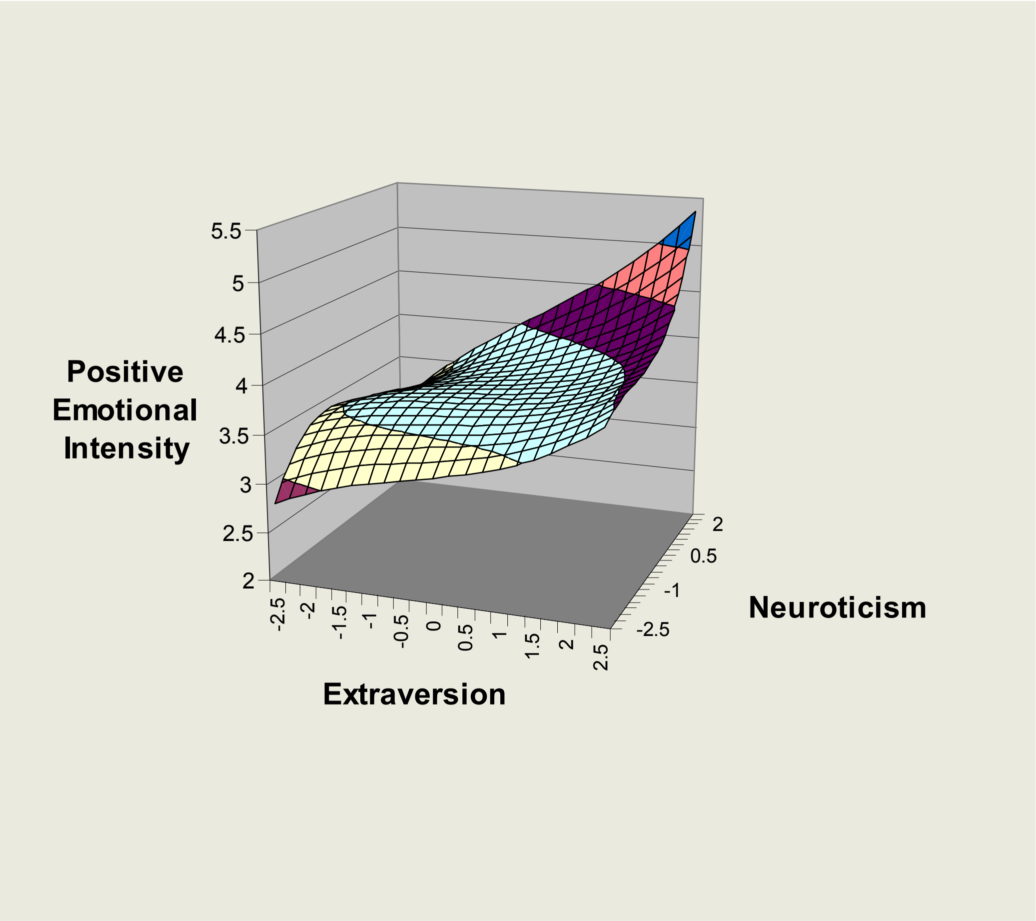
Psychology 515 - Experimental Design

Class Notes

Fall 2011

ROBERT M. MCFATTER
UNIVERSITY OF LOUISIANA
Lafayette



Psychology 515 - Experimental Design

Syllabus - Fall 2011

Instructor: Bob McFatter
Office: Girard Hall 222A Phone: 482-6589 (office); 981-1694 (home)
Internet page: *www.ucs.louisiana.edu/~rmm2440/*
E-mail: *mcfatter@louisiana.edu*

Texts: Keppel, G. and Wickens, T.D. (2004). *Design and analysis, 4th ed.,* Englewood Cliffs, NJ: Prentice-Hall.
JMP 9, Statistical Discovery Software, by SAS Institute.
(available through UL site license at http://helpdesk.louisiana.edu)
Class Notes (handout)

Objective of course: The purpose of this course is to provide the student with a working knowledge of commonly used techniques for designing and analyzing experiments in psychology. Primary emphasis is given to both practical and theoretical considerations in the use of analysis of variance procedures in designing experiments and interpreting the results. Emphasis will be given to using the JMP statistical package for carrying out analyses discussed. The first part of the course includes an introduction to the JMP statistical package. Students use JMP to do familiar descriptive statistics, data screening, histograms, scatterplots, *t*-tests, as well as analysis of variance.

Topics to be covered:

Random variables, expected values, sampling distributions, notation, logic, & computations for completely randomized one-way design
Effects of, and tests for, violations of assumptions, transformations to meet assumptions
Planned and post hoc comparisons, orthogonal contrasts including trend analyses
Power analysis of hypothesis tests, magnitude of effects estimation, expected mean squares
Factorial designs, estimation and interpretation of main effect and interaction parameters, interpretation of complex contrasts
Crossed and nested factors, fixed vs. random effects designs, randomized blocks designs
Generation of models and expected mean squares for complex designs including repeated measures and split-plot designs, choice of appropriate error terms in complex designs

Additional useful sources in the library:

Lindman, H. R. (1974). *Analysis of variance in complex experimental designs.* San Francisco: Freeman.
Maxwell, S.E. & Delaney, H.D. (1990). *Designing experiments and analyzing data.* Belmont, CA: Wadsworth.
Myers, J. (1979). *Fundamentals of experimental design, 3rd ed.* Boston: Allyn & Bacon.
Winer, B.J., Brown, D.R. & Michels, K.M. (1991). *Statistical principles in experimental design, 3rd ed.,* New York: McGraw-Hill.

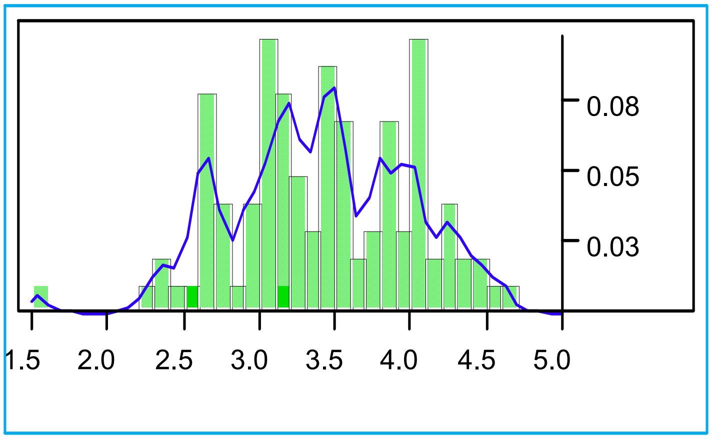
Grades will be based on three equally weighted exams.

Emergency Evacuation Procedures

A map of this floor is posted near the elevator marking the evacuation route and the *Designated Rescue Areas.* These are areas where emergency service personnel will go first to look for individuals who need assistance in exiting the building. Students who may need assistance should identify themselves to the teaching faculty.

Statistical Review

Random or stochastic variable, X: A symbol for set of scores, each of which has a particular probability of occurring when one samples at random from the population. The probabilities can always be shown graphically as a probability distribution. Associated with *every* random variable is some distribution which gives *exhaustive* information about the population of scores.



Because the probability distribution gives complete, exhaustive information about the variable, we can describe *all* properties of the variable by describing properties of the picture (graph) of that variable.

Population: The entire universe of scores on the random variable in which we are interested. We usually assume the number of scores in the population to be *infinite.*

Sample: A subset of scores from the population.

Population *parameters:* Numbers which describe characteristics of the *population* of scores. We use Greek letters: μ, σ, σ^2.

Sample *statistics:* Numbers which describe characteristics of the *sample* of scores. We use English (Roman) letters: M, s, s^2. (*NB.* a caret, ^, over a Greek letter changes it into an English letter; e.g., $\hat{\sigma}$ is the same as s.)

Major characteristics of random variables (or equivalently, distributions):

1) *Central tendency*
  Mean - arithmetic average, center of gravity.
  Median - 50th %-ile.
  Mode - most frequent score.

2) *Variability*
  Range - difference between largest and smallest scores.
  Standard deviation - intuitively, the "typical" deviation of scores from the mean
  Variance - the square of the standard deviation; not as intuitively meaningful, but more mathematically tractable. Contains *exactly* the same information as the standard deviation.

3) *Shape*
  Skewness, kurtosis, etc.

Mean or Expected Value of a Distribution:

The *mean* of a random variable (i.e., of a *population* [not a sample] of scores) is called the *expected value* of the variable and denoted $E(X)$.

This is the center of gravity or 'balance point' of the distribution and has a precise technical definition. For a continuous random variable the definition is

$$\mu = E(X) = \int_{-\infty}^{+\infty} Xf(X)dX = \frac{\sum X}{N},$$

where N = number of observations in the *population* (i.e., infinity).

Rules for algebraic manipulation of the ‘expectation operator,’ E:

a = any constant; X and Y = any random variables.

$E(aX) = a\,E(X)$

$E(a) = a$

$E(X + Y) = E(X) + E(Y) \Longrightarrow E(\Sigma X_i) = \Sigma E(X_i)$

$E(X - Y) = E(X) - E(Y)$

If X and Y are uncorrelated then $E(XY) = E(X)\,E(Y)$.

Variance and Standard Deviation of a Distribution:

The *variance* of a random variable (*population*) is technically defined as

$$\sigma^2 = Var(X) = E\left[X - E(X)\right]^2 = \frac{\sum (X-\mu)^2}{N}.$$

where again N = number of observations in the *population* (i.e., infinity).

Note that sample statistics are very different things from population parameters even though the formulas may appear superficially similar.

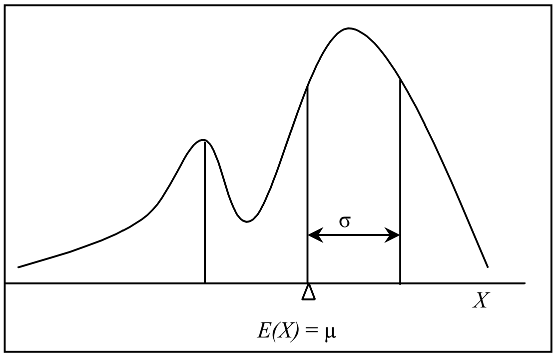
$M = \frac{\sum X}{n}$ and $\mu = \frac{\sum X}{N}$ are very different, as are

$$s^2 = \frac{\sum (X-M)^2}{n-1} \quad \text{and} \quad \sigma^2 = \frac{\sum (X-\mu)^2}{N}.$$

One difference: We *always* know sample statistics, we *never* know population parameters. We use sample statistics to *estimate* population parameters.

Sampling distributions:

Consider beginning with a population of X scores and taking an *infinite* number of samples of size n, and calculating M, s, and s^2 for *each* sample.



$n_1 = 5,\ M_1 = \_\_\_,\ s_1 = \_\_\_,\ s_1^2 = \_\_\_$

$n_2 = 5,\ M_2 = \_\_\_,\ s_2 = \_\_\_,\ s_2^2 = \_\_\_$

$n_3 = 5,\ M_3 = \_\_\_,\ s_3 = \_\_\_,\ s_3^2 = \_\_\_$

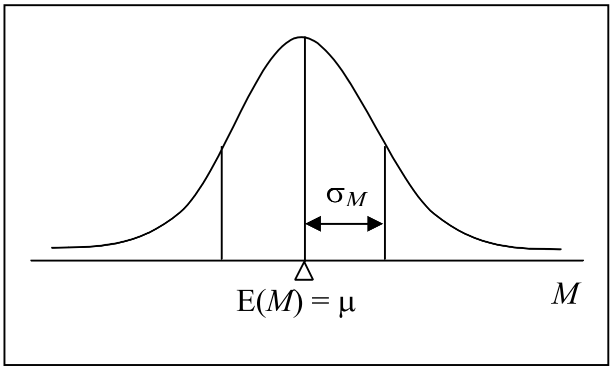
. . . .

. . . .

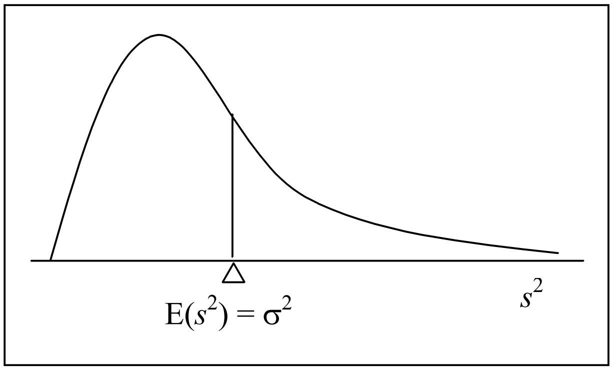
to infinity

Original Population

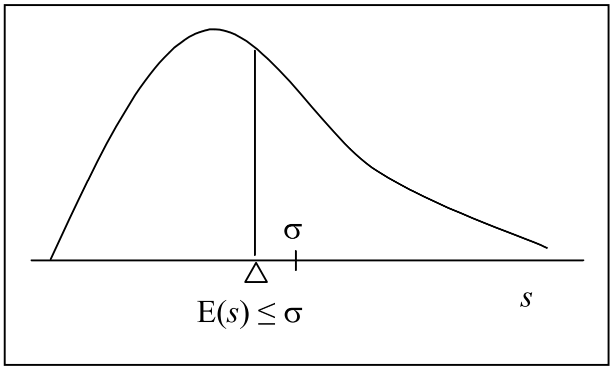
The distributions (*populations*) of the infinite number of M's, s's, and s^2's would look something like the distributions below, which are called *sampling distributions* of these sample statistics.



Sampling Distribution of M



Sampling Distribution of s^2



Sampling Distribution of s

Note that an *unbiased* statistic is one whose expected value equals the parameter it is estimating, e.g., $E(\hat{\theta}) = \theta$. Thus, M and s^2 are unbiased, but s is biased.

Central Limit Theorem:

In the *sampling distribution of means,*

1) $E(M) = \mu$ where $\mu \equiv E(X)$.

2) $\sigma_M = \dfrac{\sigma}{\sqrt{n}}$ or $\sigma_M^2 = \dfrac{\sigma^2}{n}$ where *n* = the # of observations *each M* is based on (in the example above, *n* = 5).

3) As $n \to \infty$ the sampling distribution of *M'*s becomes increasingly normal, *regardless* of the shape of the population of *X'*s.

Completely Randomized One-Way ANOVA

In this simple design several (two or more) groups of **different** subjects are formed and a score on the dependent variable obtained for each subject. The groups are thought of as being the several levels of the **one** independent variable or "factor." It is essential that scores in one group be **independent** of scores in other groups. That is, there should be no sense in which scores in one group could be meaningfully paired with scores in any other group.

Structural layout:

Example: 4 groups, 5 subjects per group.

	Group		
1	2	3	4
S_1	S_6	S_{11}	S_{16}
.	.	.	.
.	.	.	.
.	.	.	.
S_5	S_{10}	S_{15}	S_{20}

Randomized experimental form: Subjects are **randomly assigned** to be in one of the experimental conditions. The experimenter **manipulates** the independent variable by determining the specific treatment each group receives. **Example**: Subjects are randomly assigned to receive one of 4 levels of drug dosage: placebo, 50mg, 100mg, or 200mg. Scores on a cognitive performance test are obtained for each subject.

Quasi-experimental (correlational) form: Each group of subjects is treated as a **random sample** from a preexisting population. The experimenter does **not** manipulate which group the subject is in, but simply **observes** which group the subject is in. **Example**: Reading ability scores are obtained along with a measure of socioeconomic status (SES) for a large sample of subjects. Subjects are divided into 4 levels of SES in order to compare average reading ability for different SES levels.

Analysis of Variance: a = # of groups; n = # of subjects/group

Source	*df*	*Error Term for F-test*	*Comp. Formula*
A	$a - 1$	S(A)	[A] - [T]
S(A)	$a(n - 1)$	None	[Y] - [A]
Total	$an - 1$		[Y] - [T]

where $[A] = \Sigma A^2/n$, $[T] = T^2/an$, $[Y] = \Sigma(Y)^2$.

Example 1.

	a_1	a_2	a_3
	Placebo	Drug A	Drug B
$a = 3$	9	5	2
$n = 5$	8	4	4
	8	5	3
	6	8	1
	9	3	5
A_i	40	25	15
$\overline{Y}_i$	8	5	3

$[A] = (40^2 + 25^2 + 15^2)/5 = 490$

$[T] = 80^2/15 = 426.6667$

$[Y] = 9^2 + 8^2 + ... = 520$

SS_A = 63.33

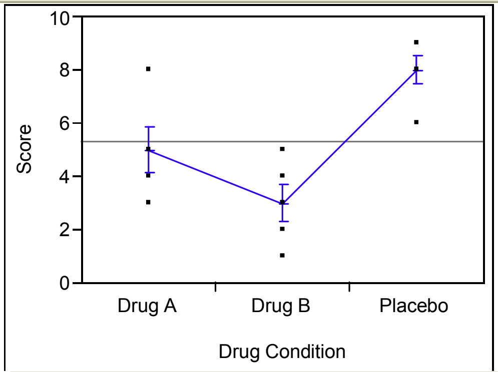
$SS_{S(A)}$ = 30.000

SS_T = 93.333

Here's a JMP analysis:

Score	Drug
9	Placebo
8	Placebo
8	Placebo
6	Placebo
9	Placebo
5	Drug A
4	Drug A
5	Drug A
8	Drug A
3	Drug A
2	Drug B
4	Drug B
3	Drug B
1	Drug B
5	Drug B

Oneway Analysis of Score By Drug Condition



Oneway Anova

Summary of Fit

Rsquare	0.678571
Adj Rsquare	0.625
Root Mean Square Error	1.581139
Mean of Response	5.333333
Observations (or Sum Wgts)	15

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Drug Condition	2	63.333333	31.6667	12.6667	0.0011
Error	12	30.000000	2.5000		
C. Total	14	93.333333			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Drug A	5	5.00000	0.70711	3.4593	6.5407
Drug B	5	3.00000	0.70711	1.4593	4.5407
Placebo	5	8.00000	0.70711	6.4593	9.5407

Std Error uses a pooled estimate of error variance

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
Drug A	5	5.00000	1.87083	0.83666	2.6771	7.3229
Drug B	5	3.00000	1.58114	0.70711	1.0368	4.9632
Placebo	5	8.00000	1.22474	0.54772	6.4793	9.5207

Means Comparisons

Comparisons for each pair using Student's t

t	Alpha
2.17881	0.05

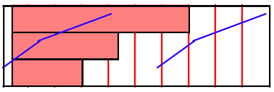
Abs(Dif)-LSD

	Placebo	Drug A	Drug B
Placebo	-2.1788	0.8212	2.8212
Drug A	0.8212	-2.1788	-0.1788
Drug B	2.8212	-0.1788	-2.1788

Level			Mean
Placebo	A		8.0000000
Drug A		B	5.0000000
Drug B		B	3.0000000

Levels not connected by same letter are significantly different

Level	- Level	Difference	Lower CL	Upper CL	p-Value
Placebo	Drug B	5.000000	2.82119	7.178813	0.0003093
Placebo	Drug A	3.000000	0.82119	5.178813	0.0110667
Drug A	Drug B	2.000000	-0.17881	4.178813	0.0686550



Theoretical Rationale

Consider the above example from an experiment with $a = 3$ and $n = 5$. The question to be answered is does the treatment have an effect on the dependent variable (DV)? The general approach is to assume that the treatment has *no* effect (null hypothesis), construct a model based on that assumption, see what the model implies we should find in an experiment if that assumption were true, and compare what we actually found with the model's predictions. If what we found is different from what the model predicts, we reject the model's assumption that there is no treatment effect.

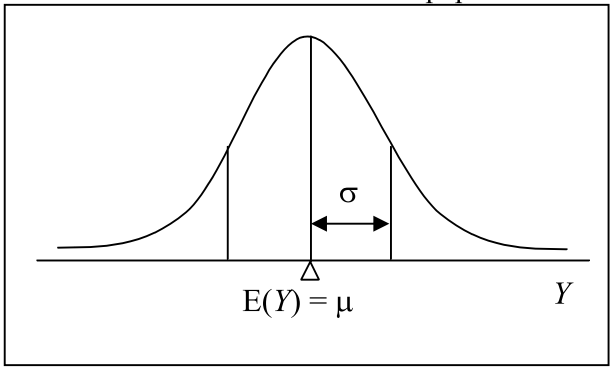
H_0: $\mu_1 = \mu_2 = \mu_3 = \mu$ (this is the assumption of no treatment effect)

Additional assumptions:

1) Each of the 3 *populations* are normally distributed.
2) $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$ (homogeneity of variance).
3) Independent random samples from each population.

If null hypothesis and assumptions are true then the 3 populations are all the same population:

If we wished to estimate σ^2 for this population from our experiment there are several ways to do it:



$\hat{\sigma}_T^2 = \dfrac{\sum Y^2 - \dfrac{\left(\sum Y\right)^2}{an}}{an-1}$ = sample variance of all 15 scores = ***MS***$_T$.

$\hat{\sigma}_W^2 = \dfrac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \hat{\sigma}_3^2}{3}$ = average of sample variances within groups = $\boldsymbol{MS_W}$.

$\hat{\sigma}_B^2 = n\hat{\sigma}_M^2$ = n (here 5) times the variance of the 3 means = $\boldsymbol{MS_B}$.

The third estimate is based on a recognition of the fact that if the assumptions and null hypothesis are true, then the 3 sample means are a random sample of size 3 from the *sampling distribution of means*. And, of course, we know the relation between the variance of the sampling distribution of means and the variance of the original population of Y's from the central limit theorem. It is: $\sigma_M^2 = \sigma^2/n$, where n = # of observations each mean is based on. Thus, $\sigma^2 = n\sigma_M^2$.

It can be shown that the last 2 estimates of σ^2 (between and within) are totally independent of each other. If all assumptions and H_0 are true (i.e., the model is correct), the 2 estimates should be close to one another. And if we form a ratio ($F = MS_B/MS_W$) of the 2 estimates, the ratio should be close to 1. If it's not, we reject the model with its assumption that there is no treatment effect. We conclude that there *is* a treatment effect.

Computing the last 2 estimates of σ^2 given above for the data in our experiment yields:

$$MS_W = \hat{\sigma}_W^2 = (1.5 + 3.5 + 2.5)/3 = 2.5$$

$$MS_B = \hat{\sigma}_B^2 = 5(6.3333) = 31.6667$$

Thus $F = MS_B/MS_W = 31.6667/2.5 = 12.67$, which is hardly close to 1, suggesting we should reject the model and conclude that there is a treatment effect.

Logically, it's possible that it's not the H_0 that's the problem with the model but one of the other assumptions. But it's clear that if the H_0 is false then one would expect to find an F-ratio *larger* than 1. This can be seen by assuming all assumptions are met except the null hypothesis. The three populations are now no longer a single population, but 3 populations whose means are *more spread out*. Thus, samples taken from these 3 populations will have *more variable* sample means than if they were taken from a single population. Thus, $\hat{\sigma}_M^2$ will be larger and so will MS_B. On the other hand, MS_W will remain unaffected by differences in population means because MS_W is simply the average of the sample variances *within* the groups.

The different mean squares (*MS*) are all *sample* variance estimates of the form *SS/df,* and thus have degrees of freedom associated with them.

$df_T = an - 1 = 14$ in example $=$ (# of obs. - 1)
$df_A = a - 1 = 2$ " $=$ (# of means - 1)
$df_{S(A)} = a(n - 1) = 12$ " $= (\Sigma$ of df for each within $\hat{\sigma}^2)$

ANOVA Summary Table

Source	*SS*	*df*	*MS*	*F*	*p*
Between	63.333	2	31.667	12.67	0.0011
Within	30.000	12	2.500		
Total	93.333	14	6.667		

Violation of Assumptions

The between subjects ANOVA (including one-way designs and higher order factorial designs) F-test is robust with respect to violations of the normality assumption. That is, the actual Type I and II error rates are usually not too different from the nominal error rates even when the population distributions are markedly nonnormal.

The same holds true generally of the homogeneity of variance assumption *as long as the sample sizes in each group are equal.* Sometimes, however, the assumption is so badly violated that one may need to take steps to take that into account. My rule of thumb is that unless the ratio of the largest to the smallest sample variance among the groups is greater than about 8 or 9 to 1, I don’t worry about it too much unless the *n*s are unequal.

In the one-way ANOVA, several tests of the homogeneity of variance assumption are available along with modifications of the usual F-test designed to correct for those violations. The usual correction is called a Welch ANOVA and is available in many statistics packages, including JMP. The next section is simply reproduced from the JMP help file on Homogeneity of Variance Tests in the one-way ANOVA:

Homogeneity of Variance Tests

When the variances across groups are not equal, the usual analysis of variance assumptions are not satisfied and the ANOVA *F* test is not valid. JMP gives four tests for equality of group variances and an ANOVA that is valid when the group sample variances are unequal. The concept behind the first three tests of equal variances is to perform an analysis of variance on a new response variable constructed to measure the spread in each group. The fourth test is Bartlett’s test, which is similar to the likelihood ratio test under normal distributions.

O'Brien's test constructs a dependent variable so that the group means of the new variable equal the group sample variances of the original response. An ANOVA on the O'Brien variable is actually an ANOVA on the group sample variances (O'Brien 1979, Olejnik and Algina, 1987).
The Brown-Forsythe test shows the *F* test from an ANOVA in which the response is the absolute value of the difference of each observation and the group median (Brown and Forsythe 1974a).
Levene's test shows the *F* test from an ANOVA in which the response is the absolute value of the difference of each observation and the group mean (Levene 1960).
Bartlett's test is a weighted geometric average of the group sample variances multiplied by a correction factor to give it a Chi Square distribution. Dividing the Bartlett Chi Square test statistic by the degrees of freedom gives the *F* value shown in the table. Bartlett's test is valid only under normality (Bartlett and Kendell 1946).

The Tests that the Variances are Equal table shows the differences between group means to the grand mean and to the median, and gives a summary of testing procedures.

If the tests of equal variances reveal that the group variances are significantly different, the Welch ANOVA for the means should be used in place of the usual ANOVA in the Fit Means table. The Welch statistic is based on the usual ANOVA *F* test: however, the means have been weighted by the reciprocal of the sample variances of the group means (Welch 1951; Brown and Forsythe 1974b; Asiribo, Osebekwin, and Gurland 1990).

If there are only two levels, the Welch ANOVA is equivalent to an unequal variance *t* test (the *t* value is the square root of the *F*).

Transformations

Sometimes it is desirable to transform scores on the dependent variable so as to make the data conform more closely to the assumptions underlying the analysis of variance. When the scores on the dependent variable are correlation coefficients or binomial proportions, carrying out certain transformations prior to performing the ANOVA is fairly routine and standard practice. Below are the transformations that are commonly used for these situations.

Dependent Variable	***Transformation***	***Baseline Error Variance***
Correlations, r	Fisher's Z $Z_r = \tanh^{-1}(r)$ $= .5 \ln[(1+r)/(1-r)]$	$1/(n-3)$ where n= # of obs. each r is based on
Binomial proportions, p	$Y = \sin^{-1}(p^{1/2})$	$821/n$ (for Y in degrees) or $1/(4n)$ (for Y in radians) where n= # of obs. each p is based on

Example with correlations as the dependent variable.

Consider an experiment with 3 different task difficulty (easy, moderate, hard) conditions and n = 33 subjects in each condition. The experimenter collects data on both anxiety and performance for each subject in each condition and calculates the correlation between anxiety and performance in each condition. The three correlations are given below:

	Easy	Moderate	Hard
r	.4	.1	-.5

To test the H_0: $\rho_1 = \rho_2 = \rho_3$, we may treat this as a one-way ANOVA with *one* observation per cell. (NB. Even though there are 33 subjects in each condition the dependent variable here is a correlation computed from those 33 subjects, and there is only one of those in each condition.) Thus, there will be no within subjects *MS* in the ANOVA. The baseline error will be used to provide the error variance after transforming to Fisher's *Z*s. The JMP analysis is below:

```
      r      Task Difficulty      Fisher's Z
    0.4                1Easy        0.423649
    0.1            2Moderate        0.100335
   -0.5                3Hard        -0.54931

Oneway Analysis of Fisher's Z By Task Difficulty
Oneway Anova
Summary of Fit

Rsquare                                  .
Adj Rsquare                              .
Root Mean Square Error                   .
Mean of Response                         -0.00844
Observations (or Sum Wgts)               3
Analysis of Variance
Source            DF         Sum of Squares          Mean Square           F Ratio    Prob > F
Task Difficulty   2          0.49106911              0.245535              .          .
Error             0          0.00000000              .
C. Total          2          0.49106911
Means for Oneway Anova
Level          Number        Mean           Std Error          Lower 95%          Upper 95%
1Easy          1             0.42365        .                  .                  .
2Moderate      1             0.10034        .                  .                  .
3Hard          1             -0.54931       .                  .                  .
Std Error uses a pooled estimate of error variance
```

Note that there is no error term or F-ratio in the output because there is only one observation per cell. The ANOVA summary table is (where A is Task Difficulty):

ANOVA Summary Table

Source	*SS*	*df*	*MS*	*F*
A	.491069	2	.245535	7.37 p = .0006
~~S(A)~~	~~0~~	~~0~~	~~0~~	
Baseline error variance		∞	.033333 [= 1/(33 - 3)]	

Note. $F(2, \infty)$ = .245535/.033333 = 7.37

An approximate *p*-level for the *F* may be obtained from an *F*-table or an exact *p*-level may be obtained (as I did here) from a spreadsheet (e.g., Excel, QuattoPro) FDIST function. The *df* for the *F* here are 2 and ∞. Infinity may be approximated in the FDIST function by 9^10.

We reject the H_0 and conclude that the population correlations between anxiety and performance differ among the groups.

Examples with binomial proportions as the dependent variable.

Example 1. Suppose an experiment on the effects of fear-arousing messages has 3 conditions (High Fear, Moderate Fear, Low Fear) with 50 subjects randomly assigned to each condition. The messages try to induce subjects to get a tetanus shot. We can treat each subject as an independent Bernoulli trial (2 choices: did get a shot, didn't get a shot). Thus, each condition has one proportion that can be treated as a binomial proportion from 50 Bernoulli trials. We are interested in whether the proportion getting a shot differs between Fear conditions. This problem could be analyzed with a chi-square analysis, but it can also be analyzed using a one-way ANOVA with the arcsin transformation. Suppose the data are as follows:

Fear Condition	# got shot	p
High Fear	15	0.3
Med Fear	30	0.6
Low Fear	20	0.4

Note again, the number of observations in each group is only *one* (because the dependent variable here is proportion getting a shot), so there will be no within term. The H_0 is $\pi_1 = \pi_2 = \pi_3$. Here's the JMP analysis:

```
Fear Condition        # got shot      p               ArcSin transform
1High Fear            15              0.3             0.57964
2Med Fear             30              0.6             0.886077
3Low Fear             20              0.4             0.684719

Oneway Analysis of ArcSin transform By Fear Condition
Oneway Anova
Summary of Fit

Rsquare                                 .
Adj Rsquare                             .
Root Mean Square Error                  .
Mean of Response                        0.716812
Observations (or Sum Wgts)              3
Analysis of Variance
Source            DF          Sum of Squares        Mean Square       F Ratio     Prob > F
Fear Condition    2           0.04849686            0.024248          .           .
Error             0           0.00000000            .
C. Total          2           0.04849686
Means for Oneway Anova
Level          Number       Mean          Std Error        Lower 95%       Upper 95%
1High Fear     1            0.579640      .                .               .
2Med Fear      1            0.886077      .                .               .
3Low Fear      1            0.684719      .                .               .
Std Error uses a pooled estimate of error variance
```

The ANOVA summary table would be as follows:

Source	*SS*	*df*	*MS*	*F*
A	.048497	2	.024248	4.85 $p = .0078$
~~S(A)~~	~~0~~	~~0~~	~~0~~	
Baseline error variance		∞	.005 [= 1/(4*50)]	

Note. $F(2, \infty) = .024248/.005 = 4.85$

Because JMP computes arcsin's using radians rather than degrees, we use the baseline error for radians, which is $1/(4n)$. We reject the H_0 that the population proportions are the same for the three conditions.

Example 2. Suppose 3 groups of subjects are formed by randomly assigning 4 subjects to each condition. The 3 groups each study the same list of 20 words and then are given different drugs. After 30 minutes, each subject is given a recognition task. The score is the proportion correct (out of 20). Assume that it is sensible to treat each word on the list as an independent 2-choice (recognized or not) trial so that the proportions are binomial proportions. The number correct (out of 20) for each subject is given in the table below:

	Drug	
Placebo	Drug A	Drug B
12	15	4
15	18	6
11	11	7
8	12	12

Note that we now have 4 proportions in each condition, so there will be a within-groups source in this analysis, in contrast to the previous 2 analyses. We use JMP to calculate proportions, transform them and carry out the analysis:

Drug Condition	# correct	Proportion	Arcsin transform
Placebo	12	0.6	0.886077
Placebo	15	0.75	1.047198
Placebo	11	0.55	0.835482
Placebo	8	0.4	0.684719
Drug A	15	0.75	1.047198
Drug A	18	0.9	1.249046
Drug A	11	0.55	0.835482
Drug A	12	0.6	0.886077
Drug B	4	0.2	0.463648
Drug B	6	0.3	0.57964
Drug B	7	0.35	0.633052
Drug B	12	0.6	0.886077

Oneway Analysis of Arcsin transform By Drug Condition
Oneway Anova
Summary of Fit

Rsquare	0.502476
Adj Rsquare	0.391915
Root Mean Square Error	0.172099
Mean of Response	0.836141
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Drug Condition	2	0.26921665	0.134608	4.5448	0.0432
Error	9	0.26656314	0.029618		
C. Total	11	0.53577980			

Means for Oneway Anova

```
Level     Number    Mean       Std Error   Lower 95%   Upper 95%
Drug A    4         1.00445    0.08605     0.80979     1.1991
Drug B    4         0.64060    0.08605     0.44595     0.8353
Placebo   4         0.86337    0.08605     0.66871     1.0580
Std Error uses a pooled estimate of error variance
```

The ANOVA summary table would be as follows:

Source	*SS*	*df*	*MS*	*F*
A	.269217	2	.134608	4.54 $p = .0432$
S(A)	.266563	9	.029618	2.37 $p = .0113$
Baseline error variance		∞	.0125 [= 1/(4*20)]	

Note. $F(2, 9) = .134608/.029618 = 4.54$; $F(9, \infty) = .029618/.0125 = 2.37$

The first *F*-test leads us to conclude that the mean proportions of words correctly recognized differ among the drug conditions.

The second *F*-ratio tests whether there are individual differences among subjects within each group in long-run probability of correctly recognizing a word on any given trial. Note that ‘noise’ (error) variability (the denominator) in a particular *F*-ratio may have several sources. Here the S(A) term contains at least two sources of variability: (1) variability resulting from individual differences in long-run propensity to correctly recognize a word on any given trial, as well as (2) variability in proportion correct that one would expect *even if all subjects within a group were identical in long-run probability of correctly recognizing a word in each trial.* It is this latter variability that the baseline error variance measures.

Thus, the second *F*-test leads us to conclude that there are individual differences even *within* the groups in long-run probability of correctly recognizing a word on each trial.

Contrasts in ANOVA

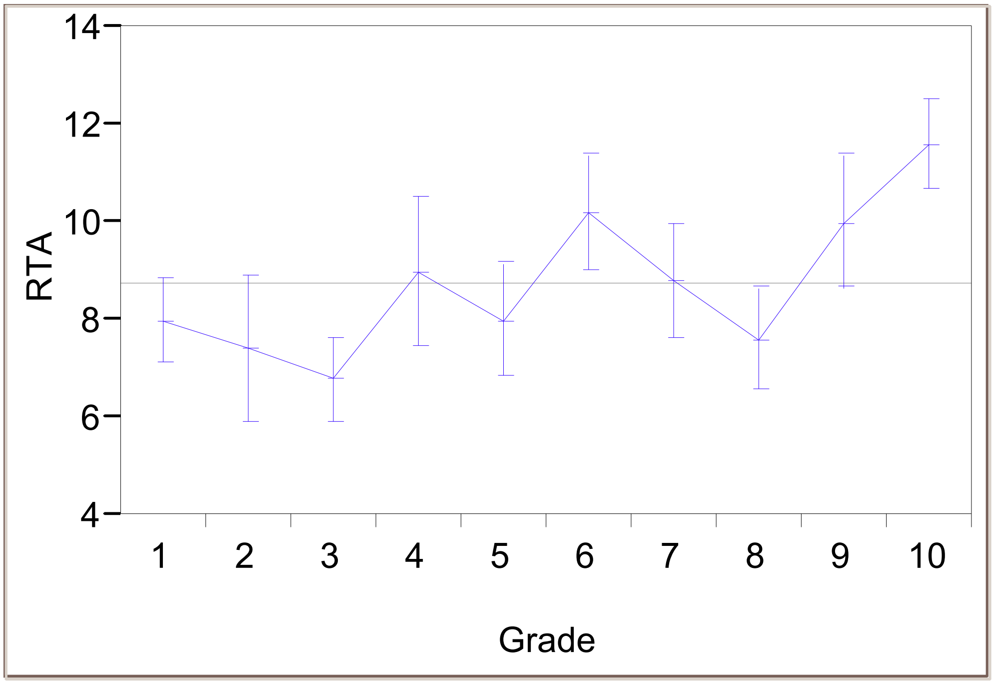
Consider an experiment designed to look at differences in role-taking ability (RTA) of children in different grades. Five subjects are sampled at random from each of grades 1 through 10, and measured on a role-taking ability scale. The totals for the 10 groups are given below. [Note: $a = 10$; $n = 5$]

Grade	1	2	3	4	5	6	7	8	9	10
	6	3	8	12	12	11	10	11	14	14
	8	7	4	5	7	13	8	5	8	10
	11	6	6	13	8	8	9	8	6	12
	7	9	9	7	5	12	5	6	11	9
	8	12	7	8	8	7	12	8	11	13
Total	40	37	34	45	40	51	44	38	50	58
Mean	8	7.4	6.8	9	8	10.2	8.8	7.6	10	11.6

A one-way ANOVA is carried out on the data and yields the following results:

Source	*SS*	*df*	*MS*	*F*	
A	99.62	9	11.07	1.60	$p = .1472$ *ns*
S(A)	276.00	40	6.90		

Note the nonsignificant omnibus *F*, suggesting that there are no differences in role-taking ability (RTA) among the different grades. However, if one plots mean role-taking scores as a function of Grade the following plot is obtained:



Note that the points seem to generally trend upward suggesting that RTA increases as grade level increases. It is important to understand that the omnibus *F* cannot adequately test the idea that there is a linear trend in this data because no account is taken of the fact that grades 1 & 2 are closer together than grades 1 & 10. All grades are simply treated as different groups.

Generally if we have a hypothesized *pattern* of means that we are interested in detecting, we will be able to do much better than simply testing the omnibus H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots$, which takes no account of the pattern we are interested in.

An improved hypothesis test can be carried out by testing a *contrast* or *comparison* among the means. [We use these 2 terms interchangeably.] In this case, the contrast we are interested in is a *linear trend contrast.* Contrasts are constructed by specifying a set of weights [i.e., c_i; i = 1,2, ... a] (one for each group mean) which sum to zero [i.e., $\Sigma c_i = 0$]. The pattern of the weights (called contrast coefficients) should reflect the hypothesized pattern of interest in the means.

Since we are interested in a linear increase in RTA with grade level, the contrast coefficients should reflect the pattern of the grades. To get a set of weights that do this and also sum to zero, we can turn the grades into deviation scores by subtracting the mean grade from each.

The mean grade is 5.5, so the deviation scores are:

Grade	1	2	3	4	5	6	7	8	9	10
c_i:	-4.5	-3.5	-2.5	-1.5	-.5	+.5	+1.5	+2.5	+3.5	+4.5

These would work fine for contrast coefficients since $\Sigma c_i = 0$, but it's awkward to have decimals. Multiplying by a constant doesn't affect contrasts, so let's make nicer coefficients by multiplying them by 2:

Grade	1	2	3	4	5	6	7	8	9	10
c_i:	-9	-7	-5	-3	-1	+1	+3	+5	+7	+9

To calculate SS for a contrast, use either of the 2 equivalent formulas:

$$SS_{A_\psi} = \frac{\left(\sum c_i A_i\right)^2}{n\sum c_i^2} \text{ or } SS_{A_\psi} = \frac{n\left(\sum c_i \overline{Y}_i\right)^2}{\sum c_i^2} = \frac{n\hat{\psi}^2}{\sum c_i^2} \text{ where } \hat{\psi} = \sum c_i \overline{Y}_i\,.$$

For the above data we have (using the 1st formula):

$$SS_{A_{linear}} = \frac{5(56.2)^2}{330} = 47.85515$$

The SS_A can thus be partitioned as follows:

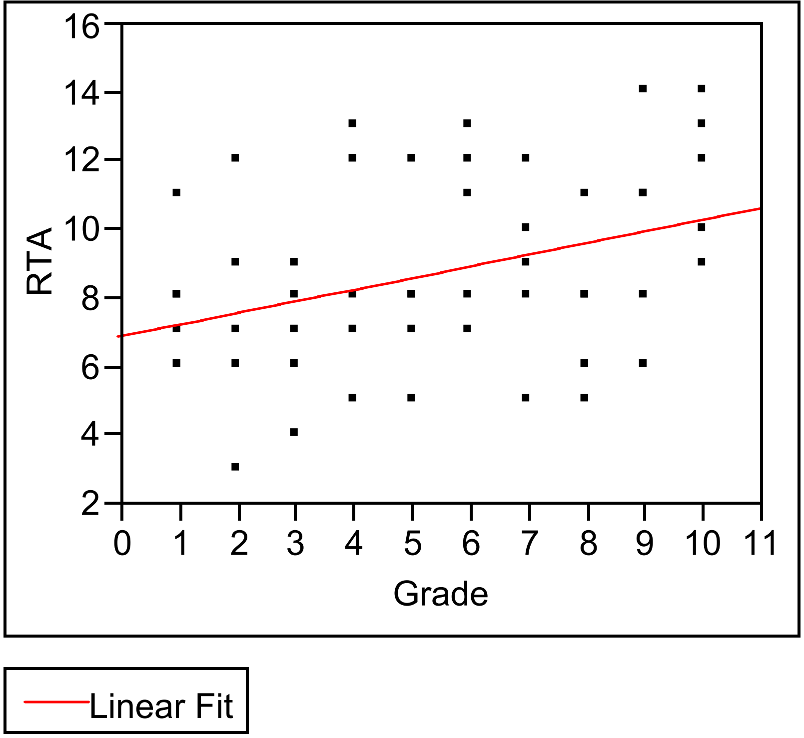
Source	*SS*	*df*	*MS*	*F*	
A	99.62	9	11.07	1.60	$p = .1472$
linear	47.86	1	47.86	6.94	$p = .0120$
residual	51.76	8	6.47	.94	$p = .4969$
S(A)	276.00	40	6.90		

Note, any contrast has only 1 *df.* We conclude here that there is a highly significant linear trend in the data, i.e., RTA does significantly increase with increasing grade level. The nonsignificant residual says that after taking account of differences among means due to linear changes in grade level there is no evidence of systematic (i.e., nonchance) differences among means. Here is the JMP analysis using the Fit Model platform:

```
Response RTA
Summary of Fit

RSquare                         0.265215
RSquare Adj                     0.099888
Root Mean Square Error          2.626785
Mean of Response                8.74
Observations (or Sum Wgts)      50

Analysis of Variance
Source     DF     Sum of Squares        Mean Square      F Ratio
Model      9      99.62000              11.0689          1.6042
Error      40     276.00000             6.9000           Prob > F
C. Total   49     375.62000                              0.1472


Effect Tests
Source    Nparm     DF     Sum of Squares      F Ratio     Prob > F
Grade     9         9      99.620000           1.6042      0.1472
Effect Details
Grade
Least Squares Means Table
Level   Least Sq Mean          Std Error      Mean
1       8.000000               1.1747340      8.0000
2       7.400000               1.1747340      7.4000
3       6.800000               1.1747340      6.8000
4       9.000000               1.1747340      9.0000
5       8.000000               1.1747340      8.0000
6       10.200000              1.1747340      10.2000
7       8.800000               1.1747340      8.8000
8       7.600000               1.1747340      7.6000
9       10.000000              1.1747340      10.0000
10      11.600000              1.1747340      11.6000
Contrast
Test Detail

1            -0.36
2            -0.28
3            -0.2
4            -0.12
5            -0.04
6            0.04
7            0.12
8            0.2
9            0.28
10           0.36
Estimate     2.248
Std Error    0.8536
t Ratio      2.6335
Prob>|t|     0.012
SS           47.855


Sum of Squares        47.855151515
Numerator DF          1
Denominator DF        40
F Ratio               6.9355292051
Prob > F              0.0119555973
```

An alternative way to quickly get the linear contrast *SS* is to do a regression analysis predicting RTA from the *continuous* predictor Grade. The SS_{Model} from this analysis will be the Grade linear contrast *SS*. Here's the JMP analysis using the Fit Y by X platform:

Bivariate Fit of RTA By Grade



Linear Fit

RTA = 6.8666667 + 0.3406061 Grade

Summary of Fit

RSquare	0.127403
RSquare Adj	0.109224
Root Mean Square Error	2.613127
Mean of Response	8.74
Observations (or Sum Wgts)	50

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	8	51.76485	6.47061	0.9378
Pure Error	40	276.00000	6.90000	Prob > F
Total Error	48	327.76485		0.4969
				Max RSq
				0.2652

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	47.85515	47.8552	7.0082
Error	48	327.76485	6.8284	Prob > F
C. Total	49	375.62000		0.0109

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	6.8666667	0.798324	8.60	<.0001
Grade	0.3406061	0.128661	2.65	0.0109

Other patterns among means may be of interest in other experiments, and *it is almost always better in real experiments to test precise hypotheses specified by contrasts, rather than omnibus hypotheses.*

The null hypothesis tested by a contrast may be written as follows:

$$H_0: \psi = \sum c_i \mu_i = 0$$

This fact suggests that other interesting hypotheses may be tested with contrasts. Consider an experiment with 2 experimental groups and 1 control group. If one wished to test the hypothesis that the mean for the control group was equal to the average of the means for the 2 experimental groups, one could write that hypothesis as follows:

$$\text{H}_0\text{: } \mu_C = \frac{\mu_{E_1} + \mu_{E_2}}{2} \quad \text{or, equivalently, } \quad \psi = 2\mu_C - \mu_{E_1} - \mu_{E_2} = 0$$

yielding

$$\begin{array}{lccc} & \text{C} & \text{E}_1 & \text{E}_2 \\ c_i\text{:} & 2 & -1 & -1 \end{array} \quad \text{for contrast weights.}$$

Another meaningful contrast here would be to compare the 2 experimental groups.

$$\text{H}_0\text{: } \mu_{E_1} = \mu_{E_2} \quad \text{or} \quad \mu_{E_1} - \mu_{E_2} = 0\text{, yielding}$$

$$\begin{array}{lccc} & \text{C} & \text{E}_1 & \text{E}_2 \\ c_i\text{:} & 0 & 1 & -1 \end{array} \quad \text{for contrast coefficients.}$$

Example 2. Test the 2 hypotheses above in an experiment with the following means and $n = 4$:

		A	
	C	E_1	E_2
	9	12	6
	4	6	8
	6	8	12
	5	10	14
Total	24	36	40
Mean	6	9	10

		A	
	C	E_1	E_2
	6	9	10
c_{1i}:	2	-1	-1
c_{2i}:	0	1	-1

$$SS_1 = \frac{4[2(6) - 1(9) - 1(10)]^2}{6} = 32.667$$

$$SS_2 = \frac{4[0(6) + 1(9) - 1(10)]^2}{2} = 2.000$$

The ANOVA summary table would include the following:

Source	*SS*	*df*	*MS*	*F*	*p*
A	34.667	2	17.333	2.11	.1775
ψ_1	32.667	1	32.667	3.97	.0774
ψ_2	2.000	1	2.000	0.24	.6337
S(A)	74.000	9	8.222		

The JMP analysis that would produce these results is below (Note that the above *F*s for the contrasts are equal to t^2s from the JMP analysis):

```
Response Y
Summary of Fit

RSquare                          0.319018
RSquare Adj                      0.167689
Root Mean Square Error           2.867442
Mean of Response                 8.333333
Observations (or Sum Wgts)       12

Analysis of Variance
Source     DF      Sum of Squares       Mean Square      F Ratio
Model      2       34.66667             17.3333          2.1081
Error      9       74.00000             8.2222           Prob > F
C. Total   11      108.66667                             0.1775

Effect Tests
Source     Nparm      DF      Sum of Squares        F Ratio        Prob > F
A          2          2       34.666667             2.1081         0.1775

Effect Details
A
Least Squares Means Table
Level      Least Sq Mean          Std Error          Mean
C          6.000000               1.4337209          6.0000
E1         9.000000               1.4337209          9.0000
E2         10.000000              1.4337209          10.0000

Contrast
Test Detail

C          1          0
E1         -0.5       1
E2         -0.5       -1
Estimate   -3.5       -1
Std Error  1.7559     2.0276
t Ratio    -1.993     -0.493
Prob>|t|   0.0774     0.6337
SS         32.667     2


Sum of Squares          34.666666667
Numerator DF            2
Denominator DF          9
F Ratio                 2.1081081081
Prob > F                0.1774635581
```

Note that 32.667 + 2.000 = 34.667 so the 2 contrasts *partition* the SS_A. This is because the two contrasts are *orthogonal.* Two contrasts are orthogonal if and only if $\Sigma c_{1i}c_{2i} = 0$. To say two contrasts are orthogonal means that they test patterns in the group means that are independent of one another. If the df_A is $a - 1$ then the SS_A can always be partitioned into a set of $a - 1$ (but no more than $a - 1$) mutually orthogonal contrasts. There are, of course, many different sets of mutually orthogonal contrasts that could be obtained for any given number of groups (as long as $a > 2$). One set that is always easy to mechanically generate is the set of Helmert contrasts. For example, with 5 groups, a set of Helmert contrasts would be:

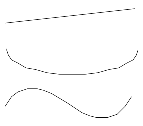
		A			
1	2	3	4	5	
4	-1	-1	-1	-1	
0	3	-1	-1	-1	Helmert contrasts
0	0	2	-1	-1	
0	0	0	1	-1	

If the levels of A are *quantitative* (e.g., Grade levels as in the above RTA example) then it is often sensible to do a *trend analysis.* Trend analyses are carried out by using sets of orthogonal polynomial contrast coefficients. If the quantitative levels of A are *equally spaced* then the coefficients in Keppel's table of orthogonal polynomial coefficients may be used.

If the levels are not equally spaced, *linear* trend coefficients are still easy to compute (just convert the quantitative levels to deviation scores), but the higher order trend coefficients are more difficult to compute. If a higher order trend analysis is desired in this case, it's best to use regression analysis to carry this out.

Note that the contrast coefficients reflect the patterns of means they detect. For example,

Group	1	2	3	4	5	6
Linear	-5	-3	-1	1	3	5
Quadratic	-5	1	4	4	1	-5
Cubic	-5	7	4	-4	-7	5
etc.						



It's important to recognize that the proportion of SS_A that is due to any contrast is equal to the squared correlation between the contrast coefficients and the group means. That is,

$$r^2_{cA} = \frac{SS_{A_\psi}}{SS_A}.$$

This implies that the more closely the pattern of the contrast coefficients matches the pattern of the means, the greater the proportion of between mean variation the contrast accounts for. In the RTA example above, the squared correlation between RTA and Grade level is equal to 47.855/99.62 = .48, indicating that 48% of the variation in RTA means is accounted for by a linear trend over Grade levels.

Using JMP for Calculating SS for Hypotheses When Only Means Are Given

It is often useful to be able to calculate the *SS* for a contrast or omnibus hypothesis test when only the group means are available, i.e., no within group data is available. To do this in JMP, simply put the group means (not group totals) into a column and make another column (perhaps call it Freq) that has the *n* for each group in it. Many platforms will allow the specification of a frequency column (Freq) along with the *X* and *Y* variables. Just specify the column that has the *n*s in it as the Freq variable and the *SS* obtained for both omnibus and contrast hypotheses will the same as those given by Keppel's formulas. This technique will also come in handy for doing power calculations using JMP a later section.

Of course, the within (error) terms in the output will all be zero because the program assumes that each group has *n* identical scores (viz., the mean), and the variance of *n* identical scores is zero. Thus, no *F*s can be computed by the program. However, if within group information is otherwise available (e.g., the *MSE* is given in a journal article or problem, or perhaps the standard deviations for each group are available from which the *MSE* can be computed), *F*-tests may be computed using the numerator *SS* generated by the JMP analysis.

It will be a very useful exercise for you to practice this technique by taking only the means and *n*s from the last two examples (the RTA example and the two experimental-one control group example) and reproducing both the omnibus and all the contrast *SS* given above.

Multiple Comparisons

If a number of contrasts or comparisons are made, the question of possible increased likelihood of Type I errors becomes relevant. Let α_{pc} = the per comparison α level, i.e., the probability on any comparison of making a Type I error if the H_0 is true. Let α_{fw} = the family-wise α level, i.e., the probability of making at least one Type I error in a set or family of hypothesis tests if H_0 is true.

If *k independent* tests are made using α_{pc}, then $\alpha_{fw} = 1 - \left(1 - \alpha_{pc}\right)^k$.
For example, using this formula, if $k = 10$, and $\alpha_{pc} = .05$ then $\alpha_{fw} = .40$. That is, if one makes 10 independent hypothesis tests, each at the .05 level, the probability of making at least one Type I error is .40 if H_0 is true.

Multiple comparison procedures are designed to try and control this increased family-wise error rate, and they do this generally by making it harder for individual comparisons to be declared significant. Of course, in doing this the Type II error rate necessarily increases.

A distinction is often made here between planned or *a priori* comparisons and *post hoc* (data fishing) comparisons. If one wants to test a relatively small number of planned comparisons, I recommend usually not making any correction for family-wise Type I errors.

If one is doing post hoc comparisons or a large number of planned comparisons or is interested for some reason in controlling α_{fw}, Bonferroni's (also called Dunn's) test is a good general procedure. The steps are:

Bonferroni's Test

1) Decide which *k* hypotheses will be tested (This must be done independently of the results).
2) Choose an acceptable α_{fw}'.
3) Test each of the *k* hypotheses using $\alpha_{pc} = \frac{\alpha_{fw}'}{k}$.
4) Then $\alpha_{fw} < \alpha_{fw}'$, i.e., the actual family-wise Type I error rate is less than or equal to α_{fw}'.

The Šidák-Bonferroni test, which is slightly more powerful, and is to be preferred, simply replaces the formula in step (3) by $\alpha_{pc} = 1 - \left(1 - \alpha_{fw}'\right)^{\frac{1}{k}}$. This can be easily calculated using a calculator or Excel. Here are some other commonly discussed multiple comparison procedures:

Scheffe's Test

Calculate the usual *F* for the contrast, but use $(a - 1)\, F[a\text{-}1, a(n\text{-}1)]$ as the critical value instead of the usual tabled *F*. This is the most conservative procedure available, i.e., regardless of how many contrasts or why you're looking at them, one is guaranteed that α_{fw} is less than the α you use. Of course, the Type II error rate soars with this procedure. I don't recommend this procedure under any common circumstances. Note also that the minimum significant difference between any pair of means, using Scheffe's test would be

$$\bar{d}_S = \sqrt{\frac{(a-1)F[df_A, df_{Error}]2MS_{Error}}{n}}$$

where n = the # of observations each mean you're comparing is based on. Of course, the critical F here and elsewhere can be easily obtained using a spreadsheet FINV(α, *dfnum*, *dfden*) function.

Tukey's HSD

This test is designed for the special circumstance in which one wishes to compare *all* possible pairs of means. It's very close in power to Bonferroni's test, but just a hair more powerful under this special circumstance.

$$\bar{d}_T = q\sqrt{\frac{MS_{Error}}{n}}$$

where q = the studentized range statistic and n = the # of observations each mean you're comparing is based on. Tukey’s HSD test is an option provided in both JMP’s ‘Fit Y by X|Oneway’ and ‘Fit Model’ platforms.

Fisher's LSD (sometimes called 'protected *t*' test)

1) Check omnibus *F.* If significant then continue, else stop.
2) Compute

$$\bar{d}_F = t(df_{Error})\sqrt{\frac{2MS_{Error}}{n}}$$

where n = the # of observations each mean you’re comparing is based on.

Note that except for step (1) this procedure is exactly equivalent to making no special correction for family-wise Type I errors. Consequently, this is the least conservative (most powerful) procedure. Whether one would use the HSD or the LSD would depend pretty much on one's relative concern with Type I and Type II errors.

Fisher’s LSD is called ‘Student’s *t*’ in JMP and is an option in both the ‘Fit Y by X|Oneway’ and ‘Fit Model’ platforms.

Mathematical Model for ANOVA

The model underlying the one-way ANOVA follows the general form: ***Data* = *Fit* + *Residual.*** One way this idea can be applied to the one-way ANOVA is to describe each individual's score in the experiment as being composed of 2 pieces:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \text{ where the } \varepsilon_{ij} \text{ are N(0, } \sigma^2 \text{).}$$

Y_{ij} is the score for the *j*th person in the *i*th group, and μ_i represents the population mean for the *i*th group. The ε_{ij} are random errors or residuals. The group means μ_i are often "reparameterized" in the following way:

$$\mu_i = \mu + \alpha_i \text{ and } \sum \alpha_i = 0,$$

where the α_is are called *treatment effects* or *'main' effects.* Note that this implies that

$$\alpha_i = \mu_i - \mu.$$

That is, the α_is are simply deviation scores of the group means around the grand mean of the experiment. The entire model for any score may thus be written:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}.$$

Adopting this model means that one is thinking of the scores in an experiment as composed of 3 pieces: a grand mean, μ, a treatment effect, α_i, and a residual, ε_{ij}. For example, in an experiment with $a = 3$, $n = 5$, $\mu = 10$, $\alpha_1 = 3$, $\alpha_2 = -2$, and $\alpha_3 = -1$, the scores in the experiment would be composed as follows:

Group	1	2	3
	$10 + 3 + \varepsilon_{11}$	$10 - 2 + \varepsilon_{21}$	$10 - 1 + \varepsilon_{31}$
	$10 + 3 + \varepsilon_{12}$	$10 - 2 + \varepsilon_{22}$	$10 - 1 + \varepsilon_{32}$
	$10 + 3 + \varepsilon_{13}$	$10 - 2 + \varepsilon_{23}$	$10 - 1 + \varepsilon_{33}$
	$10 + 3 + \varepsilon_{14}$	$10 - 2 + \varepsilon_{24}$	$10 - 1 + \varepsilon_{34}$
	$10 + 3 + \varepsilon_{15}$	$10 - 2 + \varepsilon_{25}$	$10 - 1 + \varepsilon_{35}$
μ_i	13	8	9

Note that the variance of the scores within any group would simply be the variance of ε_{ij}. Thus, the MS_{Within}, which is the average of the variances within the groups would estimate σ^2, i.e., the variance of the residuals, ε_{ij}.

The null hypothesis, which is often written, H_0: $\mu_1 = \mu_2 = \ldots$, may also be written in terms of the α_is as follows: H_0: $\alpha_i = 0$. That is, if the H_0 is true, all the treatment effects are zero.

It is often of interest to estimate the parameters of the model above. Commonly this is done by finding estimates of the parameters in any given data set that minimize the sum of squared residuals. This is called the 'least squares' method of estimation. That is, one finds estimates of μ and the α_i which minimize the quantity:

$$Q = \sum\sum \varepsilon_{ij}^2 = \sum\sum \left(Y_{ij} - \mu - \alpha_i\right)^2.$$

This kind of minimization is a standard problem in differential calculus. One finds the partial derivatives of Q with respect to μ and α_i and sets them equal to zero. Solving the resulting equations leads to the following 'least squares' estimators for the model parameters.

$$\hat{\mu} = \overline{T}$$
$$\hat{\alpha}_i = \overline{Y}_i - \overline{T}$$
$$\hat{\varepsilon}_{ij} = Y_{ij} - \overline{Y}_i \,.$$

The 'Fit Model' platform in JMP will automatically provide the least squares parameter estimates, $\hat{\mu}$ and $\hat{\alpha}_i$. It directly provides only the first a - 1 of the $\hat{\alpha}_i$, but because the $\hat{\alpha}_i$ are constrained to sum to zero, the last value is easily obtained by summing the others and multiplying the result by -1. Below is the JMP analysis of the data from homework #2. Notice that the grand mean estimate, $\hat{\mu}$ = 11.22222, is given as the Intercept in the Parameter Estimates table, while $\hat{\alpha}_1$ = -1.555556 and $\hat{\alpha}_2$ = -1.11111 are given just below that. The third treatment effect, $\hat{\alpha}_3$= +2.6666667 is the negative of the sum of the first two. Of course, these parameter estimates could just as easily have been computed from the means of the three groups. The 'Expanded Estimates' option in JMP will provide the last treatment effect as well.

Response Score
Summary of Fit

```
RSquare                          0.412879
RSquare Adj                      0.363952
Root Mean Square Error           2.395984
Mean of Response                 11.22222
Observations (or Sum Wgts)       27
```

Analysis of Variance

```
Source    DF    Sum of Squares      Mean Square     F Ratio
Model     2     96.88889            48.4444         8.4387
Error     24    137.77778           5.7407          Prob > F
C. Total  26    234.66667                           0.0017
```

Parameter Estimates

```
Term              Estimate     Std Error    t Ratio    Prob>|t|
Intercept         11.222222    0.461107     24.34      <.0001
DrugL[Drug A]     -1.555556    0.652104     -2.39      0.0253
DrugL[Drug B]     -1.111111    0.652104     -1.70      0.1013
```

Effect Tests

```
Source    Nparm    DF    Sum of Squares    F Ratio    Prob > F
DrugL     2        2     96.888889         8.4387     0.0017
```

Effect Details
DrugL
Least Squares Means Table

```
Level      Least Sq Mean     Std Error      Mean
Drug A     9.666667          0.79866143     9.6667
Drug B     10.111111         0.79866143     10.1111
Placebo    13.888889         0.79866143     13.8889
```

Power of Hypothesis Tests

The power of an omnibus hypothesis test in ANOVA can be determined by using the formula for ϕ^2 to obtain a value for ϕ and using the power charts available in Keppel and many other texts. The value of ϕ depends on sample size, the magnitude of the treatment effects, as indexed by $\Sigma\alpha_i^2$, the number of groups, and the variance within groups, σ^2. The formula for ϕ^2 is:

$$\phi_A^2 = \frac{n\sum \alpha_i^2}{a\sigma^2} = \frac{n\sum\left(\mu_i - \mu\right)^2}{a\sigma^2} = \frac{n\delta^2}{\sigma^2}.$$

δ^2 here would be $\Sigma(\mu_i - \mu)^2 / a = \Sigma\alpha_i^2 / a$, that is, simply the variance of the population group means, or, equivalently, the variance of the treatment effects, and is a commonly used measure of the size of the treatment effect. JMP uses the values of δ, σ, and n to compute the power of hypothesis tests. (Actually, JMP uses the total $N = an$ in the experiment. Make sure you keep this in mind when you do power calculations in JMP.)

Example. Suppose one were interested in the effects of 2 weeks of computer assisted instruction (CAI) on math performance of 6th graders. One contemplates doing an experiment with 3 groups–one experimental and two controls. Math scores are measured in grade equivalent scores. If the experimenter believes it would be important to be able to detect in an experiment an effect that raised kids' scores by one full grade level in the two weeks of instruction, he/she might contemplate doing an experiment with $n = 20$ subjects per group, and be interested in the power of detecting the following size effect:

	Regular Class	CAI	Control Computer Class
Grade equivalent Math means	6	7	6

Suppose previous research suggests that grade equivalent math scores have a standard deviation of around 2 within regular classes. What is the probability (power) in an experiment with $n = 20$ of being able to detect an effect this large using $\alpha = .05$?

$$\mu = 6.33,\ \alpha_1 = -.33,\ \alpha_2 = .67,\ \alpha_3 = -.33,\ \sigma^2 = 4,\ n = 20.$$

$$\phi_A^2 = \frac{20\left[(-.33)^2 + .67^2 + (-.33)^2\right]}{3(4)} = 1.111$$

Therefore, $\phi_A = \sqrt{1.111}$ = 1.054, $df_{\text{den}} = 57$, and $df_{\text{num}} = 2$. Using Keppel's charts we see that here the power ≅ .30 for rejecting the null hypothesis at $\alpha = .05$. This is probably an unacceptable power. Usually, one would like to have at least a .80 chance of rejecting H_0 for an effect of interest. JMP can be used to get a more precise value for the power, which is .34. The JMP analysis is below:

```
Oneway Analysis of Mean By Group
Freq: n
Oneway Anova Power Details
Test: Group

Power
Alpha     Sigma     Delta       Number    Power
0.0500    2         0.471405    60        0.3371
0.0500    2         0.471405    70        0.3895
0.0500    2         0.471405    80        0.4403
0.0500    2         0.471405    90        0.4892
0.0500    2         0.471405    100       0.5357
0.0500    2         0.471405    110       0.5796
0.0500    2         0.471405    120       0.6207
0.0500    2         0.471405    130       0.6590
0.0500    2         0.471405    140       0.6944
0.0500    2         0.471405    150       0.7269
0.0500    2         0.471405    160       0.7567
0.0500    2         0.471405    170       0.7838
0.0500    2         0.471405    180       0.8084
```

```
Alpha      Sigma      Delta       Number     Power
0.0500     2          0.471405    190        0.8307
0.0500     2          0.471405    200        0.8507
0.0500     2          0.471405    210        0.8686
0.0500     2          0.471405    220        0.8846
0.0500     2          0.471405    230        0.8989
0.0500     2          0.471405    240        0.9116
```

The power can be increased by increasing sample size, *n.* The table above indicates that for a power of .80 the total *N* would need to be around 180 or 60 subjects/group. One can also use the DOE (Design of Experiments) platform in JMP to do power calculations for the omnibus test in a one-way ANOVA. Using that platform gives a sample size of *N* = 176 or 59 subjects/group as the sample size required for a power of .80 in this experiment. [*Caution:* Be sure you understand that the LSN (Least Significant Number) is *not* the sample size needed to attain whatever power you are looking for. See the JMP help files on that issue.]

To use Keppel's charts to find how large *n* needs to be for power = .80, one works backwards. Solving the power equation for *n* yields:

$$n = \frac{a\sigma^2\phi_A^2}{\sum \alpha_i^2} .$$

From the charts, we estimate $\phi_A = 1.8$. Thus $\phi_A^2 = 3.24$.

$$n = \frac{(3)(4)(3.24)}{.667} = 58.3 \approx 60 \text{ subjects/condition.}$$

Power for comparisons:

The power of the hypothesis test for a single *df* contrast may be obtained using JMP from the 'Fit Model' platform analysis, but not the 'Fit Y by X' platform. One first does a contrast analysis on the means of interest. [Be sure to include the appropriate *n*s in a Freq column for the analysis.] Then under Effect Details|Least Squares Means Table|LS Means Contrast|Power Analysis a Power Details Dialog box will have the correct value of δ calculated for the contrast. One needs only put in the assumed value for σ and proceed as usual.

In the example above, a sensible *a priori* contrast would be to compare the CAI group to the mean of the other 2 groups, so that the contrast coefficients would be: c_i: -1 2 -1. The output from a JMP analysis of this contrast is below:

Response Mean

Freq: n

Effect Details

Group Contrast

Test Detail

```
CAI          1
Control      -0.5
Regular      -0.5
Estimate     1
Std Error    1.4e-8
t Ratio      7.07e7
Prob>|t|     0
SS           13.333


Sum of Squares         13.333333333
Numerator DF           1
Denominator DF         57
F Ratio                5.0039996e15
Prob > F               0
```

Power Details

Test Contrast

Power

Alpha	Sigma	Delta	Number	Power
0.0500	2	0.471405	60	0.4346
0.0500	2	0.471405	70	0.4936
0.0500	2	0.471405	80	0.5485
0.0500	2	0.471405	90	0.5993
0.0500	2	0.471405	100	0.6457
0.0500	2	0.471405	110	0.6879
0.0500	2	0.471405	120	0.7260
0.0500	2	0.471405	130	0.7602
0.0500	2	0.471405	140	0.7908
0.0500	2	0.471405	150	0.8180
0.0500	2	0.471405	160	0.8421
0.0500	2	0.471405	170	0.8634
0.0500	2	0.471405	180	0.8820

Note that the power of the hypothesis test for this contrast with 20 subjects/group is .4346 (i.e., higher than the omnibus test power). Note also that the total *N* required for a power of .80 is between 140 and 150, or about 48 subjects/group.

The power of the hypothesis test for a single *df* contrast can also be obtained from Keppel's charts by using the formula (see also McFatter, R.M. & Gollob, H.F., 1986, *Educational and Psychological Measurement, 46,* 883-886):

$$\phi_\psi^2 = \frac{n\psi^2}{2\sigma^2\left(\sum c_i^2\right)} \text{ where } \psi = \sum c_i \mu_i \; .$$

As above, a sensible *a priori* contrast would be to compare the CAI group to the mean of the other 2 groups, so that the contrast coefficients would be: c_i: -1 2 -1.

Thus, in the example, ψ = (-1)(6) + (2)(7) + (-1)(6) = 2. And

$$\phi_\psi^2 = \frac{(20)(2^2)}{(2)(4)(6)} = 1.667, \qquad \phi_\psi = 1.29, \qquad df_{\text{num}} = 1$$

From the charts, *power* ≅ .43 for $\alpha = .05$.

Note the increased power when one tests for a specific pattern of means. To find the number of subjects needed to have power = .80, one solves to obtain:

$$n = \frac{2\phi_\psi^2\sigma^2\left(\sum c_i^2\right)}{\psi^2}.$$

Solving for *n* in the above example yields *n* = 48 subjects/condition.

Expected Mean Squares

MS_A and $MS_{S(A)}$ for an experiment are, of course, sample statistics, hence random variables. That is, they will vary from sample to sample (or experiment to experiment) even if the population parameters (i.e., μ and α_i) remain constant. Therefore, it is appropriate to consider the sampling distributions (over an infinite number of experiments) of MS_A and $MS_{S(A)}$.

In particular, it is of interest to consider the means or expected values of these two sampling distributions. It is easy to see what $E[MS_{S(A)}]$ would be by remembering what the $MS_{S(A)}$ is, namely, the average of the variances within the groups, and using the rules for expected values.

$$\begin{aligned} E\left[MS_{S(A)}\right] &= E\left[\frac{\hat{\sigma}_1^2+\hat{\sigma}_2^2+\hat{\sigma}_3^2+\ldots+\hat{\sigma}_a^2}{a}\right] \\ &= \frac{1}{a}\left(E\left[\hat{\sigma}_1^2\right]+E\left[\hat{\sigma}_2^2\right]+E\left[\hat{\sigma}_3^2\right]+\ldots+E\left[\hat{\sigma}_a^2\right]\right) \\ &= \frac{1}{a}\left(a\sigma^2\right)=\sigma^2. \end{aligned}$$

It can be shown by similar but slightly more involved algebra that

$$E\left[MS_A\right]=n\theta_A^2+\sigma^2 \text{ where } \theta_A^2=\frac{\sum\alpha_i^2}{a-1}.$$

θ_A^2 is called the *variance component* of A. In a *fixed* effects model (which is what we have been considering) the α_i's are considered to be the entire *population* of treatment effects we're interested in. The *variance* of the main effect parameters, α_i, would therefore be

$$\delta_A^2=\frac{\sum\alpha_i^2}{a}=\left(\frac{a-1}{a}\right)\theta_A^2.$$

Note that if H_0 is true, then $\alpha_i = 0$ and $\theta_A^2 = 0$. Note also that

$$\frac{E\left(MS_A\right)}{E\left(MS_{S(A)}\right)}=\frac{n\theta_A^2+\sigma^2}{\sigma^2}.$$

This ratio (of the same form as the *F*-ratio) will be equal to one when the null hypothesis is true and will be greater than one when H_0 is false.

It can also be shown that the population variance of the scores on the dependent variable of an experiment can be partitioned into treatment variance and error variance pieces. (This is another instance of *Data = Fit + Residual.*)

$$\sigma_Y^2=\delta_A^2+\sigma^2.$$

The proportion of the total variance of the dependent variable that is due to the treatment effect is called omega-squared and is equal to

$$\omega_A^2=\frac{\delta_A^2}{\delta_A^2+\sigma^2}.$$

Estimates of θ_A^2 and σ^2 can be obtained as follows:
Set

$$\begin{aligned} MS_A &= n\hat{\theta}_A^2+\hat{\sigma}^2 \\ MS_{S(A)} &= \hat{\sigma}^2 \end{aligned}$$

and solve to get:

$$\hat{\theta}_A^2 = \frac{MS_A - MS_{S(A)}}{n}$$ [If the numerator is negative, set it equal to zero.]

Estimates of δ_A^2 and ω_A^2 can then be obtained as follows:

$$\hat{\delta}_A^2 = \left(\frac{a-1}{a}\right)\hat{\theta}_A^2$$

$$\hat{\omega}_A^2 = \frac{\hat{\delta}_A^2}{\hat{\delta}_A^2 + \hat{\sigma}^2}.$$

Examples. Let $a = 3$, $n = 20$.

Source	*SS*	*df*	*MS*	*F*	
A	100	2	50	5.00	$p < .01$
S(A)	570	57	10		
Total	670	59	11.36		

$$\hat{\theta}_A^2 = \frac{50-10}{20} = 2.00, \qquad \hat{\delta}_A^2 = \left(\frac{3-1}{3}\right)2.00 = 1.333$$

$$\hat{\sigma}^2 = 10 \qquad \hat{\omega}_A^2 = \frac{1.333}{1.333+10} = .12 \text{ or } 12\%.$$

The treatment accounts for 12% of the variance of the dependent variable.
Now consider the following experiment: $a = 3$, $n = 5$.

Source	*SS*	*df*	*MS*	*F*	
A	40	2	20	2.00	*ns.* Note nonsignificant *F*.
S(A)	120	12	10		
Total	160	14	11.43		

$$\hat{\theta}_A^2 = \frac{20-10}{5} = 2.00, \qquad \hat{\delta}_A^2 = \left(\frac{3-1}{3}\right)2.00 = 1.333$$

$$\hat{\sigma}^2 = 10 \qquad \hat{\omega}_A^2 = \frac{1.333}{1.333+10} = .12 \text{ or } 12\%.$$

Note that the size of effect estimates are identical for these two experiments even though one F is significant while the other is not. This illustrates the fact that the question of statistical significance is an entirely different question from that of whether there is a large effect or not.

Sometimes one will see $R^2 = SS_A/SS_{Tot}$ used as an index of the proportion of variance accounted for by the treatment. In the 2 examples above these values would be, respectively, .15 and .25—both larger than the more appropriate omega-squared estimates. In general R^2 will overestimate the proportion of variance due to the treatment, and omega-squared is preferred.

ω_A^2 can be misleading, however, especially in experiments in which the levels of the independent variable are under the experimenter's control. The value of ω_A^2 can be raised or lowered simply by including different levels in the experiment. Note Keppel's arguments concerning this issue. In many cases it may be safest to simply talk about size of effect by referring to mean differences or α_i's rather than ω_A^2.

Two-way Factorial ANOVA

In this design several groups of **different** subjects are formed from a "factorial" arrangement of two independent variables. A score on the dependent variable is obtained for each subject. Each factor has two or more levels. It is essential that scores in one group be **independent** of scores in other groups. That is, there should be no sense in which scores in one group could be meaningfully paired with scores in any other group.

***Example structural layout**:* 3 subjects per group in a 2 x 3 factorial design

		FACTOR B		
	LEVEL	1	2	3
FACTOR A	1	S_1 S_2 S_3	S_7 . .	S_{13} . .
	2	S_4 S_5 S_6	. . S_{12}	. . S_{18}

***Randomized experimental form**:* Subjects are **randomly assigned** to be in one of the experimental cells. (There are 6 cells in the above example.) The experimenter **manipulates** the independent variables by determining the specific treatment combination each group receives.

Example: Subjects are randomly assigned to receive one of 2 sets of instructions (Factor A) and one of 3 levels of drug dosage: placebo, 50mg, or 100mg (Factor B). Scores on a cognitive performance test are obtained for each subject.

***Quasi-experimental (correlational) form**:* Each group of subjects is treated as a **random sample** from a preexisting population. The experimenter does **not** manipulate which group the subject is in, but simply **observes** which group the subject is in.

Example: Reading ability scores are obtained along with a measure of socioeconomic status (SES) for a large sample of subjects. Subjects are divided into males and females (Factor A) and 3 levels of SES (Factor B) in order to compare average reading ability for different SES levels and sex groups.

***Analysis of Variance**:* n = # of subjects per cell; a = # of levels of Factor A; b = # of levels of Factor B

Source	*df*	*Error Term for **F**-test*
A	$a - 1$	S(AB)
B	$b - 1$	S(AB)
A × B	$(a - 1)(b - 1)$	S(AB)
S(AB)	$ab(n - 1)$	None
Total	$abn - 1$	

The model underlying the two-way ANOVA is

$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, where the ε_{ijk} are $N(0, \sigma^2)$, and

$i = 1,2, \ldots a$
$j = 1,2, \ldots b$
$k = 1,2, \ldots n$

Because of the factorial structure of the means, the μ_{ij} may be parameterized as follows:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \text{ where}$$

$$\sum \alpha_i = 0, \qquad \sum \beta_j = 0$$

$$\sum_i (\alpha\beta)_{ij} = 0 \text{ for all } j, \text{ and } \sum_j (\alpha\beta)_{ij} = 0 \text{ for all } i.$$

The α_i and β_j are called *'main effect'* parameters; the $(\alpha\beta)_{ij}$ are called *'interaction'* parameters.

Consider a two-way factorial ($a = 3$, $b = 4$) experimental population with the following parameters:

$\mu = 5$
$\alpha_i = \{-3, 1, 2\}$
$\beta_j = \{3, 0, -1, -2\}$

$(\alpha\beta)_{ij}$	b_1	b_2	b_3	b_4
a_1	1	-2	0	1
a_2	1	0	-2	1
a_3	-2	2	2	-2

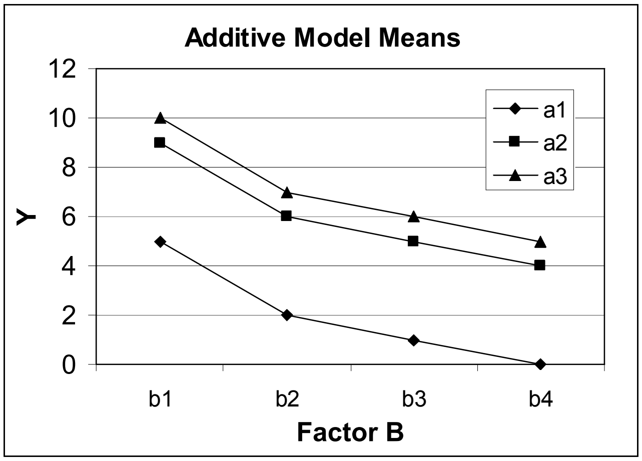
Note. Rows and columns must sum to zero.

The *"additive"* (no interaction) model means that would be implied by the main effect and grand mean parameters above would be as follows:

Additive Model Means

μ_{ij}	b_1	b_2	b_3	b_4
a_1	5	2	1	0
a_2	9	6	5	4
a_3	10	7	6	5

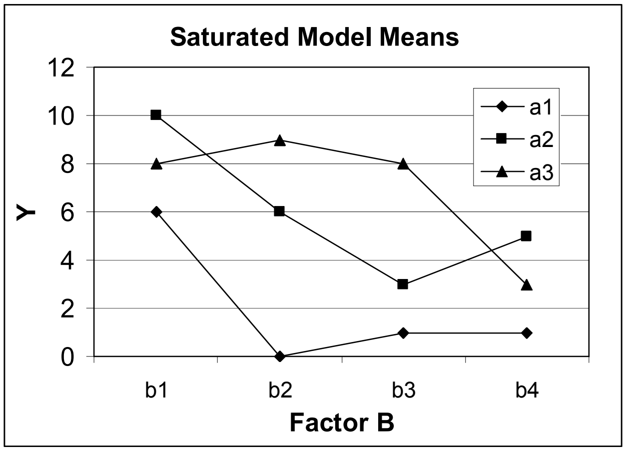
A plot of these means would look like the following:



Note the parallel lines. This will always be the case with an "additive" model, i.e., one with no interaction.

Now adding in the $(\alpha\beta)_{ij}$ interaction parameters leads to the *"saturated"* model means below:
Saturated Model Means

μ_{ij}	b_1	b_2	b_3	b_4
a_1	6	0	1	1
a_2	10	6	3	5
a_3	8	9	8	3



Note the nonparallel lines. This will always be the case when there is a nonzero interaction.

It is possible to obtain least squares estimators of the model parameters. They turn out to be:

$$\hat{\mu} = \overline{T}$$

$$\hat{\alpha}_i = \overline{A}_i - \overline{T}$$

$$\hat{\beta}_j = \overline{B}_j - \overline{T}$$

$$\widehat{(\alpha\beta)}_{ij} = \overline{AB}_{ij} - \overline{A}_i - \overline{B}_j + \overline{T}$$

$$\hat{\varepsilon}_{ijk} = Y_{ijk} - \overline{AB}_{ij}$$

Example. Consider a social cognition experiment. Ss in 4 (2×2) conditions make good-bad ratings (-7 to +7) of adjective-noun combinations. The following mean ratings are obtained:

		B		
		Crook	Nurse	$\overline{A}_i$
A	Cruel	-5	-3	-4
	Friendly	-3	3	0
	$\overline{B}_j$	-4	0	$\overline{T} = -2$

$$\hat{\mu} = \overline{T} = -2$$

$$\hat{\alpha}_i = \overline{A}_i - \overline{T} = \{-2, +2\}$$

$$\hat{\beta}_j = \overline{B}_j - \overline{T} = \{-2, +2\}$$

$$\widehat{(\alpha\beta)}_{ij} = \overline{AB}_{ij} - \overline{A}_i - \overline{B}_j + \overline{T} =$$

	$\widehat{(\alpha\beta)}_{ij}$	B	
		Crook	Nurse
A	Cruel	1	-1
	Friendly	-1	1

For example,

$$\widehat{(\alpha\beta)}_{11} = -5-(-4)-(-4)+(-2)=1$$

The $\hat{\alpha}_i$ main effect estimates may be interpreted as the effects of "cruel" and "friendly," respectively, on the ratings of the combinations. The $\hat{\beta}_j$ main effect estimates are similarly interpreted as the effects of "crook" and "nurse" on the ratings. The $\widehat{(\alpha\beta)}_{ij}$ interaction estimates may be interpreted as the additional effects on ratings due to idiosyncrasies arising from the particular adjective-noun combinations. For example, the ones for "cruel crook" and "friendly nurse" might suggest that subjects evaluate these combinations more highly than one would expect solely on the basis of the adjectives and nouns alone because the adjectives are consonant with role expectations. The minus ones for "cruel nurse" and "friendly crook" might suggest that unnerving combinations are evaluated more negatively than might be expected based solely on the simple adjectives and nouns. This can be clearly seen by considering what the *additive* model predicts the means should be:

		B	
		Crook	Nurse
A	Cruel	-6	-2
	Friendly	-2	2

Above are means predicted solely from *additive* effects. NB. Actual means are perfectly predicted by the saturated (including interaction) model.

The null hypotheses tested in a two-way factorial design are as follows:

A main effect: H$_0$: $\mu_{a_1} = \mu_{a_2} = \ldots$ or $\alpha_i = 0$.

B main effect: H$_0$: $\mu_{b_1} = \mu_{b_2} = \ldots$ or $\beta_j = 0$.

AB interaction: H$_0$: parallel lines or $(\alpha\beta)_{ij} = 0$.

Expected Mean Squares

The expected mean squares for the two-way fixed effects factorial ANOVA are given below:

$E(MS_A) = bn\,\theta_A^2 + \sigma^2$

$E(MS_B) = an\,\theta_B^2 + \sigma^2$

$E(MS_{AB}) = n\,\theta_{AB}^2 + \sigma^2$

$E(MS_{S(AB)}) = \sigma^2$

Note that if the H$_0$ is true then the θ^2s would equal zero. The $E(MS)$'s show that using the $MS_{S(AB)}$ as the denominator of the F-ratios would lead to F's of about 1 when the H$_0$'s are true.

where $$\theta_A^2 = \frac{\sum \alpha_i^2}{a-1}$$

$$\theta_B^2 = \frac{\sum \beta_j^2}{b-1}$$

$$\theta_{AB}^2 = \frac{\sum\sum (\alpha\beta)_{ij}^2}{(a-1)(b-1)}.$$

Estimates of the θ^2 s may be obtained as in the one-way case by solving the equations below:

$$MS_A = bn\hat{\theta}_A^2 + \hat{\sigma}^2 \qquad MS_B = an\hat{\theta}_B^2 + \hat{\sigma}^2$$

$$MS_{AB} = n\hat{\theta}_{AB}^2 + \hat{\sigma}^2 \qquad MS_{S(AB)} = \hat{\sigma}^2$$

This leads to:

$$\hat{\theta}_A^2 = \frac{MS_A - MS_{S(AB)}}{bn} \qquad \hat{\theta}_B^2 = \frac{MS_B - MS_{S(AB)}}{an}$$

$$\hat{\theta}_{AB}^2 = \frac{MS_{AB} - MS_{S(AB)}}{n} \qquad \hat{\sigma}^2 = MS_{S(AB)}.$$

If a numerator is negative, set it equal to zero.

And

$$\hat{\delta}_A^2 = \left(\frac{a-1}{a}\right)\hat{\theta}_A^2 \qquad \hat{\omega}_A^2 = \frac{\hat{\delta}_A^2}{\hat{\delta}_A^2 + \hat{\delta}_B^2 + \hat{\delta}_{AB}^2 + \hat{\sigma}^2}, \text{ etc.}$$

$$\hat{\delta}_B^2 = \left(\frac{b-1}{b}\right)\hat{\theta}_B^2$$

$$\hat{\delta}_{AB}^2 = \frac{(a-1)(b-1)}{ab}\hat{\theta}_{AB}^2$$

Simple Main Effects Analyses

In interpreting interactions it is sometimes useful to do simple effects analyses.
Example. Two categories of patient (A) and three drugs (B); $n = 3$.

		Drug (B)		
		b_1	b_2	b_3
Category	Schizophrenics	8 4 0	10 8 6	8 6 4
(A)	Normals	14 10 6	4 2 0	15 12 9

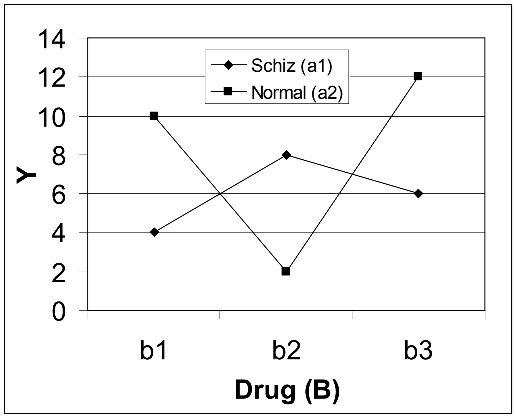
The usual ANOVA summary table for this data is:

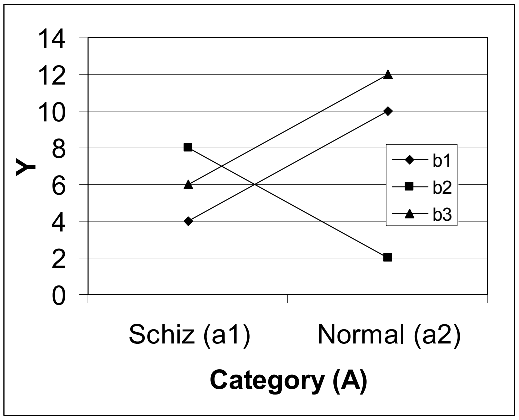
Source	*SS*	*df*	*MS*	*F*
A	18	1	18.00	2.04
B	48	2	24.00	2.72
AB	144	2	72.00	8.15, $p < 01$
S(AB)	106	12	8.83	

AB Summary Table

	b_1	b_2	b_3	Σ
a_1	12	24	18	54
a_2	30	6	36	72
Σ	42	30	54	126

Note profiles of Drug simple effects and Category simple effects:





$$SS_{B\,at\,a_1} = \frac{12^2 + 24^2 + 18^2}{3} - \frac{54^2}{9} = 24.00$$

$$SS_{B\,at\,a_2} = 168.00$$

ANOVA Summary for simple B main effects:

Source	*SS*	*df*	*MS*	*F*
B at a_1	24	2	12.00	1.36 *ns.*
B at a_2	168	2	84.00	9.51 $p < .01$
S(AB)	106	12	8.83	

ANOVA Summary for simple A main effects:

Source	*SS*	*df*	*MS*	*F*
A at b_1	54	1	54.00	6.11 $p < .01$
A at b_2	54	1	54.00	6.11 $p < .01$
A at b_3	54	1	54.00	6.11 $p < .01$
S(AB)	106	12	8.83	

Note well from the simple B main effects table that

$$\begin{array}{rcll} 24 + 168 & = & 48 + 144 & \text{or} \\ \Sigma SS_{B\ at\ ai} & = & SS_B + SS_{AB} & \end{array}$$

Note also from the simple A main effects table that

$$\begin{array}{rcll} 54 + 54 + 54 & = & 18 + 144 & \text{or} \\ \Sigma SS_{A\ at\ bj} & = & SS_A + SS_{AB} & \end{array}$$

The interpretation of the interaction that would be derived from the simple B main effects analysis would be that there is *no* significant drug effect for schizophrenics ($F = 1.36$, *ns.*), whereas there *is* a significant drug effect for normals ($F = 9.51$, $p < .01$).

The interpretation of the interaction that would be derived from the simple A main effects analysis would be that there are significant differences between mean scores for schizophrenics and normals for all three drugs (all *F*s = 6.11, $p < .01$), but that the difference for drug b_2 is just the opposite from that of the other two drugs.

Simple effects analyses are carried out in JMP by using the 'Test Slices' option associated with the Least Squares Means table from the interaction. This will be under 'Effect Details' in the 'Fit Model' analysis.

The JMP output for the data above is as follows:

```
Response Y

Analysis of Variance
Source      DF     Sum of Squares          Mean Square        F Ratio
Model       5      210.00000               42.0000            4.7547
Error       12     106.00000               8.8333             Prob > F
C. Total    17     316.00000                                  0.0126

Parameter Estimates
Term                                        Estimate      Std Error     t Ratio     Prob>|t|
Intercept                                   7             0.700529      9.99        <.0001
Category (A)[Normal]                        1             0.700529      1.43        0.1789
Drug (B)[b1]                                0             0.990697      0.00        1.0000
Drug (B)[b2]                                -2            0.990697      -2.02       0.0664
Category (A)[Normal]*Drug (B)[b1]           2             0.990697      2.02        0.0664
Category (A)[Normal]*Drug (B)[b2]           -4            0.990697      -4.04       0.0016

Effect Tests
Source                    Nparm      DF      Sum of Squares          F Ratio       Prob > F
Category (A)              1          1       18.00000                2.0377        0.1789
Drug (B)                  2          2       48.00000                2.7170        0.1063
Category (A)*Drug (B)     2          2       144.00000               8.1509        0.0058
Effect Details
Category (A)
Least Squares Means Table
Level      Least Sq Mean           Std Error          Mean
Normal     8.0000000               0.99069747         8.00000
Schizo     6.0000000               0.99069747         6.00000
Drug (B)
Least Squares Means Table
Level    Least Sq Mean           Std Error          Mean
b1       7.0000000               1.2133516          7.00000
b2       5.0000000               1.2133516          5.00000
b3       9.0000000               1.2133516          9.00000
```

```
Category (A)*Drug (B)
Least Squares Means Table
Level            Least Sq Mean                Std Error
Normal,b1        10.000000                    1.7159384
Normal,b2        2.000000                     1.7159384
Normal,b3        12.000000                    1.7159384
Schizo,b1        4.000000                     1.7159384
Schizo,b2        8.000000                     1.7159384
Schizo,b3        6.000000                     1.7159384
Slice Category (A)=Normal

Sum of Squares           168
Numerator DF             2
Denominator DF           12
F Ratio                  9.5094339623
Prob > F                 0.0033522146
Slice Category (A)=Schizo

Sum of Squares           24
Numerator DF             2
Denominator DF           12
F Ratio                  1.358490566
Prob > F                 0.293883415
Slice Drug (B)=b1

Sum of Squares           54
Numerator DF             1
Denominator DF           12
F Ratio                  6.1132075472
Prob > F                 0.0293592734
Slice Drug (B)=b2

Sum of Squares           54
Numerator DF             1
Denominator DF           12
F Ratio                  6.1132075472
Prob > F                 0.0293592734
Slice Drug (B)=b3

Sum of Squares           54
Numerator DF             1
Denominator DF           12
F Ratio                  6.1132075472
Prob > F                 0.0293592734
```

Note, also, the Parameter Estimates table above. Make sure that you know what each estimate means and how you could get the complete set by using the ones given.

This page intentionally left blank.

Interaction Contrasts

Consider the following example. A researcher is interested in studying reading ability as a function of socioeconomic status (S) and grade (G) in school for children in grades 1 - 4. Below are cell *totals* on a reading ability test. There are 5 subjects/cell.

		Grade (G)				
		1	2	3	4	Σ
	High	10	20	30	40	100
SES (S)	Med	5	10	15	20	50
	Low	3	5	7	9	24
	Σ	18	35	52	69	174

The cell and marginal means associated with this data set would be:

		Grade (G)				
		1	2	3	4	Mean
	High	2	4	6	8	5
SES (S)	Med	1	2	3	4	2.5
	Low	0.6	1	1.4	1.8	1.2
	Mean	1.2	2.333	3.467	4.6	2.9

The usual main effect and interaction omnibus effects may be partitioned into trend components in much the same way as in a one-way design. Assume equally spaced levels for SES. The summary table would look like the following:

Source	*SS*	*df*	*MS*
S	149.20	2	74.60
linear	144.40	1	144.40
quadratic	4.80	1	4.80
G	96.33	3	32.11
linear	96.33	1	96.33
quadratic	0.00	1	0.00
cubic	0.00	1	0.00
SxG	32.67	6	5.44
$S_{lin}\times G_{lin}$	32.00	1	32.00
$S_{quad}\times G_{lin}$	0.67	1	0.67
$S_{lin}\times G_{quad}$	0.00	1	0.00
$S_{quad}\times G_{quad}$	0.00	1	0.00
$S_{lin}\times G_{cub}$	0.00	1	0.00
$S_{quad}\times G_{cub}$	0.00	1	0.00

The student should verify that these *SS* are correct for the above data. Note that the within cell data are not available so computations in JMP would need to use the means as the *Y* variable with the Freq column specified with the cell *n*'s.

The computational formula for computing the *SS* for any contrast by calculator is as follows:

$$SS_{\psi} = \frac{\left(\sum c_i T_i\right)^2}{n \sum c_i^2},$$

where T_i is some cell or marginal total, and *n* is the # of observations summed to get T_i.

If one uses means instead of totals, the *n* just goes in the numerator instead of the denominator.

Example. *SS* for S_{linear} would be calculated as follows:

$$SS_{S_{linear}} = \frac{\left[1(100) + 0(50) + (-1)(24)\right]^2}{20(2)} = 144.40$$

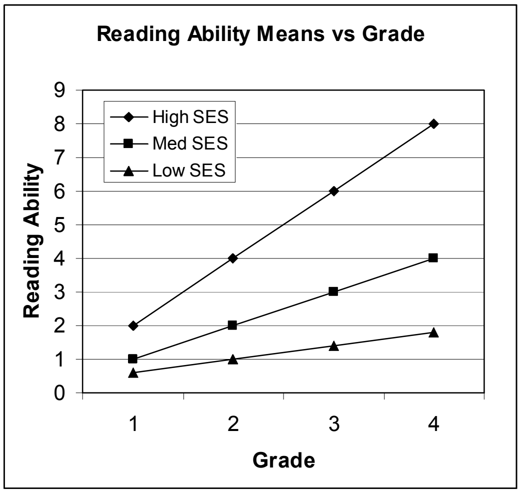
In JMP, one would need to go to the Effects Detail|SES Effect and specify a LSMeans contrast.

The interpretation of this source, if significant, would be that, *averaged over all grades,* reading scores tend to increase linearly with SES level. That is, this tests the linear trend in the three *marginal* SES means. The interpretation of the $G_{quadratic}$ source, if significant, would be that, *averaged over all levels of SES,* reading scores show a quadratic trend over the four Grade levels. That is, it tests the quadratic trend in the *marginal* Grade means.

The interpretation of the interaction single *df* contrasts requires one to have a clear understanding of what an interaction is and how to interpret it. An S×G interaction means that the pattern of the effect of SES on Reading differs depending on which Grade level one looks at. By partitioning the 6 *df* associated with the S×G interaction into single df contrasts one can specify precisely how the pattern of the effect of SES on Reading differs depending on Grade level.

For example, a significant $G_{lin} \times S_{lin}$ source would imply that the *rate of increase* in reading scores over grades 1-4 is *linearly* related to SES.

Consider the plot of cell means for the above data:



Note that the slopes of the lines (over Grade levels, i.e., the G_{lin} trends) increase *linearly* with SES (i.e., from low SES to high SES). This pattern would lead to a strong $G_{lin} \times S_{lin}$ component of the interaction. That this is true can be seen from the summary table above for this data. Almost all the interaction between G and S is concentrated in this pattern.

If, on the other hand, one had means in which the *rate of increase* (slope) in reading scores over Grade levels was high for middle class (Med SES) kids, but relatively flatter for both High and Low SES kids, then one would expect the $G_{lin} \times S_{quad}$ component of the G×S interaction to be significant.

The computation of *SS* for interaction contrasts uses the same formula as for other contrasts. The weights are applied to cell means (or totals), however, rather than marginal means (or totals). The weights are obtained by multiplying the coefficients of the individual contrasts. For example, the weights for the $G_{lin} \times S_{quad}$ would be:

		Grade (G) *Linear*			
		-3	-1	+1	+3
	High -1	+3	+1	-1	-3
SES (S)	Med +2	-6	-2	+2	+6
Quadratic	Low -1	+3	+1	-1	-3

It is also possible to partition the 6 *df* for the S×G interaction into components as follows:

Source	*df*		*Source*	*df*
S × G	6		S × G	6
$S \times G_{linear}$	2	or	$G \times S_{linear}$	3
$S \times G_{quad}$	2		$G \times S_{quad}$	3
$S \times G_{cubic}$	2			

The interpretation of a significant $S \times G_{linear}$ source would be that the linear trend (slope) over grades in reading ability differs among SES levels, but in an unspecified pattern.

To compute the *SS* for an $A_\psi \times B$ source one could use the following formula (based on the simple effects analyses described earlier):

$$SS_{A_\psi \times B} = \sum SS_{A_\psi \, at \, b_j} - SS_{A_\psi}$$

or simply sum the *SS* for the single *df* contrasts that partition the $A_\psi \times B$ source.

Multiple Comparisons for Factorial Designs

The issues are the same for multiple comparisons for higher order designs as they are for the one-way design. The formulas are also the same except that in place of *n* in the one-way formulas use the number of observations each mean you are comparing is based on.

Three-way Factorial Designs

The basic principles of two-factorial designs are easily generalized to higher order factorial designs. The three-way factorial design [S(ABC)] would have the following ANOVA summary table:

Source	*df*	*Computational Formulas*
A	*a* - 1	[A] - [T]
B	*b* - 1	[B] - [T]
C	*c* - 1	[C] - [T]
AB	(*a* - 1)(*b* - 1)	[AB] - [A] - [B] + [T]
AC	(*a* - 1)(*c* - 1)	[AC] - [A] - [C] + [T]
BC	(*b* - 1)(*c* - 1)	[BC] - [B] - [C] + [T]
ABC	(*a* - 1)(*b* - 1)(*c* - 1)	[ABC] - [AB] - [AC] - [BC] + [A] + + [B] + [C] - [T]
S(ABC)	*abc*(*n* - 1)	[Y] - [ABC]

In interpretation, the highest order interactions are interpreted first. A significant three-way interaction, e.g., ABC, may be interpreted by noting that it means that the BC interaction differs in pattern from one level of A to another. Or, equivalently, the AB interaction differs in pattern from one level of C to another. For example, if there were no AB interaction at c_1, but there was a large AB interaction at c_2, this would lead to a three-way ABC interaction.

More Complex Designs

In considering more complex ANOVA designs it is necessary to distinguish between *fixed* and *random* factors, and between *crossed* and *nested* factors.

Fixed vs. Random

A *fixed* factor is a factor whose levels are considered to be the entire population of levels about which one is interested in drawing inferences. A *random* factor, on the other hand, is one whose levels are considered to be a random sample from the population of levels about which one is interested in drawing inferences.

For example, when Subjects is considered as a factor in a design, it is almost always considered as a random factor because one is usually interested in drawing inferences about some population of subjects from which the subjects in the experiment are considered to be a random sample. When, on the other hand, subjects are randomly assigned to different conditions of an experiment (e.g., different levels of drug dosage) the Condition factor is usually considered to be a fixed factor because the levels of Condition have not been randomly selected from a large population of levels, but *are* the entire population of levels about which the experimenter wishes to draw inferences.

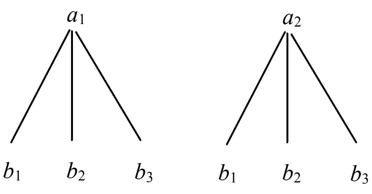
Whether a factor is fixed or random has implications for the construction of appropriate F-ratios as will be shown later. All of the factors (other than Subjects) that have been considered up until now have been assumed to be fixed. Some factors that are sometimes considered to be random factors in experiments are Classes, Schools, Experimenters, Nouns, Verbs, Lists, Faces, Replications, etc.

Crossed vs. Nested

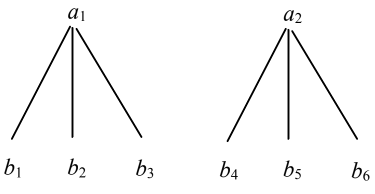
Two factors, A and B, are said to be *crossed* if and only if ***every level of A occurs at every level of B***. The two factors are written A x B or AB. Factor B is said to be *nested* in factor A if and only if ***several different levels of B occur at each level of A***. The factors are written B(A).

Example.

A and B Crossed
A×B



B Nested in A
B (A)



Note that in the simple one-way design, S(A), subjects are nested in treatments. In the two-way factorial design, S(AB), A is crossed with B, and subjects are nested in A and in B.

There are many possible designs that result from various combinations of crossing and nesting. For example:

S(A)	- ordinary one-way between-subjects design
S(AB)	- two-way factorial between-subjects design
S(ABC)	- three-way factorial between-subjects design
S×A	- one-factor repeated measures (within-subjects) design
S×A×B	- two repeated measure factors design
S(A(B))	- hierarchical (nested) design
S(A)B	- split-plot design (one between- and one within-subjects factors)

Consider the following situation: A researcher carries out an experiment looking at all 4 schools in his district. From each school 3 classrooms are selected at random. From each classroom 2 children are selected at random and given a pre-test prior to treatment and a post-test after treatment. What is the ANOVA design that is appropriate for analyzing this experiment?

Let A = classroom
B = school
S = subject
P = pre- vs post-test
r = random
f = fixed

The design would be:

r r f f
S(A(B)) × P
2 3 4 2 <---- # of levels of each factor
NB. These multiply to equal the total # of obs. in the experiment.

Effect of Random Factors on the Analysis

It's quite instructive to think carefully about Keppel's discussion of the effect of random factors on the analysis. He shows that when a fixed factor is crossed with a random factor the *MS* for the fixed factor is contaminated by any interaction that exists between the two factors. Thus, the *E*(*MS*) for the fixed factor will be increased by the addition of a term for the interaction. Consider the *E*(*MS*) for this design:

r f r
S(AB)
n a b

Source	*E(MS)*			
A	$nb\,\theta_A^2$ +	$n\,\sigma_{AB}^2$	+	$\sigma^2_{S(AB)}$
B	$na\,\sigma_B^2$		+	$\sigma^2_{S(AB)}$
AB		$n\,\sigma_{AB}^2$	+	$\sigma^2_{S(AB)}$
S(AB)				$\sigma^2_{S(AB)}$

Note that whenever a source contains *only* fixed factors, its variance component is labeled with a θ^2. If a source contains *any* random factors, its variance component is labeled with a σ^2. Note also that the appropriate error term for testing the fixed A main effect is the AB interaction because if the H_0 for the A main effect were true (i.e., $\theta_A^2 = 0$) the ratio of the $E(MS_A)$ to the $E(MS_{AB})$ would be equal to one. The correct error term for testing the random B main effect, on the other hand, would still be the $MS_{S(AB)}$ for the same reason.

JMP Analysis of Designs with Random Factors

Although JMP does allow the specification of random effects in the Fit Model platform (In the Construct Model box, there is a red triangle next to the word 'Attributes'), it has difficulty in picking the correct error term for random effects when there are fixed factors also in the design. In the case of the design above, S(AB) with A fixed and B random, the correct *F* is computed for the fixed factor A (i.e., the AB term is used as the denominator of the *F*-ratio), but the test for the random B factor incorrectly also uses the AB term for the denominator of the *F*-ratio. The computation of *SS*, *df*, and *MS* are correct, but the *F*-tests may (depending on the specific design used) use the wrong effect for an error term in the denominator of the *F*-ratio. This is usually not as big a problem as it might seem because one is generally interested primarily in tests of the fixed effects in the model, which are correctly computed.

JMP Analysis of Designs with Nested Factors

JMP also allows the specification of nested factors in the Construct Model box with the 'Nest' button using a notation that is similar the one used above. In general, one should always specify factors that are nested in other factors as random and pay attention to what error terms were used in computing the *F* values that are reported in the output.

For both random and nested factors, there appears to be no substitute for generating the *E(MS)* yourself for the design and knowing which terms are the appropriate error terms for each effect. It is easy to tell whether JMP has used the appropriate error term because it tells you in the output which effect it used for the denominator of each *F*-test when there are random factors in the model. Because the *MS* for the effects are correct, it is always possible to compute the correct *F* by using the appropriate *MS* for the denominator of the *F*. See the *Rules of Thumb* section at the back for rules in generating the ANOVA summary table, *E*(*MS*), error terms, etc. for common designs.

Repeated Measures Designs

A repeated measure (or 'within-subjects') factor in a design is a factor that is crossed with subjects. That is, each subject receives all levels of that factor. Designs with these kinds of factors are quite common in psychology experiments.

The simplest repeated measure design is the S × A design. Each subject receives all levels of A and is measured once at each level of A. Factor A is usually considered to be fixed, (S is random) and we will assume this in what follows.

The model that underlies the simple S × A repeated measures design is:

$$Y_{ij} = \mu + \alpha_i + \pi_j + \varepsilon_{ij}$$

where the π_j are Subject (i.e., Person) main effect parameters and are assumed to be a *random* sample from a population of Subject main effect parameters. Thus, Subjects is a random factor while A (Treatment) is assumed to be fixed. The terms π_j and ε_{ij} are assumed to be normally distributed with means of zero.

Actually, to be conceptually complete we probably ought to include a term in the model for the $(\alpha\pi)_{ij}$ interaction, but because we have only one observation in each cell of the two-way S × A table, it's impossible to get separate estimates of the interaction and the within-cells error. Therefore, we must lump them together and call them both ε_{ij}.

If we had more than one replication of each subject's scores in each treatment, i.e., a R(SA) design, then it would be possible to separate interaction from error, and a model that includes a separate interaction term would be appropriate.

An additional assumption of *any design that has a fixed factor crossed with a random factor*, such as a repeated measures design, is an assumption variously called "homogeneity of covariance," "sphericity," "compound symmetry," and probably other names as well. If one thinks about the matrix of covariances among the levels of the fixed factor, a sufficient (though not always necessary) condition for this assumption to be met is that all entries in the covariance matrix be equal.

For example, if $a = 4$ then the covariance and correlation matrices of the levels of A will be 4 × 4 matrices:

$$\Sigma = \begin{array}{ccccc} & a_1 & a_2 & a_3 & a_4 \\ a_1 & \sigma_{11} & \text{(symm.)} & & \\ a_2 & \sigma_{21} & \sigma_{22} & & \\ a_3 & \sigma_{31} & \sigma_{32} & \sigma_{33} & \\ a_4 & \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{array} \qquad R = \begin{array}{ccccc} & a_1 & a_2 & a_3 & a_4 \\ a_1 & 1 & \text{(symm.)} & & \\ a_2 & \rho_{21} & 1 & & \\ a_3 & \rho_{31} & \rho_{32} & 1 & \\ a_4 & \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{array}$$

NB. ρ here indicates a population correlation.

If all entries in the covariance matrix are equal, then the variances in each level of A are all equal, and the correlations in the correlation matrix are all equal to one another. Note that these conditions are likely to be violated in many common types of psychology experiments.

Unlike the homogeneity of variance assumption associated with between-subjects ANOVA, within-subjects F-tests are *not* robust with respect to violations of the sphericity assumption.

Three ways of dealing with violations of sphericity assumption

1) Greenhouse-Geisser (or Box) test.

a) Example: For testing the C main effect in the S(G)C design, compute usual F-ratio, but look up critical F using $df = [1, g(n - 1)]$ instead of the usual $[(c - 1), g(n - 1)(c - 1)]$, i.e., replace $(c - 1)$ with 1. This is conservative.

b) Find an estimate $\hat{\varepsilon}$, using either formulas provided by Greenhouse & Geisser or Huynh & Feldt, and look up critical F using $df = [\hat{\varepsilon}\ (c - 1),\ \hat{\varepsilon}\, g(n - 1)(c - 1)]$. This seems to be a very satisfactory approach and is provided routinely by many computer programs. A value of $\hat{\varepsilon}$ close to one means the assumption is not badly violated, while values closer to zero indicate more severe violation of the sphericity assumption.

2) Construct *a priori* single *df* contrasts for the repeated measures factor and use the appropriate partitioned error term to test each one individually. This is also a good solution when there are a sufficient number of *df* in the partitioned error terms to provide reasonable power.

For example, in Problem #1 of homework 7, one would appropriately test the researcher's hypothesis by testing the $C_{\text{linear}} \times G$ component of the CG interaction. The error term would be the *MS* due to $S(G)C_{\text{linear}}$. The ANOVA summary might look like this:

Source	*df*
G	1
S(G)	38
C	6
CG	6
$G \times C_{linear}$	1
$G \times C_{residual}$	5
S(G)C	228
$S(G)C_{linear}$	38
$S(G)C_{residual}$	190

Note: Only fixed factors that are crossed with random factors require this partitioning of error terms. Moreover, factors with only 2 levels automatically meet the sphericity assumptions.

3) Use MANOVA (multivariate ANOVA) treating individual single *df* contrasts as multiple dependent variables. Not usually as powerful as above approaches, and somewhat more complex to carry out and interpret.

Using JMP to Analyze Designs with Within-Subjects Factors

There are several ways to analyze designs with within-subjects factors (fixed factors crossed with the random factor, Subjects) in JMP. The most satisfactory and flexible approach is to have the data laid out with *all* the scores for a given subject on the same row (i.e., one row per subject). Thus the columns would represent levels of the repeated measures factor(s). All of the columns representing the levels of the within-subjects factor(s) are listed as Y variables in the Fit Model platform. If there are between-subjects factors, their effects are listed in the Construct Model box. The Manova 'Personality' is selected to carry out the analysis.

[This approach is also applicable whenever there is a random factor crossed with a fixed factor in the design. Simply treat the levels of the random factor as if they were subjects in a repeated measures design, and arrange the data so that the levels of the fixed factor are the columns in the table with each level of the random factor taking up only one row. The analysis would be directly analogous to the repeated measures analysis described below.]

Below is an example of the analysis of a simple S×A repeated measures design using JMP.

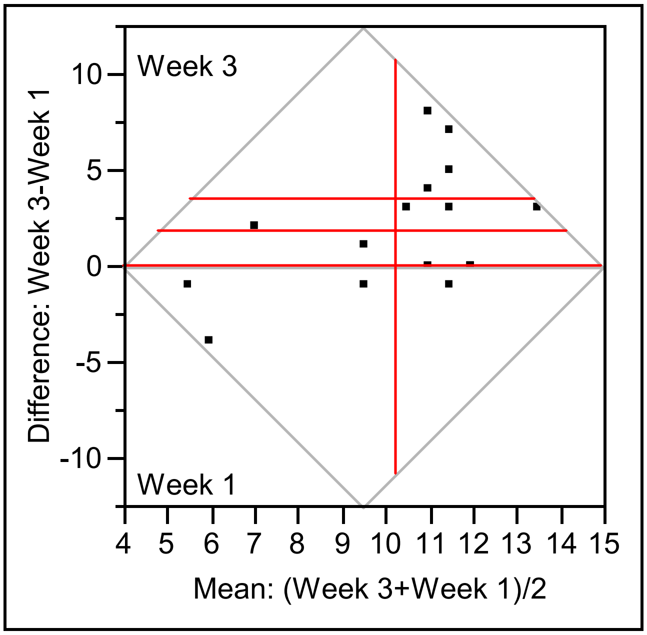
REPEATED MEASURES EXAMPLE USING JMP

The following are hypothetical scores on a role-taking ability test that has been administered at the end of each week in a 3-week training procedure for young children. The program was designed to improve children's ability to adopt the point of view of another person. We are interested in whether the program has any effect.

Week 1	Week 2	Week 3	Subj	Gender
9	9	14	1	M
9	13	13	2	M
12	14	15	3	M
9	13	12	4	M
6	8	8	5	M
10	7	9	6	M
8	13	15	7	M
9	12	10	8	M
6	4	5	9	F
11	9	11	10	F
10	10	13	11	F
12	10	12	12	F
12	10	11	13	F
12	12	12	14	F
7	9	15	15	F
8	6	4	16	F

If we wanted to compare week 1 with week 3, we could do a *t*-test, but the *t*-test that would be appropriate would be a dependent samples *t*-test. Why?

Matched Pairs
Difference: Week 3-Week 1



```
Week 3              11.1875       t-Ratio     2.274248
Week 1              9.375         DF          15
Mean Difference     1.8125        Prob > |t|  0.0381
Std Error           0.79697       Prob > t    0.0190
Upper95%            3.51119       Prob < t    0.9810
Lower95%            0.11381
N                   16
Correlation         0.3774
```

Note that it would be possible to do an ordinary one-way ANOVA on the 3 groups, but that would be inappropriate because we have repeated measures data, i.e., non-independent samples. Nonetheless, let's mindlessly do it anyway just to see what happens. We need to have the data in a different form (We can use the Tables|Stack Columns... command).

Subj	Gender	Week	RTA
1	M	Week 1	9
2	M	Week 1	9
3	M	Week 1	12
4	M	Week 1	9
5	M	Week 1	6
6	M	Week 1	10
7	M	Week 1	8
8	M	Week 1	9
9	F	Week 1	6
10	F	Week 1	11
11	F	Week 1	10
12	F	Week 1	12
13	F	Week 1	12
14	F	Week 1	12
15	F	Week 1	7
16	F	Week 1	8
1	M	Week 2	9
2	M	Week 2	13
3	M	Week 2	14
4	M	Week 2	13
5	M	Week 2	8
6	M	Week 2	7
7	M	Week 2	13
8	M	Week 2	12
9	F	Week 2	4
10	F	Week 2	9
11	F	Week 2	10
12	F	Week 2	10
13	F	Week 2	10
14	F	Week 2	12
15	F	Week 2	9
16	F	Week 2	6
1	M	Week 3	14
2	M	Week 3	13
3	M	Week 3	15
4	M	Week 3	12
5	M	Week 3	8
6	M	Week 3	9
7	M	Week 3	15
8	M	Week 3	10
9	F	Week 3	5
10	F	Week 3	11
11	F	Week 3	13
12	F	Week 3	12
13	F	Week 3	11
14	F	Week 3	12
15	F	Week 3	15
16	F	Week 3	4

Oneway Analysis of RTA By Week
Oneway Anova
Summary of Fit

```
Rsquare                         0.073119
Adj Rsquare                     0.031925
Root Mean Square Error          2.785378
Mean of Response                10.16667
Observations (or Sum Wgts)      48
```

Analysis of Variance

```
Source    DF    Sum of Squares    Mean Square    F Ratio    Prob > F
Week      2     27.54167          13.7708        1.7750     0.1811
Error     45    349.12500         7.7583
C. Total  47    376.66667
```

```
Means for Oneway Anova
Level     Number    Mean       Std Error   Lower 95%    Upper 95%
Week 1    16        9.3750     0.69634     7.9725       10.778
Week 2    16        9.9375     0.69634     8.5350       11.340
Week 3    16        11.1875    0.69634     9.7850       12.590
Std Error uses a pooled estimate of error variance
```

The above analysis is INAPPROPRIATE and INCORRECT. For an appropriate analysis we need to do a repeated measures ANOVA, which is a two-way Subjects × Weeks ANOVA. One way to do it is as follows. Note that we use Subjects as a factor in the design and make it a 'Random effect.'

```
Response RTA
Summary of Fit

RSquare                        1
RSquare Adj                    .
Root Mean Square Error         .
Mean of Response               10.16667
Observations (or Sum Wgts)     48

Analysis of Variance
Source     DF     Sum of Squares     Mean Square     F Ratio
Model      47     376.66667          8.01418         .
Error      0      0.00000            .               Prob > F
C. Total   47     376.66667                          .

Expected Mean Squares
The Mean Square per row by the Variance Component per column

EMS              Intercept      Week        Subj&Random     Subj*Week&Rando
                                                            m
Intercept        0              0           0               0
Week             0              16          0               1
Subj&Random      0              0           3               1
Subj*Week&Rando  0              0           0               1
m
  plus 1.0 times Residual Error Variance
Variance Component Estimates
Component              Var Comp Est     Percent of Total
Subj&Random            4.120833         53.115
Subj*Week&Random       3.6375           46.885
Total                  7.758333         100.000
 These estimates based on equating Mean Squares to Expected Value.
Test Denominator Synthesis
Source                 MS Den      DF Den      Denom MS Synthesis
Week                   3.6375      30          Subj*Week&Random
Subj&Random            3.6375      30          Subj*Week&Random
Subj*Week&Random       0           .
Tests wrt Random Effects
Source                 SS          MS Num      DF Num     F Ratio     Prob > F
Week                   27.5417     13.7708     2          3.7858      0.0342
Subj&Random            240         16          15         4.3986      0.0003
Subj*Week&Random       109.125     3.6375      30         .           .

Effect Details
Week
Effect Test
Sum of Squares       F Ratio       DF      Prob > F
27.541667            3.7858        2       0.0342
Denominator MS Synthesis:  Subj*Week&Random
Least Squares Means Table
Level     Least Sq Mean          Std Error         Mean
Week 1    9.375000               0.47680578        9.3750
Week 2    9.937500               0.47680578        9.9375
Week 3    11.187500              0.47680578        11.1875
```

Subj&Random
Effect Test

```
Sum of Squares          F Ratio         DF      Prob > F
240.00000               4.3986          15      0.0003
Denominator MS Synthesis:  Subj*Week&Random
```

Subj*Week&Random
Effect Test

```
Sum of Squares          F Ratio         DF      Prob > F
109.12500               .               30      .
Denominator MS Synthesis:
```

The above is a possible way to do the analysis, but the more common way to do the repeated measures ANOVA is to use the data in its original form with each column representing one level of the repeated measures factor, in this case, Week 1, Week 2, and Week 3. Notice that all the scores for each subject are in a single row. We use the Fit Model platform and make all 3 columns as Y variables with no between subject effects in the model. Choose MANOVA as the 'Personality.' When the model is run, the 'Response Specification' box appears. Here we want to check 'Univariate Tests Also' and then go to 'Click Here.' Select 'Repeated Measures' and name the repeated measures factor some sensible name. 'Time' is the default. I've changed that to 'Week' in what follows. The output is given below. The *F*-test for Week is in the 'Within Subjects' section in the subsection labeled 'Week', and the one we want is in the row labeled 'Univariate unadj.' From the output below we find $F(2, 30) = 3.79$, $p = .0342$, the same as from the analysis above.

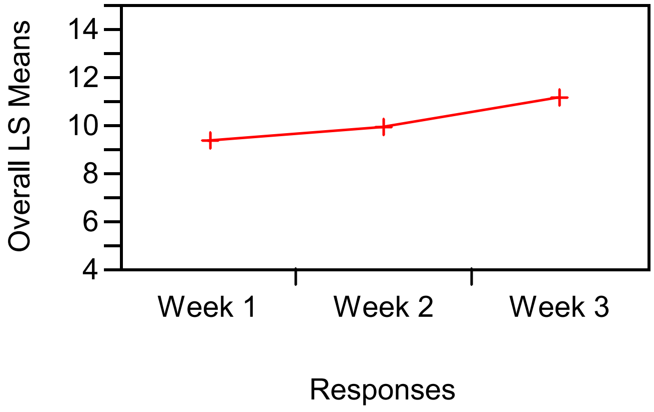
Manova Fit
Response Specification

```
To construct the linear combinations across responses,


N       16
DFE     15
```

Parameter Estimates

```
                    Week 1              Week 2              Week 3
Intercept           9.375               9.9375              11.1875
```

Least Squares Means
Overall Means



```
Overall             Week 2              Week 3
MeansWeek 1
9.375               9.9375              11.1875
```

```
Between Subjects
Sum
All Between
Test      Value            Exact F         NumDF      DenDF      Prob>F
F Test    0                .               .          .          .

Intercept
Test      Value            Exact F         NumDF      DenDF      Prob>F
F Test    20.672222        310.0833        1          15         <.0001

Within Subjects
Contrast
Sphericity Test

Mauchly Criterion      0.819025
ChiSquare              2.7949702
DF                     2
Prob >Chisq            0.2472179

All Within Interactions
Test                         Value            Approx. F       NumDF      DenDF      Prob>F
Wilks' Lambda                1                .               .          .          .
Pillai's Trace               0                .               .          .          .
Hotelling-Lawley             0                .               .          .          .
Roy's Max Root               0                .               .          .          .
Univar unadj Epsilon=        1                .               0          30         .
Univar G-G   Epsilon=        0.8467579        .               0          25.403     .
Univar H-F   Epsilon=        0.9430085        .               0          28.29      .

Week
Test                         Value            Exact F         NumDF      DenDF      Prob>F
F Test                       0.4335173        3.0346          2          14         0.0804
Univar unadj Epsilon=        1                3.7858          2          30         0.0342
Univar G-G   Epsilon=        0.8467579        3.7858          1.6935     25.403     0.0427
Univar H-F   Epsilon=        0.9430085        3.7858          1.886      28.29      0.0371

Column1
Column1
M Matrix
MWeek 1             Week 2              Week 3
-1                  1                   0

Whole Model
Test                         Value            Exact F         NumDF      DenDF      Prob>F
F Test                       0                .               .          .          .
Univar unadj Epsilon=        1                .               0          15         .
Univar G-G   Epsilon=        1                .               0          15         .
Univar H-F   Epsilon=        1                .               0          15         .

Intercept
Test                         Value            Exact F         NumDF      DenDF      Prob>F
F Test                       0.0496628        0.7449          1          15         0.4017
Univar unadj Epsilon=        1                0.7449          1          15         0.4017
Univar G-G   Epsilon=        1                0.7449          1          15         0.4017
Univar H-F   Epsilon=        1                0.7449          1          15         0.4017
```

Column2

```
Column2
```

M Matrix

```
MWeek 1             Week 2              Week 3
-1                  0                   1
```

Whole Model

```
Test                          Value          Exact F        NumDF     DenDF     Prob>F
F Test                        0              .              .         .         .
Univar unadj Epsilon=         1              .              0         15        .
Univar G-G   Epsilon=         1              .              0         15        .
Univar H-F   Epsilon=         1              .              0         15        .
```

Intercept

```
Test                          Value          Exact F        NumDF     DenDF     Prob>F
F Test                        0.3448134      5.1722         1         15        0.0381
Univar unadj Epsilon=         1              5.1722         1         15        0.0381
Univar G-G   Epsilon=         1              5.1722         1         15        0.0381
Univar H-F   Epsilon=         1              5.1722         1         15        0.0381
```

Now if we wanted to include the between subjects factor Gender, we would simply put that factor in the ‘Construct Model Effects’ box and repeat the analysis, leading to the following output:

Manova Fit

Response Specification

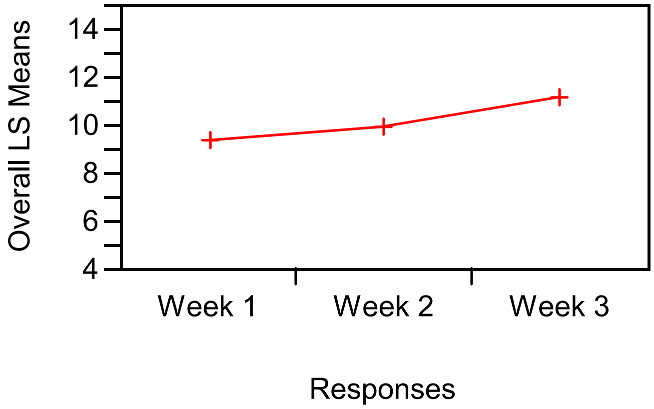
```
To construct the linear combinations across responses,


N       16
DFE     14
```

Parameter Estimates

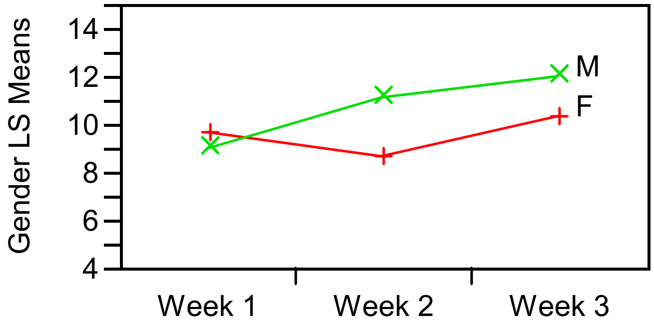
```
                    Week 1              Week 2              Week 3
Intercept           9.375               9.9375              11.1875
Gender[F]           0.375               -1.1875             -0.8125
```

Least Squares Means

Overall Means



```
Overall             Week 2              Week 3
MeansWeek 1
9.375               9.9375              11.1875
```



```
Responses
Gender

Gender              Week 1              Week 2              Week 3
F                   9.75                8.75                10.375
M                   9                   11.125              12

Between Subjects
Sum
All Between
Test      Value           Exact F          NumDF      DenDF      Prob>F
F Test    0.0623386       0.8727           1          14         0.3660

Intercept
Test      Value           Exact F          NumDF      DenDF      Prob>F
F Test    21.9609         307.4526         1          14         <.0001

Gender
Test      Value           Exact F          NumDF      DenDF      Prob>F
F Test    0.0623386       0.8727           1          14         0.3660

Within Subjects
Contrast
Sphericity Test

Mauchly Criterion       0.768212
ChiSquare               3.4279641
DF                      2
Prob >Chisq             0.180147

All Within Interactions
Test                        Value           Exact F         NumDF       DenDF       Prob>F
F Test                      0.6456817       4.1969          2           13          0.0392
Univar unadj Epsilon=       1               3.3937          2           28          0.0479
Univar G-G   Epsilon=       0.811828        3.3937          1.6237      22.731      0.0600
Univar H-F   Epsilon=       0.9687229       3.3937          1.9374      27.124      0.0497

Week
Test                        Value           Exact F         NumDF       DenDF       Prob>F
F Test                      0.4437082       2.8841          2           13          0.0919
Univar unadj Epsilon=       1               4.3899          2           28          0.0220
Univar G-G   Epsilon=       0.811828        4.3899          1.6237      22.731      0.0310
Univar H-F   Epsilon=       0.9687229       4.3899          1.9374      27.124      0.0232

Week*Gender
Test                        Value           Exact F         NumDF       DenDF       Prob>F
F Test                      0.6456817       4.1969          2           13          0.0392
Univar unadj Epsilon=       1               3.3937          2           28          0.0479
Univar G-G   Epsilon=       0.811828        3.3937          1.6237      22.731      0.0600
Univar H-F   Epsilon=       0.9687229       3.3937          1.9374      27.124      0.0497
```

Column1

Column1

M Matrix

MWeek 1	Week 2	Week 3
-1	1	0

Whole Model

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.6212724	8.6978	1	14	0.0106
Univar unadj Epsilon=	1	8.6978	1	14	0.0106
Univar G-G Epsilon=	1	8.6978	1	14	0.0106
Univar H-F Epsilon=	1	8.6978	1	14	0.0106

Intercept (Note: This is the test for the $\text{Week}_{\text{contrast}}$)

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.0805169	1.1272	1	14	0.3063
Univar unadj Epsilon=	1	1.1272	1	14	0.3063
Univar G-G Epsilon=	1	1.1272	1	14	0.3063
Univar H-F Epsilon=	1	1.1272	1	14	0.3063

Gender (Note: This is the test for the $\text{Week}_{\text{contrast}}$ × Gender interaction)

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.6212724	8.6978	1	14	0.0106
Univar unadj Epsilon=	1	8.6978	1	14	0.0106
Univar G-G Epsilon=	1	8.6978	1	14	0.0106
Univar H-F Epsilon=	1	8.6978	1	14	0.0106

Column2

Column2

M Matrix

MWeek 1	Week 2	Week 3
-1	0	1

Whole Model

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.1737247	2.4321	1	14	0.1412
Univar unadj Epsilon=	1	2.4321	1	14	0.1412
Univar G-G Epsilon=	1	2.4321	1	14	0.1412
Univar H-F Epsilon=	1	2.4321	1	14	0.1412

Intercept

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.4047161	5.6660	1	14	0.0321
Univar unadj Epsilon=	1	5.6660	1	14	0.0321
Univar G-G Epsilon=	1	5.6660	1	14	0.0321
Univar H-F Epsilon=	1	5.6660	1	14	0.0321

Gender

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.1737247	2.4321	1	14	0.1412
Univar unadj Epsilon=	1	2.4321	1	14	0.1412
Univar G-G Epsilon=	1	2.4321	1	14	0.1412
Univar H-F Epsilon=	1	2.4321	1	14	0.1412

Rules of Thumb for Writing the ANOVA Table

These rules of thumb are taken (with minor modifications) from unpublished notes by Harry F. Gollob. Similar sets of rules may be found in most standard texts—e.g., Myers, p. 205-206; Winer, p. 371-375; Searle, p. 389-394.

I. Generate lines of the ANOVA table *(i.e., the sources of variation).*

a) Find the "simple sources," i.e., write the simplest term for each letter.
1. If a factor is nested it must ***always*** be written with the factor(s) in which it is nested; e.g., if **S** is nested in an **A×B** matrix, the "simple source" involving **S** is **S(AB)**. **S** should not be written alone.
2. Factors that are not nested are written alone. In the **S(AB)** design **A** and **B** are not nested so they are "simple sources."
3. There is a "simple source" for each factor in the design.

Example. For the **S(G)C** design the simple sources are **G, S(G)**, and **C**.

b) Let k = the number of factors in the design. Now consider the list of all possible combinations of simple sources taking $1,2,\ldots,k$ simple sources at a time. Each combination which ***does not use any letter more than once*** identifies a source of variation in the ANOVA table. When the usual assumptions are met, every source identified in this way is statistically independent (when cell sizes are equal) of every other source.

Example. For the **S(G)C** design we obtain the sources **G, S(G), C, CG,** and **S(G)C.** The two combinations that *do not* identify lines of the ANOVA table are **GS(G)** and **GS(G)C** because **G** is repeated in each of these combinations.

II. Generate df for the ANOVA table.

a) Let the number of levels of any factor in the design be designated by the lower case letter corresponding to that factor; e.g., the number of levels of **S** is s. If the factor is nested the number of levels refers to the number of levels *within each level* or combination of levels of the factors in which it is nested.

Example. In the **S(AB)** design s refers to the number of subjects within each **AB** cell, i.e., the number per cell. In the **S(A(B))** design, a refers to the number of levels of **A** within each level of **B**.

b) Think of the letters used to identify the sources of variance as defining a product for each source. For letters *not* in parentheses substitute the corresponding lower case letter minus 1, and for letters in parentheses substitute simply the lower case letter. The resulting product gives the *df* for the source in question.

Example. For **S(G)C**:

Source	*df*
G	$g - 1$
S(G)	$g(s - 1)$
C	$c - 1$
CG	$(c - 1)(g - 1)$
S(G)C	$g(s - 1)(c - 1)$

III. Generating computing formulas in <u>balanced</u> designs.

[NB. All of the following sections apply only to <u>balanced</u> designs.]

Just apply Keppel's rules to the *df* expressions.

IV. Generating E(MS).

a) Identify variance components which contribute to the MS for each source. (In the following, an *"extra"* letter refers to a letter not used in writing the source; e.g., in an **S(A)** design, **S** would be an extra letter with respect to the **A** source of variance.)
1. The source of interest is always included.
2. An additional component is added if
i) it includes all letters that are used in the source being considered ***and if***
ii) all *extra* letters ***either*** represent ***random*** factors ***or*** are in parentheses or both.
3. Many authors refer to the source that uses all the letters in the design as σ_e^2. However, σ_e^2 is also sometimes used to refer to measurement error (as opposed to sampling error) even in cases where no estimate of measurement error is available in the data at hand (e.g., in an **S×A** repeated measures design with **S** random and **A** fixed).

b) Determine the coefficients for the variance components. Each variance component identified in the previous step is multiplied by the product of the number of levels of each factor whose letter is not used in identifying the component.

Example. **S(G)C** with **G** fixed and **S** random:

Source	***df***	***E(MS)***
G	$g - 1$	$sc\,\theta^2_{\text{G}} + c\,\sigma^2_{\text{S(G)}}$
S(G)	$g(s - 1)$	$c\,\sigma^2_{\text{S(G)}}$
C	$c - 1$	$sg\,\theta^2_{\text{C}} + \sigma^2_{\text{S(G)C}}$
CG	$(c - 1)(g - 1)$	$s\,\theta^2_{\text{CG}} + \sigma^2_{\text{S(G)C}}$
S(G)C	$g(s - 1)(c - 1)$	$\sigma^2_{\text{S(G)C}}$

θ^2 is used to indicate that a variance component involves ***only*** fixed factors.

V. Choose the best error term for each source of interest.

To test the H_o: $\theta^2_Z = 0$ or $\sigma^2_Z = 0$ where Z refers to any source of variance in the model, find an appropriate error term and compute the ratio, $F = MS_Z/MS_{\text{error}}$. An appropriate error term is one whose *E*(*MS*) equals the *E*(*MS*) for Z when the null hypothesis, H_o, is true.

When an appropriate error term is not available from any single source in the table, an approximate test of significance can often be obtained by use of a ***Quasi-F*** ratio or perhaps some judicious pooling. Myers, J.L. (1979) *Fundamentals of experimental design, 3rd ed.,* p. 188-194, discusses these possibilities for the **S×A×B** design.

VI. Estimating variance components.

Generate the table of *E*(*MS*), substitute the obtained *MS* for the *E*(*MS*) in each line of the table, and use the resulting set of equations to compute estimates of the variance components. If a negative estimate is obtained it is usually advisable to set the estimate equal to zero (but see Searle, 1971, p. 406-408, for other alternatives).

Example. **S(G)C**. From the table of E(MS) obtained above (IVb) we make the substitutions and obtain:

$$MS_G = cs\hat{\theta}_G^2 + c\hat{\sigma}_{S(G)}^2 \qquad (1)$$

$$MS_{S(G)} = c\hat{\sigma}_{S(G)}^2 \qquad (2)$$

$$MS_C = gs\hat{\theta}_C^2 + \hat{\sigma}_{S(G)C}^2 \qquad (3)$$

$$MS_{CG} = s\hat{\theta}_{CG}^2 + \hat{\sigma}_{S(G)C}^2 \qquad (4)$$

$$MS_{S(G)C} = \hat{\sigma}_{S(G)C}^2 \qquad (5)$$

From (1) and (2) we obtain

$$MS_G = cs\hat{\theta}_G^2 + MS_{S(G)}$$

so that

$$\hat{\theta}_G^2 = \frac{MS_G - MS_{S(G)}}{cs}$$

From (2) we see that $\hat{\sigma}_{S(G)}^2 = \dfrac{MS_{S(G)}}{c}$.

Working in a similar manner with Equations (3), (4), and (5) we get:

$$\hat{\theta}_C^2 = \frac{MS_C - MS_{S(G)C}}{gs}$$

$$\hat{\theta}_{CG}^2 = \frac{MS_{CG} - MS_{S(G)C}}{s}.$$

VII. Estimating ω^2 (omega-squared) for any source.

a) The proportion of total variation in the population that is accounted for by a source of variation is referred to as omega-squared: ω^2. To estimate omega-squared we first obtain estimates of the variance of the parameters associated with each source of variance. The variance of the parameters of any effect can be easily estimated by multiplying the corresponding ***variance component*** estimate by a 'multiplier.' In the string of letters which identify each source, let A, B, ... denote ***fixed factors that are not in parentheses***. Then the multiplier for a source is given by

$$\frac{(1)(a-1)(b-1)\ldots}{(1)\,(a)\,(b)\ldots}.$$

Note that if no factors in the source meet the condition (fixed factors not in parentheses), then the multiplier is one.

b) Having obtained estimates of the variances for each source, it simple to compute an estimate of ω^2:

Let $\hat{\sigma}_i^2$ denote the estimated variance for the *i*th source of variance. Then

$$\hat{\sigma}_Y^2 = \sum \hat{\sigma}_i^2 \quad \text{and} \quad \hat{\omega}_i^2 = \frac{\hat{\sigma}_i^2}{\hat{\sigma}_Y^2}.$$

Example. **S(G)C; G** and **C** fixed; **S** random.

Source	Estimated Variance	=	(Multiplier)(Est. Variance Component)
G	$\hat{\sigma}_G^2$	=	$\frac{(g-1)}{g}\hat{\theta}_G^2$
S(G)	$\hat{\sigma}_{S(G)}^2$	=	$(1)\hat{\sigma}_{S(G)}^2$
C	$\hat{\sigma}_C^2$	=	$\frac{(c-1)}{c}\hat{\theta}_C^2$
CG	$\hat{\sigma}_{CG}^2$	=	$\frac{(c-1)(g-1)}{cg}\hat{\theta}_{CG}^2$
S(G)C	$\hat{\sigma}_{S(G)C}^2$	=	$\frac{(c-1)}{c}\hat{\theta}_{S(G)C}^2$

$$\hat{\omega}_G^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \hat{\sigma}_{S(G)}^2 + \hat{\sigma}_C^2 + \hat{\sigma}_{CG}^2 + \hat{\sigma}_{S(G)C}^2}, \text{ etc.}$$

VIII. Power of statistical tests.

The procedure used in determining the power of a test depends on whether the source of variance being tested is a fixed or random effect. A source is a fixed effect if its name is one in which *all* letters outside parentheses represent fixed factors. In all other cases the source is random.

a) Computing the power of tests of fixed effects.

First, compute the noncentrality parameter, ϕ, by using the equation

$$\phi^2 = \left[\frac{df_{num}}{1+df_{num}}\right]\left[\frac{E(MS_{num}) - E(MS_{den})}{E(MS_{den})}\right].$$

After computing ϕ, use the commonly available Pearson & Hartley power charts to find the power of the test.

Example. **S(G)C** with **S** random and other factors fixed.
For this case the noncentrality parameter for computing the power of the test of the **C** main effect is the square root of

$$\phi_C^2 = \left[\frac{c-1}{c}\right]\left[\frac{sg\theta_C^2}{\sigma_{S(G)C}^2}\right].$$

b) Computing the power of tests of random effects.

Lindman (1974, *Analysis of variance in complex experimental designs.* San Francisco: Freeman, pp. 118, 135, & 185) tells how to do power calculations for random effects.

Additional Examples Using JMP

Example 1. Below is the JMP analysis for Keppel & Wickens' fully nested design example in Chapter 25 (p. 552). They call it a 'Completely between-subjects design.' The design would be *S(B(A))*.

Y	A	B
14	a1	b11
15	a1	b11
12	a1	b11
13	a1	b11
13	a1	b11
13	a1	b12
12	a1	b12
15	a1	b12
16	a1	b12
12	a1	b12
15	a1	b13
17	a1	b13
15	a1	b13
16	a1	b13
15	a1	b13
7	a2	b21
8	a2	b21
7	a2	b21
7	a2	b21
8	a2	b21
8	a2	b22
6	a2	b22
11	a2	b22
11	a2	b22
11	a2	b22
13	a2	b23
10	a2	b23
11	a2	b23
11	a2	b23
10	a2	b23

Model Specification

A
B[A]& Random

Response Y
Summary of Fit

RSquare	0.821343
RSquare Adj	0.784123
Root Mean Square Error	1.443376
Mean of Response	11.73333
Observations (or Sum Wgts)	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	229.86667	45.9733	22.0672
Error	24	50.00000	2.0833	Prob > F
C. Total	29	279.86667		<.0001

Test Denominator Synthesis

Source	MS Den	DF Den	Denom MS Synthesis
A	11.8333	4	B[A]&Random
B[A]&Random	2.08333	24	Residual

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob > F
A	182.533	182.533	1	15.4254	0.0171
B[A]&Random	47.3333	11.8333	4	5.6800	0.0023

```
Effect Details
A
Effect Test
Sum of Squares           F Ratio        DF     Prob > F
182.53333                15.4254        1      0.0171
Denominator MS Synthesis:
 B[A]&Random
Least Squares Means Table
Level    Least Sq Mean               Std Error            Mean
a1       14.200000                   0.88819417           14.2000
a2       9.266667                    0.88819417           9.2667
B[A]&Random
Effect Test
Sum of Squares           F Ratio        DF     Prob > F
47.333333                5.6800         4      0.0023
Denominator MS Synthesis:
 Residual
Least Squares Means Table
Level       Least Sq Mean               Std Error
[a1]b11     13.400000                   0.64549722
[a1]b12     13.600000                   0.64549722
[a1]b13     15.600000                   0.64549722
[a2]b21     7.400000                    0.64549722
[a2]b22     9.400000                    0.64549722
[a2]b23     11.000000                   0.64549722
```

Example 2. This the JMP analysis of the $S \times A \times B$ extended example in Keppel & Wickens' chapter 18. The highlighted items refer to the tests carried out in the chapter. You will note that Keppel's numbers contain some round-off errors.

The analysis below was carried out using the data layout as given with the 'Compound' Response option of the Manova analysis from the 'Fit Model' platform. The final B_{linear} analysis was carried out using the 'Contrast' Response option.

Subject	a1b1	a1b2	a1b3	a1b4	a2b1	a2b2	a2b3	a2b4	a3b1	a3b2	a3b3	a3b4
1	1	1	2	2	1	2	2	3	1	2	2	3
2	2	2	3	4	2	3	4	5	3	3	3	5
3	2	3	3	3	3	4	3	4	2	3	3	4
4	0	1	0	2	1	2	2	3	2	2	2	3
5	1	1	2	2	1	1	2	3	1	1	2	2
6	1	2	3	3	2	3	3	5	2	3	4	4
7	1	1	2	3	2	1	2	3	1	1	2	3
8	1	1	2	3	1	2	2	3	2	2	3	4

Manova Fit

Response Specification

To construct the linear combinations across responses,

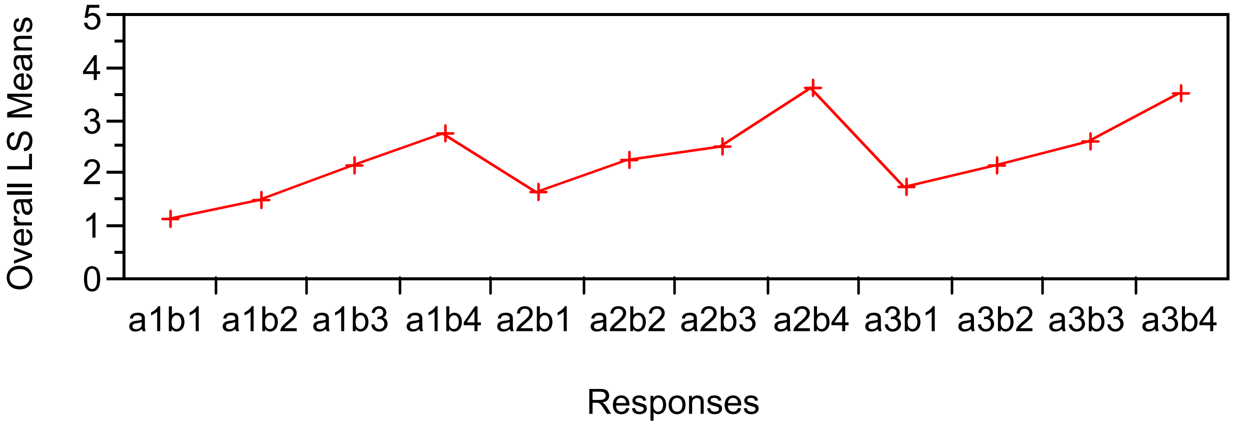
N	8
DFE	7

Parameter Estimates

	a1b1	a1b2	a1b3	a1b4	a2b1	a2b2	a2b3	a2b4	a3b1	a3b2	a3b3	a3b4
Intercept	1.125	1.5	2.125	2.75	1.625	2.25	2.5	3.625	1.75	2.125	2.625	3.5

Least Squares Means

Overall Means



```
Overall Meansa1b1  a1b2  a1b3   a1b4  a2b1   a2b2  a2b3  a2b4   a3b1  a3b2   a3b3   a3b4
1.125              1.5   2.125  2.75  1.625  2.25  2.5   3.625  1.75  2.125  2.625  3.5
```

Compound

B

Sphericity Test

Mauchly Criterion	0.6958111
ChiSquare	2.0753189
DF	5
Prob >Chisq	0.8386282

```
Whole Model
Test                         Value         Approx. F     NumDF       DenDF       Prob>F
Wilks' Lambda                1             .             .           .           .
Pillai's Trace               0             .             .           .           .
Hotelling-Lawley             0             .             .           .           .
Roy's Max Root               0             .             .           .           .
Univar unadj Epsilon=        1             .             0           21          .
Univar G-G   Epsilon=        0.7901        .             0           16.592      .
Univar H-F   Epsilon=        1             .             0           21          .

Intercept
Test                         Value         Exact F       NumDF       DenDF       Prob>F
F Test                       23.72029      39.5338       3           5           0.0007
Univar unadj Epsilon=        1             44.7468       3           21          <.0001
Univar G-G   Epsilon=        0.7901        44.7468       2.3703      16.592      <.0001
Univar H-F   Epsilon=        1             44.7468       3           21          <.0001
Compound
A

Sphericity Test

Mauchly Criterion       0.75
ChiSquare               1.7260924
DF                      2
Prob >Chisq             0.421875

Whole Model
Test                         Value         Approx. F     NumDF       DenDF       Prob>F
Wilks' Lambda                1             .             .           .           .
Pillai's Trace               0             .             .           .           .
Hotelling-Lawley             0             .             .           .           .
Roy's Max Root               0             .             .           .           .
Univar unadj Epsilon=        1             .             0           14          .
Univar G-G   Epsilon=        0.8           .             0           11.2        .
Univar H-F   Epsilon=        1             .             0           14          .

Intercept
Test                         Value         Exact F       NumDF       DenDF       Prob>F
F Test                       3.125         9.3750        2           6           0.0142
Univar unadj Epsilon=        1             10.9375       2           14          0.0014
Univar G-G   Epsilon=        0.8           10.9375       1.6         11.2        0.0033
Univar H-F   Epsilon=        1             10.9375       2           14          0.0014
Compound
B*A

Sphericity Test

Mauchly Criterion       0.0293562
ChiSquare               16.857197
DF                      20
Prob >Chisq             0.662226

Whole Model
Test                         Value         Approx. F     NumDF       DenDF       Prob>F
Wilks' Lambda                1             .             .           .           .
Pillai's Trace               0             .             .           .           .
Hotelling-Lawley             0             .             .           .           .
Roy's Max Root               0             .             .           .           .
Univar unadj Epsilon=        1             .             0           42          .
Univar G-G   Epsilon=        0.591002      .             0           24.822      .
Univar H-F   Epsilon=        1             .             0           42          .

Intercept
Test                         Value         Exact F       NumDF       DenDF       Prob>F
F Test                       1.2222222     0.4074        6           2           0.8336
Univar unadj Epsilon=        1             0.8235        6           42          0.5582
Univar G-G   Epsilon=        0.591002      0.8235        3.546       24.822      0.5105
Univar H-F   Epsilon=        1             0.8235        6           42          0.5582
```

Contrast

M Matrix

Ma1b1	a1b2	a1b3	a1b4	a2b1	a2b2	a2b3	a2b4	a3b1	a3b2	a3b3	a3b4
-3	-1	1	3	-3	-1	1	3	-3	-1	1	3

Whole Model

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0	.	.	.	.
Univar unadj Epsilon=	1	.	0	7	.
Univar G-G Epsilon=	1	.	0	7	.
Univar H-F Epsilon=	1	.	0	7	.

Intercept

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	18.846154	131.9231	1	7	<.0001
Univar unadj Epsilon=	1	131.9231	1	7	<.0001
Univar G-G Epsilon=	1	131.9231	1	7	<.0001
Univar H-F Epsilon=	1	131.9231	1	7	<.0001

Example 3. The following are several analyses of a data set which has been expanded (I added three wolves) from the 'Animals' example from the JMP manual. Nine animals from three species were tracked, and the diameter of the area that each animal wandered was recorded. Each animal was measured four times, once per season. The design is thus *Subject (Species) Season*.

species	subject	fall	spring	summer	winter
FOX	1	0	5	3	0
FOX	2	3	5	4	1
FOX	3	4	6	2	3
COYOTE	1	4	7	8	2
COYOTE	2	5	6	6	4
COYOTE	3	7	8	9	5
WOLF	1	3	7	10	2
WOLF	2	5	6	8	3
WOLF	3	4	7	9	2

Analysis of Animals Expanded data set using the MANOVA layout and analysis type. In the Response Specification menu the Repeated Measures option was used.

Manova Fit

Response Specification

To construct the linear combinations across responses,

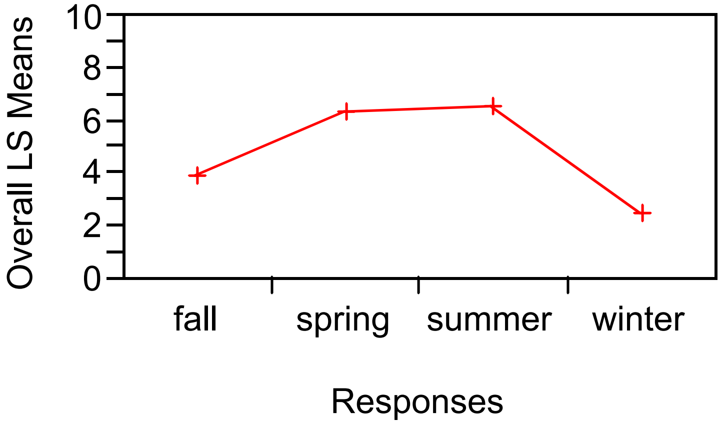
N	9
DFE	6

Parameter Estimates

	fall	spring	summer	winter
Intercept	3.88888889	6.33333333	6.55555556	2.44444444
species[COYOTE]	1.44444444	0.66666667	1.11111111	1.22222222
species[FOX]	-1.5555556	-1	-3.5555556	-1.1111111

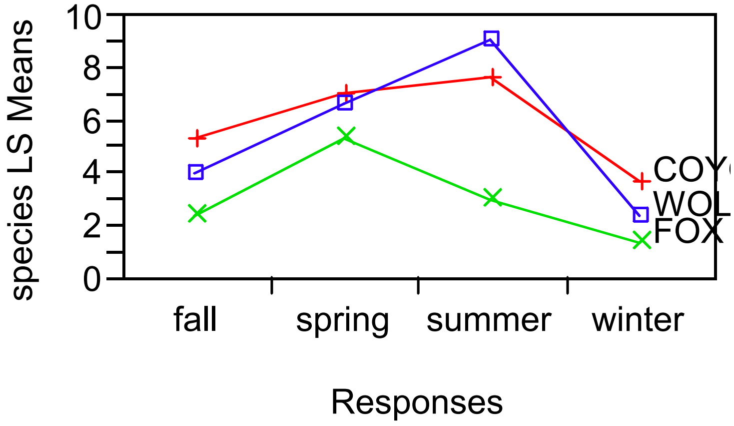
Least Squares Means

Overall Means



Overall Meansfall	spring	summer	winter
3.88888889	6.33333333	6.55555556	2.44444444

species



species	fall	spring	summer	winter
COYOTE	5.33333333	7	7.66666667	3.66666667
FOX	2.33333333	5.33333333	3	1.33333333
WOLF	4	6.66666667	9	2.33333333

Between Subjects

Sum

All Between

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	3.4789644	10.4369	2	6	0.0111

Intercept

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	48.428803	290.5728	1	6	<.0001

species

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	3.4789644	10.4369	2	6	0.0111

Contrast

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	3.4182848	20.5097	1	6	0.0040

Contrast

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.0606796	0.3641	1	6	0.5683

Within Subjects

Contrast

Sphericity Test

Mauchly Criterion	0.099309
ChiSquare	10.906061
DF	5
Prob >Chisq	0.0532747

All Within Interactions

Test	Value	Approx. F	NumDF	DenDF	Prob>F
Wilks' Lambda	0.096602	2.9566	6	8	0.0794
Pillai's Trace	1.0380715	1.7986	6	10	0.1965
Hotelling-Lawley	7.9576427	3.9788	6	6	0.0586
Roy's Max Root	7.7784151	12.9640	3	5	0.0086
Univar unadj Epsilon=	1	3.9091	6	18	0.0112
Univar G-G Epsilon=	0.4564045	3.9091	2.7384	8.2153	0.0556
Univar H-F Epsilon=	0.7430647	3.9091	4.4584	13.375	0.0235

Season

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	30.525783	40.7010	3	4	0.0019
Univar unadj Epsilon=	1	31.6281	3	18	<.0001
Univar G-G Epsilon=	0.4564045	31.6281	1.3692	8.2153	0.0003
Univar H-F Epsilon=	0.7430647	31.6281	2.2292	13.375	<.0001

Season*species

Test	Value	Approx. F	NumDF	DenDF	Prob>F
Wilks' Lambda	0.096602	2.9566	6	8	0.0794
Pillai's Trace	1.0380715	1.7986	6	10	0.1965
Hotelling-Lawley	7.9576427	3.9788	6	6	0.0586
Roy's Max Root	7.7784151	12.9640	3	5	0.0086
Univar unadj Epsilon=	1	3.9091	6	18	0.0112
Univar G-G Epsilon=	0.4564045	3.9091	2.7384	8.2153	0.0556
Univar H-F Epsilon=	0.7430647	3.9091	4.4584	13.375	0.0235

Analysis of Animals Expanded data set using the Manova layout and analysis type. In the Response Specification menu the Sum and Contrast options were used. In the Contrast dialog the contrasts shown in the M-Matrices below were specified. The *F*-tests in this analysis for both between and within

Manova Fit

Response Specification

To construct the linear combinations across responses,

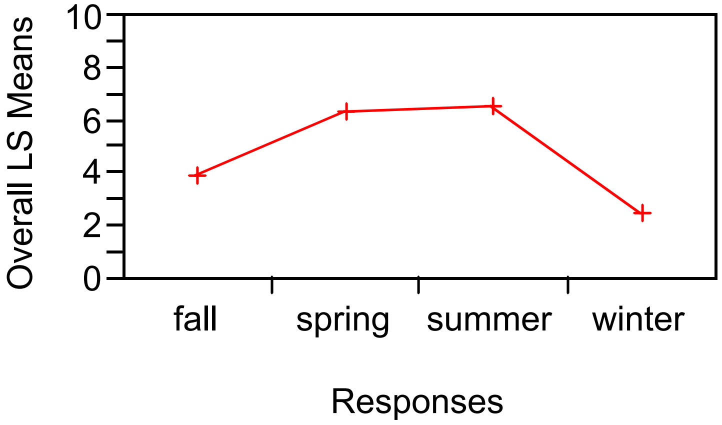
N	9
DFE	6

Parameter Estimates

	fall	spring	summer	winter
Intercept	3.88888889	6.33333333	6.55555556	2.44444444
species[COYOTE]	1.44444444	0.66666667	1.11111111	1.22222222
species[FOX]	-1.5555556	-1	-3.5555556	-1.1111111

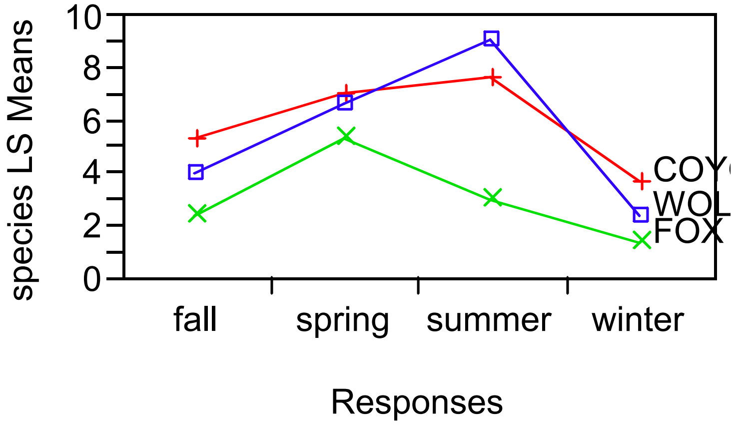
Least Squares Means

Overall Means



Overall Meansfall	spring	summer	winter
3.88888889	6.33333333	6.55555556	2.44444444

species



species	fall	spring	summer	winter
COYOTE	5.33333333	7	7.66666667	3.66666667
FOX	2.33333333	5.33333333	3	1.33333333
WOLF	4	6.66666667	9	2.33333333

Sum

Whole Model

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	3.4789644	10.4369	2	6	0.0111

Intercept

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	48.428803	290.5728	1	6	<.0001

species

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	3.4789644	10.4369	2	6	0.0111

Contrast

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	3.4182848	20.5097	1	6	0.0040

Contrast

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.0606796	0.3641	1	6	0.5683

```
Contrast


Whole Model
Test                Value          Approx. F      NumDF       DenDF       Prob>F
Wilks' Lambda       0.0575059      4.2268         6           8           0.0325
Pillai's Trace      1.2093054      2.5490         6           10          0.0918
Hotelling-Lawley    11.749808      5.8749         6           6           0.0245
Roy's Max Root      11.340686      18.9011        3           5           0.0037

Intercept
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     75.463818      100.6184       3           4           0.0003

species
Test                Value          Approx. F      NumDF       DenDF       Prob>F
Wilks' Lambda       0.0575059      4.2268         6           8           0.0325
Pillai's Trace      1.2093054      2.5490         6           10          0.0918
Hotelling-Lawley    11.749808      5.8749         6           6           0.0245
Roy's Max Root      11.340686      18.9011        3           5           0.0037
Column1
Column1



Whole Model
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     0.8666667      2.6000         2           6           0.1537

Intercept
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     12.675439      76.0526        1           6           0.0001

species
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     0.8666667      2.6000         2           6           0.1537
Column2
Column2

Whole Model
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     2.463964       7.3919         2           6           0.0241

Intercept
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     1.5225225      9.1351         1           6           0.0233

species
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     2.463964       7.3919         2           6           0.0241
Column3
Column3

Whole Model
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     1.3981481      4.1944         2           6           0.0725

Intercept
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     39.185185      235.1111       1           6           <.0001


species
Test       Value          Exact F        NumDF       DenDF       Prob>F
F Test     1.3981481      4.1944         2           6           0.0725
```

Contrast

Sphericity Test

```
Mauchly Criterion          0.099309
ChiSquare                  10.906061
DF                         5
Prob >Chisq                0.0532747
```

Whole Model

```
Test                       Value          Approx. F      NumDF      DenDF      Prob>F
Wilks' Lambda              0.096602       2.9566         6          8          0.0794
Pillai's Trace             1.0380715      1.7986         6          10         0.1965
Hotelling-Lawley           7.9576427      3.9788         6          6          0.0586
Roy's Max Root             7.7784151      12.9640        3          5          0.0086
Univar unadj Epsilon=      1              3.9091         6          18         0.0112
Univar G-G   Epsilon=      0.4564045      3.9091         2.7384     8.2153     0.0556
Univar H-F   Epsilon=      0.7430647      3.9091         4.4584     13.375     0.0235
```

Intercept

```
Test                       Value          Exact F        NumDF      DenDF      Prob>F
F Test                     30.525783      40.7010        3          4          0.0019
Univar unadj Epsilon=      1              31.6281        3          18         <.0001
Univar G-G   Epsilon=      0.4564045      31.6281        1.3692     8.2153     0.0003
Univar H-F   Epsilon=      0.7430647      31.6281        2.2292     13.375     <.0001
```

species

```
Test                       Value          Approx. F      NumDF      DenDF      Prob>F
Wilks' Lambda              0.096602       2.9566         6          8          0.0794
Pillai's Trace             1.0380715      1.7986         6          10         0.1965
Hotelling-Lawley           7.9576427      3.9788         6          6          0.0586
Roy's Max Root             7.7784151      12.9640        3          5          0.0086
Univar unadj Epsilon=      1              3.9091         6          18         0.0112
Univar G-G   Epsilon=      0.4564045      3.9091         2.7384     8.2153     0.0556
Univar H-F   Epsilon=      0.7430647      3.9091         4.4584     13.375     0.0235
```

Column1

```
Column1
```

M Matrix

```
Mfall            spring           summer           winter
-1               -1               -1               3
```

Whole Model

```
Test                       Value          Exact F        NumDF      DenDF      Prob>F
F Test                     0.8666667      2.6000         2          6          0.1537
Univar unadj Epsilon=      1              2.6000         2          6          0.1537
Univar G-G   Epsilon=      1              2.6000         2          6          0.1537
Univar H-F   Epsilon=      1              2.6000         2          6          0.1537
```

Intercept

```
Test                       Value          Exact F        NumDF      DenDF      Prob>F
F Test                     12.675439      76.0526        1          6          0.0001
Univar unadj Epsilon=      1              76.0526        1          6          0.0001
Univar G-G   Epsilon=      1              76.0526        1          6          0.0001
Univar H-F   Epsilon=      1              76.0526        1          6          0.0001
```

species

```
Test                       Value          Exact F        NumDF      DenDF      Prob>F
F Test                     0.8666667      2.6000         2          6          0.1537
Univar unadj Epsilon=      1              2.6000         2          6          0.1537
Univar G-G   Epsilon=      1              2.6000         2          6          0.1537
Univar H-F   Epsilon=      1              2.6000         2          6          0.1537
```

```
Contrast
Test                      Value          Exact F        NumDF       DenDF       Prob>F
F Test                    0.5482456      3.2895         1           6           0.1197
Univar unadj Epsilon=     1              3.2895         1           6           0.1197
Univar G-G   Epsilon=     1              3.2895         1           6           0.1197
Univar H-F   Epsilon=     1              3.2895         1           6           0.1197

Contrast
Test                      Value          Exact F        NumDF       DenDF       Prob>F
F Test                    0.3184211      1.9105         1           6           0.2162
Univar unadj Epsilon=     1              1.9105         1           6           0.2162
Univar G-G   Epsilon=     1              1.9105         1           6           0.2162
Univar H-F   Epsilon=     1              1.9105         1           6           0.2162
Column2
Column2

M Matrix
Mfall            spring           summer           winter
-1               -1               2                0


Whole Model
Test                      Value          Exact F        NumDF       DenDF       Prob>F
F Test                    2.463964       7.3919         2           6           0.0241
Univar unadj Epsilon=     1              7.3919         2           6           0.0241
Univar G-G   Epsilon=     1              7.3919         2           6           0.0241
Univar H-F   Epsilon=     1              7.3919         2           6           0.0241

Intercept
Test                      Value          Exact F        NumDF       DenDF       Prob>F
F Test                    1.5225225      9.1351         1           6           0.0233
Univar unadj Epsilon=     1              9.1351         1           6           0.0233
Univar G-G   Epsilon=     1              9.1351         1           6           0.0233
Univar H-F   Epsilon=     1              9.1351         1           6           0.0233

species
Test                      Value          Exact F        NumDF       DenDF       Prob>F
F Test                    2.463964       7.3919         2           6           0.0241
Univar unadj Epsilon=     1              7.3919         2           6           0.0241
Univar G-G   Epsilon=     1              7.3919         2           6           0.0241
Univar H-F   Epsilon=     1              7.3919         2           6           0.0241

Contrast
Test                      Value          Exact F        NumDF       DenDF       Prob>F
F Test                    1.893018       11.3581        1           6           0.0150
Univar unadj Epsilon=     1              11.3581        1           6           0.0150
Univar G-G   Epsilon=     1              11.3581        1           6           0.0150
Univar H-F   Epsilon=     1              11.3581        1           6           0.0150


Contrast
Test                      Value          Exact F        NumDF       DenDF       Prob>F
F Test                    0.5709459      3.4257         1           6           0.1137
Univar unadj Epsilon=     1              3.4257         1           6           0.1137
Univar G-G   Epsilon=     1              3.4257         1           6           0.1137
Univar H-F   Epsilon=     1              3.4257         1           6           0.1137
```

Column3

Column3

M Matrix

Mfall	spring	summer	winter
-1	1	0	0

Whole Model

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.2166667	0.6500	2	6	0.5552
Univar unadj Epsilon=	1	0.6500	2	6	0.5552
Univar G-G Epsilon=	1	0.6500	2	6	0.5552
Univar H-F Epsilon=	1	0.6500	2	6	0.5552

Intercept

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	4.0333333	24.2000	1	6	0.0027
Univar unadj Epsilon=	1	24.2000	1	6	0.0027
Univar G-G Epsilon=	1	24.2000	1	6	0.0027
Univar H-F Epsilon=	1	24.2000	1	6	0.0027

species

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.2166667	0.6500	2	6	0.5552
Univar unadj Epsilon=	1	0.6500	2	6	0.5552
Univar G-G Epsilon=	1	0.6500	2	6	0.5552
Univar H-F Epsilon=	1	0.6500	2	6	0.5552

Contrast

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.1041667	0.6250	1	6	0.4593
Univar unadj Epsilon=	1	0.6250	1	6	0.4593
Univar G-G Epsilon=	1	0.6250	1	6	0.4593
Univar H-F Epsilon=	1	0.6250	1	6	0.4593

Contrast

Test	Value	Exact F	NumDF	DenDF	Prob>F
F Test	0.1125	0.6750	1	6	0.4427
Univar unadj Epsilon=	1	0.6750	1	6	0.4427
Univar G-G Epsilon=	1	0.6750	1	6	0.4427
Univar H-F Epsilon=	1	0.6750	1	6	0.4427

Analysis of Animals Expanded data set using within-subjects contrast variables as *Y*'s and species as a between-subjects factor. These within-subjects contrast variables were created as $\frac{\hat{\psi}}{\sqrt{\sum c_i^2}}$. This scaling of the contrast variables yields correct SS for interactions of the contrast variables with between-subjects effects and contrasts. (NB. This analysis doesn't yield any pure between-subjects information.)

species	subject	fall	spring	summer	winter	fall v spr	w v fa.sp,su	su v fa,sp
FOX	1	0	5	3	0	-3.5355339	-2.3094011	0.40824829
FOX	2	3	5	4	1	-1.4142136	-2.5980762	0
FOX	3	4	6	2	3	-1.4142136	-0.8660254	-2.4494897
COYOTE	1	4	7	8	2	-2.1213203	-3.7527767	2.04124145
COYOTE	2	5	6	6	4	-0.7071068	-1.4433757	0.40824829
COYOTE	3	7	8	9	5	-0.7071068	-2.5980762	1.22474487
WOLF	1	3	7	10	2	-2.8284271	-4.0414519	4.0824829
WOLF	2	5	6	8	3	-0.7071068	-2.8867513	2.04124145
WOLF	3	4	7	9	2	-2.1213203	-4.0414519	2.85773803

Least Squares Fit
Response fall v spr

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.728483	0.351364	-4.92	0.0027
species[COYOTE]	0.5499719	0.496904	1.11	0.3108
species[FOX]	-0.392837	0.496904	-0.79	0.4593

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
species	2	2	1.4444444	0.6500	0.5552

Effect Details
species
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
COYOTE	-1.178511	0.60858062	-1.1785
FOX	-2.121320	0.60858062	-2.1213
WOLF	-1.885618	0.60858062	-1.8856

Contrast
Test Detail

COYOTE	0.5	1
FOX	-1	0
WOLF	0.5	-1
Estimate	0.5893	0.7071
Std Error	0.7454	0.8607
t Ratio	0.7906	0.8216
Prob>|t|	0.4593	0.4427
SS	0.6944	0.75

Sum of Squares	1.4444444444
Numerator DF	2
Denominator DF	6
F Ratio	0.65
Prob > F	0.5552456576

Response w v fa.sp,su

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-2.726376	0.312629	-8.72	0.0001
species[COYOTE]	0.1283001	0.442124	0.29	0.7814
species[FOX]	0.8018754	0.442124	1.81	0.1197

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
species	2	2	4.5740741	2.6000	0.1537

Effect Details
species
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
COYOTE	-2.598076	0.54148857	-2.5981
FOX	-1.924501	0.54148857	-1.9245
WOLF	-3.656552	0.54148857	-3.6566

Contrast
Test Detail

COYOTE	0.5	1
FOX	-1	0
WOLF	0.5	-1
Estimate	-1.203	1.0585
Std Error	0.6632	0.7658
t Ratio	-1.814	1.3822
Prob>|t|	0.1197	0.2162
SS	2.8935	1.6806

Sum of Squares	4.5740740741
Numerator DF	2
Denominator DF	6
F Ratio	2.6
Prob > F	0.1537445335

Response su v fa,sp

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.179384	0.390209	3.02	0.0233
species[COYOTE]	0.0453609	0.551839	0.08	0.9372
species[FOX]	-1.859798	0.551839	-3.37	0.0150

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
species	2	2	20.259259	7.3919	0.0241

Effect Details
species
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
COYOTE	1.224745	0.67586250	1.2247
FOX	-0.680414	0.67586250	-0.6804
WOLF	2.993821	0.67586250	2.9938

Contrast
Test Detail

COYOTE	0.5	1
FOX	-1	0
WOLF	0.5	-1
Estimate	2.7897	-1.769
Std Error	0.8278	0.9558
t Ratio	3.3702	-1.851
Prob>|t|	0.015	0.1137
SS	15.565	4.6944

Sum of Squares	20.259259259
Numerator DF	2
Denominator DF	6
F Ratio	7.3918918919
Prob > F	0.0240591292

Here's the nonManova analysis (using subject[species]&Random as a factor). This analysis would be useful if one were simply interested in getting *SS*. However, note that although the between-subjects *F*-tests are correct, the within-subjects contrast tests use the pooled error term rather than the much better partitioned error terms. Also, the omnibus within-subjects tests make no adjustment for violation of the sphericity assumption.

species	subject	Season	Miles
FOX	1	fall	0
FOX	2	fall	3
FOX	3	fall	4
COYOTE	1	fall	4
COYOTE	2	fall	5
COYOTE	3	fall	7
WOLF	1	fall	3
WOLF	2	fall	5
WOLF	3	fall	4
FOX	1	spring	5
FOX	2	spring	5
FOX	3	spring	6
COYOTE	1	spring	7
COYOTE	2	spring	6
COYOTE	3	spring	8
WOLF	1	spring	7
WOLF	2	spring	6
WOLF	3	spring	7
FOX	1	summer	3
FOX	2	summer	4
FOX	3	summer	2
COYOTE	1	summer	8
COYOTE	2	summer	6
COYOTE	3	summer	9
WOLF	1	summer	10
WOLF	2	summer	8
WOLF	3	summer	9
FOX	1	winter	0
FOX	2	winter	1
FOX	3	winter	3
COYOTE	1	winter	2
COYOTE	2	winter	4
COYOTE	3	winter	5
WOLF	1	winter	2
WOLF	2	winter	3
WOLF	3	winter	2

Response Miles
Summary of Fit

RSquare	0.912181
RSquare Adj	0.829241
Root Mean Square Error	1.058475
Mean of Response	4.805556
Observations (or Sum Wgts)	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	17	209.47222	12.3219	10.9981
Error	18	20.16667	1.1204	Prob > F
C. Total	35	229.63889		<.0001

Test Denominator Synthesis

Source	MS Den	DF Den	Denom MS Synthesis
subject[species]&Random	1.12037	18	Residual
species	2.86111	6	subject[species]&Random
Season	1.12037	18	Residual
species*Season	1.12037	18	Residual

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob > F
subject[species]&Random	17.1667	2.86111	6	2.5537	0.0574
species	59.7222	29.8611	2	10.4369	0.0111
Season	106.306	35.4352	3	31.6281	<.0001
species*Season	26.2778	4.37963	6	3.9091	0.0112

Effect Details

species

Effect Test

Sum of Squares	F Ratio	DF	Prob > F
59.722222	10.4369	2	0.0111

Denominator MS Synthesis:
subject[species]&Random

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
COYOTE	5.9166667	0.48828877	5.91667
FOX	3.0000000	0.48828877	3.00000
WOLF	5.5000000	0.48828877	5.50000

Contrast

Test Detail

COYOTE	0.5	1
FOX	-1	0
WOLF	0.5	-1
Estimate	2.7083	0.4167
Std Error	0.598	0.6905
t Ratio	4.5288	0.6034
Prob>|t|	0.004	0.5683
SS	58.681	1.0417

Sum of Squares	59.722222222
Numerator DF	2
Denominator DF	6
F Ratio	10.436893204
Prob > F	0.0111292825

Season

Effect Test

Sum of Squares	F Ratio	DF	Prob > F
106.30556	31.6281	3	<.0001

Denominator MS Synthesis:
Residual

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
fall	3.8888889	0.35282516	3.88889
spring	6.3333333	0.35282516	6.33333
summer	6.5555556	0.35282516	6.55556
winter	2.4444444	0.35282516	2.44444

Contrast

Test Detail

fall	-0.333	-0.5	-1
spring	-0.333	-0.5	1
summer	-0.333	1	0
winter	1	0	0
Estimate	-3.148	1.4444	2.4444
Std Error	0.4074	0.4321	0.499
t Ratio	-7.727	3.3427	4.899
Prob>|t|	4e-7	0.0036	0.0001
SS	66.898	12.519	26.889

Sum of Squares	106.30555556
Numerator DF	3
Denominator DF	18
F Ratio	31.628099174
Prob > F	2.1733284e-7

species*Season

Effect Test

Sum of Squares	F Ratio	DF	Prob > F
26.277778	3.9091	6	0.0112

Denominator MS Synthesis:
Residual

Contrast

Test Detail

COYOTE,fall	0.0833	0.125	0.25	-0.167	-0.25	-0.5
COYOTE,spring	0.0833	0.125	-0.25	-0.167	-0.25	0.5
COYOTE,summer	0.0833	-0.25	0	-0.167	0.5	0
COYOTE,winter	-0.25	0	0	0.5	0	0
FOX,fall	-0.167	-0.25	-0.5	0	0	0
FOX,spring	-0.167	-0.25	0.5	0	0	0
FOX,summer	-0.167	0.5	0	0	0	0
FOX,winter	0.5	0	0	0	0	0
WOLF,fall	0.0833	0.125	0.25	0.1667	0.25	0.5
WOLF,spring	0.0833	0.125	-0.25	0.1667	0.25	-0.5
WOLF,summer	0.0833	-0.25	0	0.1667	-0.5	0
WOLF,winter	-0.25	0	0	-0.5	0	0
Estimate	0.6944	-1.708	0.4167	0.6111	-1.083	0.5
Std Error	0.4321	0.4583	0.5292	0.499	0.5292	0.6111
t Ratio	1.6071	-3.727	0.7873	1.2247	-2.047	0.8182
Prob>|t|	0.1254	0.0015	0.4414	0.2365	0.0555	0.424
SS	2.8935	15.565	0.6944	1.6806	4.6944	0.75

Sum of Squares	26.277777778
Numerator DF	6
Denominator DF	18
F Ratio	3.9090909091
Prob > F	0.0112461007

The complete ANOVA summary table in the format we have been using is:

Animals expanded data set

	SS	*df*	*MS*	*F*	*p*
Species	59.72222	2	29.8611	10.44	0.0111
coyote, wolf v. fox (A1)	58.68056	1	58.6806	20.51	0.004
coyote v. wolf (A2)	1.041667	1	1.04167	0.36	0.5683
Subject[Species]&Random	17.16667	6	2.86111	2.55	0.0574
Season	106.3056	3	35.4352	31.63	0
w v fa,sp,su (S1)	66.898	1	66.898	76.05	0.0001
su v fa,sp (S2)	12.519	1	12.519	9.14	0.0233
fall v. spring (S3)	26.88889	1	26.8889	24.2	0.0027
Species*Season	26.27778	6	4.37963	3.91	0.0112
A1 x S1	2.893536	1	2.89354	3.29	0.1197
A1 x S2	15.56546	1	15.5655	11.36	0.015
A1 x S3	0.694444	1	0.69444	0.63	0.4593
A2 x S1	1.680529	1	1.68053	1.91	0.2162
A2 x S2	4.694676	1	4.69468	3.43	0.1137
A2 x S3	0.75	1	0.75	0.68	0.4427
Error	20.16667	18	1.12037		
S1 Error	5.277768	6	0.87963		
S2 Error	8.22257	6	1.37043		
S3 Error	6.666667	6	1.11111		