**STAT 501 – Homework 5 – Fall 2015 – Due Oct 4**

**Instructions**: Use Word to type your answers within this document. Then, submit your answers in the appropriate dropbox in ANGEL by the due date. The point distribution is located next to each question.

1. (**8x5=40 points**) Suppose that X1, X2, and X3 are potential predictor variables in a model used to predict Y. Use the given SSE values in the following table to answer the questions below.

|  |  |  |
| --- | --- | --- |
| SSE(X1) = 510 | SSE(X1, X2) = 450 | SSE(X1, X2, X3) = 330 |
| SSE(X2) = 905 | SSE(X1, X3) = 400 |  |
| SSE(X3) = 720 | SSE(X2, X3) = 640 |  |

The notation SSE(X1, X2) means the sum of squares for a multiple regression model that includes X1 and X2 as predictors (and does not include X3).

1. Calculate SSR(X3|X1). Show your work.

SSR(X3|X1) = SSE(X1) - SSE(X1, X3) = 510 – 400 = 110

1. Explain in words what is measured by the quantity calculated in the previous part.

SSR(x3 | x1) is the reduction in the error sum of squares — or the increase in the regression sum of squares — when you add x3 to a model already containing x1

It is the marginal effect of adding X3 to the regression model when X1 is already in the model.

1. Calculate SSR(X1|X2, X3). Show your work.

SSR(X1|X2, X3) = SSE(X2, X3) - SSE(X1, X2, X3) = 640 – 330 = 310

1. Explain in words what is measured by the quantity calculated in the previous part.

SSR(X1|X2, X3) is the reduction in the error sum of squares — or the increase in the regression sum of squares — when you add x1 to a model already containing x2 and x3.

It is the marginal effect of adding X1 to the regression model when X2 and X3 are already in the model.

1. Consider the “full” model, Yi = β0 + β1 Xi,1 + β2 Xi,2 + β3 Xi,3 + εi. What is the “reduced” model associated with the null hypothesis H0: β2 = β3 = 0?

The reduced model is: Yi = β0 + β1 Xi,1 + εi

1. Suppose that n = 70. Calculate the value of an F-statistic for testing H0: β2= β3 = 0 for the model Yi = β0 + β1 Xi,1 + β2 Xi,2 + β3 Xi,3 + εi. It is not necessary to carry out the test – just calculate the value of F. Show your work. (Hint: Use the general linear F-test.)

The general linear F-Test is: 

F\* = [{SSE(X1) – SSE(X1, X2, X3)} / {(n-2) – (n-4)}] / [SSE(X1, X2, X3) / (n-4)]

= [(510-330)/2] / [330/66] = 90 / 5 = 18

1. Calculate the value of the coefficient of partial determination R2Y,2|1.

R2Y,2|1 = SSR(X2 | X1) / SSE(X1) = (SSE(X1) – SSE(X1, X2)) / SSE(X1) = (510-450) / 510 = 0.1176

1. Write a sentence that interprets the value calculated in the previous part.

It is the proportion of variation in Y that is explained by the predictor X2 that cannot be explained by the predictor X1.

Stated differently, it measures the proportionate reduction in the variation in Y remaining after X1 is included in the model that is gained by also including X2 in the model.

1. **(10+6x5=40 points)**
   1. Fill in the blanks in the following tables. The column labeled “Seq SS” represents “sequential sums of squares” (measures the reduction in the SS when a term is added to a model that contains only the terms before it), while the column labeled “Adj SS” represents “adjusted sums of squares” (measures the reduction in the SS for each term relative to a model that contains all of the remaining terms). *[Hint: The t-statistics in the Coefficients table assume all other predictors are included in the model, so if we square these we get the F-statistics in the Anova table based on Adjusted Sums of Squares.]*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source | df | Seq SS | Adj SS | Adj MS | F-statistic  based on Adj SS | p-value  based on Adj SS |
| Regression | 3 | 100.866 | 100.866(same as left) | 33.622 (100.866/3) | 35.14 | 0.000 |
| X1 | 1 | 67.444 | 33.031 |  | 34.52 | 0.000 |
| X2 | 1 | 3.883 | 0.1594 | 0.1594 | 0.1666 (-0.4082^2) | 0.6841 (1-0.315910) |
| X3 | 1 | 29.539 | 29.539 (same as left) | 29.539 | 30.88 | 0.000 |
| Error | 93 | 88.976 (189.842-100.866) | 88.976 (same as left) | 0.9567  SS/df | ---- | ------- |
| Total | 96 | 189.842 | 189.842 |  | ---- | ------- |

Coefficients

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | Coef | SE coef | t-statistic | p-value |
| Constant | 0.58 | 1.24 | 0.45 | 0.652 |
| X1 | 0.34 | 0.058 | 5.88 | 0.000 |
| X2 | –0.01 | 0.0245 | -0.4082 (-0.01/0.0245) | 0.684 (2\* t-stat for df=94 & -0.4082) |
| X3 | 0.06 | 0.0103 | 5.56 | 0.000 |

* 1. Calculate SSR(X3|X1), that is the sequential sum of squares obtained by adding X3 to a model already containing only the predictor X1. Show your work.

We know SSR(x1, x2, x3) = SSR(x1) + SSR(x3|x1) + SSR(x2 | x1, x3)

So, SSR(x3|x1) = SSR(x1, x2, x3) – (SSR(x1) + SSR(x2 | x1, x3))

= 100.866 – (67.444 + 0.1594) = 33.2626

* 1. Explain in words what is measured by the quantity calculated in the previous part.

SSR(x3 | x1) is the reduction in the error sum of squares — or the increase in the regression sum of squares — when you add x3 to a model already containing x1

It is the marginal effect of adding X3 to the regression model when X1 is already in the model.

* 1. Discuss the conceptual difference between the sequential sum of squares (Seq SS) and adjusted sum of squares (Adj SS) in terms of the predictor X2. For this data, what are the numerical values of these sums of squares for the predictor X2?

Sequential sums of squares depend on the order the factors are entered into the model. It is the unique portion of SS Regression explained by a factor, given any previously entered factors.

Adjusted sums of squares does not depend on the order the factors are entered into the model. It is the unique portion of SS Regression explained by a factor, given all other factors in the model, regardless of the order they were entered into the model. For example, if you have a model with three factors, X1, X2, and X3, the adjusted sum of squares for X2 shows how much of the remaining variation X2 explains, given that X1 and X3 are also in the model.

**For this data, the sequential sum of squares for X2 = SSR(X2|X1) = 3.883 = increase in the regression sum of squares when X2 is added to the model which has only X1 as a predictor. Sequential sum of squares depends on the order the predictors are entered into the model. It is the portion of the regression sum of squares explained by a predictor, after accounting for the previously entered predictors. The adjusted sum of squares for X2 = SSR(X2|X1, X3) = 0.163 = increase in the regression sum of squares when X2 is added to the model which already contains all other predictors, X1 and X3 in this case. Adjusted sum of squares does not depend on the order the predictors are added to the model.**

* 1. Calculate the value of an F-statistic for testing H0: β2= β3 = 0 within the model Yi = β0 + β1 Xi,1 + β2 Xi,2 + β3 Xi,3 + εi. It is not necessary to carry out the test – just calculate the value of F. Show your work.

The general linear F-Test is: 

F\* = [{SSE(X1) – SSE(X1, X2, X3)} / {(n-2) – (n-4)}] / [SSE(X1, X2, X3) / (n-4)]

= [SSR(X2, X3 | X1) / 2] / [SSE(X1, X2, X3) / (n-4)]

= [SSR(X2, X3 | X1) / 2] / MSE(full) = [SSR(X2, X3 | X1) / 2] / [88.976 / 93] = [SSR(X2, X3 | X1) / 2] / 0.9567

Now SSR(X2, X3 | X1) = SSR(X2 I X1) + SSR(X3 I X1, X2) = 3.883 + 29.539 = 33.422

Therefore F\* = (33.422 / 2) / 0.9567 = 17.4673



* 1. Calculate the value of the coefficient of partial determination R2Y,2|1.

R2Y,2|1 = SSR(X2 | X1) / SSE(X1) = SSR(X2 | X1) / (SSTO – SSR(X1)) = 3.883 / (189.842 - 67.444) = 0.03172



* 1. Write a sentence that interprets the value calculated in the previous part.

It is the proportion of variation in Y that is explained by the predictor X2 that cannot be explained by the predictor X1.

Stated differently, it measures the proportionate reduction in the variation in Y remaining after X1 is included in the model that is gained by also including X2 in the model.

**X2 explains 3.2% of the variation in Y that cannot be explained by X1.**

1. **(4x5=20 points)**  Consider the “GroceryRetailer” dataset. A large, national grocery retailer tracks productivity and costs of its facilities closely. Data were obtained from a single distribution center for a one-year period. Each data point for each variable represents one week of activity. The variables included are the number of cases shipped (X1), the indirect costs of the total labor hours as a percentage (X2), a qualitative predictor called holiday that is coded 1 if the week has a holiday and 0 otherwise (X3), and the total labor hours (Y).

1. Obtain the ANOVA table that decomposes the regression sum of squares into extra sums of squares associated with X1; X3 given X1; and with X2 given X1 and X3. Give their values along with the associated dfs.

The output from Minitab:

Analysis of Variance

Source DF Seq SS Seq MS F-Value P-Value

Regression 3 2176606 725535 35.34 0.000

X1 1 136366 136366 6.64 0.013

X3 1 2033565 2033565 99.04 0.000

X2 1 6675 6675 0.33 0.571

Error 48 985530 20532

Total 51 3162136

|  |  |  |
| --- | --- | --- |
|  | extra sums of squares | df |
| X1 | 136366 | 1 |
| X3 given X1 | 2033565 | 1 |
| X2 given X1 and X3 | 6675 | 1 |

1. Test whether X2 can be dropped from the regression model given that X1, and X3 are retained. Use the F-test statistic and α = 0.05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

We test the hypotheses:

*H*0 : *β*2 = 0

*H*A : *β*2 ≠ 0

The full model is the model containing all of the possible predictors: yi=(β0+β1xi1+β2xi2+β3xi3)+ϵi

The reduced model is: yi=(β0+β1xi1+β3xi3)+ϵi

The general linear statistic: 

= ((SSE(X1, X3) – SSE(X1, X2, X3)) / 1) / MSE(X1, X2, X3)

= SSR(X2 | X1, X3) / MSE(X1, X2, X3) = 6675 / 20532 = 0.325

The P-value is the probability — if the null hypothesis were true — that we would get an F-statistic larger than 0.325. Comparing our partial F-statistic to an F-distribution with 1 numerator degree of freedom and 48 denominator degrees of freedom, Minitab tells us that the probability is 0.428723 that we would observe an F-statistic smaller than 0.325:

F distribution with 1 DF in numerator and 48 DF in denominator

x P( X ≤ x )

0.325 0.428723

Therefore, the probability that we would get an F-statistic larger than 0.325 is 0.571. That is, the P-value is > 0.05. There is not sufficient evidence (F = 0.325, P > 0.05) to conclude that X2 is significantly related to the Y.

X2 can be dropped from the regression model given that X1, and X3 are retained.



1. Now test H0: β2 = 0 vs. Ha: β2 ≠ 0 in the model E(Y) = β0 + β1 X1 + β3 X3 + β2 X2 using a t-test. What is the value of the test statistic? What is the P-value and conclusion?

The minitab output is:

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant 4150 196 21.22 0.000

X1 0.000787 0.000365 2.16 0.036 1.01

X3 623.6 62.6 9.95 0.000 1.01

X2 -13.2 23.1 -0.57 0.571 1.02

The T-statistic is computed using T\* = Coeff / SE Coeff = -13.2 / 23.1 = -0.571

The p-value(computed using 2\* t-stat for df=50 & -0.571) = 0.571. That is, the P-value is > 0.05.

There is not sufficient evidence to conclude that X2 is significantly related to the Y. Thus it is feasible to drop the variables X2 from the model.

1. From part (a), use sequential sum of squares to test H0: β2 = β3 = 0 in the model E(Y) = β0 + β1 X1 + β3 X3 + β2 X2. Give the test statistic, P-value and conclusion.

Here we test the hypotheses:

H0 : β2 = β3 = 0

HA : At least one βi ≠ 0 (for i = 2, 3)

The full model is the model containing all of the possible predictors: yi=(β0+β1xi1+β2xi2+β3xi3)+ϵi

The reduced model is: yi=(β0+β1xi1)+ϵi

The general linear statistic: 

becomes a partial F-test: 

Now SSR(X2, X3 | X1) = SSR(X3 | X1) + SSR(X2 | X1, X3)

From the table above we get: SSR(X2, X3 | X1) = 2033565 + 6675 = 2040240

Therefore F\* = (2040240 / 2) / 20532 = 1020120 / 20532 = 49.684

The P-value is the probability — if the null hypothesis were true — that we would observe a partial F-statistic more extreme than 49.684. The following Minitab output:

F distribution with 2 DF in numerator and 48 DF in denominator

x P( X ≤ x )

49.684 1.00000

tells us that the probability of observing such an F-statistic that is smaller than 49.684 is close to 1.0. Therefore, the probability of observing such an F-statistic that is larger than 49.684 is close to 0.

The P-value is very small. There is sufficient evidence (F = 49.684, P < 0.001) to conclude that Y is significantly related to either X2 or X3 or both — after taking into account X1.

We reject the null hypothesis.

