Lecture 9

9.1. Analysis of Variance

Summary

In this lecture we showed how analysis of variance can be used to test for differences among means of several populations or treatments. We introduced the completely randomized design, the randomized block design, and the two-factor factorial experiment. The completely randomized design and the randomized block design are used to draw conclusions about differences in the means of a single factor. The primary purpose of blocking in the randomized block design is to remove extraneous sources of variation from the error term. Such blocking provides a better estimate of the true error variance and a better test to determine whether the population or treatment means of the factor differ significantly.

We showed that the basis for the statistical tests used in analysis of variance and experimental design is the development of two independent estimates of the population variance σ^2. In the single-factor case, one estimator is based on the variation between the treatments; this estimator provides an unbiased estimate of σ^2 only if the means μ_1, μ_2,.. μ_k are all equal. A second estimator of σ^2 is based on the variation of the observations within each sample; this estimator will always provide an unbiased estimate of σ^2. By computing the ratio of these two estimators (the F statistic) we developed a rejection rule for determining whether to reject the null hypothesis that the population or treatment means are equal. In all the experimental designs considered, the partitioning of the sum of squares and degrees of freedom into their various sources enabled us to compute the appropriate values for the analysis of variance calculations and tests. We also showed how Fisher's LSD procedure and the Bonferroni adjustment can be used to perform pairwise comparisons to determine which means are different.

Glossary

ANOVA table A table used to summarize the analysis of variance computations and results It contains columns showing the source of variation, the sum of squares, the degrees of freedom, the mean square, and the *F* value(s).

Partitioning The process of allocating the total sum of squares and degrees of freedom to the various components.

Multiple comparison procedures Statistical procedures that can be used to conduct statistical comparisons between pairs of population means.

Comparisonwise Type I error rate The probability of a Type I error associated with a single pairwise comparison.

Experimentwise Type I error rate The probability of making a Type I error on at least one of several pairwise comparisons.

Factor Another word for the independent variable of interest.

Treatments Different levels of a factor.

Single-factor experiment An experiment involving only one factor with *k* populations or treatments.

Experimental units The objects of interest in the experiment.

Completely randomized design An experimental design in which the treatments are randomly assigned to the experimental units.

Blocking The process of using the same or similar experimental units for all treatments. The purpose of blocking is to remove a source of variation from the error term and hence provide a more powerful test for a difference in population or treatment means.

Randomized block design An experimental design employing blocking.

Factorial experiment An experimental design that allows statistical conclusions about two or more factors.

Replications The number of times each experimental condition is repeated in an experiment.

Interaction The effect produced when the levels of one factor interact with the levels of another factor in influencing the response variable.

Key formulas

Testing for the Equality of *k* Population Means

Sample Mean for Treatment *j*

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \tag{9.1}$$

Sample Variance for Treatment *j*

$$s_j^2 = \frac{\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}_j\right)^2}{n_j - 1} \tag{9.2}$$

Overall Sample Mean

$$\bar{x} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n_r} \tag{9.3}$$

$$n_T = n_1 + n_2 + \cdots + n_k \tag{9.4}$$

Mean Square Due to Treatments

$$MSTR = \frac{SSTR}{k-1} \tag{9.5}$$

Sum of Squares Due to Treatments

$$SSTR = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2 \tag{9.6}$$

Mean Square Due to Error

$$MSE = \frac{SSE}{n_r - k} \tag{9.7}$$

Sum of Squares One to Error

$$SSE = \sum_{j=1}^{k}(n_j - 1)s_j^2 \tag{9.8}$$

Test Statistic for the Equality of it Population Means

$$F = \frac{MSTR}{MSE} \tag{9.9}$$

Total Sum of Squares

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}\right)^2 \tag{9.10}$$

Partitioning of Sum of Squares

$$\text{SST} = \text{SSTR} + \text{SSE} \tag{9.11}$$

Multiple Comparison Procedures

Test Statistic for Fisher's LSD Procedure

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \tag{9.12}$$

Fisher's LSD

$$LSD = t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{9.13}$$

Completely Randomized Designs

Mean Square Due to Treatments

$$MSTR = \frac{\sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2}{k-1} \tag{9.14}$$

Mean Square Due to Error

$$MSE = \frac{\sum_{j=1}^{k} (n_j - 1) s_j^2}{n_T - k} \tag{9.15}$$

F Test Statistic

$$F = \frac{MSTR}{MSE} \tag{9.9}$$

Randomized Block Designs

Total Sum of Squares

$$SST = \sum_{i=1}^{b} \sum_{j=1}^{k} \left(x_{ij} - \bar{\bar{x}}\right)^2 \tag{9.16}$$

Sum of Squares Due to Treatments

$$SSTR = b \sum_{j=1}^{k} \left(\bar{x}_j - \bar{\bar{x}}\right)^2 \tag{9.17}$$

Sum of Squares Due to Blocks

$$SSBL = k \sum_{i=1}^{b} \left(\bar{x}_i - \bar{\bar{x}}\right)^2 \tag{9.18}$$

Sum of Squares Due to Error

$$SSE = SST - SSTR - SSBL \tag{9.19}$$

Factorial Experiments

Total Sum of Squares

$$SST = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} \left(x_{ijk} - \bar{\bar{x}}\right)^2 \tag{9.20}$$

Sum of Squares for Factor A

$$SSA = br \sum_{i=1}^{a} \left(\bar{x}_i - \bar{\bar{x}}\right)^2 \tag{9.21}$$

Sum of Squares for Factor B

$$SSB = ar \sum_{j=1}^{b} \left(\bar{x}_j - \bar{\bar{x}}\right)^2 \tag{9.22}$$

Sum of Squares for Interaction

$$SSAB = r \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}}\right)^2 \tag{9.23}$$

Sum of Squares for Error

$$SSE = SST - SSA - SSB - SSAB \tag{9.24}$$

Lecture supplementary material

Table. Anova table for a completely randomized design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Treatments	SSTR	k-1	$MSTR = \frac{SSTR}{k-1}$	$\frac{MSTR}{MSE}$
Error	SSE	n_τ-k	$MSE = \frac{SSE}{n_\tau - k}$	
Total	SST	n_τ-1		