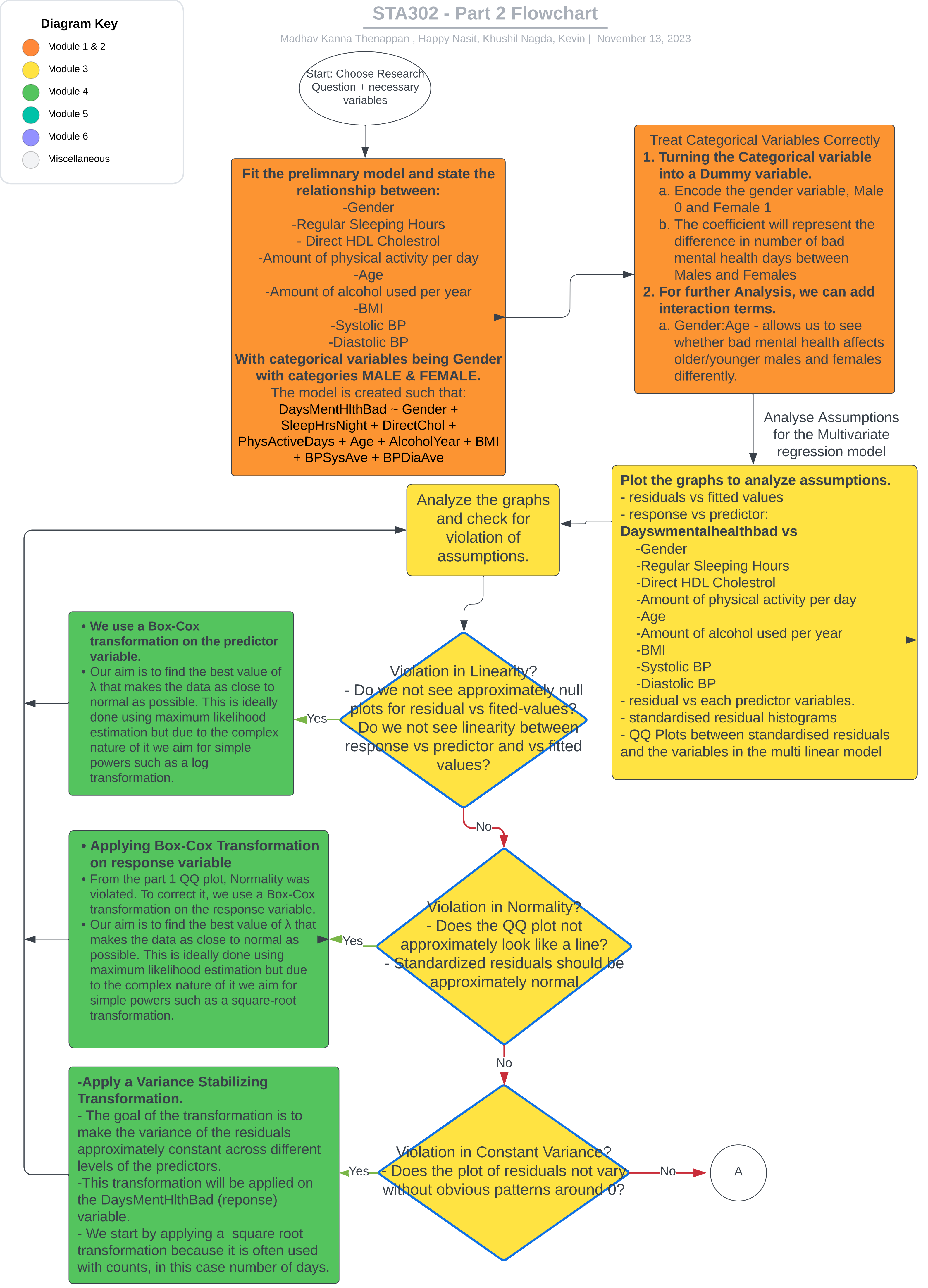
STA302 - Part 2 Flowchart

Madhav Kanna Thenappan , Happy Nasit, Khushil Nagda, Kevin | November 13, 2023



STA302 - Part 2 Flowchart

Madhav Kanna Thenappan , Happy Nasit, Khushil Nagda, Kevin | November 13, 2023

Diagram Key

- Module 1 & 2
- Module 3
- Module 4
- Module 5
- Module 6
- Miscellaneous

A

Confirm that both Linearity Assumptions and Conditionas have been satisfied.
- If we calculate a confidence interval and/or a prediction interval using data that violates the linearity assumption, predictions made for new observations will be systematically understated or overstated because the model does not represent the true relationship between the variables.

We want to estimate the effect of the predictors on the population parameters.

Calculate the 95% confidence interval for each predictor in the multlinear regression model.
The confidence interval tells us the range within which we can be confident that the true population parameter lies given a certain level of confidence.

- The confidence interval tells us that If we were to repeatedly take random samples of the same size from the population and construct a 95% confidence interval for a particular regression coefficient each time using the same model, then we would expect 95% of those confidence intervals to contain the true population coefficient for that predictor

- The value of the confidince interval that we obtain can tell us two things.
1. A CI that doesn't have zero means the effect of the predictor on the response is statistically significant.
2. The width of the interval gives an indication of the precision of the estimate i.e. narrow CI suggests more precise estimate.

We are now close to answering our research question:
To what extent can biological and lifestyle markers predict depressive symptoms among
survey respondents 18 - 80 years old?"

We'd like to know the range within which we expect future individual observations of depressive symptoms to fall, with a 95% level of confidence.

So, we calculate a prediction interval
Calculate the 95% prediction interval for the response variable.
- Prediction interval is a range of values within which we expect a single new observation to fall, with a 95% level of confidence.
- If we were to repeatedly sample from the population according to the regression model and also a new Y*, x* and compute a 95% prediction interval each time, then 95% of the intervals would include the population response value.

Start with i = 1 th coefficeint when considering hypothesis tests for individual coefficients

Perform hypothesis tests for individual coefficients of predictor variables.
1) For the response - predictor variable pair, construct:
- Null Hypothesis: coefficient i = 0
- Alternate Hypothesis: coefficient i not = 0

2) The test statistic follows the T distribution:

$$T = \frac{\hat{\beta}_i}{\hat{se}(\hat{\beta}_i)} \sim t_{n-(p-1)}$$

3) Find the p value, the probability that T* (computed test statistic from data) is at least or more extreme than the random test statistic. We assume a value of alpha = 0.05.
4) Conclusion based on p-value:
- If p value < 0.05, then we reject the null hypothesis
- If p value >= 0.05, then we do not reject the null hypothesis.

If we do not reject the null hypothesis, we do not include the predictor variable in our reduced model for the F test, and we include the predictor variable in the reduced model if we reject the null hypothesis.

i = total number of individual coefficients in full regression model? (in our model, the total number of individual coefficients = 9)

No

i = i + 1

Yes

Perform partial F test with the reduced model:
1) Construct Null and Alternative hypotheses:
- Null Hypothesis: the reduced model is sufficient
- Alternative Hypothesis: the reduced model is not sufficient

2) Define and compute the test statistic:
The random test statistic is the ratio of mean sums of squares with sampling distributions given by:

$$F = \frac{MS_{reg}}{MSR} \sim F(p-r, n-p-1)$$

F* is the computed value of the test statistic with the data
3) Calculate the p value, the probability that F* is at least as large or larger than the random test statistic. We asume alpha = 0.05.
4) Conclusion based on p value:
- If p value < 0.05, then we reject the null hypothesis, and thus the reduced model is not sufficient.
- If p value >= 0.05, then we do not reject the null hypothesis, and thus the reduced model is sufficient.

End