**MIDTERM REVIEW**

**PROOFS** Please prove the following:

1. a) Use the normal equations to prove(remember you should start by formulating the SSR and them minimizing it )



b)



c) Prove that B is an unbiased estimators of β

d) Prove RegSS=B2sxx

**SIMPLE LINEAR REGRESSION**

1. Conduct a simple linear regression analysis for the following data:

We know that the population Linear model is .This estimated using the equation fitted model that we obtain using the sample data.

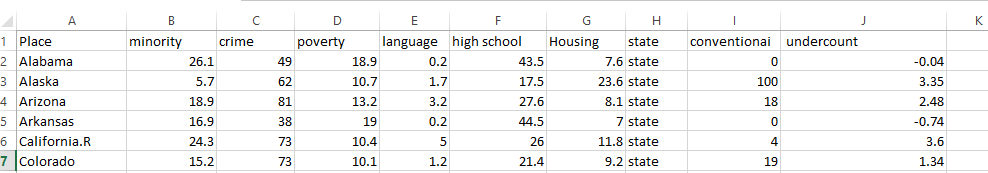
The following data gives information regarding

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Number of courses taken in CS | 1 | 2 | 3 | 4 | 5 |
| Number of courses in Statistics | 1 | 1 | 2 | 2 | 4 |

1. Calculate the Fitted Model coefficients A and B
2. Calculate RSS.
3. Interpret A and B in this context. Does A make sense in this case study
4. Is the linear regression between number of courses taken in CS and number of courses taken in Statistics relationship significant? Conduct the appropriate Hypothesis test. Use the significance level α=.05.
5. Create the labeled ANOVA table for this example
6. Conduct the F test to decide if the aforementioned linear relationship is significant. Use the significance level α =.05.
7. Did you get the same result in part d and part f
8. Calculate t2 is it approximately equal to F?
9. Calculate R2.Interpret this in context of this study.
10. Find the 95% confidence interval for B in this case study.

**MULTIPLE LINEAR REGRESSION**

A multiple regression study was conducted using the following data. The analyst wishes to analyze the effect of minority and poverty on crime.



> EricksenFinalmod<-lm(crime~poverty+minority,data=Ericksen)

> EricksenFinalmod

Call:

lm(formula = crime ~ poverty + minority, data = Ericksen)

Coefficients:

(Intercept) poverty minority

58.670 -1.398 1.195

> summary(EricksenFinalmod)

Call:

lm(formula = crime ~ poverty + minority, data = Ericksen)

Residuals:

Min 1Q Median 3Q Max

-34.026 -13.138 -1.683 8.688 70.643

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 58.6701 8.2382 7.122 1.23e-09 \*\*\*

poverty -1.3984 0.7642 -1.830 0.072 .

minority 1.1949 0.1955 6.111 6.86e-08 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.62 on 63 degrees of freedom

Multiple R-squared: 0.4577, Adjusted R-squared: 0.4405

F-statistic: 26.59 on 2 and 63 DF, p-value: 4.249e-09

1. Does poverty and minority significantly effect crime? Justify
2. Does poverty by itself significantly effect crime? Justify
3. Does minority by itself significantly effect crime? Justify
4. Without seeing the scatterdot how can you say that the scatter dots (data values) are mostly aligned with the fitted line or not? Provide two measures and explain.
5. How much of the variation in the crimes can be attributed to poverty and minority? Justify
6. Would you use R squared or Adjusted R squared. Give your rationale for this decision.
7. Seeing the outputs please write the equation of the fitted model.
8. Predict the crime for a minority score of 25 and poverty of 20.
9. Do you think there might be pairwise correlation if so justify your answer.