Homework 3

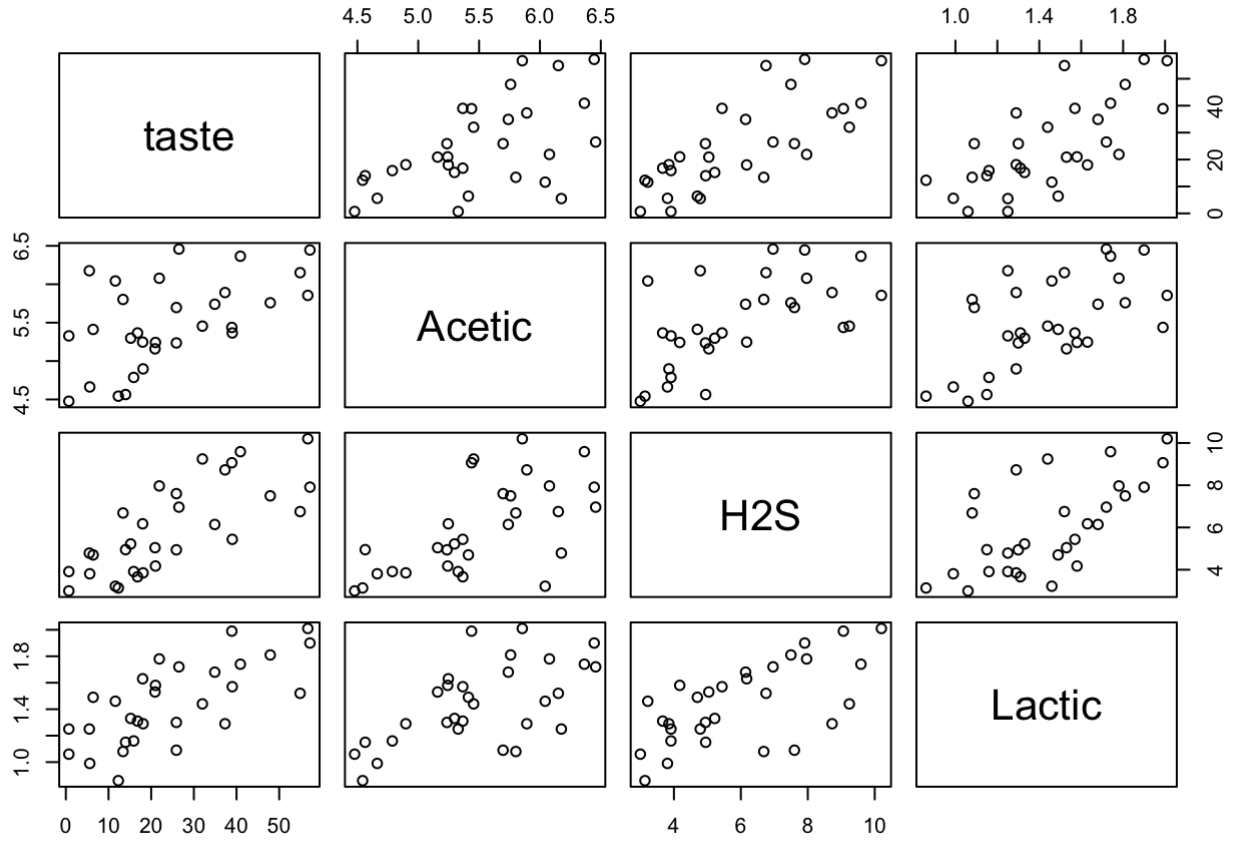
Felicity Escarzaga and Tyler Thatcher

09/25/2019

Problem 1

(a)

```
data(cheddar, package="faraway")
pairs(cheddar)
```



(b)

```
taste_model <- lm(cheddar$taste ~ cheddar$Acetic + cheddar$H2S + cheddar$Lactic)
sum_taste <- summary(taste_model)
sum_taste
```

```
##
## Call:
## lm(formula = cheddar$taste ~ cheddar$Acetic + cheddar$H2S + cheddar$Lactic)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.390  -6.612  -1.009   4.908  25.449
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -28.8768    19.7354  -1.463  0.15540
## cheddar$Acetic   0.3277     4.4598   0.073  0.94198
## cheddar$H2S      3.9118     1.2484   3.133  0.00425 **
## cheddar$Lactic  19.6705     8.6291   2.280  0.03108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.13 on 26 degrees of freedom
## Multiple R-squared:  0.6518, Adjusted R-squared:  0.6116
## F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

(c)

```
cor_taste_model <- cor(fitted(taste_model), cheddar$taste)
cor_taste_model
```

```
## [1] 0.8073256
```

(d)

```
taste_model_2 <- lm(cheddar$taste ~ cheddar$Acetic + cheddar$H2S + cheddar$Lactic - 1)
sum_taste <- summary(taste_model_2)
sum_taste
```

```
##
## Call:
## lm(formula = cheddar$taste ~ cheddar$Acetic + cheddar$H2S + cheddar$Lactic -
##     1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.4521  -6.5262  -0.6388   4.6811  28.4744
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## cheddar$Acetic   -5.454      2.111  -2.583  0.01553 *
## cheddar$H2S       4.576      1.187   3.854  0.00065 ***
## cheddar$Lactic   19.127      8.801   2.173  0.03871 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.34 on 27 degrees of freedom
## Multiple R-squared:  0.8877, Adjusted R-squared:  0.8752
## F-statistic: 71.15 on 3 and 27 DF,  p-value: 6.099e-13
```

The R Squared from the summary is: 0.8877
To calculate R squared, you would square the r (coefficent of determination) of the of the model.

```
r_squared <- cor(fitted(taste_model), cheddar$taste)^2
r_squared
```

```
## [1] 0.6517747
```

Problem 2

(a)

```
x <- model.matrix( ~ Acetic + H2S + Lactic, cheddar)
y <- cheddar$taste
```

(b)

```
xtxi <- solve(t(x) %*% x)
as.vector(betahat<- xtxi %*% t(x) %*% y)
```

```
## [1] -28.8767696   0.3277413   3.9118411  19.6705434
```

(c)

```
H_mat <- x %*% (solve(t(x) %*% x)) %*% t(x)
```

The number of parameters for the regression model is 4 parameters.

```
sum_diag <- sum(diag(H_mat))
sum_diag
```

```
## [1] 4
```

(d)

```
y_hat <- H_mat %*% y
y1 <- y_hat[1]
y2 <- y_hat[2]
yn <- tail(y_hat, n=1)

# Results for y1, y2, and yn from the homework
results <- c(y1, y2, yn)
results
```

```
## [1]  1.792448 22.637394 16.461523
```

(e)

```
residuals <- y - y_hat
# Display only the even indicies of the residuals
even <- function(x) x%%2 == 0
even_ind <- even(1:length(residuals))
residuals[even_ind]
```

```
##  [1]  -1.7373936   9.9624634   1.7348312 -17.3905351   0.7512666
##  [6]  15.6552777   8.1566913  -4.0427813 -14.1917289  -8.6067963
## [11]  -4.2412368   4.2231851  -2.8149807 -12.0607374 -10.9615230
```

(f)

```
n<- length(na.omit(y))
p<- dim(x)[2]
sigmahat<- sum(residuals^2)/(n-p)
sigmahat
```

```
## [1] 102.6312
```

(g)

```
verified_r_2 <- cor(y_hat, cheddar$taste)^2
verified_r_2
```

```
##           [,1]
## [1,] 0.6517747
```

```
r_squared
```

```
## [1] 0.6517747
```

(h)

```
varbetahat<- sigmahat * xtxi
varbetahat
```

```
##             (Intercept)     Acetic       H2S     Lactic
## (Intercept)  389.486737 -77.979215  8.960302  -7.332453
## Acetic       -77.979215  19.889429 -2.089205 -13.148377
## H2S            8.960302  -2.089205  1.558578  -4.670244
## Lactic        -7.332453 -13.148377 -4.670244  74.460587
```

```
varbetahat<- vcov(taste_model)
varbetahat
```

```
##                (Intercept) cheddar$Acetic cheddar$H2S cheddar$Lactic
## (Intercept)     389.486737     -77.979215    8.960302      -7.332453
## cheddar$Acetic  -77.979215      19.889429   -2.089205     -13.148377
## cheddar$H2S       8.960302      -2.089205    1.558578      -4.670244
## cheddar$Lactic   -7.332453     -13.148377   -4.670244      74.460587
```

(i)

```
estimated_se <- sqrt(diag(varbetahat))
estimated_se
```

```
##    (Intercept) cheddar$Acetic    cheddar$H2S cheddar$Lactic
##      19.735418       4.459757       1.248430       8.629055
```

(j)

These errors are that R estimated from the regression model that includes the intercepts.

(k)

(i)

```
tstat <- (betahat - 0) / estimated_se[2]
tstat
```

```
##                   [,1]
## (Intercept) -6.4749654
## Acetic       0.0734886
## H2S          0.8771423
## Lactic       4.4106765
```

(ii)

```
t_test <- 2*pt(q=abs(tstat), df=n-p, lower=FALSE)
t_test[2]
```

```
## [1] 0.9419798
```

```
summary(taste_model)
```

```
##
## Call:
## lm(formula = cheddar$taste ~ cheddar$Acetic + cheddar$H2S + cheddar$Lactic)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.390  -6.612  -1.009   4.908  25.449
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -28.8768    19.7354  -1.463  0.15540
## cheddar$Acetic   0.3277     4.4598   0.073  0.94198
## cheddar$H2S      3.9118     1.2484   3.133  0.00425 **
## cheddar$Lactic  19.6705     8.6291   2.280  0.03108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.13 on 26 degrees of freedom
## Multiple R-squared:  0.6518, Adjusted R-squared:  0.6116
## F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

(iii)

The results from the hypothesis test, lead us to believe that we should exclude acetic from the model.

(iv)

```
tcrit<- qt(0.975, n-p)
tcrit
```

```
## [1] 2.055529
```

(v)

```
lower <- betahat[2] - tcrit * estimated_se[2]
upper <- betahat[2] + tcrit * estimated_se[2]
ci <- c(lower, upper)
ci
```

```
## cheddar$Acetic cheddar$Acetic
##      -8.839420       9.494902
```

```
# Verify Results
confint(taste_model)
```

```
##                     2.5 %    97.5 %
## (Intercept)    -69.443503 11.689964
## cheddar$Acetic  -8.839420  9.494902
## cheddar$H2S      1.345656  6.478026
## cheddar$Lactic   1.933267 37.407820
```

Problem 3

The interval is gonna be roughly the same, becausethe fact that estimating B is a linear function, meaning it is similar to Least Squares.