Individual Project Proposal

Beining Yin

Area of Research

How does an individual's social connection affect his/hers self evaluation? We want to investigate how living, working, as well as social situations affect personal well-being.

Teammates

Mingyuan Shi
Alejandro

Research Question 1

Question: How does a person's living situation (living by themselves or sharing living space with others) affect their mental well-being?

Formal Hypothesis:
H0: p1 - p2 = 0,
Ha: p1 - p2 ≠ 0, where p1 refers to average mental well-being of people living alone and p2 refers to average mental well-being of people sharing the living space.

Informal Hypothesis: Whether a person lives by themselves or not does not affect their mental well-being.

Variables

CSCS_data_anon dataset collected 'GEO_housing_household_size'. This variable collects that excluding the interviewee themselves, how many people share the living space. This includes a roommate, friend and/or family that shares the household with the person. From this variable, we will select observations that have a count higher than 0, and mark them as 'shared living'. For the rest of the variables (excluding the NA rows), we will mark them as 'living alone' since they have a count of 0 suggesting they don't have any roommates.

'WELLNESS_life_satisfaction' measures how well a person feels about his/hers life as a whole at the moment. This measures the self recognized state of mental well-being.

Visualization

'GEO_housing_household_size' is a numerical variable. Therefore, we will use a boxplot to visualize the distribution of average days per week that a person works from home.

‘WELLNESS_life_satisfaction’ is a categorical variable, where self evaluation is marked on a scale of 1 to 10. We will visualize the relationship of life satisfaction with work from home days by a side-by-side boxplot, so that we will directly visualize the different distribution (in individual boxplots) of online working days with regards to a peron’s life satisfaction in one x-y axis.

Analysis

Method: Two Group Hypothesis Testing

We will conduct a two group hypothesis testing to evaluate if there is a difference in mental well being between people who live by themselves from people who live with their family, friends and/or roommates. From our dataset, we wil separate the sample pool to two groups: those who live alone and those who do not. Then, we will calculate the average rate of mental wellbeing for each group. Using simulated sample and calculated simulated statistics, we will visualize the simulated distribution under the H0 assumption and compare the results with our test statistic.

Results & Discussion

We expect to see very few simulated statistics that are at least as or more extreme than the test statistics, which suggests a smaller p-value. To our expectation, this will be smaller than significance level 0.05, which suggests that there is sufficient evidence against H0, which in favor of that there is a difference in living situation (alone or not) in their average mental well-being. This is because, to my understanding, people who live alone typically have less social time and are more used to staying by themselves. This might affect their mental well-being in a negative way.

Research Question 2

Question: How does a person’s working situation (working online or working in person) affect their mental well-being?

Formal Hypothesis:

H0: $p1 - p2 = 0$,

Ha: $p1 - p2 \neq 0$, where p1 refers to average mental well-being of people working mostly from home and p2 refers to average mental well-being of people working in an office.

Informal Hypothesis: Whether a person works from home or not does not affect their mental well-being.

Variables

CSCS_data_anon dataset collected ‘WORK_shift_days_from_home_num’. This variable stores the average days per week that a person works from home. We will utilize this variable and filter individuals who work from home over 2.5 days (which is half of total weekdays in a week) and mark them as working mostly online; while the individuals who work less than 2.5 days will be marked as mostly in-person. We will separate the sample pool into these two groups: work mostly online versus work mostly in-person.

‘WELLNESS_life_satisfaction’ measures how well a person feels about his/hers life as a whole at the moment. This measures the self recognized state of mental well-being.

Visualization

‘WORK_shift_days_from_home_num’ is a numerical variable. Therefore, we will use a boxplot to visualize the distribution of average days per week that a person works from home.

‘WELLNESS_life_satisfaction’ is a categorical variable, where self evaluation is marked on a scale of 1 to 10. We will visualize the relationship of life satisfaction with work from home days by a side-by-side boxplot, so that we will directly visualize the different distribution (in individual boxplots) of online working days with regards to a peron’s life satisfaction in one x-y axis.

Analysis

Method: Two Group Hypothesis Testing

We will conduct a two group hypothesis testing to evaluate if there is a difference in mental well being between people who work online from people who work in-person. From our dataset, we will separate the sample pool into two groups: those who work from home during most weekdays (days working from home > 2.5 days) and those who do not. Then, we will calculate the average rate of mental wellbeing for each group. Using simulated samples and calculated simulated statistics, we will visualize the simulated distribution under the H0 assumption and compare the results with our test statistic.

Results & Discussion

Moving from the first research question which focuses on the effect of living situations on personal well-being, now we want to study the effect of working situations. Since the pandemic, there has been a gradual shift from in-person work to hybrid working environments, even to complete online working situations. This, while saving time and energy from commuting, does cause some mental issues as reported by many. People get a sense of disconnection when they do not get to see colleagues on a face to face

basis over the long term, and to my perspective a mostly online working situation affects personal well-being in a negative way.
We expect to see not many simulations that generate results over the test statistics line. This indicates a statistically significant p-value which suggests against H0 and in favor of Ha that there is a difference in working situation (online or not) in their average mental well-being. This finding will help us gain a better understanding of what affects an individual's mental well-being. Moreover, this will help companies build policies and guidelines in hybrid and online working that is favorable to employee's mental health.

Research Question 3

Question: How does a person's social situation (their connection with friends, family, colleagues and neighbors) affect their mental well-being?

Formal Hypothesis:
H0: $b1 = 0$,
Ha: $b1 \neq 0$, where b1 refers to the slope of the estimated line of best fit. Failure to reject H0 would suggest that the slope is 0, and there is no significant linear relationship between the predictor and response variables. Reject H0 would be in favor of Ha, which stands for a linear relationship between predictor and response.

Variables

Predictor:

1. 'CONNECTION_social_time_family_p7d_grouped': this variable collects the total hours of social activities spent with family over the past week
2. 'CONNECTION_social_time_friends_p7d_grouped': this variable collects the total hours of social activities spent with friends over the past week
3. 'CONNECTION_social_time_coworkers_and_classmates_p7d_grouped': this variable collects the total hours of social activities spent with colleagues or classmates over the past week
4. 'CONNECTION_social_time_neighbours_p7d_grouped': this variable collects the total hours of social activities spent with neighbors over the past week

From these variables collected in *CSCS_data_anon* dataset, we will calculate the total number of social activities with family, friends, colleagues/classmates as well as neighbors and store this result in a new variable 'social'. We will use this as the predictor variable for our linear regression analysis.

Response:
'WELLNESS_self_rated_mental_health' measures the rating of a person's mental health by themselves. We will be using this variable to indicate a person's mental well being at the time of survey.

Visualization

‘Social’ is a numerical variable, and we can visualize the distribution of social hours over the past week in a histogram.

‘WELLNESS_self_rated_mental_health’ is a numerical variable, and we can visualize the relationship of this variable with ‘social’ by a scatter plot, where we can look at the 2 dimensional distribution of these 2 variables and visualize for a linear trend.

Analysis

Method: Simple Linear Regression

We will use the sample data and select ‘Social’ as our explanatory variable and ‘WELLNESS_self_rated_mental_health’ as our response variable. Under the assumption of a linear trend between these two variables, we estimate the simple linear regression model. Based on the model summary, we will be able to examine the predictive power of the model, and check the p value for the b1 hypothesis test to see if the linear trend assumption still stands.

To our understanding, social connection, measured by total hours of activities per week, should have a positive linear relationship with mental well-being rating. Individuals who connect more with people, whether from personal groups to work groups, should have a healthier mental state compared to people with less socialling. Therefore, we predict to get a model with high performance and low p value that rejects H0 and in favor of Ha suggesting for a significant linear relationship of the predictor and response variable.

Result & Discussion:

In addition to our current analysis, we should further study how more personal social activities, those with friends and families, versus more work-related social activities, those with colleagues, classmates, and neighbors, affect differently in personal well-being. This is because many social activities involving work are mandated by company and not by choice. The conversations and activities usually are formatted in a way that does not allow for deeper connection between people. In comparison, personal activities with friends and families are more spontaneous and more in-depth. This would aid more in promoting mental wellbeing, to my perspective.

Moreover, we should study how covid 19 reshaped an individual's lifestyle, and in what way did that affect an individual’s mental well-being over the long run. To my understanding, since the pandemic many people have been more adapted to a hybrid or even on-line prompted lifestyle, and that affected their social activities timeframe.

These analyses in whole are vital to address the major players in maintaining a healthy mental wellbeing.