36-401 Modern Regression HW #5 Solutions

DUE: 10/20/2017 at 3PM

Problem 1 [20 points]

(a)

```r
pairs(stackloss, font.labels = 3, font.axis = 5, pch = 19)
```

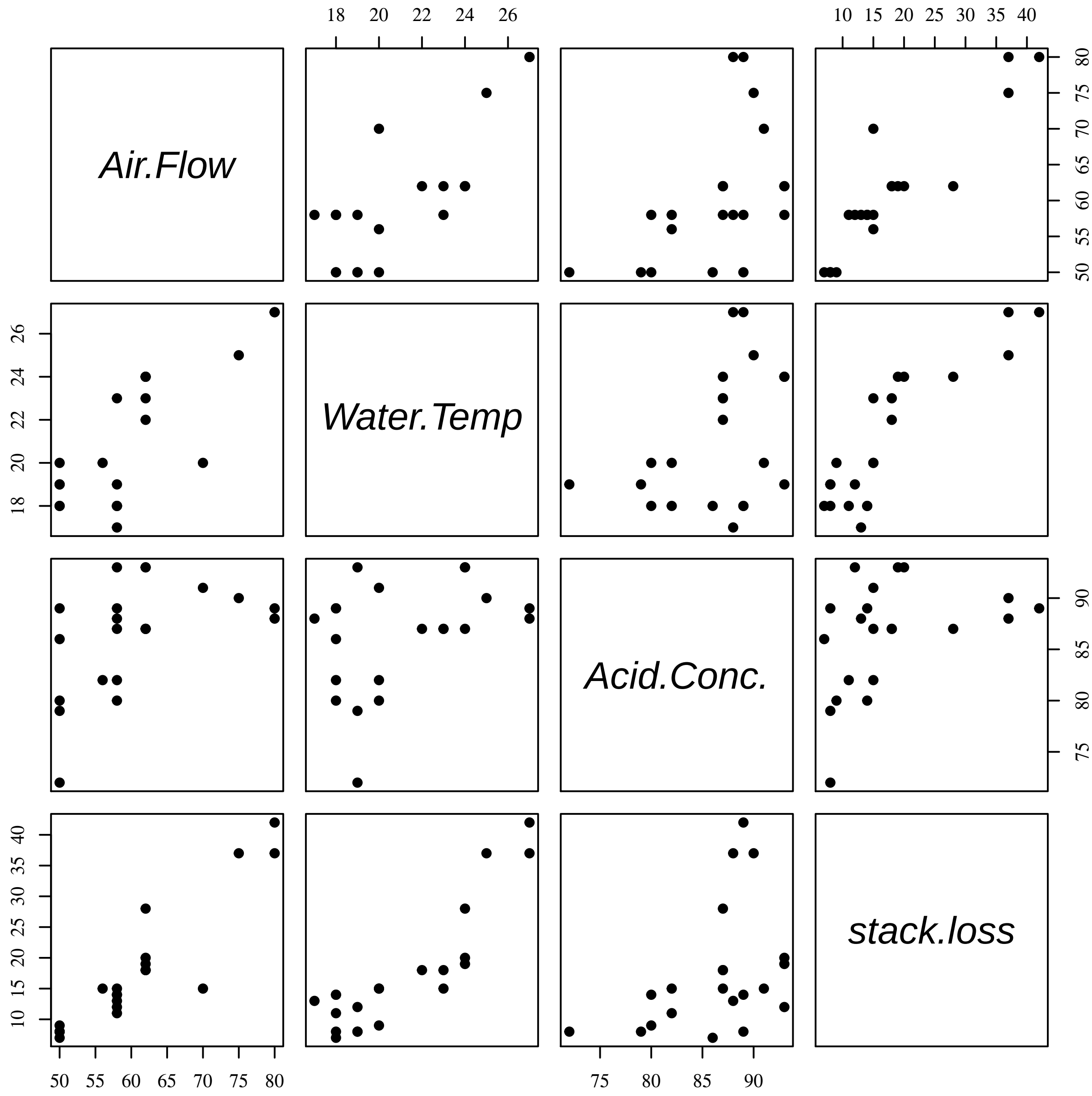

Figure 1: Pairwise associations of variables from *stackloss* data set

(b)

```r
model <- lm(stack.loss ~ Air.Flow + Water.Temp + Acid.Conc., data = stackloss)
summary(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 F-test yields a p-value of 3.016×10^{-9}, which strongly suggests that *at least one* of the predictors has a significant association with `Stack Loss`. The univariate t-tests suggest that both `Air Flow` and `Water Temperature` have significant associations with `Stack Loss`, even after accounting for all other predictors.

(c)

```r
library(knitr)
kable(confint(model, level = 0.9), digits = 2,
      caption = "90% Confidence Intervals for Regression Coefficients")
```

Table 1: 90% Confidence Intervals for Regression Coefficients

	5 %	95 %
(Intercept)	-60.61	-19.23
Air.Flow	0.48	0.95
Water.Temp	0.66	1.94
Acid.Conc.	-0.42	0.12

(d)

```r
kable(predict(model, newdata = data.frame(Air.Flow = 58, Water.Temp = 20, Acid.Conc. = 86),
      interval = "prediction", level = 0.99), digits = 3,
      caption = "99% Prediction Interval for Stack Loss given Airflow = 58,
      Water temperature = 20 and Acid = 86", col.names = c("Prediction", "Lower bound",
      "Upper bound"))
```

Table 2: 99% Prediction Interval for Stack Loss given Airflow = 58, Water temperature = 20 and Acid = 86

Prediction	Lower bound	Upper bound
14.411	4.76	24.061

(e)

In part (b) we saw the p-value for the t-test testing this hypothesis is 0.3440, so we fail to reject H_0.

This hypothesis can also equivalently be tested using a partial F-test. Table 3 shows the ANOVA table for the regression and the partial F-test of H_0 again yields the p-value of 0.3440.

```r
kable(anova(model), digits = 3, caption = "ANOVA Table for Regression")
```

Table 3: ANOVA Table for Regression

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Air.Flow	1	1750.122	1750.122	166.371	0.000
Water.Temp	1	130.321	130.321	12.389	0.003
Acid.Conc.	1	9.965	9.965	0.947	0.344
Residuals	17	178.830	10.519	NA	NA

Problem 2 [20 points]

$$\begin{aligned}
\widehat{\beta} &= \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} \\
&= (X^T X)^{-1} X^T Y \\
&= \left[\begin{pmatrix} X_{11} & \cdots & X_{1n} \\ X_{21} & \cdots & X_{2n} \end{pmatrix} \begin{pmatrix} X_{11} & X_{21} \\ \vdots & \vdots \\ X_{1n} & X_{2n} \end{pmatrix} \right]^{-1} \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ X_{21} & \cdots & X_{2n} \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i} X_{2i} \\ \sum_{i=1}^n X_{1i} X_{2i} & \sum_{i=1}^n X_{2i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n X_{1i} Y_i \\ \sum_{i=1}^n X_{2i} Y_i \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n X_{1i}^2 & 0 \\ 0 & \sum_{i=1}^n X_{2i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n X_{1i} Y_i \\ \sum_{i=1}^n X_{2i} Y_i \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sum_{i=1}^n X_{1i}^2} & 0 \\ 0 & \frac{1}{\sum_{i=1}^n X_{2i}^2} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n X_{1i} Y_i \\ \sum_{i=1}^n X_{2i} Y_i \end{pmatrix} \\
&= \begin{pmatrix} \frac{\sum_{i=1}^n X_{1i} Y_i}{\sum_{i=1}^n X_{1i}^2} \\ \frac{\sum_{i=1}^n X_{2i} Y_i}{\sum_{i=1}^n X_{2i}^2} \end{pmatrix}
\end{aligned}$$

As we saw in Homework 2, these are the least square estimators for the two separate univariate regressions through the origin.

Problem 3 [20 points]

(a)

```
X <- matrix(c(1,1,1,1,4,3,10,7,5,4,9,10), ncol = 3)
Y <- matrix(c(25,20,57,51), ncol = 1)
model3 <- lm(Y ~ X - 1)
summary(model3)
```

```
##
## Call:
## lm(formula = Y ~ X - 1)
##
## Residuals:
##        1        2        3        4
## -0.70098  0.57353  0.06373  0.06373
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## X1  -2.6029     1.3382  -1.945   0.3023
## X2   3.0686     0.3249   9.444   0.0672 .
## X3   3.2059     0.3490   9.185   0.0690 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9102 on 1 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9995
## F-statistic:  2766 on 3 and 1 DF,  p-value: 0.01398
```

```
mytable <- summary(model3)$coefficients
row.names(mytable) <- c("Intercept","X1","X2")
kable(mytable)
```

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-2.602941	1.3382353	-1.945055	0.3023200
X1	3.068627	0.3249375	9.443746	0.0671615
X2	3.205882	0.3490389	9.184886	0.0690397

The F-test yields a p-value of 0.0139, which suggests that *at least one* of X_1 and X_2 has a significant association with Y. However, as indicated by the t-tests, there is not enough evidence to conclude individually that X_1 has a significant association with Y (after accounting for X_2), or that X_2 has a significant association with Y (after accounting for X_1).

(b)

```r
library(xtable)
# you can use this package to convert R objects into nice LaTeX tables/matrices
print(xtable(t(X) %*% X), tabular.environment = "pmatrix",
      include.rownames = FALSE, include.colnames = FALSE, hline.after = NULL)
print(xtable(solve(t(X) %*% X)), tabular.environment = "pmatrix",
      include.rownames = FALSE, include.colnames = FALSE, hline.after = NULL)
```

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} 4.00 & 24.00 & 28.00 \\ 24.00 & 174.00 & 192.00 \\ 28.00 & 192.00 & 222.00 \end{pmatrix}$$

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{pmatrix} 2.16 & 0.06 & -0.32 \\ 0.06 & 0.13 & -0.12 \\ -0.32 & -0.12 & 0.15 \end{pmatrix}$$

(c)

```r
print(xtable(solve(t(X) %*% X) %*% t(X) %*% Y), tabular.environment = "pmatrix",
      include.rownames = FALSE, include.colnames = FALSE, hline.after = NULL)
```

$$\begin{aligned} \widehat{\beta} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \\ &= \begin{pmatrix} -2.60 \\ 3.07 \\ 3.21 \end{pmatrix} \end{aligned}$$

(d)

```r
print(xtable(X %*% solve(t(X) %*% X) %*% t(X)), tabular.environment = "pmatrix",
      include.rownames = FALSE, include.colnames = FALSE, hline.after = NULL)
```

$$\begin{aligned} \boldsymbol{H} &= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \\ &= \begin{pmatrix} 0.41 & 0.49 & 0.05 & 0.05 \\ 0.49 & 0.60 & -0.04 & -0.04 \\ 0.05 & -0.04 & 1.00 & -0.00 \\ 0.05 & -0.04 & -0.00 & 1.00 \end{pmatrix} \end{aligned}$$

(e)

$$\begin{aligned} \mathrm{Var}(\widehat{\beta}) &= \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\ &= \sigma^2 \begin{pmatrix} 2.16 & 0.06 & -0.32 \\ 0.06 & 0.13 & -0.12 \\ -0.32 & -0.12 & 0.15 \end{pmatrix} \end{aligned}$$

Note: This is the *variance-covariance* matrix of $\widehat{\beta}$. Notice it depends on the true (unknown) distribution of the data. The standard errors provided by R's `summary` command are plug-in estimates for the square root of the diagonal elements of $\text{Var}(\widehat{\beta})$. That is, R gives you

$$\begin{aligned}
\left\{\widehat{\text{se}}(\widehat{\beta}_k)\right\}_{k=0}^{p} &= \left\{\widehat{\sigma}\sqrt{(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{jj}}\right\}_{j=1}^{p+1} \\
&= \sqrt{\frac{1}{n-(p+1)}\sum_{i=1}^{n} e_i^2}\left\{\sqrt{(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{jj}}\right\}_{j=1}^{p+1} \\
&= 0.9101821 \cdot \begin{pmatrix} 1.4702941 \\ 0.3570028 \\ 0.3834825 \end{pmatrix} \\
&= \begin{pmatrix} 1.3382353 \\ 0.3249375 \\ 0.3490389 \end{pmatrix}
\end{aligned}$$

Problem 4 [20 points]

$$\begin{aligned}
\text{tr}(\boldsymbol{H}) &= \text{tr}(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T) \\
&= \text{tr}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}) \\
&= \text{tr}((\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{X})) \\
&= \text{tr}(\boldsymbol{I}_{p+1}) \\
&= p+1
\end{aligned}$$

where we have used the cyclic property of the trace operator and $\boldsymbol{I}_{p+1}$ is the $(p+1)\times(p+1)$ identity matrix.

Problem 5 [20 points]

$$\begin{aligned}
\widehat{\boldsymbol{Y}}^T\boldsymbol{e} &= (\boldsymbol{H}\boldsymbol{Y})^T(\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{Y}) \\
&= \boldsymbol{Y}^T\boldsymbol{H}^T(\boldsymbol{I}_n-\boldsymbol{H})\boldsymbol{Y} \\
&= \boldsymbol{Y}^T\boldsymbol{H}(\boldsymbol{I}_n-\boldsymbol{H})\boldsymbol{Y} \\
&= \boldsymbol{Y}^T(\boldsymbol{H}-\boldsymbol{H}^2)\boldsymbol{Y} \\
&= \boldsymbol{Y}^T(\boldsymbol{H}-\boldsymbol{H})\boldsymbol{Y} \\
&= \boldsymbol{Y}^T(\boldsymbol{0}_{n\times n})\boldsymbol{Y} \\
&= 0
\end{aligned}$$