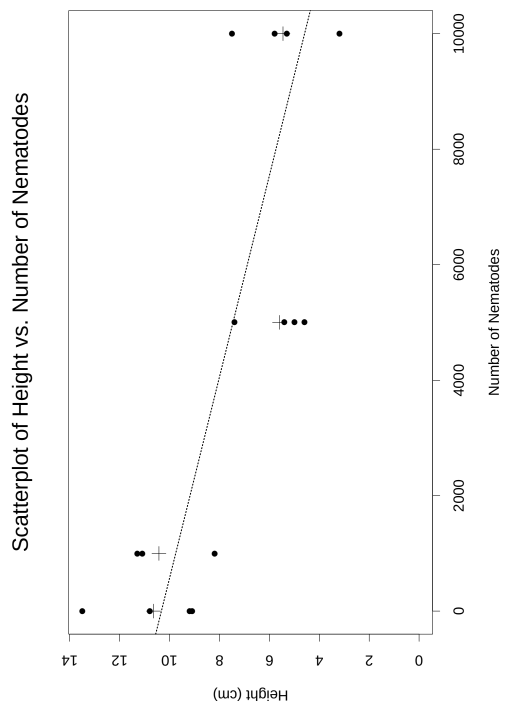

Analysis of Variance Model

- Let Y_{ij}= height of plant i in treatment group j, $j = 1, 2, 3, 4.$
- Let $\bar{Y}_j$= mean of plant heights in group j.
- **ANOVA model**: $Y_{ij} = \bar{Y}_j + \varepsilon_{ij}$
- *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 (*response* variable)
- Let X= number of nematodes given to each plant. *Explanatory variable* X measured at (*fixed*) values of 0, 1000, 5000 or 10,000.
- **Linear Regression Model**: probability model relating Y to X

$$Y = \beta_o + \beta_1 X + \varepsilon \tag{1}$$

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

Regression Model

- **Line of Means**:

$$\mu\{Y|X\} = \beta_o + \beta_1 X \tag{2}$$

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

Regression estimation

- Estimated mean function

$$\hat{\mu}\{Y|X\} = \hat{\beta}_o + \hat{\beta}_1 X \tag{3}$$

- From Splus, for the nematode data this is:

$$\hat{\mu}\{Y|X\} = 10.33 - 0.0006X \tag{4}$$

- 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	1000	11.1	9.75	1.35	1.82
6	1000	11.1	9.75	1.35	1.82
7	1000	8.2	9.75	-1.55	2.41
8	1000	11.3	9.75	1.55	2.39
9	5000	5.4	7.46	-2.06	4.23
10	5000	4.6	7.46	-2.86	8.17
11	5000	7.4	7.46	-0.06	0.00
12	5000	5.0	7.46	-2.46	6.04
13	10000	5.8	4.59	1.21	1.47
14	10000	5.3	4.59	0.71	0.51
15	10000	3.2	4.59	-1.39	1.93
16	10000	7.5	4.59	2.91	8.48

Splus output for nematode regression

```
Coefficients:
                 Value Std. Error  t value Pr(>|t|)
(Intercept)  10.3264   0.6890    14.9876   0.0000
     number  -0.0006   0.0001    -4.6740   0.0004

Residual standard error: 1.933 on 14 df
Multiple R-Squared: 0.6094
F-stat: 21.85 on 1 and 14 df, p-val is 0.0003584

Correlation of Coefficients:
       (Intercept)
number -0.7127
```


Presentation of Regression Results

$$\hat{Y} = \underset{(0.69)}{10.32} + \underset{(0.0001)}{-0.0006}\, X$$

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

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 $H_A : \beta_1 > 0$)

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

- Tests regarding β_o:

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