1. The standard linear model.

\[ Y = \beta_0 + \beta_1 x + \epsilon \]

where

(a) \( \epsilon \) is a random variable with mean 0 and variance \( \sigma^2 \),
(b) \( \beta_0, \beta_1, \sigma^2 \) are (unknown) parameters,
(c) and \( \epsilon \) has a normal distribution.

2. The data \((x_1, y_1), \ldots, (x_n, y_n)\) comes from \(n\) independent and identically distributed \(\epsilon_i\).

3. Estimating \(\beta_0, \beta_1\). Use the least squares estimates!

\[
\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.
\]

4. The distribution of \(\beta_1\).

(a) \( \hat{\beta}_1 \) is an unbiased estimator of \( \beta_1 \).
(b) \( \text{Var}(\beta_1) = \frac{\sigma^2}{S_{xx}} \).
(c) \( \hat{\beta} \) has a normal distribution.

5. To make a confidence interval for \(\hat{\beta}_1\), we need to estimate \(\sigma^2\). First, using

\[
\text{SSResid} = \sum (y_i - \hat{y}_i)^2 = \sum \left( y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right)^2
\]

we define

\[
\text{MSResid} = \frac{\text{SSResid}}{n-2}
\]

MSResid is an unbiased estimator of \(\sigma^2\). We call \( s = \sqrt{\text{MSResid}} \) the residual standard error and use it to estimate \( \sigma \).

6. The standard error of \(\hat{\beta}_1\) is \( s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}} \).

7. Confidence interval for \(\beta_1\):

\[
\hat{\beta}_1 \pm t^* s_{\hat{\beta}_1} \quad t^* = t_{\alpha/2, n-2}
\]

**Homework - due Monday, May 5, 2008**

1. Read Section 8.2.
2. Do problems 8.1, 8.2ab, 8.3ab, 8.5a.
Useful R

```r
> fo=read.csv('http://www.calvin.edu/~stob/data/fruitohms.csv')
> l=lm(ohms~juice,data=fo)
> xyplot(ohms~juice,data=fo,type=c("p","r"))
> summary(l)
Call:
lm(formula = ohms ~ juice, data = fo)

Residuals:
    Min     1Q Median     3Q    Max
-4198.386 -793.855   3.758  614.348  3139.493

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7519.404    233.779   32.16  <2e-16 ***
juice         -89.877     6.022   -14.93  <2e-16 ***

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1122 on 126 degrees of freedom
Multiple R-Squared: 0.6387,    Adjusted R-squared: 0.6358
F-statistic: 222.7 on 1 and 126 DF,  p-value: < 2.2e-16

> confint(l)
     2.5 %      97.5 %
(Intercept)   7056.7633   7982.04545
juice        -101.7949   -77.96007
```