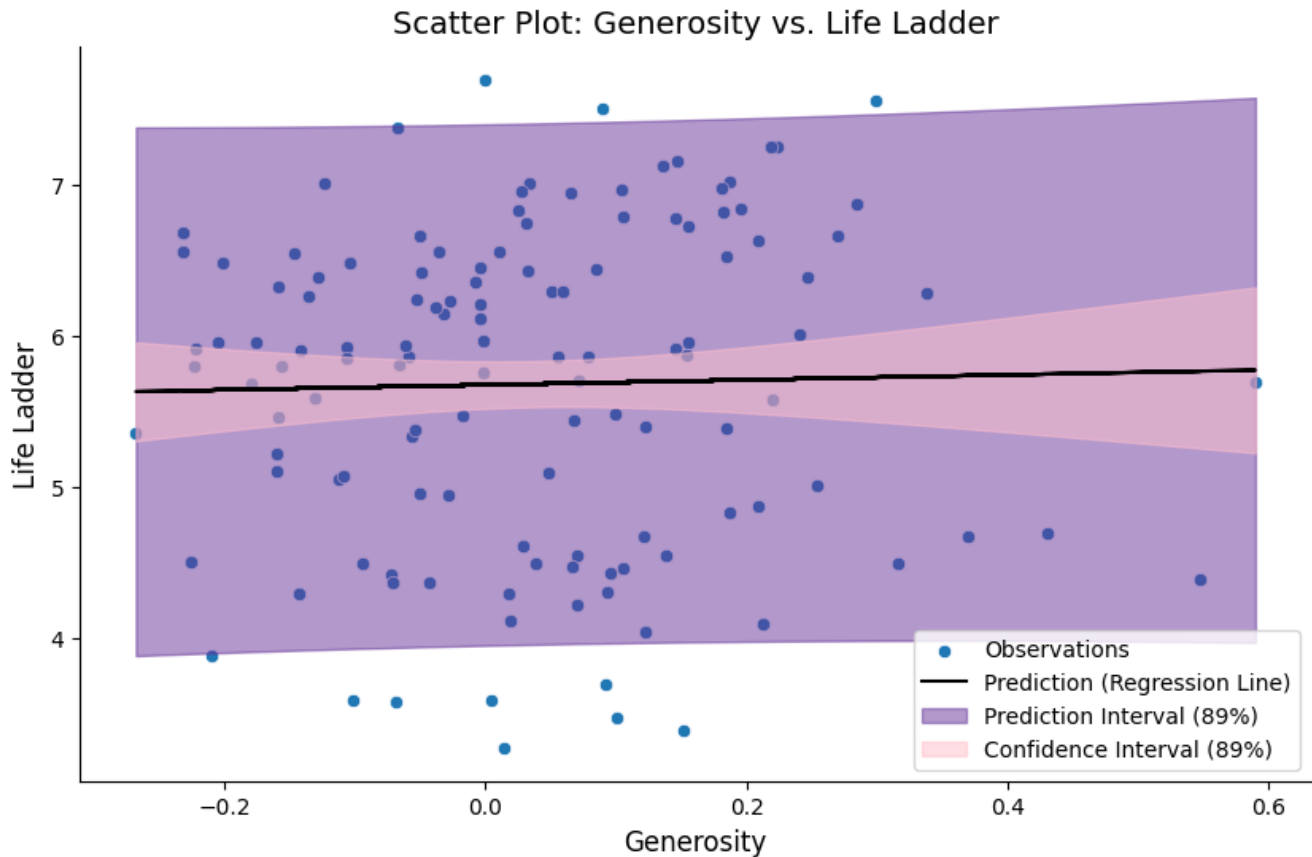


MAT-206:00030

Name: Win (Thanawin Pattanaphol)

Student ID: 01324096

1.



Answer the questions presented in the instructions

How much of the variance of the response can the model explain?

- This can be checked by using the R-squared value 0.001 which shows that only 0.1% of the variance in Life Ladder is evident in the model Generosity as the predictor.

Is there any variable that seems irrelevant when we include the others?

- The p-value for Generosity is 0.780 which suggests that Generosity does not have a significant relationship with the Life Ladder variable.

Do we have evidence to consider all the variables irrelevant at the same time?

- From the F-statistic value of 0.07861 and the p-value for the F-statistic 0.780, there is not enough evidence to reject the null hypothesis that the model as a whole is irrelevant.

Do we have evidence to consider that the errors follow a normal distribution?

- The omnibus test p-value is 0.004 which shows a notable deviation from normal along with the Jarque-Bera (JB) test's p-value of 0.0443 – showing that there are some non-normality in the data. Thus, there are evidence that shows that it does not follow a normal distribution.

MAT-206:00030

Name: Win (Thanawin Pattanaphol)

Student ID: 01324096

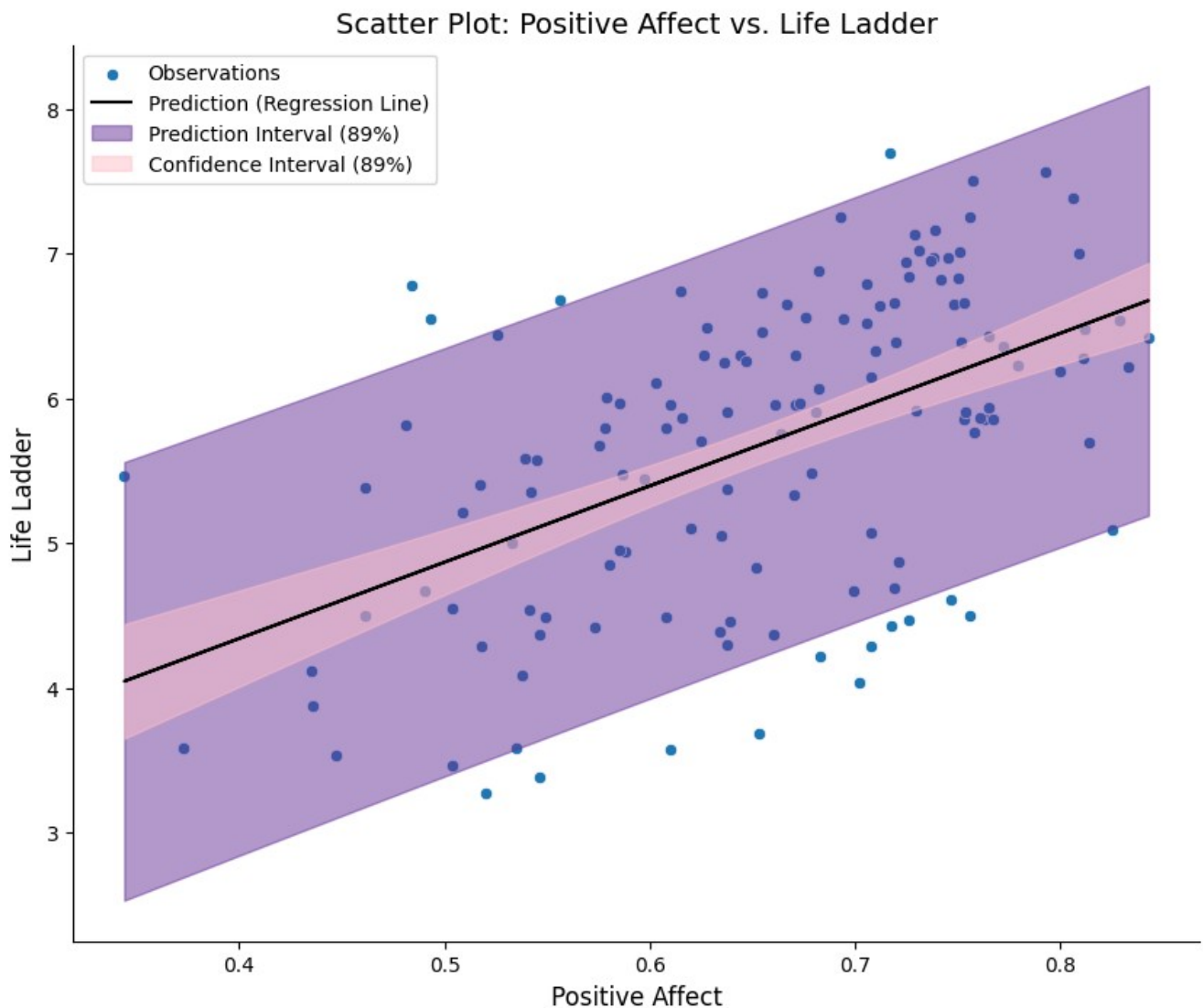
Do we have evidence to consider that the errors are independent?

There is no strong evidence against the assumption of independent errors shown in the Durbin-Watson statistic with the value of 1.906 which suggests that the errors do not show notable autocorrelation.

Do we have evidence of numerical problems in the design matrix?

There is no evidence of numerical problems in design matrix, the condition number is 6.24 which is relatively low and does not indicate any instability in the design matrix.

2.



How much of the variance of the response can explain the model?

R-squared value of 0.271 shows that around 27% of the variance in Life Ladder can be explained by Positive Affect and suggests that the model shows a moderate variability in the response.

MAT-206:00030

Name: Win (Thanawin Pattanaphol)

Student ID: 01324096

Is there any variable that seems irrelevant when we include the others?

Positive Affect has a p-value of 0.000 which is very low and is a notable predictor of Life Ladder. Thus, positive affect appears to be relevant to Life Ladder.

Do we have evidence to consider all the variables irrelevant at the same time?

F-statistic's low p-value shows that at least of the predictors (Positive Affect) is related to the response variable of Life Ladder. Thus, reject the null hypothesis that all variables are irrelevant.

Do we have evidence to consider that the errors follow a normal distribution?

There is mild evidence that the errors may not be perfectly normal such as the Omnibus test where its p-value is 0.047 and the Jarque-Bera test with the p-value of 0.102 – both shows a small deviation from normality, however, they are not strong enough to say that there is non-normality.

Do we have evidence to consider that the errors are independent?

There is no strong evidence to suggest that the errors are not independent such as the Durbin-Watson statistic of 1.815 which is close to 2 meaning that there is no autocorrelation, thus, not independent.

Do we have evidence of numerical problems in the design matrix?

Again, there is no evidence of numerical problems in the design matrix – condition number of 13.7 which is below 30 – does not demonstrate instability.

3.

How much of the variance of the response can explain the model?

- R-squared value of 0.856 shows that 85.6% of the variance in Life Ladder is shown by the predictors in the model – thus – suggests that the model explains most of the variability in the response variable.

Is there any variable that seems irrelevant when we include the others?

- Healthy life expectancy at birth with the p-value of 0.516 and the Generosity p-value of 0.103 shows that they seem to be less relevant in explaining Life Ladder in the presence of other predictors.

Do we have evidence to consider all the variables irrelevant at the same time?

- F-statistic value of 80.73 and the p-value of 2.99e-42 shows that the overall model is statistically significant which means that we reject the null hypothesis that all variables are irrelevant.

Do we have evidence to consider that the errors follow a normal distribution?

- There is no strong evidence to say that the errors do not follow a normal distribution – The omnibus test with a p-value of 0.83 and Jarque-Bera test p-value of 0.105 does not suggest major problems with normality.

Do we have evidence to consider that the errors are independent?

- The Durbin-Watson statistic has a value of 2.170 which shows that the errors appear to be independent as the statistic is near 2.

MAT-206:00030

Name: Win (Thanawin Pattanaphol)

Student ID: 01324096

Do we have evidence of numerical problems in the design matrix?

- There is evidence of numerical problems in the design matrix shown by the condition number of $1.51e+03$ which is relatively large – showing the potential presence of instability (both multicollinearity and numerical) in the design matrix.

4.

(a) There are several differences in the columns: Healthy life expectancy at birth and the Low GDP per capita – the instruction `scaler.fit_transform` essentially shrinks the two columns and scales the data into a 0 – 1 scale.

(b) The condition number before the function was applied: 1508.9221

The condition number after the function was applied: 43.9715 – It shows that after the scaling function has been applied, the condition number dramatically decreases.

5.

How much of the variance of the response can explain the model?

- The R-squared value of 0.841 shows that 84.1% of the variance in Life Ladder can be explained via the predictors in the model, this high value indicates that the model can explain the variation in the response variable very well.

Is there any variable that seems irrelevant when we include the others?

- All the variables (Social Support, Freedom to make life choices, Perceptions of corruption, Positive affect and Log GDP scaled) is shown to have a p-value that is below 0.05, thus, none of the variables appear to be insignificant or irrelevant and that they all contribute to the Life Ladder variable.

Do we have evidence to consider all the variables irrelevant at the same time?

- Due to the fact that F-statistic variable (119.1) and p-value of $2.26e-43$ has a very low value shows that the overall model is statistically significant and thus we reject the null hypothesis that all variables are irrelevant at the same time.

Do we have evidence to consider that the errors follow a normal distribution?

- There is no evidence to suggest that the errors do not follow a normal distribution (omnibus test p-value of 0.486) and Jarque-Bera test with the p-value of 0.607 have value that is greater than 0.05 which shows that there is no significant departure from normality.

Do we have evidence to consider that the errors are independent?

- The Durbin-Watson statistic of 2.108 with a value close to 2 shows that the errors are likely to be independent.

Do we have evidence of numerical problems in the design matrix?

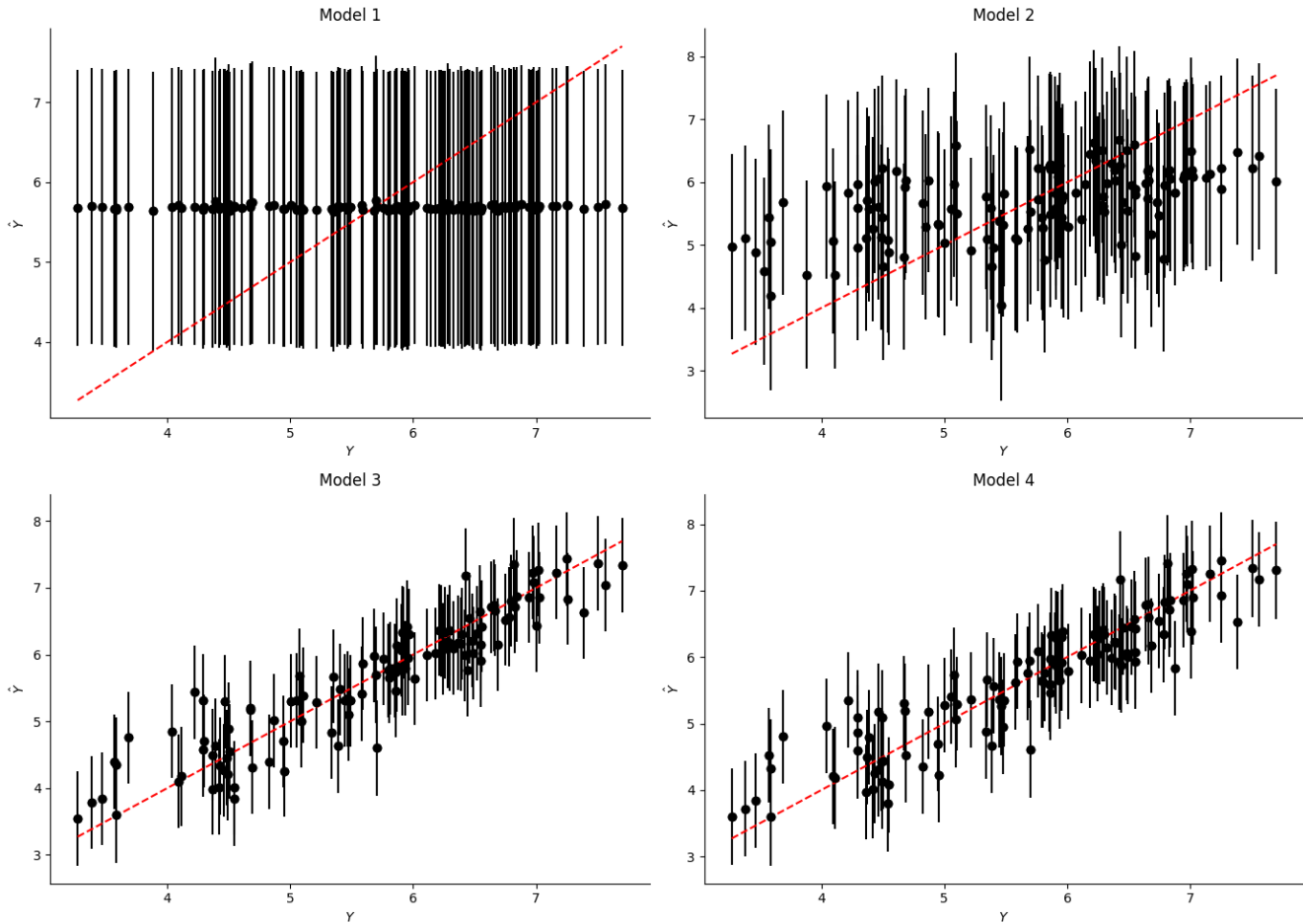
- There is no evidence of numerical problems in the design matrix shown in the condition number with the value of 15.6.

MAT-206:00030

Name: Win (Thanawin Pattanaphol)

Student ID: 01324096

6.



7. The model that is the most preferable to explain the variable Life Ladder is the 3rd model since the condition number in a regression model encapsulates how sensitive the model is to small changes in the input data, thus, a lower condition number means the model is more flexible, thus, is why the 3rd model is the most preferable.