Experimental Design
Note 3
Introduction to ANOVA

Keunbaik Lee

Sungkyunkwan University

What if there are more than two levels of a single factor?

- The t-test does not directly apply.
- There are lots of practical situations where there are either more than two levels of interest, or there are several factors of simultaneous interest.
- The analysis of variance (ANOVA) is the appropriate analysis "engine" for these types of experiments.

ANOVA - Analysis of Variance I

- Extends independent-samples t test.
- Compares the means of groups of independent observations
  - Do not be fooled by the name
  - ANOVA does not compare variances
  - The name "analysis of variance" stems from a partitioning of the total variability in the response variable into components
  - Can compare more than two groups
- The ANOVA was developed by Fisher in the early 1920s, and initially applied to agricultural experiments. Now it is used widely.

Say the sample contains K independent groups

ANOVA - Analysis of Variance II

- ANOVA tests the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$$

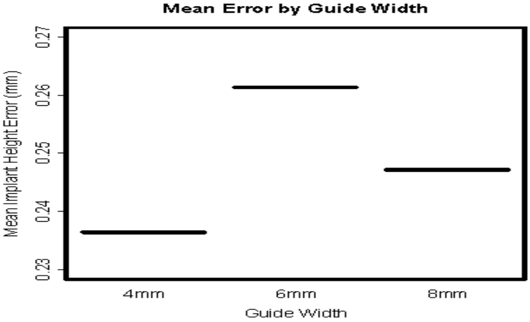
  That is, “the group means are all equal”
- The alternative hypothesis is

$$H_1 : \mu_i \neq \mu_j \text{ for some } i, j$$

  or, “the group means are not all equal”

Example: Accuracy of Implant Placement I

- Implants were placed in a manikin using placement guides of various widths.
- 15 implants were placed using each guide.
- Error (discrepancies with a reference implant) was measured for each implant



Example: Accuracy of Implant Placement II

- Does changing the guide change the mean height error?
- Is there an optimum level for guide?
- We would like to have an objective way to answer these questions
- The t-test really does not apply here - more than two factor levels
    - Pairwise comparisons will inflate type I error

The ANOVA Statistic

- Each time a hypothesis test is performed at significance level α, there is probability rejecting in error.
- Performing multiple tests increases the chances of rejecting in error at least once.
- For example:
  - if you did 3 independent hypothesis tests at the $\alpha = 0.05$
  - If, in truth, H_0 were true for all three.
  - The probability that at least one test rejects H_0 is 14.3% (P(at least one rejection)=1-P(no rejections)$=1 - .95^3 = 0.143$)

Why pairwise comparisons inflates type I error?

- To combine the differences from the grand mean we
  - Square the differences
  - Multiply by the numbers of observations in the groups
  - Sum over the groups

$$SS_B = 15(\bar{X}_{4mm} - \bar{\bar{X}})^2 + 15(\bar{X}_{6mm} - \bar{\bar{X}})^2 + 15(\bar{X}_{8mm} - \bar{\bar{X}})^2$$

where $\bar{X}_*$ are the group means and $\bar{\bar{X}}$ is the grand mean.

$$SS_B = \text{Sum of Squares Between groups}$$

Note: This looks a bit like a variance.

How big is big?

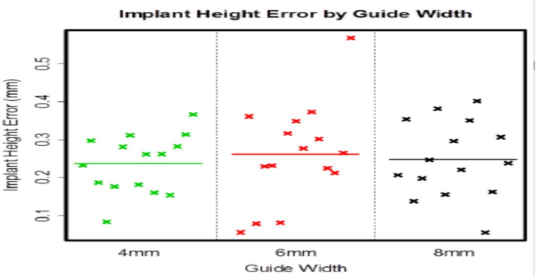
- For the Implant Accuracy Data, $SS_B = 0.0047$
- Is that big enough to reject H_0?
- As with the t test, we compare the statistic to the variability of the individual observations.
- In ANOVA the variability is estimated by the Mean Square Error, or MSE

MSE: Mean Square Error I

The Mean Square Error is a measure of the variability after the group effects have been taken into account.

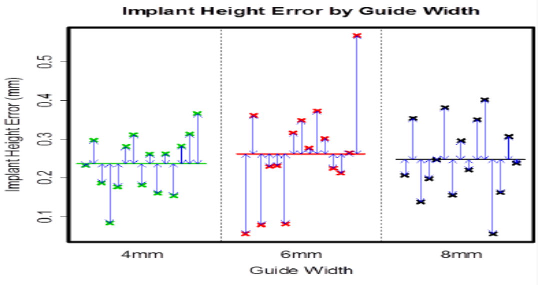
$$MSE = \frac{1}{N-a} \sum_j \sum_i (x_{ij} - \bar{x}_j)^2$$

where x_{ij} is the ith observation in the jth group.



MSE: Mean Square Error II

Note that the variation of the means seems quite small compared to the variance of observations within groups



Notes on MSE

- If there are only two groups, the MSE is equal to the pooled estimate of variance used in the equal-variance t test.
- ANOVA assumes that all the group variances are equal.
- Other options should be considered if group variances differ by a factor of 2 or more.

ANOVA F Test

- The ANOVA F test is based on the F statistic

$$F = \frac{SS_B/(a-1)}{MSE}$$

  where a is the number of groups.

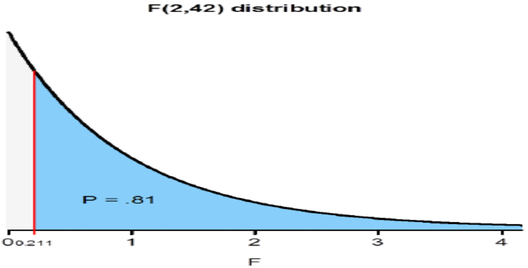
- Under H_0 the F statistic has an "F" distribution, with $a-1$ and $N-a$ degrees of freedom (N is the total number of observations). In this case N=45

F Test p-value

To get a p-value we compare our F statistic to an $F(2, 42)$ distribution. In our example, The p-value is

$$F = \frac{.0047/2}{.4664/42} = .211$$

The p-value is $P(F(2, 42) > .211) = 0.81$.



ANOVA Table I

Results are often displayed using an ANOVA Table

Source of Variation	Sum of Squares	df	Mean Square	F	P-value
Between Groups	.005	2	.002	.211	.811
Within Groups	.466	42	.011		
Total	.470	44			

- The name “analysis of variance” stems from a partitioning of the total variability in the response variable into components that are consistent with a model for the experiment

ANOVA Table II

- The basic single-factor ANOVA model is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

where $\mu =$an overall mean, $\tau_i = i$th treatment effect (τ_i is constant and $\sum_i \tau_i = 0$), $\epsilon_{ij} =$experiemental error, $NID(0, \sigma^2)$ for $i = 1, 2, \cdots, a$ and $j = 1, 2, \cdots, n$ (Balanced design).

Models for the Data

There are two ways to write a model for the data:

$y_{ij} = \mu + \tau_i + \epsilon_{ij}$ is called the **effects model**.

Let $\mu_i = \mu + \tau_i$. Then

$y_{ij} = \mu_i + \epsilon_{ij}$ is called the **mean model**.

Regression models can also be employed.

Notations for ANOVA I

- Total variability is measured by the total sum of squares:

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2.$$

Notations for ANOVA II

- The basic ANOVA partitioning is:

$$\begin{aligned}
\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\cdot\cdot})^2 &= \sum_{i=1}^{a}\sum_{j=1}^{n}\{(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})+(y_{ij}-\bar{y}_{i\cdot})\}^2 \\
&= n\sum_{i=1}^{a}(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2, \\
SS_T &= SS_{Treatment}+SS_E \\
\text{or } SS_T &= SS_B+SS_E
\end{aligned}$$

- A large value of $SS_{Treatments}$ reflects large differences in treatment means.

Notations for ANOVA III

- A small value of $SS_{Treatments}$ likely indicates no differences in treatment means.
- Formal statistical hypotheses are:

$$
\begin{aligned}
H_0 &: \mu_1 = \mu_2 = \cdots = \mu_a & H_0 &: \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\
H_1 &: \text{ at least one ``$=$'' does not hold.} & H_1 &: \text{at least one is not } 0
\end{aligned}
$$

- While sums of squares cannot be directly compared to test the hypothesis of equal means, mean squares can be compared.

Notations for ANOVA IV

- A mean square is a sum of squares divided by its degrees of freedom:

$$
\begin{aligned}
& df_{Total} = df_{Treatments} + df_{Error}, \\
& an - 1 = a - 1 + a(n-1), \\
& MS_{Treatment} = \frac{SS_{Treatment}}{a-1}, \quad MS_E = \frac{SS_E}{a(n-1)}.
\end{aligned}
$$

- If the treatment means are equal, $MS_{Teatments} = 0$.

The Analysis of Variance is Summarized in a Table I

TABLE 3.3

The Analysis of Variance Table for the Single-Factor, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between treatments	$SS_{\text{Treatments}} = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$	$a-1$	$MS_{\text{Treatments}}$	$F_0 = \dfrac{MS_{\text{Treatments}}}{MS_E}$
Error (within treatments)	$SS_E = SS_T - SS_{\text{Treatments}}$	$N-a$	MS_E	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	$N-1$		

- The reference distribution for F_0 is the $F_{a-1,a(n-1)}$ distribution
- Reject the null hypothesis (equal treatment means) if $F_0 > F_{\alpha,a-1,a(n-1)}$.

The Analysis of Variance is Summarized in a Table II

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N},$$

$$SS_{Treatments} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N},$$

$$SS_E = SS_T - SS_{Treatments}$$

Some words for coding the observations

- If every observation subtracts the same constant, then sums of squares do not change, so we can get the same conclusion.
- If we multiply each observation by the same constant, then the sums of squares change. But the F ratio is equal to the F ratio for the original data. It implies that we can still get the same conclusion.

Parameter estimation I

- Estimates for parameters:

$$\hat{\mu} = \bar{y}_{\cdot\cdot}, \quad \hat{\tau}_i = (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}),$$
$$\hat{\epsilon}_{ij} = y_{ij} - \bar{y}_{i\cdot} \quad \text{(residual)}$$

  So $y_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\epsilon}_{ij}$.

- So estimator of mean of the ith treatment is:

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i\cdot}$$

Parameter estimation II

- And the $100(1-\alpha)\%$ confidence interval for the ith treatment mean:

$$\bar{y}_{i\cdot} - t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_{i\cdot} + t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}}$$

- C.I. for the difference of two treatment means:

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} - t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}} \le \mu_i - \mu_j \le \bar{y}_{i\cdot} - \bar{y}_{j\cdot} - t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}}$$

Model for unbalanced experiment I

- More general model for unbalanced experiment:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \text{ for } i = 1, 2, \cdots, a; j = 1, 2, \cdots, n_i,$$

where $\sum_{i=1}^{a} n_i \tau_i = 0$.

- Notation:

$$y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij} \;\Rightarrow\; \bar{y}_{i\cdot} = y_{i\cdot}/n_i \text{ (treatment sample mean, or row mean}$$

$$y_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij} \;\Rightarrow\; \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/N \text{ (grand sample mean)}$$

Model for unbalanced experiment II

- Decomposition of y_{ij}: $y_{ij} = \bar{y}_{\cdot\cdot} + (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{ij} - \bar{y}_{i\cdot})$
- Estimates for parameters:

$$\begin{aligned} \hat{\mu} &= \bar{y}_{\cdot\cdot}, \\ \hat{\tau}_i &= (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}), \\ \hat{\epsilon}_{ij} &= y_{ij} - \bar{y}_{i\cdot}, \ \text{(residual)} \end{aligned}$$

  So $y_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\epsilon}_{ij}$.
- It can be verified that

$$\sum_{i=1}^{a} n_i \hat{\tau}_i = 0, \quad \sum_{j=1}^{n_i} \hat{\epsilon}_{ij} = 0, \ \text{for all } i$$

Example: Tensile Strength

Investigate the tensile strength of a new synthetic fiber. The factor is the weight percent of cotton used in the blend of the materials for the fiber and it has five levels.

Percent of cotton	Tensile Strength 1	2	3	4	5	Total	Average
15	7	7	11	15	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8

See Tensile.SAS.