hw4 Q5

Lisong he

2024-03-27

a)

```r
data(stackloss, package = 'datasets')
full_model <- lm(stack.loss ~ Air.Flow + Water.Temp + Acid.Conc., data=stackloss)
summary(full_model)
```

```
##
## Call:
## lm(formula = stack.loss ~ Air.Flow + Water.Temp + Acid.Conc.,
##     data = stackloss)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2377 -1.7117 -0.4551  2.3614  5.6978
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -39.9197    11.8960  -3.356  0.00375 **
## Air.Flow      0.7156     0.1349   5.307  5.8e-05 ***
## Water.Temp    1.2953     0.3680   3.520  0.00263 **
## Acid.Conc.   -0.1521     0.1563  -0.973  0.34405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.243 on 17 degrees of freedom
## Multiple R-squared:  0.9136, Adjusted R-squared:  0.8983
## F-statistic:  59.9 on 3 and 17 DF,  p-value: 3.016e-09
```

The estimated function is Li = -39.9197 + 0.7156*Ai + 1.2953*Wi - 0.1521Ci. From the regression model, the predictor Acid.Conc has t*statistic above 0.05, which proves to be insignificant. So air flow and water tempreature are important.

b)

```r
reduced_model <- lm(stack.loss ~ Air.Flow, data=stackloss)
summary(reduced_model)
```

```
##
## Call:
## lm(formula = stack.loss ~ Air.Flow, data = stackloss)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.2896  -1.1272  -0.0459   1.1166   8.8728
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -44.13202    6.10586  -7.228 7.31e-07 ***
## Air.Flow      1.02031    0.09995  10.208 3.77e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.098 on 19 degrees of freedom
## Multiple R-squared:  0.8458, Adjusted R-squared:  0.8377
## F-statistic: 104.2 on 1 and 19 DF,  p-value: 3.774e-09
```

The reduced model has the following estimated fit function: Li = -44.13202 + 1.02031*Ai

c)

```r
SSEf <- sum(residuals(full_model)^2)
SSEr <- sum(residuals(reduced_model)^2)
print(SSEf)
```

```
## [1] 178.83
```

```r
print(SSEr)
```

```
## [1] 319.1161
```

DF of SSE(f) is n - p which is 21 - 4 = 17. DF of SSE(r) is n - p which is 21 - 2 = 19. The full model has smaller SSE, indicating a better fit. The degrees of freedom reflect the number of independent predictors used to estimate variance, with the full model having fewer degrees of freedom due to more parameters being estimated.

d)

```r
MSEf <- SSEf / 17
MSEr <- SSEr / 19
F <- ((SSEr - SSEf) / (19 - 17)) / MSEf
print(F)
```

```
## [1] 6.667967
```

```r
print(pf(F, 2, 17, lower.tail = FALSE) < 0.05)
```

```
## [1] TRUE
```

The p-value of the F statistic is way less than 0.05, hence rejecting the null hypothesis in favor of the alternative hypothesis.

e)

```r
pvalue <- pf(F, 2, 17, lower.tail = FALSE)
print(pvalue)
```

```
## [1] 0.007280786
```