Analysis of Variance Model

- Let $Y_{ij}$ = measured height of plant $i$ in treatment group $j$, $j = 1, 2, 3, 4$. Treatment groups are “factors”.
- Let $\bar{Y}_j$ = calculated average of plant heights in group $j$.
- ANOVA model: $Y_{ij} = \mu_j + \varepsilon_{ij}$
- Unknown parameters: group means $\mu_1, \mu_2, \mu_3, \mu_4$ and single standard deviation $\sigma$.
- ANOVA hypothesis and inference
- Splus output:

```
Df  Sum.Sq   M.Sq  F  Pr(F)
Betw   3   100.65  33.55 12.08 0.00062
With  12    33.33   2.78
```

Regression Model

- Let $Y$ = random height of plant (random, response variable)
- Let $X$ = number of nematodes given to each plant. fixed, explanatory
  Explanatory variable $X$ measured at (fixed) values of 0, 1000, 5000 or 10,000.
- Linear Regression Model: probability model relating $Y$ to a treatment level $X$, which is now treated as continuous

$$Y_{ij} = \beta_0 + \beta_1 X_j + \varepsilon$$  \hspace{1cm} (1)

- Unknown parameters: $\beta_0, \beta_1$ and $\sigma$
- Assumptions about $\varepsilon$

- Line of Means:

$$\mu \{ Y_j | X_j \} = \beta_0 + \beta_1 X_j$$  \hspace{1cm} (2)

- Assumptions
- How to find the best fitting values of $\beta_0, \beta_1$

Regression estimation

- Estimated mean function

$$\hat{\mu} \{ Y_j | X_j \} = \hat{\beta}_0 + \hat{\beta}_1 X_j$$  \hspace{1cm} (3)

- Let $X$ be measured in 1000’s of nematodes. $X=0, 1, 5, 10$. From Splus, for the nematode data this is:

$$\hat{\mu} \{ Y | X \} = 10.33 - 0.6X$$  \hspace{1cm} (4)

- What are the properties of $\hat{\beta}_0, \hat{\beta}_1$?
Presentation of Regression Results

\[ \hat{Y} = 10.33 + (-573.79 \times X) \]

(0.69) (122.76)

\[ \hat{\sigma} = 1.93 \text{ (14 df)} \]

Why can’t we just estimate \( \hat{\sigma} \) with \( SD(Y) \)?

Hypothesis tests for \( \beta_0 \) and \( \beta_1 \)

- Does \( X \) contribute any information for prediction of \( Y \)?
  
  Test: \( H_0: \beta_1 = 0 \) vs. \( H_A: \beta_1 \neq 0 \)
  
  (or \( H_A: \beta_1 > 0 \))

  \[ \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim t_{n-2} \] (5)

- Tests regarding \( \beta_0 \):

  \[ \frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \sim t_{n-2} \] (6)