Homework #1

Dustin Leatherman

1/19/2020

1

Identify the variable level of measurement in each situation

a

Students' scores on a data science course test

Interval if **score** is *continuous*. Ordinal if **score** is a *discrete* set of values. (i.e. integers)

b

ZIP code is an example of which scale of measurement

Nominal

c

Pain scale in a doctor's office

Ordinal

d

Blood type

Nominal

e

Temperature in Fahrenheit

Interval

f

A film critic lists the top 50 greatest movies of all time

Ordinal

g

Shades of lipstick available in Wholefood store

Ordinal

2

Indicate if positive correlation, negative correlation, or not correlation for each of the following pairs of variables exist

i

Calories eaten per day and IQ.

No inherent correlation unless there are studies that indicate otherwise.

ii

Calories eaten per day and weight

Generally a positive correlation exists between calories per day and weight. Outliers to this rule would be individuals with extremely high metabolisms.

3

The prediction equation resulting from fitting Y = Base price in dollars to X = horsepower for a sample of jet boats (horsepower ranging from 90 to 160) gives $Y = 5937 + (30.73)X$.

a

Explain what the slope of 30.73 represents in terms of these data.

For every 1 additional horsepower in a jet boat, the mean base price increases by $30.73.

b

Predict the base price of a jet boat with 160 horsepower.

5937 + 30.73 * 160 = $10853.80

c

Would you have any concern about using the prediction equation to predict the base price for a jet boat with 50 horsepower? Why or why not?

Predicting the base price of a jet boat outside the range of 90-160 is considered an extrapolation. This is dangerous since the model in question did not account for any trends that may occur outside that range. It is not safe to assume that a linear trend exists outside the scope of the linear model.

e

$\beta_0 = 5937$. Does this estimate by itself have a meaningful interpretation in this study? Why or why not?

β_0 represents the base price of a jetboat with 0 horsepower. In this study, this is not a meaningful interpretation since a jetboat with 0 horsepower very likely does not exist.

3

A multiple regression equation was fit for n = 15 observations using 5 independent variables X1, X2,. . . , X5 gave SST = 1477.83 and SSE = 354.68.

a

Calculate the value of the coefficient of determination and interpret your result.

$R^2 = 1 - \frac{354.68}{1477.83} = 0.76$

The Multiple Regression model in question explains 76% of the overall variance found in the 15 observations. This model has decently fits the data though without knowing the source or scope of the model, no more can be said on whether or not this is acceptable.

b

Give the null and alternative hypotheses for testing the global utility of the model.

$$H_0 : \beta_i = 0, \ \forall \ i \in [1, 5]$$
$$H_A : \text{At Least one } \beta_i \neq 0, \ i \in [1, 5]$$

c

Conduct the test, in part b (Test the hypothesis that all the slopes are zero) $\alpha = 0.01$ and give the appropriate conclusion.

Source	SS	DF	MS	F	p-value
Regression	1123.15	4	280.7875	7.9167	0.00382
Error	354.68	10	35.468		
Total	1477.83	14			

There is convincing evidence that at least one predictor is non-zero (Sum of Squares F-Test. p-value = 0.00382). Thus at least one of the predictors is associated with the response of the model.

4

Fitting a multiple regression using 20 observations gave the following results

R-Square	Root MSE	Y Mean
0.993733	0.434604	6.335000

Source	DF	Type I SS	Type III SS
x1	1	270.5529453	2.16717818
x2	1	143.6314628	2.21046747
x3	1	31.2677868	26.08887272
x4	1	3.8000902	3.80009024

Parameter	Estimate	Standard Error
Intercept	0.475	0.45845003
x1	-.578	0.17068791
x2	0.009	0.00265354
x3	0.007	0.00061762
x4	0.087	0.01933796

a

Is there significant evidence that $\beta_1 < 0$ at $\alpha = 0.05$?

$\frac{-0.578}{0.17068791} = -3.386297$

```
pt(-3.386297, 19)
```

```
## [1] 0.001549511
```

There is convincing evidence that β_1 is less than zero (one-tailed t-test. p-value = 0.0015).

b

Give a 90% confidence interval for β_3. Interpret your result.

$\mu \pm t_{0.95,n-1}SE$

$0.007 \pm 1.729133 \times 0.00061762 = [0.005932053, 0.008067947]$

With 90% confidence, the estimated mean value of β_3 is between 0.0059 and 0.0081.