Analysis - Predicting the Weather

Code

Soren Laney

September 27, 2022

Code

Hypothesis

This data was collected from this website(https://www.timeanddate.com/weather/@5605242/historic?month=1&year=2022 (https://www.timeanddate.com/weather/@5605242/historic?month=1&year=2022)) on the 15th of September 2022. I gathered the date, Wind speed, Temp and Humidity from the 1st of September until the 15th of September. The data was collected for the purpose of predicting the weather on Monday the 19th of September. The temperature will be estimated by using the lag data from two days prior to predict the temperature using a linear regression model.

Hypothesis - The Null hypothesis is that there not be a correlation between the high temperature two days in the future and today's current high temperature. The Alternative being that there is a correlation that indicates what the temperature will be two days in the future from today's high.

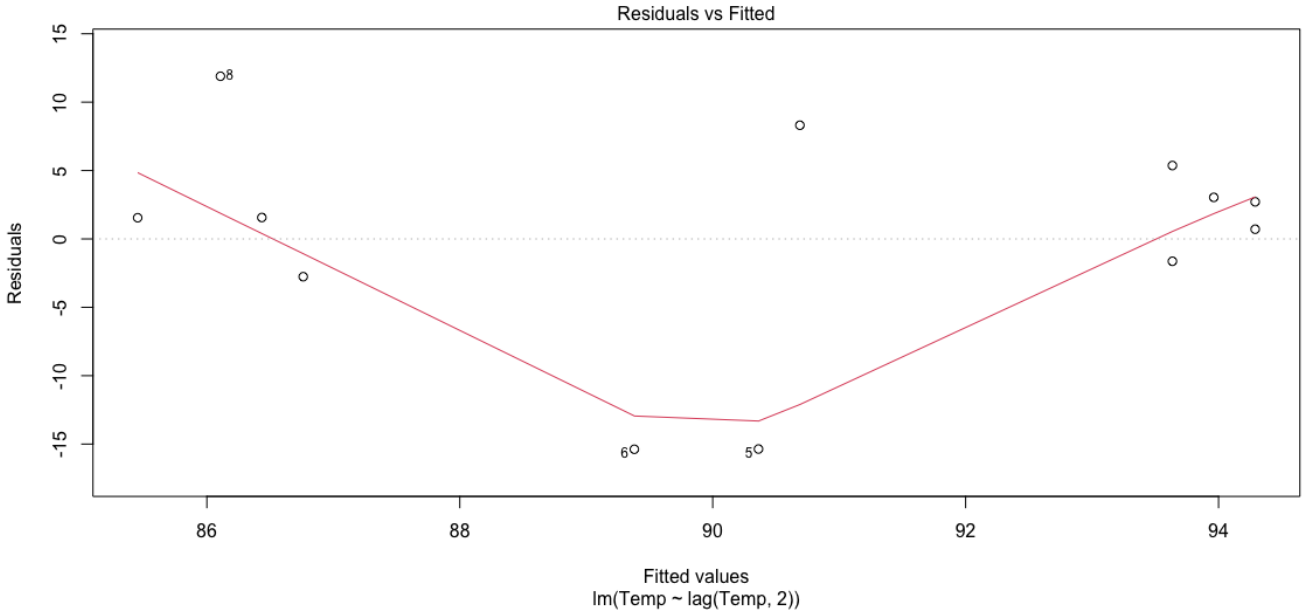
The true regression model assumed by a regression analysis is given by

$$\underbrace{Y_i}_{\text{Predicted Temperature}} = \overbrace{\beta_0}^{\text{y-int}} + \overbrace{\beta_1}^{\text{slope}} \underbrace{X_i}_{\text{High Temp}} + \epsilon_i \quad \text{where } \epsilon_i \sim N(0, \sigma^2)$$

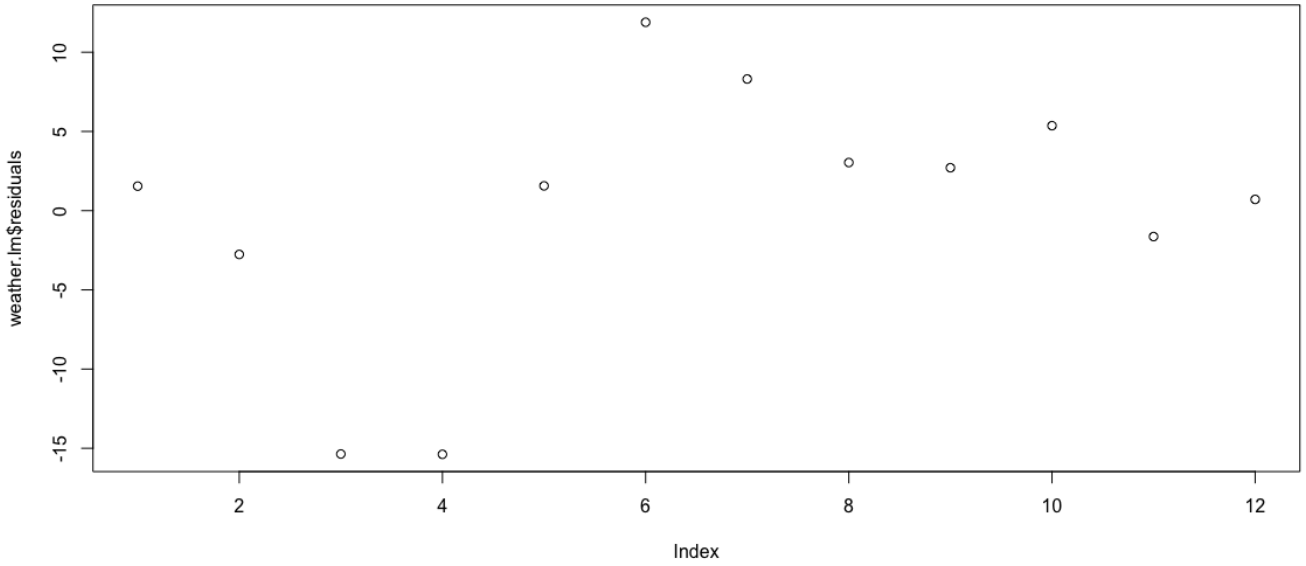
With Beta 0 representing the temperature high and the Beta 1 representing the lag temperature.

Regression Appropriateness

Code



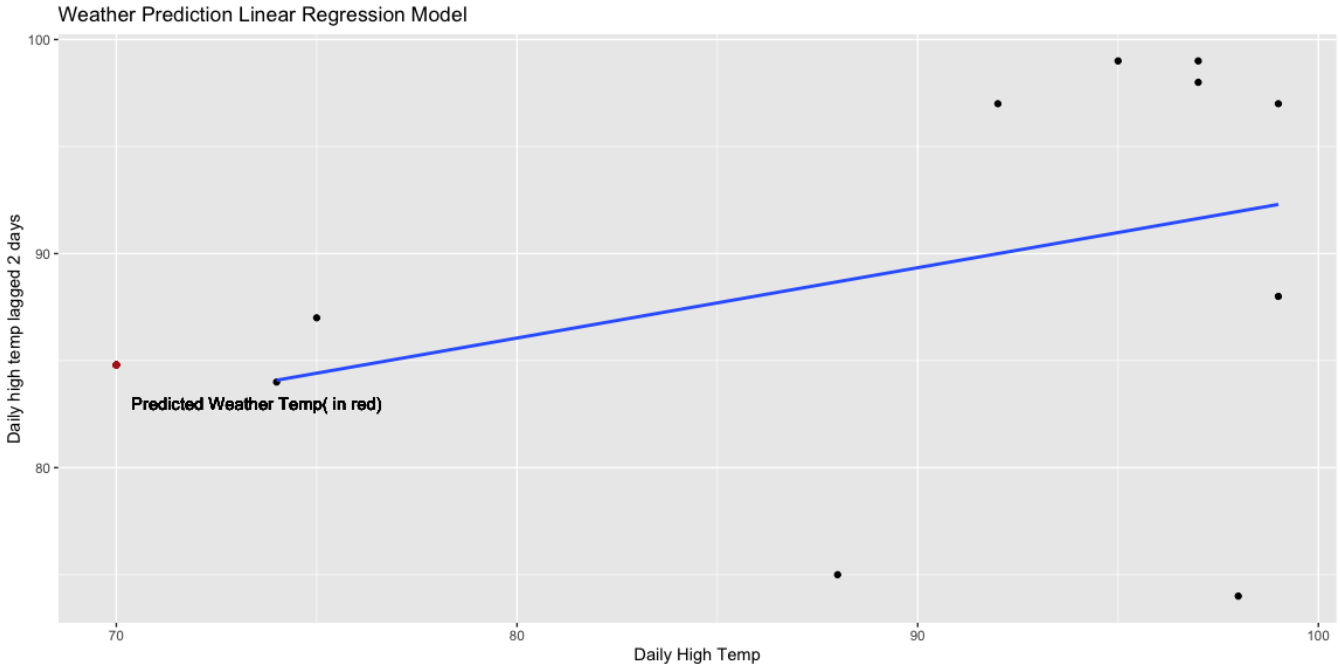
Code



Code

Analysis

Code



Code

Code

Observations	12 (2 missing obs. deleted)
Dependent variable	Temp
Type	OLS linear regression

F(1,10)	1.84
R²	0.16
Adj. R²	0.07

	Est.	S.E.	t val.	p
(Intercept)	61.89	21.18	2.92	0.02
lag(Temp, 2)	0.33	0.24	1.36	0.20

Standard errors: OLS

Code

As we can see above, the graph represents the temperature plotted against what the temperature will be in two days. The regression is represented by β_0 being equal to 61.89 or the base temperature on a given day in the data set. Therefore β_1 is equal to 0.33 which indicates that for every degree, plus or minus of temperature two days prior, the high temp will be multiplied by that or X_i to predict the additional temperature change of that days high.

Conclusions

I predict based on the linear model that the high temperature monday will be 84.8 degrees Fahrenheit.

Upon analyzing the data it appears that to be normal under a QQ plot and scatter. However, under the residuals, it appeared to not be normal. It curves sharply down and then back up making a v-shape. Due to this I do not beleive that my data bring a significant result.

To predict the high temp on Monday I used the temp plotted against itself on a 2 day lag so that I could use the high from Saturday to predict the temp on Monday. The estimated equation from the analysis is as seen below.

The estimated regression line obtained from a regression analysis, pronounced “y-hat”, is written as

$$\underbrace{\hat{Y}_i}_{\text{Predicted Temp}} = \overbrace{b_0}^{\text{est. y-int}} + \overbrace{b_1}^{\text{est. slope}} \underbrace{X_i}_{\text{High Temp}}$$

In this equation Beta zero would be being 61.88 and the est. slope of Beta One being 0.3273. The equation that I used to predict the high on Monday was sum(coef(weather.lm)*c(1, 70)) = 84.7982. After preforming a summary statistic on the linear regression, I found there to be a p-value of .2049 and a standard error of8.626, making the result not significant or highly concentrated. This would also be a Type 1 error as the p-value does not have significance and I will fail to reject the null hypothesis. This is because of the data I choose to use in my linear regression. I picked the the dates of the 1st to the 15th to use. The temperature of the last two weeks was not similar to the weather leading into this weekend. This leads me to believe that my prediction of the high being roughly 84.8 degrees on monday will not be accurate.