13 Analysis of Variance

OBJECTIVE

In this chapter, useful analysis of variance (ANOVA) techniques for comparing group means are presented. Specifically, the one-way ANOVA, two-way ANOVA, randomized block, Latin-square, repeated measures, and analysis of covariance techniques are treated in depth. Statistical assumptions and their robustness are likewise discussed. Tests of planned or complex comparisons of means are also illustrated.

13.1 Basic Concepts in Analysis of Variance

The term *analysis of variance* probably sounds familiar to you, especially if you have been schooled in at least one quantitative methodology course or have been working in the field of social sciences for some time. Analysis of variance (ANOVA), as the name implies, is a statistical technique that is intended to analyze variability in data in order to infer the inequality among population means. This may sound illogical, but there is more to this idea than just what the name implies.

The ANOVA technique extends what an independent-samples t test can do to multiple means. The null hypothesis examined by the independent-samples t test is that two population means are equal. If more than two means are compared, repeated use of the independent-samples t test will lead to a higher Type I error rate (the experiment-wise α level) than the α level set for each t test. A better approach than the t test is to consider all means in one null hypothesis—that is, examining the plausibility of the null hypothesis with a single statistical test. In doing so, researchers not only save time and energy, but more important, they can exercise a better control of the probability of falsely declaring significant differences among means. Such an idea was conceived by Sir R. A. Fisher more than 50 years ago. In his honor, the statistic used in ANOVA is called an F statistic.

The F statistic is a ratio. Its numerator and denominator are both estimates. When the null hypothesis of equal population means holds up, both estimates should be similar because they are estimates of the same quantity, that is, the variance of sampling errors. Under the alternative hypothesis, though, the numerator estimates not only the variance of sampling errors but also the squared treatment effect. And the denominator still estimates the error variance. Thus, the F ratio under the alternative hypothesis is noticeably larger than 1. The extent to which the observed F ratio is larger than 1 provides the basis for rejecting the null hypothesis in ANOVA.

Suppose that data were obtained from a typical state university on students' drinking behavior. The university had a policy banning hard liquors and beer from university properties, including dorms and Greek houses. But everybody knew somebody who drank while living on campus at this university. Students living off campus were even more likely to drink, perhaps. Let's look at weekly average drinks consumed by four groups of students and their variability:

```
                    Example 13.0 Average drinks and variability                                  1

                                The MEANS Procedure

             Analysis Variable : score1 no. of drinks in spring break

Four housing        N
conditions         Obs            Mean         Std Dev         Maximum         Minimum
-------------------------------------------------------------------------------------
Dorm                 8       3.0000000       1.5118579       6.0000000       1.0000000

Greek                8       3.5000000       0.9258201       5.0000000       2.0000000

Off-campus apt       8       4.2500000       1.0350983       6.0000000       3.0000000

Rented house         8       6.2500000       1.2817399       8.0000000       5.0000000
-------------------------------------------------------------------------------------
```

Notice from the printout that all sample means are different; so are sample standard deviations. To what extent can one know that the variation among these four means is not merely the variation that already existed among individuals, even in the same housing condition? The answer lies in an *F* test. The *F* test is formed from the mean square between groups or conditions divided by the mean square within groups. Both mean squares estimate the variance of sampling errors under the null hypothesis, as alluded to before. Under the alternative hypothesis, though, the mean square between groups will be larger than the mean square within groups. This is so because the mean square between groups, in this case, reflects not only sampling errors but also the varying numbers of drinks consumed by students living in four conditions. Thus, a significant *F* is indicated by a magnitude that is larger than 1 and statistically significant (see Example 13.1 for the *F* result and its *p* level).

The *F* test introduced in this chapter is associated with three statistical assumptions. The first assumption is that observations are randomly or independently selected from their respective populations. The second is that the shape of population distributions is normal. And the third is that these normal populations have identical variances. The consequences of violating any or all of these assumptions are discussed in **Section 13.5: Tips.** Suggestions on how to compensate for violating the assumptions are also included in the same section.

13.2 An Overview of the GLM Procedure for ANOVA

The GLM procedure is particularly well suited for analyzing data collected in any ANOVA design. The procedure name, GLM, stands for general linear models, which is the type of statistical models imposed on data in all ANOVA designs. A general linear model accounts for data in terms of main effects, interaction effects, nested effects, time-related effects, or merely sampling errors (or random errors). Correspondingly, types of ANOVA designs specified in the GLM procedure include completely randomized (Example 13.1), randomized factorial (Examples 13.2 and 13.3), randomized block (Example 13.4), Latin-square (Examples 13.5 and 13.6), repeated measures (Example 13.7), analysis of covariance (ANCOVA) (Examples 13.8 and 13.9), and any combination of these designs. Designs can be balanced (or orthogonal) or unbalanced. A balanced design is a design in which groups or cells have an equal number or a proportional number of data points in them. An unbalanced design does not have this property. Whenever possible, you should strive for a balanced design. Reasons for this suggestion are given in **Section 13.5: Tips**.

Two approaches, the univariate and the multivariate tests, for data collected from repeated measures designs are available in PROC GLM. Both are illustrated in Example 13.7.

Besides testing various null hypotheses with an *F* test, the GLM procedure offers a variety of multiple comparison procedures for the means. These include Dunn's (or the Bonferroni *t*) test, the Dunn-Šidák test, the one- and two-tailed Dunnett tests, the Scheffé test, the Newman-Keuls test, and Tukey's Honestly Significant Difference (or HSD) test. All are illustrated in this chapter. Other comparison procedures are presented in the online documentation at www.sas.com under the GLM procedure. Each test can be performed with a user-specified α level (see **Section 13.4**). Alternatively, you may request that a confidence interval be constructed for each pair of means. Tests of cell means for interactions or planned orthogonal contrasts are also available in PROC GLM. These are demonstrated in **Section 13.5.**

13.3 Examples

Data used in the following nine examples are from the raw data file design.dat. They are analyzed according to various ANOVA designs so as to illustrate certain data analysis techniques. All examples assume that the effects are fixed. Because of this, the interpretations of results presented in this chapter are for illustrative purposes only.

Example 13.1 One-Way Analysis of Variance

Do college students drink on campus, even against university policy? You bet, speaking from personal observations and the literature! But just how much do they drink? Let's investigate this issue by interviewing 32 students from a state university. These 32 students were randomly selected in equal numbers from (a) university dorms, (b) Greek houses, (c) off-campus apartments, and (d) rented houses. These students were asked to keep an honest record of drinks consumed during the spring-break week. To encourage these students to be honest, they were told that their data would remain confidential and be part of a national survey of college students' life on campus.

One intriguing question regarding college students' drinking is whether students in different housing arrangements exercised varying degrees of constraints on their drinking behavior and, hence, they drank varying amounts during the spring break. This question can be answered by a one-way ANOVA.

The program below addresses the question of how housing arrangements are related to weekly consumption of beer and hard liquor by college students during the spring break (score1). It consists of four statements. The first statement, **PROC GLM**, identifies a SAS data set design to be analyzed. The second statement, **CLASS**, lists one independent variable, indep1. The third statement, **MODEL**, specifies the design to be a one-way ANOVA design. Following the MODEL statement, the **MEANS** statement is used to carry out comparisons of group means. The two comparison procedures listed after slash (/) are **BON** and **TUKEY**. BON stands for Bonferroni *t* test, or the Dunn procedure, whereas TUKEY stands for Tukey's Honestly Significant Difference (or HSD) test.

```
/* The following bolded SAS statements establish the SAS data set 'design' */

PROC FORMAT;
     VALUE resident 1='Dorm' 2='Greek' 3='Off-campus apt' 4='Rented house';
RUN;

DATA design;
     INFILE 'd:\data\design.dat';
     INPUT indep1 id score1 score2 score3 sex $ major;
     LABEL indep1='four housing conditions'
           id='student id no.'
           score1='no. of drinks during the spring break'
           score2='no. of drinks during the final week'
           score3='no. of drinks after the final week'
           major='student academic major';
     FORMAT indep1 resident.;
RUN;

TITLE 'Example 13.1 One-way analysis of variance';

PROC GLM DATA=design;
     CLASS indep1;
     MODEL score1=indep1;
     MEANS indep1 / BON TUKEY;
RUN; QUIT;
```

Output 13.1 One-Way Analysis of Variance

```
           Example 13.1 One-way analysis of variance                          1

                       The GLM Procedure

                    Class Level Information

Class          Levels    Values

indep1              4    Dorm Greek Off-campus apt Rented house


            Number of Observations Read          32
            Number of Observations Used          32
```

Example 13.1 One-way analysis of variance 2

The GLM Procedure

Part (A)

Dependent Variable: score1 no. of drinks in spring break

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	49.00000000	16.33333333	11.15	<.0001
Error	28	41.00000000	1.46428571		
Corrected Total	31	90.00000000			

Part (B)

R-Square	Coeff Var	Root MSE	score1 Mean
0.544444	28.47239	1.210077	4.250000

Part (C)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
indep1	3	49.00000000	16.33333333	11.15	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
indep1	3	49.00000000	16.33333333	11.15	<.0001

Example 13.1 One-way analysis of variance 3

The GLM Procedure

Tukey's Studentized Range (HSD) Test for score1

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	28
Error Mean Square	1.464286
Critical Value of Studentized Range	3.86125
Minimum Significant Difference	1.6519

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	indep1
A	6.2500	8	Rented house
B	4.2500	8	Off-campus apt
B			
B	3.5000	8	Greek
B			
B	3.0000	8	Dorm

```
                        Example 13.1 One-way analysis of variance                        4

                                    The GLM Procedure

                           Bonferroni (Dunn) t Tests for score1

NOTE: This test controls the Type I experimentwise error rate, but it generally has a
                          higher Type II error rate than REGWQ.


                         Alpha                                 0.05
                         Error Degrees of Freedom                28
                         Error Mean Square                 1.464286
                         Critical Value of t                2.83893
                         Minimum Significant Difference      1.7177


                Means with the same letter are not significantly different.


             Bon Grouping          Mean      N    indep1

                         A        6.2500      8    Rented house

                         B        4.2500      8    Off-campus apt
                         B
                         B        3.5000      8    Greek
                         B
                         B        3.0000      8    Dorm
```

Page 1 of the output summarizes the ANOVA design: four levels (or groups) of the indep1 factor and 32 data points. According to page 2 of the output, the *F* test of average drinks reaches a significance level of 0.0001. This means that students living in various environments did drink unequal amounts of beer and hard liquor during the spring break. This conclusion is confirmed by Tukey's HSD test (page 3) and the Bonferroni *t* test (page 4). Both tests reveal that "Rented house" is the hardest drinking group, which is followed, to a lesser degree, by "Off-campus apt", "Greek", and "Dorm", in that order. The average drink in the "Off-campus apt" group was found to be statistically significantly different from "Rented house" but not significantly different from the other two groups. Likewise, the "Greek" group was not statistically significantly different from the "Dorm" group. These differences are identified by different letters, such as A and B, printed under **Tukey Grouping** and **Bon Grouping**. Groups with the same letter are considered to be not statistically significantly different from each other.

Is it necessary to apply two comparison procedures, such as Tukey and Bonferroni *t*? For exploration of data and for illustration of these procedures in SAS, the answer is yes. For confirming a theory or cross-validating other findings, no. Because this chapter is intended to expose you to various comparison procedures available in the GLM procedure, two procedures were specified in the program. Tukey's HSD test was specifically developed to examine all possible simple (or pairwise) differences. It controls the Type I error rate at the family-wise level, namely, for the set of all pairwise comparisons. The Bonferroni *t* test (the Dunn procedure) is more flexible. It can be used to test differences between two means as well as among three or more means. Both procedures can handle equal as well as unequal group sizes. Perhaps you'd ask, "If the Bonferroni *t* test is more flexible than Tukey's test, why will anyone need Tukey's procedure at all?" The answer lies in the statistical power. The statistical power of each test is best understood by the heading **Minimum Significant Difference.** This value sets the criterion by which an observed mean difference is judged to be statistically significant. So the smaller this number, the greater is the power. For the current data, Tukey's test is more powerful because its Minimum Significant Difference (or MSD) of 1.6519 is smaller than 1.7177 for the Bonferroni test. The latter procedure is definitely more flexible; but its flexibility comes at a price. In general, Tukey's test is the most powerful test for all pairwise comparisons, and it controls the experiment-wise Type I error rate at or below the α level specified by the researcher. The Bonferroni test is well suited to a mixture of simple and complex comparisons, especially when the total number of comparisons is neither too few nor too many, say, between 10 and 15. It is important to note that all comparison procedures programmed into GLM examine pairwise differences only. If complex comparisons of means are desired, alternative specifications are needed (see **Section 13.5: Tips**).

Let's now return to page 2 of the output and pick up the rest of the information. **Part (A)** assesses the overall significance with an *F* test (= 11.15) and its *p* level (< 0.0001). Both **Type I** and **Type III SS** in **Part (C)** offer identical information as **Part (A)**. These two parts are identical only in a one-way ANOVA design, because there is only one effect to be tested. Therefore, **Part (C)** can be ignored for a one-way design. **Part (B)** presents four descriptive statistics. The first is **R-Square** (= 0.544444), which is the ratio of SS_{model} to SS_{total}, or 49/90. The R-Square value indicates that 54.4444% of the variability of the number of drinks consumed by students is explained by this one-way ANOVA model. The second is **Coeff Var (C.V.),** which stands for coefficient of variation or the ratio of standard deviation divided by the overall mean times 100 (= 1.210077 ÷ 4.25 × 100 = 28.47239). The third is **Root MSE** or the square root of Mean Square Error $(=\sqrt{1.46428571}=1.210077)$. The root MSE is the sample estimate for the population standard deviation. It is used to calculate the MSD reported on pages 3 and 4 of the output. The fourth statistic, **score1 Mean** (= 4.25), is the grand average of the dependent variable, that is, the average number of drinks consumed by 32 college students in this study.

Example 13.2 Two-Way Analysis of Variance

Because there is a common perception that men drink more than women, let's see if gender is a factor in the student survey described above. Let's suppose that out of eight students randomly selected from each of the four housing conditions, half were women and half were men. Hence, it is possible to study the gender effect, the housing condition, and the joint effect of both factors on college students' drinking behavior. The SAS program written below is much like the one presented in Example 13.1 except for the CLASS and the MODEL statements. The CLASS statement now lists <u>indep1</u> and <u>sex</u>

as independent variables. The MODEL statement has three terms listed on the right side of the equal sign (=): indep1, sex, and indep1*sex, which represent two main effects and one interaction, respectively. Thus, the corresponding design is a two-way ANOVA.

```
/* See Example 13.1 for the DATA step in creating the SAS data set 'design' */

TITLE 'Example 13.2 Two-way analysis of variance';

PROC GLM DATA=design;
     CLASS indep1 sex;
     MODEL score1=indep1 sex indep1*sex;
     MEANS sex indep1 / BON;
RUN; QUIT;
```

Output 13.2 Two-Way Analysis of Variance

```
Example 13.2 Two-way analysis of variance                                                    1

            The GLM Procedure

         Class Level Information

   Class         Levels    Values

   indep1             4    1 2 3 4

   sex                2    Female Male


 Number of Observations Read          32
 Number of Observations Used          32
```

```
                         Example 13.2 Two-way analysis of variance                           2

                                    The GLM Procedure
  Part (A)

  Dependent Variable: score1   no. of drinks in spring break

                                            Sum of
     Source                      DF        Squares     Mean Square    F Value    Pr > F

     Model                        7    60.00000000      8.57142857       6.86    0.0002

     Error                       24    30.00000000      1.25000000

     Corrected Total             31    90.00000000

  Part (B)
                     R-Square     Coeff Var      Root MSE    score1 Mean

                     0.666667      26.30668      1.118034       4.250000

  Part (C)

     Source                      DF      Type I SS     Mean Square    F Value    Pr > F

     indep1                       3    49.00000000     16.33333333      13.07    <.0001
     sex                          1     8.00000000      8.00000000       6.40    0.0184
     indep1*sex                   3     3.00000000      1.00000000       0.80    0.5061


     Source                      DF    Type III SS     Mean Square    F Value    Pr > F

     indep1                       3    49.00000000     16.33333333      13.07    <.0001
     sex                          1     8.00000000      8.00000000       6.40    0.0184
     indep1*sex                   3     3.00000000      1.00000000       0.80    0.5061
```

```
                   Example 13.2 Two-way analysis of variance                        3

                                The GLM Procedure
Part (D)
                         Bonferroni (Dunn) t Tests for score1

 NOTE: This test controls the Type I experimentwise error rate, but it generally has a
                        higher Type II error rate than REGWQ.


                    Alpha                                    0.05
                    Error Degrees of Freedom                   24
                    Error Mean Square                        1.25
                    Critical Value of t                   2.06390
                    Minimum Significant Difference         0.8158


             Means with the same letter are not significantly different.


               Bon Grouping          Mean      N    sex

                          A        4.7500     16    Male

                          B        3.7500     16    Female
```

```
                   Example 13.2 Two-way analysis of variance                        4

                                The GLM Procedure
Part (E)
                         Bonferroni (Dunn) t Tests for score1

 NOTE: This test controls the Type I experimentwise error rate, but it generally has a
                        higher Type II error rate than REGWQ.


                    Alpha                                    0.05
                    Error Degrees of Freedom                   24
                    Error Mean Square                        1.25
                    Critical Value of t                   2.87509
                    Minimum Significant Difference         1.6072


             Means with the same letter are not significantly different.


          Bon Grouping          Mean      N    indep1

                     A        6.2500      8    Rented house

                     B        4.2500      8    Off-campus apt
                     B
                     B        3.5000      8    Greek
                     B
                     B        3.0000      8    Dorm
```

Output 13.2 has the same appearance as Output 13.1. Therefore, there is no need to explain many of the concepts again; only new terms are discussed here. Page 2 of the output is divided into three parts. **Part (A)** presents the *F* test for the overall design, its value (= 6.86), and the *p* level (= 0.0002); all are indicative of some effect being statistically significant in the data. Hence, **Part (C)** becomes relevant at this point. It shows that both main effects are significant at the $p < 0.0001$ and 0.0184 levels, respectively, yet the interaction is not. Look for these results under the heading **Type I SS** and **Pr > F**.

Out of the two significant main effects, the sex effect is new and is followed up by the Bonferroni *t* test—**Part (D)** on page 3—that shows males (mean = 4.75) indeed drank significantly more than females (mean = 3.75). One question for you to think over is this: Is it necessary to perform the Bonferroni *t* test on the sex difference, if the *F* test of the same variable is already statistically significant at $\alpha = 0.05$, based on $p = 0.0184$?

The other statistically significant effect due to indep1 has a larger *F* ratio (= 13.07) in **Part (C)**, compared with 11.15 from **Output 13.1**, though the significance level is identical ($p < 0.0001$). The Bonferroni *t*-test result reaches the same conclusion as that shown in **Output 13.1**, namely, the 4th group, living in rented houses, drank significantly more than the other three groups [**Part (E)** on page 4].

Example 13.3 Confirming No Interaction With a Plot of Cell Means

How can you cross-validate the lack of significant interactions in data? There is an easy way: Calculate eight cell means and plot these means using the symbols of the sex variable. Here is a program written for this purpose:

```
/* See Example 13.1 for the DATA step in creating the SAS data set 'design' */

TITLE 'Example 13.3 Confirming no interaction with a plot of cell means';

PROC MEANS DATA=design NOPRINT;
     VAR score1;
     OUTPUT OUT=out MEAN=meandrnk;
     CLASS sex indep1;
RUN;

PROC PRINT DATA=out;
RUN;

PROC PLOT DATA=out;
     PLOT meandrnk*indep1=sex / HPOS=50 VPOS=20;
RUN;
```

The program uses three SAS procedures: MEANS, PRINT, and PLOT. The purpose of PROC MEANS is to compute cell means and save them in a SAS data set called out. Note that no printout is requested by the MEANS procedure. Instead, PROC PRINT is used to list the grand mean, eight cell means plus four group means of indep1 and two means of sex. This output (page 1 below) is much simpler than what would have been generated by PROC MEANS. The last procedure, PLOT, is used to graphically display eight cell means under four housing conditions using symbols “F” or “M” of the sex variable. Two options, HPOS= and VPOS=, are specified primarily to control the frame of the plot.

Output 13.3 Confirming No Interaction With a Plot of Cell Means

Example 13.3 Confirming no interaction with a plot of cell means 1

Obs	sex	indep1	_TYPE_	_FREQ_	meandrnk
1		.	0	32	4.25
2		1	1	8	3.00
3		2	1	8	3.50
4		3	1	8	4.25
5		4	1	8	6.25
6	Female	.	2	16	3.75
7	Male	.	2	16	4.75
8	Female	1	3	4	2.50
9	Female	2	3	4	3.50
10	Female	3	3	4	3.50
11	Female	4	3	4	5.50
12	Male	1	3	4	3.50
13	Male	2	3	4	3.50
14	Male	3	3	4	5.00
15	Male	4	3	4	7.00

```
            Example 13.3 Confirming no interaction with a plot of cell means        2

                  Plot of meandrnk*indep1.  Symbol is value of sex.

meandrnk |
    7.00 +                                             M
    6.75 +
    6.50 +
    6.25 +
    6.00 +
    5.75 +
    5.50 +                                             F
    5.25 +
    5.00 +                              M
    4.75 +
    4.50 +
    4.25 +
    4.00 +
    3.75 +
    3.50 +M              F              F
    3.25 +
    3.00 +
    2.75 +
    2.50 +F
         -+--------------+--------------+--------------+-
        Dorm           Greek     Off-campus apt   Rented house

                          four housing conditions

NOTE: 3 obs had missing values.  1 obs hidden.
```

Notice how, on the page 2 plot, the letter M always lies above F, except for Greek houses where F and M collide because their corresponding means are identical. As long as one gender group (males in this case) constantly drank more than, or at least as much as, the other gender group (females) across the four housing conditions, there is likely to be no statistically significant interaction. Graphing cell means is a good way to infer the presence or the absence of an interaction effect. Of course, if there is no interaction in the population, these two groups will differ by the same magnitude across the four housing conditions. As a general observation, if both main effects are statistically significant, the interaction is unlikely to be also significant. If the interaction is statistically significant, one or both main effects are unlikely to be significant.

Example 13.4 Randomized Block Design

One tactic in conducting experimental or quasi-experimental studies is to control for the impact of extraneous variables that are not the researcher's main interest. One way to handle an extraneous variable is to match subjects on such a variable so that its presence is well represented in all groups of the independent variable. This type of design is called a randomized block design.

Suppose that the amount of drinks consumed by students could be a function of their academic majors. We, therefore, need to control for the variation of majors in each housing condition. Let's factor students' major (major) into the analysis while keeping the housing arrangements (indep1) as the sole independent variable in the study. Both variables are listed on the CLASS statement as sources of effects.

The MODEL statement specifies indep1 and major as the two effects that account for the variation in the dependent variable. There is no interaction of indep1 by major listed on the MODEL statement because, in a block design, the interaction between the independent variable and the matching (or the blocking) variable is assumed nonexistent.

The MEANS statement specifies indep1 to test the mean differences due to housing arrangements, and SIDAK requests the Dunn-Šidák comparison procedure to test the mean differences.

```
/* See Example 13.1 for the DATA step in creating the SAS data set 'design' */

TITLE 'Example 13.4 Randomized block design';

PROC GLM DATA=design;
     CLASS indep1 major;
     MODEL score1=indep1 major;
     MEANS indep1 / SIDAK;
RUN; QUIT;
```

Output 13.4 Randomized Block Design

```
  Example 13.4 Randomized block design                                    1

           The GLM Procedure

        Class Level Information

Class          Levels     Values

indep1              4     1 2 3 4

major               8     1 2 3 4 5 6 7 8


Number of Observations Read          32
Number of Observations Used          32
```

```
                              Example 13.4 Randomized block design                        2

                                       The GLM Procedure
Part (A)

Dependent Variable: score1   no. of drinks in spring break

                                                Sum of
   Source                       DF             Squares     Mean Square    F Value    Pr > F

   Model                        10         70.50000000      7.05000000       7.59    <.0001

   Error                        21         19.50000000      0.92857143

   Corrected Total              31         90.00000000

Part (B)
                  R-Square     Coeff Var      Root MSE    score1 Mean
                  0.783333      22.67351      0.963624       4.250000

Part (C)
   Source                       DF           Type I SS     Mean Square    F Value    Pr > F

   indep1                        3         49.00000000     16.33333333      17.59    <.0001
   major                         7         21.50000000      3.07142857       3.31    0.0156


   Source                       DF         Type III SS     Mean Square    F Value    Pr > F

   indep1                        3         49.00000000     16.33333333      17.59    <.0001
   major                         7         21.50000000      3.07142857       3.31    0.0156
```

```
                            Example 13.4 Randomized block design                          3

                                     The GLM Procedure

                                  Sidak t Tests for score1

NOTE: This test controls the Type I experimentwise error rate, but it generally has a
                       higher Type II error rate than REGWQ.
```

```
                    Alpha                                 0.05
                    Error Degrees of Freedom                21
                    Error Mean Square                 0.928571
                    Critical Value of t                2.90270
                    Minimum Significant Difference   1.3986


          Means with the same letter are not significantly different.


      Sidak Grouping          Mean       N    indep1

                    A        6.2500      8    Rented house

                    B        4.2500      8    Off-campus apt
                    B
                    B        3.5000      8    Greek
                    B
                    B        3.0000      8    Dorm
```

Page 2 of the output is divided into three parts for easy explanation. **Part (A)** shows the overall significance ($F = 7.59$, $p < 0.0001$) of the design model to account for variance in score1. **Part (B)** supports the significant finding with a high **R-Square** (= 0.783333) and a small **Root MSE** (= 0.963624). **Part (C)** presents the *F* test of indep1 (= 17.59) and its *p* level (< 0.0001). This *F* value is larger than the one reported in **Output 13.1**. It is so because the denominator of the present *F* is slightly smaller than the one before, due to model differences. In other words, by matching students on their majors, we have effectively reduced the sum of squares of errors to such an extent that its mean square (or the reduced SS divided by its reduced degrees of freedom) is still smaller than the value derived from the one-way ANOVA model. Thus, the effort to match subjects was fruitful.

The question, "How effective is the matching?" can also be answered by the *F* test of the major effect. In **Part (C),** under **Type I SS,** it shows that such an *F* test is statistically significant at $\alpha = 0.05$ ($p = 0.0156$). Thus, we conclude that matching students on majors effectively reduced the **Mean Square Error** from 1.46428571 (from Output 13.1) to 0.92857143, reported in Output 13.4.

On page 3 of the output, the SIDAK procedure follows up on the significant *F* of indep1 by examining all pairwise differences in means. This test result reaches the same conclusion as Output 13.1 or Output 13.2, namely, the 4th group, living in rented houses, drank significantly more than the other three groups. The Dunn-Šidák test is an improvement over the Bonferroni *t* test (also called the Dunn procedure) because it requires a smaller critical value in computing the MSD than the Bonferroni *t* test.

Example 13.5 Latin-Square Design

Have you heard of the phrase, "Statistics is Greek to me!"? Well, add Latin on top of the Greek! In ANOVA, there is actually a design called the Latin-square (or LS) design. The LS design is an extension of the randomized block design. In a randomized block design, only one extraneous variable is being controlled, whereas in a LS design, two are controlled. Here is the layout of a LS design—suppose that in the data file design.dat, variable a is the old indep1 variable, that is, the four housing arrangements. Two other variables, b and c, denote two extraneous variables, academic standing and majors, respectively. Let's further suppose that the 32 data items were collected according to the 4 × 4 LS design depicted below:

	c1	c2	c3	c4
b1	a1 3 2	a2 4 4	a3 4 4	a4 5 5
b2	a2 3 3	a3 3 3	a4 6 6	a1 2 3
b3	a3 4 6	a4 7 8	a1 1 3	a2 2 3
b4	a4 5 8	a1 4 6	a2 4 5	a3 5 5

As you probably recall from a statistics textbook, a LS design is one in which the number of levels (or groups) of the treatment variable, as well as that of the two extraneous variables, ought to be identical. For this reason, variables b and c were artificially created to also contain four groups, like the four housing conditions under variable a.

In the SAS program, the rearranged data are first read into a SAS data set called ls, and then analyzed by the GLM procedure. On the MODEL statement, three main effects plus one three-way interaction are specified. These are followed by a MEANS statement with the **SCHEFFE** post hoc procedure specified after the slash (/). You should be forewarned that the three-way interaction is not supposed to reach significance because LS designs assume that no interaction exists between the treatment factor and one or all of the extraneous variables.

```
/* The following bolded statements establish the SAS data set 'ls' */

DATA ls;
     INPUT a b c score @@;
     LABEL a='Four housing conditions'
           b='academic standing'
           c='major'
           score='no. of drinks in spring break';
DATALINES;
1 1 1 3  1 1 1 2  2 2 1 3  2 2 1 3  3 3 1 4  3 3 1 6  4 4 1 5  4 4 1 8
2 1 2 4  2 1 2 4  3 2 2 3  3 2 2 3  4 3 2 7  4 3 2 8  1 4 2 4  1 4 2 6
3 1 3 4  3 1 3 4  4 2 3 6  4 2 3 6  1 3 3 1  1 3 3 3  2 4 3 4  2 4 3 5
4 1 4 5  4 1 4 5  1 2 4 2  1 2 4 3  2 3 4 2  2 3 4 3  3 4 4 5  3 4 4 5
RUN;

TITLE 'Example 13.5 Latin-square design';

PROC GLM DATA=ls;
     CLASS a b c;
     MODEL score=a b c a*b*c;
     MEANS b c / SCHEFFE;
RUN; QUIT;
```

Output 13.5 Latin-Square Design

```
    Example 13.5 Latin-square design                                    1

           The GLM Procedure

        Class Level Information

    Class          Levels    Values

    a                   4    1 2 3 4

    b                   4    1 2 3 4

    c                   4    1 2 3 4


Number of Observations Read          32
Number of Observations Used          32
```

Example 13.5 Latin-square design 2

The GLM Procedure

Part (A)

Dependent Variable: score no. of drinks in spring break

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	77.00000000	5.13333333	6.32	0.0003
Error	16	13.00000000	0.81250000		
Corrected Total	31	90.00000000			

Part (B)

R-Square	Coeff Var	Root MSE	score Mean
0.855556	21.20913	0.901388	4.250000

Part (C)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
a	3	49.00000000	16.33333333	20.10	<.0001
b	3	12.25000000	4.08333333	5.03	0.0121
c	3	5.25000000	1.75000000	2.15	0.1335
a*b*c	6	10.50000000	1.75000000	2.15	0.1031

Source	DF	Type III SS	Mean Square	F Value	Pr > F
a	3	49.00000000	16.33333333	20.10	<.0001
b	3	12.25000000	4.08333333	5.03	0.0121
c	3	5.25000000	1.75000000	2.15	0.1335
a*b*c	6	10.50000000	1.75000000	2.15	0.1031

Part (D) Example 13.5 Latin-square design 3

The GLM Procedure

Scheffe's Test for score

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	16
Error Mean Square	0.8125
Critical Value of F	3.23887
Minimum Significant Difference	1.4049

Means with the same letter are not significantly different.

Scheffe Grouping		Mean	N	b
	A	5.2500	8	4
	A			
B	A	4.2500	8	3
B	A			
B	A	3.8750	8	1
B				
B		3.6250	8	2

```
Part (E)                     Example 13.5 Latin-square design                          4

                                     The GLM Procedure

                                  Scheffe's Test for score

                  NOTE: This test controls the Type I experimentwise error rate.


                           Alpha                                   0.05
                           Error Degrees of Freedom                  16
                           Error Mean Square                     0.8125
                           Critical Value of F                  3.23887
                           Minimum Significant Difference        1.4049


                   Means with the same letter are not significantly different.


                    Scheffe Grouping          Mean      N    c

                                   A        4.8750      8    2
                                   A
                                   A        4.2500      8    1
                                   A
                                   A        4.1250      8    3
                                   A
                                   A        3.7500      8    4
```

Page 1 and **Parts (A)** and **(B)** of page 2 should be familiar to you by now; therefore, there is no need to explain them again. Beginning with **Part (C)**, **Type I SS**, four *F* tests of main effects and the interaction effect are presented. The main effect of a (the four housing conditions) on drinking behavior is statistically significant as before. The *F* value is larger than before due to a smaller mean square error. The effect of b is also statistically significant at 0.0121, but the effect of c is not significant (p = 0.1335). This means that factor b, but not factor c, was an effective matching variable that accounted for a substantial portion of variance in the number of drinks. The significant *F* test for factor b is followed up by the Scheffé post hoc test. **Part (D)** on page 3 reveals that the Scheffé test found that the fourth level (seniors) of factor b (academic standing) yielded a significantly higher average number of drinks than the second level (sophomores). So it would be interesting to trace back to data and figure out who were these seniors and sophomores that contributed to this significant difference. In **Part (E)** on page 4, analysis of factor c did not detect any pair of means to be significantly different, as the overall *F* test of the same effect is not significant.

Earlier in this example, it was pointed out that any LS design assumes that no interaction exists. Fortunately, the interaction was not significant for the present data (p = 0.1031). Therefore, the assumption is met.

Example 13.6 Collapsing the Interaction With Residuals in a Latin-Square Design

Because the three-way interaction is tested to be nonsignificant, it becomes another estimate for the variance of sampling errors. One estimate already exists; it is the mean square error, printed in **Part (A)**. Some statistics textbooks suggest that these two be combined in order to increase the degrees of freedom. This recommendation can be easily implemented in a SAS program. Note here that the three-way interaction is removed from the MODEL statement. The removal implies that the three-way interaction is pooled with the error term. The combined mean square may be called the *residual mean square* or *mean square residual*.

```
/* See Example 13.5 for the DATA step in creating the SAS data set 'ls' */

TITLE 'Example 13.6 Collapsing the interaction with residuals in a Latin-square design';

PROC GLM DATA=ls;
     CLASS a b c;
     MODEL score=a b c;
     MEANS b c / SCHEFFE;
RUN; QUIT;
```

Output 13.6 Collapsing the Interaction With Residuals in a Latin-Square Design

```
[Page 1 output is omitted]

    Example 13.6 Collapsing the interaction with residuals in a Latin-square design      2

                                 The GLM Procedure

Dependent Variable: score   no. of drinks in spring break

                                          Sum of
   Source                      DF        Squares     Mean Square    F Value    Pr > F

   Model                        9    66.50000000      7.38888889       6.92    0.0001

   Error                       22    23.50000000      1.06818182

   Corrected Total             31    90.00000000


                    R-Square     Coeff Var      Root MSE    score Mean

                    0.738889      24.31833      1.033529      4.250000


   Source                      DF      Type I SS     Mean Square    F Value    Pr > F

   a                            3    49.00000000     16.33333333      15.29    <.0001
   b                            3    12.25000000      4.08333333       3.82    0.0241
   c                            3     5.25000000      1.75000000       1.64    0.2093


   Source                      DF    Type III SS     Mean Square    F Value    Pr > F

   a                            3    49.00000000     16.33333333      15.29    <.0001
   b                            3    12.25000000      4.08333333       3.82    0.0241
   c                            3     5.25000000      1.75000000       1.64    0.2093
```

```
Example 13.6 Collapsing the interaction with residuals in a Latin-square design     3

                              The GLM Procedure

                           Scheffe's Test for score

         NOTE: This test controls the Type I experimentwise error rate.


                    Alpha                                  0.05
                    Error Degrees of Freedom                 22
                    Error Mean Square                  1.068182
                    Critical Value of F                 3.04912
                    Minimum Significant Difference       1.5629


           Means with the same letter are not significantly different.


              Scheffe Grouping          Mean      N    b

                             A        5.2500      8    4
                             A
                        B    A        4.2500      8    3
                        B    A
                        B    A        3.8750      8    1
                        B
                        B             3.6250      8    2
```

```
Example 13.6 Collapsing the interaction with residuals in a Latin-square design     4

                              The GLM Procedure

                           Scheffe's Test for score

         NOTE: This test controls the Type I experimentwise error rate.


                    Alpha                                  0.05
                    Error Degrees of Freedom                 22
                    Error Mean Square                  1.068182
                    Critical Value of F                 3.04912
                    Minimum Significant Difference       1.5629


          Means with the same letter are not significantly different.


           Scheffe Grouping          Mean      N    c

                           A       4.8750      8    2
                           A
                           A       4.2500      8    1
                           A
                           A       4.1250      8    3
                           A
                           A       3.7500      8    4
```

The output conveys identical messages, as in Output 13.5, in terms of significant results of a and b main effects. One thing is different, though; the Model *F* value increases from 6.32 to 6.92, yet the *F* values of a, b, and c decrease in magnitude. The reduction in these *F* values is due to an increase in MS for the error term, which is not offset by an increase in degrees of freedom.

Example 13.7 Repeated Measures Design ($SPF_{p.q}$)

This example illustrates analytical approaches for a repeated measures design. Let's suppose that three data points were collected from each student: one during the spring break (score1), one during the final week (score2), and another after the final week (score3). With these additional measures, it is possible to determine whether college students' drinking habits were related to their stress, assuming greater stress was felt at the end of a semester than during the spring break or after the finals. A repeated measures design is a type of split plot factorial design for which between-block and within-block differences and their interactions are investigated. **Plot** is an agricultural term that refers to a parcel of land, divided into sub-plots that are called **blocks**. Within a block, the soil condition, irrigation, plants, and so on are homogeneous. By the same token, a repeated measures design regards observations in the same treatment level (or group) to be homogeneous. Differences observed within blocks are explained by the repeated factor (time in this example). Differences observed between blocks are explained by the between-block factor, or the four housing arrangements coded as indep1. A repeated measures design with one between-block factor and one within-block factor is denoted as $SPF_{p.q}$, where p is the number of levels for the between-block factor (p = 4 in this example) and q is the number of levels for the within-block factor (q = 3 in this example). An $SPF_{p.q}$ design yields three effects to be examined: two main effects of the between-block factor and the within-block factor and one interaction effect of these two factors.

In the program below, the CLASS statement lists indep1 as the sole independent variable. The MODEL statement lists score1, score2, and score3 as dependent variables on the left and indep1 on the right-hand side of the equal sign (=). This statement will cause PROC GLM to apply multivariate analyses to the three dependent variables. The next statement, REPEATED, applies univariate analyses to the data. The repeated factor, time, is the overarching variable under which score1, score2, and score3 are its three levels.

```
/* See Example 13.1 for the DATA step in creating the SAS data set 'design' */

TITLE 'Example 13.7 Repeated measures design (SPF p.q)';

PROC GLM DATA=design;
     CLASS indep1;
     MODEL score1-score3=indep1;
     REPEATED time;
RUN; QUIT;
```

Output 13.7 Repeated Measures Design ($SPF_{p.q}$)

```
Example 13.7 Repeated measures design (SPF p.q)                             1

               The GLM Procedure

            Class Level Information

        Class          Levels    Values

        indep1              4    1 2 3 4


    Number of Observations Read          32
    Number of Observations Used          32
```

```
                  Example 13.7 Repeated measures design (SPF p.q)                    2

                                 The GLM Procedure

Dependent Variable: score1   no. of drinks in spring break

                                            Sum of
   Source                      DF          Squares     Mean Square    F Value    Pr > F

   Model                        3      49.00000000     16.33333333      11.15    <.0001

   Error                       28      41.00000000      1.46428571

   Corrected Total             31      90.00000000

                   R-Square     Coeff Var      Root MSE    score1 Mean

                   0.544444      28.47239      1.210077       4.250000

   Source                      DF        Type I SS     Mean Square    F Value    Pr > F

   indep1                       3      49.00000000     16.33333333      11.15    <.0001

   Source                      DF      Type III SS     Mean Square    F Value    Pr > F

   indep1                       3      49.00000000     16.33333333      11.15    <.0001
```

Example 13.7 Repeated measures design (SPF p.q) 3

The GLM Procedure

Dependent Variable: score2 no. of drinks in final week

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	52.7500000	17.5833333	7.11	0.0011
Error	28	69.2500000	2.4732143		
Corrected Total	31	122.0000000			

R-Square	Coeff Var	Root MSE	score2 Mean
0.432377	44.93273	1.572646	3.500000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
indep1	3	52.75000000	17.58333333	7.11	0.0011

Source	DF	Type III SS	Mean Square	F Value	Pr > F
indep1	3	52.75000000	17.58333333	7.11	0.0011

Example 13.7 Repeated measures design (SPF p.q) 4

The GLM Procedure

Dependent Variable: score3 no. of drinks after final week

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	29.6250000	9.8750000	2.11	0.1219
Error	28	131.2500000	4.6875000		
Corrected Total	31	160.8750000			

R-Square	Coeff Var	Root MSE	score3 Mean
0.184149	46.18802	2.165064	4.687500

Source	DF	Type I SS	Mean Square	F Value	Pr > F
indep1	3	29.62500000	9.87500000	2.11	0.1219

Source	DF	Type III SS	Mean Square	F Value	Pr > F
indep1	3	29.62500000	9.87500000	2.11	0.1219

Example 13.7 Repeated measures design (SPF p.q) 5

The GLM Procedure
Repeated Measures Analysis of Variance

Repeated Measures Level Information

Dependent Variable	score1	score2	score3
Level of time	1	2	3

Part (A)

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no time Effect
H = Type III SSCP Matrix for time
E = Error SSCP Matrix

S=1 M=0 N=12.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.23666320	43.54	2	27	<.0001
Pillai's Trace	0.76333680	43.54	2	27	<.0001
Hotelling-Lawley Trace	3.22541397	43.54	2	27	<.0001
Roy's Greatest Root	3.22541397	43.54	2	27	<.0001

Part (B)

MANOVA Test Criteria and F Approximations for the Hypothesis of no time*indep1 Effect
H = Type III SSCP Matrix for time*indep1
E = Error SSCP Matrix

S=2 M=0 N=12.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.79355122	1.10	6	54	0.3727
Pillai's Trace	0.21628459	1.13	6	56	0.3561
Hotelling-Lawley Trace	0.24776343	1.10	6	34.278	0.3844
Roy's Greatest Root	0.17821404	1.66	3	28	0.1975

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

Example 13.7 Repeated measures design (SPF p.q) 6

The GLM Procedure
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
indep1	3	128.2083333	42.7361111	5.74	0.0034
Error	28	208.4166667	7.4434524		

```
                  Example 13.7 Repeated measures design (SPF p.q)                    7

                                The GLM Procedure
                       Repeated Measures Analysis of Variance
              Univariate Tests of Hypotheses for Within Subject Effects
Part (C)
                                                                            Adj Pr > F
Source                     DF   Type III SS   Mean Square  F Value  Pr > F   G - G   H - F

time                        2   23.08333333   11.54166667    19.54  <.0001  <.0001  <.0001
time*indep1                 6    3.16666667    0.52777778     0.89  0.5062  0.4817  0.4916
Error(time)                56   33.08333333    0.59077381


                            Greenhouse-Geisser Epsilon    0.7068
                            Huynh-Feldt Epsilon           0.8131
```

This output probably causes your eyes to cross! Let's begin with page 2. This page is identical to page 2 of Output 13.1, based on a one-way ANOVA design. Thus, you can conclude that during the spring break, students drank more or less liquor depending on where they lived.

Pages 3 and 4 display the second and third one-way ANOVA result based on score2 and score3, respectively. Like score1, the *F* test of students' drinking during the final week is statistically significant at $\alpha = 0.05$ ($F = 7.11$, $p = 0.0011$). The **R-Square** is lower and **MSE** is higher in score2, compared with score1. However, the *F* test of score3 (i.e., the number of drinks after the final week) is not statistically significant ($p = 0.1219$).

Page 5 is devoted entirely to the multivariate analysis of score1, score2, and score3. **Part (A)** presents four multivariate tests of the main effect, time. **Part (B)** presents tests of the interaction between time and indep1. Each of the four multivariate tests is based on a slightly different alternative hypothesis. The time factor was tested to be statistically significant at $\alpha = 0.05$ by all four multivariate tests. However, none uncovers statistically significant differences in the number of drinks due to the interaction between time and indep1.

The univariate tests are presented on pages 6 and 7. Page 6 displays test results of the between-block factor (indep1). According to the magnitude of the *F* value (= 5.74) and its *p* level (= 0.0034), the four housing conditions had an impact on the students' drinking behaviors. This finding has been shown in previous examples.

Page 7 of the output contains univariate analyses of the repeated factor, time, and its interaction with the between-block factor, indep1. Both are tested using the denominator called **Error (time)**. This term is usually referred to in statistics textbooks as the interaction of the repeated factor, time, with the error term of the between-block factor. This error term is smaller than the between-block error term. Verify this by comparing 0.59077381 (page 7) with 7.4434524 (page 6). Using this smaller error term as the denominator, the *F* test for the time factor in **Part (C)** is significant ($F = 19.54$, $p < 0.0001$). However, the *F* test for the time*indep1 interaction is not significant ($F = 0.89$, $p = 0.5062$).

You may have noticed that there are three *p* values listed after the *F* value in **Part (C)** on page 7. Besides the one you are familiar with (i.e., **Pr > F**), there are two additional column headings that read as "**Adj. Pr > F**" according to the "**G-G**" and "**H-F**" correction formulae, respectively. The **G - G** correction formula refers to the conservative approach proposed by Geisser-Greenhouse, whereas **H - F** refers to the Huynh-Feldt approach. Both approaches seek to correct the *p* levels of univariate *F* tests performed on the repeated factor and its interaction with the between-block factor. The corrections are needed because both *F* tests assume that the variance-covariance matrix of repeated measures is of a certain type. Violation of this assumption results in a positive bias in the *F* statistic; hence, it is inflated. These correction formulae adjust the significance level downward, by multiplying the degrees of freedom with the **Epsilon** coefficient (Epsilon = 0.7068 for the G-G correction formula, and Epsilon = 0.8131 for the H-F formula), when data do not satisfy this assumption. And data almost always violate this structural requirement assumed for the variance-covariance matrix. In our example, the corrections do not change the significant conclusion reached for the time factor or the nonsignificant conclusion for the time*indep1 interaction.

Example 13.8 Analysis of Covariance (ANCOVA)

Given the purpose of Example 13.7 and its null hypotheses, there exists an alternative way of examining the data to determine if, in fact, time makes a difference in students' drinking behavior. This example demonstrates this alternative analysis strategy, namely, the analysis of covariance, or ANCOVA. To demonstrate this strategy, the first measure, score1, is treated as a covariate. The second measure, score2, is treated as the dependent variable, and indep1 is the independent variable or the treatment factor.

The idea behind ANCOVA is simple. If a variable, namely, the covariate, is linearly related to the dependent variable, yet it is not the main focus of a study, its effect can be partialled out from the dependent variable through the least-squares regression equation. The remaining, or the adjusted, portion of the dependent variable is subsequently analyzed according to the usual ANOVA designs. In this example, students' drinking during the final week is adjusted for their spring break drinking. The adjusted number of drinks is subsequently analyzed by four housing arrangements in a one-way ANOVA.

In programming an ANCOVA design into PROC GLM, it is better to write score1 (the covariate) before indep1 (the independent variable) on the MODEL statement. In doing so, you will only need to interpret the **TYPE I** sum of squares result from page 2 of the output. Furthermore, the **LSMEANS** statement replaces the MEANS statement. LSMEANS stands for the least-squares means. The least-squares means are average number of drinks during the final week after they are adjusted for average number of drinks consumed during the spring break (the covariate). Two options, **PDIFF** and

STDERR, are specified to make a comparison between each pair of adjusted means. PDIFF requests significance levels for tests of all pairs of adjusted means. STDERR requests the *t* test of each adjusted mean against 0 and prints the significance level of the *t* test.

```
/* See Example 13.1 for the DATA step in creating the SAS data set 'design' */

TITLE 'Example 13.8 Analysis of covariance (ANCOVA)';

PROC GLM DATA=design;
     CLASS indep1;
     MODEL score2=score1 indep1;
     LSMEANS indep1 / PDIFF STDERR;
RUN; QUIT;
```

Output 13.8 Analysis of Covariance (ANCOVA)

```
Example 13.8 Analysis of covariance (ANCOVA)                                1

              The GLM Procedure

           Class Level Information

       Class         Levels    Values

       indep1             4    1 2 3 4


   Number of Observations Read          32
   Number of Observations Used          32
```

```
                    Example 13.8 Analysis of covariance (ANCOVA)                    2

                                The GLM Procedure
Part (A)

Dependent Variable: score2   no. of drinks in final week

                                       Sum of
   Source                  DF         Squares    Mean Square    F Value    Pr > F

   Model                    4     104.3597561     26.0899390      39.93    <.0001

   Error                   27      17.6402439      0.6533424

   Corrected Total         31     122.0000000

Part (B)
                  R-Square     Coeff Var      Root MSE    score2 Mean

                  0.855408      23.09417      0.808296       3.500000
```

```
Part (C)

   Source                     DF       Type I SS     Mean Square    F Value    Pr > F

   score1                      1     102.4000000     102.4000000     156.73    <.0001
   indep1                      3       1.9597561       0.6532520       1.00    0.4080

Part (D)

   Source                     DF     Type III SS     Mean Square    F Value    Pr > F

   score1                      1     51.60975610     51.60975610      78.99    <.0001
   indep1                      3      1.95975610      0.65325203       1.00    0.4080
```

```
                          Example 13.8 Analysis of covariance (ANCOVA)                    3

                                        The GLM Procedure
    Part (E)                          Least Squares Means

                                score2         Standard                  LSMEAN
                  indep1        LSMEAN            Error    Pr > |t|      Number

                  1         3.77743902       0.32644527      <.0001           1
                  2         3.21646341       0.30105039      <.0001           2
                  3         3.75000000       0.28577578      <.0001           3
                  4         3.25609756       0.38132468      <.0001           4


    Part (F)                    Least Squares Means for effect indep1
                                 Pr > |t| for H0: LSMean(i)=LSMean(j)

                                     Dependent Variable: score2

                   i/j              1              2              3              4

                      1                        0.1815         0.9500         0.3733
                      2        0.1815                         0.2096         0.9412
                      3        0.9500         0.2096                         0.3092
                      4        0.3733         0.9412         0.3092


    NOTE: To ensure overall protection level, only probabilities associated with pre-planned
          comparisons should be used.
```

Pages 2 and 3 of Output 13.8 are part and parcel of ANCOVA, although not all results are equally relevant. The MS error (= 0.6533424) and its df (= 27) in **Part (A)** are relevant; they will be referred to later. **Part (B)** depicts four descriptive statistics. The first (**R-Square=0.855408**) describes a strong linear relationship between the dependent variable (<u>score2</u>) and the independent variable (<u>indep1</u>) and the covariate (<u>score1</u>) jointly.

Part (C) tells us that the covariate, <u>score1</u>, is an effective covariate because it accounts for a substantial portion of the sum of squares (**Type I**) in the dependent measure, <u>score2</u>. The substantial sum of squares translates into a large F value (=156.73), significant at $p < 0.0001$. The remaining variance in <u>score2</u> that is explained by <u>indep1</u> is, therefore, negligible ($F = 1.00$, $p = 0.4080$).

The nonsignificant effect of indep1 on score2 is confirmed by comparisons of least squares means (**Part (F)** of page 3). None of these comparisons reaches the α level of 0.05 or even 0.10. **Part (E)** displays the least squares means (or adjusted means) of score2. All are above 3 (ounces or bottles?). Each is further tested against a null hypothesis of zero adjusted mean in the underlying population. All tests yield a highly significant result at $p < 0.0001$. These results indicate that students' drinking during the final week was definitely prevailing in all four housing conditions. The drinking recorded at the end of the semester was evident even after it was adjusted for the amount consumed during the spring break. Too much stress, maybe?

- On the LSMEANS statement, there can be other options besides PDIFF and STDERR. Specifically, the option **ALPHA=** (a small probability, such as 0.10) can be used to specify the confidence level (which equals $1 - p$) of each adjusted mean or difference in a pair of adjusted means. The default is 0.05. The ALPHA= option is specified simultaneously with the PDIFF or the **CL** option. The CL option is similar to the PDIFF option in that the CL option computes a confidence interval for each adjusted mean, whereas the PDIFF option computes the confidence interval for the difference in each pair of adjusted means.

- If you wish to control the Type I error rate in simultaneous tests of adjusted means, you may specify the **ADJUST=** option on the LSMEANS statement, after the slash (/). If ADJUST= **SIDAK**, then the adjusted means are tested by the Dunn-Šidák procedure with a family-wise Type I error controlled at 0.05 (the default) or the level specified by the ALPHA= option. If ADJUST=**DUNNETT**, adjusted means are tested by the Dunnett procedure, which compares each adjusted mean with a reference mean (the default is the adjusted mean of the last group), at a family-wise α level of 0.05 or the level specified by the ALPHA= option.

Example 13.9 Examining ANCOVA Assumptions

The ANCOVA approach comes with a price. It requires (a) that a linear relationship exist between the covariate and the dependent measure and (b) that there be no interaction between the covariate and the independent variable. The first assumption can be checked by drawing a scatter plot based on score1 and score2 and computing a Pearson correlation to determine if the relationship is indeed linear and substantial. The second assumption needs to be examined by a statistical test. This example demonstrates how both assumptions can be examined. Note that the interaction of score1 with indep1 is added to the MODEL statement and the option **SOLUTION** is inserted after the slash (/).

```
/* See Example 13.1 for the DATA step in creating the SAS data set 'design' */

TITLE 'Example 13.9 Examining ANCOVA assumptions';

PROC GLM DATA=design;
     CLASS indep1;
     MODEL score2=score1 indep1 score1*indep1 / SOLUTION;
RUN; QUIT;
```

Output 13.9 Examining ANCOVA Assumptions

```
[Page 1 output is not shown])

                       Example 13.9 Examining ANCOVA assumptions                        2

                                   The GLM Procedure

Dependent Variable: score2   no. of drinks in final week

                                              Sum of
   Source                      DF          Squares     Mean Square    F Value    Pr > F

   Model                        7      105.2498188      15.0356884      21.54    <.0001

   Error                       24       16.7501812       0.6979242

   Corrected Total             31      122.0000000


                   R-Square     Coeff Var      Root MSE    score2 Mean

                   0.862703      23.86910      0.835419       3.500000


   Source                      DF        Type I SS     Mean Square    F Value    Pr > F

   score1                       1      102.4000000     102.4000000     146.72    <.0001
   indep1                       3        1.9597561       0.6532520       0.94    0.4386
   score1*indep1                3        0.8900627       0.2966876       0.43    0.7368 ← NS


   Source                      DF      Type III SS     Mean Square    F Value    Pr > F

   score1                       1      45.24687984     45.24687984      64.83    <.0001
   indep1                       3       0.69189216      0.23063072       0.33    0.8034
   score1*indep1                3       0.89006274      0.29668758       0.43    0.7368


                                                     Standard
        Parameter                Estimate               Error    t Value    Pr > |t|

        Intercept             -1.565217391 B       1.56777135      -1.00      0.3281
        score1 (β weight)      1.130434783 B       0.24635150       4.59      0.0001
        indep1          1      0.752717391 B       1.71398072       0.44      0.6645
        indep1          2      0.731884058 B       1.99250499       0.37      0.7166
        indep1          3     -0.634782609 B       2.05571926      -0.31      0.7601
        indep1          4      0.000000000 B        .                .         .
        score1*indep1 1       -0.067934783 B       0.32296954      -0.21      0.8352
        score1*indep1 2       -0.213768116 B       0.42072528      -0.51      0.6160
        score1*indep1 3        0.269565217 B       0.39210410       0.69      0.4984
        score1*indep1 4        0.000000000 B        .                .         .

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to
      solve the normal equations.  Terms whose estimates are followed by the letter 'B'
      are not uniquely estimable.
```

The *F* test of the interaction effect is, fortunately, not statistically significant. This implies that there is no sufficient evidence in the present data to support an interaction between the covariate (score1) and the independent variable (indep1). In the section where you find "**Parameter**" and "**Estimate**", the label (β weight) is inserted next to score1. This label is meant to draw your attention to the estimate (1.130434783), which is the regression weight of score2 (the dependent variable) regressing on score1. Technically speaking, this regression weight is β_w, which stands for the regression weight that is assumed equal in all treatment conditions. Suffice it to say, the magnitude of β_w suggests a strong and linear relationship between the covariate and the dependent variable.

13.4 How to Write the PROC GLM Codes

Based on the examples presented so far, you probably have recognized that the GLM procedure is more complex than the TTEST procedure, even though both are used to compare means. The GLM procedure is versatile for a variety of experimental designs and linear models. It provides diverse comparison procedures to examine pairwise as well as complex contrasts among means. The GLM procedure consists of eight essential statements. Seven are explained here; the eighth statement, CONTRAST, is explained in **Section 13.5: Tips**. Statements not introduced here can be found from the online documentation at www.sas.com.

```
PROC  GLM        DATA= sas_dataset_name  <options>;
      CLASS      independent_ or blocking_variable(s);
      MODEL      dependent_variable(s) = effects;
      MEANS      main_effects / comparison_procedures <options>;
      LSMEANS    main_effects / <options>;
      REPEATED   repeated_factor(s);
      TEST       H= effects E= error_term;
      BY         classification_variable(s);
```

The first statement, **PROC GLM**, initializes the procedure and specifies the data set to be analyzed. In addition, you may specify the option **MANOVA**. This option requests that the GLM procedure rely on a multivariate method of removing observations from the analysis, namely, the listwise deletion method. In other words, if an observation has a missing value on one or more independent or dependent variables, the SAS system removes such an observation from the analysis. This option is applied in multivariate analyses, such as Example 13.7, or in the interactive mode of data analysis.

The second statement, **CLASS**, is to identify independent or blocking variables in a design. This statement is required; it must precede the MODEL statement.

The third statement, **MODEL**, is to specify an ANOVA design, also a linear model, for the data. On the left side of the equal sign (=), dependent variable(s) are listed. On the right side, effects such as main effects, interactions, blocking effects, nested effects, and covariates are listed. These effects decompose the total sum of squares of the dependent variable. Below are examples of the MODEL syntax for several commonly used designs:

Main-Effect Design

```
MODEL score=a b;   (two-way ANOVA)   or
MODEL score=a b c; (three-way ANOVA)
```

Completely Factorial Design

```
MODEL score=a b a*b;   same as MODEL score=a | b;
(both are two-way)
MODEL score=a | b | c; same as MODEL score=a b c a*b a*c b*c a*b*c;
(both are three-way)
```

Nested Design

```
MODEL score=a c(b) a*c(b)       same as MODEL score=a | c(b);
MODEL score=a c a*c b(a) c*b(a) same as MODEL score=a | b(a) | c;
MODEL score=a(b) c(b) a*c(b)    same as MODEL score=a(b) | c(b);
```

Randomized Block Design

```
MODEL score=a block;
```

It is sometimes necessary to examine differences among group means. This is accomplished by the **MEANS** statement. A variety of comparison procedures are available; each is sensitive to mean differences under a particular circumstance. These procedures are listed after a slash (/). A few other options are likewise listed after the slash. Interaction effects listed on the MEANS statement, before the slash, will not be tested, however; they are described instead in terms of cell means.

Below is a list of comparison procedures and options for the MEANS statement, listed after the slash (/):

BON performs a two-tailed Dunn's procedure based on the Bonferroni inequality.

DUNNETT performs a two-tailed Dunnett's procedure that compares a control group with any other group. The control group is defaulted to the first group. If you wish to change the control group from the first to another, you specify the control group in parentheses as follows:

```
MEANS  drug / DUNNETT (2);
```

According to the statement above, the second group is specified to be the control group of the drug factor. For character factors, single quotes are needed around the group name. For example, the statement below identifies the placebo group as the control group.

```
MEANS  drug / DUNNETT ('placebo');
```

A one-tailed Dunnett's test is also possible with a minor modification of the keyword to **DUNNETTL** or **DUNNETTU.**

DUNNETTL executes a one-tailed Dunnett's test with the alternative hypothesis stating that the experimental group mean is less than the control mean.

DUNNETTU executes a one-tailed Dunnett's test with the alternative hypothesis stating that the experimental group mean is greater than the control mean.

SCHEFFE performs a two-tailed Scheffé procedure. The Scheffé procedure is based on the same *F* distribution as the overall *F* test. So if the overall *F* test is significant at, say, α = 0.05, the Scheffé test will surely find either a pair of means or three or more means to be different at the same α level.

SIDAK performs a two-tailed Dunn-Šidák procedure, based on the *t* distribution.

SNK performs a two-tailed Newman-Keuls' modified *t* test of ordered mean differences.

TUKEY performs a two-tailed Tukey's HSD test.

HOVTEST performs the Levene test of homogeneity of variance.

CLDIFF builds a 95% confidence interval for each pair of means for all comparison procedures, except for the **SNK** procedure. The 95% confidence can be changed using the next option, **ALPHA=**.

ALPHA= a small probability that specifies the α level for carrying out all comparison procedures listed above. The specification also changes the confidence level for the **CLDIFF** option since confidence level = (1 – ALPHA) × 100%.

E= specifies the denominator for all comparison procedures listed above. If omitted, the default is the mean square residual ($MS_{Residual}$).

The fifth statement, **LSMEANS**, tests single or pairs of least-squares means. This statement is relevant to ANCOVA designs and comparisons of adjusted means (i.e., least-squares means) between groups. Two options are illustrated in Example 13.8: **PDIFF** and **STDERR.** The other three options are the following:

ALPHA= a small probability specifies the α level for the test of least-squares means; the default is 0.05.

CL requests the (1 – ALPHA) × 100% confidence level to be constructed for each least-squares mean.

ADJUST=T or BON or SIDAK or TUKEY or DUNNETT requests that a *t* test (specified by **T**), or Bonferroni *t* test (**BON**), or the Dunn-Šidák test (**SIDAK**), or Tukey's HSD test (**TUKEY**), or the DUNNETT test (**DUNNETT**) be applied to pairs of least-squares means.

The sixth statement, **REPEATED**, names a factor for which repeated measures are analyzed by either a univariate or a multivariate approach (see Example 13.7 for an illustration).

The seventh statement, **TEST**, is used to specify effects to form the numerator and the denominator of an *F* ratio. In Example 13.5, it was mentioned that for the 4 × 4 × 4 LS design, two estimates for the variance of sampling errors could be considered. One is the mean square of the three-way interaction and the other is the mean square residuals. The latter was used as a denominator for all *F* tests carried out in Example 13.5. Had we been interested in using the second estimate as the denominator, we would have specified the TEST statement as follows on the next page.

```
TEST H=a b c  E=a*b*c;
```

Finally the last statement, **BY**, serves the same purpose as in all other SAS procedures. It divides the data set into subgroups according to diverse values of the BY variable. Within each subgroup, the same ANOVA design is applied and the same analysis follows accordingly. If more than one BY variable is listed, all possible combinations of the BY variables' values are used in dividing up the entire data set. Be sure to presort the data set in the ascending order of all the BY variables, if the BY statement is included in the GLM procedure. Presorting a data set can be accomplished using the SORT procedure.

13.5 Tips

- How to handle missing or invalid data

By default, PROC GLM does not include observations that have missing information on either the dependent variable(s) or any of the CLASS variables.

When the REPEATED statement is specified to analyze data from a repeated measures design, you are advised to also specify the MANOVA option in the PROC GLM statement.

- What are the statistical assumptions associated with the *F* test conducted in one-way fixed-effects ANOVA?

The *F* test carried out in a one-way fixed-effects ANOVA is closely related to the independent-samples *t* test introduced in Chapter 12. If the one-way linear model presumed for data captures all sources of variations in the dependent variable, the *F* test assumes, first of all, that subjects are randomly selected from their respective populations, or that they are randomly assigned to conditions of the independent variable. Second, the underlying populations are normally distributed. Third, variances of normal populations are assumed to be equal. These assumptions are referred to in the literature as the independence assumption, the normality assumption, and the equal variance assumption.

Beyond the one-way fixed-effects ANOVA, factorial ANOVA designs, randomized block ANOVA designs, LS designs, repeated measure designs, and ANCOVA make additional statistical assumptions. For a detailed discussion of these assumptions and their robustness, refer to Box (1954), Clinch and Keselman (1982), Glass, Peckham, and Sanders (1972), Kirk (1995), Rogan and Keselman (1977), Tan (1982), and Tomarken and Serlin (1986).

- What to do if data do not satisfy the statistical assumptions in one-way fixed-effects ANOVA

For one-way fixed-effects ANOVAs, statisticians in general agree that the independence assumption is not robust to its violation. It is an important assumption because its violation renders the interpretation of the *F* test inexact and biased.

The normality assumption is quite robust, especially when the underlying populations are symmetric and sample sizes are equal and greater than 12 in all conditions. Even if population distributions are asymmetric and/or more peaked or flatter than the normal curve, the normality assumption is still robust as long as the population distributions are shaped the same and sample sizes are equal. One way to check the normality assumption is demonstrated in Chapter 9, Example 9.4.

The equal variance assumption is robust in balanced designs if samples are taken from underlying normal populations in which the ratio of the largest variance to the smallest variance is no more than 3. Unfortunately, this assumption is not robust when the ratio of the largest to the smallest variances exceeds 3, even if equal sample sizes are maintained. Under these conditions, alternative parametric tests, such as the Brown-Forsyth test, exist to compensate for the violation of the equal variance assumption. These alternative parametric tests are discussed and illustrated in Clinch and Keselman (1982).

In the worst possible scenario, in which sample sizes are unequal and terribly small and the populations are far from normal, you can still fall back on nonparametric tests. These are explained in Chapter 14.

- What if the research design is unbalanced?

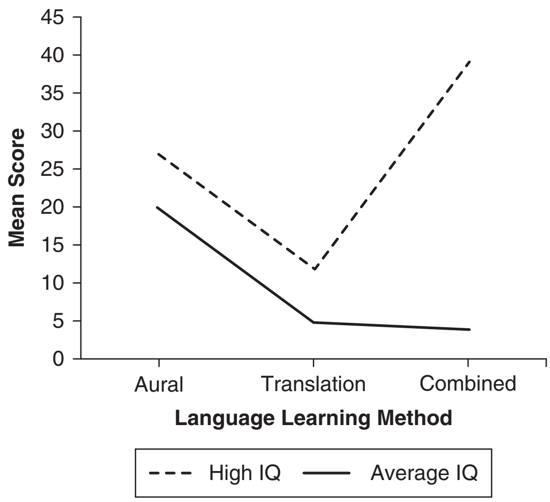
An unbalanced design is a design in which cell sizes are unequal, or some cells have missing observations. For the unbalanced designs, tests of main effects and of interactions are nonorthogonal or statistically dependent. For discussions of these designs and their treatments in SAS, refer to the Four Types of Estimable Functions and the GLM chapters in *SAS/STAT 9.1 User's Guide* (SAS Institute Inc., 2004d) or the online documentation at www.sas.com.

- How to test planned contrasts in PROC GLM

As stated before, PROC GLM is capable of carrying out planned contrasts of main effects and interactions. These planned contrasts are specified by the CONTRAST statement. Suppose a 2 × 3 factorial design includes IQ as the row factor and the method of learning a foreign language as the column factor. The row factor, <u>iq</u>, has two levels, (high and average), and the column factor <u>method</u>, has three levels: the aural method, the translation method, and the combined method. The dependent <u>score</u> is students' comprehension of a passage written in the foreign language they studied. The diagram below may help you grasp the 2 × 3 design and six hypothetical cell means:

	Language Learning		
	Aural	*Translation*	*Combined*
High IQ	27	12	39
Average IQ	20	5	4

The graph below depicts hypothetical means of the six cells:



Based on the design and means graphed above, let's suppose that five orthogonal contrasts are of interest:

$$\psi_1 = \overline{Y}_{\text{High IQ}} - \overline{Y}_{\text{Average IQ}}$$

$$\psi_2 = \overline{Y}_{\text{Aural}} - \overline{Y}_{\text{Translation}}$$

$$\psi_3 = \overline{Y}_{\text{Aural}} + \overline{Y}_{\text{Translation}} - 2 \times \overline{Y}_{\text{Combined}}$$

$$\psi_4 = (\overline{Y}_{\text{Aural}} - \overline{Y}_{\text{Translation}})_{\text{High IQ}} - (\overline{Y}_{\text{Aural}} - \overline{Y}_{\text{Translation}})_{\text{Average IQ}}$$

$$\psi_5 = (\overline{Y}_{\text{Aural}} + \overline{Y}_{\text{Translation}} - 2 \times \overline{Y}_{\text{Combined}})_{\text{High IQ}} - (\overline{Y}_{\text{Aural}} + \overline{Y}_{\text{Translation}} - 2 \times \overline{Y}_{\text{Combined}})_{\text{Average IQ}}$$

The first contrast is a test of the main effect of iq, the second and the third are tests of main effects of method, and the last two are tests of interactions. To implement these planned orthogonal contrasts into PROC GLM, five CONTRAST statements are written as follows:

```
PROC GLM DATA=ortho ORDER=DATA;
     CLASS iq method;
     MODEL score=iq method iq*method;
     CONTRAST   'psy1'    iq            1    -1;
     CONTRAST   'psy2'    method        1    -1    0;
     CONTRAST   'psy3'    method        1     1   -2;
     CONTRAST   'psy4'    iq*method     1    -1    0    -1    1   0;
     CONTRAST   'psy5'    iq*method     1     1   -2    -1   -1   2;
```

Note that each CONTRAST statement is independent of all others; thus, each ends with a semi-colon (;). Each statement is written according to the following syntax:

CONTRAST 'title of the contrast' effect_name
coefficients_to_be_applied_to_group_means;

For a main effect, it is easy to figure out how coefficients are applied to each group (level) under that main effect. Simply multiply successive coefficients, from left to right, with group means that are ordered according to the way data were read. This is the reason why, in the PROC GLM statement, the option ORDER=DATA is included.

It is tricky, however, with interaction effects. Take the iq*method interaction, for example. How does SAS know to multiply –2 in ψ_5 with the mean of the High IQ students in the combined condition? The key lies in the order in which the two variables (or factors) are listed. In the program above, iq precedes method. Therefore, the first three coefficients, namely, 1, 1, and –2, are applied to the high iq group, whereas the last three, –1, –1, and 2, are applied to the average iq group. Within the high iq group, coefficients 1 and 1 are further applied to the first two conditions of method, whereas –2 is applied to the last condition, that is, the combined method. Try using this logic to interpret the coefficients in ψ_4 to make sure that you can write CONTRAST statements for interactions on your own.

After executing the five contrasts, the output shows the following results. Each contrast is tested with 1 and 24 degrees of freedom—the degrees of freedom for the MS error. Four contrasts are statistically significant at α = 0.01, but ψ_4 is not. This nonsignificant result is confirmed by the graph and by the cell mean difference (27 – 12 = 20 – 5).

```
Contrast                    DF   Contrast SS   Mean Square  F Value  Pr > F

 psy1-on iq                  1   2000.833333   2000.833333   800.33  <.0001
 psy2-on method              1   1125.000000   1125.000000   450.00  <.0001
 psy3-on method              1    201.666667    201.666667    80.67  <.0001
 psy4-on iq by method        1      0.000000      0.000000     0.00  1.0000
 psy5-on iq by method        1   1306.666667   1306.666667   522.67  <.0001
```

The CONTRAST statement is applicable to (a) orthogonal contrasts, such as the five tested here, (b) nonorthogonal contrasts, (c) simple or pair-wise contrasts, and (d) complex contrasts of means.

- How to use ODS with the GLM procedure

To use the ODS, you need to know ODS table names corresponding with various portions of the output. Table 13.1 presents selected ODS table names for the GLM procedure and their descriptions.

Table 13.1 Selected ODS Table Names and Descriptions for the GLM Procedure

ODS Table Name	*Description*	*GLM Procedure Statement*
OverallANOVA	Overall ANOVA	(default)
Fitstatistics	R-square, C.V., Root MSE, and dependent variable's mean	(default)
ModelANOVA	ANOVA for model terms	(default)
Means	Group means	MEANS
MCLinesInfo	Multiple comparison information	MEANS / comparison procedure options
MCLines	Multiple comparison output	MEANS / comparison procedure options
MultStat	Multivariate statistics	REPEATED or MANOVA
Epsilons	Greenhouse-Geisser and Huynh-Feldt epsilons	REPEATED
LSMeans	Least-squares means	LSMEANS
Diff	Significance levels for tests of all pairs of least-squares means	LSMEANS / PDIFF

Based on the list of ODS table names, you may select certain results to be displayed in the Output window. For example, the following program selects the BON procedure's result of Example 13.1 to be included in the output:

```
ODS  SELECT Bon.MCLinesInfo Bon.MCLines;
PROC GLM DATA=design;
     CLASS indep1;
     MODEL score1=indep1;
     MEANS indep1 / BON TUKEY;
RUN;
```

Likewise, you may select certain result(s) to be exported as a SAS data set. For example, the following program exports R-square, C.V., Root MSE, and dependent variable's mean of Example 13.1 to the SAS data set fit:

```
ODS  OUTPUT FitStatistics = fit;
PROC GLM DATA=design;
     CLASS indep1;
     MODEL score1=indep1;
     MEANS indep1 / BON TUKEY;
RUN;
```

Furthermore, you may select certain results to be saved in file formats other than the SAS standard output. For example, the following program saves the output of Example 12.1 in HTML format in its default style:

```
ODS HTML BODY = 'd:\result\Example13_1Body.html'
         CONTENTS = 'd:\result\Example13_1TOC.html'
         PAGE = 'd:\result\Example13_1Page.html'
         FRAME = 'd:\result\Example13_1Frame.html';

PROC GLM DATA=design;
     CLASS indep1;
     MODEL score1=indep1;
     MEANS indep1 / BON TUKEY;
RUN;

ODS HTML CLOSE;
RUN;
```

For additional information about the ODS feature, consult with *SAS 9.1.3 Output Delivery System: User's Guide* (SAS Institute Inc., 2006c) and *SAS/STAT 9.1 User's Guide* (SAS Institute Inc., 2004d) or the online documentation at www.sas.com.

13.6 Summary

Haven't you felt like you have had enough of ANOVA? Almost! The ANOVA technique is versatile for testing population mean differences, and so is the GLM procedure—a comprehensive tool for handling a variety of ANOVA designs. The null hypothesis tested in these designs is always the same: that population means are equal. In other words, there is no effect of any kind. The alternative hypothesis states that some means are unequal. The statistic conceptualized by Sir R. A. Fisher to test the null hypothesis is an *F* value. The *F* value is a ratio of two estimates. These two estimates should give the same variance of sampling errors under the null hypothesis. Under the alternative hypothesis, though, the numerator should be larger than the denominator because it contains a portion that reflects the effect being tested under the null hypothesis.

Once the null hypothesis is rejected by an *F* test at a preset α level, one concludes that some means are most likely different from each other. At this point, it is necessary to apply a comparison procedure to pinpoint the specific source of differences among means. PROC GLM provides many such procedures for testing pairs of means. All are performed as a two-tailed test, except for the DUNNETT procedure, which can be performed as a one-tailed test.

If an ANOVA design is balanced, PROC ANOVA can also be specified to test null hypotheses and compare mean differences. And the syntax illustrated in this chapter is equally valid for the ANOVA procedure. There are, however, differences between ANOVA and GLM procedures. In the case of ANCOVA, the GLM procedure can treat a continuous variable as an independent variable, whereas the ANOVA procedure cannot. The GLM procedure provides the CONTRAST statement for testing planned comparisons of main effects and of interactions. These planned comparisons can be complex, based on three or more means. They can be orthogonal as well as nonorthogonal. Yet the CONTRAST statement is not available in the ANOVA procedure, although PROC ANOVA is efficient and versatile for analyzing data collected from a balanced ANOVA design.

13.7 Exercises

1. Four department stores, Macy's, J. C. Penney, Sears, and Target, were selected for a marketing research study of their advertising success. Advertising success was operationally defined as the number of items purchased by four typical customers randomly selected at each store on the second Saturday in July. The following data represent their purchasing behavior:

Subject	*Macy's*	*J. C. Penney*	*Sears*	*Target*
1	3	0	1	4
2	7	2	3	6
3	5	0	4	2
4	5	10	8	8

a. What is the average number of items purchased by all customers?
b. What are the values of $MS_{between}$ and MS_{within}?
c. Is there any significant difference in the number of items purchased by customers at these four stores?
d. Use the Tukey's method to assess the significance in the number of items bought at Macy's versus J. C. Penney. Write a sentence to help your grandma understand this statistical result.

2. A curious and bright graduate student carried out an investigation of a possible link between the size and wall colors of professors' offices and professors' research productivity. She constructed a reliable and valid measure to quantify the productivity and used it to gather the following data; the higher the number, the greater was the professor's productivity:

		Room Color			
		Peach	*Cream*	*Gray*	*Blue*
Room Size	Small	71 80	50 63	104 112	110 105
	Medium	175 164	159 152	133 128	154 141
	Large	105 103	109 113	79 83	66 58

a. What is the average productivity by professors located in gray offices?
b. If the president of the unnamed university wished to standardize all professors' offices, what size of offices should this graduate student recommend?
c. Overall, which office wall color is most helpful to professors' productivity?
d. Does size of offices interact with room color in affecting the professors' research productivity? If so, how strong is the interaction?
e. If your answer to (d) above is yes, which combination of room color and size is most conducive to professors' productivity and which combination is the least?

3. A teacher wants to know if computerized instruction is better than the traditional method for teaching elementary school students. After applying these two methods in two different classes for one semester, the teacher administered tests in three subjects, arithmetic, arts, and reading, and obtained the following scores:

	Computerized		*Traditional*	
	Boys	*Girls*	*Boys*	*Girls*
Arithmetic	85, 70, 90, 82, 63, 84	68, 72, 65, 72, 79, 80	95, 89, 92, 66, 75, 60	75, 82, 78, 69, 70, 75
Arts	77, 89, 69, 82, 70, 87	92, 65, 75, 83, 82, 78	92, 88, 86, 70, 96, 60	77, 82, 79, 85, 72, 80
Reading	68, 75, 85, 92, 66, 80	74, 82, 76, 93, 82, 87	72, 74, 69, 85, 60, 83	82, 86, 77, 72, 74, 88

Answer the following questions based on these data:

a. Do students score differently in the three subjects? (*Hint*: One-way ANOVA)
b. Apply Tukey's procedure to examine differences in three subjects.
c. Do students score differently under two teaching methods? (*Hint*: One-way ANOVA)
d. Do boys and girls score differently?
e. Is there an interaction between and among teaching methods, subject matters, and gender? (*Hint*: Three-way ANOVA)

4. A group of young children has recently been diagnosed as severely depressed beyond the normal acceptable level. A study is therefore proposed and funded by the First Lady's Pocket Grant to investigate whether three antidepression drugs can improve children's depression. Three hospitals are randomly selected (a_1, a_2, a_3) to administer these drugs (t_1, t_2, t_3) to depressed children who come from either single-parent homes (b_1), divorced-then-remarried homes (b_2), or intact families (b_3). Data show the following trend (the higher the score, the better is the drug's effect):

	b_3	b_2	b_1
a_2	6 (t_1)	7 (t_2)	8 (t_3)
a_1	2 (t_2)	1 (t_3)	5 (t_1)
a_3	0 (t_3)	4 (t_1)	1 (t_2)

Perform a suitable statistical analysis on these data and summarize your results in an ANOVA table with $\alpha = 0.05$. Write a sentence to interpret the results.

5. Eight graduate students living on midwestern university campus were surveyed with regard to the government policy on phone wiretapping as a mechanism against terrorism. The survey was carried out at two times: on September 11, 2007, and shortly after Thanksgiving, also in 2007. The instrument used to collect data asked students about their attitude toward the necessity of such a government policy to fight against terrorism. The higher the score, the more supportive was the attitude. In addition, the researcher also collected information from each student regarding his or her stand on a national gun control law. Data exhibit the following trend:

Subject ID	*Group*	*On September 11, 2007*	*After Thanksgiving, 2007*
1	For gun control	4	7
2	For gun control	7	8
3	For gun control	3	5
4	For gun control	2	5
5	Against gun control	10	11
6	Against gun control	8	10
7	Against gun control	9	9
8	Against gun control	7	5

What are different analysis strategies that a data analyst can employ to find out if differences in students' attitudes could be explained by their stands on the national gun control law, time of the survey, and an interaction of these two?

6. In a computer literacy class, the instructor wished to determine if students' learning was different due to different teaching methods. Three methods (encouragement, practice and drill, and self-directed learning) were used in three classes. To better account for the teaching method effect, the instructor decided to measure students' IQ as a covariate. He administered an IQ test at the beginning of the study and a computer literacy test after the study was concluded. Is there any difference in students' computer literacy from three classes after IQ is taken into consideration?

Encouragement		*Practice and Drill*		*Self-Directed Learning*	
Test Score	*IQ*	*Test Score*	*IQ*	*Test Score*	*IQ*
16	124	17	137	13	112
15	123	15	116	11	104
14	115	18	148	14	111
15	120	17	135	11	105
17	136	19	147	12	103
13	104	18	135	14	113

13.8 Answers to Exercises

1. a. The average number of items purchased by all subjects (also customers) = 4.25.
   b. $MS_{between}$ = 3.66666667 and MS_{within} = 10.16666667.
   c. No, because the *F* test, $F(3, 12) = 0.36$, $p = 0.7825$ is not statistically significant at the α level of 0.05.
   d. Tukey's test is used to examine if pairs of means are statistically significantly different from each other. In this case, the difference between these two means must be at least 6.6935 (= HSD = MSD) in order to be considered statistically significant. Since the mean difference between Macy's (= 5) and J. C. Penney (= 3) is 2, they are not considered significantly different from each other. Therefore, grandma, customers at Macy's and J. C. Penney bought approximately the same amount of stuff on a Saturday in July. Where do you want me to take you to shop?

2. a. 106.50
   b. The medium-sized offices should be recommended because these offices yielded the highest mean level of productivity (= 150.75).
   c. The color "peach" should be recommended for office walls because professors in peach-colored offices produced the most research (mean = 116.333), compared with professors in offices painted in cream, gray, or blue.
   d. Yes, the room size did interact statistically significantly with room colors in affecting professors' research productivity, $F(6, 12) = 39.86$, $p < 0.0001$. ω^2 for the interaction effect = 0.9067, effect size = 3.117. Statistical power for detecting the significant interaction effect is nearly 100%. Both ω^2 and the statistical power were obtained by hand calculation, not from SAS directly.
   e. The most conducive combination is a peach-colored and medium-sized office (mean productivity = 169.5); the least is a cream-colored and small office (mean productivity = 56.5).

3. a. No, because the *F* test of the subject factor, $F(2, 69) = 0.83$, $p = 0.4404$ is not statistically significant at the α level of 0.05.

   b. Tukey's test is used to examine if pairs of means are statistically significantly different from each other. To be considered statistically significant, the observed difference between any two group means should be at least as large as 6.2211 (= HSD = MSD). Results from Tukey's test indicate that none of the pairwise comparisons is statistically significant. These results are consistent with the overall *F* test.

   c. No, because the *F* test of the method factor, $F(1, 70) < 0.01$, $p = 0.9586$ is not statistically significant at the α level of 0.05.

   d. No, because the *F* test of the sex factor, $F(1, 70) = 0.06$, $p = 0.3313$ is not statistically significant at the α level of 0.05.

   e. No, because the result of the *F* test of the three-way interaction among subject, method, and sex, $F(2, 60) = 0.12$, $p = 0.8878$ is not statistically significant at the α level of 0.05. Furthermore, none of the two-way interactions is statistically significant at $\alpha = 0.05$: (i) subject*method, $F(2, 60) = 0.49$, $p = 0.6132$; (ii) subject*sex, $F(2, 60) = 2.10$, $p = 0.1313$; and (iii) method*sex, $F(1, 60) = 0.00$, $p = 1.0$.

4. This research project calls for the application of the Latin-square (LS) design for which factors a and b are nuisance variables and factor t is the treatment factor. According to this LS design, the SS total is decomposed as follows:

Source	*SS*	*df*	*MS*	*F*	*p*
a	48.222	2	24.111	7.75	.1143
b	6.222	2	3.111	1.00	.5000
t	6.889	2	3.444	1.11	.4746
Error	6.222	2	3.111		
Total	67.556	8			

   Because the *F* test of the t main effect is not statistically significant, it is concluded that three antidepression drugs did not produce noticeable differences in improving children's depression after controlling for differences in hospitals and family backgrounds.

5. Strategy A: Apply an $SPF_{p.q}$ design for which the Group variable is the between-block factor and the two measures as levels of the within-block factor.

   Strategy B: Perform a one-way ANOVA using the Group variable as the independent variable and the difference between the two measures as the dependent variable.

Strategy C: Perform two one-way ANOVAs using the Group variable as the independent variable and each of the two measures as the dependent variable. Discuss any discrepancy in findings due to the time of the measures.

Strategy D: Perform a one-way ANCOVA for which the first measure, taken on September 11, 2007, is the covariate and the second measure, taken after Thanksgiving 2007, is the dependent variable. The Group variable is the independent variable.

Strategy E: Apply the nonparametric test of Strategy B.

Strategy F: Apply the nonparametric test of Strategy C.

6. The ANCOVA result is summarized as follows:

Source	*SS*	*df*	*MS*	*F*	*p*
IQ	48.2862	1	48.2862	89.48	<0.0001
Method	2.0633	2	1.0317	1.91	0.2096
Error	4.3172	8	0.5396		
Total	54.6667	11			

From the ANCOVA result, we can conclude that IQ is an effective covariate, $F(1, 8) = 89.48$, $p < 0.0001$. After adjusting for IQ, the effect of methods is not statistically significant, $F(2, 8) = 1.91$, $p = 0.2096$. However, the appropriateness of using ANCOVA to analyze data for this study is questionable because IQ is found to interact with the method, $F(2, 6) = 6.59$, $p = 0.0306$.