One-way ANOVA: Example

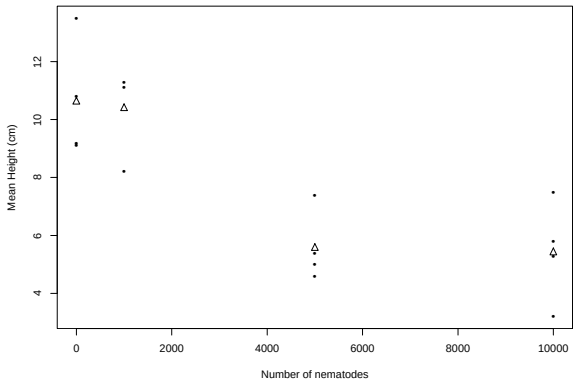
How do nematodes (microscopic worms) affect plant growth? A botanist prepares 16 identical planting pots and randomly assigns each pot to receive different numbers of nematodes. A tomato seedling is transplanted into each plot.

Table 1: Increase in the height of tomato seedlings 16 days after planting

Nematodes	Seedling Growth (cm)			
0	10.8	9.1	13.5	9.2
1000	11.1	11.1	8.2	11.3
5000	5.4	4.6	7.4	5.0
10000	5.8	5.3	3.2	7.5

Exploratory Data Analysis

Nematodes	$\bar{X}$	s
0	10.650	2.053
1000	10.425	1.486
5000	5.600	1.244
10000	5.450	1.771



One-Way ANOVA

- Goal: comparison of k population or treatment means $\mu_1, \cdots, \mu_k$
- Assumptions
  - Each of the k populations is normally distributed.
  - $\sigma_1 = \cdots = \sigma_k$
  - Independence of observations both within and between the k groups.
- Test of Hypotheses: (k = 4 here) Let μ_i = mean growth (cm) for nematode level i, where $i = 1, 2, 3, 4$.
  - H_o : $\mu_1 = \mu_2 = \mu_3 = \mu_4$
  - H_A : at least two of the μ_i are different.

ANOVA model

$$y_{ij} = \mu_j + \varepsilon_{ij}$$
$$i = 1, \cdots, n_j$$
$$j = 1, \cdots, K$$

- y_{ij}'s varying about their population mean μ_j.
- $\varepsilon_{ij} \sim N(0, \sigma)$ where ε_{ij} represent random variation from fit of ANOVA model.

Components of Variation

- Variation *between* k sample means: Extra Sum of Squares (SSG)
- Variation *within* the k samples: RSS$_{full}$ (SSE)
- Variation of data around overall mean: RSS$_{reduced}$ (SST)

ANOVA Decomposition

$$\begin{aligned} \text{RSS}_{reduced} &= \text{RSS}_{full} + \text{ESS} \\ \sum_{j=1}^{K}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2 &= \sum_{j=1}^{K}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_j)^2 + \sum_{j=1}^{K}\sum_{i=1}^{n_j}(\bar{y}_j-\bar{y})^2 \end{aligned}$$

Analysis of Variance Table

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-statistic	*p*-value
Between Groups	ESS = $\sum_{j=1}^{K}\sum_{i=1}^{n_j}(\bar{y}_j-\bar{y})^2$	K-1	ESS/(K-1)	$\frac{\text{ESS}/(K-1)}{\text{RSS}_{full}/(N-K)}$	
Within Groups	RSS_{full} = $\sum_{j=1}^{K}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_j)^2$	N-K	RSS_{full}/(N-K) $(= s^2{}_{pooled})$		
Total	$\text{RSS}_{reduced}$ = $\sum_{j=1}^{K}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2$	N-1			

Nematodes: Analysis of Variance Table

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-statistic	*p*-value
Between Groups	100.65	3	33.55	12.08	.0006
Within Groups	33.33	12	2.78		
Total	133.98	15			

Steps in Splus

1. Number of nematodes is a *factor*. Go to **Data**$\rightarrow$ **Change Data Type**. Change the variables for number of nematodes to **factor**.
2. **Statistics** $\rightarrow$ **ANOVA** $\rightarrow$ **Fixed Effects**.
   - Dependent Variable: Height
   - Independent Variable: Number of nematodes

This produces the Splus formula: height $\sim$ nematodes

```
*** Analysis of Variance Model ***
Short Output:
Call:
  aov(formula = height ~ num.nema, data = nematode)
Terms:
                num.nema Residuals
 Sum of Squares 100.6469   33.3275
Deg. of Freedom        3        12

Residual standard error: 1.666521
Estimated effects are balanced

            Df Sum of Sq  Mean Sq  F Value   Pr(F)
 num.nema    3 100.6469 33.54896 12.07974 0.000616
 Residuals  12  33.3275  2.77729
```


Example: Insects and colors

The presence of harmful insects in farm fields is detected by erecting boards covered wtih a sticky material and then examining the insects trapped on the board. Which colors are most attractive to cereal leaf beetles? 6 boards in 4 colors were placed in a field of oats.

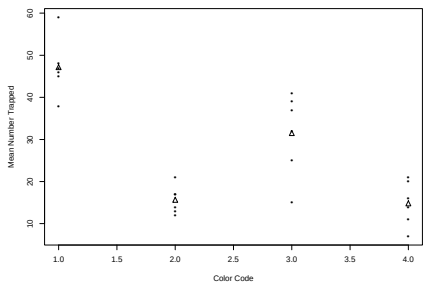
Table 2: Number of cereal leaf beetles trapped

Color	Number trapped					
Yellow	45	59	48	46	38	47
White	21	12	14	17	13	17
Green	37	32	15	25	39	41
Blue	16	11	20	21	14	7

Summary statistics

Color	n	$\bar{X}$	s
Yellow	6	47.17	6.79
White	6	15.67	3.33
Green	6	31.50	9.91
Blue	6	14.83	5.34

Figure 1: 1=Yellow, 2=White, 3=Green, 4=Blue



- Sum of squares: Between: 4218.458, Within: 920.500

Analysis of Variance Table

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-statistic	*p*-value
Between Groups					
Within Groups					
Total					