STA 674

Exam 1 key

Spring 2022

1. A. 4 points. Explanatory variable is mileage. Response variable is price. 2 points each. No partial credit.
2. 10 points. Predicted(price) = 1.95e+04 – 7.11e-02\*miles

1.95e+04 is the y-intercept, which is the value of the predicted y at miles=0 ($19,500 for a new Prius).

-7.11e-02 is the change in price for each additional mile added to the car. ($0.07 decrease for each mile).

Partial credit given.

1. 16 points. Four assumptions:
2. E(ei)=0 or linearity: scatterplot shows mostly linear with the exception of possible outliers. The residual vs fitted is a bit suspicious.
3. ei have constant variance: with the exception of the two outliers, residual plot vs. fitted values does not display a pattern.
4. ei are normally distributed: QQplot is close to a straight line.
5. ei are independent: No reason to think data are not independent.

Partial credit given for all parts.

1. A. 10 points. Both an outlier (because of y value) and leverage or influential point as they have potential to affect regression line.

Partial credit given.

B. 20 points. Compare regression equations 1. See above. 2. Pred(y) = 2.12e+04 -9.99e-02

Compare F-value: 1. 37.35 with very small p-value 2. 102.9

Compare Root MSE: 1. 3572 2. 2501

Compare R-squared: 1. 0.55 2. .79

Compare t-value for exp. Variable: 1. -6.11 with very small p-value 2. -10.14 with very small p-value

In all of the above, the model is performing better without the outlier/leverage points.

Partial credit given.

1. A. 4 points. R-squared is .80. The model does appear to be useful due to the high Rsq.

Partial credit given.

B. 6 points. The graphs show issues remain with the outliers present that were present with one variable.

Partial credit given.

1. 4 points Null is B1=B2=0, Alt is at least one is not = 0. p-value: 5.835e-11, reject Ho.
2. 6 points. Both variables are significant, so when holding miles constant, and increase of 1 year of age results in a decrease of $843.80 dollars. When holding age constant, an increase in one mile results in a decrease in price of about $.03. Both on average.

Partial credit given.

E. 6 points. Two var model, Adj Rsq increase significantly and p-value for model decreased. Be sure to state those values for full credit. Partial credit given.

F. 4 points. To answer this question, should refer to the fact that both t-values for the two predictors produce very small p-values, demonstrating that both variables are important in the model.

Partial credit given.

1. Suppose that the manager of a local Kroger wishes to collect data on the following variables from a random sample of transactions/receipts at their store:

* Total amount spent
* Number of items purchased
* Day of week
* Time of day (morning, afternoon, evening)
* Payment type (credit card, cash, other)

1. 4 points. State a research question that could be addressed by applying linear regression analysis to (some of) the data.  Be sure to specify the response and explanatory variables. Use only the variables presented above.

SOLUTION: Answers will vary, but there should be a research question such as: How does total amount spent at Kroger depend on number of items, day of the week, time of day and payment type?

Response variable could be total spent, with the rest being explanatory. Partial credit given.

1. 3 points. State an additional variable for which data could be collected and classify it as categorical or quantitative, being sure not to change the observational units and response variable from part a.

SOLUTION: Again, answers will vary, but an additional variable could be organic product (y or n). Categorical. MUST be a new variable not listed above for full credit. Partial credit given

1. 3 points. State the linear model for your final decision of predictors from parts a and b.

Again, answers will vary, but should roughly be in the form

Partial Credit given