Individual Proposals

1. Introduction

To explore the Canadian Social Connection Survey (CSCS) data, and based on the general idea of studying the relation between social connection and personal health which can be understood as well-being, I will research the question given below in this proposal:

Will more negative feelings affect the overall personal well-being rate in a negative way?

Will more social actions help to improve the overall well-being rate?

How will living with pets or other people improve the overall well-being rate?

2. Analysis Plan

2.1 Will more negative feelings affect the overall personal well-being rate in a negative way?

2.1.1 Considered Plan

In this section, I plan to use simulation by bootstrapping to evaluate the question, with data of how many negative feelings people have. Bootstrapping with replacement to create more data, then estimate the coefficient of correlation between the paired data, plot a histogram, then judge it with a confidence interval.

2.1.2 Data Used

1. Q23_1 WELLNESS_malach_pines_burnout_measure_tired
2. Q23_2 WELLNESS_malach_pines_burnout_measure_disappointed
3. Q23_3 WELLNESS_malach_pines_burnout_measure_hopeless
4. Q23_4 WELLNESS_malach_pines_burnout_measure_trapped
5. Q23_5 WELLNESS_malach_pines_burnout_measure_helpless
6. Q23_6 WELLNESS_malach_pines_burnout_measure_depressed
7. Q23_7 WELLNESS_malach_pines_burnout_measure_sick
8. Q23_8 WELLNESS_malach_pines_burnout_measure_worthless
9. Q23_9 WELLNESS_malach_pines_burnout_measure_difficulty_sleeping
10. Q23_10 WELLNESS_malach_pines_burnout_measure_had_it
11. Q16 WELLNESS_life_satisfaction

2.1.3 Variables and Pre-assumption

Depression rate: The summation of how many negative feelings a person has by adding Q23_1 to Q23_10, no smaller than 1, a diagram will clearly show its distribution.

Life Satisfaction: The actual value of rate of well-being answered for Q16, from 1 to 10. Directly from Q16, may use a box diagram to visualize.

Assumption:

All the data in the sample should be independent and randomly selected.

2.1.4 Analysis Plan

As variables 1 to 10 are answered with the degree of the feeling, not in numerical values directly, so the first step is to transfer them into numerical values. Since there are seven different kinds from "always" to "never", I will give each a value from 1 to 7 in acceding way, so "never" is 1, "always" is 7. Adding these values from variables 1 to 10, we get a summation we can call "depression rate" which basically describes how many negative feelings one person has.

Then group the data based on the life satisfaction rate we have from Q16: 1 to 5 is marked as unhappy, while 6-10 is marked as happy.

To get the answer, bootstrap the sample in pair of the "Depression rate" and "Life Satisfaction" for each person with replacement, we have for 1000 times and calculate every sample coefficient of correlation to get a new distribution.

2.1.5 Possible Results and Evaluation

Considering a 95% confidence interval, apply to the distribution by using a histogram representing it. For 0 lies within the aera of the confidence interval, then there is a possibility that the negative feeling does not affect the overall well-being of a person, vice versa.

2.2 Will more social actions help to improve the overall well-being rate?

2.2.1 Considered Plan

In this section, I aim to apply hypothesis test to assess the question, using data on the frequency of social interactions among individuals. Setting a null hypothesis and test the data given null hypothesis. A p-value will be calculated under this condition, based on the p-value, we can get the

result.

2.2.2 Data Used

1. Q117_1 CONNECTION_activities_greeted_stranger_p3m
2. Q117_2 CONNECTION_activities_greeted_neighbour_p3m
3. Q117_3 CONNECTION_activities_talked_day_p3m
4. Q117_4 CONNECTION_activities_talked_family_p3m
5. Q117_5 CONNECTION_activities_talked_job_p3m
6. Q117_6 CONNECTION_activities_talked_hobbies_p3m
7. Q117_7 CONNECTION_activities_phone_p3m
8. Q117_8 CONNECTION_activities_letter_or_email_p3m
9. Q117_9 CONNECTION_activities_checked_in_p3m
10. Q117_10 CONNECTION_activities_text_or_messaged_p3m
11. Q117_11 CONNECTION_activities_chat_p3m
12. Q117_12 CONNECTION_activities_video_chat_p3m
13. Q117_13 CONNECTION_activities_group_video_chat_p3m
14. Q117_14 CONNECTION_activities_walk_p3m
15. Q117_15 CONNECTION_activities_coffee_p3m
16. Q117_16 CONNECTION_activities_board_games_p3m
17. Q117_17 CONNECTION_activities_computer_games_p3m
18. Q117_18 CONNECTION_activities_onlinegames_p3m
19. Q117_19 CONNECTION_activities_visited_friends_p3m
20. Q117_20 CONNECTION_activities_visited_family_p3m
21. Q117_21 CONNECTION_activities_community_p3m
22. Q117_22 CONNECTION_activities_helped_p3m
23. Q117_23 CONNECTION_activities_meeting_work_p3m
24. Q117_24 CONNECTION_activities_meeting_organization_p3m
25. Q117_25 CONNECTION_activities_discussion_group_p3m
26. Q117_26 CONNECTION_activities_group_exercise_p3m
27. Q117_27 CONNECTION_activities_church_p3m
28. Q117_28 CONNECTION_activities_new_friend_p3m
29. Q117_29 CONNECTION_activities_hug_p3m
30. Q117_30 CONNECTION_activities_kissed_p3m
31. Q117_31 CONNECTION_activities_sex_p3m
32. Q16 WELLNESS_life_satisfaction

2.2.3 Variable and Pre-assumption

Social Score: The summation of the quantized answers from Q117_1 to Q117_31. There were 7

different answers in total, so given each answer from 1 to 7 in acceding, so 1 for least frequent which

is "less than monthly", 7 for "daily or almost daily".

Life Satisfaction: The actual value of rate of well-being answered, from 1 to 10. Directly from Q16,

may use a box diagram to visualize.

2.2.4. Analysis Plan

To analyze the data, a two-tailed hypothesis test is applied. Set the null hypothesis H0: "More social actions do not help to improve the overall well-being rate". For testing the null hypothesis, "Social score" and "Life satisfaction" are introduced. "Social score" measures how active a person does social; "Life satisfaction" is the self-graded well-being rate.

2.2.5 Possible Results and Evaluation

In general, we tend to believe that increased social activities contribute to better mental and physical health, leading to a greater sense of well-being. Thus, in hypothesis testing, we are inclined to reject the null hypothesis (H0), which states that there is no significant relationship between the frequency of social activities and happiness. However, if the results show that we do not have sufficient statistical evidence to reject the null hypothesis, this may suggest that the link between increased social interaction and well-being is not as strong as we anticipated. In other words, more social engagement does not necessarily lead to greater happiness.

2.3 How will living with pets or other people improve the overall well-being rate?

2.3.1 Considered Plan

In this section, a simple linear regression will be used to evaluate a model to estimate the relation between well-being rate and other member in a family. A scatter plot will be produced to produce a visualized result, and multiple graphs will be used to test the model.

2.3.2Data Used
1. Q41_1 GEO_housing_live_with_partner
2. Q41_2 GEO_housing_live_with_children
3. Q41_3 GEO_housing_live_with_grandkids
4. Q41_4 GEO_housing_live_with_parent
5. Q41_5 GEO_housing_live_with_in_laws
6. Q41_6 GEO_housing_live_with_siblings
7. Q41_7 GEO_housing_live_with_roommate
8. Q41_8 GEO_housing_live_with_other
9. Q42_1 GEO_housing_live_with_dogs

10. Q42_2 GEO_housing_live_with_cats
11. Q42_3 GEO_housing_live_with_ferrets
12. Q42_4 GEO_housing_live_with_birds
13. Q42_5 GEO_housing_live_with_fish
14. Q42_6 GEO_housing_live_with_other_pet
15. Q16 WELLNESS_life_satisfaction

2.3.3 Variables

x: The independent variable: "Accompany rate". The variable for how many pets a person has and how many people live with the person. It is a summation from Q41_1 to Q42_6 in 2.3.1 and a histogram can be plotted to see its tendency clearly.

y: The dependent variable: value for life satisfaction. Directly from Q16, may use a box diagram to visualize.

Assumption:

- The parameters are linear

- Every data is independent

- Sample has common variance

- The errors are normally distributed

2.3.4 Analysis Plan

For this part a simple linear regression model is considered. When it comes to a simple linear regression, the following assumptions are made: parameters are linear; observations are independent; there is common error variance; errors are normally distributed with mean 0 and variance $\sigma 2$.

The variables are x and y as shown in 2.3.2. Analysis starts with plotting every pair of data on a cross-coordinate diagram, with accompany rate x in x-axis, and value for life satisfaction y in y-axis. The expected model for the sample will be:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + e_i$$

$\hat{y}$: Estimated y value
$\hat{\beta}_0$: Intercept at y-axis for x = 0
$\hat{\beta}_1$: Slope when there is one unit increase in x, how many increases in $\hat{y}$
e_i: Random error

Check the assumption of the model with different diagram is indeed: the residual versus each predictor, residual versus estimated y value, residual versus x value and normal quantile quantile

plot, to see whether there is any pattern, if not, test passed.

To calculate the specific parameter, we need to use the method of least square: find the line that gives the minimum overall squared residuals.

2.3.5 Possible Results and Evaluation

It is foreseeable that the presence of more pets or family members will have a positive impact on one's sense of happiness. This is because companionship often provides emotional support and reduces feelings of loneliness, thereby enhancing overall well-being. Therefore, in this context, parameter $\hat{\beta}_1$ is expected to be positive, so the regression formula shows a positive correlation.

Additionally, tests for significance also needed to be considered. A F-test after getting the full formula will do it great. For a high significance, it shows the correlation is strong, but for a low significance, some adjustment should be made.