STA 332
Laboratory/Fieldwork on Experimental Design I

Ibadan Distance Learning Centre Series

STA 332
Laboratory/Fieldwork on Experimental Design I

By
Christopher Godwin Udomboso
Department of Statistics
University of Ibadan

Published by
Distance Learning Centre
University of Ibadan

First Published 2011

General Editor: *Prof. Bayo Okunade*
Series Editor: *Mrs. Temitope A. Omoloye*

Typesetted @ Distance Learning Centre, University of Ibadan

Table of Content

General Introduction

Analysis of variance (ANOVA) represents a set of models that can be fit to data, and also a set of methods that can be used to summarize an existing fitted model.

Analysis of variance is particularly an effective tool for analyzing highly structured experimental data (in agriculture, multiple treatments applied to different batches of animals or crops; in psychology, multi-factorial experiments manipulating several independent experimental conditions and applied to groups of people; industrial experiments in which multiple factors can be altered at different times and in different locations).

At the end of this course, students should be able to

1. design a simple layout for experimental data.
2. solve simple analysis of variance data.
3. conduct simple statistical tests and carry out inferences on experimental data.
4. carry out simple research problems involving experimental data.

LECTURE ONE

Overview of Experimentation

Introduction

Analysis of variance (ANOVA) represents a set of models that can be fit to data, and also a set of methods that can be used to summarize an existing fitted model. We shall first consider ANOVA as it applies to classical linear models (the context for which it was originally devised; Fisher, 1925) and then discuss how ANOVA has been extended to generalized linear models and multilevel models. Analysis of variance is particularly effective tool for analyzing highly structured experimental data (in agriculture, multiple treatments applied to different batches of animals or crops; in psychology, multi-factorial experiments manipulating several independent experimental conditions and applied to groups of people; industrial experiments in which multiple factors can be altered at different times and in different locations).

Objectives

At the end of this lecture, you should be able to

1. have a knowledge of experimentation.
2. learn about the basic principles of experimentation.
3. understand what the completely randomized design (CRD) is all about.

Pre Test

1. What is experimentation?
2. What are the principles of experimentation?

CONTENT

In statistics, analysis of variance (ANOVA) is a collection of statistical models, and their associated procedures, in which the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form ANOVA provides a statistical test of whether or not the means of several groups are all equal, and therefore generalizes *t*-test to more than two groups. Doing multiple two-sample t-tests would result in an increased chance of committing a type I error. For this reason, ANOVAs are useful in comparing two, three or more means.

Basic Principles

The basic principles of experimental designs are randomization, replication and local control. These principles make a valid test of significance possible. Each of them is described below:

1. **Randomization**: The first principle of an experimental design is randomization, which is a random process of assigning treatments of the experimental units. The random process implies that every possible allotment of treatment has the same probability. An experimental unit is the smallest division of the experimental condition whose effect is to be measured and compared. The purpose of randomization is to remove bias and other sources of extraneous variation or external factors, which are not controllable. Another advantage of randomization (accompanied by replication) is that it forms the basis of any valid statistical test. Hence the treatment. Must be assigned at random to the experimental units. Randomization is usually done by drawing numbered cards from a well-shuffled pack of cards; or by drawing numbered balls from a well-shaken container or by using tables of random numbers.

2. **Replication**: The second principle of an experimental design is replication; which is a repetition of the basic experiment. In other words, it is a complete run for all the treatments to be tested in the experiment. In all experiments, some variation is introduced because of the fact that the experimental units such as individuals or plots of land in agricultural experiments cannot be physically identical. This type of variation can be removed by using a number

of experimental units. We therefore perform the experiment more than once, i.e. we have repeated the basic experiment. An individual repetition is called a replicate. The number of the shape and the size of replicates depend upon the nature of the experimental material. A replication is used

i. To secure more accurate estimate of the experimental error, a term which represents the difference that would be observed if the same treatments were applied several times to the same
ii. To decrease the experimental error and thereby to increase precision which is a measure of the variability of the experimental error and;
iii. To obtain more precise estimate of the mean effect of a treatment.

3. **Local control**: It has been observed that all experiments source of variation are not removed by randomization and replication. These necessities a refinement in the experimental technique. In other words, we need to choose a design in such a manner that all extraneous sources of variation are brought under control. For this purpose, we make use of local control, a term referring to the amount of balancing, blocking and grouping of the experimental unit in such a way that the result is a balanced arrangement of the treatments. Blocking means that similar experimental units shows be collected together to form a relatively homogeneous group. A block is also a replicate. The main purpose of the principle of local control is to increase the efficiency of an experimental design by decreasing the experimental error. The point to remember here is that the term "local control" should not be confused with the word "control". The word "control" in an experimental design is used for a treatment.

Completely Randomised Design (CRD)

A completely randomized (CR) design, which is the simplest type of the basic designs, may be defined as a design in which the treatments are assigned to the experimental units completely at random, that is the randomization is done without any restrictions. The design is completely flexible. Any number of treatments may be used. Moreover, the number

of units per treatment needs to be equal. A completely randomized design is considered to be more useful in situations where

i. The experimental units are homogeneous
ii. The experiments are small such as laboratory equipments and
iii. Some experimental units are likely to be destroyed or to fail to respond.

Experimental layout

The layout of an experiment is the actual placement of the treatments of the experimental units, which may pertain to time, space or type of material. Suppose we have k treatments and the experimental material is divided into n-experimental units. We shall then assign the k-treatments at random to the n-experimental units in such a way that the treatment $T_j\ (j = 1, 2, 3, \dots, k)$ is applied r_j times, with $\sum r_j = n$. When each treatment is applied the same number of times, then $r_1 = r_2 = \cdots = r_k = r$ and $\sum r_j = r_k = n$. Usually, each treatment is applied (or replicated) and equal number of times.

An example of the experimental layout for a completely randomized design (CRD) using four treatments A, B, C, and D, each replicated 3 times, is shown below:

Table 1.1

C	A	B	D
C	B	C	A
A	D	D	B

Table 1.2

D	D	A	C
C	B	D	B
C	B	A	A

The result or response of a treatment, which may be a real yield, the gain in weight, the gain in weight, the ability, and so on, is generally called yield and is represented by the letter Y.

In a completely randomized design, there is only one primary factor under consideration in the experiment. The test subjects are assigned to treatment levels of the primary factor at random.

Advantages of a CRD

1. Very flexible design (i.e. number of treatments and replicates is only limited by the number of experimental units)
2. Statistical analysis is simple compared to other designs
3. Loss of information due to the large number of degrees of freedom for the error source of variation.

Disadvantages of a CRD

1. if experimental units are not homogeneous and you fail to minimize this variation using blocking, there may be a loss of precision.
2. Usually the least efficient design unless experimental units are homogeneous
3. Not suited for a large number of treatments

Mathematical Model for a CRD

$Y_{ij} = \mu + T_i + \varepsilon_{ij}$

Where Y_{ij} is the j th observation of the i th treatment,

μ is the population/grand mean

T_i is the treatment effect of the i th treatment

and ε_{ij} is the random error $\sim NID(0, \sigma^2)$

Example 1.1

Given the following data

REPLICATE	**TREATMENTS**		
1	*A*	*B*	*C*
2	23	42	47
3	31	47	43
4	33	34	39

Test for the significance of the treatment effects at $\alpha = 5\%$

Solution

The steps to follow are

1. Write hypotheses to be tested
2. Calculate correction factor
3. Calculate total sum of squares (SST)
4. Calculate treatment sum of squares (SSt)
5. Calculate error sum of squares (SSe)
6. Complete the ANOVA table
7. Look up to the F -table or use the p -value
8. Calculate coefficient of variation

ANOVA TABLE (SKETCH) FOR CRD

Source of variation	df	SS	MS	F
Treatment	$t-1$	SSt	$\frac{SSt}{t-1} = A$	$\frac{A}{B}$
Error	$t(r-1)$	SSe	$\frac{SSt}{t(r-1)} = B$	
Total	$tr-1$	SST		

Solution

Hypothesis

H_0 : treatment effects are not significant or H_0: $\mu_1 = \mu_2 = \mu_3$

H_1 : treatment effects are significant or H_1: $\mu_1 \neq \mu_2 \neq \mu_3$

or H_1: $\mu_i \neq \mu$ for at least one i.

Replicate	A	B	C	Y_j
1	23	42	47	112
2	36	26	43	105
3	31	47	43	121
4	33	34	39	106
Y_i	123	149	172	144

$$CF = \frac{Y^2}{N} = \frac{442^2}{12} = 16428$$

$$SST = \Sigma\Sigma\, yy^2 \quad CF$$

$$= 23^2 + 42^2 + 47^2 + 36^2 + 26^2 + 43^2 + 31^2 + 47^2 + 43^2 + 33^2 + 34^2 + 39^2 - 16428$$

$$= 17108 - 16428 = 680$$

$$SSt = \Sigma \frac{Yi^2}{r} \quad CF$$

$$= \frac{123^2 + 149^2 + 172^2}{4} \quad CF$$

$$= 16278.5 - 16428$$

$$= 300.5$$

$$SSe = SST - SSt$$

$$= 680 - 300.5$$

$$= 379.5$$

ANOVA TABLE (CRD) for the problem

Source of variation	df	SS	MS	F
Treatment	2	300.5	150.25	3.563
Error	9	379.5	42.167	
Total	11	680.0		

F_{tab} = F(v_1, v_2, α) = F(2, 9, 5%) = 4.26

Decision rule: Reject H_0 if $F_{cal} > F_{tab}$, otherwise accept H_0

Conclusion: H_0 cannot be rejected at 5% level of significance because $F_{tab} > F_{cal.}$. In other words, the treatment performs the same way.

Example 1.2

Given the following data

REPLICATE	**TREATMENT**			
1	2.0	1.7	2.0	2.1
2	2.2	1.9	2.4	2.2
3	1.8	1.5	2.7	2.2
4	2.3		2.5	1.9
5	1.7		2.4	

Test for the significance of the treatment

Summary

In this lecture, you have been able to

1. understand the concept of experimentation.
2. learn about the basic principles of experimentation.
3. learn about the concept of the completely randomized design (CRD).
4. learn how to develop a simple CRD layout.
5. solve some simple problems involving a CRD.

Post Test

1. What is experimentation?
2. Discuss the basic principles of experimentation.
3. What is a completely randomized design (CRD)?
4. What are the advantages and disadvantages of the CRD?
5. In a study, subjects are randomly assigned to one of three groups: ***control***, ***experiment*** A, or ***experiment*** B. After treatment, achievement test scores for the three groups are compared. What is the appropriate statistical test for this comparison?
6. Dr. Martha Bergen studied attitudes toward computer-enhanced learning for seminary education among full-time professors at Southwestern Baptist Theological Seminary in 1989.14 One of her hypotheses was that there would be a "significant difference [in attitude toward computer-enhanced learning] between the professors in the religious education, theology, and church music schools." Scores were generated from an attitude scale Dr. Bergen developed for the study. The mean attitude scores for the three schools were 118 (highest) in the Religious Education faculty, 117 in the church music faculty, and 114 (lowest) in the theology faculty. But were these differences in attitude significant? Here is the ANOVA table she generated:

SOURCE OF VARIATION	SUM OF SQUARES	df	MS	F	p
Between	323.387	2	161.694	.472	.626
Within	25018.652	73	342.721		
Total	25342.039	75			

Using the problem and printout above, answer these questions:

a. Is the F -ratio significant? Explain why you say this.

b. Explain this F -ratio in terms of the three group means: 114, 117, 118.

c. How do you explain the differences in the school mean scores?

d. Dr. Bergen did not apply multiple comparisons tests to see if any single mean was significantly different from the others. Why? Was she correct in doing so?

7. A report on an investigation includes the following information related to the influence of several mouth washes to length of time that breath remains "great".

Analysis of Variance

Source	*df*	*SS*	*MS*
Treatments	3	500	166.67
Error	196	19,600	100
Total	199	20,100	

Means (hours that breath remained "great")

Whiskey	12
Brand X	11
Water	9
Brand L	8

F(critical) (.05) = 3.92 with 196 df

You are to develop plans for a follow-up study. In particular, you are to re-examine the difference between Brand X and Brand L. In looking at methods for estimating number of replicates needed you find that you need values for -

1. the size of the difference to be detected
2. the anticipated standard deviation
3. the anticipated variance

On the basis of the above report what values will you use for each of the above 3 items? Why?

8. An investigator randomly assigns 30 college students into three equal size study groups (early morning, afternoon, late night) to determine if the period of the day at which people study has an

effect on their retention. The students live in a controlled environment for one week, on the third day of which the experimental treatment (study of predetermined material) is administered. The seventh day the investigator tests for retention, and in computing his analysis he sees that his MS within groups is larger than his MS among groups. What is the indication of this result?

a. An error in calculation was made.
b. There was more than the expected variability between groups.
c. There was more variability between subjects within the same group than there was between groups.
d. That there should have been additional controls in the experiment.

References

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. Biometrics 3, 1–21.

Fisher, R. A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

Gelman, A. (2005). Analysis of variance: why it is more important than ever (with discussion). Annals of Statistics 33, 1–53.

Yates, F. (1967). A fresh look at the basic principles of the design and analysis of experiments. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 4, 777–790.

LECTURE TWO

The Basic Principles of Analysis of Variance

Introduction

The aim of this lecture is to look at an example in some detail. This will be done by actually working through the numerical mechanics, and relating it to the output. Once the origin of the output has been derived from first principles, it will not be necessary to do this again. This section will also provide you with your first introduction to model formulae.

Objectives

At the end of this lecture, you should be able to

1. explain the concept of ***analysis of variance*** (ANOVA).
2. understanding the concepts of partitioning the variability.
3. design a simple ANOVA table.
4. solve simple problems.

Pre Test

1. What do you understand by ***analysis of variance*** (ANOVA)?
2. What sort of problems does ANOVA solve?
3. What is the meaning of the term, ***degree of freedom***?

CONTENT

One-way within-subject ANOVA

In this model we have k measurement per subject. The treatment effects for subject $n = 1...N$ are measured relative to the average response made

by subject n on all treatments. The *k*th response from the *n*th subject is modeled as

$$y_{nk} = \tau_k + \pi_n + e_{nk} \qquad (2.1)$$

Where τ_k are the treatment effects (or *within-subject effects*), π_n are the *subject effect* and e_{nk} are the residual errors. We are not normally interested in π_n, but its explicit modeling allows us to remove variability due to differences in average responsiveness of each subject. It is also possible to express the full model in terms of differences between treatments.

To test whether the experimental factor is significant we compare the full model in equation 11 with the reduced model.

$$y_{nk} = \pi_n + e_{nk} \qquad (2.2)$$

A Simple Example

If we have three fertilizers, and we wish to compare their efficacy, this could be done by a field experiment in which each fertilizer is applied to 10 plots, and then the 30 plots are later harvested, with the crop yield being calculated for each plot. We now have three groups of ten figures, and we wish to know if there are any differences between these groups. The data were recorded in the fertilizers dataset as shown in Table 2.1.

When these data are plotted on a graph, it appears that the fertilizers do differ in the amount of yield produced, but there is also a lot of variation between plots given the same fertilizer. Whilst it appears that fertilizer 1 produces the highest yield on average, a number of plots treated with fertilizer 1 did actually yield less than some of the plots treated with fertilizers 2 or 3.

We now need to compare these three groups to discover if this apparent difference is statistically significant. When comparing two samples, the first step was to compute the difference between the two sample means. However, because we have more than two samples, we do not compute the differences between the group means directly. Instead, we focus on the variability in the data. At first this seems slightly counter-intuitive: we are going to ask questions about the means of three groups by analyzing the variation in the data. How does this work?

Table 2.1 Raw data from the fertilizers dataset

Fertiliser	Yields (in tonnes) from the 10 plots allocated to that fertiliser
1	6.27, 5.36, 6.39, 4.85, 5.99, 7.14, 5.08, 4.07, 4.35, 4.95
2	3.07, 3.29, 4.04, 4.19, 3.41, 3.75, 4.87, 3.94, 6.28, 3.15
3	4.04, 3.79, 4.56, 4.55, 4.53, 3.53, 3.71, 7.00, 4.61, 4.55

What happens when we calculate a variance?

The variability in a set of data quantifies the scatter of the data points around the mean. To calculate a variance, first the mean is calculated, then the deviation of each point from the mean. Deviations will be both positive and negative; and the sum will be zero. (This follows directly from how the mean was calculated in the first place). This will be true regardless of the size of the dataset, or amount of variability within a dataset, and so the raw deviations are not useful as a measure of variability. If the deviations are squared before summation then this sum is a useful measure of variability, which will increase the greater the scatter of the data points around the mean. This quantity is referred to as a sum of squares (SS), and is central to our analysis.

The SS however cannot be used as a comparative measure between groups, because clearly it will be influenced by the number of data points in the group; the more data points, the greater the SS. Instead, this quantity is converted to a variance by dividing by $n-1$, where n equals the number of data points in the group. A variance is therefore a measure of variability, taking account of the size of the dataset.

Why use $n-1$ rather than n ?

If we wish to calculate the average squared deviation from the mean (i.e. the variance) why not divide by n ? The reason is that we do not actually have n independent pieces of information about the variance. The first step was to calculate a mean (from the n independent pieces of data collected). The second step is to calculate a variance with reference to that mean. If $n-1$ deviations are calculated, it is known what the final deviation must be, for they must all add up to zero by definition. So we have only $n-1$ independent pieces of information on the variability

about the mean. Consequently, you can see that it makes more sense to divide the SS by $n-1$ than n to obtain an average squared deviation around the mean. The number of independent pieces of information contributing to a statistic is referred to as the degrees of freedom.

PARTITIONING

Partitioning the variability

In an ANOVA, it is useful to keep the measure of variability in its two components; that is, a sum of squares, and the degrees of freedom associated with the sum of squares. Returning to the original question: what is causing the variation in yield between the 30 plots of the experiment? Numerous factors are likely to be involved: e.g. differences in soil nutrients between the plots, differences in moisture content, many other biotic and abiotic factors, and also the fertilizer applied to the plot. It is only the last of these that we are interested in, so we will divide the variability between plots into two parts: that due to applying different fertilizers, and that due to all the other factors. To illustrate the principle behind partitioning the variability, first consider two extreme datasets. The first step would be to calculate a grand mean, and there is considerable variation around this mean. The second step is to calculate the three group means that we wish to compare: that is, the means for the plots given fertilizers A , B and C . It can be seen that once these means are fitted, little variation is left around the group means. In other words, fitting the group means has removed or explained nearly all the variability in the data. This has happened because the three means are distinct.

Now consider the other extreme, in which the three fertilizers are, in fact, identical. Once again, the first step is to fit a grand mean and calculates the sum of squares. Second, three group means are fitted, only to find that there is almost as much variability as before. Little variability has been explained. This has happened because the three means are relatively close to each other (compared to the scatter of the data).

The amount of variability that has been explained can be quantified directly by measuring the scatter of the treatment means around the grand mean. In the first of the two examples, the deviations of the group means around the grand mean are considerable, whereas in the second example these deviations are relatively small. The dataset given in Table 2.1 represents an intermediate situation in which it is not immediately obvious

if the fertilizers have had an influence on yield. When the three group means are fitted, there is an obvious reduction in variability around the three means (compared to the one mean). But at what point do we decide that the amount of variation explained by fitting the three means is significant? The word significant, in this context, actually has a technical meaning. It means 'When is the variability between the group means greater than that we would expect by chance alone?'

At this point it is useful to define the three measures of variability that have been referred to. These are:

SST = Total sum of squares.

Sum of squares of the deviations of the data around the grand mean.

This is a measure of the total variability in the dataset.

SSe = Error sum of squares.

Sum of squares of the deviations of the data around the three separate group means. This is a measure of the variation between plots that have been given the same fertiliser.

SSt = Fertiliser sum of squares.

Sum of squares of the deviations of the group means from the grand mean.

This is a measure of the variation between plots given different fertilisers.

Variability is measured in terms of sums of squares rather than variances because these three quantities have the simple relationship:

$$SST = SSt + SSe$$

So the total variability has been divided into two components; that due to differences between plots given different treatments, and that due to differences between plots given the same treatment. Variability must be due to one or other of these two causes. Separating the total SS into its component SS is referred to as partitioning the sums of squares.

A comparison of SSt and SSe is going to indicate whether fitting the three fertilizers means accounts for a significant amount of variability in the data. The greater the number of means fitted to the data, the greater SSt would be, because more variance would have been explained. Taken to the limit, if our aim was merely to maximize SSt, we should fit a mean

for every data point, because in that way we could explain all the variability. For a valid comparison between these two sources of variability, we need to compare the variability per degree of freedom, that is, the variances.

Partitioning the degrees of freedom

Every SS was calculated using a number of independent pieces of information. The first step in any analysis of variance is to calculate SST. It has already been discussed that when looking at the deviations of data around a central grand mean, there are $n-1$ independent deviations: that is, in this case $n-1=29$ degrees of freedom (df). The second step is to calculate the three treatment means. When the deviations of two of these treatment means from the grand mean have been calculated, the third is predetermined, as again by definition, the three deviations must sum to zero. Therefore, SSt, which measures the extent to which the group means deviate from the grand mean, has two df associated with it. Finally, SSe measures variation around the three group means. Within each of these groups, the ten deviations must sum to zero. Given nine deviations within the group, the last is predetermined. Thus SSe has $3 \times 9 = n - 3 = 27$ df associated with it. Just as the SS are additive, so are the df.

Mean squares

Combining the information on SS and df, we can arrive at a measure of variability per df. This is equivalent to a variance, and in the context of ANOVA is called a mean square (MS). In summary:

Fertiliser Mean Square $(MSt) = {SSt}/{2}$

The variation (per df) between plots given different fertilizers.

Error Mean Square $(MSe) = {SSe}/{27}$

The variation (per df) between plots given the same fertilizer.

Total Mean Square $(MST) = {SST}/{29}$

The total variance of the dataset.

Unlike the SS, the MS are not additive.

So now the variability per df due to differences between the fertilizers has been partitioned from the variability we would expect due to all other factors. Now we are in the position to ask: by fitting the treatment means, have we explained a significant amount of variance?

F -ratios

If none of the fertilizers influenced yield, then the variation between plots treated with the same fertilizer would be much the same as the variation between plots given different fertilizers. This can be expressed in terms of mean squares: the mean square for fertilizer would be the same as the mean square for error: i.e. the ratio of these two mean squares is the F -ratio, and is the end result of the ANOVA. Even if the fertilizers are identical, it is unlikely to equal exactly 1, it could by chance take a whole range of values. The F distribution represents the range and likelihood of all possible F -ratios under the null hypothesis (i.e. when the fertilizers are identical).

If the three fertilizers were very different, then the MSt would be greater than the MSe, and the F -ratio would be greater than 1. F -ratio can be quite large even when there are no treatment differences. At what point do we decide that the size of the F-ratio is due to treatment differences rather than chance?

Just as with other test statistics, the traditional threshold probability of making a mistake is 0.05. In other words, we accept that the F -ratio is significantly greater than 1 if it will be that large or larger under the null hypothesis only 5% of the time. If we had inside knowledge that the null hypothesis was in fact true, then 5% of the time we would still get an F -ratio that large. When we conduct an experiment, however, we have no such inside knowledge, and we are trying to gather evidence against it. Our p-value is a measure of the strength of evidence against the null hypothesis. Only when it is less than 0.05 do we consider the evidence great enough to accept.

It should be mentioned that the exact F distribution will depend upon the df with which the F -ratio was constructed. In this case, the df are 2 and 27, associated with the numerator and the denominator of the F -ratio respectively. The general shape will vary from a decreasing curve to a

humped distribution, skew to the right. When doing an ANOVA table in most packages the F -ratio, degrees of freedom and the p -value are provided in the output, or occasionally you are left to look up the F -ratio in statistical tables.

Summary

In this lecture, you have been able to

1. learn about the concept of ***analysis of variance*** (ANOVA).
2. know what happens when variance is calculated, and why $n-1$ is used rather than n .
3. understand the concept of partitioning of the ***sum of squares***.
4. learn how to compute degrees of freedom, mean square error and the F ratio.

Post Test

1. What is analysis of variance?
2. What happens when variance is computed?
3. Why do we use $n-1$ rather than n in the computation of variance?
4. If X has an F distribution with $df = 4, 5$, what is the value of c where $P(X < c) = 0.95$?
5. In a simple analysis of variance problem, which of the following is an estimate of the variance of individual measurements (after the various effects have been accounted for)? (MS means SS/df so each of answers is a Mean Square.)
   a. MS(between)
   b. MS(within)
   c. MS(total)
   d. none of the above
6. Given the following observed number of pigs for 8 litters, the numerator of the formula for s**2 is called the corrected sum of squares as illustrated.

s**2 = (Sum of Squares)/(n-1). Find the sum of squares.

X(1) = 9 X(5) = 10
X(2) = 6 X(6) = 7
X(3) = 14 X(7) = 8
X(4) = 9 X(8) = 9

a) 5 b) SQRT(5) c) 40 d) 80

7. An experiment was conducted as a oneway random ANOVA design yielding K sample means, each based on n scores. If the between and within mean squares are represented by S(m)**2 and S(p)**2, respectively, what is the number of degrees of freedom for S(m)**2?
   a. $n-1$
   b. $k-1$
   c. $n-k$
   d. $(n-1)(k-1)$
   e. none of the above

8. In reading a scientific article you encounter the following table:

Analysis of Variance

Source	*SS*	*df*	*MS*	*F*
Between samples	722.7	4	180.68	15.3**
Within samples	473.3	40	11.83	
Total	1196.0	44		

Further reading indicates that all sample sizes are equal. Then we know that the experimenter used

a. 4 samples of size 10.
b. 5 samples of size 10.
c. 4 samples of size 9.
d. 5 samples of size 9.
e. None of these

9. i. What is meant by randomization?

An undesirable effect of some antihistamines is drowsiness, which is a consequence of the effect of the drugs on the central nervous system. These data come from an experiment of Hedges, Hills, Maclay, Newman-Taylor and Turner (1971) to compare the effect on the central nervous system of a placebo and two antihistamines. This was done by measuring the flicker frequency 3 some time after drug administration in four volunteers who have taken the three treatments. The data presented here are scaled measures based on the flicker frequency.

Subject Number	**Meclastine**	**Promethazine**	**Placebo**
1	112	112	131
2	48	37	61
3	106	93	112
4	51	46	70

ii. Plot these data in a meaningful way and comment.

iii. Carry out an appropriate analysis to examine whether there is a difference between the effects of the different drugs, stating clearly your hypotheses, conclusions and any assumptions made.

iv. State two precautions which should have been taken in running this experiment.

v. Give two benefits which would have resulted if more than one measurement for each drug for each subject had been obtained.

(3;4;7;3;3)

[The numbers in the bottom right hand corner of the question are as they appear in the Oxford Prelims exam paper and indicate to the candidate how many marks each part of the question is potentially worth if answered correctly.]

References

Charles McCreery 2007 Analysis of Variance OXFORD FORUM Magdalen College Oxford Psychological Paper No. 2007-3

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. Biometrics 3, 1–21.

Fisher, R. A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

Gelman, A. (2005). Analysis of variance: why it is more important than ever (with discussion). Annals of Statistics 33, 1–53.

Howell, David C. (1997). Statistical Methods for Psychology (4th edition). London: Duxbury Press. Chapters 11-14.

Howell, David C. (1997). Statistical Methods for Psychology (4th edition). London: Duxbury Press.

Kirk, R. E. (1995). Experimental Design: Procedures for the Behavioral Sciences, third edition. Brooks/Cole.

Plackett, R. L. (1960). Models in the analysis of variance (with discussion). Journal of the Royal Statistical Society B 22, 195–217.

Snedecor, G. W., and Cochran, W. G. (1989). Statistical Methods, eighth edition. Iowa State University Press.

Spiegel, Murray R. (1992). Schaum's Outline of Theory and Problems of Statistics (2nd edition). New York: McGraw-Hill. Chapter 16.

Yates, F. (1967). A fresh look at the basic principles of the design and analysis of experiments. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 4, 777–790.

LECTURE THREE

Practical Steps in Solving An Analysis of Variance Problem

Introduction

Having explained the principles behind an analysis of variance, this lecture will provide an example of a one-way ANOVA. The student is advised to note carefully this procedure.

Objective

At the end of this lecture, you should be able to understand the practical steps employed in solving problems in analysis of variance.

Pre-Test

List the stepwise procedure in solving problems in analysis of variance.

CONTENT

Step 1: The data

The first point is to represent the two variables in a form that a statistical program will understand. To do this, the data should be converted from Table 2.1 to the 'samples and subscripts' form shown in Table 3.1. It can be seen here that FERTIL is represented by the subscripts 1, 2 and 3 which correspond to the three different fertilizers. This variable is categorical, and in this sense the values 1, 2 and 3 are arbitrary. In contrast, YIELD is continuous, the values representing true measurements.

Table 3.1 Data presented as samples and subscripts

FERTIL	YIELD (tonnes)
1	6.27
1	5.36
1	6.39
1	4.85
1	5.99
1	7.14
1	5.08
1	4.07
1	4.35
1	4.95
2	3.07
2	3.29
2	4.04
2	4.19
2	3.41
2	3.75
2	4.87
2	3.94
2	6.28
2	3.15
3	4.04
3	3.79
3	4.56
3	4.55
3	4.55
3	4.53
3	3.53
3	3.71
3	7.00
3	4.61

Step 2: The question

This is the first use of model formulae—a form of language that will prove to be extremely useful. The question we wish to ask is: 'Does fertilizer affect yield?'.

This can be converted to the word equation

YIELD = FERTIL.

This equation contains two variables: YIELD, the data we wish to explain and FERTIL, the variable we hypothesise might do the explaining.

YIELD is therefore the response (or dependent) variable, and FERTIL the explanatory (or independent) variable. It is important that the data variable is on the left hand side of the formula, and the explanatory variable on the right hand side. It is the right hand side of the equation that will become more complicated as we seek progressively more sophisticated explanations of our data.

Having entered the data into a worksheet in the correct format, and decided on the appropriate model formula and analysis, the specific command required to execute the analysis will depend upon your package. A typical output is presented here in a generalized format.

Analysis of variance with one explanatory variable

Word equation: YIELD = FERTIL

FERTIL is categorical

One-way analysis of variance for YIELD

Source	df	SS	MS	F	p
FERTIL	2	10.8227	5.4114	5.70	0.009
Error	27	25.6221	0.9490		
Total	29	36.4449			

Output

The primary piece of output is the ANOVA table, in which the partitioning of SS and df has taken place. This will either be displayed directly, or can be constructed by you with the output given. The total SS have been partitioned between treatment (FERTIL) and error, with a

parallel partitioning of degrees of freedom. Each of the columns ends with the total of the preceding terms.

The calculation of the SS is displayed in Table 3.2. Columns M , F and Y give the grand mean, the fertiliser mean and the plot yield for each plot in turn.

Table 3.2 Calculating the SS and the DF

Datapoint	FERTIL	M	F	Y	MY	MF	FY
1	1	4.64	5.45	6.27	1.63	0.80	0.82
2	1	4.64	5.45	5.36	0.72	0.80	−0.09
3	1	4.64	5.45	6.39	1.75	0.80	0.94
4	1	4.64	5.45	4.85	0.21	0.80	−0.60
5	1	4.64	5.45	5.99	1.35	0.80	0.54
6	1	4.64	5.45	7.14	2.50	0.80	1.69
7	1	4.64	5.45	5.08	0.44	0.80	−0.37
8	1	4.64	5.45	4.07	−0.57	0.80	−1.38
9	1	4.64	5.45	4.35	−0.29	0.80	−1.10
10	1	4.64	5.45	4.95	0.31	0.80	−0.50
11	2	4.64	4.00	3.07	−1.57	−0.64	−0.93
12	2	4.64	4.00	3.29	−1.35	−0.64	−0.71
13	2	4.64	4.00	4.04	−0.60	−0.64	0.04
14	2	4.64	4.00	4.19	−0.45	−0.64	0.19
15	2	4.64	4.00	3.41	−1.23	−0.64	−0.59
16	2	4.64	4.00	3.75	−0.89	−0.64	−0.25
17	2	4.64	4.00	4.87	0.23	−0.64	0.87
18	2	4.64	4.00	3.94	−0.70	−0.64	−0.06
19	2	4.64	4.00	6.28	1.64	−0.64	2.28
20	2	4.64	4.00	3.15	−1.49	−0.64	−0.85
21	3	4.64	4.49	4.04	−0.60	−0.16	−0.45
22	3	4.64	4.49	3.79	−0.85	−0.16	−0.70
23	3	4.64	4.49	4.56	−0.08	−0.16	0.07

24	3	4.64	4.49	4.55	−0.09	−0.16	0.06
25	3	4.64	4.49	4.55	−0.09	−0.16	0.06
26	3	4.64	4.49	4.53	−0.11	−0.16	0.04
27	3	4.64	4.49	3.53	−1.11	−0.16	−0.96
28	3	4.64	4.49	3.71	−0.93	−0.16	−0.78
29	3	4.64	4.49	7.00	2.36	−0.16	2.51
30	3	4.64	4.49	4.61	−0.03	−0.16	0.12
df		1	3	30	29	2	27
SS					36.44	10.82	25.62

Column MY then represents the deviations from the grand mean for each plot. If these values are squared and summed, then the result is the total SS of 36.44. FY then represents the deviations from the group mean for each plot; these values squared and summed give the error SS.

Finally, MF represents the deviations of the fertilizer means from the grand mean; squaring and summing giving the treatment SS. Dividing by the corresponding df gives the mean square. Comparison of the two mean squares gives the F -ratio of 5.70. The probability of getting an F -ratio as large as 5.70 or larger, if the null hypothesis is true, is the p-value of 0.009. That is sufficiently small to conclude that these fertilizers probably do differ in efficacy.

Presenting the results

Having concluded that there is a significant difference between the fertilizers, it would be interesting to know where this difference lies. One useful way of displaying the results would be to tabulate the means for each group, and their 95% confidence intervals. What do we mean by a confidence interval, and how are they constructed?

A confidence interval is an expression of how confident we are in our estimates (in this case, the three group means). For each confidence interval, we would expect the true mean for that group to lie within that range 95% of the time.

To construct a confidence interval, both the parameter estimate, and the variability in that estimate are required. In this case, the parameters estimated are means—we wish to know the true mean yield to be expected when we apply fertilizer 1, 2 or 3—which we will denote µA, µB, and µC respectively. These represent true population means, and as such we cannot know their exact values—but our three treatment means represent estimates of these three parameters. The reason why these estimates are not exact is because of the unexplained variation in the experiment, as quantified by the error variance which we previously met as the error mean square, and will refer to as s^2.

Table 3.3 Constructing confidence intervals

Fertiliser	**B**	t_{crit} **with 27 *df* for 95% confidence**		**Confidence interval**
1	5.445	2.0518	0.3081	(4.81, 6.08)
2	3.999	2.0518	0.3081	(3.37, 4.63)
3	4.487	2.0518	0.3081	(3.85, 5.12)

The 95% confidence interval for a population mean is:

$$B \pm t_{crit}$$

The key point is where our value for s comes from. If we had only the one fertilizer, then all information on population variance would come from that one group, and s would be the standard deviation for that group. In this instance however there are three groups, and the unexplained variation has been partitioned as the error mean square. This is using all information from all three groups to provide an estimate of unexplained variation—and the degrees of freedom associated with this estimate are 27—much greater than the 9 which would be associated with the standard deviation of any one treatment. So the value of s used is $\sqrt{MSe} = \sqrt{0.949} = 0.974$. This is also called the pooled standard deviation. Hence the 95% confidence intervals are as shown in Table 3.3.

These intervals, combined with the group means, are an informative way of presenting the results of this analysis, because they give an indication of how accurate the estimates are likely to be.

It is worth noting that we have assumed it is valid to take one estimate of s and apply it to all fertilizer groups. However, consider the following scenario. Fertilizer 1 adds nitrate, while Fertilizer 2 adds phosphate (and Fertilizer 3 something else altogether). The plots vary considerably in nitrate levels, and Fertilizer 1 is sufficiently strong to bring all plots up to a level where nitrate is no longer limiting. So Fertilizer 1 reduces plot-to-plot variation due to nitrate levels. The phosphate added by Fertilizer 2 combines multiplicatively with nitrate levels, so increasing the variability arising from nitrate levels. The mean yields from plots allocated to Fertilizer 2 would be very much more variable, while those allocated to Fertilizer 1 would have reduced variability, and our assumption of equal variability between plots within treatments would be incorrect. The 95% confidence interval for Fertilizer 2 will have been underestimated.

Fortunately in this case the group standard deviations do not look very different (Table 3.4), so it is unlikely that we have a problem.

Table 3.4

Descriptive Statistics for YIELD by FERTIL

FERTIL	*N*	Mean	Standard Deviation
1	10	5.445	0.976
2	10	3.999	0.972
3	10	4.487	0.975

Summary

In this lecture, you have been able to learn about the stepwise procedure in solving analysis of variance problems.

Post Test

1. Discuss the steps to be taken in solving an analysis of variance problem.
2. How do you set up a confidence interval in an analysis of variance problem?

3. Samples of size 11 are taken from each of 5 populations. Complete the following analysis of variance table:

Source	SS	df	MS	F
------	----	----	----	-
Between means	1000	a	c	e
Within samples	5000	b	d	
Total	6000			

   a. a = 4 b = 44 c = 250 d = 113.6 e = 2.2
   b. a = 4 b = 44 c = 250 d = 113.6 e = 0.2
   c. a = 5 b = 55 c = 200 d = 90.9 e = 0.2
   d. a = 5 b = 50 c = 200 d = 100 e = 2.0
   e. a = 4 b = 50 c = 250 d = 100 e = 2.5

4. In a single factor experiment with four levels if the mean square (between)=25, mean square(within)=10, n(1)=n(2)=n(3)=8 and n(4)=10, what is the value of (corrected) total sums of squares?
   a. 435
   b. 786
   c. 1221
   d. Insufficient information
   e. Sufficient information but correct value is not given

5. In the ANOVA for a single factor experiment with four levels all n 's equal 5 and $\bar{Y}_1 = 22, \bar{Y}_2 = 24, \bar{Y}_3 = 20$, and $\bar{Y}_4 = 18$. What is the sum of squares for between groups?
   a. 25.00
   b. 33.33
   c. 100.00
   d. Cannot be determined without more data

References

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Fisher, R. A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

Green, B. F., and Tukey, J. W. (1960). Complex analyses of variance: general problems. Psychometrika 25 127–152.

Kirk, R. E. (1995). Experimental Design: Procedures for the Behavioral Sciences, third edition. Brooks/Cole.

Snedecor, G. W., and Cochran, W. G. (1989). Statistical Methods, eighth edition. Iowa State University Press.

Spiegel, Murray R. (1992). Schaum's Outline of Theory and Problems of Statistics (2nd edition). New York: McGraw-Hill. Chapter 16.

Yates, F. (1967). A fresh look at the basic principles of the design and analysis of experiments. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 4, 777–790.

LECTURE FOUR

The Geometrical Approach for Analysis of Variance

Introduction

The analysis that has just been conducted can be represented as a simple geometrical picture. One advantage of doing this is that such pictures can be used to illustrate certain concepts. In this illustration, geometry can be used to represent the partitioning and additivity of the sum of squares (SS).

Objective

At the end of this lecture, you should be able to use the geometrical approach in solving an analysis of variance problem.

Pre Test

Explain the meaning of the following terms:

1. 2D plane
2. 3D plane

CONTENT

The geometrical approach is actually a two-dimensional representation of multidimensional space. One dimension is represented by the position of a point on a line—one coordinate can be used to define that position. Two dimensions may be pictured as a graph, with a point being specified by two coordinates. This can be extended to three dimensions, in which the position of a point in a cube is specified by three coordinates. Beyond three dimensions it is no longer possible to visualise a geometrical

picture to represent all dimensions simultaneously. It is possible however to take a slice through multidimensional space and represent it in two dimensions. For example, if a cube has axes x, y, and z, the position of three points can be specified by their x, y, and z coordinates. A plane could then be drawn through those three points, so allowing them to be represented on a piece of paper. There are still three coordinates associated with each point (and so defining that point), but for visual purposes, the three dimensions have been reduced to two. In fact, it is possible to do this for any three points, however many dimensions they are plotted in. This trick is employed by the geometrical approach.

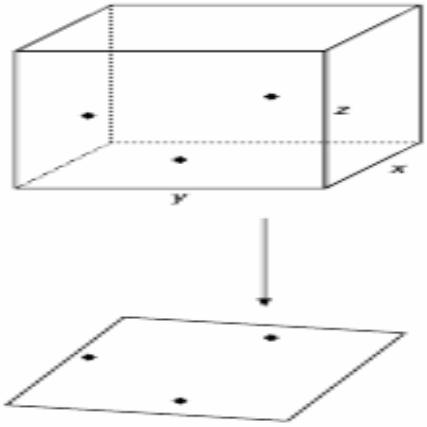
In this case, there are as many dimensions as there are data points in the dataset (30). Each point is therefore represented by 30 coordinates. The three points themselves are the columns 3, 4 and 5 (M, F and Y) of Table 3.2.

Point Y

This point represents the data, so the 30 coordinates describing this point are the 30 measurements of yield.

Point M

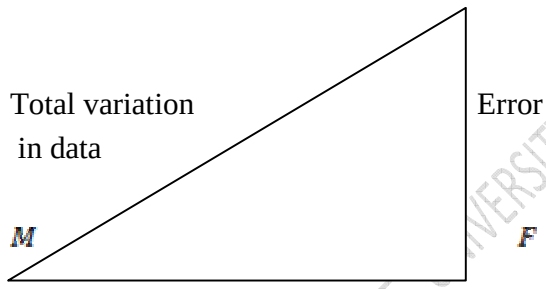
This point represents the grand mean. Because we are dealing with 30-dimensional space (as dictated by the size of the dataset), this point also has 30 coordinates specifying its position in multidimensional space. However, the values of these 30 coordinates are all the same (the grand mean).



Representing 3D in 2D

Point F

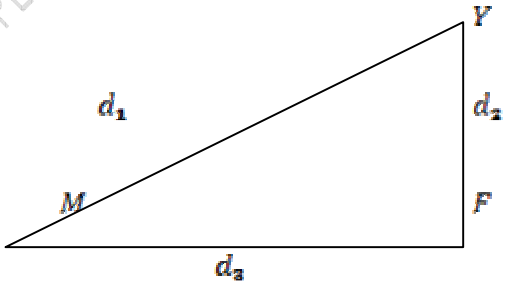
This point represents the treatment means. While still 30 elements long, the first ten elements are the mean for treatment 1 (and are therefore the same value), the second ten the mean for treatment two etc. Therefore the first part of the geometrical approach is that the three variables, M , F and Y , are represented as points. These three points may be joined to form a triangle in two dimensional space as follows:



Effect of fertilizer

Variables represented as points, sources as vectors

The triangle has been drawn with F at a right angle. The lines joining the points are vectors, and these represent sources of variability. For example, the vector MY represents the variability of the data (Y) around the grand mean (M). In the same way that a vector can be decomposed into two components, so can the variability be partitioned into (i) FY —the variability of the data around their group means, and (ii) MF —the variability of the group means around the grand mean. The implication here is that sources of variability are additive. While this assumption is crucial in our approach, it is not necessarily true.



The Pythagoras theorem

The third part of the geometrical approach relies on the fact that the triangle is right-angled. The squared length of each vector is then equivalent to the SS for that source.

Pythagoras states that:

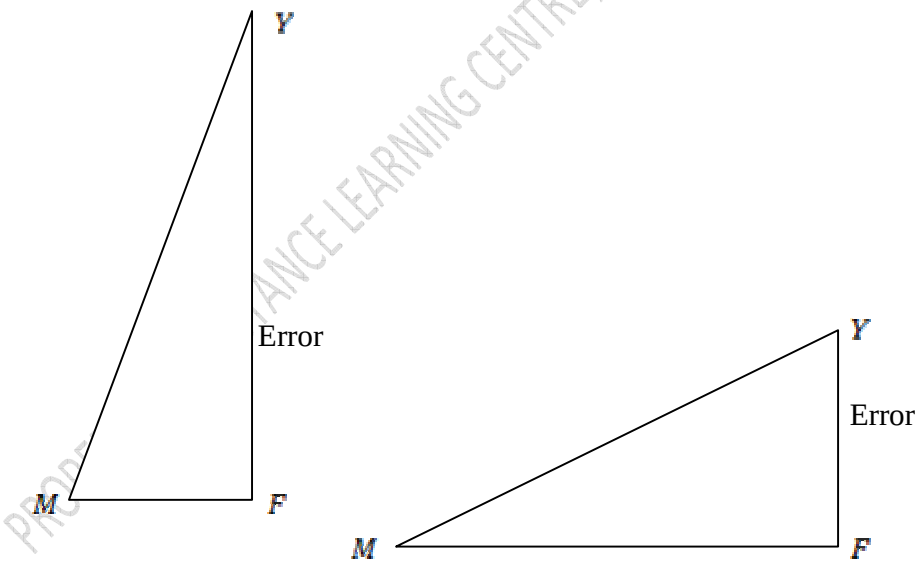
$$d_1^2 = d_2^2 + d_3^2$$

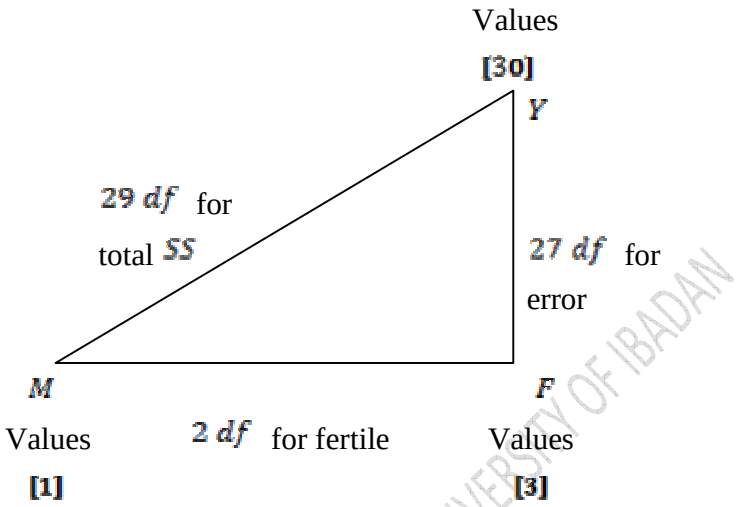
This is equivalent to:

$$SST = SSt + SSe$$

This illustrates geometrically the partitioning of sums of squares. It is precisely because the sums of squares can be summed in this way that they are central to the analysis of variance. Other ways of measuring variation (e.g. using variances) would not allow this, because the variances of the components would not add up to the variance of the whole.

The shape of the triangle can then provide information on the relative sizes of these different components.





Partitioning the degrees of freedom

It is also possible to represent the parallel partitioning of degrees of freedom in a similar manner. At each apex, are the numbers of values in each variable (30 different data points, 3 treatment means, and 1 grand mean). The difference between these values gives the number of degrees of freedom associated with moving from one point to another. For example, calculating a grand mean is equivalent to moving from Y to M, and in doing so, twenty nine degrees of freedom are lost. Moving from M to F is equivalent to exchanging one grand mean for three treatment means, the difference being two degrees of freedom. These degrees of freedom are associated with the corresponding vectors and therefore with thesources represented by these vectors.

Nonsphericity

Due to the nature of the level in an experiment, it may be the case that if a subject responds strongly to level i he may respond strongly to level j. In other words there may be a correlation between responses. These show that for some pairs of conditions there does not seem to be a correlation. This correlation can be characterized graphically by fitting a Gaussian to each 2D data cloud and then plotting probability contours. If these contour form a sphere (a circle, in two dimensions) then the data is Independent

and Identically Distributed (IID), i.e, same variance in all dimensions and there is no correlation. The more these contours look like ellipses, the more 'nonsphericity' there is in the data.

The possible nonsphericity can be taken into account in the analysis using a correction to degree of freedom. In the above example, a Greenhouse Geisser (GG) correction estimates $e = .7$, giving DFs of $[2.1, 23.0]$ and a p-value (with GG we use the same F -statistics i.e. F = 6.89) of p =*0.004*. Assuming sphericity, as before, we computed p = *0.001*. Thus the presence of nonsphericity in the data makes us less confident of the significance of the effect.

An alternative representation of the within-subjects model is given in the appendix. This shows how one can take into account nonsphericity. Various other relevant terminologies is also defined in the appendix.

Summary

In this lecture, you have been able to solve an analysis of variance problem using a geometric approach.

Post Test

Use the geometric approach to solve problems in lectures 1, 2 and 3.

References

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Gelman, A., Pasarica, C., and Dodhia, R. M. (2002). Let's practice what we preach: using graphs instead of tables. The American Statistician 56, 121–130.

LECTURE FIVE

Two-Way Analysis of Variance (1)

Introduction

Sometimes a researcher might want to simultaneously examine the effects of two treatments (where both treatments have nominal-level measurement). This lecture, thus, introduces the concept of two-way analysis of variance.

Objectives

At the end of this lecture, you should be able to

1. understand the concept of a ***two-way analysis of variance***.
2. know how to solve problems involving two-way analysis of variance.
3. understand and solve problems involving ***interactions***.

Pre Test

1. What do you understand by a two-way analysis of variance?
2. What is interaction?

CONTENT

Examples:

The effect of sex and race on wages

The effects of the level of pollution and the level of city services on housing prices

The effects of religion and region on income

To elaborate: with sex and race, we might wonder if

There are differences because of sex alone

There are differences because of race alone

There are differences attributable to particular combinations of sex and race - that is, are there interaction effects? For example, white males, white females, and black males may all have similar wages, but black females could have much lower wages. We'll discuss interaction effects more shortly.

Two-Way Anova with a Balanced Design and the Classic Experimental Approach

We can use Analysis of Variance techniques for these and more complicated problems. These techniques can get fairly involved and employ several different options, each of which has various strengths and weaknesses. In a balanced design, *all cell frequencies are equal*, i.e. the number of observations in each combination of treatments is the same. So, for example, there would be 5 white males, 5 black males, 5 white females, and 5 black females. Balanced designs are unlikely in survey research but they are quite common (and often encouraged) in experimental studies. Equal cell frequencies make it easier to disentangle the effects of the row and column variables (e.g. sex and race) and also minimizes the effect of non-homogenous population variances if they exist.

In addition, note that several programs give various options for the "Method" to use for Anova. If the design is balanced, it does not matter what method you use. But, if you choose what SPSS calls the *Classic Experimental Approach*, many of the formulas that follow will be valid even when the design is not balanced. The *Regression Approach* and the *Hierarchical Approach* are other options (and several other options, with varying names, are also listed in different procedures). The SPSS manual and other sources have more information if you find yourself needing to know about these.

As noted below, these assumptions are not required for everything we will be talking about. These assumptions will affect how computations are done with the raw data but, once that is done, the hypothesis testing procedures will be largely the same. Ergo, the most critical parts of our discussion will apply even when designs are not balanced.

The Model

When we have 2 treatments, the model can be written as

$$y_{ij} = \mu + \tau_j + \lambda_k + (\tau\lambda)_{jk} + \varepsilon_{ijk}$$

where μ = the grand mean, τ_j is the treatment effect for the jth category of the row variable, λ_k is the treatment effect for the k th category of the column variable, $(\tau\lambda)_{jk}$ is the interaction effect for the combination of the j th row category and the k th column category.

Example: Suppose the overall average income is N20,000, the average black income is N15,000, the average female income is N17,000, and the average black woman's income is N10,000. This means that μ = N20,000, τ_B = -N5,000, λ_W = -N3,000, $(\tau\lambda)_{BW}$ = -N2,000.

As before, we want to partition the variance. Note that

$$s_y^2 = \frac{\Sigma\Sigma\Sigma\overline{y_{ijk} - (\bar{y})}^2}{N-1} = \frac{Total\ SS}{N-1} = \frac{TSS}{N-1} = MS\ Total$$

Further, note that

Component	Description
$\overline{y_{ijk} - (\bar{y})}$ =	Deviation of the individual score from the overall mean
$\overline{y_{ijk} - (\bar{y}_{jk})}$ =	Deviation of the individual score from the group mean, i.e. ε_{ijk}
+ $\overline{y_j - (\bar{y})}$	Deviation of the jth row's mean from the overall mean, i.e. τ_j

$+\overline{\overline{y_k} - (\bar{y})}$	Deviation of the kth column's mean from the overall mean, i.e. $\hat{\lambda}_k$
$\overline{\overline{y_{jk}} - \overline{y_j} - \overline{y_k} + \bar{y}}$	Deviation of "combination" mean from row and column means; the interaction, i.e. $(\tau\hat{\lambda})_{jk}$

Note that we are using the same trick we did before of adding and then subtracting the same terms.

Hence, $\Sigma\Sigma\Sigma\,\overline{y_{ijk} - (\bar{y})}^2$ can be broken out as follows (any seemingly omitted terms conveniently work out to be zero):

$$\Sigma\Sigma\Sigma\,\overline{y_{ijk} - (y_{jk})}^2 = \Sigma\Sigma\Sigma\,\hat{\varepsilon}^2_{ijk} = SS\ Error,$$

$$df = n - jk$$

This is analogous to SS Within from 1-way ANOVA. This represents the deviation of individuals from the means of others who have the same value on the row and column variables (e.g. are of the same sex and race); that is, this represents the component of the scores that cannot be accounted for by group membership. The df arise from the fact that there are N cases, and $j * k$ means have to be estimated.

Also,

$$\Sigma\Sigma\Sigma\,\overline{y_j - (\bar{y})}^2 = \Sigma\Sigma\Sigma\,\tau_j^2 = SS\ Rows,$$

$$df = j - 1$$

$$\overline{\overline{y_k} - (\bar{y})}^2 = \Sigma\Sigma\Sigma\,\lambda_k^2 = SS\ Columns,$$

$$df = k - 1$$

$$\Sigma\Sigma\Sigma\,\overline{y_{jk} - \overline{y_j} - \overline{y_k} + \bar{y}}^2 = \Sigma\Sigma\Sigma\,(\tau\lambda)_{jk}^2 = SS\ Interaction,$$

$$df = (j-1)(k-1)$$

Other useful partitioning include

$$SS\ Main = SS\ Total - SS\ Interaction - SS\ Residual$$
$$df = j + k - 2$$

Note also that, *when all cell frequencies are equal*, i.e. the number of observations in each combination of treatments is the same,

SS Main = SS Columns + SS Rows.

This will not necessarily be true otherwise. The fact that it is true in a balanced design is one of its main advantages.

Another useful partitioning is

SS Cells = SS Explained = SS Main + SS Interaction = SS Total – SS Error

$$df = jk - 1$$

When all cell frequencies are equal,

SS Cells = SS Columns + SS Rows + SS Interaction.

Finally, note that,

Total SS = SS Main + SS Interaction + SS Error = SS Explained + SS Error

$$df = j - 1 + k - 1 + jk + 1 - j - k + n - jk = n - 1$$

Again, *when all cell frequencies are equal,* Total SS = SS Columns + SS Rows + SS Interaction + SS Error.

When doing statistical inference, we assume that for each treatment combination JK, the random error terms ε_{ijk} are - N(0, σ^2); the variance σ^2 is the same for each treatment combination. The random error terms are independent

I. **Tests OF Interest**:

A. $H_0: (\tau\lambda)_{jk} = 0$ for all j, k

$H_A: (\tau\lambda)_{jk} <> 0$ for at least 1 j, k

This is a test of whether there are any <u>interaction effects</u>; the appropriate <u>test statistic</u> is

$$F_{(J-1)(k-1)N-jk} = \frac{\textbf{SS Interaction}(J-1)(k-1)}{\textbf{SS Error}(N-jk)} = \frac{\textbf{MS Interaction}}{\textbf{MS Error}}$$

If the null hypothesis is true, $F_{cal} < F_{tab}$

B. $H_0: \tau_1 = \tau_2 = \ldots = \tau_J = 0$

$H_A:$ At least 1 $\tau_j <> 0$

This tests whether there are any <u>row effects</u>. The appropriate <u>test statistic</u> is

$$F_{J-1N-jk} = \frac{\textbf{SS Rows}\ (J-1)}{\textbf{SS Error}(N-jk)} = \frac{\textbf{MS Rows}}{\textbf{MS Error}}$$

If the null hypothesis is true, $F_{cal} < F_{tab}$

C. $H_0: \lambda_1 = \lambda_2 = \ldots = \lambda_K = 0$

H_A: At least 1 $\lambda_k <> 0$

This tests whether there are any column effects. The appropriate test statistic is

$$F_{k-1N-jk} = \frac{\text{SS Columns } (k-1)}{\text{SS Error}(N-jk)} = \frac{\text{MS Columns}}{\text{MS Error}}$$

If the null hypothesis is true, $F_{cal} < F_{tab}$

Note: The last two tests are primarily of interest if you conclude that interaction effects are not significant. If, on the other hand, you conclude that the interaction effects do not equal zero, then you know both treatments (i.e. the row and column effects) are significant.

D. H_0: All τ's and λ's = 0

H_A: At least one τ or λ does not equal 0

This tests whether any of the main effects (i.e. row or column effects; or, non-interaction effects) are nonzero. The appropriate test statistic is

$$F_{j+k-2,N-jk} = \frac{\text{SS Main } (j+k-2)}{\text{SS Error}(N-jk)} = \frac{\text{MS Main}}{\text{MS Error}}$$

If the null hypothesis is true, $F_{cal} < F_{tab}$

H_0: All τ's, λ's, and (τλ)'s = 0

H_A: At least one τ, λ, or (τλ) does not equal 0

This tests whether there are any effects at all. If the null hypothesis is true, then every cell in the table will have the same true mean. The appropriate test statistic is

$$F_{jk-1,N-jk} = \frac{\textbf{SS Cells } (jk-1)}{\textbf{SS Error}(N-jk)} = \frac{\textbf{MS Cells}}{\textbf{MS Error}}$$

If the null hypothesis is true, F - **F**([JK - 1], N - JK).

III. Row, Column, and Interaction Effects – Examples

What are interaction effects? Here are some substantive examples:

Medicines A and B may have no effect when either is taken alone. But, the two together may have an effect. "The whole is different from the sum of the parts."

Another example: we might find that greater income leads to greater fertility for those who want children, and lower fertility for those who do not want children. We say that the effect of income is dependent on desires, or that desires and income interact in determining fertility.

Good teachers and small classrooms might both encourage learning. A good teacher in a small classroom might be especially effective. The whole is greater than the sum of the parts.

Following are hypothetical 2-way ANOVA examples. The dependent variable is income (in thousands of dollars), the row variable is gender (Male or Female), the column variable is type of occupation (A, B, or C). Unless otherwise stated, assume that frequencies are equal for all cells.

1. Row (Gender) effects only.

	Occ A	**Occ B**	**Occ C**	
Male	$\mu_{MA} = 18$ $\tau\lambda_{MA} = 0$	$\mu_{MB} = 18$ $\tau\lambda_{MB} = 0$	$\mu_{MC} = 18$ $\tau\lambda_{MC} = 0$	$\mu_{M} = 18$ $\tau_{M} = 2$

Female	$\mu_{FA} = 14$ $\tau\lambda_{FA} = 0$	$\mu_{FB} = 14$ $\tau\lambda_{FB} = 0$	$\mu_{FC} = 14$ $\tau\lambda_{FC} = 0$	$\mu_{F} = 14$ $\tau_{F} = -2$
	$\mu_{A} = 16$ $\lambda_{A} = 0$	$\mu_{B} = 16$ $\lambda_{B} = 0$	$\mu_{C} = 16$ $\lambda_{C} = 0$	$\mu = 16$

The 2 rows differ, but the three columns are all the same. Within each occupation, men make ₦4,000 more on average than do women; each of the three occupations pays equally well.

2. Column (Occupation) effects only.

	Occ A	**Occ B**	**Occ C**	
Male	$\mu_{MA} = 12$ $\tau\lambda_{MA} = 0$	$\mu_{MB} = 16$ $\tau\lambda_{MB} = 0$	$\mu_{MC} = 20$ $\tau\lambda_{MC} = 0$	$\mu_{M} = 16$ $\tau_{M} = 0$
Female	$\mu_{FA} = 12$ $\tau\lambda_{FA} = 0$	$\mu_{FB} = 16$ $\tau\lambda_{FB} = 0$	$\mu_{FC} = 20$ $\tau\lambda_{FC} = 0$	$\mu_{F} = 16$ $\tau_{F} = 0$
	$\mu_{A} = 12$ $\lambda_{A} = -4$	$\mu_{B} = 16$ $\lambda_{B} = 0$	$\mu_{C} = 20$ $\lambda_{C} = 4$	$\mu = 16$

The three columns differ, but the two rows are the same. Occupation C pays better than B and B pays better than A. Within each occupation, however, men and women make the same.

3. Row and column effects.

	Occ A	**Occ B**	**Occ C**	
Male	$\mu_{MA} = 14$ $\tau\lambda_{MA} = 0$	$\mu_{MB} = 18$ $\tau\lambda_{MB} = 0$	$\mu_{MC} = 22$ $\tau\lambda_{MC} = 0$	$\mu_{M} = 18$ $\tau_{M} = 2$
Female	$\mu_{FA} = 10$ $\tau\lambda_{FA} = 0$	$\mu_{FB} = 14$ $\tau\lambda_{FB} = 0$	$\mu_{FC} = 18$ $\tau\lambda_{FC} = 0$	$\mu_{F} = 14$ $\tau_{F} = -2$
	$\mu_{A} = 12$ $\lambda_{A} = -4$	$\mu_{B} = 16$ $\lambda_{B} = 0$	$\mu_{C} = 20$ $\lambda_{C} = 4$	$\mu = 16$

Both the rows and columns differ. Within each occupation, men make ₦4,000 more on average than women do. Within each gender, those in occupation C average ₦4,000 more than those in B, and those in B average ₦4,000 more than those in A.

4. Interaction effects I.

	Occ A	**Occ B**	**Occ C**	
Male	$\mu_{MA} = 15$ $\tau\lambda_{MA} = -1$	$\mu_{MB} = 15$ $\tau\lambda_{MB} = -1$	$\mu_{MC} = 21$ $\tau\lambda_{MC} = 2$	$\mu_{M} = 17$ $\tau_{M} = 1$
Female	$\mu_{FA} = 15$ $\tau\lambda_{FA} = 1$	$\mu_{FB} = 15$ $\tau\lambda_{FB} = 1$	$\mu_{FC} = 15$ $\tau\lambda_{FC} = -2$	$\mu_{F} = 15$ $\tau_{F} = -1$
	$\mu_{A} = 15$ $\tau_{A} = -1$	$\mu_{B} = 15$ $\tau_{B} = -1$	$\mu_{C} = 18$ $\tau_{C} = 2$	$\mu = 16$

Five of the six cells have the same mean. However, for some reason, the combination of males and occupation C results in high male earnings.

5. Interaction effects II - differing magnitudes of effects.

	Occ A	**Occ B**	**Occ C**	
Male	$\mu_{MA} = 12$ $\tau\lambda_{MA} = -1$	$\mu_{MB} = 16$ $\tau\lambda_{MB} = -1$	$\mu_{MC} = 26$ $\tau\lambda_{MC} = 2$	$\mu_{M} = 18$ $\tau_{M} = 2$
Female	$\mu_{FA} = 10$ $\tau\lambda_{FA} = 1$	$\mu_{FB} = 14$ $\tau\lambda_{FB} = 1$	$\mu_{FC} = 18$ $\tau\lambda_{FC} = -2$	$\mu_{F} = 14$ $\tau_{F} = -2$
	$\mu_{A} = 11$ $\lambda_{A} = -5$	$\mu_{B} = 15$ $\lambda_{B} = -1$	$\mu_{C} = 22$ $\lambda_{C} = 6$	$\mu = 16$

Men make more than women, and the advantage is especially great in occupation C. Or, those in occupation C make more than those in other occupations, and the advantage is especially great for men.

6 Interaction effects III - differing directions of effects.

	Occ A	**Occ B**	**Occ C**	
Male	$\mu_{MA} = 18$ $\tau\lambda_{MA} = +2$	$\mu_{MB} = 16$ $\tau\lambda_{MB} = 0$	$\mu_{MC} = 14$ $\tau\lambda_{MC} = -2$	$\mu_{M} = 16$ $\tau_{M} = 0$
Female	$\mu_{FA} = 14$ $\tau\lambda_{FA} = -2$	$\mu_{FB} = 16$ $\tau\lambda_{FB} = 0$	$\mu_{FC} = 18$ $\tau\lambda_{FC} = 2$	$\mu_{F} = 16$ $\tau_{F} = 0$
	$\mu_{A} = 16$ $\lambda_{A} = 0$	$\mu_{B} = 16$ $\lambda_{B} = 0$	$\mu_{C} = 16$ $\lambda_{C} = 0$	$\mu = 16$

In this example, the effect of gender depends on occupation. Males do better than women in Occupation A but worse in occupation C; in Occupation B there is no difference. Or, occupation C is better paying for women but not for men, whereas for occupation A the opposite is true. Note that, if you only looked at the main effects, you would erroneously conclude that gender and occupation have no effects on income, when in reality they do have effects but the effects work in opposing directions.

Computational Procedures - Two-Way ANOVA – Balanced Designs

Let A = row variable, B = column variable, J = number of categories for A, K = number of categories for B, T_{Aj} = the sum of the scores in group A_j, T_{Bk} = the sum of the scores in group B_k, T_{AjBk} is the sum of the scores for the observations which fall in both groups A_j and B_k (there are J*K of these totals), n_{Aj} = number of observations in group A_j, n_{Bk} = number of observations in group B_k, n_{AjBk} is the number of observations which fall in both groups A_j and B_k. [NOTE: While I will show you how to do the raw data calculations, in practice they are tedious enough that I generally would not expect you to do them by hand, at least on an exam. You should know how to do the other formulas, however, as they show how the different parts of the ANOVA table are related to each other.]

Note that many (albeit not all) of the formulas for raw data calculations and Sums of Squares assume a *balanced design*, i.e. all cell frequencies are equal for each possible combination of values for the row and column variables. Computations are somewhat more complicated when designs are not balanced. *The Mean Square formulas and the F tests are accurate regardless of whether the design is balanced or not.*

Formula	Explanation
Raw Data Calculations (Balanced Design)	
(1) = $(\Sigma\Sigma\Sigma y_{ijk})^2/n = N\bar{}\mu^2$	Sum all the observations. Square the result. Divide by the total number of observations.
(2) = $\Sigma\Sigma\Sigma y_{ijk}^2$	Square each observation. Sum the squared observations.
(3) = $\Sigma T_{Aj}^2/n_{Aj}$	Add up the values for the observations for group A_1. Square the result. Divide by the number of observations in group A_1. Repeat for each category of A. Add the results for each of the J groups together.
(4) = $\Sigma T_{Bk}^2/n_{Bk}$	Add up the values for the observations for group B_1. Square the result. Divide by the number of observations in group B_1. Repeat for each category of B. Add the results for each of the K groups together.
(5) = $\Sigma\Sigma T_{AjBk}^2/n_{AjBk}$	Add up the values for the observations which fall in both group A_1 and B_1. Square this value, and divide by n_{A1B1}. Repeat for each of the J*K combinations, and sum the results.

Sums of Squares Calculations (Balanced Design)	
SS Total = (2) - (1)	Total sum of squares
SS Rows = (3) - (1)	Row sum of squares. This is also sometimes called SS_A.
SS Columns = (4) - (1)	Column sum of squares. Also called SS_B.
SS Interaction = (5) + (1) - (3) - (4) = SS Total - SS Rows - SS Columns - SS Error = SS Total – SS Main – SS Error	Interaction sum of squares. Also called SS_{AB}. It may be easier to use the second formula.
SS Error = (2) - (5) = SS Total - SS Cells	Error sum of squares. It is analogous to SS Within in one-way ANOVA. Also called SS Residual.

SS Main = (3) + (4) – [2 * (1)] = SS Columns + SS Rows = SS Total – SS Error – SS Interaction	Main effects Sum of Squares. Also called SS_{A+B}
SS Cells = (5) - (1) = SS Main + SS interaction = SS Total - SS Error.	This is analogous to SS Between in one-way ANOVA. Also called SS Explained.
Mean Square Calculations (Balanced or unbalanced)	
MS Total = s^2 = SS Total/(n-1)	Remember that MS Total = s^2
MS Rows = SS Rows/(J-1)	Also called MS_A.
MS Columns = SS Columns/(K-1)	Also called MS_B.
MS Interaction = SS Interaction/((J-1)(K-1))	Also called MS_{AB}
MS Main = SS Main/(J + K - 2)	Also called MS_{A+B}
MS Cells = SS Cells/((J*K)-1)	Also called MS Explained.
MS Error = SS Error/ (n - J*K)	Also called MS Residual.
Possible F Tests (Balanced or unbalanced):	
MS Rows/MS Error	Do means differ across categories of the row variable, i.e. do tau's differ? d.f. = J-1, n-J*K
MS Columns/MS Error	Do means differ across categories of the column variable, i.e. do lambdas differ? d.f. = K-1, n-J*K
MS Interaction/MS Error	Do any of the interaction effects differ from zero? d.f. = (J-1)(K-1), n-J*K
MS Main/MS Error	Are any of the row or column effects nonzero? d.f. = J + K - 2, n-J*K
MS Cells/MS Error	Are there any differences anywhere across groups? d.f. = (JK-1), N-JK.

An ANOVA table often looks something like this (with the computed values substituted).

Source	SS	D.F.	Mean Square	F
A + B (or Main Effects)	SS Main	J + K - 2	SS Main / (J + K - 2)	MS Main / MS Error
A (or main effect of A)	SS Rows	J - 1	SS Rows / (J - 1)	MS Rows / MS Error
B (or main effect of B)	SS Columns	K - 1	SS Columns / (K - 1)	MS Columns / MS Error
AB (or 2-way interaction)	SS Interaction	(J - 1) * (K - 1)	SS Interaction / (J -1) (K - 1)	MS Interaction / MS Error
A + B + AB (or explained)	SS Cells	(J * K) - 1	SS Cells / (J * K) - 1	MS Cells / MS Error
Error (or residual)	SS Error	N - (J * K)	SS Error / (N - J * K)	
Total	SS Total	N - 1	SS Total / (N - 1)	

Examples

1. A researcher is interested in differences in income by Region (North, South, East, and West) and Religion (Catholic, Protestant, Other). She draws a sample of ten people for each combination of region and religion. She finds that SS Rows = 200, SS Columns = 170, SS Interaction = 100, and s^2 = 16.81. Construct the Anova Table, and indicate which effects are significant at the .05 level. (NOTE: Region is the row variable.)

Solution

Again the design is balanced. You don't have to do any work with the raw data here; instead, you have to understand how the different parts of the ANOVA table are related to each other. Let us begin with what we are told:

Source	SS	D.F.	Mean Square	F
A + B (or Main Effects)	SS Main =	J + K - 2 =	SS Main = (J + K - 2)	MS Main = MS Error
A (or main effect of A)	SS Rows = **200**	J - 1 =	SS Rows = (J - 1)	MS Rows = MS Error
B (or main effect of B)	SS Columns = **170**	K - 1 =	SS Columns = (K - 1)	MS Columns = MS Error
AB (or 2-way interaction)	SS Intraction = **100**	(J - 1) * (K - 1) =	SS Intrction = (J -1)(K - 1)	MS Intrction = MS Error
A + B + AB (or explained)	SS Cells =	(J * K) - 1 =	SS Cells = (J * K) - 1	MS Cells = MS Error
Error (or residual)	SS Error =	N - (J * K) =	SS Error = (N - J * K)	
Total	SS Total =	N - 1 =	SS Total = **16.81** (N - 1)	

We are also told J = 4 (there are 4 regions), K = 3 (3 religions).

We can deduce that N = J*K*10 = 120.

Recall that s^2 = MS Total, and that MS Total = SS Total/(n-1)

==> SS Total = s^2 * (N - 1) = 16.81 * 119 = 2000.

SS Main is obtained by adding SS Rows + SS Columns = 200 + 170 = 370.

SS Cells is obtained by adding up SS Columns + SS Rows + SS Interactions

= 200 + 170 + 100 = 470.

SS Error is obtained by computing SS Total - SS Cells = 2000 - 470 = 1530.

The remaining quantities in the table are obtained by filling in the appropriate values for the formulas. Hence, we get (* = significant at the .05 level):

Source	SS	D.F.	Mean Square	F
A + B (or Main Effects)	SS Main = 370	J + K - 2 = 5	SS Main = 74.00 (J + K - 2)	MS Main = 5.22* MS Error
A (or main effect of A)	SS Rows = 200	J - 1 = 3	SS Rows = 66.67 (J - 1)	MS Rows = 4.71* MS Error
B (or main effect of B)	SS Columns = 170	K - 1 = 2	SS Columns = 85.00 (K - 1)	MS Columns = 6.0* MS Error
AB (or 2-way interaction)	SS Intraction = 100	(J - 1) * (K - 1) = 6	SS Intrction = 16.67 (J -1)(K - 1)	MS Intrction = 1.18 MS Error
A + B + AB (or explained)	SS Cells = 470	(J * K) - 1 = 11	SS Cells = 42.73 (J * K) - 1	MS Cells = 3.02* MS Error
Error (or residual)	SS Error = 1530	N - (J * K) = 108	SS Error = 14.17 (N - J * K)	
Total	SS Total = 2000	N - 1 = 119	SS Total = 16.81 (N - 1)	

2. A consumer research firm wants to compare three brands of radial tires (X, Y, and Z) in terms of tread life over different road surfaces. Random samples of four tires of each brand are selected for each of three surfaces (asphalt, concrete, gravel). A machine that can simulate road conditions for each of the road surfaces is used to find the tread life (in thousands of miles) of each tire. Construct an ANOVA table and conduct F-tests for the presence of nonzero brand effects, road surface effects, and interaction effects.

Surface/ Brand	X	Y	Z
Asphalt	36, 39, 39, 38	42, 40, 39, 42	32, 36, 35, 34
Concrete	38, 40, 41, 40	42, 45, 48, 47	37, 33, 33, 34
Gravel	34, 32, 34, 35	34, 34, 30, 31	36, 35, 35, 33

although on an exam I'd be more likely to give you something like problem 1 and/or give you finished results and ask you to interpret them.

Note that the design is balanced. Let A = Road surface, B = Brand. HINT: It is legitimate to subtract a constant from EVERY observation. This will not affect any of the values in the ANOVA table, and it often makes the calculations simpler. I will subtract 30 from each observation, yielding the following table:

Surface/ Brand	**X**	$\mathbf{T_{AjBk}}$	**Y**	$\mathbf{T_{AjBk}}$	**Z**	$\mathbf{T_{AjBk}}$	$\mathbf{T_{Aj}}$
Asphalt	6 9 9 8	32	12 10 9 12	43	2 6 5 4	17	92
Concrete	8 10 11 10	39	12 15 18 17	62	7 3 3 4	17	118
Gravel	4 2 4 5	15	4 4 0 1	9	6 5 5 3	19	43
$\mathbf{T_{Bk}}$		86		114		53	253

1. $= (\Sigma\Sigma\Sigma y_{ijk})^2/n = 253^2/36 = 1778.03$
2. $= \Sigma\Sigma\Sigma y_{ijk}^2 = 6^2 + 9^2 + 12^2 + ... + 3^2 = 2451$
3. $= \Sigma\, T_{Aj}^2/n_{Aj} = 92^2/12 + 118^2/12 + 43^2/12 = 2019.75$
4. $= \Sigma\, T_{Bk}^2/n_{Bk} = 86^2/12 + 114^2/12 + 53^2/12 = 1933.42$
5. $= \Sigma\Sigma\, T_{AjBk}^2/n_{AjBk} = 32^2/4 + 39^2/4 + ... + 19^2/4 = 2370.75$

SS Total = (2) - (1) = 2451 - 1778.03 = 672.97

SS Rows = (3) - (1) = 2019.75 - 1778.03 = 241.72

SS Columns = (4) - (1) = 1933.42 - 1778.03 = 155.39

SS Interaction = (5) + (1) - (3) - (4) =

2370.75 + 1778.03 - 2019.75 - 1933.42 = 195.61

SS Main = SS Rows + SS Columns = 397.11

SS Cells = (5) - (1) = 592.72

SS Error = (2) - (5) = 80.25

Anova Table:

Source	Ss	D.f.	Mean square	F
A + B	397.11	4	99.28	33.43*
A	241.72	2	120.86	40.69*
B	155.39	2	77.70	26.16*
AB	195.61	4	48.90	16.46*
A+B+AB	592.72	8	74.09	24.95*
Error	80.25	27	2.97	
Total	672.97	35	19.23	

- = significant at the .05 level

Note:

- • To test for the presence of nonzero road effects, the degrees of freedom = 2,27 and we accept H_0 if F # 3.34.
- • To test for the presence of nonzero brand effects, d.f. = 2,27 and we accept H_0 if F # 3.34.
- • To test for the presence of nonzero interaction effects, d.f. = 4,27 and we accept H_0 if F # 2.72.
- • To test for the presence of any nonzero effects, d.f. = 8, 27 and we accept H_0 if F # 2.21.

Summary

In this lecture, you have been able to

1. learn about the concept of a two-way analysis of variance.
2. learn about interactions.
3. solve problems relating to two-way analysis of variance.

Post Test

1. Investigators studying the biology of cell death carried out an experiment in rats that explored neuroprotection associated with varying doses of troglitazone and varying ion forms. The outcome measured was percent cell death relative to glutamate (GLUT). Higher values of GLUT indicate greater cell death. The study design utilized a fully factorial 2 way analysis of variance model. Factor I is dose troglitazone at 3 levels: 1.3, 4.5 and 13.5. Factor II is ion form at two levels: 0=negative and 1=positive. Using the data below,
   a. State an appropriate ANOVA model, defining all terms. State appropriate null and alternative hypotheses.
   b. Test the assumption of equality of variances.
   c. Construct the ANOVA model analysis of variance table.
   d. Carry out the analysis of variance. Report your findings in a sentence or two that summarizes your conclusions. Report any limitations.

glut	dose	ion
73.61	1.3	0
130.69	1.3	0
118.01	1.3	0
140.2	1.3	0
97.11	1.3	1
114.26	1.3	1
120.26	1.3	1
92.39	1.3	1
26.95	4.5	0
53.23	4.5	0
59.57	4.5	0
53.23	4.5	0
28.51	4.5	1
30.65	4.5	1
44.37	4.5	1
36.23	4.5	1

-8.83	13.5	0
25.14	13.5	0
20.16	13.5	0
34.65	13.5	0
-35.8	13.5	1
-7.93	13.5	1
-19.08	13.5	1
5.36	13.5	1

References

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Fisher, R. A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

Green, B. F., and Tukey, J. W. (1960). Complex analyses of variance: general problems. Psychometrika 25 127–152.

Hodges, J. S., Cui, Y., Sargent, D. J., and Carlin, B. P. (2005). Smoothed ANOVA. Technical report, Department of Biostatistics, University of Minnesota.

Kirk, R. E. (1995). Experimental Design: Procedures for the Behavioral Sciences, third edition. Brooks/Cole.

Plackett, R. L. (1960). Models in the analysis of variance (with discussion). Journal of the Royal Statistical Society B 22, 195–217.

Snedecor, G. W., and Cochran, W. G. (1989). Statistical Methods, eighth edition. Iowa State University Press.

Spiegel, Murray R. (1992). Schaum's Outline of Theory and Problems of Statistics (2nd edition). New York: McGraw-Hill. Chapter 16.

LECTURE SIX

Two-Way Analysis of Variance (2)

Introduction

This lecture looks at the two factor classification problem. For two factor classification, we have two different types of treatments. In table form, one set of treatment will be list in the top row and the second set of treatments will be listed on the first column.

Objectives

At the end of this lecture, you should be able to

1. understand the concept of ***factor classification*** in a two-way analysis of variance
2. solve problems relating to factor classification.

Pre Test

1. What is a ***factor***?
2. What is factor classification?

CONTENT

The following is an example of a two-factor classification of analysis of variance

Assume the track coach of a local high school wishes to test, among three brands of running shoes, the best performing shoes for freshmen, sophomore, and senior students. The numbers in the table below, represent the average running times for the 100 yard dash.

	Brand A (seconds)	Brand B (seconds)	Brand C (seconds)
Freshmen	10.80	10.27	10.63
Sophomore	11.66	11.84	11.11
Senior	10.71	12.06	12.06

Since we have two different types of treatments, we have two null hypothesis tests:

$H_0^{(1)}$: There is no difference between brands of running shoes (between columns).

$H_0^{(2)}$: There is no difference between class of runners (between rows)

In deciding to reject $H_0^1 or H_0^2$, we first must learn how to compute four types of variations:

1. Total variation (S_T^2)
2. Variation between rows (S_R^2)
3. Variation between columns (S_C^2)
4. Variation due to chance (S_E^2)

   The equation for the relationship between these four variations $S_E^2 = S_T^2 - S_E^2 - S_C^2$.

Solved problem 6.1

a. total variation.
b. variation between rows.
b. variation between columns.
a. random variation.

Solutions:

(a)

Step 1: From the table above compute the row totals, column totals, row means, column means, table total and mean of the table:

	Brand A (seconds)	Brand B (seconds)	Brand C (seconds)	Row Total	Mean Total
Freshmen	10.80	10.27	10.63	31.7	10.57
Sophomore	11.66	11.84	11.11	34.61	11.54
Senior	10.71	12.06	12.06	34.83	11.61
Column Total	33.17	34.17	33.80	**Table Total** 101.14	**Table Mean** 11.24
Column Mean	11.06	11.39	11.27		

Step 2: Compute the following table by subtracting the grand mean from each value of the table and squaring these differences:

	Brand A (seconds)	Brand B (seconds)	Brand C (seconds)
Freshmen	$(10.80-11.24)^2=0.19$	$(10.27-11.24)^2=0.94$	$(10.63-11.24)^2=0.37$
Sophomore	$(11.66-11.24)^2=0.18$	$(11.84-11.24)^2=0.36$	$(11.11-11.24)^2=0.02$
Senior	$(10.71-11.24)^2=0.28$	$(12.06-11.24)^2=0.67$	$(12.06-11.24)^2=0.68$

Step 3: Total variation is the total of all these numbers:

$S_T^2=3.69$

b.

The formula for the variation between rows is by summing the values in the following table:

$(\text{Row Mean - Table Mean})^2$
$(10.57-11.24)^2=0.45$ $(11.54-11.24)^2=0.19$ $(11.61-11.24)^2=0.14$
Sum= 0.68
$S_R^2=cSum=3(0.68)=2.04$

c.

The formula for the variation between columns is by summing the values in the following table:

$(\text{Column Mean - Table Mean})^2$
$(11.06-11.24)^2 = 0.032$ $(11.39-11.24)^2 = 0.023$ $(11.27-11.24)^2 = 0.00$
Sum = 0.055
$S_C^2 = rSum = 3(0.055) = 0.165$

d.

To compute the random variation (S_E^2), we use the formula:

$$S_E^2 = S_T^2 - S_E^2 - S_C^2 = 3.69 - 2.04 - 0.165 \approx 1.49$$

Solved Problems

Solved Problem 6.2 : A large petroleum company wishes to test five new gasoline additives for increased fuel efficiency. Their research department purchased 15 new model sedans and drove each car 100 miles, over the same track. Each additive was mixed with three octane gasolines: regular , premium and super. The following table is the mileage recorded for each car in this test. Here, mileage is measured for each car as to the number of gallons consumed to travel 100 miles.

Additve octane	**A**	**b**	**C**	**D**	**E**
Regular	5.11	5.23	6.13	5.00	6.18
Premium	4.76	5.00	4.02	5.11	4.87
Super	4.01	4.50	3.98	4.98	5.00

From this table, compute:

a. total variation
b. variation between rows.
c. variation between columns.
d. random variation.

Solutions:

(a)

Step 1: From the table above, compute the row totals, column totals, row means, column means, grand total and mean of the grand total:

Additive octane	**A**	**b**	**C**	**D**	**E**	**Row Total**	**Row Mean**
Regular	5.11	5.23	6.13	5.00	6.18	27.65	5.53
Premium	4.76	5.00	4.02	5.11	4.87	23.76	4.75
Super	4.01	4.50	3.98	4.98	5.00	22.47	4.49
Column Total	13.88	14.73	14.13	15.09	16.05	**Table Total**	**Table Mean**
Column Mean	4.63	4.91	4.71	5.03	5.35	73.88	4.93

Step 2: Compute the following table by subtracting the table mean value of the and squaring these differences:

Total variation is the total of all these numbers:

$S_T^2 = 5.97$

b.

The formula for variation between rows is by summing the values in the following table:

(Row Means – Table Means) ^{2}
$(5.53-4.93)^2 = 0.36$
$(4.75-4.93)^2 = 0.03$
$(4.49-4.93)^2 = 0.19$
$\text{Sum} \approx 0.58$
$S_R^2 = \text{c(sum)} = 5(0.58) = 2.90$

c.

The formula for variation between column is summing the values in the following table:

Additive octane	**A**	**B**	**C**	**D**	**E**
Regular	$(5.11-4.93)^2 \approx 0.03$	$(5.23-4.93)^2 \approx 0.09$	$(6.13-4.93)^2 \approx 1.44$	$(5.00-4.93)^2 \approx 0.01$	$(6.18-4.93)^2 \approx 1.57$
Premium	$(4.76-4.93)^2 \approx 0.03$	$(5.00-4.93)^2 \approx 0.01$	$(4.02-4.93)^2 \approx 0.82$	$(5.11-4.93)^2 \approx 0.03$	$(4.87-4.93)^2 \approx 0.00$
Super	$(4.01-4.93)^2 \approx 0.84$	$(4.45-4.93)^2 \approx 0.18$	$(3.98-4.93)^2 \approx 0.89$	$(4.98-4.93)^2 \approx 0.00$	$(5.00-4.93)^2 \approx 0.01$

(Column Means – Table Means)^2
$(4.63-4.93)^2 = 0.09$ $(4.91-4.93)^2 = 0.00$ $(4.71-4.93)^2 = 0.05$ $(5.03-4.93)^2 = 0.01$ $(5.35-4.93)^2 = 0.18$
Sum ≈ 0.33
S_C^2 = c(sum)= 3(0.33)= 0.99

d.

To compute the random variation (S_E^2), we use the formula

$$S_E^2 = S_T^2 - S_E^2 - S_C^2 = 5.79 - 2.90 - 0.99 \approx 2.08$$

Unsolved Problems with answers

Problem 1: A medical research laboratory wishes to test if there is a different drug that promote weight loss for women and men over 200

pounds. The following table is the resulting weight loss (in pounds) after 60 days.

Drug/Gender	A	B	C
MALE	22.45	20.65	24.11
FEMALE	27.22	28.00	28.11

From this table, compute

a. total variation.
b. variation between rows.
c. variation between columns.
d. random variation.

Answers:

a. 49.77
b. 43.44
c. 3.37
d. 2.96

Testing Hypothesis between rows and between columns using the F Distribution

We need to test two hypothesis:

$H_0^{(1)}$: There is no statistical difference between the columns.

$H_0^{(1)}$: There is no statistical difference between the rows.

To test $H_0^{(1)}$, we use the F distribution where

$$F = \frac{(r-1)S_C^2}{S_E^2}$$

With $d_2 = c-1$ and $d_1 = (r-1)(c-1)$ degrees of freedom.

To test $H_0^{(2)}$, we use the F distribution where $F = \frac{(c-1)S_r^2}{S_E^2}$

With $d_2 = r-1$ and $d_1 = (r-1)(c-1)$ degrees of freedom.

Example 1: For 6.2

a. Find F.

   Using the level of significance of 0.05 and 0.01, determine if there is a statistical difference between brands of running shoes.

b. Using the level of significance of 0.05 and 0.01 , determine if there is a statistical difference between class year.

Solutions:

a.

Here , we are testing across the columns.

Step 1: To find F we use the formula:

$$F = \frac{(r-1)S_C^2}{S_E^2}$$

Where

r = the number of rows

Step 2: From example 3.1, we computed:

$S_C^2 = 0.15,\ S_E^2 = 1.56.$

Since r = 3,

$$F = \frac{(r-1)S_C^2}{S_E^2} = \frac{(3-1)0.15}{1.56} \approx 0.19$$

Step 3: We have d_2 = c-1 = 3-1 =2 degrees of freedom and

d_1 =(r-1)(c-1) = (3-1)(3-1) = 4 degrees of freedom.

Step 4: Using the F distribution table for α =0.05, $F_{0.05}$ = 6.94.

Step 5: Since F = 0.019 < 6.94, we conclude there is no significant difference between running shoes.

Step 6: Using the F distribution table for α = 0.01, $F_{0.01}$=18

Step 7: Since F = 0.019 < 18, we conclude there is no significant difference between the shoes.

b.

Step 1: To find F we use the formula:

$$F = \frac{(c-1)S_R^2}{S_E^2}$$

Where

c= the number of columns

Step 2: From example 3.1, we computed:

$S_R^2 = 1.98,\ S_E^2 = 1.56.$

Since c = 3,

$$F = \frac{(c-1)S_C^2}{S_E^2} = \frac{(3-1)1.98}{1.56} \approx 2.54.$$

Step 3: We have d_2 = r-1 = 3-1 =2 degrees of freedom and

d_1 =(r-1)(c-1) = (3-1)(3-1) = 4 degrees of freedom.

Step 4: Using the F distribution table for α =0.05, $F_{0.05}$ = 6.94.

Step 5: Since F = 2.54 < 6.94, we conclude there is no significant difference between class years.

Step 6: Using the F distribution table for α = 0.01, $F_{0.01}$=18

Step 7: Since F = 0.019 < 18, we conclude there is no significant difference between the shoes.

Solved Problems

Problem 1: For Solved Problem 6.2 ,

a. Find F. Using the level of significance of 0.05 and 0.01, determine if there is a statistical difference between gasoline additives.
b. Find F Using the level of significance of 0.05 and 0.01, determine if there is a statistical difference between octanes.

Solutions:

a.

Here, we are testing across the columns.

Step 1: To find F, we use the formula:

$$F = \frac{(r-1)S_C^2}{S_E^2}$$

Where

r = the number of rows

Step 2: From 42.3 – solved problem 1, we computed:

$S_C^2 = 0.99,\ \ S_E^2 = 2..08.$

Since r = 3,

$$F = \frac{(r-1)S_C^2}{S_E^2} = \frac{(3-1)0.99}{2.08} \approx 0.94$$

Step 3: We have d_2 = c-1 = 5-1 =4 degrees of freedom and

d_1 =(r-1)(c-1) = (3-1)(5-1) = 8 degrees of freedom.

Step 4: Using the F distribution table for α =0.05, $F_{0.05}$ = 3.84.

Step 5: Since F = 0.95 <3.84, we conclude there is no significant difference between gasoline additives.

Step 6: Using the F distribution table for α = 0.01, $F_{0.01}$=7.01

Step 7: Since F = 0.95 < 7.08, we conclude there is no significant difference between gasoline additives.

b.

Here, we are testing down rows.

Step 1: To find F we use the formula:

$$F = \frac{(c-1)S_R^2}{S_E^2}$$

Where

c= the number of columns

Step 2:

$S_R^2 = 2.9,\ S_E^2 = 2.08.$

Since c = 5,

$$F = \frac{(c-1)S_C^2}{S_E^2} = \frac{(5-1)2.90}{2.08} \approx 5.57.$$

Step 3: We have d_2 = r-1 = 3-1 =2 degrees of freedom and

d_1 =(r-1)(c-1) = (3-1)(5-1) = 8 degrees of freedom.

Step 4: Using the F distribution table for α =0.05, $F_{0.05}$ = 4.46.

Step 5: Since F = 5.57 < 4.46, we conclude there is a significant difference between octane which affects mileage.

Step 6: Using the F distribution table for α = 0.01, $F_{0.01}$=8.65

Step 7: Since F = 5.57 < 8.65, we conclude there is no significant difference in octane.

Unsolved Problems with answers

Problem 1: Problem 6.2**,**

a. Find F. Using a level of significance of 0.05 and 0.01, determine if there is a statistical difference between diet drugs.
b. Using a level of significance of 0.05 and 0.01, determine if there is a statistical difference between men and women.

Answers:

a. Since F = 1.14, using a level of significance of 0.05 and 0.01, we conclude there is significant difference between diet drugs.
b. F = 29.35. Since F = 29.35 > 19, we conclude there is a significant difference in weight loss between men and women.

Summary

In this lecture, you have been able to

1. learn about factor classification in a two-way analysis of variance.
2. solve problems relating to facto classification.

Post Test

1. On your own, solve the problems tackled in this lecture.
2. A physiologist reports an investigation of potential plant hormones. He reports the following averages for lengths of 20 stem segments treated.

   Compound X 1.18
   Compound Y 1.17
   Compound Z 1.15
   Control 1.14

   The conclusion is that there are no treatment differences. Are you satisfied with this conclusion? Why or why not?

3. Ms. Romeo teaches three sections of beginning latin at a local high school. The following table gives the average grade over the past five years.

Latin Section/ Average grade	Section 1	Section 2	Section 3
1991	78.20	76.50	79.20
1992	79.10	77.90	79.87
1993	79.60	77.60	78.66
1994	80.10	81.78	79.10
1995	81.10	80.10	85.78

Using one – factor classification to see if there is a significant difference between sections, find:

a. Total variation (S_T^2).
b. Variation between treatments (S_B^2).
c. Variation within treatment (S_W^2).
d. F.
e. Using a 0.05 level of significance, is there a difference in grades between the class sections. Applying two factor classification to the above table data, find
f. Total variation.
g. Variation between rows
h. Variation between columns.
i. Random variation.
j. F. Using a 0.05 level of significance, is there a difference in grades between the class sections?
k. F. Using a 0.05 level of significance, is there a difference in grades between the five years?

References

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Fisher, R. A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

Green, B. F., and Tukey, J. W. (1960). Complex analyses of variance: general problems. Psychometrika 25 127–152.

Hodges, J. S., Cui, Y., Sargent, D. J., and Carlin, B. P. (2005). Smoothed ANOVA. Technical report, Department of Biostatistics, University of Minnesota.

Kirk, R. E. (1995). Experimental Design: Procedures for the Behavioral Sciences, third edition. Brooks/Cole.

Plackett, R. L. (1960). Models in the analysis of variance (with discussion). Journal of the Royal Statistical Society B 22, 195–217.

Snedecor, G. W., and Cochran, W. G. (1989). Statistical Methods, eighth edition. Iowa State University Press.

Spiegel, Murray R. (1992). Schaum's Outline of Theory and Problems of Statistics (2nd edition). New York: McGraw-Hill. Chapter 16.

LECTURE SEVEN

Two-Way Analysis of Variance (3)

Introduction

This lecture is a continuation of the last two lectures. Here, we split the two-way model into ***pooled*** and ***partitioned*** errors. We also introduce the concept of ***contrast***.

Objectives

At the end of this lecture, you should be able to learn more about two-way analysis of variance with respect to factor classification, interactions and contrast.

Pre Test

What do you understand by a ***contrast***?

CONTENT

The full model for a two way, K_1-*by*-K_2 repeated measure ANOVA, with $P = K_1 K_2$ measurements taken from each of *N* subjects, can be written as

$$y_{nkl} = \tau_{kl} + \pi_n + e_{nkl}$$

Where $k = 1 \ldots K_1$ and $l = 1 \ldots K_2$ index the levels of factor A and factor B respectively. Here we can think of indicator function $k = g_k(i), l = g_l(i)$ and $n = g_n(i)$ that return the levels of both factors and subject identity for

the *i*th scan. Again, π_n are subject effects and e_{nkl} are residual errors. This equation can be written in matrix form

$$y = XB + e$$

Where $X = [I_p \otimes 1_N, 1_N \otimes I_P]$ is the design matrix and $\beta = [\tau_{kl}, \pi_n]^T$ are the regression coefficients. This is identical to the one-way within-subject design but with P instead of K treatment effects.

Model and Null Hypotheses

The difference between pooled and partitioned error models can be expressed by specifying the relevant models and null hypotheses

Pooled Errors

For the main effect of A we test the null hypothesis $\mathrm{H}_0 : \tau_q^A = 0$ for all *q*. Similarly for the main effect of B. For an interaction we test the null hypothesis

$$\mathrm{H}_0 : \tau_{qr}^{AB} = 0 \text{ for all } q, r.$$

For example, for the 3-by-3 design there are *q* = 1…2 differential effects for factor A and *r* = 1…2 for factor B. the pooled error model therefore has regression coefficients

$$\hat{\beta} = [\tau_1^A, \tau_2^A, \tau_1^B, \tau_2^B, \tau_{11}^{AB}, \tau_{12}^{AB}, \tau_{21}^{AB}, \tau_{22}^{AB}, m, \pi_n]^T \qquad (7.1)$$

For the main effect of A we test the null hypothesis $\mathrm{H}_0 : \tau_1^A = \tau_2^A = 0$. For the interaction we test the null hypothesis $\mathrm{H}_0 : \tau_{11}^{AB} = \tau_{12}^{AB} = \tau_{21}^{AB} = \tau_{22}^{AB} = 0$.

Partitioned Errors

Partitioned error model can be implemented by applying contrasts to the data, and then creating a separate model (that is, separate GLM analysis) to test each effect. In other words, a two-stage approach can be taken. The first stage is to create contrasts of the condition for each subject, and the second stage is to put these contrasts or 'summary statistics' into a model with a block-diagonal design matrix.

For partitioned errors we first transform our data set y_{nkl} into a set of differential effects for each subject and then model these as a GLM. This set of differential effect for each subject is created using appropriate contrast at the 'first level'. The models that we describe below then correspond to a 'second-level' analysis. The difference between first and second level analyses is described in the previous chapter on random effects analysis.

To test for the main effect of A, we first create the new data points ρ_{nq} which are the differential effects between the levels in A for each subject n (see e.g. section). We then compare the full model

$$\rho_{nq} = \tau_q^A + e_{nq}$$

to the reduced model

$$\rho_{nq} = e_{nq}.$$

We are therefore testing the null hypothesis, $\mathrm{H}_0 : \tau_q^A = 0$ for all q.

Similarly for the main effect of B. To test for an interaction we first create the new data points ρ_{nqr} which are the differences of differences of differential effects for each subject. For a K_1 and K_2 ANOVA there will be $(K_1 - 1)(K_2 - 2)$ of these. We then compute the full model

$$\rho_{nq} = \tau_q^{AB} + e_{nqr}$$

To reduce model $\rho_{nqr} = e_{nqr}$. We are therefore testing the null hypothesis $\mathrm{H}_0 : \tau_{qr}^{\mathbf{AB}} = 0$ for all q, r.

For example, for a 3-by-3, there are q= 1..2 differential effects for factor A and r= 1..2 for factor B. We first create the differential ρ_{nq}.To test for the main effect of A we compare the full model

$$\rho_{nq} = \tau_1^A + \tau_2^A + e_{nq}$$

To the reduce model $\rho_{nq} = e_{nq}$. We are therefore testing the null hypothesis, $\mathrm{H}_0 : \tau_{qr}^A = \tau_2^A = 0$ Similarly for the main effect of B.

To test for an interaction we first create the differences of differential effects for each subject. There are 2×2 = 4 of these. We then compare the full model

$$\rho_{nqr} = \tau_{11}^{AB} + \tau_{12}^{AB} + \tau_{21}^{AB} + \tau_{22}^{AB} + e_{nqr}$$

To the reduced model $\rho_{nq_r} = e_{nq_r}$. We are therefore testing the null hypothesis, $H_0 : \tau_{11}^{AB} = \tau_{12}^{AB} = \tau_{21}^{AB} = \tau_{22}^{AB} = 0$ ie. that all the 'simple' interactions are zero.

Note how

1. the degrees of freedom have been reduced, being split equally among the three effects,
2. there is no need for a nonsphericity correction in this case(since $K_1 = K_2 = 2$), and
3. the p-value for some of the effects have decreased, while those for the other effects have increased.

Whether p-value increase or decrease depends on the nature of the data (particularly correlation between conditions across subjects), but in many real dataset partitioned error comparisons yield more sensitive inferences. This is why, for repeated-measures analyses, the preferred over using a pooled error [Howell 1992]. But the partitioned error approach requires a new model to be specified for every effect we want to test.

Numerical Example

Pooled Error

Consider a 2X2 ANOVA of the same data used in the previous example, with $K_1 = K_2 = 2, P = K_1 K_2 = 4, N = 12, J = PN = 48$. Assuming that the four columns/conditions are ordered;

1	2	3	4	
A_1B_1	A_1B_2	A_2B_1	A_2B_2	(7.2)

Where A_1 represents the first level of factor A, B_2 represents the second level of factor B e.t.c, and the row are ordered; all subjects data for cell A_1B_1; all for A_1B_2 and so on.

Main effects are not really meaningful in the presence of a significant interaction. In the presence of an interaction. In the presence of an interaction, one does not normally report the main effects, but proceeds by

testing the differences between the level of one factor for each of the level of the other factor in the interaction (so-called *simple effects*). In the case, the presence of a significant interaction could be used to justify further simple effect contrasts (see above), e.g. the effect B at the first and second levels of A are given by the contrasts $e = [1, -1, 0, 0]$ and $e = [0, 0, 1, -1]^T$.

Equivalent result would be obtained if the design matrix were rotated so that the first three columns reflect the experimental effects plus a constant term in the fourth column (only the first columns would be rotated). This is perhaps a better conception of the ANOVA approach, reflecting the conception of factorial designs in terms of the experimental effects rather than the individual conditions. This rotation is achieved by setting the new design matrix.

$$X_r = X \begin{bmatrix} C^T & 0_{4,12} \\ 0_{12,4} & I_{12} \end{bmatrix} \tag{7.3}$$

Where

$$C^T = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{7.4}$$

Notice that the rows of C^T are identical to the contrasts for the main effects and interactions plus a constant term. The three experimental effects can now be tested by the contrasts weight $[1, 0, 0, 0]^T, [0, 1, 0, 0]^T, [0, 0, 0, 1]^T$ (again, padded with zeros).

In this example, each factor only has two levels which results in one-dimensional contrasts for testing main effects and interactions.

Summary

In this lecture, you have been able to learn how to solve more two-way analysis of variance problems involving factor classifications, interactions and ***contrast***.

Post Test

1. Discuss the differences and/or similarities between
   a. factor classification, interactions and contrast.
   b. pooled errors and partitioned errors
2. Write down the appropriate hypothesis to be used in testing
   a. the main effect
   b. the interaction effect

   in the case of pooled errors.
3. Describe how you would implement a partitioned error model.

References

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Green, B. F., and Tukey, J. W. (1960). Complex analyses of variance: general problems. Psychometrika 25 127–152.

Plackett, R. L. (1960). Models in the analysis of variance (with discussion). Journal of the Royal Statistical Society B 22, 195–217.

Snedecor, G. W., and Cochran, W. G. (1989). Statistical Methods, eighth edition. Iowa State University Press.

LECTURE EIGHT

Statistical Inference Theory

Introduction

Solving any statistical problem will involve carrying out an inference. This lecture introduces the concept of statistical inference in an analysis of variance problem.

Objective

At the end of this lecture, you should be able to conduct simple inference on problems in analysis of variance.

Pre Test

1. What is Statistical Inference?
2. True or False? If False, correct it.
   In a one-way classification ANOVA, when the null hypothesis is false, the probability of obtaining an F-ratio exceeding that reported in the F table at the .05 level of significance is greater than .05.

CONTENT

Analysis of Variance (ANOVA) is a method of testing the hypothesis of the difference between three or more population means. In the case where we are testing three population means, $H_O : \mu_1 = \mu_2 = \mu_3$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$ or

$\mu_1 \neq \mu_3$ or

$\mu_2 \neq \mu_3$

To use ANOVA, we need to make the following assumptions

1. The populations are assumed to be (approximately) normally distributed.
2. The populations have equal variances.
3. The samples from each population are independent of each other.

The following example is a typical problem to be solved by ANOVA:
Assume the track coach of a local high school wishes to test, among three brands of running shoes, the best performing shoes. He decides to select 15 track runners to run the 100 yard dash. In this race, each brand is worn by five runners. The following tables give the timing outcome of the race of each brand:

Brand A (seconds)	**Brand B (seconds)**	**Brand C (seconds)**
11.30	12.17	11.43
12.06	12.14	10.11
11.78	11.66	11.66
12.11	10.99	10.28
10.98	10.89	11.55

One- factor classification

Problems of this type are called one-factor classification because only one variable (factor) is considered: the brand of running shoes. The variables are called treatments. For the above example we have three treatments, the three brands of shoes. In this lesson we will study both one-factor classification of ANOVA where the samples are of equal size for each treatment.

In deciding to reject H_O or not, we first must learn to compute three types of variations:

1. Total variation (S_T^2)
2. Variation within treatments (S_W^2)
3. Variation between treatments (S_B^2)

The equation for the relationship between total variation, within variation and between variation is $(S_W^2 = S_T^2 - S_B^2)$

Example 8.1: For the example above, compute

a. total variation.
b. variation between treatments.
c. variation within treatments.

Solutions:

a.

Step 1: Compute the mean $\overline{X}$ for all the numbers in the table:

Brand A (seconds)	**Brand B (seconds)**	**Brand C (seconds)**
11.30 12.06 11.78 12.11 10.98	12.17 12.14 11.66 10.99 10.89	11.43 10.11 11.66 10.28 11.55
$\overline{X} = 11.41$		

Step 2: **a.** Subtract $\overline{X}$ from each of the number in the table.

b. Square each of these.

c. The total variation of the sum of the values computed in the table:

Brand A (seconds)	**Brand B (seconds)**	**Brand C (seconds)**
$(11.30-11.41)^2 \approx 0.01$	$(12.17-11.41)^2 \approx 0.58$	$(11.43-11.41)^2 \approx 0.0$
$(12.06-11.41)^2 \approx 0.43$	$(12.14-11.41)^2 \approx 0.54$	$(10.11-11.41)^2 \approx 1.68$
$(11.78-11.41)^2 \approx 0.14$	$(11.66-11.41)^2 \approx 0.06$	$(11.66-11.41)^2 \approx 0.06$
$(12.11-11.41)^2 \approx 0.49$	$(10.99-11.41)^2 \approx 0.17$	$(10.28-11.41)^2 \approx 1.27$
$(10.98-11.41)^2 \approx 0.18$	$(12.11-11.41)^2 \approx 0.49$	$(11.55-11.41)^2 \approx 0.02$

Sum of Table Values: $S_T^2 = 5.92$

b.

Step 1: Compute the mean $\overline{X}$ for each column

Brand A (seconds)	**Brand B (seconds)**	**Brand C (seconds)**
11.30 12.06 11.78 12.11 10.98	12.17 12.14 11.66 10.99 10.89	11.43 10.11 11.66 10.28 11.55
$\overline{X}$ = 11.65	$\overline{X}$ =11.57	$\overline{X}$ =11.00

Step 2: Subtract $\overline{X}$ (computed in step 1) from each of the $\overline{X}$ s' in the above table.

Square each of these differences.

The between variation is the sum of the values from b multiplied by the number of rows.

$(11.66-11.41)^2 \approx 0.06$ $(11.57-11.41)^2 \approx 0.03$ $(11.01-11.41)^2 \approx 0.16$
S_B^2= 5x(column sum)= 5(0.25)= 1.25

a.

To compute the within variation, we use the formula:

$S_W^2 = S_T^2 - S_B^2 = 5.92 - 1.25 = 4.67$

Solved Problems

Solved Problem 1: A large petroleum company wishes to test five new gasoline additives for increase fuel efficiency. Their researched department purchased 35 new modern sedans and drove each car 100 miles, over the same track. Each additive was mixed with the gasoline of seven sedans.

The following table is the mileage recorded for each car in the test. Here, mileage is measured for each car as the number of gallons consumed to travel 100 miles

Additive	**A**	**B**	**C**	**D**	**E**
Number of gallons	5.56	4.97	5.01	4.11	6.87
	5.87	6.00	4.25	5.11	5.76
	5.00	6.77	5.98	5.12	6.02
	5.73	6.09	7.11	6.11	7.00
	3.99	4.78	5.55	6.00	6.51
	4.12	5.00	5.44	5.51	6.09
	4.78	5.11	5.54	7.00	7.11

Compute:

a. total variations.
b. variation between treatments.
c. variation within treatments.

Solutions:

a.

Step 1: Compute the mean $\overline{X}$ for all the numbers in the table:

A	**B**	**C**	**D**	**E**
5.56	4.97	5.01	4.11	6.87
5.87	6.00	4.25	5.11	5.76
5.00	6.77	5.98	5.12	6.02
5.73	6.09	7.11	6.11	7.00
3.99	4.78	5.55	6.00	6.51
4.12	5.00	5.44	5.51	6.09
4.78	5.11	5.54	7.00	7.11

$\overline{X}$ = 5.63

Step 2: **a.** Subtract $\overline{X}$ from each of the number in the table.

b. Square each of these.

c. The total variation of the sum of the values computed in the table:

A	B	C	D	E
$(5.56-5.63)^2 \approx 0.00$	$(4.97-5.63)^2 \approx 0.43$	$(5.01-5.63)^2 \approx 0.38$	$(4.11-5.63)^2 \approx 2.30$	$(6.87-5.63)^2 \approx 1.54$
$(5.87-5.63)^2 \approx 0.06$	$(6.00-5.63)^2 \approx 0.14$	$(4.25-5.63)^2 \approx 1.90$	$(5.11-5.63)^2 \approx 0.27$	$(5.76-5.63)^2 \approx 0.01$
$(5.00-5.63)^2 \approx 0.39$	$(6.77-5.63)^2 \approx 1.30$	$(5.98-5.63)^2 \approx 0.12$	$(5.12-5.63)^2 \approx 0.26$	$(6.02-5.63)^2 \approx 0.15$
$(5.73-5.63)^2 \approx 0.01$	$(6.09-5.63)^2 \approx 0.21$	$(7.11-5.63)^2 \approx 2.20$	$(6.11-5.63)^2 \approx 0.23$	$(7.00-5.63)^2 \approx 1.88$
$(3.99-5.63)^2 \approx 2.68$	$(4.78-5.63)^2 \approx 0.72$	$(5.55-5.63)^2 \approx 0.00$	$(6.00-5.63)^2 \approx 0.14$	$(6.51-5.63)^2 \approx 0.78$
$(4.12-5.63)^2 \approx 2.27$	$(5.00-5.63)^2 \approx 0.39$	$(5.44-5.63)^2 \approx 0.04$	$(5.51-5.63)^2 \approx 0.01$	$(6.09-5.63)^2 \approx 0.21$
$(4.78-5.63)^2 \approx 0.72$	$(5.11-5.63)^2 \approx 0.27$	$(5.54-5.63)^2 \approx 0.01$	$(7.00-5.63)^2 \approx 1.88$	$(7.11-5.63)^2 \approx 2.20$

Sum of the table values: $S_T^2 = 26.15$

b.

Step 1: Compute the mean $\overline{X}$ for each column:

A	B	C	D	E
5.56	4.97	5.01	4.11	6.87
5.87	6.00	4.25	5.11	5.76
5.00	6.77	5.98	5.12	6.02
5.73	6.09	7.11	6.11	7.00
3.99	4.78	5.55	6.00	6.51
4.12	5.00	5.44	5.51	6.09
4.78	5.11	5.54	7.00	7.11
$\overline{X} \approx 5.01$	$\overline{X} \approx 5.53$	$\overline{X} \approx 5.55$	$\overline{X} \approx 5.57$	$\overline{X} \approx 6.48$

Step 2: Subtract $\overline{X}$ (computed in step 1) from each of the $\overline{X}$ s' in the above table.

Square each of these differences.

The between variation is the sum of the values from b multiplied by the number of rows.

$(5.01-5.63)^2 \approx 0.39$
$(5.53-5.63)^2 \approx 0.01$
$(5.55-5.63)^2 \approx 0.01$
$(5.57-5.63)^2 \approx 0.00$
$(6.48-5.63)^2 \approx 0.73$
S_B^2= 7x(column sum)= 7(1.13)= 7.91

C. To compute the within variation, we use the formula:

$S_W^2 = S_T^2 - S_B^2$ = 26.15 - 7.91 = 18.24.

Summary

At the end of this lecture, you have been able to conduct simple inference on problems relating to analysis of variance.

Post Test

1. A survey of 114 men and 126 women produced the result that the mean amount of chicken soup consumed by the men in a month's time was .67 liters, compared with a mean of .54 liters for the women. The variance of the chicken soup consumption for the men was 25% greater than that for the women. A 95% confidence interval for the difference between the means (men's mean - women's) was found to be -.07 to +.33 liters. At the .05 level, is the difference in mean chicken soup consumption between sexes statistically significant?

a. no
b. yes
c. Can't tell from the data given

2. A carefully designed experiment has just been concluded. Execution of the experiment was flawless. Unfortunately use of ALPHA (.05) indicated no significant differences among treatments.
   What useful information can you supply future investigators when you report on this experiment?
   Indicate agreement with yes or disagreement with no for each of the following items.
   a. No useful information. These are negative results and there is nothing useful to report.
   b. The estimated variance (and its df) can be useful to future investigators.
   c. A careful description of experimental conditions and treatments may be useful as an indication of circumstances where responses are about the same.
   d. Significant differences can be reported by changing the Type I error rate from .05 to .10, .20 or whatever is needed to declare significance.
3. ***Note***: Items i. thru iii. are based on a teaching experiment involving four elementary statistics classes. Below are scores for 24 students who took the same final examination.

Statistics Cookbook	**Statistics with Humor**	**Statistics Made Useful**	**Statistics in Story Form**
78	51	64	54
78	57	54	61
79	64	61	79
70	75	66	69
83	42	57	69
74	83	71	65
462	372	373	397

Suppose further that calculations yield SS(total) = 800, and SS(between) = 300.

i. What is the observed value of the statistic one computes to test H(0): MU(1) = MU(2) = MU(3) = MU(4) against H(1): not all 4 means are equal?
   a. 2.5 b. 5.2 c. 4.0 d. 11.1 e. None of these

ii. If ALPHA = .01, then the critical value of the statistic is
   a. 3.10 b. 4.94 c. 8.66 d. 26.7 e. None of these

iii. Suppose that the observed value of the statistic in Item i. is 5.6 while the critical value in Item ii. is 7.21. With only this information, which of the following conclusions is most logical?
   a. The four populations do not all have the same mean.
   b. The four populations have the same mean.
   c. The four populations do not all have the same mean. Statistics Cookbook and Statistics in Story Form produce higher means than the other two books.
   d. Statistics Cookbook has the highest population mean.
   e. The four population means may be different but these samples fail to demonstrate any difference.

4. Solve the examples illustrated in this lecture.

References

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Fisher, R. A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

Green, B. F., and Tukey, J. W. (1960). Complex analyses of variance: general problems. Psychometrika 25 127–152.

Kirk, R. E. (1995). Experimental Design: Procedures for the Behavioral Sciences, third edition. Brooks/Cole.

Snedecor, G. W., and Cochran, W. G. (1989). Statistical Methods, eighth edition. Iowa State University Press.

Spiegel, Murray R. (1992). Schaum's Outline of Theory and Problems of Statistics (2nd edition). New York: McGraw-Hill. Chapter 16.

LECTURE NINE

Hypotheses Testing

Introduction

In analysis of variance, like other statistical models, it is important to test assertions and claims. In this lecture, we aim at testing some conjectures put forward in an analysis of variance problem.

Objective

At the end of this lecture, you should be able to test simple hypothesis in an analysis of variance problem.

Pre Test

1. What do you understand by the term ***critical region***?
2. Suppose the critical region for a certain test of hypothesis is of the form F > 9.48773 and the computed value of F from the data is .86. (F refers to an F statistic.) Then:
   a. H_0 should be rejected.
   b. H_A is two-tailed.
   c. The significance level is given by the area to the right of 9.48773 under the appropriate F distribution.
   d. None of these.

CONTENT

Unsolved problems with answers

Problem 9.1: A medical student research laboratory t wishes to test if there is a difference between three different drugs that promote weight loss for women over 200 pounds. The client randomly divide up fifteen over- weight into three equal groups. Each group take only one the drugs. The following table is the resulting loss (in pounds) after 60 days:

Drug A	**Drug B**	**Drug C**
21.30	22.11	33.43
32.06	19.18	29.00
21.78	17.66	21.26
22.11	19.79	30.58
30.98	20.89	21.55

Compute:

a. total variation.
b. variation between treatments.
c. variation within treatments.

Answers:

a. $S_T^2 = 394$
b. $S_B^2 = 145$
c. $S_W^2 = 248$

Testing Hypothesis on Means using the F Distribution

To test the null and alternative hypothesis:

$H_O : \mu_1 = \mu_2 = \mu_3 = ... = \mu_n$

$H_1 : \mu_1 \neq \mu_2$

or

$\mu_1 \neq \mu_3$

Or

$\mu_2 \neq \mu_3, etc,$

We use F distribution

Using the values from analysis of variance, we need to test the F distribution for the value

$$F=\frac{S_B^2 c(r-1)}{S_W^2(c-1)}, \text{ where}$$

c = the number of treatments (number of columns of the table)

r = the sample size for each treatment (number of rows of the tables)

$d_2 =$ c-1 degrees of freedom

$d_1 =$ c(r-1) degrees of freedom

Example 1: For Problem 8.1

a. State H_0 and H_1.

b. Compute F.

c. Find $F_{0.05}$

Would you reject H_0?

State your conclusions.

d. Find $F_{0.01}$

Would you reject H_0?

State your conclusions.

Solutions:

a.

$H_O : \mu_A = \mu_B = \mu_C$

H_a : at least one of the μ values is different from the other two.

b. From Example 1.

$S_B^2 = 1.25$

$S_W^2 = 4.67$

C=3 and r=5

$$F=\frac{S_B^2 c(r-1)}{S_W^2(c-1)} = \frac{(1.25)3(5-1)}{(4.67)(3-1)} = \frac{15}{9.34} \approx 1.61$$

c.

$d_2 =$ c-1 = 3 – 1 = 2

$d_1 =$ r (c-1)= 3(5-1) = 12

From the distribution table for 0.05, we find $F_{0.05} = 3.89$

Since F=1.61 < 3.89 , H_O is not rejected. For a level of significance of 0.05 we have not statistical basis to conclude that the make of the running shoes improve the performance of the runners.

d.

$d_2 =$ c-1 = 3 – 1 = 2

$d_1 =$ r (c-1)= 3(5-1) = 12

From the F distribution table for 0.01, we find $F_{0.01} = 6.93$.

Since F=1.61 < 6.93 , H_O is not rejected. For a level of significance of 0.01 we have not statistical basis to conclude that the make of the running shoes improve the performance of the runners.

Solved Problems

Solved Problem 9.1: For Problem 9.1

a. State H_0 and H_1.

b. Compute F.

c. Find $F_{0.05}$

Would you reject H_0 ?

State your conclusions.

d. Find $F_{0.01}$

Would you reject H_0 ?

State your conclusions.

Solutions:

a.

$H_O : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$

H_a : at least one of the μ values is different from the other two.

b. From Solved Problem 9.1.

$S_B^2 = 7.91$

$S_W^2 = 18.24$

C=5 and r=7,

F= $\frac{S_B^2 c(r-1)}{S_W^2 (c-1)} = \frac{(7.91)3(7-1)}{(18.24)(5-1)} = \frac{237.3}{72.96} \approx 3.25.$

c.

$d_2 =$ c-1 = 5 – 1 = 4

$d_1 =$ r (c-1)= 5(7-1) = 30

From the F distribution table for 0.01, we find $F_{0.01} = 4.02.$

Since F=3.25 < 4.02 , H_O is not rejected. For a level of significance of 0.01 we have no statistical basis to conclude that the make of the running shoes improve the performance of the runners.

Unsolved problems with answers

Solved Problem 1: For unsolved Problem 9.1

a. State H_0 and H_a.

b. Compute F.

c. Find $F_{0.05}$

Would you reject H_0?

State your conclusions.

d. Find $F_{0.01}$

Would you reject H_0?

State your conclusions.

Answers:

a. $H_O : \mu_A = \mu_B = \mu_C$

H_a :at least one of the μ values is different from the other two.

b. F= 3.49

c. $F_{0.05} = 3.89.$

Since F=3.49 < 3.89 , we do not reject H_O. There is no statistical basis for assuming among the three diet drugs for reducing weight.

d. $F_{0.01} = 6.93.$

Since F=3.49 < 6.93 , we do not reject H_O. There is statistical basis for assuming among the three diet drugs for reducing weight.

Summary

In this lecture, you have been able to test simple hypothesis in an analysis of variance problem.

Post Test

1. Consider the following ANOVA table:

Source	SS	df
-----	--	--
Between	30.5	4
Within		
Total	165.0	99

What decision would be made regarding H_0: population means are equal?

a. Reject H_0 at the .05 level
b. Fail to reject H_0 at the .01 level
c. Insufficient information is given to answer

2. Samples of size 5 are taken from 3 populations and the following analysis of variance table found. Test the hypothesis that the three populations have the same means.

Source	d.f.	M.S.	F
------	----	----	-
Between means		350	
Within samples		100	
Total			

a. F = 3.5 so hypothesis is rejected at 5% level
b. F = 3.5 so hypothesis is not rejected at 5% level
c. F = 3.5 so hypothesis is not rejected at 10% level
d. F = 3.5 so hypothesis is rejected at 1% level
e. Do not have enough information to perform test.

3. A fisheries researcher wishes to conclude that there is a difference in mean weights of three species of fish caught in a large lake near Lincoln, Nebraska. The data are as follows: (Use ALPHA = .05.)

	SPECIES	

X	Y	Z
1.5	1.5	6.0
4.0	1.0	4.5
4.5	4.5	4.5
3.0	2.0	5.5

ANOVA Table (incomplete):

Source of Variation	SS	df	MS	F
Between Groups	17.04	2	8.52	
Within Groups	14.19	9	1.58	
TOTAL	31.23	11		

i. The null hypothesis is:
   a. H(O): BETA = 0
   b. H(O): MU = 0
   c. H(O): MU(X) = MU(Y) = MU(Z)
   d. H(O): BETA(X) = BETA(Y)= BETA(Z)

ii. The test statistic is:
   a. t(calc) = 2.52
   b. t(calc) = 3.09
   c. F(calc) = 1.20
   d. F(calc) = 5.40

iii. The critical value is:
   a. t(.05,9) = 2.262
   b. t(.10,9) = 1.833
   c. F(.05,2,9) = 4.26
   d. F(2.5,2.9) = 5.71

iv. What is your conclusion?
   a. Reject H(O) because F(calc) > F(crit), (at least 1 pair has different means).
   b. Reject H(O) because t(calc) > t(crit), (all means are different).
   c. Fail to reject H(O) because F(calc) < F(crit), (insufficient evidence that means are different).
   d. Fail to reject H(O) because t(calc) < t(crit), (means are equal).

4. A one-way classification analysis of variance is performed on experimental data for which there were 10 subjects in each of two groups. The .95 confidence interval around the difference YBAR(1) - YBAR(2) is -0.10 to 1.5. Which one of the following statements is true?
   a. The F ratio obtained in the analysis of variance was less than 4.41
   b. The F ratio obtained in the analysis of variance was greater than 8.28
   c. The true difference MU(1) - MU(2) must lie between -0.10 and 1.50.
   d. The best estimate of MU(1) - MU(2) possible from the results is 1.50.

5. A researcher assigns each of his interviewers a list of 7 families, drawn randomly from a region, to be interviewed. Each interviewer is instructed to administer a successful parenting scale (SPS) to each parent in his sample. The SPS scores, Y(i), are defined as ranging from 0 (no parenting skills deemed successful) to 100 (successful parenting skills consistently and skillfully applied). An interviewer returns with data for both parents. Use this data to test, using classical analysis of variance, at the 90% level of confidence, the hypothesis that "mothers are more likely to be successful parents".

Mothers	**Fathers**
Y(i)	Y(j)
68	63
72	48
48	30
54	52
83	55
92	41
87	57

MBAR = 72.00
FBAR = 49.43

ANOVA Table (incomplete):

Source of Variation	SS	df	MS	F(calc)
Between Groups	1783.143	1		
Within Groups	2391.714	12	199.310	
Total	**4174.857**	**13**		

6. Use one-way analysis of variance, with an F test, to test the hypothesis that "The wealthier a person, the more likely he will be relatively politically conservative," at the 90% level of confidence. Note that, for purposes of research, the researcher operationally defined "wealthy" as those with an annual income of ₦7,000,000,

while "poor" subjects received less than ₦50,000/year. Note, too, that the "political conservatism" scale used produced scores of 0 for "extremely liberal", and 100 for "extremely conservative". Sample data:

Income	**Category**
Wealthy	**Poor**
90	50
80	60
70	40
60	50
90	30

ANOVA Table (incomplete):

Source of variation	SS	df	MS	F(calc.)
Between groups	2560	1		
Within groups	1200	8		
Corrected total	3760	9		

7. A sociologist conducted a study of assertion by having one of her top students, after appropriate training, note the number of assertive acts performed in a day by each of 10 randomly selected coeds, producing the following sample of data, in acts per day: [5,3,10,6,4,9,5,5,7,5]. Another sociologist wonders whether the male and female students at Bedrock indeed differ in assertiveness and, by a similar procedure, gathers the following data for male students, in acts per day: [8,3,5,8,12,10,7,7,9,7]. Use a one-way analysis of variance to test the hypothesis that male students are more assertive than female students at Bedrock College, at the 90% level of confidence.

   ANOVA table (incomplete):

Source of variation	SS	df	MS	F(calc.)
Between Groups	14.45	1		
Within Groups	99.30	18		
Total	113.75	19		

8. Mr. Martin can drive to work along four different routes, and the following are the number of minutes in which he timed himself on five different occasions for each route:

	Route 1	Route 2	Route 3	Route 4
	22	25	26	26
	26	27	29	28
	25	28	33	27
	25	26	30	30
	31	29	33	30
T.(j) =	129	135	151	141

ANOVA Table (incomplete)

Source	df	SS	M.Sq.	F ratio
Routes	.3	52.8		
Error	16	100.4		
Total	19	153.2		

Complete the ANOVA Table and test if all routes are equally fast (ALPHA = 5%).

9. An imaginary study has been conducted on the effects of three brands of laxatives on regularity of TV actresses where each brand was tested by one actress belonging to each of 10 age groups. Results obtained included: F= (brand M.Sq.)/(Error M.Sq.) = 2.1 with 2 and 18 df.
   a. What hypothesis is tested using this F ratio?
   b. Interpret these results using a significance level of 5%.

References

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Fisher, R. A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

Green, B. F., and Tukey, J. W. (1960). Complex analyses of variance: general problems. Psychometrika 25 127–152.

Kirk, R. E. (1995). Experimental Design: Procedures for the Behavioral Sciences, third edition. Brooks/Cole.

Snedecor, G. W., and Cochran, W. G. (1989). Statistical Methods, eighth edition. Iowa State University Press.

Spiegel, Murray R. (1992). Schaum's Outline of Theory and Problems of Statistics (2nd edition). New York: McGraw-Hill. Chapter 16.

LECTURE TEN

Generalisation to M-Way Analysis of Variance

Introduction

In this lecture, we introduce the generalization of an analysis of variance problem to an M -way ANOVA.

Objective

At the end of this lecture, you should be able to

1. generalize a two-way analysis of variance problem.
2. solve a generalized analysis of variance problem.

Pre Test

1. What is an M -way analysis of variance?
2. What is
   a. an effect?
   b. a contrast?

CONTENT

The examples in lecture seven can be generalized to M-way ANOVAs, For a K_1-by- K_2-by- K_M design, there are

$$\mathrm{P} = \prod_{m=1}^{M} K_m$$

Conditions. An M-way ANOVA has $2^M - 1$ experimental effects in total, consisting of M main effect plus $M!/(M-r)!r!$ interactions of order r= 2…M. A 3-way ANOVA for example has 3 main effects (A, B, C), three second-order interactions (AxB, BxC, AxC) and one third-order interaction (AxBxC). Or more generally, an 0th-order interaction is equivalent to a main effect.

We consider model where every cell has its own coefficient (like Equation 13). We will assume these conditions are ordered in a GLM so that the first factor *rotates* slowest, the second factor next slowest, etc, so that for a 3-way ANOVA with factor A, B, C.

$$\begin{array}{ccccccc} 1 & & 2 \ldots & & K_3 & \ldots & P \\ A_1B_1C_1 & & A_1B_1C_2 \ \ldots & & A_1B_1C_{K_3} & \ldots & A_{K_1}B_{K_2}C_{K_3} \end{array} \qquad (10.1)$$

The data is ordered all subject for cell $A_1B_1C_1$, all subjects for cell $A_1B_1C_2$ etc.

The F-contrasts for testing main effects and interactions can be constructed in an iterative fashion as follows. We define initial component contrast[2]

$$C_m = 1_{K_m} \qquad D_m = -diff(I_{K_m})^T$$

Where diff(A) is a matrix of column differences of A (as in the Matlab function diff). So for a 2-by-a ANOVA

$$C_1 = C_2 = [1, \ 1]^T \qquad D_1 = D_2 = [1, \ -1]^T \qquad (10.2)$$

The term C_m can be thought of as the *common effect* for the *m*th factor and D_m as the *differential effect*. Then contrasts for each experimental effect can be obtained by the Kronecker product of C_m's and D_m's for each factor m = 1…M. for a 2-by-2 ANOVA, for example, the two main effects and interaction are respectively

$$D_1 \otimes C_2 = [1 \quad 1 \quad -1 \quad -1]^T$$
$$C_1 \otimes D_2 = [1 \quad -1 \quad 1 \quad -1]^T$$
$$D_1 \otimes D_2 = [1 \quad -1 \quad -1 \quad 1]^T$$

This also illustrates why an interaction can be thought of as a *difference of differences*. The product $C_1 \otimes C_2$ represents the constant term.

For a 3-by-3 ANOVA

$$C_1 = C_2 = [1, 1, 1]^T \quad D_1 = D_2 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}^T \tag{10.3}$$

And the two main effects and interaction are respectively

$$D_1 \otimes C_2 = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}^T \tag{10.4}$$

$$D_1 \otimes C_2 = \begin{bmatrix} 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix}^T \tag{10.5}$$

$$D_1 \otimes C_2 = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix}^T \tag{10.6}$$

The four rows of this interaction contrast correspond to four 'simple interactions' $\tau_{11}^{AB}, \tau_{12}^{AB}, \tau_{21}^{AB}$ and τ_{22}^{AB}. This reflects the fact that an interaction can arise from the presence of one or more simple interactions.

Two-stage procedure for partitioned errors

Repeated measures M-way ANOVAs with partitioned error can be implemented using the following Summary, Statistics approach.

1. Set up first level design matrices where each cell is modeled separately as indicated in equation 6.2.
2. Fit first level models.
3. For the effect you wish to test the Kronecker product rules outlined in the previous section to see what F-contrast you'd need to use to test the effect at the first level. For example, to test for an

interaction in a 3 x 3 ANOVA you'd use the F-contrast in equation 30 (application of this contrast to subject *n's* data you how significant that effect is in that subject)

4. If the F-contrast in the previous step has R_c rows then for each subject, create the corresponding R_c contrast image. For N subjects this then gives a total of NR_c contrast images that will be modeled at the second-level.
5. Set up a second-level design matrix, $X_2 = I_{R_c} \otimes 1_N$. The number of conditions is R_c. for example, in a 3 x 3 ANOVA, $X_2 = I_4 \otimes 1_N$ as shown in figure 9.
6. Fit the second level model.
7. Test for the effect using the F-contrast. $C_2 = I_{R_C}$.

   For each effect we wish to test we must get the appropriate contrast images from the first level (step 3) and implement a new 2nd level analysis (step 4 to 7). Because we are talking *differential* effects to the second level we don't need to include subject effects at the second level.

Summary

In this lecture, you have been able to

1. understand the basics of a generalized analysis of variance.
2. solve a generalized analysis of variance problem.

Post Test

Describe the procedure for setting up the partitioned errors.

References

Cochran, W. G., and Cox, G. M. (1957). Experimental Designs, second edition. New York: Wiley.

Green, B. F., and Tukey, J. W. (1960). Complex analyses of variance: general problems. Psychometrika 25 127–152.

Howell D.C (1992). Statistical methods for psychology. Duxbury Press, 1922

Plackett, R. L. (1960). Models in the analysis of variance (with discussion). Journal of the Royal Statistical Society B 22, 195–217.

Snedecor, G. W., and Cochran, W. G. (1989). Statistical Methods, eighth edition. Iowa State University Press.

Winer B.J, Brown D.R and Michels K.M. (1991). Statistical principle in experimental design. McGraw-Hill 1991.

Appendices

A: The Kronecker Product

If A is $m_1 \times m_2$ matrix and B is an $n_1 \times n_2$ matrix. Then the Knonecker product of A and B is
$(m_1 n_1) \times (m_2 n_2)$ matrix

$$\begin{bmatrix} a_{11}B & \cdots & a_{1m_2}B \\ \cdots & & \\ a_{m_1 1}B & & a_{m_1 m_2}B \end{bmatrix}$$

Circularity

A covariance matrix $\sum$ is circular if

$$\sum_{ii} + \sum_{jj} - 2\sum_{ij} = 2\lambda$$

For all *I, j*.

Compound Symmetry

If all the variance are equal to λ_1 and all the covariances are equal to λ_2 then we have *compound symmetry.*

Nonsphericity

If $\sum$ is K x K covariance matrix and the first K – 1 eigenvalues are identically equal to

$$\lambda = 0.5\left(\sum_{ii} + \sum_{jj} - 2\sum_{ij}\right)$$

Then $\sum$ is *spherical*. Every other matrix is non-spherical or has *nonsphericity.*

Greenhouse-Geisser correction

For a 1-way ANOVA between subjects with N subjects and K levels the overall F statistics is approximately distributed as

$$F[(K-1)e, (N-1)(K-1)e]$$

$$e = \frac{\left(\sum_{i=1}^{K-1} \lambda_i\right)^2}{(k-1)\sum_{i=1}^{K-1} \lambda_i^2}$$

Where

and λ_1 are the eigenvalues of the normalized matrix $\sum_Z$ where

$$\sum_Z = M^T \sum_{ij} M$$

and M is a K by K – 1 matrix with orthogonal columns (eg. The columns are the first K – 1 eigenvalues of $\sum_{ij}$).

B: Within-subject models

The model in equation 11 can also be written as

$$y_n = 1_K \pi_n + \tau + e_n$$

Where y_n is now K x 1 vector of measurements from the *n*th subject, 1_K is a K x 1 vector of 1's, and τ is a K x 1 vector with *k*th entry τ_k and e_n is a K x 1 vector with *k*th entry e_{nk} where

$$p(e_n) = N(0, \sum_e)$$

We have choice as to whether to treat the subject effects π_n as fixed-effects or random-effects. If we choose random-effects then

$$p(e_n) = N(\mu, \sigma_\pi^2)$$

and overall we have a mixed-effects model as the typical response for subject *n,* π_n , is viewed as a random variable. The reduced model is

$$y_n = 1_k \pi_n + e_n$$

For the full model we can write

$$p(y) = \prod_{n=1}^{N} p(y_n)$$

$$p(y_n) = N(m_y, \sum_y)$$

and

$$m_y = 1_{K\mu} + \tau$$
$$\sum_y = 1_{K\sigma_\pi^2} 1_K^T + \sum_e$$

If the subject effects are random-effects, and $\sum_y = \sum_e$, otherwise. If $\sum_y = \sigma_e^2 1_K$ then $\sum_y$ has *compound symmetry*. It is also spherical (see Appendix A) for K = 4 for example

$$\sum_y = \begin{bmatrix} \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 \end{bmatrix}$$