Problems Chapter 6, set 2.

1. Consider the regression model for i=1,2,3. With , , draw the data in three dimensional space and identify the orthogonal projection of y onto L(X) the space spanned by the vector **1** and **x**. Explain geometrically and
2. After fitting the regression model on 15 cases, it is found that the mean square error and .

Find

* 1. The estimate of .
  2. The estimate of = 3 \* (.4)=1.2.
  3. The estimate of .
  4. The estimate of .

1. Consider a multiple regression model of the price of houses (y) on three explanatory variables: taxes paid , number of bathrooms , and square feet . The incomplete output for a regression on n=28 houses is as follows:

Call:

lm(formula = price ~ taxes+nbath+sqft)

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | T value | Pr( >|t|) |
| (Intercept) | -10.65 | 24.02 |  | 0.6617331 |
| taxes | 0.18966 | 0.05623 | 3.37 | 0.002537815 |
| nbath | 81.87 | 47.82 |  | 0.1001654 |
| sqft | 0.10063 | 0.03125 |  | 0.003659009 |

p-value: 2P(T> |my value|) 2\*(1-pt( ,df)) df=n-p-1=24

5.23 e-4 =

Analysis of Variance Table

Response: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| Regression | 3=p | 504541 | 504541/3  =168180.3 | 168180.3/1524.083  =110.3485 | 0 |
| Errors | 24=n-p-1 | 36578 | 36578/24  =1524.083 |  |  |
| Total | 27 = n-1 | 541119 |  |  |  |

**p-value:** P(F>110.35) = 1-pf(110.35,3,24)

* 1. Calculate *0.932403*

and adjusted-*0.9239534*

.

* 1. Calculate s=*39.03951*
  2. Test the practical utility of the regression model (F-test) (write down the hypothesis, the statistic of the test, the p-value and the conclusion of the test).

H0: (reduced model)

Ha: (full model)

Test stat: F=110.35 with a p-value of 0

* 1. Test the statistical utility of the individual terms in the model (t- tests for the coefficients). (Write down the hypothesis, the statistic of the test, the p-value and the conclusion of the test)

H0: vs Ha: test statistic: t= With p-value 0.6617331

H0: vs Ha: test statistic: t= 3.37 with a p-value 0.002537815 hence “taxes” is significant in the model

H0: vs Ha: test statistic: t=

H0: vs Ha: test statistic: t=

* 1. Find 95% confidence intervals for the regression coefficients for taxes, nbath and sqft.

qt(.975,24)

0.18966 2.063899 0.05623

Same procedure for the others

1. The incomplete output for a regression of the price of houses (y) on two explanatory variables: taxes paid and square feet is

Call:

lm(formula = price ~ taxes+sqft)

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | T value | Pr( >|t|) |
| (Intercept) | 4.89 | 23.08 |  |  |
| taxes | 0.2437 | 0.04884 |  |  |
| sqft | 0.13397 | 0.02537 |  |  |

Analysis of Variance Table

Response: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| Regression | 2 | 500074 | 250037 |  |  |
| Errors | 25 | 41045 | 1641.8 |  |  |
| Total | 27 | 541119 |  |  |  |

Calculate 0.9241479

and adjusted-0.9180798

* 1. .

Calculate s= 40.51913

.

* 1. Test the practical utility of the regression model (F-test) (write down the hypothesis, the statistic of the test, the p-value and the conclusion of the test).
  2. Comment on the fact that the regression coefficients for taxes and sqft are different in the two models.

1. Compare the two models in problems 3 and 4.
2. The following matrices were computed for a certain regression problem:
3. Write down the estimated regression equation. Obtain the standard error of the regression coefficients.
4. Compute the t statistics to the test the simple hypothesis that each regression coefficient is equal to 0. Carry out the test and state your conclusion.