1-Way Completely Randomized ANOVA

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

1-Way Completely Randomized Design

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 it as a special case of a test for comparing variances.

An Introductory Example

- Consider the following extremely simple, artificial data set representing three groups, each with three subjects. The scores are

	Group1	Group2	Group3
1	1.00	4.00	7.00
2	2.00	5.00	8.00
3	3.00	6.00	9.00

ANOVA:The Basic Idea

- At its foundation, ANOVA attempts to assess whether a group of means are all the same. If there are k groups, then the null hypothesis might be written

$$H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_k \tag{1}$$

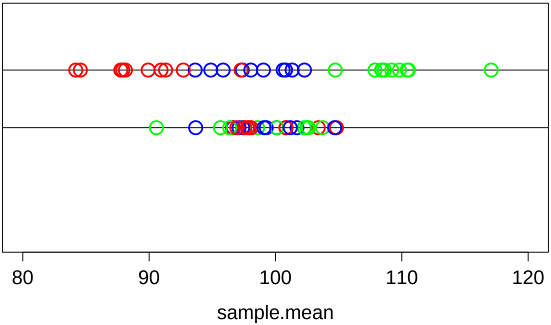
ANOVA:The Basic Idea

- ANOVA does this by comparing the variability of the sample means to what it *should be* if the population means are all the same.
- Suppose, for the time being, that 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 means represent repeated samples from the same population.
- In that case, as we recall from our earlier discussion of the Z test, the variance of the sample means should be approximately $\sigma_M^2 = \sigma^2/n$.
- It is crucial to realize that, under all the assumptions, that is the smallest long run variance that these means can show for any set of population mean values.
- 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.

ANOVA:The Basic Idea

- The plot on the next side shows 10 sample means taken from each of 3 groups. All populations had standard deviations of 15 and sample sizes of 25. The lower line shows 30 means from 3 populations with identical means of 100,100,100. Group 1 is red, Group 2 in blue, Group 3 in green. The dispersion that you see along the line is totally due to within group sampling variation.
- The upper line shows 30 means sampled from populations with means of 90,100,and 110. The dispersion of the means on the top line reflects both within group sampling variation and the spread of the means between groups.

ANOVA:The Basic Idea



ANOVA:The Basic Idea

- As a consequence of the facts from the preceding slides, we can conceptualize the null hypothesis in ANOVA in an entirely different (but equivalent) way from Equation 1
- To emphasize that this is a *one-tailed hypothesis*, I write both the null and alternative hypotheses.

$$H_0 : \sigma^2_M = \frac{\sigma^2}{n} \qquad H_1 : \sigma^2_M > \frac{\sigma^2}{n} \tag{2}$$

ANOVA:The Basic Idea

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

$$F_{k-1,k(n-1)} = \frac{s_M^2}{\hat{\sigma}^2/n} = \frac{ns_M^2}{\hat{\sigma}^2} \tag{3}$$

- 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.

ANOVA:The Basic Idea

- Here are the calculations in R.

```
> Means <- c(mean(Group1),mean(Group2),mean(Group3))
> Variances <- c(var(Group1),var(Group2),var(Group3))
> n <- 3
> k <- 3
> MS.between <- n*var(Means)
> MS.within <-  mean(Variances)
> F <- MS.between/MS.within
> df.between <- k-1
> df.within <- k * (n-1)
> c(MS.between,MS.within,F,df.between,df.within)
[1] 27  1 27  2  6
```

Calculation in R

- The preceding method of calculation is extraordinarily simple. But it is only valid if each group has the same n.
- The Gravetter-Walnau textbook gives computational formulas for the analysis of variance in the general case when n is not necessarily the same in each group. I also provide computational instructions for the case of unequal group n in a handout on the course website.
- However, it is unlikely that, in practice, you will ever calculate an ANOVA by hand. R makes it really easy.
- First, have to enter your data in the correct, special format.
- Instead of entering the data for our 3 groups in 3 columns, we enter each score on a second line, accompanied by a group code. The data entry process is shown on the next slide.

Calculation in R

- It is vitally important that the factor variables are typed as factors.
- In this case, we are ready to go. The following simple commands generate the ANOVA source table, as discussed in your textbook.

```
> Group <- c(1,1,1,2,2,2,3,3,3)
> Score <- c(1,2,3,4,5,6,7,8,9)
> summary(aov(Score ~ factor(Group)))
              Df Sum Sq Mean Sq F value Pr(>F)
factor(Group)  2     54      27      27  0.001 ***
Residuals      6      6       1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The F statistic has a value of 27.00, with 2 and 6 degrees of freedom.
- Remember that this is a 1-tailed test, so the critical value with $\alpha = 0.05$ is at the 0.95 quantile. To find the critical value, we use

```
> qf(0.95,2,6)
[1] 5.143253
```