

Statistical Design of Experiments-Part III Analysis of Variance

Joseph J. Nahas



UNIVERSITY OF
NOTRE DAME

© Joseph J. Nahas 2012

10 Dec 2012 1

References

- NIST ESH [1.3.5.4](#) One-Factor ANOVA and [1.3.5.5](#) Multi-factor Analysis of Variance
- NIST ESH [7.4.3](#) Are the means equal? and subsections
 - [7.4.3.1](#) 1-Way ANOVA overview
 - [7.4.3.2](#) The 1-way ANOVA model and assumptions
 - [7.4.3.3](#) The ANOVA table and tests of hypotheses about means
 - [7.4.3.4](#) 1-Way ANOVA calculations
 - [7.4.3.5](#) Confidence intervals for the difference of treatment means
 - [7.4.3.6](#) Assessing the response from any factor combination
 - [7.4.3.7](#) The two-way ANOVA
 - [7.4.3.8](#) Models and calculations for the two-way ANOVA
- NIST ESH [7.4.4](#) What are variance components?



UNIVERSITY OF
NOTRE DAME

10 Dec 2012
© Joseph J. Nahas 2012

2

Outline

- **Analysis of Variance (ANOVA) Basics**
- 2 Way ANOVA
- 2 Way ANOVA Example
- Significance Testing using the F distribution



UNIVERSITY OF
NOTRE DAME

10 Dec 2012
© Joseph J. Nahas 2012

3

Analysis of Variance

- In an Analysis of Variance, the variation in the response measurements is partitioned into components that correspond to different sources of variation.
- The primary components are:
 - Input factors
 - Random variation



UNIVERSITY OF
NOTRE DAME

10 Dec 2012
© Joseph J. Nahas 2012

4

Sum of Squares

- The variance of n measurements is given by

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}$$

- Analysis of Variance concentrates on the numerator of the variance calculation.
 - This term is called the Total Sum of Squares or SS(Total)
- The numerator or SS(Total) is factored into component parts.

Single Factor Experiment

- For a single factor experiment with k treatments and n_i samples at each level:

$$SS(\text{Total}) = SST + SSE$$

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

where

- SS(Total) is the Total Sum of Squares
- SST is the Treatment Sum of Squares
- SSE is the Error Sum of Squares
- k is the number of treatments
- n_i is the number of observations at each treatment level.
- A treatment is a specific combination of factor levels whose effect is to be compared to other treatments.

Outline

- Analysis of Variance (ANOVA) Basics
 - 2 Way ANOVA**
- 2 Way ANOVA Example
- Significance Testing using the F distribution

Two Factor Model

- Two Factor Analysis of Variance
 - Two factors A and B
 - A has levels 1, 2, ... a
 - B has levels 1, 2, ... b
- Model each sample data point, Y_{ijk} as made up of four parts:
 - The mean value: μ
 - The response to factor A: α_i
 - The response to factor B: β_j
 - An error term associated with that sample: ϵ_{ijk}

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

$$i = 1, 2, \dots, a; j = 1, 2, \dots, b; k = 1, 2, \dots, r$$

The Sample Data Structure

		Factor B			
		1	2	...	b
Factor A	1	$Y_{111}, Y_{112}, \dots, Y_{11r}$	$Y_{121}, Y_{122}, \dots, Y_{12r}$...	$Y_{1b1}, Y_{1b2}, \dots, Y_{1br}$
	2	$Y_{211}, Y_{212}, \dots, Y_{21r}$	$Y_{221}, Y_{222}, \dots, Y_{22r}$...	$Y_{2b1}, Y_{2b2}, \dots, Y_{2br}$

	a	$Y_{a11}, Y_{a12}, \dots, Y_{a1r}$	$Y_{a21}, Y_{a22}, \dots, Y_{a2r}$...	$Y_{ab1}, Y_{ab2}, \dots, Y_{abr}$

The Model (cont.)

- We can define the deviations from the mean so that they sum to zero.

$$\sum_{i=1}^a \alpha_i = 0$$

$$\sum_{j=1}^b \beta_j = 0$$

- We can estimate the value of α_i from the average over the sample containing that factor:

$$\hat{\alpha}_i = \frac{\sum_{k=1}^r \sum_{j=1}^b y_{ijk}}{rb} - \bar{y} = \bar{y}_i - \bar{y}$$

Contributions to the Total Variation

- Because the deviations are defined around the mean, the variation in the sampled data from the global mean can be made up of three parts:
 - The variation due to Factor A,
 - The variation due to Factor B,
 - The variation due to noise, i.e. the error.
- We can break the variation down into its component parts by looking at a sum of the squares:

$$SS(Total) = SS(A) + SS(B) + SSE$$

where

- SS(A) is the weighted sum of squares for factor A,
- SS(B) is the weighted sum of squares for factor B, and
- SSE is the sum of squares for the error

Sum of Square Formulation

- The four components of the Sum of Squares are:

$$SS(A) = rb \sum_{i=1}^a \hat{\alpha}_i = rb \sum_{i=1}^a (\bar{y}_i - \bar{y})^2 \quad \text{where } rb \text{ is the number of samples in } y_i$$

$$SS(B) = ra \sum_{j=1}^b \hat{\beta}_j = ra \sum_{j=1}^b (\bar{y}_j - \bar{y})^2 \quad \text{where } ra \text{ is the number of samples in } y_j$$

$$SSE = \sum_{k=1}^r \sum_{j=1}^b \sum_{i=1}^a (y_{ijk} - \bar{y}_{ij})^2$$

$$SS(Total) = \sum_{k=1}^r \sum_{j=1}^b \sum_{i=1}^a (y_{ijk} - \bar{y})^2$$

Note the similarities in the formulations

Expansion of a Variance Calculation

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

$$= \frac{\sum_{i=1}^n \left(y_i - \frac{\sum_{j=1}^n y_j}{n} \right)^2}{n-1}$$

Raw Sum of Squares
RSS

Corrected
Sum of Squares
CSS

Correction
for the mean
CM

Degrees of Freedom
DF

Calculating SS

$$CM = \frac{(\text{SumOfAllObservations})^2}{rab} = \frac{\left(\sum_{k=1}^r \sum_{j=1}^b \sum_{i=1}^a y_{ijk} \right)^2}{rab}$$

$$SS_{total} = \sum (\text{EachObservation})^2 - CM$$

$$SS(A) = \frac{\sum_{i=1}^a A_i^2}{rb} - CM \quad \text{where} \quad A_i = \sum_{k=1}^r \sum_{j=1}^b y_{ijk}$$

$$SS(B) = \frac{\sum_{j=1}^b B_j^2}{ra} - CM \quad \text{where} \quad B_j = \sum_{k=1}^r \sum_{i=1}^a y_{ijk}$$

$$SSE = SS_{total} - SS(A) - SS(B)$$

ANOVA Table

Source	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square Variation (MS)
Factor A	SS(A)	(a-1)	SS(A)/(a-1)
Factor B	SS(B)	(b-1)	SS(B)/(b-1)
Error	SSE	(N-a-b-1)	SSE/(N-a-b-1)
Total	SS(Total)	(N-1)	

↑
Relative contribution
to the total variation

Outline

- Analysis of Variance (ANOVA) Basics
- 2 Way ANOVA
- **2 Way ANOVA Example**
- Significance Testing using the F distribution

An Example

		Materials (B)		
		1	2	3
Labs (A)	1	4.1	3.1	3.5
		3.9	2.8	3.2
		4.3	3.3	3.6
Labs (A)	2	2.7	1.9	2.7
		3.1	2.2	2.3
		2.6	2.3	2.5

Summations

r	3
a	2
b	3
r a b	18

Summations in each box on previous slide

		Materials (B) Sum			Total (Ai)
		1	2	3	
Lab (A) Sum	1	12.3	9.2	10.3	31.8
	2	8.4	6.4	7.5	22.3
Total(Bj)		20.7	15.6	17.8	54.1

Summation over rows

Summation over Columns

Total Sum

Summations

CM	162.6
SS(Total)	7.93
SS(A)	5.0139
SS(B)	2.1811
SSE	0.7344

Calculations based on Slide 14

ANOVA Table

Source	SS	df	MS
A	5.0139	1	5.0139
B	2.1811	2	1.0906
Error	0.7344	12	0.0612
Total	7.93	17	

Outline

- Analysis of Variance (ANOVA) Basics
- 2 Way ANOVA
- 2 Way ANOVA Example
- Significance Testing using the F distribution

Significance of the Relative Contribution

- Are the various factor contributions significant relative to the noise in the measurement.
- Look at variance of factor relative to variance of the noise.

ANOVA Table

Source	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square Variation (MS)	F Statistic	F Reference
Factor A	SS(A)	(a-1)	SS(A)/(a-1)	MS(A)/MSE	$F_{0.05, a-1, N-a-b-1}$
Factor B	SS(B)	(b-1)	SS(B)/(b-1)	MS(B)/MSE	$F_{0.05, b-1, N-a-b-1}$
Error	SSE	(N-a-b-1)	SSE/(N-a-b-1)		
Total	SS (Total)	(N-1)			

Variance of Factor relative to Variance of Error

F Distribution Comparing Factor Variance to Error Variance

Expanded ANOVA Table

Source	SS	df	MS	Fstat	Fref	
A	5.0139	1	5.0139	81.92	4.75	Significant
B	2.1811	2	1.0906	17.82	3.89	Significant
Error	0.7344	12	0.0612			
Total	7.93	17				
Alpha = 0.05						

In Excel: Fref = FINV(Alpha, a-1, N-a-b-1)
 = FINV(0.05, 2-1, 18-2-3-1)
 = FINV(0.05, 1, 12)
 = 4.75