1-Way Fixed Effects ANOVA

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

1-Way Fixed Effects ANOVA

Introduction

- In this module, we introduce the single-factor completely randomized analysis of variance design.
- We discover that there are several ways to conceptualize the design.
- For example, we can see the design as a generalization of the 2-sample t-test on independent groups.
- Or, we can see the design as a special case of multiple regression with fixed regressors.

Introduction

- We will investigate several key aspects of any design we investigate:
  1. Understanding the major hypotheses the design can test;
  2. Measures of effect size we can extract from the statistical analysis;
  3. Relationships between power, precision, sample size, and effect size, and how we exploit them in planning our experiments;
  4. Statistical assumptions, restrictions, and robustness properties.

An Introductory Example

MWL present data from a hypothetical study of memory in their Table 8.1.

40 participants were divided randomly into 4 groups.

Each participant studied a list of 20 words and was tested for recall a day later.

Each of the 3 experimental groups was instructed to use a different special memorization strategy.

A 4th group was simply told to try to memorize the list.

Data are shown in Table 8.1 (MWL, p. 171). (There is an error in the Table: $\overline{Y}_{\bullet j} = 9.95$ should be $\overline{Y}_{\bullet\bullet} = 9.95$.

An Introductory Example

Table 8.1 Recall scores from a hypothetical memory study

	Control	Loci	Image	Rhyme	
	11	10	13	16	
	4	18	16	9	
	8	6	3	7	
	3	20	6	10	
	11	15	13	9	
	8	9	10	14	
	2	8	13	16	
	5	11	9	3	
	8	12	5	9	
	5	12	19	12	
$\bar{Y}_{.j} =$	6.5	12.1	10.7	10.5	$\bar{Y}_{.j} = 9.95$
$s_j^2 =$	10.056	19.433	25.567	16.722	

The Basic Idea behind ANOVA

As we saw in detail in Psychology 310, at its foundation, ANOVA attempts to assess whether a group of means are all the same.

It does this by comparing the variability of the sample means to what it *should be* if the population means are all the same.

If the population means are all the same, the sample sizes are all the same (n per group), and the populations all have the same variance σ^2, then the variance of the sample means should be approximately σ^2/n.

On the other hand, if the population means are not all the same, then the variance of the sample means should be larger than σ^2/n, because the spread of the means reflects more than sampling variability.

The Basic Idea behind ANOVA

As a consequence of this, and a lot of statistical machinery, we arrive at an F statistic that may be written, for a groups,

$$F_{a-1,a(n-1)} = \frac{s^2_{\bar{X}}}{\hat{\sigma}^2/n} = \frac{ns^2_{\bar{X}}}{\hat{\sigma}^2} \quad (1)$$

We simply compute the sample variance of the sample means, and divide by an estimate of σ^2/n. With equal sample sizes, $\hat{\sigma}^2$ is simply the mean of the group variances.

The Basic Idea behind ANOVA

Here are the calculations in R.

```
> memory.data <- read.csv("Table 8_1 Memory Data.csv")
> means <- with(memory.data, aggregate(Score, by = list(Method), FUN = mean
> var.xbars <- var(means)
> variances <- with(memory.data, aggregate(Score, by = list(Method), FUN =
> var.e <- mean(variances)
> n <- 10
> MS.A <- n * var.xbars
> MS.e <- var.e
> F <- MS.A/MS.e
> c(MS.A, MS.e, F)

[1] 57.966667 17.944444  3.230341
```

The ANOVA Structural Model

An alternative way of conceptualizing ANOVA begins with a *structural model* that includes extensive assumptions and restrictions.

Following the notation in RDASA3, we shall assume a balanced design with n subjects per cell, and a cells.

Let Y_{ij} be the score for the ith person in the jth cell. Then

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \tag{2}$$

where

1. μ is the "grand mean" or overall population mean,
2. $\alpha_j = \mu_j - \mu$ is the difference between the mean of population j and the overall mean,
3. ε_{ij} is independent, normal error. The ε_{ij} are i.i.d. $N(0, \sigma_e^2)$.

These assumptions imply that the observations in each group are normally distributed, and that group variances are equal.

MWL summarize the model assumptions in their Box 8.1.

The ANOVA Structural Model

Box 8.1 Parameter Definitions and Assumptions

1. *The parent population mean*, μ. This is the grand mean of the treatment populations selected for this study and is a constant component of all scores in the a populations. It is the average of the treatment population means:

$$\mu = \sum_{j=1}^{a} \mu_j / a$$

2. The effect of treatment A_j, α_j. This equals $\mu_j - \mu$ and is a constant component of all scores obtained under A_j but may vary over treatments (levels of j).

2.1 Because the deviation of all scores about their mean is zero, $\Sigma_j \alpha_j = 0$.

2.2 If the null hypothesis is true, all $\alpha_j = 0$.

2.3 The population variance of the treatment effects is $\sigma_A^2 = \sum_{j=1}^{a} \alpha_j^2 / a$.

3. *The error*, ε_{ij}. This is the deviation of the i^{th} score in group j from μ_j and reflects uncontrolled, or chance, variability. It is the only source of variation within the j^{th} group, and if the null hypothesis is true, the only source of variation within the data set. We assume that

3.1 The ε_{ij} are independently distributed; i.e., the probability of sampling some value of ε_{ij} does not depend on other values of ε_{ij} in the sample.

3.2 The ε_{ij} are normally distributed in each of the a treatment populations. Also, because $\varepsilon_{ij} = Y_{ij} - \mu_j$, the mean of each population of errors is zero; i.e., $E(\varepsilon_{ij}) = 0$.

3.3 The distribution of the ε_{ij} has variance σ_e^2 (error variance) in each of the a treatment populations; i.e., $\sigma_1^2 = \ldots = \sigma_j^2 = \ldots = \sigma_a^2$. This is the assumption of *homogeneity of variance*. The error variance is the average squared error; $\sigma_e^2 = E(\varepsilon_{ij}^2)$.

Computations

Computational Formulas

Table 8.3 The analysis of variance for the one-factor between-subjects design

(a) General form of the ANOVA

Source	*df*	*SS*	*MS*	*F*
Total	$an-1$	$\sum_{j=1}^{a}\sum_{i=1}^{n}(Y_{ij}-\bar{Y}_{..})^2$		
A	$a-1$	$n\sum_{j=1}^{a}(\bar{Y}_{.j}-\bar{Y}_{..})^2$	SS_A/df_A	$MS_A/MS_{S/A}$
S/A	$a(n-1)$	$\sum_{j=1}^{a}\sum_{i=1}^{n}(Y_{ij}-\bar{Y}_{.j})^2$	$SS_{S/A}/df_{S/A}$	

(b) ANOVA of the data of Table 8.1

Source	Sum of squares	*df*	Mean square	*F*	*p*-value
Method	173.90	3	57.967	3.230	.034
Error	646.00	36	17.944		
Total	819.90	39			

Computations

Calculation in R

Let's load in the data from Table 8.1.

```
> memory.data <- read.csv("Table 8_1 Memory Data.csv")
> head(memory.data)

   Method Score
1 Control    11
2 Control     4
3 Control     8
4 Control     3
5 Control    11
6 Control     8

> str(memory.data)

'data.frame': 40 obs. of  2 variables:
 $ Method: Factor w/ 4 levels "Control","Image",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Score : int  11 4 8 3 11 8 2 5 8 5 ...
```

Computations

Calculation in R

We see from the above that memory.data has two variables, one of which is an integer variable, the other a "factor."

It is vitally important that the factor variables are typed as factors.

In this case, we are ready to go.

```
> summary(aov(Score ~ Method, data = memory.data))

            Df Sum Sq Mean Sq F value Pr(>F)
Method       3  173.9   57.97    3.23 0.0336 *
Residuals   36  646.0   17.94
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Distribution of the F Statistic

The F statistic with $a-1$ and $a(n-1)$ degrees of freedom has a noncentral F distribution with noncentrality parameter

$$\lambda = n\sum_{j=1}^{a}\left(\frac{\alpha_j}{\sigma}\right)^2 \tag{3}$$

Just as we had standardized measures of fit in the t test situation, we have related measures in ANOVA.

A widely-used measure in the population is

$$\omega^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2} \tag{4}$$

where

$$\sigma_A^2 = \frac{\sum_{j=1}^a (\mu_j - \mu)^2}{a} = \frac{\sum_{j=1}^a \alpha_j^2}{a} \tag{5}$$

Note the similarity between this measure and R^2.

A related measure is f, which compares the standard deviation of effects to σ_e^2.

$$f = \frac{\sigma_A}{\sigma_e} \tag{6}$$

The RMSSE

Measures of Overall Fit

ω^2 and f

The Root Mean Square Standardized Effect (RMSSE) (Steiger & Fouladi, 1997) averages the sum of squared standardized effects by $a-1$ instead of a, then takes the square root to return to the original metric.

$$RMSSE = \sqrt{\frac{\sum_{j=1}^{a}(\alpha_j/\sigma)^2}{a-1}} \tag{7}$$

Measures of Overall Fit

ω^2 and f

A primary reason for dividing by $a-1$ instead of a is that actually only $a-1$ standardized effects are free to vary.

Consider the very simple special case of a two-group experiment involving a treatment group and a placebo control. Suppose the population standard deviation is 1, and the control group has a population mean of 0, the experimental group has a population mean of 2.

In this case, there is one standardized experimental effect, and it is 2 standard deviations in size. On the other hand, the analysis of variance defines two effects, and they are $\alpha_1 = -1$ and $\alpha_2 = +1$, respectively. So, if we average the sum of squared ANOVA effects, we come up with an average squared effect of 1. Clearly, this is misleading, an artifact of the way ANOVA effects are defined.

Measures of Overall Fit

ω^2 and f

There seems to be no simple, universally acceptable solution to this problem. However, averaging with a appears to underestimate effect levels consistently.

Consequently, we propose to average by the number of independent effects, that is, $a-1$. With this stipulation, and in view of Equation 3, the root mean square standardized effect can be written as

$$RMSSE = \sqrt{\frac{\lambda}{n(a-1)}} \tag{8}$$

Some Class Participation Questions

Suppose you have 4 groups with $n = 12$ per group.

If the population RMSSE is 0.60, what would be the power of the F-test?

In the same situation you observe an F statistic of 3.29. Construct a 90% confidence intervalfor λ.

Construct a 90% confidence interval for the population RMSSE.

Can you construct a 90% confidence interval for what the actual power was in your experiment?