

# Research Design - - Topic 3

## Completely Randomized Single Factor Analysis of Variance

© 2010 R.C. Gardner, Ph.D.

- General Description, Example and Purpose
- SPSS Analysis
- Experimental Design Approach (Kirk, Ch. 3)
  - The Model
  - Definitional Formulae
  - Expected Mean Squares
  - Assumptions
  - Effect Size and Power
- General Linear Model Approach (Kirk, Ch. 5)

# General Description, Example, and Purpose

Fisher (1925) introduced analysis of variance as a technique to assess the effects of treatments by comparing variation attributed to treatments to variation due to random factors. In time, very formal models were developed based on the complexity of various designs. Today, we can distinguish between two broad approaches:

- a. The traditional (Experimental Design)
- b. The General Linear Model (GLM)

This lecture, and subsequent ones for the next few weeks, will focus on the Experimental Design Approach. The General Linear Model Approach is reviewed on slides 16-18, and the multiple regression approach to GLM will be discussed in later lectures, and will be referred to as MRC.

The following example has  $A = 4$  levels of a treatment and  $n = 8$  observations in each treatment. The values below the table are the means and variances for each treatment as well as the grand mean (mean of the means) and the mean of the variances.

## Example from Kirk (1995, p. 230)

	$A_1$	$A_2$	$A_3$	$A_4$
	4	4	5	3
	6	5	6	5
	3	4	5	6
	3	3	4	5
	1	2	3	6
	3	3	4	7
	2	4	3	8
	2	3	4	10

Means	3.00	3.50	4.25	6.25	4.25
Variances	2.286	.857	1.071	4.500	2.179

The purpose of the analysis is to determine whether the means are more variable than can be reasonably attributed to chance. If so, it might be concluded that the treatments had some effect. In this case, one would be interested in determining:

1. The size of the effect, and/or the power of the test.

2. Whether there are differences between any pairs of means (a contrast) or any other type of contrast (see Topic 4).

# SPSS Analysis

## Using Clope to perform the analysis

Enter SPSS and put data in the Data Editor. The data will consist of one column for the level of A and one for the dependent variable.

## Two Programs are available

**GLM Univariate.** This program will perform one analysis at a time, yields estimates of  $\eta^2$  and power, and permits tests of means if requested.

Click on: Analyze→GLM→Univariate

**Oneway.** This program will perform many single factor analyses of variance in the same run and permits tests of means if requested.

Click on: Analyze→Compare Means→Oneway

# SPSS GLM Run

GET

FILE='\\tsclient\E\PSYCH540\kirldata171.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

UNIANOVA

x BY b

/METHOD = SSTYPE(3)

/INTERCEPT = INCLUDE

/EMMEANS = TABLES(b) COMPARE ADJ(BONFERRONI)

/PRINT = ETASQ OPOWER HOMOGENEITY

/CRITERIA = ALPHA(.05)

/DESIGN = b .

## Estimates

Dependent Variable: x

b	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	3.000	.522	1.931	4.069
2.00	3.500	.522	2.431	4.569
3.00	4.250	.522	3.181	5.319
4.00	6.250	.522	5.181	7.319

### Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: x

F	df1	df2	Sig.
1.293	3	28	.296

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+b

### Tests of Between-Subjects Effects

Dependent Variable: x

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Corrected Model	49.000 <sup>b</sup>	3	16.333	7.497	.001	.445	22.492	.972
Intercept	578.000	1	578.000	265.311	.000	.905	265.311	1.000
b	49.000	3	16.333	7.497	.001	.445	22.492	.972
Error	61.000	28	2.179					
Total	688.000	32						
Corrected Total	110.000	31						

a. Computed using alpha = .05

b. R Squared = .445 (Adjusted R Squared = .386)



# Experimental Design Approach

## The Model:

Based on Fisher's (1925) formulation, it proposes that an observation can be defined in terms of the parameters accounting for variation in sample values.

Thus: 
$$X_{ai} = \mu + \alpha_a + \varepsilon_{ai}$$

where:

$\mu$  = grand mean in the population

$\alpha_a$  = effect of treatment a (i.e.,  $\mu_a - \mu$ )

$\varepsilon_{ai}$  = random error (i.e.,  $X_{ai} - \mu_a$ )

# Definitional Formulae

Source	Sum of Squares	df	F
A	$n \sum_1^a (\bar{X}_a - \bar{G})^2$	a - 1	$\frac{MS_A}{MS_{S/A}}$
S/A	$\sum_1^a \sum_1^n (X_{ai} - \bar{X}_a)^2$	a(n-1)	
Total	$\sum_1^a \sum_1^n (X_{ai} - \bar{G})^2$	an - 1	

## Expected Mean Squares

The parameters in the model estimated by the values of the statistics. They will differ depending on whether the model is fixed or random.

**Fixed Effects Model:** All levels of the factor in the population are included in the experiment, or generalizations are limited to the levels in the population that are included in the experiment.

**Random Effects Model:** A random sample of the levels of the factor in the population are included in the experiment.

Source	Fixed Effects	Random Effects
A	$\frac{n \sum \alpha_a^2}{a-1} + \sigma_\varepsilon^2$	$n\sigma_a^2 + \sigma_\varepsilon^2$
S/A	$\sigma_\varepsilon^2$	$\sigma_\varepsilon^2$

# Assumptions

## Experimental Design Model:

**Independent Random Sampling.** Observations are independently and randomly obtained from the population.

**Homogeneity of Variance.** The variances in the treatment populations are equal.

**Normality.** The treatment populations are normally distributed.

**Null Hypothesis.** The treatment population means are equal.  
Ho:  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$

# Proportions of Variance

## Experimental Design Model:

- $\omega^2$  (Omega-squared for Fixed Effects)

$$\omega^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_\varepsilon^2} = \frac{\frac{(a-1)}{na} (MS_A - MS_{S/A})}{\frac{(a-1)}{na} (MS_A - MS_{S/A}) + MS_{S/A}} = \frac{v_1(F-1)}{v_1(F-1) + N}$$

- $\rho_I$  (Intraclass Correlation for Random Effects)

$$\rho_I = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_\varepsilon^2} = \frac{\frac{1}{n} (MS_A - MS_{S/A})}{\frac{1}{n} (MS_A - MS_{S/A}) + MS_{S/A}} = \frac{(F-1)}{(F-1) + n}$$

## General Linear Model

- Eta-Squared

$$\eta^2 = \frac{SS_A}{SS_A + SS_{S/A}} = \frac{v_1 F}{v_1 F + v_2}$$

# Effect Size

Cohen (1988) defines Effect Size as:

$$f = \sqrt{\frac{\omega^2}{1 - \omega^2}}$$

where:

Effect Size	f	$\omega^2$
Small	.10	.01
Medium	.25	.06
Large	.40	.14

Relation of  $\omega^2$  to  $\eta^2$

$$\omega^2 = \frac{\eta^2(v_1 + v_2) - v_1}{\eta^2(v_1 + v_2) - v_1 + N(1 - \eta^2)}$$

# Options available in G\*Power 3.1

## Test family

F-ratios

## Statistical Test

ANOVA: fixed effects, omnibus, one way

## Types of Calculations

**A priori:** Compute required sample size - - given  $\alpha$ , power and effect size

**Criterion:** Compute required  $\alpha$  - - given power, effect size, and sample size

**Post hoc:** Compute achieved power - - given  $\alpha$ , sample size, and effect size

**Sensitivity:** Compute required effect size - - given  $\alpha$ , power, and sample size

**Compromise:** Compute implied  $\alpha$  and power - - given  $\beta$  to  $\alpha$  ratio, sample size, and effect size

## General Linear Model Approach

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \varepsilon_i$$

where:  $Y_i = DV$  for individual  $i$

$\beta_0, \beta_1, \beta_2 \dots =$  unknown parameters

$X_{i1}, X_{i2}, X_{i3} \dots =$  known IV's

$\varepsilon =$  random error with mean = 0, variance =  $\sigma_\varepsilon^2$

**Matrix Notation:**  $Y = X\beta + \varepsilon$

$$\hat{\beta} = (X'X)^{-1} X'Y$$



The next slide shows one example of GLM with the sample data. The X-matrix is referred to as the Structural Matrix and is used to identify the treatment conditions. The values of the  $\beta$ -vector will depend on how the treatment groups are identified.

Dummy coding is used in the example. The first column identifies the constant while columns 2-4 indicate the treatment level. As a consequence the  $\beta$  values equal mean 4, mean 1 – mean 4, mean 2 – mean 4, and mean 3 – mean 4, respectively. All types of coding will produce the same values of SSTO (Total SS), SSR (Treatment SS) and SSE (Within SS), and hence the same analysis of variance summary table.

SPSS uses GLM with Dummy coding.

**Table 6.5-1** ■ Computational Procedures for CR-4 Design Using a Regression Model with Dummy Coding

Data and basic matrices [ $N = 32, h = 4, n = 8, p = 4$ ]

	$y$ $N \times 1$	$X$ $N \times h$		
		$x_0 \quad x_1 \quad x_2 \quad x_3$		
$a_1$	$\begin{bmatrix} 4 \\ 6 \\ 3 \\ 3 \\ 1 \\ 3 \\ 2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$	$\mathbf{J} = \begin{bmatrix} \mathbf{1}_{N \times 1} \\ \mathbf{1}'_{1 \times N} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}$	
$a_2$	$\begin{bmatrix} 4 \\ 5 \\ 4 \\ 3 \\ 2 \\ 3 \\ 4 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$	$(\mathbf{X}'\mathbf{X})^{-1}_{h \times h} = \begin{bmatrix} \frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{2}{8} & \frac{1}{8} & \frac{1}{8} \\ -\frac{1}{8} & \frac{1}{8} & \frac{2}{8} & \frac{1}{8} \\ -\frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{2}{8} \end{bmatrix}$	$\mathbf{X}'\mathbf{y}_{h \times 1} = \begin{bmatrix} 136 \\ 24 \\ 28 \\ 34 \end{bmatrix}$
$a_3$	$\begin{bmatrix} 5 \\ 6 \\ 5 \\ 4 \\ 3 \\ 4 \\ 3 \\ 4 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$	$\hat{\boldsymbol{\beta}}_{h \times 1} = \begin{bmatrix} 6.25 \\ -3.25 \\ -2.75 \\ -2.00 \end{bmatrix}$	$SSTO = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{J}\mathbf{y}N^{-1} = 688.000 - 578.000 = 110.000$
$a_4$	$\begin{bmatrix} 3 \\ 5 \\ 6 \\ 5 \\ 6 \\ 7 \\ 8 \\ 10 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$	$SSR = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{J}\mathbf{y}N^{-1} = 627.000 - 578.000 = 49.000$	$SSE = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 688.000 - 627.000 = 61.000$

# References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Fisher, R. A. (1925). *Statistical methods for research workers*. London, GB: Oliver & Boyd.