Statistics and Methodology Group Project
Group 4

F. Giuliani – 2035273 - 684603
L. Martorano – 2033590 - 306264
L. Minarini – 2035217 - 736034
R. Rampoldi – 2035931 - 129700
S. Srinivasan – 2034106 - 336695

1 - Multiple Linear Regression

1.1 – Which countries are represented in these data?
The countries represented in these data are five:
- China: 156
- Germany: 276
- India: 356
- Russia: 643
- USA: 840

1.2 – What are the sample size for each country represented in these data?
- China: 2300
- Germany: 2046
- India: 5659
- Russia: 2500
- USA: 2232

1.3 – Overall, is there a significant effect of country on feelings of happiness?
There is a significant effect of country on feelings of happiness (β = -0.0000798, SE = 0.0000253 , t = -3.16, p = 0.0016)

1.4 – Which country has the highest level of feelings of happiness?
The country with the highest level of feelings of happiness is Russia (M = 2.109).

1.5 – Which country has the lowest level of feelings of happiness?
The country with the lowest level of feelings of happiness is USA (M = 1.728).

1.6 – How do the country-specific levels of feelings of happiness change after controlling for subjective state of health?
After controlling for subjective state of health, the levels of feeling of happiness increase in Germany, India and USA, while they decrease in China and Russia.
(To calculate the difference between the two models, we subtracted the estimate slope of the model with two predictors (Country and Feelings of Happiness) from the estimate slope of the model with three predictors (Country, Feelings of Happiness and Subjective State of Health), resulting in:
- China: -0.6509
- Germany: 0.0078
- India: 0.0254
- Russia: -0.1439
- USA: 0.0668)

2 - **Continuous Variable Moderation**

2.1 – After controlling for country, does the importance people afforded to democracy (*DemImp*) significantly predict the extent to which they think their country is being run democratically (*DemRun*)?

> Yes, the importance people afforded to democracy does significantly predict the extent to which they think their country is being run democratically (β = 0.35, *SE* = 0.008, *t* = 44.66, *p* < 0.001)

2.2 – After controlling for country, does the *DemImp* → *DemRun* effect vary as a function of peoples' satisfaction with their lives (SWL)?

> Yes, the effect varies slightly (β = -0.006, *SD* = 0.003, *t* = -2.02, *p* = 0.04)

2.3 - Within what range of SWL is the *DemImp* → *DemRun* simple slope from Question 2 statistically significant?
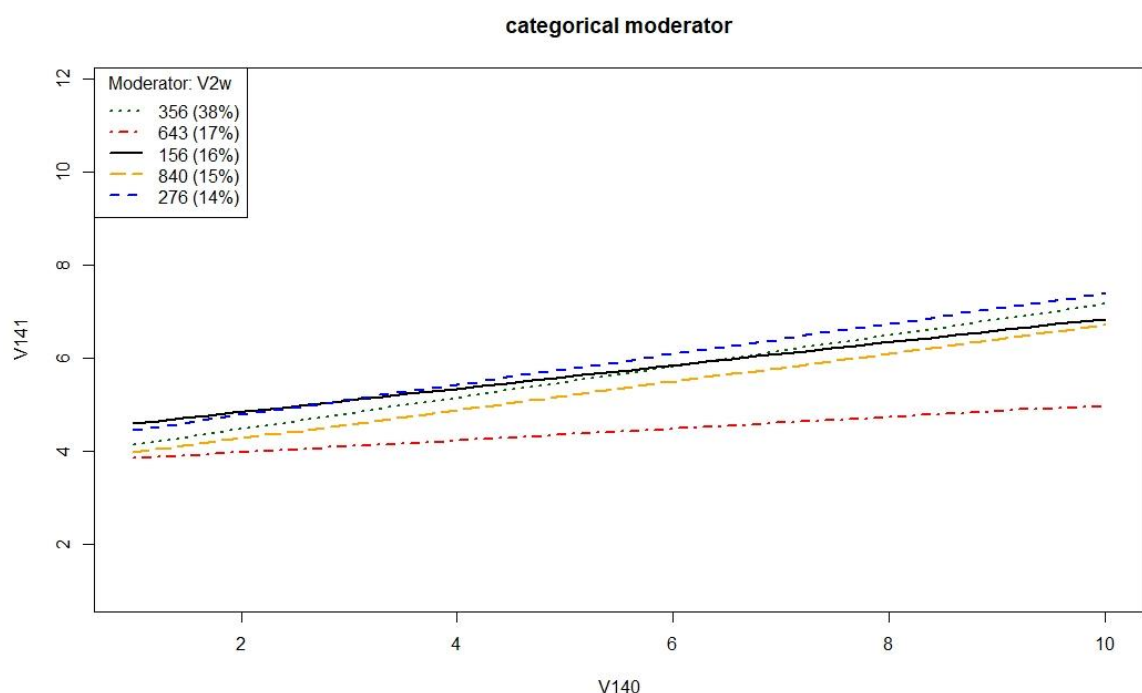
> The SWL simple slope is significant between 29.84 and 1581.38

3 - **Categorical Variable Moderation**

3.1 - After controlling for SWL, does the *DemImp* → *DemRun* effect vary significantly by country?

> Yes, the effect varies by country (*F* = 28.65, *p* < 0.001)

3.2 - Visualize the results from Question 1 in a suitable way.

categorical moderator

3.3 - For which country is the effect of *DemImp* on *DemRun* <u>strongest</u>, after controlling for SWL?

      The strongest effect happens in India (β = 0.3356, *SE* = 0.012, *t* = 28.85, *p* < 0.001)

3.4 - For which country is the effect of *DemImp* on *DemRun* <u>weakest</u>, after controlling for SWL?

      The weakest effect happens in Russia (β = 0.126, *SE* = 0.017, *t* = 7.58, *p* < 0.001)

3.5 - Are the simple slopes referenced in Questions 3 and 4 statistically significant?

      Yes, they both are (For India β = 0.34 and for Russia β = 0.13; both have *p* < 0.001)

4 – **Predictive Modeling**

4.1 - Select and list three (theoretically justified) sets of predictors (or functions thereof, e.g., interactions or polynomials) to use in predicting *FinSat*.

      We run three different models with the following sets of predictors:
- Social Class (V238), Scale of Income (V239) and Family Savings (V237);
- How much freedom of choice and control over own life (V55), Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair (V56) and Secure Neighborhood (V170);
- How many children do you have (V58), Marital Status (V57) and Highest education level attained (V248).

4.2 - Briefly explain why you expect the three sets of predictors you chose in Question 1 to perform well. That is, explain your rationale for defining these three sets.

      While trying to select the right sets of predictors, we found ourselves working with a long list of possibilities; first, we proceeded by trying to find the predictors with the highest correlation coefficients related to the "Financial Situation of their household" variable. Then, from the apprised list, we tried to put together the predictors which were theoretically justifiable. From here we created the first two models; the first one has just two predictors from the above-mentioned list, as we added "Social Class" since we thought it could be strictly intertwined both to "Scale of Income" and "Family savings". For the second model, we then tried to choose predictors whose outcome on "Financial situation" would not be so obvious, even though the correlation coefficient is still high for all of the three.

      As for the last model, we simply chose three predictors that could be logically justified without looking at any correlation coefficients.

4.3 - Use 10-fold cross-validation to compare the predictive performance of the three models define in Question 1.

      The three models have the following *CVE*s:
- For model 1 is 4.921
- For model 2 is 4.862
- For model 3 is 5.785

4.4 - Which of the three models compared in Question 3 performed best?

      The model that performed better is the second one (How much freedom of choice and control over own life (V55), "Do you think most people would try to take

advantage of you if they got a chance, or would they try to be fair" (V56) and Secure in Neighborhood (V170)) because it has the lowest *MSE* (4.86) and highest *R-squared* (0.178)

4.5 - What is the estimated prediction error of the best model?
> The prediction error is 4.862

4.6 - Based on the selection you made in Question 4, what can you say about the attributes that are important for predicting financial satisfaction?
> When the independent predictors we selected in the model are equal to zero, "Financial Satisfaction" is equal to 3.58.
> For each additional point of "How much freedom of choice and control over own life", we see an increase in "Financial Satisfaction" by 0.31 points, after controlling for the other two predictors.
> For each additional point of "Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair", "Financial Satisfaction" is expected to increase by 0.17 points, after controlling for the other two predictors.
> For each additional point of "Secure Neighborhood", we see a decrease in "Financial Satisfaction" by 0.43 points, after controlling for the other two predictors.
> The *p-values* are all less than .001 and the model explains 18% of the variance.