Linear Regression Analysis

Final Exam

Nick Morris

12/14/15

Model Description

Fitted Equation:

The final model chosen to predict Thermal Load is given below. There are 8 terms, the intercept, and a transformed response. The first term is Compactness, with a negative effect on the response, there is an expected 0.7614 unit decrease in the square root of Thermal Load for every 1 unit increase in Compactness. This negative effect is supported by the drop in Thermal Load for the largest values of Compactness. The second term is Orientation(O3), with a negative effect on the response, there is an expected 0.005314 unit decrease in the square root of Thermal Load if the building has an O3 orientation The third term is Orientation(O4), with a negative effect on the response, there is an expected 0.002463 unit decrease in the square root of Thermal Load if the building has an O4 orientation. The fourth term is Orientation(O5), with a negative effect on the response, there is an expected 0.02791 unit decrease in the square root of Thermal Load if the building has an O5 orientation. The fifth term is a Compactness and Surface interaction, with a positive effect on the response, there is an expected 0.1006 unit increase in the square root of Thermal Load for every 1 unit increase in Compactness and Surface. This positive effect is supported by the positive slopes seen in Figure 1 below. The sixth term is a Compactness and Glazing interaction, with a positive effect on the response, there is an expected 3.564 unit increase in the square root of Thermal Load for every 1 unit increase in Compactness and Glazing. This positive effect is support by the positive slopes in Figure 2 below. The seventh term is a Compactness, Surface, and Wall interaction, with a positive effect on the response, there is an expected 0.000008294 unit increase in the square root of Thermal Load for every 1 unit increase in Compactness, Surface, and Wall. This slight positive effect is supported by the even spread of points in Figure 4 below. The eighth term is a Glazing, Wall, and Roof interaction, with a negative effect on the response, there is an expected 0.00002878 unit decrease in the square root of Thermal Load for every 1 unit increase in Glazing, Wall, and Roof. This slight negative effect is supported by the even spread of points in Figure 3 below. There is a set of 3 dummy variables, Orientation(O3), Orientation(O4), and Orientation(O5). They're described below the fitted equation.

$$\begin{aligned}\sqrt{\hat{Y}_{Thermal\ Load}} = & -46.53 - 0.7614 * X_{Compactness} - 0.005314 * X_{Orientation(O3)} \\ & - 0.002463 * X_{Orientation(O4)} - 0.02791 * X_{Orientation(O5)} \\ & + 0.1006 * X_{Compactness} * X_{Surface} + 3.564 * X_{Compactness} * X_{Glazing} \\ & + 0.000008294 * X_{Compactness} * X_{Surface} * X_{Wall} \\ & - 0.00002878 * X_{Glazing} * X_{Wall} * X_{Roof}\end{aligned}$$

$$s.t.\ X_{Orientation(O3)} = 1\ if\ O3; 0\ o.w.$$

$$s.t.\ X_{Orientation(O4)} = 1\ if\ O4; 0\ o.w.$$

$$s.t.\ X_{Orientation(O5)} = 1\ if\ O5; 0\ o.w.$$

$$s.t. X_{Orientation(O3)} = X_{Orientation(O3)} = X_{Orientation(O3)} = 0\ if\ O2$$

Interaction Plots:

Figure 1 below is a two-way interaction plot of the Compactness:Surface term. This significant interaction is supported by the non-parallel lines. The slope of higher values of Compactness increase at a larger rate across the values of Surface in terms of Thermal Load in original units. Figure 2 below is a two-way interaction plot of the Compactness:Glazing term. This significant interaction is supported by the non-parallel lines. The slope of higher values of Compactness increase at a slightly larger rate across the values of Glazing in terms of Thermal Load in original units. Figure 3 below is a scatterplot of Thermal Load in original units v. Wall, colored by Glazing, and faceted by Roof. This significant interaction is supported by the distinct groups across Wall values due to Roof values, and groups across Thermal Load values due to Glazing values. Figure 4 below is a scatterplot of Thermal Load in original units v. Wall, colored by Compactness, and faceted by Surface. This significant interaction is supported by the distinct groups across Wall values and Thermal Load values due to the Compactness and Surface values.

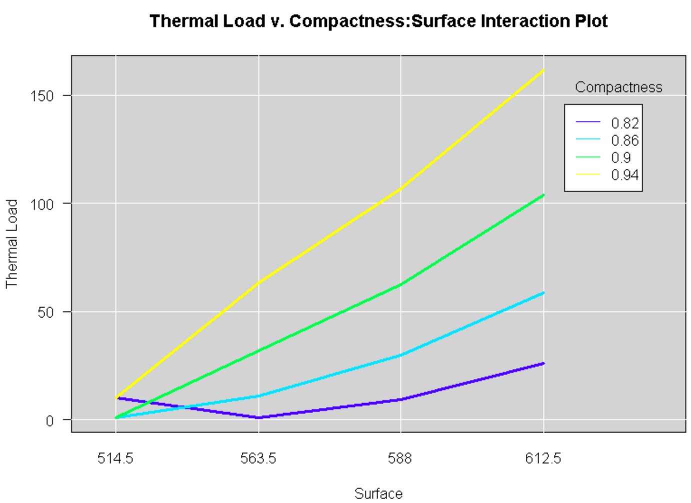

Figure 1: Compactness:Surface Interaction

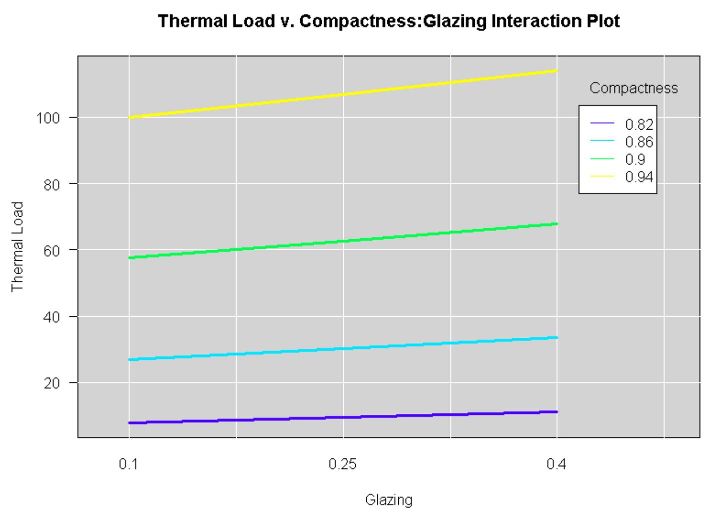

Figure 2: Compactness:Glazing Interaction

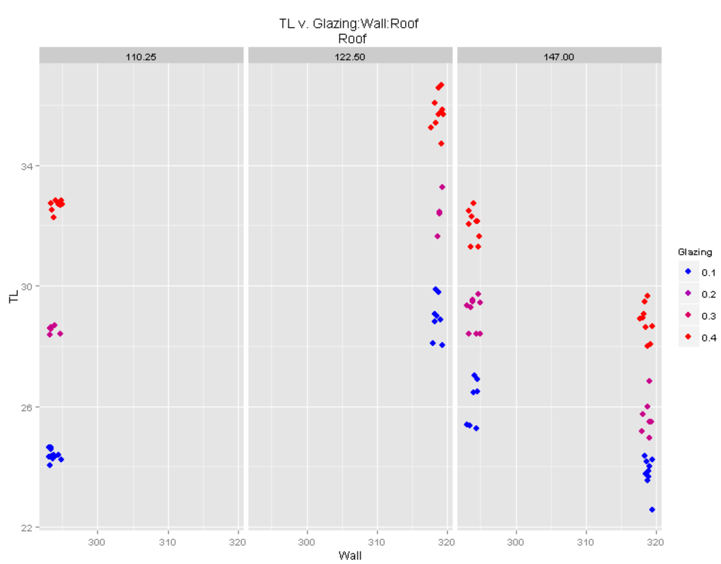

Figure 3: Glazing:Wall:Roof Interaction

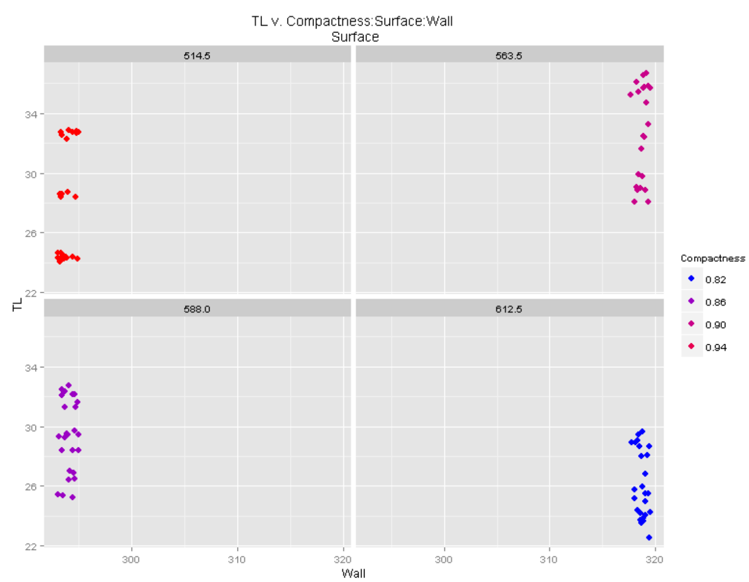

Figure 4: Compactness:Surface:Wall Interaction

Major Assumptions:

Table 1: Major Assumptions Conclusions

Assumption	Violation
2	FALSE
3	FALSE
4	FALSE
6	FALSE

The results of the six major assumptions are given on the right. Assumption 2, constant variance, is met due to the graph on the top-left corner of Figure 5 below. This residuals vs. fitted plot shows that the distribution of residuals is broken into five distinct groups due to the discrete nature of the continuous variables. There is a constant vertical dispersion across fitted values, supporting this assumption being met. Assumption 3, normality, is met due to the graphs on the bottom-left and bottom-right corners of Figure 5. The tails of the normal q-q plot trail off slightly and this is supported by the kernel density plot with more density in the tails than the ideal normal distribution. The center of the data falls along the normal q-q plot line and the kernel density plot shows a bell shape, supporting the normality of the residuals. Assumption 4, the residuals are uncorrelated, is met due to the graph in the top-middle of Figure 5. The Residuals v. Residuals Lag plot visualizes the relationship within the residuals as it is evaluated in the Durbin Watson test with a Lag of one. This graph shows insignificant correlation within the residuals.

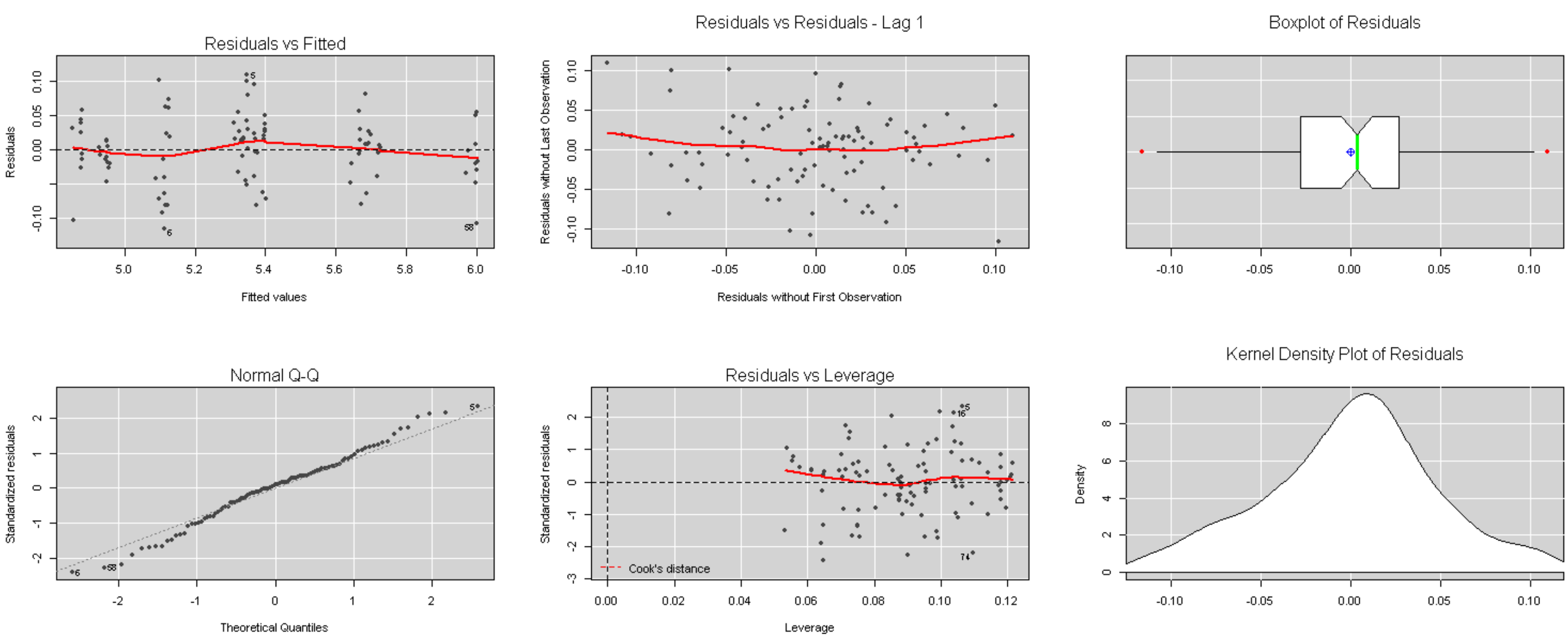

Figure 5: Residuals Plots of Thermal Load Linear Regression Model

The analytical tests performed for the major assumptions are shown below in Table 2. The results are in agreement with the final conclusions made in Table 1. The Chi Square Test for non-constant variance fails to reject the null hypothesis and concludes that the residuals have constant variance. The Shapiro-Wilk normality test fails to reject the null hypothesis and concludes that the residuals come from a normal distribution. The Durbin Watson test fails to reject the null hypothesis and concludes that the residuals are uncorrelated. The general variance inflation factors calculated in the right hand column show moderate inflation for terms Compactness:Glazing and Glazing:Wall:Roof, but are less than the rule of thumb value 10, which if over 10 would indicate a violation of assumption 6 due to severe inflation.

Table 2: Analytical Tests for Major Model Assumptions

```
Assumption 2 - Residuals have Constant Variance - P-Value > 0.05

Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.4643517    Df = 1     p = 0.4955972

Assumption 3 - Residuals are Normally Distributed - P-Value > 0.05

      Shapiro-Wilk normality test

data:  LM$residuals
W = 0.98825, p-value = 0.5265

Assumption 4 - Residuals are Uncorrelated - P-value > 0.05

 lag Autocorrelation D-W Statistic p-value
   1      -0.1251253      2.244524   0.226
 Alternative hypothesis: rho != 0

Assumption 6 - The Regressors are Independent - All vif < 10

                              GVIF Df GVIF^(1/(2*Df))
Compactness               5.122022  1        2.263188
Orientation               1.092425  3        1.014842
Compactness:Surface       1.133111  1        1.064477
Compactness:Glazing      25.417744  1        5.041601
Compactness:Surface:Wall  1.438233  1        1.199264
Glazing:Wall:Roof        27.906859  1        5.282694
```