East Carolina University

**Spring 2020 REGRESSION MODEL**

**Project 2 MIS 6913**

**Liberte Kamdem**

**Statistic for business Analytics MIS6913**

**Dr. Richard Hauser**

# Table of Contents

# Introduction……………………………………………………………………………………….3

# Background……………………………………………………………………………………….4

# Research Methodology…...………………………………………………………………………5

# State Linear mode……………………………………………………………………………….7

# Limitation of the model…………………………………………………………………………13

# Conclusion……………………………………………………………………………………....14

# Appendices………………………………………………………………………………………15

# References……………………………………………………………………………………….23

**Introduction**

Since the millennium, the earth has been covered with nearly 71% of water. This situation forced human beings to create means of transport like the primitive boat as a means of transportation since the primary era to conquer this vulnerable space. Since then, these primitive boats have gone from the simple tree trunk used to balance the user when crossing the river to the best means of transporting goods in intercontinental trade. With the industrial revolution and the popularization of the new mode of transport like planes, trains, and cars at the beginning of the XIXth century, the boat considerably reduced its hegemony in the transportation of merchandise. This reduction in freight transport did not mark the end of the importance of the boat in modern society, but on the contrary, it propelled the boat towards the luxury tool. The boat has become more useful as a luxury tool for the millionaire, billionaire, and those who pretend. It allows raising the social status of those who own it like Jeff Bezos' yacht by 400 million dollars.

Knowing how expensive the boat has become and how valuable it is to maintain, only the wealthy can dream of affording it.

This reality is even more remarkable, with a reduction in the line of credit offered by the various financial institutions for the purchase of boats. And the high and considerable risk of maritime accidents. This scarcity of funding leads some less wealthy to fall back on used boats to belong to the minimal club of boat owners.

The right question is whether it is worth falling back on the used boat market?

Is it necessary, or do they have real criteria that can guide us in buying a specific brand with long-term equity?

Knowing that the average amortization period varies between four and seven years, we will focus on two brands Bayliner and Grady white, to analyze the result.

**Background**

The Bayliner brand is the larger manufacturer of the recreational boat in the world. It was established in 1957 and represent within 6o countries with 400 dealers. It offers 25 different models with 175 bowriders as a favorite among first-time boat buyers. As a leader for more than 60years, it has of outstanding value and loyal customer who already come back for more significant and innovative

Grady White has a reputation for design and producing exceptional fiberglass boats. Since 1959, Grady White has been built in Greenville, NC. Their goal from the beginning is to deliver quality, reliable, safety, and performance boats. Also, develop a significant relationship between company and customer. The company has the most comprehensive line up of center console. Specialize of a fishing sport boat, Grady white boat is designed around 15 models with a lot of fisherman need in mind, as generous storage space, extra freeboard, roomy interior, and sometimes walking around to make it easy to bring in big fish.

**Research Methodology**

Data used in the report to analyze boat purchase was defined by the instructor as follows in the table below. We focused on specific brands: Bayliner and Grady White. Due to time and financial constraints, we focus on secondary data. We have collected this data on various websites specializing in the sale of boats such as www.boattrader.com, www.baot.com, [www.ebay.com](http://www.ebay.com).

Table 1 Boat characteristic

|  |  |  |
| --- | --- | --- |
| Criteria | Parameter | Coding |
| Age of boat | YEAR | QUANTITATIVE |
| Length | FEET | QUANTITATIVE |
| Motor size (HP) | HORSEPOWER | QUANTITATIVE |
| Number of Engines | TOTAL AMOUNT OF ENGINE | QUANTITATIVE |
| Powerboat Type | CLASS | QUALITATIVE |
| Brand of Boat | MANUFACTURE | QUALITATIVE |
| Motor type (code IB, OB) | MOTOR LOCATION | QUALITATIVE |
| Asking price (in $) | US DOLLARS | QUANTITATIVE |

Table 2 statistic table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Categorie by  brand | Mean | median | Standard Deviation | range | Max | min |
| Age  of  BL | 3.29 | 2 | 2.782 | 9 | 9 | 0 |
| Age  of  GW | 4.19 | 4 | 1.96 | 8 | 9 | 1 |
| Length  of BL | 20.04 | 20 | 1.55 | 5 | 23 | 18 |
| Length of GW | 21.38 | 150 | 2.20 | 8 | 26 | 18 |
| Size of Motor  BL | 158.75 | 200 | 52.14 | 170 | 260 | 90 |
| Size of Motor  GW | 233.33 | 27974.5 | 59.86 | 150 | 300 | 150 |
| Asking price  BL | 29583.5 | 27947.5 | 29896.75 | 42819 | 58719 | 159000 |
| Asking price  GW | 74776 | 65900 | 29896.75 | 119000 | 15900 | 40000 |

**T-test result**

A t-test of mean was executed to compare the differences between brands Bayliner and Grady White on age, length, total Horsepower, and price.

The hypotheses for each of the four tests was the same and held an alpha of .05:

Ho: MBL-MGW= 0 Ha: MBL -MGW≠0

Alpha = .05

The table below summarizes the result on different t-text

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Statistic** | **Age** | | | **Length** | | | **TotalHP\*\*** | | | **Price** | | |
| Brand | Bayliner | Grady White | Is GW statistical older | Bayliner | Grady White | Is GW statistical Longer | Bayliner | Grady White | Does GW statistical more HP | Bayliner | Grady White | Does GW statistical cost more |
| Mean | 3.29 | 4.190 | 20.04 | 21.380 | 158.75 | 233.3 | 29583.5 | 74776 |
| Median | 2 | 4 | **yes** | 20 | 20 | yes | 150 | 200 | Yes | 27947.5 | 65900 | yes |
| Std Dev | 2.786 | 1.96 | **significance** | 1.556 | 2.201 | significance | 52.140 | 59.86 | significance | 9664.44 | 29896.754 | significance |

**conclusion**

Significance differences appear on the t-text mean of age, length, total horsepower, and price between Bayliner and Grady Whites. The outcome of the t-text suggests the importance difference these characteristics affect the market value of the different brands of the boat. For that, a customer needs to consider all these characteristics tested when it comes to purchasing a second-hand boat.

**State Linear model**

At the beginning of the study, specific boat characteristics were chosen to define the price tag of the boat, as represented in Table (1). For a potential customer to be able to predict the right price in the future, we must transform this character into a predictor in a linear equation of the boat price Y. such that this predictor can define range price accurately according to specific character customer requests.

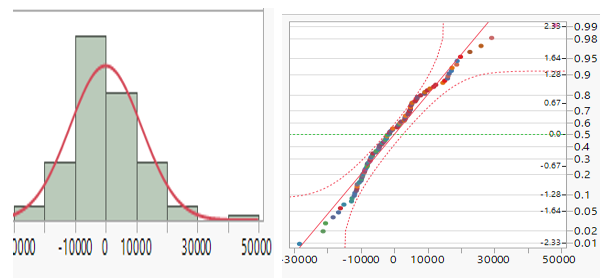
Let beginning by defining a linear equation y with all predictor associates.

Y =β0+β1Xi+β2Xj+ β3Xm + β4Xp + β5xz+€ (1)

where xi, xj, xk, Xm, Xp,and xz, correspond to respectively to Age of boat, length of the boat, Motor size, Motor type, powerboat type and Brand of the boat, with β1, β2 β3, β4, β5 respective coefficient of different predictor.β0 price intercept

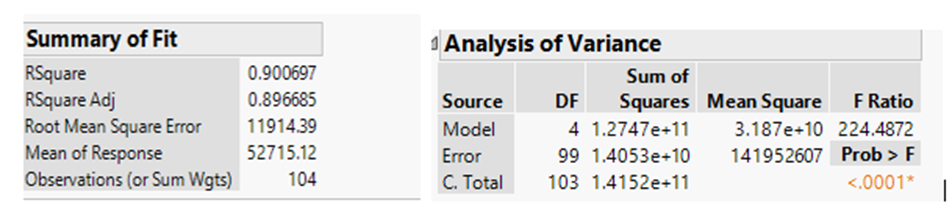
Running the regression model, with predictors enumerated above, we will be able to determine the relationship between a dependent variable and an independent variable.

**State the assumption**



The table above shows that residual is normally distributed. Quantile plot show the set of observation concentrate around the straight line confirming the normal distribution of residual. With a mean of zero and they are independent of one another. We can assert that assumption is respected.

**Text of overall Model utility**



Analysis of variance is used to determine whether the model is significant

**H0 : B1=B2=B3=B4=B5**

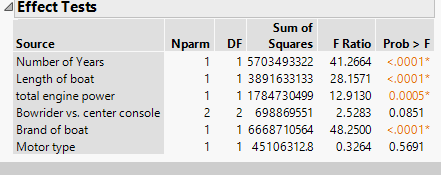
Ha: at least one Bi not equal zero

Alpha=.05 reject if prob>F<.05

Prob>F=.0001, Show that the model is significant with (F<0001) at point .05 level.

The R-square Adj =89.66%, explain the proportion of the variance for a dependent variable that explained by an independent variable in regression model.

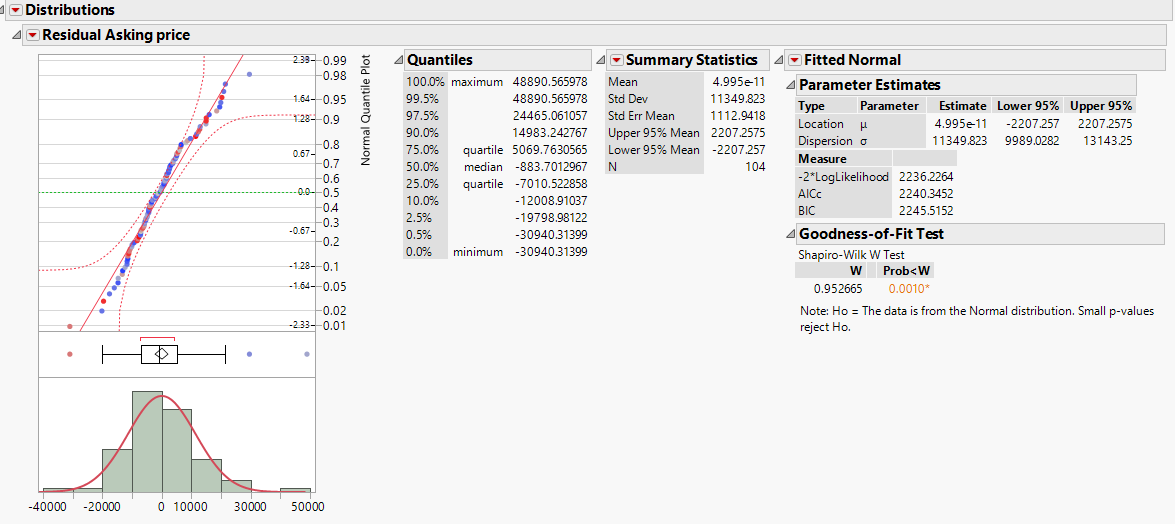
individual independent text term



However, distribution model table show predictors: motor type and powerboat type insignificant and require to be remove from model.

As far as predictor correlation is concern, multivariable analysis table do not present any form strong correlation between predictor. This mean none of predictor cannot be remove from model due to multicollinearity.

**Adjusted model**

First text

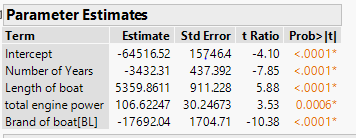
After identifying four significant predictors [Number of years, Boat length, Boat power (HP), and brand of the boat], it becomes easy to state the new linear regression boat price model Y1. The new regression model defining boat price Y1 can be determined as:

Predicted boat price: Y1 =β0+β1Xi+β2Xj+ β3Xm + β4Xp + € (2)

where Xy, Xl, Xe and Xb represented respectively the number of years of the boat, the length of boat, The engine power and the boat brand. Also, β1, β2, β3, β4 represent respectively the slope of Xy, Xl, Xe, Xb and β0 the equation intercept.

According to distribution below, model. price model equation can be state as:

Y**=-64516 - 3432.31** Xy **+** **5359.89** Xl **+ 106.62** Xe **- 17692.04** Xb (BL) **+** €



Hypothesize the model

Let state that the average price of boat is function of number of years, motor side and brand of boat so

let H0= β1= β2= β3= β4=0

Ha= at least 1 β ≠ 0

Α=0.05 reject if prob>F=0.05, Prob>F=.0001

By rejecting the null hypothesize, we conclude that the model is significant.

How good the model is?

Because we have more than one independent variable R2adj is needed to analysis the model.

The R2adj =89.66% mean independent variable explain 89.66% of dependent variable.

The Text of the individual independent term confirm the significance of each

Quantile plot show the set of observation concentrate around the straight line show the normal distribution of residual. As far as predictor correlation is concern, multivariable analysis table do not present any form strong correlation between predictor. This mean none of predictor cannot be remove from model due to multicollinearity.

**Interpretation of individual model**

β1 is negative meaning Age has a negative impact of price boat this confirms the fast price depreciation of the boats.

β2 is positive meaning the length of boat has increase the price of boat.

β3 is positive meaning the engine power has the positive impact on boat price

β4 is negative when it comes to Bayliner. A as far as GW is concern brand does have any impact on predict price.

Limitation of the model:

**Insignificant factor**

The residual factor € independently expresses 11% of the predictor of the linear equation. This expression reduces the precise projection of the linear equation.

Some predictors are insignificant and not represented in the final linear regression model. We may quote, for example, the type of engine, which is very important for anyone who wishes to buy motor vehicles.

**neglect factor**

Some specific predictors that could influence boat prices were not considered when collecting data for this study. We can quote.

The cost of living affects luxury goods in certain regions. If we consider the boat as a luxury item, the cost of living can change the price of a boat depending on the seller's postal code. We can notice that a boat in Seattle with the same character can be more expensive than in Raleigh.

Geography location such as beach town boost demand and supply and therefore increase price this factor was not take in consideration

**conclusion**

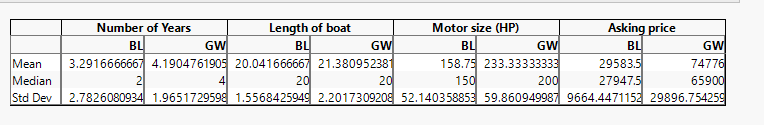
The linear model shows that the predictor factor of the price is robust and may be useful for decision making. This equation can be necessary for any customer who would like to own the boat. However, the potential customer must take seriously into account the factors listed in the model limitation concerning the actual price of the boat.

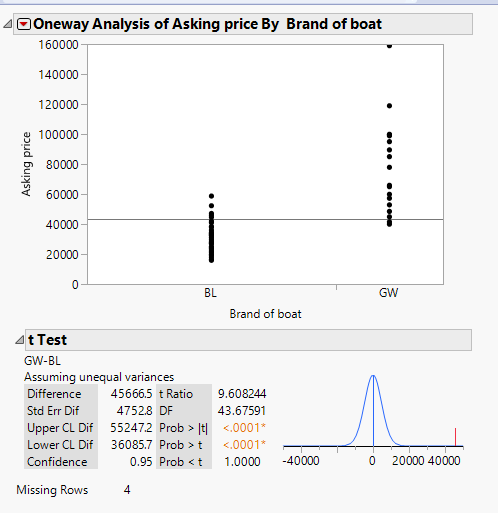
**Appendices**

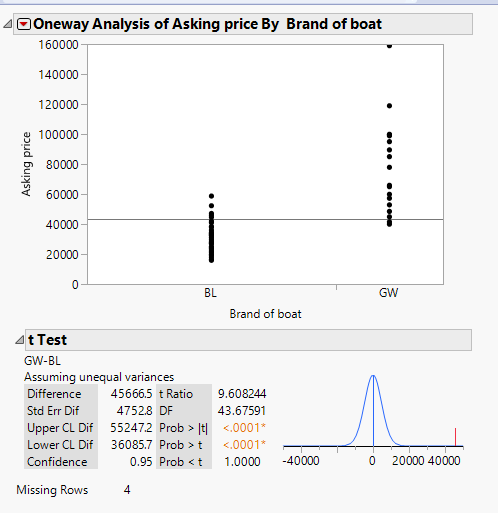
**Data set used**

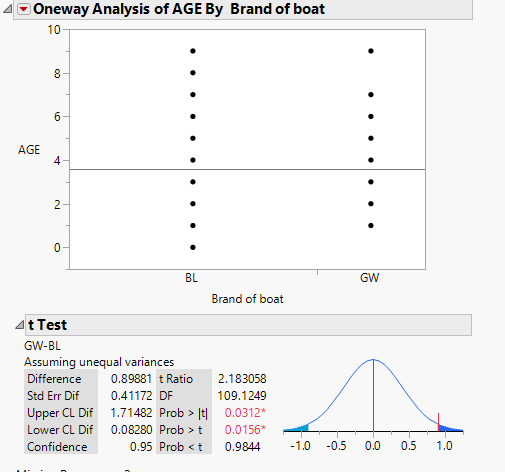
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age of boat | Number of Years | Length of boat | Motor size (HP) | Number of engines | Bowrider vs. center console | Brand of boat | Motor type | Asking price |
| 2016 | 4 | 22 | 220 | 1 | B | BL | IB | 33500 |
| 2011 | 9 | 21 | 220 | 1 | B | BL | IB | 24995 |
| 2019 | 1 | 20 | 250 | 1 | B | BL | IB | 44599 |
| 2011 | 9 | 23 | 260 | 1 | B | BL | IB | 29900 |
| 2020 | 0 | 22 | 200 | 1 | B | BL | IB | 41995 |
| 2019 | 1 | 22 | 200 | 1 | B | BL | IB | 45999 |
| 2016 | 4 | 22 | 200 | 1 | B | BL | IB | 41000 |
| 2017 | 3 | 22 | 200 | 1 | B | BL | IB | 38400 |
| 2012 | 8 | 21 | 220 | 1 | B | BL | IB | 19975 |
| 2014 | 6 | 21 | 220 | 1 | B | BL | IB | 28900 |
| 2011 | 9 | 21 | 220 | 1 | B | BL | IB | 19250 |
| 2013 | 7 | 21 | 260 | 1 | B | BL | IB | 22900 |
| 2018 | 2 | 20 | 200 | 1 | B | BL | IB | 37990 |
| 2016 | 4 | 20 | 200 | 1 | B | BL | IB | 16500 |
| 2012 | 8 | 20 | 135 | 1 | B | BL | IB | 16500 |
| 2019 | 1 | 20 | 200 | 1 | B | BL | IB | 35975 |
| 2016 | 4 | 19 | 220 | 1 | B | BL | IB | 32650 |
| 2018 | 2 | 22 | 200 | 1 | B | BL | OB |  |
| 2020 | 0 | 22 | 200 | 1 | B | BL | OB | 52269 |
| 2020 | 0 | 22 | 200 | 1 | B | BL | OB | 58719 |
| 2019 | 1 | 22 | 200 | 1 | B | BL | OB | 47149 |
| 2015 | 5 | 21 | 150 | 1 | B | BL | OB | 27900 |
| 2019 | 1 | 21 | 115 | 1 | B | BL | OB | 27850 |
| 2018 | 2 | 21 | 115 | 1 | B | BL | OB | 26945 |
| 2017 | 3 | 21 | 115 | 1 | B | BL | OB | 27900 |
| 2015 | 5 | 19 | 115 | 1 | B | BL | OB | 20995 |
| 2015 | 5 | 19 | 115 | 1 | B | BL | OB | 20995 |
| 2018 | 2 | 19 | 115 | 1 | B | BL | OB | 21900 |
| 2014 | 6 | 19 | 150 | 1 | B | BL | OB | 19900 |
| 2013 | 7 | 19 | 150 | 1 | B | BL | OB | 19500 |
| 2020 | 0 | 19 | 115 | 1 | B | BL | OB | 31000 |
| 2019 | 1 | 19 | 115 | 1 | B | BL | OB | 27995 |
| 2015 | 5 | 18 | 115 | 1 | B | BL | OB | 15900 |
| 2014 | 6 | 18 | 135 | 1 | B | BL | OB | 17999 |
| 2018 | 2 | 18 | 115 | 1 | B | BL | OB | 28755 |
| 2020 | 0 | 18 | 90 | 1 | B | BL | OB | 28734 |
| 2017 | 3 | 24 | 300 | 1 | B | GW | OB | 99500 |
| 2018 | 2 | 23 | 300 | 1 | B | GW | OB | 118900 |
| 2019 | 1 | 21 | 250 | 1 | B | GW | OB | 77900 |
| 2016 | 4 | 20 | 200 | 1 | B | GW | OB | 59900 |
| 2016 | 4 | 20 | 150 | 1 | B | GW | OB | 52857 |
| 2016 | 4 | 20 | 150 | 1 | B | GW | OB | 59995 |
| 2013 | 7 | 20 | 200 | 1 | B | GW | OB | 44900 |
| 2015 | 5 | 19 | 150 | 1 | B | GW | OB |  |
| 2012 | 8 | 21 | 150 | 1 | CC | BL | OB | 35999 |
| 2019 | 1 | 21 | 150 | 1 | CC | BL | OB | 32995 |
| 2018 | 2 | 21 | 150 | 1 | CC | BL | OB | 29900 |
| 2020 | 0 | 21 | 150 | 1 | CC | BL | OB | 34567 |
| 2019 | 1 | 18 | 90 | 1 | CC | BL | OB | 24597 |
| 2019 | 1 | 18 | 90 | 1 | CC | BL | OB | 24997 |
| 2019 | 1 | 18 | 90 | 1 | CC | BL | OB | 24615 |
| 2019 | 1 | 18 | 90 | 1 | CC | BL | OB | 27359 |
| 2018 | 2 | 18 | 90 | 1 | CC | BL | OB | 21997 |
| 2017 | 3 | 18 | 115 | 1 | CC | BL | OB | 19500 |
| 2017 | 3 | 18 | 115 | 1 | CC | BL | OB | 32200 |
| 2018 | 2 | 18 | 90 | 1 | CC | BL | OB | 23849 |
| 2016 | 4 | 26 | 300 | 1 | CC | GW | OB | 159000 |
| 2015 | 5 | 25 | 300 | 1 | CC | GW | OB | 95000 |
| 2016 | 4 | 25 | 300 | 1 | CC | GW | OB | 89500 |
| 2014 | 6 | 23 | 300 | 1 | CC | GW | OB | 85000 |
| 2017 | 3 | 23 | 300 | 1 | CC | GW | OB | 99900 |
| 2018 | 2 | 21 | 250 | 1 | CC | GW | OB | 64995 |
| 2018 | 2 | 21 | 200 | 1 | CC | GW | OB | 98900 |
| 2014 | 6 | 20 | 200 | 1 | CC | GW | OB | 48500 |
| 2016 | 4 | 20 | 200 | 1 | CC | GW | OB | 65900 |
| 2013 | 7 | 20 | 300 | 1 | CC | GW | OB | 40000 |
| 2011 | 9 | 20 | 200 | 1 | CC | GW | OB | 41499 |
| 2017 | 3 | 20 | 200 | 1 | CC | GW | OB | 66000 |
| 2017 | 3 | 18 | 150 | 1 | CC | GW | OB | 44900 |
| 2020 | 0 | 18 |  |  |  |  |  |  |

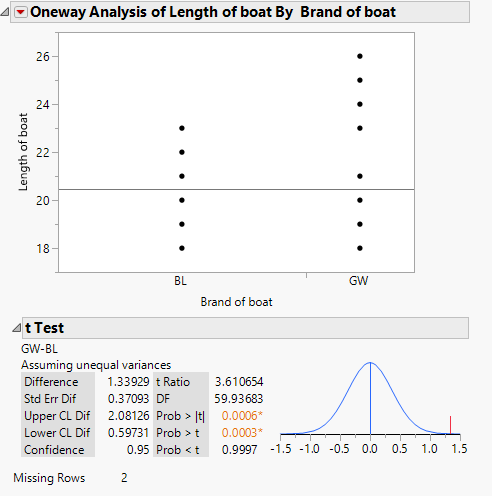
**Mean, median standard deviation table**

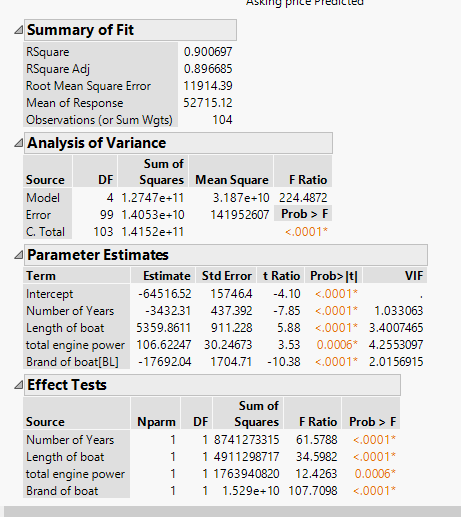


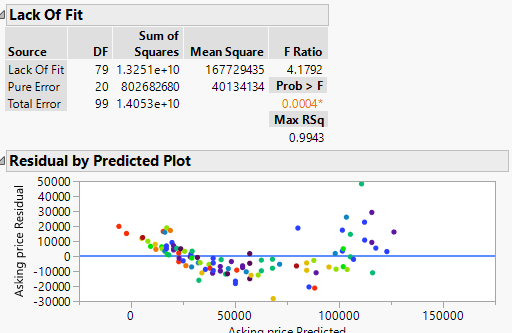












# References

Farway, j. j. (2004). *Extending the linear modrl with R.* BOLTON: CRC PRESS.

Wainwright, M. j. (2015). *High-Dimension Statistic.* london: Cambbridge Scerie.

Warne, R. T. (2014). *Statistic for socisl science.* Melbourne: Cambridge university.