|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PM592: Regression Analysis for Data Science** | | | Name: |  |  |
| **HW6** |  |  |  |  |  |
| *Variable Parameterizations* | | | | |  |

**Instructions**

* Answer questions directly within this document.
* Upload to Blackboard by the due date & time.
* Clearly indicate your answers to all questions.
* If a question requires analysis, attach all relevant output to this document in the appropriate area. Do not attach superfluous output.
* There are 2 questions and 30 points possible.

Ultrasounds are typically inaccurate for predicting the weight of a newborn baby (<https://www.verywellfamily.com/can-they-tell-how-big-the-baby-is-by-ultrasound-2758737>). Dr. Stollen was curious about whether resident physicians improve their estimates of fetal weight in their second vs. first year of residency. She collected data on estimated and actual fetal weight from 251 estimations, located in *fetal\_weight.csv*.

Dr. Stollen was concerned because the accuracy of fetal weight estimations may depend on the actual birth weight of the baby.

**Data Dictionary**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Meaning** | **Coding** |
| bw | Birthweight; weight of newborn (g) |  |
| year | Year of residency of physician making estimate |  |
| ap.sqrt | “estimation inaccuracy”; square root of absolute value of percent inaccuracy, defined as: |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question 1** |  |  |  |  | [17 points] |  |

We will first examine the functional form of birthweight in its relationship with estimation inaccuracy, and produce 3 possible models.

|  |  |
| --- | --- |
|  | 1a. [3 points] Create a scatter plot of ap.sqrt (Y) vs. bw (X) with a best-fit line. What are your impressions of this relationship? |

|  |  |
| --- | --- |
|  | 1b. [3 point] Using the hierarchical polynomials approach, create a regression model for the relationship between ap.sqrt and birthweight. Report the model, r-squared, adjusted r-squared, and predicted r-squared for Model 1 below. |

|  |  |
| --- | --- |
|  | 1c. [3 points] Using the fractional polynomials approach, create a regression model for the relationship between ap.sqrt and birthweight. Report the model, r-squared, adjusted r-squared, and predicted r-squared for Model 2 below. |

|  |  |
| --- | --- |
|  | 1d. [3 points] Create a set of dummy variables for birthweight quintile (i.e., 5 categories). Create a regression model for the relationship between ap.sqrt and birthweight category. Report the model, r-squared, adjusted r-squared, and predicted r-squared for Model 3 below.  Hint: gtools::quantcut(). |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Equation | R-squared | Adjusted R-squared | Predicted R-squared |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |

|  |  |
| --- | --- |
|  | 1e. [5 points] For each of the three models above, create a 1-2 sentence description of the relationship between estimation inaccuracy and birthweight. Use the model parameter estimates in your interpretation. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question 2** |  |  |  |  | [13 points] |  |

Determine whether year of residency is related to estimation inaccuracy.

|  |  |
| --- | --- |
|  | 2a. [8 points] Use a regression model to determine the relationship between year of residency and ap.sqrt. Then, determine if birthweight confounds the relationship between year of residency and ap.sqrt, using each of the three functional forms from Question 1. Assume that birthweight can be related to year of residency. |

|  |  |  |
| --- | --- | --- |
| Adjustment |  | P |
| Unadjusted |  |  |
| BW (Model 1 – hierarchical polynomials) |  |  |
| BW (Model 2 – fractional polynomials) |  |  |
| BW (Model 3 – categorical) |  |  |

|  |  |
| --- | --- |
|  | 2b. [2 points] Does birthweight appear to confound the relationship between year of residency and estimation inaccuracy? Why? |

|  |  |
| --- | --- |
|  | 2c. [3 points] Considering Dr. Stollen’s research question is how year of residency relates to estimation inaccuracy, which model’s parameter estimate would you report? |