Predicting Survival

Cody Frisby

10/21/2017

1)

The table below includes a summary of the linear model. The rightmost column includes the *pvalue* of the statistical test of each term in our model. Those with a small *pvalue* are deemed "significant".

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6766.548205	1272.730166	5.3165615	0.0000007
Year	-87.485499	17.730881	-4.9340750	0.0000033
Age	-1.475815	3.335759	-0.4424226	0.6591684
Survival	-541.083040	79.146193	-6.8365011	0.0000000
Surgery	304.383110	100.695803	3.0227984	0.0032038
Transplant	218.514209	70.645519	3.0931079	0.0025872

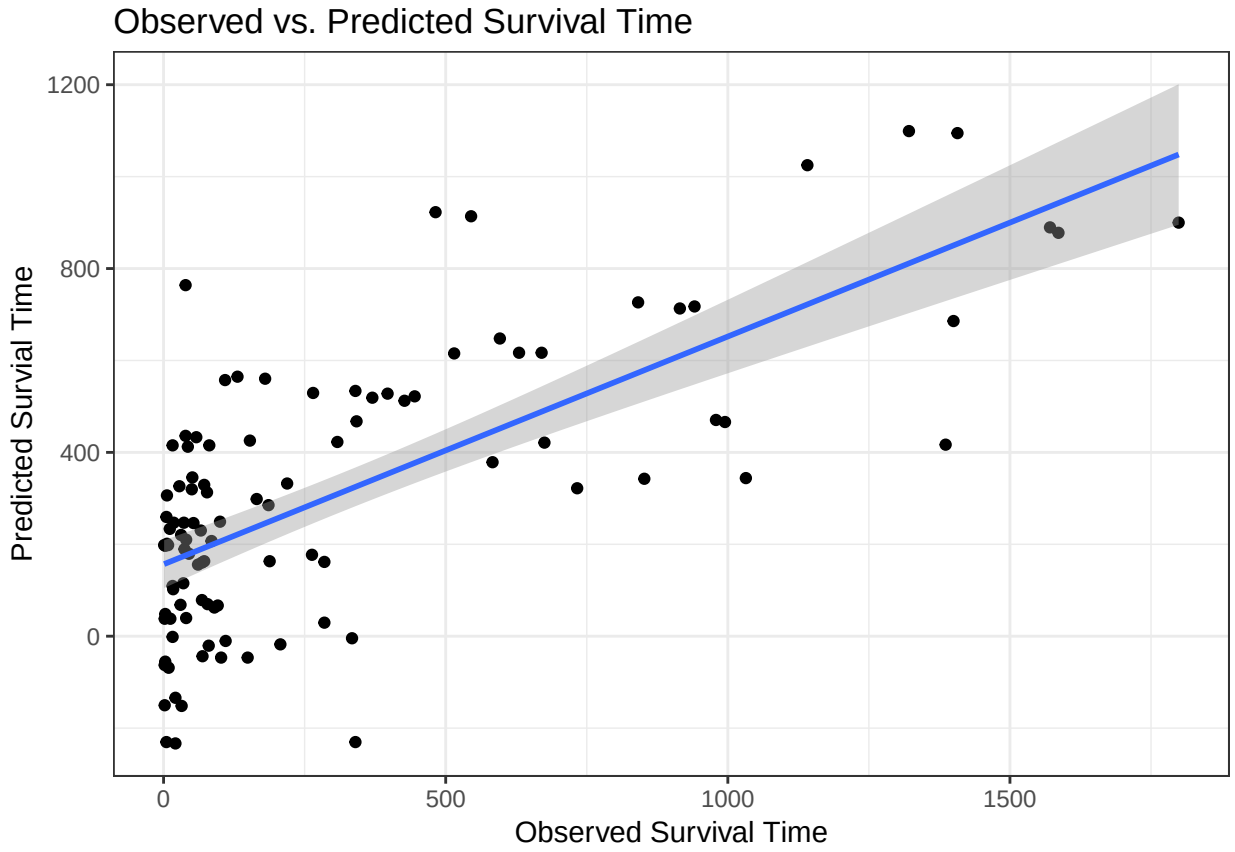
As can bee seen, **Survival** (1 means they passed away and 0 means they survived) is a significant predictor of **SurvivalTime** (this is kind of obvious), as well as **Year** of acceptance, **Surgery** (1 if they had prior surgery, 0 if not), and **Transplant** (1 if they had transplant, 0 if not). **Intercept** test at 6766.548205 vs. 0 is also significant, but the test is whether or not the intercept is 0 so this isn't suprising. Age of patient is not a strong indicator of **SurvivalTime**.

2)

Where $Age = 54$, $Year = 69$, $Survival = 0$, $Surgery = 1$, and $Transplant = 0$ the model predicts a **SurvivalTime** of 954.74.

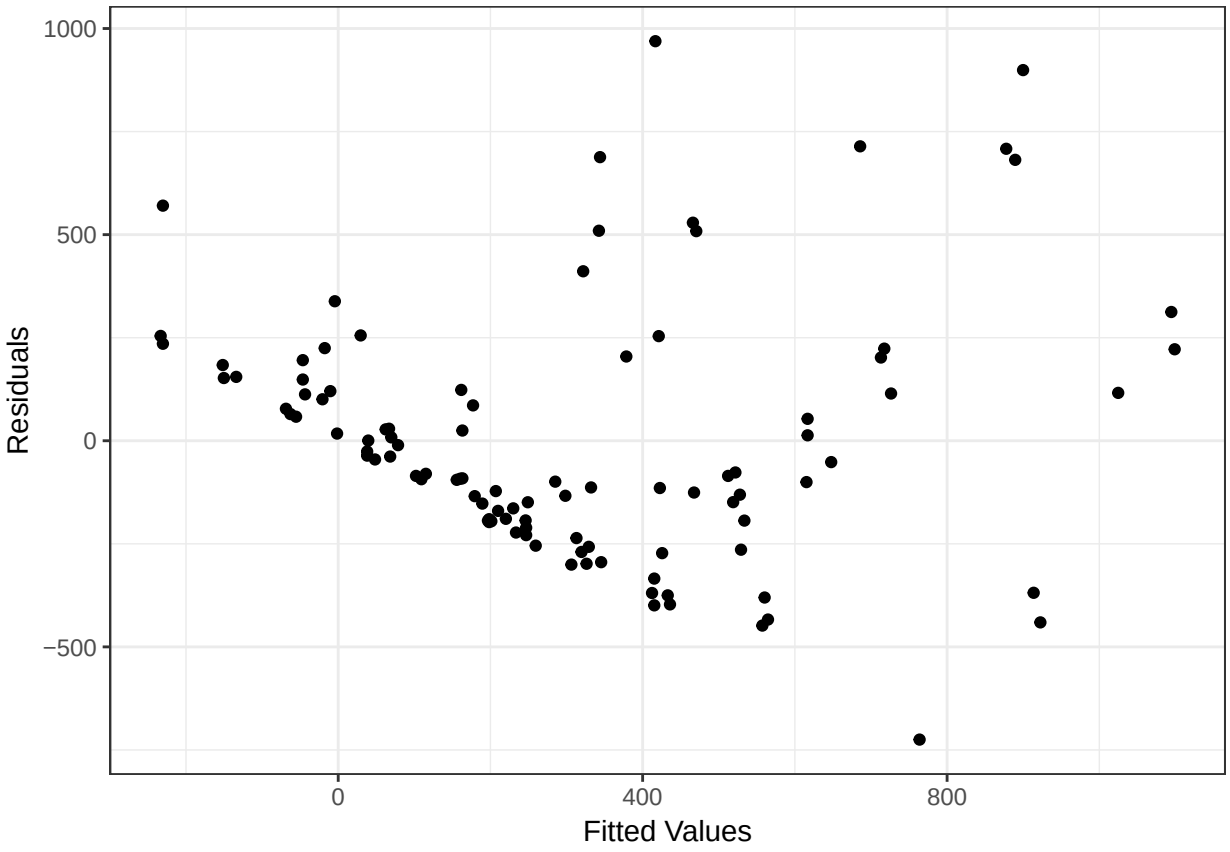
3)

Below is a plot of observed survival time vs. predicted. The gray shaded area on the plot is the confidence interval. As can be seen, the gray bar gets sider the further to the right it goes. This means that our confidence interval is getting wider. For the individual above, the prediction interval for their survival is $[273.36, 1636.12]$. This is quite a wide interval. Basically we are saying we are 95% confident that this individual will survive between 273 and 1636 days.



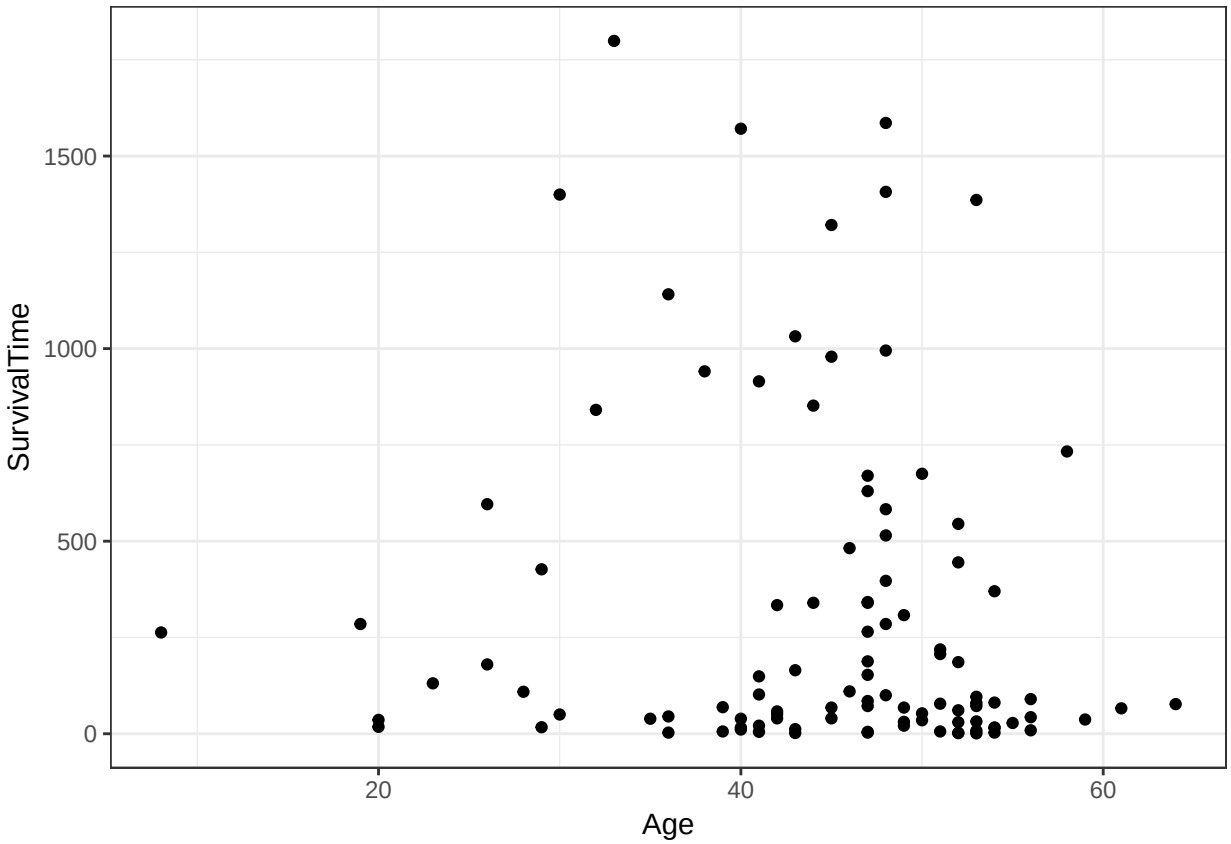
The R^2 value for *observed vs. predicted* is 0.495644.

Furthermore, the plot of the model's residuals vs. the predicted values is concerning. The assumption of equality of variance is supect here. And as the plot below shows, the model is predicting quite a few negative values for **SurvivalTime**. If we are in need of a model that has a much smaller window of prediction confidence, this model is not it.



4)

A plot of **SurvivalTime** vs. **Age** is shown below. It appears to me to be mostly noise. It is interesting that the largest survival times appear to be those who are middle aged. This could be simply due to the fact that most of our observations include ages between 41 and 52.


SurvivalTime
1500
1000
500
0
20
40
60
Age