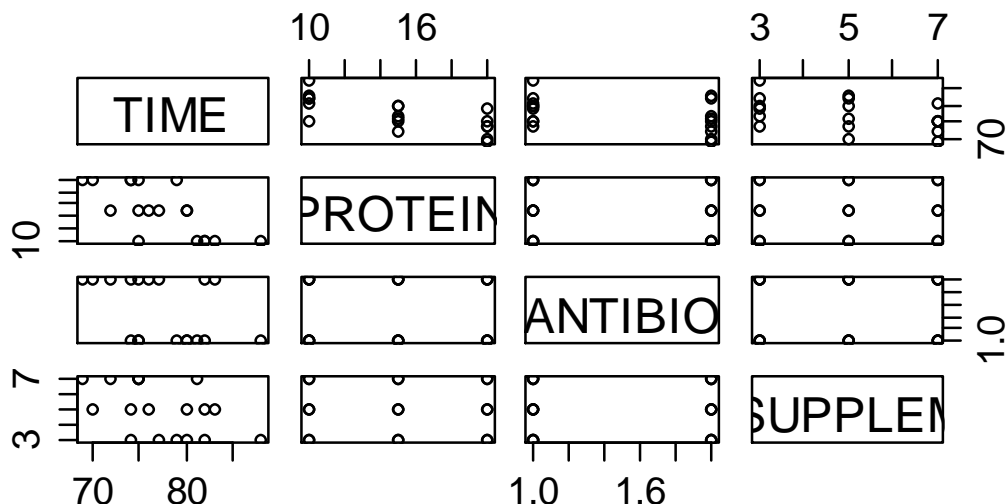## STAT 512 HW1
28 points Total, 2 points per question unless otherwise noted.

1. Scatterplots

2. Pairwise Correlations

```
                TIME      PROTEIN     ANTIBIO     SUPPLEM
TIME       1.0000000 -0.7111002 -0.4180398 -0.4693261
PROTEIN   -0.7111002  1.0000000  0.0000000  0.0000000
ANTIBIO   -0.4180398  0.0000000  1.0000000  0.0000000
SUPPLEM   -0.4693261  0.0000000  0.0000000  1.0000000
```

3. Simple Linear Regressions w/ R2 (**6 pts**)

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  89.8333      3.2022  28.054 4.92e-15 ***
PROTEIN      -0.8333      0.2060  -4.046 0.000938 ***

Multiple R-squared:  0.5057
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   83.333       3.436  24.254  4.8e-14 ***
ANTIBIO       -4.000       2.173  -1.841   0.0843 .

Multiple R-squared:  0.1748
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  84.2083      3.4019  24.753 3.49e-14 ***
SUPPLEM      -1.3750      0.6468  -2.126   0.0494 *

Multiple R-squared:  0.2203
```
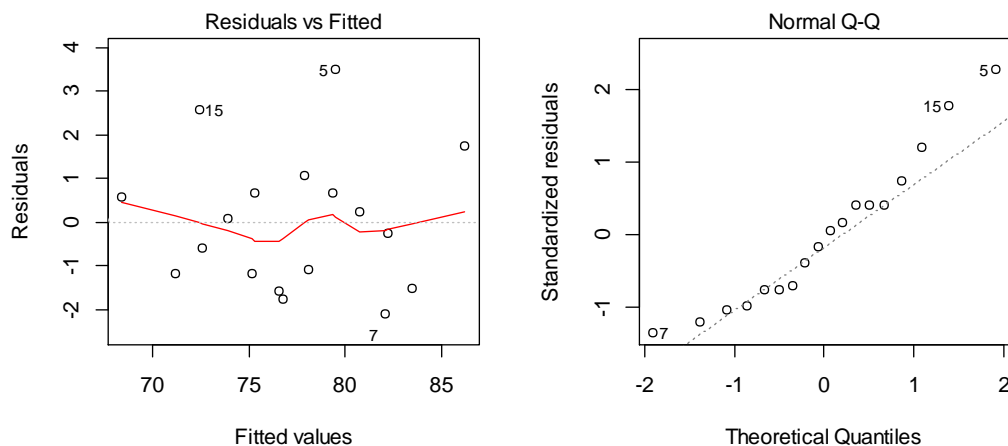
4. Multiple Regression w/ R2

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 102.7083     2.3104  44.455  < 2e-16 ***
PROTEIN      -0.8333     0.0987  -8.443 7.27e-07 ***
ANTIBIO      -4.0000     0.8059  -4.963 0.000208 ***
SUPPLEM      -1.3750     0.2467  -5.572 6.88e-05 ***

Multiple R-squared:  0.9007
```

5. This is a designed experiment. So Protein, Time and Antibio are balanced and <u>uncorrelated</u>.

6. Diagnostic plots (**4 pts**)

The Resids vs Fitted plot gives us information about the asssumptions of linearity and equal variance. No curvature or (obvious) megaphone, so assumptions satisfied.
The QQplot gives us information about the assumption of normality. There is some slight curvature (due to a few moderate outliers), but looks OK.

7. 90.07% of the variation in time is explained by the regression on protein, supplement and antibiotics.

8. Interpretation of Intercept and Partial Regression Coefficient for Antibio:
A <u>one unit increase in antibio</u> is associated with <u>a decrease of 4 units in average finishing time</u>, while <u>holding other variables (protien, supplem) constant</u>.

9. Two sided p-values

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 102.7083     2.3104  44.455  < 2e-16 ***
PROTEIN      -0.8333     0.0987  -8.443 7.27e-07 ***
ANTIBIO      -4.0000     0.8059  -4.963 0.000208 ***
SUPPLEM      -1.3750     0.2467  -5.572 6.88e-05 ***
```

10. (**4 pts**) Test $H_0: \beta_1 = -3$ versus $H_A: \beta_1 \neq -3$ (H0: Protein = -3.0)
Test Statistic: F = 481.87 or t = (-0.8333 + 3)/0.0987 = 21.95238
p-value < 0.001
Reject H0, conclude β_1 is different from -3.

R Code:

```
library(car)
InData <- read.csv("C:/hess/STAT511_FA11/ASCII-comma/CH12/ex12-
53.TXT", quote = " ' ", row.names = 1)
str(InData)
#1
pairs(InData[,c(4,1:3)])
#2
cor(InData[,c(4,1:3)])
#3
Model1 <- lm(TIME ~ PROTEIN, data = InData)
summary(Model1)
Model2 <- lm(TIME ~ ANTIBIO, data = InData)
summary(Model2)
Model3 <- lm(TIME ~ SUPPLEM, data = InData)
summary(Model3)
#4
Model4 <- lm(TIME ~ PROTEIN + ANTIBIO + SUPPLEM, data = InData)
summary(Model4)
#6
par(mfrow=c(1,2))
plot(Model4)
#10
#Using lht()
c1 <- c(0, 1, 0, 0)
lht(Model4, c1, rhs = c(-3))
#By hand
t <- (-0.8333  + 3)/0.0987
t
2*(1-pt(abs(t), df = 14))#11
```