values ranging from "Not in the past three months" to "Earlier today". They represent the answers to the questions "When was the last time you had kissed someone?" and "When was the last time you had sex with someone?". These variables can be used to determine how active someone is romantically.
- **Hypothesis:** H_0*:* There is no difference in life satisfaction between those who are romantically active and those who aren't.
- **Method:**
    - First, we come up with some sort of way to categorize somebody as romantically active or romantically inactive. (**ex:** ranking each category of the `kissed_last` and `sex_last` variables from 0 (for "Not in the past three months") to 8 (for "Earlier today"), and then take the average of those 2 values, and compare them against a picked threshold to distinguish the romantically active from the romantically inactive).
    - Then, we do a **permutation test** by shuffling the created binary romantically active label. We then infer the p-value, which represents the probability that the dataset matches what we would have expected to see if the null hypothesis is true, and use it to measure the strength of the evidence against the null hypothesis.
- **Discussion of Results:**
    - If the p-value is low, we can say there is strong evidence that there's a difference in levels of life satisfaction between those who are romantically active and those who aren't.
    - Otherwise, we can say that the evidence for there being a difference in levels of life satisfaction between those who are romantically active and those who aren't isn't strong.

Proposal 3:

- **Research Question:** Are people who live with dogs or cats happier than those who don't?
- **Variables:**
  - `WELLNESS_subjective_happiness_scale_happy:` This discrete variable can take 7 values ranging from "1 – Not a happy person" to "7 – A very happy person". It can be used to gauge how happy an individual is.
  - `GEO_housing_live_with_dogs` and `GEO_housing_live_with_cats:` These variables are integers representing how many dogs or cats an individual lives with.
- **Hypothesis:** H_0*:* There is no difference in happiness between those who live with dogs or cats and those who don't.
- **Method:**
  - First, we create a binary variable telling us whether an individual lives with at least 1 dog or cat using the `GEO_housing_live_with_dogs` and `GEO_housing_live_with_cats` variables.
  - Then, we do a **permutation test** by shuffling the newly created binary label that tells us whether an individual owns a dog or a cat. We then infer the p-value, which represents the probability that the dataset matches what we would have expected to see if the null hypothesis is true, and use it to measure the strength of the evidence against the null hypothesis.
- **Discussion of Results:**
  - If the p-value is low, we can say there is strong evidence that there's a difference in levels of happiness between those who own a dog or a cat and those who don't.
  - Otherwise, we can say that the evidence for there being a difference in happiness between those who own a dog or a cat and those who don't isn't strong.