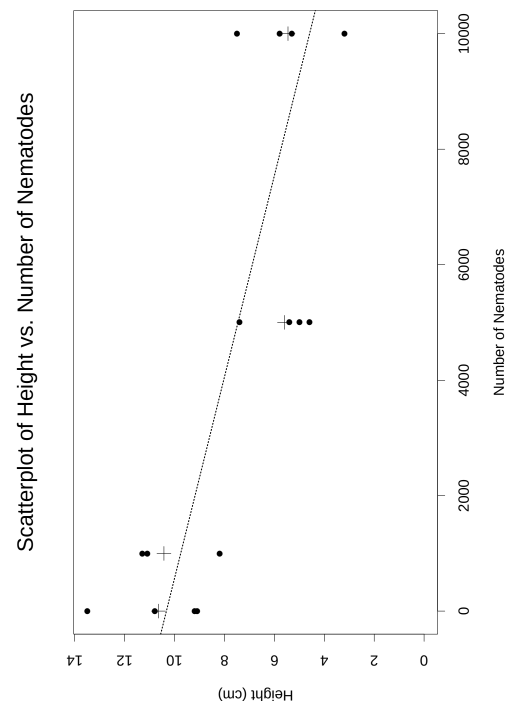

Analysis of Variance Model

- 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 σ.
- 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*

$$Y_{ij} = \beta_o + \beta_1 X_j + \varepsilon \qquad (1)$$

- *Unknown parameters*: β_o, β_1 and σ
- Assumptions about ε

Regression Model

- **Line of Means**:

$$\mu\{Y_j|X_j\} = \beta_o + \beta_1 X_j \qquad (2)$$

- Assumptions
- How to find the best fitting values of β_o, β_1?

Regression estimation

- Estimated mean function

$$\hat{\mu}\{Y_j|X_j\} = \hat{\beta}_o + \hat{\beta}_1 X_j \qquad (3)$$

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

$$\hat{\mu}\{\mathbf{Y}|\mathbf{X}\} = 10.33 - 0.6\mathbf{X} \qquad (4)$$

- What are the properties of $\hat{\beta}_o$, $\hat{\beta}_1$?

	number	height	fitted	residual	sq.resid
1	0	10.8	10.33	0.47	0.22
2	0	9.1	10.33	-1.23	1.50
3	0	13.5	10.33	3.17	10.07
4	0	9.2	10.33	-1.13	1.27
5	1	11.1	9.75	1.35	1.82
6	1	11.1	9.75	1.35	1.82
7	1	8.2	9.75	-1.55	2.41
8	1	11.3	9.75	1.55	2.39
9	5	5.4	7.46	-2.06	4.23
10	5	4.6	7.46	-2.86	8.17
11	5	7.4	7.46	-0.06	0.00
12	5	5.0	7.46	-2.46	6.04
13	10	5.8	4.59	1.21	1.47
14	10	5.3	4.59	0.71	0.51
15	10	3.2	4.59	-1.39	1.93
16	10	7.5	4.59	2.91	8.48

Presentation of Regression Results

$\hat{Y}$ = 10.33 + -573.79 X

(0.69) (122.76)

$\hat{\sigma}$ = 1.93 (14 df)

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

Hypothesis tests for β_o and β_1

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

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

- Tests regarding β_o:

$$\frac{\hat{\beta}_o - \beta_o}{SE(\hat{\beta}_o)} \sim t_{n-2} \qquad (6)$$