

Meyer, L. S., Gamst, G., & Guarino, A. J. (2006). Applied multivariate research: Design and interpretation. Thousand Oaks, CA: Sage.

## CHAPTER 9A

# MANOVA

## Comparing Two Groups

---

**A**s we saw in Chapters 8A and 8B, ANOVA is applied to designs with single dependent measures and is thus conceived as a univariate procedure. Even if researchers measured more than one dependent variable, as is often the case, they can analyze only one at a time using ANOVA. The next six chapters cross into the domain of multivariate ANOVA in which many dependent variables are simultaneously analyzed within a single ANOVA design. When we make this transition, we find ourselves in the domain of *multivariate analysis of variance*, abbreviated as MANOVA.

This chapter introduces MANOVA by focusing on the two-group between-subjects design. We are thus interested in assessing the effects of one dichotomous (two-group) independent variable on two or more quantitative dependent variables. For example, we might study the effects of gender (male, female) on worker job satisfaction and worker absenteeism, or we might be interested in the effect of treatment type (cognitive-behavioral, psychoanalytic) on global assessment of functioning (GAF) and consumer service satisfaction.

Although these research questions can certainly be addressed with univariate statistical procedures that examine each dependent variable separately (e.g., using *t* tests or one-way ANOVAs), they can also be profitably explored (or properly done) in the multivariate realm by examining the dependent measures *collectively* and *simultaneously* by means of a multivariate generalization of the *t* test called Hotelling's  $T^2$  (or MANOVA for the two-group independent variable context).

## The Use of MANOVA

The strategy of taking more than one index of the behavior of participants in a research study has much to be said in its favor. Rarely is one aspect of behavior so isolated from other aspects of the overall response that it can paint a comprehensive picture of how someone responded to a situation. To return to an example that we just mentioned, we would certainly expect that employees who were more satisfied with their job would in general have less absenteeism than those who were less satisfied. In measuring the effects of some sort of workplace intervention program, both satisfaction and absenteeism are likely to be indicators of some more general or latent variable concerning the feelings of employees toward their job. This latent variable might be called "employee contentment" and would represent the composite variable subsuming satisfaction and absenteeism. Focusing on only one aspect of employee contentment, such as job satisfaction, thus provides only part of the issue in which we as researchers are really interested.

### MANOVA and Variates

This is similar to the multiple regression situation where the weighted set of predictors can be thought of as forming a variate that is then related to the criterion variable. An analogous situation occurs in MANOVA. The dependent variables in a MANOVA design are combined into a weighted linear composite. In MANOVA, the weights are determined to be those that allow the variate to maximally distinguish (differentiate, discriminate) between the groups (the levels of the independent variable) in the study. Multiple regression yields a single weighted combination of predictors because the criterion variable is quantitatively measured. Because the target of the prediction in MANOVA is the independent variable, which is treated as a categorical variable, the number of variates derived by a MANOVA is the lesser of (a) the number of dependent variables and (b) the number of groups (levels of the independent variable) minus 1. Because this chapter explores differences between two groups, the number of derived variates will always be 1.

### MANOVA and Vectors

Single measures are often referred to as scalar measures. We use them as dependent variables in experimental designs because that single measure is considered adequate in assessing a variable of interest to the researchers. Multiple measures are often referred to as *vectors*. Vectors are

a set (combination) of numbers that describes a phenomenon and is very useful in those all-too-frequent situations where no single number is sufficient to quantify the phenomenon. Essentially, the variate can be thought of as an example of a vector where the variate value is computed as a weighted sum of its components.

Vectors are fairly commonly used by us in everyday living. An example of a vector from our general experience would be speed. We understand speed in a rather direct sense of how fast we are traveling. But speed is really composed of two separate physical variables: distance and time. Thus, if it takes us half an hour to travel 30 miles, we know that we are traveling at 60 miles per hour. In this case, we have divided distance by time to calculate speed. In the MANOVA designs we will be discussing, it is appropriate to speak of vectors to represent the dependent variable variate. Here, the vectors are computed as a weighted sum of the dependent variables.

### When MANOVA Is Appropriate

So what exactly is the benefit of conducting a simultaneous analysis of multiple dependent variables? Seven immediate advantages can be identified (Bray & Maxwell, 1985; Stevens, 2002).

First, single dependent measures seldom capture completely a phenomenon being scrutinized. Multiple measures provide the researcher with a certain amount of useful redundancy (through the correlation of the multiple measures) and the ability to broaden or enhance the conceptual domain under study. For example, job satisfaction could be tapped with a single item: "How satisfied are you with your current job?" Although such a global measure will provide useful information, a better approach would be to explore separate and unique job satisfaction facets (e.g., satisfaction with pay, benefits, coworkers, location, etc.) that can initially be combined into one global job satisfaction vector that can subsequently be decomposed into its separate job satisfaction constituents.

Second, MANOVA provides some control over the overall alpha level or Type I error rate (i.e., the chance of making a false rejection of the null hypothesis). If we were to examine gender differences (the independent variable), for example, with four job satisfaction dependent variables (pay, benefits, coworkers, and location) using four separate univariate *t* tests or one-way ANOVAs each evaluated at the .05 alpha level, we would expect a statistically significant effect 5% of the time for each dependent measure. In practice, the alpha level we wind up with across these four analyses actually lies somewhere between 5% and 18.5% (i.e.,  $1 - (.95)(.95)(.95)(.95)$ ). The point is that multiple univariate *t* tests or ANOVAs can inflate the operational alpha level

(Type I error), a state of affairs often called *probability pyramiding*; using MANOVA avoids this problem (see Hummel & Sligo, 1971).

Third, univariate statistical tests tend to ignore the intercorrelation found between dependent variables. As we will see, MANOVA considers dependent variable intercorrelation by examining the variance-covariance matrices.

Fourth, MANOVA enables researchers to examine relationships between dependent variables at each level of the independent variable.

Fifth, MANOVA provides researchers with statistical guidance to reduce a large set of dependent measures to a smaller assemblage.

Sixth, MANOVA helps to identify dependent variables that produce the most group (independent variable) separation or distinction.

And seventh, MANOVA can "tease out" group differences that may become masked with univariate statistical analyses but are discovered under conditions of increased power in the multivariate situation.

### When MANOVA Should Not Be Used

Conversely, there are at least three circumstances under which we would either not want to use MANOVA or approach MANOVA with considerable caution (see Bray & Maxwell, 1985).

First, MANOVA should not be used if the dependent variables are uncorrelated. As we will see, the ideal situation for using MANOVA is when the dependent variables are moderately correlated. Weinert (1995), for example, uses an example in which the correlations between three dependent variables ranged between .21 and .36 to illustrate the appropriateness of a MANOVA design.

Second, and this is the other side of what we just talked about, MANOVA should not be used with a set of dependent variables that is very highly correlated. Statistically, such correlations will run the risk of a multicollinearity condition, which could cause the analysis to yield improper results if SPSS allowed the analysis to proceed in the first place. Conceptually, to the extent that variables are highly correlated, they can be said to be measuring the same construct and are therefore redundant.

There are two common situations that we have found where students encounter this problem. One situation occurs when students use the subscales of an inventory together with the total inventory scores as dependent variables. The subscales here are portions of the total score and thus in combination correlate very highly (sometimes almost perfectly) with the total score. The solution to this situation is to run two analyses, one using only the subscales and another using the total score.

A corre  
one c  
for a  
dent recip  
exam well  
linea  
A  
appr  
the r  
The  
is co  
great  
their  
the b  
put c  
help

As n  
whet  
varia  
asses  
hypc  
(for  
S

Mea  
is th

A second situation in which the dependent variables are too highly correlated occurs when one of the dependent variables is computed from one or more of the others. For example, if researchers use the time it takes for a response to be made as well as the speed of the response as dependent variables, they will have created multicollinearity because speed is the reciprocal of time and the two are therefore perfectly correlated. As another example, using two variables such as GAF at Time 1 and GAF at Time 2 as well as the difference score (GAF 2 minus GAF 1) would create perfect linear dependence that would produce multicollinearity.

A third situation in which MANOVA should not be used or, perhaps more appropriately, should be used with great caution is when the target user or the recipient of the analysis has limited statistical or technical knowledge. The assumptions underlying MANOVA as well as the output of the analysis is considerably more complex than its univariate cousins, and this poses greater interpretation and communication challenges for the researchers and their audience. Using complicated statistical tools such as MANOVA places the burden squarely on the shoulders of the researchers to explain the output of the analysis in ways that do justice to the results while at the same time help the end user understand what was found.

### The Univariate *t* Test

As noted in Chapter 8A, the univariate *t* test addresses the question of whether two population means are equal. For example, the independent variable could be gender (male, female) and the dependent variable global assessment of functioning (GAF). The *t* test is a way of testing the null hypothesis ( $H_0$ ), which states that in the population there is no difference (for example) on GAF score between men and women.

Symbolically, the null hypothesis can be expressed as follows:

$$H_0: \mu_{\text{Men}} = \mu_{\text{Women}}$$

To test the null hypothesis, we calculate the following statistic:

$$t = \frac{\text{Mean}_{\text{Men}} - \text{Mean}_{\text{Women}}}{\sqrt{\text{SE}_{\text{diff}}}}$$

Where  $\text{Mean}_{\text{Men}}$  is the mean value on the dependent measure for men,  $\text{Mean}_{\text{Women}}$  is the mean value on the dependent value for women, and  $\text{SE}_{\text{diff}}$  is the standard error of the difference between two means. The *t* statistic is

evaluated with  $N - 2$  degrees of freedom, where  $N$  is the total number of scores.

The top portion of this  $t$  ratio is the mean difference between the two genders on a single dependent variable; it can be conceptualized as an index of between-group variability. The term in the denominator of this ratio is the standard error of the difference between means; it can be thought of as an index of within-group variability. This standard error is calculated by following this set of steps: (a) divide the variance of the men's scores by the number of men in the sample, (b) divide the variance of the women's scores by the number of women in the sample, (c) subtract the result of (b) from the result of (a) to achieve a single value, and (d) take the square root of this single value.

The standard error of the difference between means is, as is true for virtually all standard error statistics, an estimated standard deviation of a hypothetical sampling distribution. Here, the sampling distribution is based on mean differences. Hypothetically, if we randomly sampled the same number of men and women and assessed them on our GAF dependent variable, we would obtain a certain mean difference. If we did it again, we would find another mean GAF score difference. If we sampled an infinite number of times, we would obtain a certain amount of variability in these mean differences (greater sample sizes would result in less variability of the sample mean differences). We would also find that these difference values were distributed in a normal manner. The standard error of mean differences is the standard deviation of this normal sampling distribution. Our best estimate of that standard deviation is  $SE_{\text{diff}}$ .

The  $t$  ratio is thus a count of the distance between two means using the  $SE_{\text{diff}}$  as a metric. That is, we can determine the difference between means in  $SE_{\text{diff}}$  units. For example, if the mean for men was 60 and the mean for women was 50, the difference would be 10. With a  $SE_{\text{diff}}$  of 2, the genders can be thought of as differing by 5  $SE_{\text{diff}}$  units, which in the context of  $t$  is a relatively large value.

### The Multivariate Hotelling's $T^2$

In the previous univariate example, we hypothesized about the relationship between gender (the independent variable) and GAF score (the dependent variable). Suppose we included a second, but related, dependent variable such as service satisfaction to our study. Assume that clients' GAF score and satisfaction with the service they received are moderately correlated. Although separate univariate  $t$  tests (one for each dependent variable) are certainly an option, a more acceptable as well as more elegant approach

would be to employ *Hotelling's T*<sup>2</sup> (or MANOVA in the two-group case). In this situation, the linear composite or variate might be thought of as "therapy efficacy."

Hotelling's *T*<sup>2</sup> creates a vector (variante or weighted linear composite) that best separates the levels or categories of the independent variable. Hotelling's *T*<sup>2</sup> tests a multivariate null hypothesis of the general form:

$$H_0: \begin{bmatrix} \mu_{11} = \mu_{12} \\ \mu_{21} = \mu_{22} \\ \dots = \dots \\ \dots = \dots \\ \dots = \dots \\ \mu_{p1} = \mu_{p2} \end{bmatrix}$$

The above hypothesis indicates that the population mean vectors are equal across the two groups. Note that  $p$  is the number of dependent variables and that the two subscripted numbers refer to the dependent variable and the independent variable group or level, respectively. Hence,  $\mu_{21}$  represents the population mean of Dependent Variable 2 for Group 1. For the present example, where we have just the two dependent variables of GAF score and satisfaction, the null hypothesis simplifies to this:

$$H_0: \mu_{\text{GAF Males}} = \mu_{\text{GAF Females}}$$

$$\mu_{\text{Satis Males}} = \mu_{\text{Satis Females}}$$

As is true for the univariate  $t$ , this multivariate version produces a test statistic that can be compared with a critical value to determine statistical significance. The main difference between the two is that the univariate  $t$  test compares two population means, whereas Hotelling's *T*<sup>2</sup> multivariate analogue compares two vectors of means.

### Some Mathematical Aspects of $T^2$

The calculations that produce the  $T^2$  statistic are necessarily complex (and tedious) because of the manipulation of vectors or matrices of numbers and are beyond the scope and rationale of this text. However, some of the conceptual underpinnings of these calculations can be profitably explored.

As we noted previously, the univariate  $t$  test is essentially a ratio of between-group variability over within-group variability or error. This ratio

produces a coefficient (the  $t$  value) that can be subsequently evaluated for statistical significance. Stevens (2002) shows that Hotelling's  $T^2$  can be computed by (a) substituting the dependent variable means with a vector of means for each group and (b) replacing the univariate error term (or denominator) by its matrix analogue  $S$  (the estimated population covariance matrix).

Using matrix algebra (see Stevens, 2002), it is possible to mathematically generalize the univariate  $t$  to its multivariate counterpart. Similar to its univariate counterpart ( $t$ ), Hotelling's  $T^2$  creates a ratio of between-group variability (based on the mean vectors) to within-group variability (as represented by the inverse of the covariance matrix). Through matrix algebra, these matrices of numbers are reduced to a single value called a *determinant* that expresses the generalized variance of a matrix. Based on this, we can make the multivariate assessment of between- to within-group variance (Harris, 2001).

When Hotelling provided us with his  $T^2$ , he knew that it was related to  $F$ . Stevens (2002) notes that Hotelling (1931) was the first to demonstrate that  $T^2$  can be transformed into the  $F$  distribution using the following conversion formula:

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2$$

with  $p$  and  $(N - p - 1)$  degrees of freedom ( $p$  equals the number of dependent variables). In the above conversion formula,  $n_1$  and  $n_2$  represent the sample sizes of Groups 1 and 2.

### Interpretation of Hotelling's $T^2$

The Hotelling's  $T^2$  computed value can be compared with a critical value to determine its significance level (Hair et al., 1998). If its probability of occurrence is below a predetermined alpha level (e.g.,  $p < .05$ ), then we are able to conclude that gender differences (if that is our independent variable) exist and that there are unequal mean vectors for our dependent variables (Weinfurt, 1995). That is, we can conclude that the two groups differ significantly on the population values of the vector represented by the dependent variables used in the analysis.

In practice, most computer statistical packages do not routinely display a Hotelling's  $T^2$  value and its corresponding degrees of freedom and level of significance. Instead,  $T^2$  is transformed, by means of the previously noted formula (and its facsimiles), into  $F$  values with four multivariate statistics

commonly lambda, Pi tests asses value, whi of these f in Chapte The impo the same (Tabachni

A stat .05 level c exist. Cor example, composite isfaction i

Once a st we reject equal on to examí procedur juncture normally that the on the de

### Multipl

Perh cance is with an 1995). T which re due to t formula number the follo level we

commonly found in MANOVA statistical computer programs: Wilks's lambda, Pillai's trace, Hotelling's trace, and Roy's largest root. Each of these tests assesses the multivariate between- and within-group variability and computes a multivariate test coefficient that is in turn converted to an  $F$  value, which is evaluated much as any other  $F$  statistic. The specifics on each of these four multivariate test statistics will be discussed in greater detail in Chapter 10A when we explore MANOVA with three or more groups. The important point to remember here is that these four tests will produce the same ( $F$  test) result when the independent variable has only two levels (Tabachnick & Fidell, 2001b).

A statistically significant effect, as evidenced by the above tests (at the .05 level or less) indicates that group differences on the dependent variate exist. Concretely, if we obtained a significant  $F$  ratio for our previous example, we would conclude that gender differences exist on the weighted composite (dependent variate) of respondents' GAF scores and service satisfaction ratings.

### What to Do After a Significant Multivariate Effect

Once a statistically significant multivariate effect has been established (i.e., we reject the null hypothesis that our independent variable groupings are equal on the weighted composite dependent variate), researchers need to examine the nature of the variate. There are a number of established procedures, which we will now briefly review. It should be noted at this juncture that if the multivariate test is *not* significant ( $p > .05$ ), we would normally not proceed with any further analysis. Instead, we would conclude that the dichotomous independent variable is not differentially distributed on the dependent measures.

#### Multiple Univariate $t$ or $F$ Tests

Perhaps the most popular procedure to follow up multivariate significance is to conduct separate  $t$  tests or ANOVAs on each dependent variable with an adjusted alpha level (e.g., Hair et al., 1998; Stevens, 2002; Weinfurt, 1995). The adjustment to the alpha level is called a *Bonferroni correction*, which reduces the possibility of operating with an inflated Type I error rate due to the use of multiple univariate tests. The adjustment is made by the formula  $\alpha/p$ , where the omnibus alpha level (typically .05) is divided by the number of dependent variables. Based on our previous example, we have the following: .05/2 or .025, which becomes the new (more stringent) alpha level we would then use to evaluate each dependent measure. These

univariate tests require no additional post hoc multiple comparison tests because the independent variable has only two treatments or levels. Although we recommend this approach, other commentators (Bray & Maxwell, 1985; Tabachnick & Fidell, 2001b; Weinfurt, 1995) urge caution with this process and recommend some of the following procedures instead.

### Roy-Bargmann Step-Down Analysis

*Step-down analysis* is an alternative procedure for assessing each dependent variable separately following a statistically significant multivariate effect. Here, a univariate  $F$  value is computed for each dependent variable after controlling for the effects of the remaining dependent measures in the analysis. This procedure is analogous to hierarchical regression in that we require that the dependent measures have a logical, theoretically based, *a priori* causal ordering (Tabachnick & Fidell, 2001b; Weinfurt, 1995). The dependent variables would then be evaluated in that order.

Consider an example with the three dependent variables of GAF posttest, satisfaction with the mental health services that were provided, and number of visits clients made to the facility to illustrate this step-down process. Assume that our literature review or clinical experience led us to the understanding that GAF posttest scores, which reflect the general state of health of clients after treatment, are a major determinant of their assessment of service satisfaction. In turn, satisfaction with the services they receive is a major reason or cause for how often clients actually come to the mental health facility. Under these conditions, we would have a reasonable foundation on which to base a step-down analysis.

The first step-down  $F$  would examine GAF posttest scores and would be the same as a standard univariate  $F$  test. The second step-down  $F$  would examine service satisfaction while holding GAF posttest scores constant. Here, GAF posttest scores would essentially be treated as a covariate, and thus all remaining step-down "steps" are really separate analyses of covariance (ANCOVAs). The third step-down  $F$  would test mental health visits while holding both GAF posttest scores and service satisfaction constant. From this procedure, the investigator can determine the relative contribution of the dependent measures producing the multivariate effect.

We certainly recommend this approach if you are working with correlated multiple dependent measures *and* if these measures contain some logical hierarchical ordering. However, we will not emphasize this procedure because of issues of pragmatics and continuity. In SPSS, Roy-Bargmann step-down analyses can be produced only through the syntax-based

**MANOV**  
**Multivar**  
and Fide  
sis. We n  
a series  
GAF pos  
and thei  
depende

### Discrim

Reca  
of depe  
or categ  
is "next  
ables tak  
predicto  
or MANC  
analyses,  
entiates  
the role  
ing men  
variable  
of this y

Con  
or MAN  
variate b  
ture" of  
analog  
unique c

### Sample

Bec  
depend  
than its  
the num  
ables. So  
to achie

on tests  
levels.  
(Bray &  
caution  
procedures

ng each  
tivariate  
variable  
es in the  
in that  
ly based,  
95). The

of GAF  
provided,  
ep-down  
ed us to  
eral state  
ir assess-  
ces they  
ne to the  
asonable

id would  
 $F$  would  
constant.  
iate, and  
of covari-  
lth visits  
constant.  
contribu-

th corre-  
ain some  
is proce-  
largmann  
tax-based

**MANOVA** procedure and not through the menu-driven **GLM → Multivariate** procedure. The interested reader should consult Tabachnick and Fidell (2001b) regarding SPSS programming details for step-down analysis. We note in passing that step-down analyses can also be achieved through a series of ANCOVAs. For example, one can perform one ANCOVA using GAF posttest scores as the dependent variable and satisfaction as a covariate and then perform a second ANCOVA using GAF posttest scores as the dependent variable and both satisfaction and number of visits as covariates.

### Discriminant Analysis

Recall that the strategy used by Hotelling's  $T^2$  (MANOVA) creates a variate of dependent variables that maximizes the difference between the two levels or categories of the independent variable. As we noted in Chapter 7A, MANOVA is "next of kin" to discriminant analysis. In these two analyses, the same variables take on different roles: The continuous variables are assigned the role of predictors or independent variables in discriminant analysis. In Hotelling's  $T^2$  or MANOVA, these variables take on the role of dependent variables. In both analyses, these variables are combined into a linear composite that best differentiates the levels of the categorical variable. The categorical variable takes on the role of dependent variable in discriminant analysis in that we are predicting membership in one of the two groups. In Hotelling's  $T^2$  or MANOVA, this variable is the independent variable, and we are evaluating whether the levels of this variable (the groups) differ from each other on the variate.

Conducting a discriminant analysis following a statistically significant  $T^2$  or MANOVA allows the investigator to better understand the nature of the variate by providing both the structure coefficients to determine the "structure" of the variate and the weights of the dependent variables (which are analogous to the beta weights in multiple regression) to demonstrate the unique contributions of the measured variables (Weinfurt, 1995).

## Special Issues Concerning Hotelling's $T^2$ and MANOVA

### Sample Size Requirements

Because of the additional burden of analyzing simultaneously multiple dependent measures, Hotelling's  $T^2$  or MANOVA requires larger sample sizes than its univariate ANOVA counterpart. A minimal sample size heuristic is that the number of cases per cell must exceed the number of dependent variables. Some authors (e.g., Hair et al., 1998) argue for at least 20 cases per cell to achieve minimal levels of power.

### Dependent Variables

Hotelling's  $T^2$  or MANOVA is most efficient with high negative correlation or moderate correlation (.6) among the dependent variables (Tabachnick & Fidell, 2001b). Dependent variables that have high positive correlations are redundant, so using this procedure is considered somewhat wasteful and counterproductive. Two possible solutions to such strong correlation (.8 or .9) would be as follows:

- ▶ Create a new composite dependent variable. Although there is no particular limit on the number of dependent variables you can use to build the dependent variate, we encourage judicious restraint and recommend no more than 10 dependent conceptually related variables (Bray & Maxwell, 1985).
- ▶ Delete one (or more) of the dependent variables prior to the analysis so that such a strong correlation is no longer observed.

This multivariate analysis should also not be used if there are very low correlations between the dependent measures. Instead, separate ANOVAs with a Bonferroni adjustment would probably be in order. A significant *Bartlett's test of sphericity* ( $p < .001$ ) is indicative of sufficient correlation between the dependent variables to proceed with the multivariate analysis.

### Power of Multivariate Tests

*Power* concerns the adequacy of your study's statistical test (e.g., one of the four multivariate tests discussed previously) to detect an actual treatment effect. The power of a statistical test is a function of three parameters: the alpha level, sample size, and the effect size (i.e., the extent to which treatment groups differ on the dependent variable). There is an inverse relationship between power and alpha level. Power decreases when alpha levels become more stringent (i.e., move from .05 to .01). Thus, moving to the .01 or .001 alpha level to reduce the possibility of Type I error also reduces your power. The trick is to find a proper balance between alpha and power. This balance is informed by considering the effect size of the variables under study. In the present two-group example, effect size is the difference in the means divided by their standard deviation (Stevens, 1980). In practice, the power of a statistical test (regardless of sample size) will always increase if the effect size increases. Small effect sizes require larger samples to produce adequate statistical power of, for example, .80 (i.e., we have an 80% chance of detecting a significant result if the effect exists). Increasing sample size reduces sampling error and subsequently increases power.

Table 9:

Number Depende Variable
2
2
2
2
3
3
3
3
3
5
5
5
5
5
7
7
7
7
7

SOURCE

a.  $N_s = N$

b.  $D^2 = N$

We  
Table 9  
that sm  
to 25 p  
power  
group i

### Outline

Hc  
ANOVA  
dent m  
Type I  
becom

**Table 9a.1** Power of Hotelling's  $T^2$  at  $\alpha = .05$  for Small Through Very Large Overall Effect and Group Sizes

Number of Dependent Variables	$N_a$	Effect Size ( $D^{2b}$ )			
		Small	Medium	Large	Very Large
2	15	.26	.44	.65	.95
2	25	.33	.66	.86	.97
2	50	.60	.95	1.00	1.00
2	100	.90	1.00	1.00	1.00
3	15	.23	.37	.58	.91
3	25	.28	.58	.80	.95
3	50	.54	.93	1.00	1.00
3	100	.86	1.00	1.00	1.00
5	15	.21	.32	.42	.83
5	25	.26	.42	.72	.96
5	50	.44	.88	1.00	1.00
5	100	.78	1.00	1.00	1.00
7	15	.18	.27	.37	.77
7	25	.22	.38	.64	.94
7	50	.40	.82	.97	1.00
7	100	.72	1.00	1.00	1.00

SOURCE: Adapted from Stevens (1980).

a.  $N_s = N_1 = N_2$ .

b.  $D^2$  = Mahalanobis distance.

We demonstrate the collaborative role of these three parameters in Table 9a.1, which is adapted from Stevens (1980). From Table 9a.1, we note that small or moderate effect sizes with sample sizes greater than or equal to 25 produce relatively poor power ( $< .45$ ) and never achieve adequate power ( $> .70$ ). Conversely, when effect sizes are large, then 15 cases per group is sufficient to yield moderate or greater statistical power.

## Statistical Assumptions and Limitations

### Outliers and Missing Values

Hotelling's  $T^2$  (MANOVA) and its univariate counterpart ( $t$  test and ANOVA) are particularly sensitive to outliers or extreme values on the dependent measures. Failure to exclude outliers or transform the data could inflate Type I or II error rates. Likewise, missing values in a multivariate analysis become more problematic because of the complexity of the dependent

variate. We encourage the reader to review relevant sections in Chapters 3A and 3B for discussions on how to address these problems.

### **Independence**

The participants, respondents, or cases that compose the levels or groups of an independent variable must be independent of each other when conducting a Hotelling's  $T^2$  or MANOVA. Experimentally, independence of participants is assumed if participants have been randomly assigned to treatments or conditions of the study. Occasionally, problems occur with the use of intact groups (e.g., selecting as participants whole sets of individuals, such as classes in a program or clients of a given facility) where the entire set of individuals is exposed to or placed in one treatment level.

Sometimes dependence occurs with quasi-experimental or archival research where time-ordered effects may occur for respondents who provide information over time (e.g., GAF scores at Time 1 and Time 2). When this occurs, you are probably better off treating the dependent variables as a repeated measure and using the appropriate ANOVA design (one-way within-subjects or mixed design) to analyze the data.

### *Homogeneity of Variance-Covariance Matrices*

The assumption of equivalence or homogeneity of covariance matrices for each dependent variable across groups is the multivariate analogue to the homogeneity of variance assumption in ANOVA and multiple regression. Instead of testing for comparability of variances on a single dependent measure between the levels of the independent variable, all the coefficients in the covariance matrix of dependent variables are examined. Violation of this homogeneity of covariance matrices assumption when sample sizes are fairly equal produces minor consequences.

The standard vehicle for assessing equivalence-of-covariance matrices is the Box's  $M$  test, where statistical significance ( $p < .001$ ) is indicative of heterogeneity or inequality. Remedies for this assumption violation include the usual transformations of the dependent variables. If heterogeneity persists despite the best efforts of researchers to use the techniques described in Chapter 3A, then we recommend using Pillai's criterion to evaluate the significance of the multivariate effect (Tabachnick & Fidell, 2001b).

### **Multivariate Normality**

The Mardia's statistic is a way to assess for multivariate normality, but this statistic is not available in SPSS. Because no specific computerized test

is readily available, check for violations are typically addressed by

### **Linearity**

Linear relationships. If nonlinearity is present, it should be in order to

Because Mardia's statistic is used to discuss the assumption of homogeneity here fits the type of analysis. GAF and other scores (such as those that these variables are assigned to included in this Hotteling's test we hope that the reader will Condense

Table 3.1 shows that we note that the model is the dependent variable (therapeutic intervention had who received the deviation from the variability of the two dependent variables for the c

is readily available in SPSS to test for multivariate normality, most investigators check for univariate normality for each dependent variable, and transformations are used to address any normality departures. Note that Box's  $M$  test is highly sensitive to normality violations; thus, normality should probably be addressed before computing Box's  $M$  test.

### Linearity

Linear relationships are assumed between pairs of dependent variables. If nonlinear (curvilinear) relationships are observed, transformations may be in order.

## Numerical Hypothetical 2-Group Example

Because MANOVA can involve a certain amount of complexity, it is useful to discuss an example analysis to show how the material we have presented here fits together. Assume that we are interested in the effects of therapy type (cognitive-behavioral vs. psychoanalytic) as the independent or grouping variable on the dependent variables of community mental health clients' GAF and service satisfaction (Client Satisfaction Questionnaire-8, CSQ-8) scores (see Chapter 3B for details on these measures). Further assume that these severely mentally ill clients were randomly and independently assigned to one of the two treatment conditions. Note that we could have included more than two dependent measures, but we have chosen to keep this Hotelling's  $T^2$  example as simple as possible. The null hypothesis we hope to reject is that the vectors of means for each group are equal (i.e., therapy type has no differential effect on our dependent variate). Condensed summaries of results can be seen in Tables 9a.2 through 9a.7.

Table 9a.2 presents the data for 55 study participants. From Table 9a.2, we note that although the samples are fairly small, sufficient statistical power should be available to detect a real difference in the population if we assume that moderate effect sizes are present for our manipulation. An inspection of the dependent variable means for each group or level of the independent variable (therapy type) shows that the clients receiving cognitive-behavioral treatment had somewhat higher mean scores on GAF and CSQ-8 than did clients who received psychoanalytic intervention. The dependent variable standard deviations for each group are fairly close, suggesting comparable levels of variability for each dependent measure. Pearson  $r_s$  were computed between the two dependent measures (GAF and CSQ-8) for each group or level of the independent variable. A moderate positive correlation ( $r = .426$ ) was found for the cognitive-behavioral group, and a low zero-order correlation ( $r = .086$ )

**Table 9a.2** Client Global Assessment of Functioning (GAF) and Service Satisfaction (CSQ-8), by Therapy Type (Hypothetical Data for Hotelling's  $T^2$  or Two-Group MANOVA)

Caseid	Therapy Type					
	Group 1: Cognitive-Behavioral ( $n = 30$ ) <sup>a</sup>		Group 2: Psychoanalytic ( $n = 25$ ) <sup>b</sup>		GAF	CSQ-8
	Caseid	GAF	Caseid	GAF		
1	65	30.3	31	50	24.3	
2	69	27.1	32	55	22.1	
3	73	31.0	33	61	25.6	
4	60	29.4	34	66	25.0	
5	58	27.4	35	58	24.0	
6	70	30.1	36	50	26.0	
7	71	30.9	37	45	27.0	
8	55	27.0	38	53	26.5	
9	65	28.5	39	59	25.5	
10	60	29.1	40	61	25.0	
11	66	27.8	41	70	29.0	
12	60	29.2	42	65	26.2	
13	70	30.3	43	55	24.8	
14	70	31.0	44	50	24.0	
15	65	26.5	45	62	23.2	
16	65	27.0	46	66	26.0	
17	55	26.5	47	69	27.0	
18	72	30.0	48	50	25.1	
19	70	31.0	49	58	28.1	
20	65	28.7	50	55	25.5	
21	60	28.0	51	60	23.2	
22	55	27.0	52	60	23.0	
23	70	30.1	53	65	22.0	
24	75	29.2	54	65	24.0	
25	65	25.0	55	50	25.0	
26	62	24.3				
27	60	30.3				
28	68	26.7				
29	70	28.0				
30	68	25.5				
		<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>
GAF	65.23	5.62			58.32	6.84
CSQ-8	28.43	1.90			25.08	1.73

a.  $r = .426; p < .01$ .

b.  $r = .086; p > .05$ .

was observe variability in Box's  $M$  test

Table 9a.2 dependent combinationity violation statistics; all violations o of the covari

Box's  $M$  that the Bo

**Table 9a.3**

GAF	Not
	K
	S
	Ske
	Kur
CSQ-8	No
	F
	S
	Ske
	Kur

**Table 9a.4**

Box's  $M$   
Levene's

was observed for the psychoanalytic group. These correlations suggest some variability in the dependent variable covariance and will be assessed with Box's  $M$  test.

Table 9a.3 examines group tests of normality and diagnostics for each dependent variable. Normality tests for each group-dependent variable combination were not statistically significant, indicating no serious normality violations. This was also confirmed with the skewness and kurtosis statistics; all were within the -1 to +1 range. Because there are no normality violations observed here, we can proceed with confidence to an assessment of the covariance matrix.

Box's  $M$  and Levene's test are shown in Table 9a.4. Here, we observe that the Box's  $M$  test of the equality of variance-covariance matrices is not

**Table 9a.3** Client Global Assessment of Functioning (GAF) and Service Satisfaction (CSQ-8) Normality Tests and Diagnostics, by Therapy Type

		Group 1: Cognitive-Behavioral		Group 2: Psychoanalytic	
		Statistic	Significance	Statistic	Significance
GAF	Normality test				
	Kolmogorov-Smirnov	.150	.083	.128	.200
	Shapiro-Wilk	.939	.088	.958	.372
	Skewness	.353		.112	
	Kurtosis	.832		.928	
CSQ-8	Normality test				
	Kolmogorov-Smirnov	.129	.200	.081	.200
	Shapiro-Wilk	.946	.131	.982	.918
	Skewness	.362		.220	
	Kurtosis	.782		.033	

**Table 9a.4** Tests of Equality of Covariance Metrics and Equality of Error Variances

	Overall		GAF		CSQ-8	
	Statistic	Significance	Statistic	Significance	Statistic	Significance
Box's $M$	3.157	.338				
Levene's			1.41	.241	1.16	.286

significant, indicating equality or homogeneity. The separate Levene's tests for each dependent variable were not statistically significant either, which confirms the equal variances we believe to be true in our visual inspection of the data in Table 9a.2.

Bartlett's test of sphericity is shown to be statistically significant in Table 9a.5. This indicates that there is sufficient correlation between the two dependent variables to proceed with the analysis. A moderate positive Pearson correlation ( $r = .502$ ) was observed between GAF and CSQ-8.

We are now ready to review the multivariate test results in Table 9a.6. Instead of a Hotelling's  $T^2$  value, computer statistical programs such as SPSS translate  $T^2$  into four multivariate test statistics that are expressed as  $F$  values. In the two-group case, these  $F$  values will be identical, so we show only one of them (Roy's largest root). In the present case, this multivariate test shows that Roy's largest root yields an  $F$  value of 25.445, which is statistically significant. We may then conclude that the group mean vectors are not equal (therapy type is having an effect on the dependent variate). For the interested reader, the Hotelling's  $T^2$  coefficient can be approximated by multiplying the Hotelling's trace value (in the present example, .979) by  $N - g$ , where  $N$  is the total sample size and  $g$  is the number of groups. Thus in the present example we have the following:

$$T^2 = (.979)(55 - 2) = (.979)(53) = 51.887$$

Table 9a.6 also depicts a partial eta-squared value of .495, indicating that nearly 50% of the variance is accounted for by the combined dependent

**Table 9a.7**

<i>Dependent Measure</i>
GAF
CSQ-8

variables.  
determina  
appropriat  
Fidell, 200

Because  
separate u  
inflation, w  
Evaluating  
alpha leve  
inspection  
higher clin  
than does

**Table 9a.5** Test of Intercorrelation of Dependent Variables

	Statistic	Significance
Bartlett's test of sphericity	67.510	.000
Pearson correlation: GAF by CSQ-8	.502	.000

Bochner, A  
Technic  
Monge  
cation  
Everitt, B. S  
and tw  
48–51.  
Hakstian, A  
assump  
1255–1  
Holloway, I  
of the F  
Hotelling, F  
Statistic

**Table 9a.6** Multivariate Test Results

Effect	Multivariate Test	Value	F	Significance	Partial Eta Squared
Therapy type	Roy's largest root	.979	25.445	.001	.495

**Table 9a.7** Univariate Test Results

Dependent Measure	Independent Variable (Group) Means		Pairwise Comparison Difference		
	Cognitive-Behavioral	Psychoanalytic	F	Significance	
GAF	65.23	58.32	6.91	16.96	.00
CSQ-8	28.43	25.08	3.35	45.95	.00

variables. This partial eta-squared is calculated as 1 minus the adjusted determinant of the error matrix and is considered a more conservative and appropriate index of treatment magnitude with MANOVA (Tabachnick & Fidell, 2001b).

Because a multivariate effect is present, we can proceed to examine separate univariate  $F$  tests. These are shown in Table 9a.7. To control for alpha inflation, we perform a Bonferroni adjustment to our alpha level ( $.05/2 = .025$ ). Evaluating the  $F$  tests (one for each dependent variable) against our corrected alpha level indicates that both are statistically significant ( $p < .025$ ). A visual inspection of the means suggests that cognitive-behavioral therapy produces higher clinical outcomes (GAF scores) and greater service satisfaction (CSQ-8) than does psychoanalytic treatment.

## Recommended Readings

- Bochner, A. P., & Fitzpatrick, M. A. (1980). Multivariate analysis of variance: Techniques, models, and applications in communication research. In P. R. Monge & J. N. Cappella (Eds.), *Multivariate techniques in human communication research* (pp. 143–174). New York: Academic Press.
- Everitt, B. S. (1979). A Monte Carlo investigation of the robustness of Hotelling's one and two sample  $T^2$  tests. *Journal of the American Statistical Association*, 74, 48–51.
- Hakstian, A. R., Roed, J. C., & Lind, J. C. (1979). Two-sample  $T^2$  procedure and the assumption of homogeneous covariance matrices. *Psychological Bulletin*, 56, 1255–1263.
- Holloway, L. N., & Dunn, O. J. (1967). The robustness of Hotelling's  $T^2$ . *Journal of the American Statistical Association*, 62, 124–136.
- Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics*, 2, 360–378.

- Hummel, T. J., & Sligo, J. (1971). Empirical comparison of univariate and multivariate analysis of variance procedures. *Psychological Bulletin, 76*, 49–57.
- Lauter, J. (1978). Sample size requirements for the  $T^2$  test of MANOVA (tables for one-way classification). *Biometrical Journal, 20*, 389–406.
- Lix, L. M., & Keselman, H. J. (2004). Multivariate tests of means in independent groups designs: Effects of covariance heterogeneity and nonnormality. *Evaluation & the Health Professions, 27*, 45–69.
- Stevens, J. P. (1980). Power of the multivariate analysis of variance tests. *Psychological Bulletin, 88*, 728–737.
- Zwick, R. (1986). Rank and normal scores alternatives to Hotelling's  $T^2$ . *Multivariate Behavioral Research, 21*, 169–186.

TV  
M

This chap  
(Hotelling  
for a two-  
two levels  
related de

The p  
health da  
ill (SMI) a  
services c  
collected  
(Time 2).  
client po  
always ca

Three  
dent varia  
in the da  
Quality of  
40-item c  
objective  
clients' li  
safety, an  
such as 1  
safety, an

Meyer, L. S., Gamst, G., & Guarino, A. J. (2006). Applied multivariate research: Design and interpretation. Thousand Oaks, CA: Sage.

CHAPTER 10A

## MANOVA

### Comparing Three or More Groups

---

This chapter extends our previous two-group MANOVA discussion to the situation where researchers wish to assess the effects of one independent variable with three or more levels or treatment groups on several dependent measures. This approach is sometimes referred to as a *k*-group MANOVA or *one-way* MANOVA. For example, an investigator might be interested in the effects of type of treatment (cognitive-behavioral, psychoanalytic, brief) on global assessment of functioning (clinical outcome) and consumer service satisfaction.

The Hotelling's  $T^2$  (or two-group MANOVA) that we covered previously is a special case or generalization of the *k*-group or one-way MANOVA. As such, the previous rationale for the use of a two-group MANOVA also remains in place for the *k*-group or one-way MANOVA. Similarly, statistical assumption violation issues (independence, homogeneity of variance-covariance matrices, and normality) remain the same and can be reviewed in Chapters 9A and 9B.

### The Univariate *F* Test

Recall that in a univariate ANOVA, a single quantitative (metric) dependent variable is assessed with one or more categorical (nonmetric) independent variables. The null hypothesis takes the following form:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_i$$

This null hypothesis can be understood to represent the idea that all the population means are equal. Furthermore, in a univariate one-way ANOVA

the total variability of a study (the total sum of squares) is partitioned between two component parts: between-group variance or sum of squares (variability due to treatment effects and error) and within-group variance or sum of squares (variability due to measurement error). Symbolically, these components take the following form:

$$SS_T = SS_b + SS_w$$

These sums of square are then weighted with their respective degrees of freedom to produce variance estimates known as mean squares ( $MS_b$  and  $MS_w$ ). A ratio is formed between these two variance estimates (called the  $F$  ratio where  $F = MS_b / MS_w$ ), which is evaluated with an appropriate degrees of freedom.

### The Multivariate $F$

In the multivariate (MANOVA) situation, several continuous (metric) dependent variables are assessed with one or more categorical independent variable(s). The null hypothesis in the multivariate case takes the following form:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_i$$

This multivariate null hypothesis expresses the idea that all the population mean vectors or sets (symbolized by bold  $\mu$ s) are equal.

In MANOVA, the univariate sums of squares are replaced with sum of squares and cross-product (SSCP) matrices. These SSCP matrices consist of dependent variable sum of squares or variances along the diagonal of the matrix and covariances (cross-products) on the off-diagonal elements that represent the common variance shared between two variables (Weinfurt, 1995). Similar to its univariate cousin ANOVA, which partitions the total variability into sum of squares between and within components, so, too, MANOVA produces a similar matrix analogue bifurcation. MANOVA (through its SPSS-GLM realization) produces a total SSCP matrix ( $T$ ) that can be separated into a between-group SSCP matrix ( $B$ ) and a within-group SSCP matrix ( $W$ ). Symbolically, these matrix components form the following multivariate analogue to the univariate sum of squares partitioning:

$$T = B + W$$

Total SSCP Matrix = Between SSCP Matrix + Within SSCP Matrix

As we r  
conducted  
converted  
generalized  
sum of squ  
ization of h  
from the gi  
group sum  
treatment c  
generalizat  
within-grou  
generalizati  
how the ca  
from the de

As note  
variate test  
Roy's large  
ture is Wilk  
tice, if the  
the depend  
(the treatm  
ual or error  
inverse crit  
(Stevens, 2  
Wilks's lam  
ate (Rao's)

With th  
multivariate  
the two-gr  
or not signi  
in the litera  
are plagued

### Following

A statis  
variable is a  
Thus, we c  
process is  
we indicate  
ANOVAAs fo

As we noted in Chapters 9A and 9B, through matrix algebraic manipulations conducted within the SPSS-**GLM** program, these matrices of coefficients are converted into single values called *determinants*. Determinants reflect the generalized variance of each matrix. Thus, **T** (the determinant for the total sum of squares and cross-products matrix) reflects the multivariate generalization of how the cases in each independent variable level or group deviate from the grand mean of each dependent variable. Similarly, **B** (the between-group sum of squares and cross-product matrix) reflects the differential treatment effects on the set of dependent variables and is the multivariate generalization of the univariate between-group sum of squares. Last, **W** (the within-group sum of squares and cross-products matrix) is the multivariate generalization of the univariate within-group sum of squares and represents how the cases in each level or group of the independent variable deviate from the dependent variable means (Stevens, 2002).

As noted in Chapters 9A and 9B, there are four commonly used multivariate test statistics: Pillai's trace, Wilks's lambda, Hotelling's trace, and Roy's largest root. The most prominent of these tests in the research literature is Wilks's lambda ( $\Lambda$ ), which is basically a ratio of **W** to  $(\mathbf{B} + \mathbf{W})$ . In practice, if the independent variable has a statistically significant effect on the dependent variables—that is, if treatment effects are present—then **B** (the treatment variance-covariance) will be relatively large and **W** (the residual or error variance-covariance) will be small. Because Wilks's lambda is an inverse criterion, smaller values provide more evidence of treatment effects (Stevens, 2002). To evaluate any of the multivariate test statistics (including Wilks's lambda) SPSS translates the multivariate test value into a multivariate (Rao's) *F* statistic, which can be evaluated much as any other *F* value.

With three or more levels (groups) of the independent variable, these multivariate *F* values tend to differ slightly (remember, they are all the same in the two-group situation), but all tend to yield the same statistically significant or not significant decision. Although Wilks's lambda is most typically reported in the literature, Pillai's trace should be reported if the dependent variables are plagued by significant heterogeneity of variance-covariance matrices.

### **Following a Significant Multivariate Effect**

A statistically significant multivariate effect tells us that the independent variable is associated with differences between the vectors or sets of means. Thus, we can presume that treatment effects exist. The next step in this process is to discover which specific dependent variables are affected. As we indicated in Chapter 9A, we recommend the use of separate univariate ANOVAs for each dependent measure with a Bonferroni adjustment to the

operational alpha level (.05 divided by the number of dependent variables) to reduce the possibility of Type I error. Each statistically significant univariate  $F$  statistic can then be further evaluated with a post hoc or multiple comparison test that assesses every pairwise combination of means on each dependent measure. These post hoc tests, from the Type I error-liberal least significant difference (LSD) test to its conservative counterpart, the Scheffé test, will be reviewed in some detail in Chapter 10B.

### Hypothetical Three-Group Example

Assume that we are interested in the effects of therapy type (this time with three levels: cognitive-behavioral, psychoanalytic, brief) on clinical outcome (GAF) and service satisfaction (CSQ-8) scores. The null hypothesis is that the set or vector of means (on the dependent variate) for each group (or level of the independent variable) is equal to the others. Condensed summaries of results can be seen in Tables 10a.1 through 10a.7.

In Table 10a.1, we note the addition of a third level or group (brief psychotherapy) to the independent variable example that we used in the previous chapter. Sample size is small but acceptable for purposes of illustration. An inspection of the dependent variable means for each group suggests that the psychoanalytic treatment group generated somewhat lower clinical outcome and service satisfaction scores than did clients receiving cognitive-behavioral or brief psychotherapy.

Pearson  $r$  correlations were computed between the two dependent measures at each level (group) of the independent variable. A moderate positive correlation was observed for the cognitive-behavioral ( $r = .426$ ) and brief ( $r = .458$ ) psychotherapy groups, and a low positive correlation ( $r = .086$ ) was found for the psychoanalytic group. The differences between these correlations may suggest variability in the dependent variate covariance and will be addressed with the Box's  $M$  test.

Table 10a.2 examines normality and diagnostics for each dependent variable across levels of the independent variable. Nearly all the normality tests are not significant ( $p < .05$ ), indicating no normality violations, with the exception of the GAF scores for the brief psychotherapy group. However, the skewness and kurtosis values for this latter group were deemed adequate; thus, any normality violations were considered minor, allowing us to proceed with the analysis.

### Normality Tests and Diagnostics, by Therapy Type

Table 10a.3 depicts a statistically nonsignificant Box's  $M$  test, indicating equality of variance-covariance matrices. Separate Levene's tests for each

**Table 10a**

Group 1: G

Caseid

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

GAF

CSQ-8

a.  $n = 30, r = .426$

b.  $n = 25, r = .458$

c.  $n = 25, r = .086$

**Table 10a.1** Client Global Assessment of Functioning (GAF) and Service Satisfaction (CSQ-8), by Therapy Type (Hypothetical Data for Three-Group MANOVA)

Therapy Type								
Group 1: Cognitive-Behavioral <sup>a</sup>			Group 2: Psychoanalytic <sup>b</sup>			Group 3: Brief <sup>c</sup>		
Caseid	GAF	CSQ-8	Caseid	GAF	CSQ-8	Caseid	GAF	CSQ-8
1	65	30.3	31	50	24.3	56	69	30.1
2	69	27.1	32	55	22.1	57	55	29.0
3	73	31.0	33	61	25.6	58	70	31.0
4	60	29.4	34	66	25.0	59	68	28.0
5	58	27.4	35	58	24.0	60	72	29.9
6	70	30.1	36	50	26.0	61	68	29.0
7	71	30.9	37	45	27.0	62	65	28.2
8	55	27.0	38	53	26.5	63	67	28.0
9	65	28.5	39	59	25.5	64	66	27.1
10	60	29.1	40	61	25.0	65	56	30.1
11	66	29.2	41	70	29.0	66	72	27.9
12	60	29.2	42	65	26.2	67	70	27.0
13	70	30.3	43	55	24.8	68	61	29.0
14	70	31.0	44	50	24.0	69	65	27.0
15	65	26.5	45	62	23.2	70	55	29.0
16	65	27.0	46	66	26.0	71	69	30.0
17	55	26.5	47	69	27.0	72	66	28.3
18	72	30.0	48	50	25.1	73	70	26.5
19	70	31.0	49	58	28.1	74	70	30.0
20	65	28.7	50	55	25.5	75	66	29.1
21	60	28.0	51	60	23.2	76	68	27.9
22	55	27.0	52	60	23.0	77	69	30.0
23	70	30.1	53	65	22.0	78	70	30.0
24	75	29.2	54	65	24.0	79	65	26.0
25	65	25.0	55	50	25.0	80	66	29.0
26	62	24.3						
27	60	30.3						
28	68	26.7						
29	70	28.0						
30	68	25.5						
			<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	
GAF	65.23	5.62	58.32	6.84		65.92	4.92	
CSQ-8	28.43	1.90	25.08	1.73		28.77	1.33	

a.  $n = 30$ ,  $r = .426$ ,  $p < .01$ .b.  $n = 25$ ,  $r = .086$ ,  $p > .05$ .c.  $n = 25$ ,  $r = .458$ ,  $p > .05$ .

**Table 10a.2** Client Global Assessment of Functioning (GAF) and Service Satisfaction (CSQ-8) Normality Tests and Diagnostics, by Therapy Type

		Group 1: Cognitive- Behavioral		Group 2: Psychoanalytic		Group 3: Brief	
		Statistic	Sig.	Statistic	Sig.	Statistic	Sig.
<b>Normality tests</b>							
GAF	Kolmogorov-Smirnov	.150	.083	.128	.200	.226	.002
	Shapiro-Wilk	.939	.088	.958	.372	.858	.003
	Skewness	-.353		-.112		-1.066	
	Kurtosis	-.832		-.928		.602	
CSQ-8	Kolmogorov-Smirnov	.129	.200	.081	.200	.169	.063
	Shapiro-Wilk	.946	.131	.982	.918	.935	.115
	Skewness	-.362		.220		-.449	
	Kurtosis	-.782		.033		-.699	

**Table 10a.3** Tests of Equality of Covariance Matrices and Equality of Error Variances

	Overall		GAF		CSQ-8	
	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.
Box's <i>M</i>	7.949	.265				
Levene's test			2.34	.103	2.11	.128

dependent variable were also not statistically significant, indicating equal variances for each dependent measure across the levels of therapy type.

A statistically significant Bartlett's test of sphericity ( $p < .001$ ) as shown in Table 10a.4 indicates sufficient correlation between the dependent variables to proceed with the analysis. A moderate positive correlation ( $r = .502$ ) was observed overall between the two dependent measures.

Table 10a.5 presents the multivariate test results, all of which indicate a statistically significant effect of therapy type on the dependent variate. Because equality of variance-covariance matrices was evidenced with the nonsignificant Box's *M* test, most researchers would report the Wilks's lambda *F* value of 17.088,  $p < .001$ , with a partial eta-squared value of .310. This result tells us we can reject the null hypothesis that the group mean

**Table 10a****Table 10a**

Therapy ty

vectors ar  
dependen  
variate varBecau  
can proce  
ments to c  
that both  
result indi  
multivariatBeau  
independe  
multiple co  
results of tTable  
means for  
measure c  
is compute  
– 58.32), v  
are evaluat  
discussed  
that cognit  
nificantly h  
psychoanal

**Table 10a.4** Test of Intercorrelation of Dependent Variables

	Statistic	Significance
Bartlett's test of sphericity	100.010	.000
Pearson correlation:		
GAF by CSQ-8	.502	.000

**Table 10a.5** Multivariate Test Results

	Multivariate Test	Value	F	Sig.	Partial Eta Squared
Therapy type	Pillai's trace	.524	13.674	.000	.262
	Wilks's lambda	.476	17.088	.000	.310
	Hotelling's trace	1.102	20.654	.000	.355
	Roy's largest root	1.102	42.411	.000	.524

vectors are equal and instead conclude that therapy type influences the dependent variate. We can also note that about 31% of the total dependent variate variance is accounted for by therapy type.

Because we observed a statistically significant multivariate effect, we can proceed to examine separate univariate *F* tests with Bonferroni adjustments to our operational alpha level ( $.05/2 = .025$ ). We note in Table 10a.6 that both univariate *F* statistics are statistically significant ( $p < .025$ ). This result indicates that both dependent variables contribute to the significant multivariate effect.

Because statistically significant univariate *F*s were observed and the independent variable contained more than two levels, a Scheffé post hoc multiple comparison test was computed for each dependent measure. The results of the post hoc comparisons are shown in Table 10a.7.

Table 10a.7 summarizes pairwise comparisons between each pair of means for each dependent variable. For example, the GAF dependent measure comparison for cognitive-behavioral and psychoanalytic groups is computed by subtracting one group mean from the other (e.g., 65.23 – 58.32), which produces a difference score of 6.9133. Difference scores are evaluated by means of special multiple comparisons formulas (briefly discussed in Chapter 10B) at an alpha level of .05. These results suggest that cognitive-behavioral and brief psychotherapy produce statistically significantly higher clinical outcome and service satisfaction scores than does psychoanalytic psychotherapy.

**Table 10a.6** Univariate Test Results

	Dependent Measures	
	GAF	CSQ-8
Independent variable (group) means		
Cognitive-behavioral	65.23	28.43
Psychoanalytic	58.32	25.08
Brief	65.92	28.77
F statistic	13.29	37.35
Significance	.000	.000
Partial eta squared	.257	.492

Shaffer, J. P.  
*Educational*  
 Stevens, J. J.  
 MANOV  
 Thomas, D.  
*Multivariate*  
 Weinfurt, K.  
 (Eds.),  
 Washington  
 Wilkinson, J.  
 variance

**Table 10a.7** Multiple Comparison Post Hoc Tests (Scheffé) for GAF and Service Satisfaction

Dependent Variable	(I) Therapy Type	(J) Therapy Type	Mean Difference (I-J)
GAF	Cog-beh	Psychoanalytic	6.9133*
	Cog-beh	Brief	-.6867
	Psychoanalytic	Brief	-7.6000*
Service satisfaction	Cog-beh	Psychoanalytic	3.3460*
	Cog-beh	Brief	-.3380
	Psychoanalytic	Brief	-3.6840*

Note: Cog-beh = Cognitive-behavioral.

\* $p < .05$ .

## Recommended Readings

- Huberty, C. J. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin, 105*, 302–308.
- O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97*, 316–333.
- Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association, 69*, 894–908.
- Olson, C. L. (1976). On choosing a test statistic in MANOVA. *Psychological Bulletin, 83*, 579–586.
- Olson, C. L. (1979). Practical considerations in choosing a MANOVA test statistic: A rejoinder to Stevens. *Psychological Bulletin, 86*, 1350–1352.

- Shaffer, J. P., & Gillo, M. W. (1974). A multivariate extension of the correlation ratio. *Educational and Psychological Measurement, 34*, 521-524.
- Stevens, J. P. (1972). Four methods of analyzing between variation for the k group MANOVA problem. *Multivariate Behavioral Research, 7*, 499-522.
- Thomas, D. (1992). Interpreting discriminant functions: A data analytic approach. *Multivariate Behavioral Research, 27*, 335-362.
- Weinfurt, K. P. (1995). Multivariate analysis of variance. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 245-276). Washington, DC: American Psychological Association.
- Wilkinson, L. (1975). Response variable hypotheses in the multivariate analysis of variance. *Psychological Bulletin, 82*, 408-412.

sures

CSQ-8

28.43

25.08

28.77

37.35

.000

.492

I Service

rence (I,J)

3\*

7

0\*

0\*

30

10\*

e analyses.

ed measures

e analysis of

*Bulletin, 83,*

stic: A rejoinder

Meyer, L. S., Gamst, G., & Guarino, A. J. (2006). Applied multivariate research: Design and interpretation. Thousand Oaks, CA: Sage.

CHAPTER 11A

## MANOVA

### Two-Way Factorial

---

This chapter extends our previous discussions of two-group MANOVA (Chapters 9A and 9B) and k-group MANOVA (Chapters 10A and 10B) to the two-way (two independent variables) MANOVA situation, sometimes referred to as a factorial or two-way MANOVA, where the effects of two independent variables are examined simultaneously on two or more conceptually related dependent variables. For example, an investigator might be interested in the effects of gender (male, female) and type of treatment (psychoanalytic, cognitive-behavioral, brief) on global assessment of functioning (clinical outcome) and consumer service satisfaction.

The rationale we outlined previously (Chapters 9A and 10A) for using MANOVA is also appropriate when two (or more) independent variables are considered. Likewise, the issues relating to statistical assumption violations of independence, homogeneity of variance-covariance matrices, and normality are comparable and can be reviewed in Chapters 9A and 9B.

### The Univariate and Multivariate Factorial Design

The previous MANOVA chapters shared an important commonality: They each addressed the multivariate analysis of a single independent variable. In the univariate statistical domain, where we examine one dependent variable at a time, these single-factor analyses can provide very useful information. This is also true in the multivariate domain as well, where two or more dependent measures are profitably assessed by means of a single (categorical) independent variable.

Challenges arise when researchers work with multiple independent variables (in the univariate or multivariate context) in a study. If they opt to analyze one independent variable at a time, they will drive up Type I (false positive) error rates. Furthermore, single-factor independent variable assessments (either univariate or multivariate) do not allow researchers to determine how independent variables jointly affect the dependent measure(s).

## **Advantages of Univariate and Multivariate Factorial Designs**

Keppel et al. (1992) note several distinct advantages that univariate factorial designs have over single-factor approaches, although these advantages are also recognized by most researchers in the field. Our discussion is based on their thoughts, but we have extended the arguments to subsume the multivariate domain as well.

### **Simultaneous Manipulation of Independent Variables**

A univariate factorial design is defined as the joint or simultaneous manipulation of two or more independent variables to determine their unique and joint effect on a single dependent variable. Likewise, a multivariate factorial design examines the unique and joint effects of two or more independent variables on two or more dependent variables—both collectively (the dependent variate) and separately (as in the univariate case). Accordingly, factorial designs provide researchers with a richer context within which they may explore the phenomena under study. Adding an additional independent variable to the use of a single one can potentially increase the *ecological validity* or real-world meaningfulness of the study. This is true for at least two general reasons: First, most phenomena we wish to study in the social and behavioral sciences are observed in the presence of a host of conditions. Up to a point, the more of these conditions we can treat or manipulate as independent variables in a study, the more we are able to reproduce the real-world conditions and thus explain the operations of these variables. Second, as outlined in our discussion of multiple regression in an earlier chapter, we tend to believe that most of the phenomena we study are multiply determined. For the same reason that we would include multiple potential predictors in a regression study, we would want to include more than one independent variable in an experimental design.

### **Main Effect and Interactions**

A second advantage of univariate and multivariate factorial designs is found by examining the concepts of main effects and interaction effects.

In the  
of one  
other i  
main e  
indepe  
dent va  
might b  
the ma  
analog  
multiva

But  
present  
effect, v  
ables. Ir  
ables co  
occurs v  
second  
variate c  
to influe  
of factor  
combine

To exten  
in the ei  
and clien  
satisfacti

Table  
subjects  
type (ps  
of two le  
data for r  
Sample s  
size of 2  
dependen  
combinat  
able mea  
cognitive  
were com

In the univariate situation, main effects reflect the separate treatment effects of one independent variable averaged (or collapsed) over the levels of the other independent variable(s) on a single dependent measure. Similarly, main effects in the multivariate context refer to the separate effects of one independent variable collapsed over the levels of the other independent variable(s) on a set or vector of dependent variable means. Each independent variable is considered to be a single main effect. For example, Factor A might be type of treatment and Factor B could be gender of client. Assessing the main effects of each independent variable (Factor A and Factor B) is analogous to conducting single-factor analyses in the univariate and multivariate contexts.

But factorial designs also provide us with a new ingredient that is not present in the single-factor case. This new component is the interaction effect, which assesses the joint influence of two or more independent variables. In the univariate two-factor case, an interaction depicts how the variables combine to influence the dependent measure. This combinatory effect occurs when one independent variable changes at the different levels of the second independent variable. Similarly, an interaction effect in the multivariate context depicts how two (or more) independent variables combine to influence the composite dependent variate. Thus, a second advantage of factorial designs lies in their ability to show how independent variables combine or interact to influence the dependent measure(s).

### Numerical Hypothetical Two-Way Factorial Example

To extend our previous hypothetical example, assume that we are interested in the effects of therapy type (psychoanalytic, cognitive-behavioral, brief) and client gender (male, female) on clinical outcome (GAF) and service satisfaction (CSQ-8) scores.

Table 11a.1 depicts the raw data matrix for a  $3 \times 2$  (read 3 by 2) between-subjects factorial design with Factor A consisting of three levels of therapy type (psychoanalytic, cognitive-behavioral, brief) and Factor B consisting of two levels of gender (male, female). The first part of the table shows the data for males, and the second part of the table shows the data for females. Sample size is small but deemed adequate by meeting the minimum cell size of 20 recommended by Hair et al. (1998). An inspection of the dependent variable means (GAF and CSQ-8) for each of the six treatment combinations suggests a consistent ascending ordering of dependent variable means (for both men and women) for clients receiving psychoanalytic, cognitive-behavioral, and brief psychotherapy services. Pearson  $r$  correlations were computed between the two dependent measures for each of the six

**Table 11a.1** Client Global Assessment of Functioning (GAF) and Service Satisfaction (CSQ-8), by Therapy Type and Gender (Hypothetical Data for  $3 \times 2$  Factorial MANOVA)

Male ( $b_1$ )									Case ID	Case ID
Psychoanalytic ( $a_1$ )			Cognitive-Behavioral ( $a_2$ )			Brief ( $a_3$ )				
Case ID	GAF	CSQ-8	Case ID	GAF	CSQ-8	Case ID	GAF	CSQ-8		
1	33	22.1	21	41	25.6	41	65	30.1	61	61
2	35	23.3	22	45	25.8	42	66	29.8	62	62
3	41	21.5	23	55	26.9	43	69	27.1	63	63
4	38	22.4	24	59	27.1	44	50	26.4	64	64
5	45	22.9	25	58	27.0	45	55	27.0	65	65
6	46	24.9	26	50	24.0	46	64	30.3	66	66
7	37	20.1	27	57	25.0	47	66	30.4	67	67
8	37	25.6	28	55	24.9	48	68	29.0	68	68
9	40	21.7	29	55	26.7	49	59	28.7	69	69
10	49	25.1	30	59	27.1	50	51	27.1	70	70
11	41	22.1	31	52	22.7	51	55	26.1	71	71
12	45	23.7	32	53	23.1	52	57	28.1	72	72
13	45	25.0	33	55	25.0	53	62	30.7	73	73
14	40	25.0	34	56	26.1	54	60	30.1	74	74
15	36	21.1	35	55	26.2	55	63	29.8	75	75
16	35	22.2	36	58	28.1	56	64	28.9	76	76
17	40	23.4	37	50	23.4	57	50	26.1	77	77
18	45	24.1	38	51	24.5	58	66	28.1	78	78
19	39	20.1	39	55	25.5	59	68	29.9	79	79
20	35	21.0	40	55	26.6	60	69	30.0	80	80

treatment combinations. As can be seen in Table 11a.1, moderate positive correlations (ranging between .368 and .691) were observed between the dependent measures across conditions.

Table 11a.2 depicts normality tests and diagnostics for each dependent variable across the levels of the independent variables. Just over half of the normality tests were found to be statistically significant ( $p < .05$ ), indicating a dependent variable distribution departure from normality for some of the treatment combinations. These normality violations may be due in part to the skewness and kurtosis observed among the GAF scores in the

cognitive scores were presented with the aim of indicating for each dependent variable ( $p < .047$ ) significant differences between the two groups. These results are shown in Table 1.

	Female ( $b_2$ )								
	Psychoanalytic ( $a_1$ )			Cognitive-Behavioral ( $a_2$ )			Brief ( $a_3$ )		
	Case ID	GAF	CSQ-8	Case ID	GAF	CSQ-8	Case ID	GAF	CSQ-8
CSQ-8	61	33	21.1	81	51	25.5	101	66	30.3
30.1	62	34	21.7	82	48	24.6	102	64	30.5
29.8	63	35	22.6	83	47	26.7	103	65	31.1
27.1	64	42	23.0	84	55	28.1	104	51	29.5
26.4	65	47	22.1	85	58	25.4	105	55	28.1
27.0	66	46	23.7	86	57	26.1	106	58	27.1
30.3	67	39	22.4	87	57	27.2	107	60	30.3
30.4	68	37	22.4	88	55	27.0	108	61	30.8
29.0	69	41	23.1	89	55	28.1	109	62	31.0
28.7	70	40	24.4	90	44	25.0	110	65	30.0
27.1	71	45	22.0	91	59	26.1	111	66	31.7
26.1	72	45	20.0	92	57	27.3	112	55	29.6
28.1	73	44	23.3	93	58	28.9	113	55	29.0
30.7	74	38	23.1	94	58	28.9	114	50	28.1
30.1	75	39	22.8	95	56	26.5	115	52	28.9
29.8	76	40	24.5	96	57	26.0	116	66	29.0
28.9	77	45	22.1	97	58	25.5	117	69	30.1
26.1	78	42	23.6	98	55	25.0	118	68	30.3
28.1	79	40	23.5	99	56	26.1	119	65	29.1
29.9	80	37	21.7	100	50	27.0	120	62	28.0
30.0	$M = 40.45$		$SD = 4.07$	$n = 20$	$r = .676$	$p > .05$	$M = 54.55$		$SD = 4.24$
28.69	$M = 22.66$		$SD = 1.90$	$n = 20$	$r = .368$	$p > .05$	$M = 26.55$		$SD = 1.26$
1.57	$n = 20$		$p < .05$	$n = 20$		$p < .05$	$n = 20$		$p = .527$
$p < .001$									

cognitive-behavioral conditions. Normally, transformations of these GAF scores would probably be recommended here. However, to expedite our presentation, we will deem these normality tests adequate and proceed with the analysis.

Table 11a.3 depicts a statistically nonsignificant Box's  $M$  test ( $p < .499$ ), indicating equality of variance-covariance matrices. Separate Levene's tests for each dependent variable found a marginally statistically significant ( $p < .047$ ) effect for the GAF dependent measure and a statistically non-significant ( $p > .05$ ) test for the service satisfaction dependent measure. These results indicate that relatively equal variances are in place for each dependent variable.

**Table 11a.2** Client Global Assessment of Functioning (GAF) and Service Satisfaction (CSQ-8) Normality Tests and Diagnostics, by Therapy Type and Gender

Normality Test	Therapy Type					
	Psychoanalytic		Cognitive-Behavioral		Brief	
	Statistic	Significance	Statistic	Significance	Statistic	Significance
GAF						
Kolmogorov-Smirnov	.143	.038	.279	.000	.161	.010
Shapiro-Wilk	.958	.143	.856	.000	.906	.003
CSQ-8						
Kolmogorov-Smirnov	.075	.200	.070	.200	.146	.032
Shapiro-Wilk	.978	.619	.980	.703	.941	.036
GAF						
Skewness		.104		−1.32		−.535
Kurtosis		−.930		1.32		−.998
CSQ-8						
Skewness		.025		−.215		−.582
Kurtosis		−.437		.034		−.526
Gender						
Normality Test	Male		Female			
	Statistic	Significance	Statistic	Significance		
GAF						
Kolmogorov-Smirnov	.126	.019	.157	.001		
Shapiro-Wilk	.956	.032	.958	.036		
CSQ-8						
Kolmogorov-Smirnov	.074	.200	.101	.200		
Shapiro-Wilk	.970	.149	.958	.040		
GAF						
Skewness		−.107		−.190		
Kurtosis		−1.03		−1.03		
CSQ-8						
Skewness		−.038		−.087		
Kurtosis		−.822		−1.16		

**Table**

Box's Leven

T  
( $p <$ ).  
to pr  
meas  
two c  
be ec  
sis of  
posit  
proceT  
variab  
test v  
result  
signif  
signif**Table**Bartle  
Pearso

GAI

**Table**Thera  
Gende

Thera

Typ

**Table 11a.3** Tests of Equality of Covariance Matrices and Equality of Error Variances

	Overall		GAF		CSQ-8	
	Statistic	Significance	Statistic	Significance	Statistic	Significance
Box's <i>M</i>	15.026	.499				
Levene's Test			2.33	.047	1.86	.106

The Bartlett's test of sphericity in Table 11a.4 was statistically significant ( $p < .001$ ), indicating sufficient correlation between the dependent variables to proceed with the analysis. The correlation between the two dependent measures was a high positive one ( $r = .861$ ), which may indicate that these two dependent variables measure a common construct and may profitably be combined into a single measure and examined with a univariate analysis of variance procedure. Again, for present purposes, we judge this high positive correlation between our dependent variables to be acceptable and proceed with the multivariate analysis.

Table 11a.5 presents the multivariate test results for each independent variable separately (main effects) and their interaction. Because the Box's *M* test was not statistically significant, we report only the Wilks's lambda test results. From Table 11a.5, we can see that both main effects were statistically significant. We note in passing that the main effect of gender was marginally significant ( $p < .057$ ) and accounted for less than 5% of the total variance.

**Table 11a.4** Test of Intercorrelation of Dependent Variables

	Statistic	Significance
Bartlett's test of sphericity	177.900	.000
Pearson correlation:		
GAF by CSQ-8	.861	.000

**Table 11a.5** Multivariate Test Results

	Multivariate Test	Value	F	Significance	Partial Eta Squared
Therapy type	Wilks's lambda	.164	82.89	.000	.595
Gender	Wilks's lambda	.951	2.94	.057	.049
Therapy Type × Gender	Wilks's lambda	.943	1.68	.155	.029

The statistically significant main effect of therapy type ( $p < .000$ ) accounted for nearly 60% of the total variance. The interaction of Therapy Type  $\times$  Gender was not statistically significant. These two statistically significant main effects together with the lack of a significant interaction allow us to reject the null hypotheses that the therapy type and gender group mean vectors are equal and, instead, conclude that both therapy type and gender uniquely influence the dependent variate.

Because we observed statistically significant multivariate main effects of therapy type and gender, we can proceed to examine the two separate univariate  $F$  tests, with a Bonferroni adjustment to our alpha level, giving us an adjusted alpha level of .025 (.05 divided by the two univariate effects = .025). We note in Table 11a.6 that for the main effect of therapy type, both univariate  $F$  statistics were statistically significant ( $p < .000$ ), with each accounting for more than 75% of the total variance. However, after Bonferroni adjustment, the main effect of gender was not statistically significant ( $p > .025$ ) for either

**Table 11a.6** Univariate Test Results

	<i>Dependent Measure</i>	
	GAF	CSQ-8
Independent variable (group) means		
Therapy type		
Psychoanalytic	40.28	22.76
Cognitive-behavioral	54.12	26.06
Brief	61.05	29.16
<i>F</i> statistic	177.65	208.13
Significance	.000	.000
Partial eta squared	.757	.785
Independent variable (group) means		
Gender		
Male	51.72	25.71
Female	51.92	26.28
<i>F</i> statistic	.048	4.988
Significance	.828	.027
Partial eta squared	.000	.042
Therapy Type $\times$ Gender		
<i>F</i> statistic	.215	2.334
Significance	.807	.102
Partial eta squared	.004	.039

**Table**

Depen

GAF

Service

\* $p < .05$

of the  
(GAF)  
The ur  
counte

Be  
effect  
type h  
signific  
depend  
parisor

Tat  
on each  
is signi  
the bri  
is com  
– 40.28  
are ev  
were b  
statistic  
outcon  
psycho  
interve

Most o  
nature.

**Table 11a.7** Multiple Comparison Post Hoc Tests (Tukey) for GAF and Service Satisfaction, by Therapy Type

<i>Dependent Variable</i>	(I) <i>Therapy Type</i>	(J) <i>Therapy Type</i>	<i>Mean Difference (I-J)</i>
GAF	Brief	Psychoanalytic	20.77*
	Brief	Cognitive-behavioral	6.93*
	Cognitive-behavioral	Psychoanalytic	13.85*
Service satisfaction	Brief	Psychoanalytic	6.40*
	Brief	Cognitive-behavioral	3.10*
	Cognitive-behavioral	Psychoanalytic	3.30*

\* $p < .05$ .

of the dependent measures, indicating comparable levels of clinical outcome (GAF) and service satisfaction (CSQ-8) among men and women respondents. The univariate interaction analyses can be ignored because their multivariate counterpart was not statistically significant.

Because we observed statistically significant univariate  $F$ s for the main effect of therapy type for both dependent variables, and because therapy type has more than two levels, we can proceed with Tukey HSD (honestly significant difference) post hoc multiple comparison tests to compare each dependent variable's group means. Table 11a.7 summarizes pairwise comparisons between each pair of means for each dependent variable.

Table 11a.7 shows the difference between pairs of means for each group on each of the dependent measures. An asterisk indicates that the difference is significant. For example, the GAF dependent measure comparison with the brief and psychoanalytic groups (see Table 11a.6 for respective means) is computed by subtracting one group mean from the other (e.g., 61.05 – 40.28), which produces a difference score of 20.77. These difference scores are evaluated by means of special multiple comparisons formulas that were briefly reviewed in Chapter 10B. With all mean differences showing as statistically significant, the results suggest that the highest levels of clinical outcome and service satisfaction were achieved by clients receiving brief psychotherapy followed by cognitive-behavioral and then psychoanalytic intervention.

### The Time Dimension in Multivariate Data Analysis

Most of the statistical designs covered in this book are cross-sectional in nature. That is, the statistical analysis is employed on a set of observations

(cases) that represents a single point in time. Cross-sectional data sets run the gamut of a self-report questionnaire completed by a random sample of college freshmen on their first day of class, school districts reporting averages of standardized test scores, or exit interview data gleaned from interviews of a sample of voters exiting their precinct polling station. Such designs capture behavior, attitudes, opinions, and feelings at one moment in time, much like a photographic snapshot.

The construct of time has historically been studied by means of univariate and multivariate analysis of variance procedures. In the univariate situation, participants or cases are measured more than once on a dependent variable. For example, clients could be given a mental health evaluation at initial intake, at 6 months into their treatment, and again at the end of their first treatment year. These three longitudinal snapshots of client functioning (sometimes referred to as trend analysis) can be used to track treatment progress over time. To evaluate change, we would use a within-subjects design in which the three periodic assessments would constitute the repeated measure. Alternatively, we could incorporate a second dependent variable (e.g., client satisfaction) into this same experimental design scenario and use MANOVA to analyze the effects of these multiple dependent measures over time (e.g., Keppel & Wickens, 2004; Tabachnick & Fidell, 2001a).

A variety of longitudinal data analysis designs are gaining considerable momentum in the multivariate literature (Diggle, Heagerty, Liang, & Zeger, 2002; Hand & Crowder, 1996) that represent more complex and sophisticated approaches to the study of time-related effects. These methods include panel data analysis, cohort analysis, hierarchical linear models (HLM), and survival analysis. Although each technique is certainly unique, they also share some fundamental commonalities, such as focusing on responses or behavior over time and using methods related to multiple regression analysis (see Allison, 1990). We will briefly note each in passing.

### Panel Data Analysis

Panel studies (or linear panel analyses) are based on repeatedly measuring the same set of participants over time. For example, a metropolitan newspaper might locate a small group of undecided registered Republican and Democrat voters ( $N = 20$ ) via a telephone interview 1 year prior to a presidential election. These individuals become the panel, and their political attitudes and preferences can be assessed on a monthly basis, right up to the November election. For overviews of this method, see Cronbach and Furby (1970), Finkel (1995), Kessler and Greenberg (1981), and Menard (1991).

### Cohort I

Cohort within a d are selecte to .consisti events and of system

- Ind mai cha coh pro as r
- Ind boc thei soci

A nun cohort dat time perio be account coding an Mason & V

### Hierarch

Hierar data (level for examp programs, assumption within a cl context. Th mixed mod dependent Kenny, Bo overviews)

## Cohort Data Analysis

Cohort analysis compares one or more groups of individuals, usually within a defined age range, at different points in time. Different participants are selected from the same cohort at each test point. Cohorts are assumed to consist of individuals who have experienced similar significant life events and personal contexts; in some sense, they have entered some sort of system at the same time. Here are two examples:

- Individuals in a large corporation are selected to start a 6-month management training program. Although they may differ on many characteristics (age, geographic region of origin) they make up a cohort based on training. We could study their success in the training program and follow up every year to determine their effectiveness as managers.
- Individuals born between 1946 and 1964 are known as the “baby boomer” generation and make up a cohort based on age. A sample of these persons could be surveyed at 5- or 10-year intervals to examine their attitudes about world political events or views about domestic social policy.

A number of methodological issues affect the proper assessment of cohort data, including participant age effects, cohort status effects, and the time period, all of which influence the variability within the study and must be accounted for. These issues can be addressed in part by means of dummy coding and interaction analysis (Glenn, 2004; Mason & Fineberg, 1985; Mason & Wolfinger, 2001; Rodgers, 1982).

## Hierarchical Linear Models (HLM)

Hierarchical linear models were originally developed to study nested data (levels of variables are specific to one level of another variable)—for example, mental health clients nested within specific therapeutic programs, who are in turn nested within mental health agencies. One assumption underlying the analysis of this sort of structure is that clients within a cluster will share certain commonalities because of their shared context. The HLM approach is also referred to as multilevel models, linear mixed models, random coefficient models, or random effects models. The dependent variables in HLM can be either continuous or categorical (see Kenny, Bolger, & Kashy, 2002; Raudenbush & Bryk, 2002, for useful overviews).

HLM has also been extended to the analysis of longitudinal data where the research goal is to examine change and the factors that affect both intra- and inter-individual change (e.g., Hox, 2000; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997; Raudenbush, 2001; Singer & Willett, 2003; Weinfurt, 2000). At least three major approaches to HLM can be identified (see Diggle et al., 2002; Singer & Willett, 2003). One approach is called *marginal analysis* where the investigator builds a model that focuses on the dependent variable average and how this mean changes over time. A second approach is to develop transition models that focus on how the dependent variable is a function of or depends on previous values of the dependent measure and other variables. A third HLM longitudinal approach is to construct a random effects model where the focus becomes how regression coefficients vary among participants. Several recent and readable applications of HLM to longitudinal data can be found in O'Connell and McCoach (2004) and Lane and Zelinski (2003).

### **Survival Analysis**

Survival analysis, also called event history analysis, encompasses a number of methods (e.g., life table analysis, Kaplan-Meier method, Cox regression model) that predict the survival time between two events for one or more groups of participants or cases (see, e.g., Hosmer & Lemeshow, 2002; Lee & Wang, 2003; Singer & Willett, 1991; Wright, 2000). These methods were first developed in the medical, epidemiological, and biological fields to examine the survival times of patients undergoing various types of medical treatment and hence suggested the name of the procedure.

Survival analysis has been successfully extended to other fields, including the social and behavioral sciences as well as business and marketing. For example, Gamst (1985) used this approach to determine the length of time individuals would continue to subscribe to a newspaper under different financial incentive scenarios. But survival analysis can be applied to a host of interesting and very important topics (e.g., how long adolescents will remain in high school before dropping out, how long patients are likely to follow a medication regime before putting their medicine aside). Generally, we are interested in the length of time that cases in a target group remain "active" or "alive" (in either a literal or figurative sense).

The challenge in survival analysis is that the original number of participants can be quite variable between two points in time. Participants may continue to survive, or they may quit, drop out, or become lost to follow-up. These latter situations are called censored events and must be taken into account to produce accurate survival curve estimates. Several descriptive

methods  
for estim  
or more  
examine  
survival t

Bird, K. I  
choic  
Box, G. E  
New  
Cole, D. A  
can b  
depe  
Gabriel, K  
Manc  
New  
Hand, D.  
meas  
Harris, R.  
anal  
Dekko  
Huberty, C  
analys  
McDonald  
effect  
zatio  
Mudholka  
with f  
(pp. 4  
Spector, P  
analys

methods (e.g., life tables and Kaplan-Meier survival functions) are available for estimating survival times for a sample or comparing the survival of two or more groups. Regression models (e.g., Cox, 1972) are also available to examine the contribution of continuous (metric) independent variables to survival time.

## Recommended Readings

- Bird, K. D., & Hadzi-Pavlovic, D. (1983). Simultaneous test procedures and the choice of a test statistic in MANOVA. *Psychological Bulletin, 93*, 167-178.
- Box, G. E. P., Hunter, W. R., & Hunter, J. S. (1978). *Statistics for experimenters*. New York: Wiley.
- Cole, D. A., Maxwell, S. E., Avery, R., & Salas, E. (1994). How the power of MANOVA can both increase and decrease as a function of the intercorrelations among dependent variables. *Psychological Bulletin, 115*, 465-474.
- Gabriel, K. R. (1969). A comparison of some methods of simultaneous inference in Manova. In P. R. Krishnaiah (Ed.), *Multivariate analysis-II* (pp. 67-86). New York: Academic Press.
- Hand, D. J., & Taylor, C. C. (1987). *Multivariate analysis of variance and repeated measures*. London: Chapman & Hall.
- Harris, R. J. (1993). Multivariate analysis of variance. In L. K. Edwards (Ed.), *Applied analysis of variance in behavioral science* (pp. 255-296). New York: Marcel Dekker.
- Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin, 105*, 302-308.
- McDonald, R. A., Seifert, C. F., Lorenzen, S. J., Givens, S., & Jaccard, J. (2002). The effectiveness of methods for analyzing multivariate factorial data. *Organizational Research Methods, 5*, 255-274.
- Mudholkar, G. S., & Subbaiah, P. (1980). MANOVA multiple comparisons associated with finite intersection tests. In P. R. Krishnaiah (Ed.), *Multivariate analysis V* (pp. 467-482). Amsterdam: North-Holland.
- Spector, P. E. (1977). What to do with significant multivariate effects in multivariate analyses of variance. *Journal of Applied Psychology, 62*, 158-163.