Data Science HW4

Matt McCullough

Diagram

Description automatically generated

A picture containing letter

Description automatically generated

A picture containing letter

Description automatically generated

5a.

The explanatory variable is the market value, and the response variable is the sale price

this scatterplot does show an approximately linear relationship between the two variables

Chart, scatter chart

Description automatically generated

b.

Text

Description automatically generated

Text, letter

Description automatically generated

c.

The least squares regression line is y = 1.408271x + 1.358681

Chart

Description automatically generated

d.

It would be appropriate the slope of this linear model, as for every increase of market value (which is 1.408271 thousand dollars) to the market value of a property, there is an increase of 1.35881 thousand dollars of to the predicted sale price on average.

e.

null: B1 = 0

alternate B1 > 0

There is sufficient evidence that there is a positive linear relationship between appraised and property value and sale price for residential properties sold. β1 the slope of the straight-line model, is positive as t was 38.132 and the P value was very small as it's < .0001

The formula I used was P(T > 38.132) = <2e-16 (given by summary) < .05, so the null hypothesis can be rejected and the slope is positive

f.

The 95% confidence interval for the slope is 1.334683 - 1.481858

g.

This is a good residual plot as most of the plots are relatively near the line at 0. The more of the data near the line, the higher r^2 will be.

Chart, scatter chart

Description automatically generated

h.

Yes, this distribution of the residuals is close to a normal distribution with a mean of zero as its center is near zero and looks like a mound shape distribution

Chart, histogram

Description automatically generated

i.

Residual standard error: 68.76 on 74 degrees of freedom (s is 68.76)

Approximately 95% of the data points lie within 2s or 137.5132 thousand dollars

j.

95.16% of the variability can be explained by the regression. Yes, I would consider this regression to be a success as r^2 is .9516 which is very close to 1 and because of that can make mostly accurate predictions

k.

The 95% prediction interval of the regression line for the estimate of 300000 dollars is 285.9404 - 561.7394