MATH 232: Engineering Mathematics
Reading Guide for LAS, Section 4.4: Linear Regression

**Goals:**
1. To be able to use R to plot bivariate data in order to investigate whether there is a relationship between quantitative variables.
2. To be able to use R (and, ideally, Octave, too) to find the least-squares best-fit regression line in those instances when the data suggests a linear model between variables is warranted.
3. To be able to use R tools and output to comment on how successful the linear model accounts for the variability in the response variable.
4. To obtain a geometric understanding of the breakdown of total variation in the response variable into parts accounted for and not accounted for by the linear model:

\[ \text{SST} = \text{SSR} + \text{SSResid} \]

**Read:** Section 4.4 of LAS

**Useful R commands:**

<table>
<thead>
<tr>
<th>Commands</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>xyplot()</td>
<td>Lattice command for producing a scatterplot</td>
</tr>
<tr>
<td>lm()</td>
<td>Determine the least-squares regression line</td>
</tr>
<tr>
<td>summary()</td>
<td>When used on results of lm(), summarizes them</td>
</tr>
<tr>
<td>anova()</td>
<td>Provides analysis of variance (can use on results of lm())</td>
</tr>
<tr>
<td>predict()</td>
<td>Obtain predicted values from linear model</td>
</tr>
<tr>
<td>residuals()</td>
<td>Obtain residuals from linear model</td>
</tr>
</tbody>
</table>

**Examples (without output):**

```r
> wine = read.csv('http://www.calvin.edu/~scofield/data/comma/heartDiseaseDeathsAndWine.csv')
> names(wine)
> require(lattice)
> xyplot(hddeaths ~ winealc, data=wine, type=c('p','r'))
> lm.1 = lm(hddeaths ~ winealc, data=wine)
> summary(lm.1)
> anova(lm.1)

> predict(lm.1)
> wine$winealc - predict(lm.1)
> residuals(lm.1)
> sum(residuals(lm.1)^2)
> xyplot(residuals(lm.1) ~ wine$winealc)
```
Reading Guide for LAS, Section 4.4

Linear Regression

\[
> \text{lm.2 = lm(hddeaths } \sim \text{ winealc, data=wine, subset=(winealc < 5.5))}
\]
\[
> \text{summary(lm.2)}
\]

Terms to know:
least-squares (regression) line, model, observed, predicted/fitted values, residual, model matrix, model space, normal equations, SST, SSR, SSResid