Example of One-Way Analysis of Variance

Notation and Formulas

k treatments; n_i observations on the ith treatment; y_{ij} denotes the jth observation on the ith treatment.

$$\begin{array}{lccccc} \text{Treatment 1:} & y_{11} & y_{12} & y_{13} & \cdots & y_{1n_1} \\ \text{Treatment 2:} & y_{21} & y_{22} & y_{23} & \cdots & y_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \text{Treatment } k\text{:} & y_{k1} & y_{k2} & y_{k3} & \cdots & y_{kn_k} \end{array}$$

treatment sum: $y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$

treatment mean: $\bar{y}_{i\cdot} = y_{i\cdot}/n_i$

overall sum: $y_{\cdot\cdot} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$

overall mean: $\bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/N$ where $N = \sum_{i=1}^{k} n_i$

Sums of Squares:

$$\begin{aligned} \text{SSTot} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{\textit{all obs}} (y_{ij} - \bar{y}_{\cdot\cdot})^2 \\ \text{SSTrt} &= \sum_{i=1}^{k} n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 \\ \text{SSErr} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2 \end{aligned}$$

degrees of freedom:

$$\text{dfTot} = N - 1 \quad \text{dfTrt} = k - 1 \quad \text{dfErr} = N - k.$$

Example

data:

y_{1j}:	14	20	10	14	17
y_{2j}:	13	8	10	16	8
y_{3j}:	16	14	24	21	19
y_{4j}:	8	14	19	12	15

stem-and-leaf display:

```
Group:  1     2     3     4

     0.|     |     |     |
     0*|     |88   |     |8
     1.|044  |03   |4    |24
     1*|7    |6    |69   |59
     2.|0    |     |14   |
```

ANOVA table:

Source	df	SS	MS
Treatment	3	158.8	52.93
Error	16	232.0	14.50
Total	19	390.8	

Hypotheses: H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ versus H_A: not $\mu_1 = \mu_2 = \mu_3 = \mu_4$.
$F = \dfrac{\text{MSTrt}}{\text{MSErr}} = \dfrac{52.93}{14.50} = 3.65$ on (3, 16) df.
The p-value is 0.035 and thus we conclude that there is moderate evidence against the null hypothesis.

Assumptions Underlying ANOVA

1. Independence: For each treatment i, we have a random sample of n_i observations, Y_{ij}.
   Also, observations for each treatment i are independent of observations in all other treatments.
2. Normality: $Y_{ij} \sim \mathcal{N}(\mu_i, \sigma_i^2)$.
3. Equal Variance: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$.

The One-way ANOVA Model

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad \text{where } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad \text{and all } \varepsilon_{ij} \text{ are independent.}$$

Alternatively:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad \text{where } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)\text{ , all } \varepsilon_{ij} \text{ are independent, and } \sum \alpha_i = 0.$$

Testing for Equal Variances

Use Levene's Test (modified by Brown & Forsythe, 1974):

1. For group i, find the median; call it $\tilde{y}_i$.
2. Define $d_{ij} = |y_{ij} - \tilde{y}_i|$.
3. Perform a one-way ANOVA on the d_{ij} from step 2.
4. Reject H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ if the F-test in step 3 is significant.

A Nonparametric Test

Use a "rank transformation" (Conover and Iman, 1981):

1. Place the entire data set in order, from smallest to largest.
2. Replace each observation by its rank.
3. Analyze the rank values in a standard ANOVA.