Assignment 02

Polynomial Regression

This assignment is intended to give you experience fitting models with polynomial terms. Submit your responses to each of the questions below in a printed document. All graphics should be resized so that they do not take up more room than necessary and also should have an appropriate caption. If you are using Markdown, all syntax should be hidden (i.e., not displayed) unless specifically asked for. Any messages or warnings produced from loading packages should also be hidden. This assignment is worth 15 points. (Each question is worth 1 point unless otherwise noted.)

In this assignment, you will use the data from the file *fertility.csv* (see the data codebook to explain variation in infant mortality rates.

Linear Effect of Female Education Level on Infant Mortality Rates

1. Create a scatterplot showing the relationship between female education level and infant mortality rates.
2. Describe the relationship between female education level and infant mortality rates. Be sure to comment on the structural form, direction and strength of the relationship. Also comment on any potential observations that deviate from following this relationship (unusual observations or clusters of observations).
3. Compute and report the Pearson correlation coefficient between female education level and infant mortality rate. Based on your response to Question 2, explain whether the Pearson correlation coefficient is an appropriate summary measure of the relationship.
4. Regress infant mortality rates on female education level. For this model, posit a linear effect of female education level on infant mortality rate (Model 1). Create the scatterplot of the standardized residuals versus the fitted values from Model 1.
5. Does this plot suggest problems about meeting the assumption that the average residual is 0 at each fitted value? Explain.

Quadratic Effect of Female Education Level on Infant Mortality Rates

6. Regress infant mortality rates on female education level. For this model, posit a quadratic effect of female education level on infant mortality rate (Model 2). Write the *fitted equation* using Equation Editor (or some other program that correctly types mathematical expressions).
7. Create the scatterplot of the standardized residuals versus the fitted values from Model 2.
8. Does this plot suggest problems about meeting the assumption that the average residual is 0 at each fitted value? Explain.

Control for Differences in Gross National Income (GNI)

9. Regress infant mortality rates on female education level. For this model, posit a quadratic effect of female education level on infant mortality rate, and also control for differences in Gross National Income (Model 3). Write the *fitted equation* using Equation Editor (or some other program that correctly types mathematical expressions).

10. Create the density plot of the marginal distribution of the standardized residuals from Model 3. Add the confidence envelope for the normal distribution. Also create the scatterplot of the standardized residuals versus the fitted values from Model 3. In this plot identify any observations with extreme residuals (≤ -3 or ≥ 3) by indicating the country name of those observations in the plot. Place these plots side-by-side in your printed document and, for the purposes of captioning, etc. treat them as two subfigures within a single figure.

11. Based on the plots you created in Question 10, evaluate and comment on the tenability of each of the model assumptions. **(2pts.)**

Presenting the Results

12. Examine the structure and formatting of Table 9 at http://zief0002.github.io/epsy-8251/misc/creating-tables/creating-tables.html. Mimic the format and structure of this table to create a table to present the numerical information from the three models you fitted in this assignment. Make sure the table you create also has an appropriate caption. If the table is too wide, change the page orientation in your word processing program to "Landscape", rather than changing the size of the font. (Note: Only this table should be presented in landscape orientation…not your entire assignment!) **(2pts.)**

13. Create a publication quality plot that displays the fitted curves from Model 3. Display two separate lines to show the effect of Gross National Income. The two lines should be displayed using different linetypes or colors (or both) so that they can be easily differentiated in the plot.