Craig Shaffer

Dr. Kristin Kang

STAT 341

12-7-2022

Project Two

**Final Model:**



Table

Description automatically generated

**Reasoning for selecting the model:**

I chose this model for a few reasons. First of all, it achieved the highest adjusted R2 of any of the models. Which means it did the best at explaining the variance of charges compared to the other models. Another reason for selecting this model was the residual plot. All of the plots for the models had issues with the residual plot. But I thought this model had the least egregious violation. Ultimately, I thought it was the best model for predicting charges because it was relatively simple and ended up being the best at predicting charges.

**Model Interpretation:**

|  |  |
| --- | --- |
| **Coefficients** |  |
| Intercept | The intercept of -10864.661 tells us that when all predictors are 0 (age, bmi, etc.) then we would predict the charges to be -10864.66 dollars.  Overall, this is not a meaningful quantity because the hospital will never pay you and it is most likely impossible to have 0 in every variable. |
| age | The coefficient for age (263.154) tells us that two people sharing the same qualities (bmi, children, smoker, etc.) but differ in age by 1 will differ in charges by 263.15 dollars on average. |
| bmi | The coefficient for bmi (546.683) tells us that two people sharing the same qualities (age, children, smoker, etc.) but differ in bmi by 1 will differ in charges by 546.68 dollars on average. |
| I(bmi^2) | The coefficient for bmi^2 (-8.583) tells us that two people sharing the same qualities (age, children, smoker, etc.) but differ in bmi^2 by 1 will differ in charges by -8.58 dollars on average. |
| children | The coefficient for children (509.302) tells us that two people sharing the same qualities (age, bmi, smoker, etc.) but differ in children by 1 will differ in charges by 509.30 dollars on average. |
| smokeryes | The coefficient for smokeryes (-20260.804) tells us that two people sharing the same qualities (age, bmi, children, etc.) but differ based on yes or no smoker, smokeryes have 20260.80 dollars less in charges than the reference level (smoker no). |
| bmi:smokeryes | The slope changes for (increases) by 1436.78 for each 1 increase in bmi relationship between smokeryes and charges |
| **Predictor Significance** | All predictors in the model have p-values under 0.05. Therefore, they are significant to the model. |
| **RMSE** | Residual standard error: 4853  The RMSE of 4853 tells us that the model’s typical error is 4853 when trying to predict charges. The average error committed by the model when predicting charges is 4853 dollars. |
| **Adjusted R2** | Adjusted R-squared: 0.8394  The adjusted R-squared of 0.8394 tells us that 83.9% of the variation in charges can be explained by the model. |
| **Statistical Significance** | F-statistic: 1166 on 6 and 1331 DF, p-value: < 2.2e-16  The model’s F-stat of 1166 gives a p-value of ~0. Since the p-value is less than 0.05, the model is statistically significant. |