STAT 308 – Homework 5

For the problems in which calculations are needed, please include your `R` code with your answers, otherwise you will not be given full credit. Please upload your assignment by Thursday, October 20, 11:59 pm in a pdf file to Sakai.

- 1. Consider the r dataset `X.rds`, which contains a 60×4 matrix of predictors and `y.rds` which contains a 60×1 vector of observed responses.

To load an .RDS file into `R`, use the function `readRDS` instead of `read.csv`

a. State the least squares regression line in the form $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$

```
X <- readRDS("../Data/X.rds")
y <- readRDS("../Data/y.rds")
betahat <- solve(t(X) %*% X) %*% t(X) %*% y
betahat
```

```
##           [,1]
##     2.64178247
## x1 -3.40171357
## x2  5.07513408
## x3  0.09635578
```

$\hat{y} = 2.65 - 3.40x_1 + 5.08x_2 + 0.10x_3$

b. What is the sum of squared errors of the least squares regression line?

```
SSE <- as.vector(t(y) %*% y - t(betahat) %*% t(X) %*% y)
SSE
```

```
## [1] 77.40101
```

$SSE = 77.40$

c. What is the estimated regression variance?

```
n <- length(y)
s2 <- SSE/(n - 4) # 4 columns in X
s2
```

```
## [1] 1.382161
```

$s^2 = 1.382$ d. What is the estimate of the standard error for β_1?

```
s2 * (solve(t(X) %*% X))
```

```
##                            x1            x2           x3
##     0.049449516  0.0054795788 -0.0626490310  0.004240318
## x1  0.005479579  0.0062582926 -0.0009148134 -0.001228159
## x2 -0.062649031 -0.0009148134  0.2075926758  0.012062439
## x3  0.004240318 -0.0012281585  0.0120624394  0.025616628
```

```
sqrt(0.0062582926)
```

```
## [1] 0.07910937
```

$s_{\hat{\beta}_1} = 0.079$

2. Consider the dataset `economy.csv` which contains the following variables:

- `CRUDE`: dollars per barrel of crude oil
- `INTEREST`: interest on ten-year treasury notes
- `FOREIGN`: foreign investments in billions of dollars
- `DJIA`: Dow Jones industrial average
- `GNP`: Gross national product in billions of dollars
- `PURCHASE`: Purchasing power of U.S. dollar (in 1983 dollars)
- `CONSUMER`: Consumer debt in billions of dollars

Suppose we wish to create a linear model for crude oil price based on the other six variables in the dataset.

a. State the least squares regression line for this linear model.

```
econ <- read.csv("../Data/economy.csv")
mod <- lm(CRUDE ~ .,econ)
summary(mod)
```

```
##
## Call:
## lm(formula = CRUDE ~ ., data = econ)
##
## Residuals:
##       1       2       3       4       5       6       7       8       9      10
##  0.2113  1.1281 -1.6229 -0.7262  1.2789  0.4397 -1.6579  1.1524 -1.6288  2.7099
##      11      12
## -1.0706 -0.2140
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.648714  83.823554   0.318   0.7634
## INTEREST      2.773236   0.939115   2.953   0.0318 *
## FOREIGN       0.025056   0.179751   0.139   0.8946
## DJIA          0.001271   0.006438   0.197   0.8513
## GNP          -0.002401   0.031219  -0.077   0.9417
## PURCHASE    -16.981870  29.549097  -0.575   0.5904
```

```
## CONSUMER     -0.021833   0.073024  -0.299   0.7770
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.077 on 5 degrees of freedom
## Multiple R-squared:  0.9742, Adjusted R-squared:  0.9433
## F-statistic:  31.5 on 6 and 5 DF,  p-value: 0.0008089
```

$\hat{CRUDE} = 26.65 + 2.77 * INTEREST + 0.03 * FOREIGN + 0.001 * DJIA - 0.002 * GNP - 16.98 * PURCHASE - 0.02 * CONSUMER$

b. Interpret the parameter associated with the variable `FOREIGN` in the context of the problem.

When foreign investment increases by 1 billion dollars, we expect the price of crude oil per barrel to increase by 0.03, holding all other variables in the model constant.

c. Interpret the parameter associated with the variable `DJIA` in the context of the problem.

When the Dow Jones Industrial Average increases by 1 dollar, we expect the price of crude oil per barrel to increase by 0.001, holding all other variables in the model constant.

d. What are the error degrees of freedom for this linear model?

```
df_error <- df.residual(mod)
df_error
```

```
## [1] 5
```

$df_{error} = 5$

e. What is the estimate of the regression variance?

```
SSE <- deviance(mod)
s2 <- SSE/df_error
s2
```

```
## [1] 4.314504
```

$s^2 = 4.315$

f. What are the sum of squared errors?

```
SSE
```

```
## [1] 21.57252
```

$SSE = 21.57$

g. State the value of r^2 and interpret this value in the context of the problem.

```
summary(mod)$r.squared
```

```
## [1] 0.9742308
```

97.4% of the variation in crude oil prices can be explained by the linear model with interest rates, foreign investment, Dow Jones Industrial Average, gross national product, purchasing power of the U.S. dollar, and consumer debt included as covariates.