Final Exam

Ishita Dutta, 918193342

6/9/2021

Question 1 (25 points)

1a)

False, you can have an estimated regression line with a nonzero slope for x but still have a correlation coefficient of 0.

1b)

False, multiple linear regression can be used on nonlinear data as well as linear data so long as there is an adequate fit

1c)

True, this is a definition. We only select variables that best fit the data and help us analyze it.

1d)

```
qt(1-0.1/2,25 - 3 - 1)
```

```
## [1] 1.720743
```

False, the value 3.64 is much more outlying than our given t value at 1.721, making this residual outlying.

1e)

True, as you decrease SSE, more of that error becomes a part of the explained SSR.

Question 2 (24 points)

2a)

Ho $->$ p $>$ alpha Ha $->$ p $<$ alpha Decision $->$ If Pval $<$ alpha, conclude Ha

```
SSTO = 7.834
SSR1 = 3.681 # with SF
SSR2 = 3.673 # without SF
n = 20
SSE1 = SSTO - SSR1
SSE2 = SSTO - SSR2
R21 = SSR1 / SSTO
R22 = SSR2 / SSTO
Fstat1 = ((SSE1 - SSE2)/((n - 2)-(n - 4)))*((SSE2/n - 4)^-1)
Fstat1
```

[1] 0.001054866

Concluding Ha, as our p-value is much less than our alpha. SF is a needed parameter in the data with the current parameters.

2b)

Ho –> p > alpha Ha –> p < alpha Decision –> If Pval < alpha, conclude Ha

```
SSE3 = SSTO - 3.443
Fstat2 = ((SSE2 - SSE3)/((n - 2)-(n - 3)))*((SSE3/n - 3)^-1)
Fstat2
```

[1] 0.08272042

We fail to reject the Ho at this point as the p-value is larger than our significance level

2c)

We would need more information, like the SSR when we drop these variables and then we can test it against the alpha = 0.05 significance level.

Question 3 (27 points)

3a)

3 qualitative predictors: ST - storm windows (1 if present; 0 if absent); L1 - location (1 if zone A; 0 oth-erwise); L2 - location (1 if zone B; 0 otherwise).

3b)

3 significant regression coefficients at alpha = 0.05, 4 if we count the significant qualitative predictor.

3c)

2 at alpha = 0.05 in the direction of the sign of Bj, 3 if we count the significant qualitative predictor.

3d)

```r
# b +- t * se
B1.lower = 0.017 + 5.241 * 0.003
B1.upper = 0.017 - 5.241 * 0.003
cat("B1 Interval: [", B1.lower,",",B1.upper,"]\n")
```

```
## B1 Interval: [ 0.032723 , 0.001277 ]
```

```r
B2.lower = 3.140 + 1.984 * 1.583
B2.upper = 3.140 - 1.984 * 1.583
cat("B2 Interval: [", B2.lower,",",B2.upper,"]\n")
```

```
## B2 Interval: [ 6.280672 , -0.000672 ]
```

```r
B3.lower = -6.702 + (-3.708 * 1.807)
B3.upper = -6.702 - (-3.708* 1.807)
cat("B3 Interval: [", B3.lower,",",B3.upper,"]\n")
```

```
## B3 Interval: [ -13.40236 , -0.001644 ]
```

```r
B4.lower = 2.466 + 1.002 * 2.462
B4.upper = 2.466 - 1.002 * 2.462
cat("B4 Interval: [", B4.lower,",",B4.upper,"]\n")
```

```
## B4 Interval: [ 4.932924 , -0.000924 ]
```

```r
B5.lower = 2.253 + 1.553 * 1.451
B5.upper = 2.253 - 1.553 * 1.451
cat("B5 Interval: [", B5.lower,",",B5.upper,"]\n")
```

```
## B5 Interval: [ 4.506403 , -0.000403 ]
```

```r
B6.lower = 0.288 + 2.258 * 0.127
B6.upper = 0.288 - 2.258 * 0.127
cat("B6 Interval: [", B6.lower,",",B6.upper,"]\n")
```

```
## B6 Interval: [ 0.574766 , 0.001234 ]
```

```r
B7.lower = 5.612 + 1.835 * 3.059
B7.upper = 5.612 - 1.835 * 3.059
cat("B7 Interval: [", B7.lower,",",B7.upper,"]\n")
```

```
## B7 Interval: [ 11.22527 , -0.001265 ]
```

```r
B8.lower = 10.017 + 4.320 * 2.318
B8.upper = 10.017 - 4.320 * 2.318
cat("B8 Interval: [", B8.lower,",",B8.upper,"]\n")
```

```
## B8 Interval: [ 20.03076 , 0.00324 ]
```

```
B9.lower = 2.692 + 0.939 * 2.867
B9.upper = 2.692 - 0.939 * 2.867
cat("B9 Interval: [", B9.lower,",",B9.upper,"]\n")
```

```
## B9 Interval: [ 5.384113 , -0.000113 ]
```

```
B10.lower = 5.692 + 2.117 * 2.689
B10.upper = 5.692 - 2.117 * 2.689
cat("B10 Interval: [", B10.lower,",",B10.upper,"]\n")
```

```
## B10 Interval: [ 11.38461 , -0.000613 ]
```

Question 4 (24 points)

4a)

```
SSRA = 352.27 # x1
SSRB = 381.97 # x2
SSRC = 33.17 # x1, x2
SSRD = 11.54 # x1, x2, x3
SSTOT = 495.39
n = 20
R2A = SSRA / SSTOT
R2B = SSRB / SSTOT
R2C = SSRC / SSTOT
R2D = SSRD / SSTOT
cat("Model I: ", R2A, "\nModel II: ", R2B, "\nModel III: ", R2C, "\nModel IV: ", R2D)
```

```
## Model I:   0.7110963
## Model II:   0.7710491
## Model III:   0.06695735
## Model IV:   0.02329478
```

The best model given by R2 is Model II. This is because it has the highest R2, therefore the most correlation.

4b)

```
AICA = n * log((SSTOT - SSRA)/20) + 2
AICB = n * log((SSTOT - SSRB)/20) + 2
AICC = n * log((SSTOT - SSRC)/20) + 4
AICD = n * log((SSTOT - SSRD)/20) + 6
cat("Model I: ", AICA, "\nModel II: ", AICB, "\nModel III: ", AICC, "\nModel IV: ", AICD)
```

```
## Model I:   41.35902
## Model II:   36.70731
## Model III:   66.80617
## Model IV:   69.72085
```

Again, model 2 fits best here as it has the lowest AIC value.

4c)

```r
BICA = n * log((SSTOT - SSRA)/20) + 2* log(20)
BICB = n * log((SSTOT - SSRB)/20) + 2* log(20)
BICC = n * log((SSTOT - SSRC)/20) + 4* log(20)
BICD = n * log((SSTOT - SSRD)/20) + 6* log(20)
cat("Model I: ", BICA, "\nModel II: ", BICB, "\nModel III: ", BICC, "\nModel IV: ", BICD)
```

```
## Model I:   45.35049
## Model II:   40.69877
## Model III:   74.7891
## Model IV:   81.69525
```

Here, model 2 still shows that it is the best fit for the data, with the lowest BIC value.

4d)

No, we have not taken all three paired correlations into account, we only took x1 and x2. The other two pairs we can consider are x2 and x3, as well as x3 and x1.