Chapter 42

The TTEST Procedure

ABSTRACT 1633
INTRODUCTION 1633
SPECIFICATIONS 1634
  PROC TTEST Statement 1634
  BY Statement 1634
  CLASS Statement 1635
  VAR Statement 1635
DETAILS 1635
  Missing Values 1635
  Computational Method 1635
    The t Statistic 1635
    The Folded Form F Statistic 1636
    The Approximate t Statistic 1636
    The Cochran and Cox Approximation 1636
    Satterthwaite's Approximation 1636
  Printed Output 1636
EXAMPLES 1637
  Example 1: Comparing Group Means 1637
  Example 2: Paired Comparisons Using PROC MEANS 1638
REFERENCES 1639

ABSTRACT

The TTEST procedure performs a two sample t-test for testing the hypothesis that
the means of two groups of independent and normally distributed observations
are equal.

INTRODUCTION

PROC TTEST computes sample means for each of two groups of observations
identified by levels of a CLASS variable and tests the hypothesis that the population
means are the same. This analysis can be considered a special case of a one-
way analysis of variance with two levels of classification.

PROC TTEST computes the t statistic based on the assumption that the vari-
ces of the two groups are equal, and it computes an approximate t based on
the assumption that the variances are unequal (the Behrens-Fisher problem). The
degrees of freedom and probability level are given for each; Satterthwaite's (1946)
approximation is used to compute the degrees of freedom associated with the
approximate t. In addition, you can request the Cochran and Cox (1950) approxi-
mation of the probability level for the approximate t. An F (folded) statistic is com-
punted to test for equality of the two variances (Steel and Torrie 1980).

The TTEST procedure is not designed for paired comparisons. See Example 2 for
a description of a method for a paired-comparisons t-test.
Note that the underlying assumption of the t-test is that the observations are random samples drawn from two independent and normally distributed populations. If the assumptions for the t-test are not satisfied, you should analyze your data using PROC NPAR1WAY.

SPECIFICATIONS

The following statements are available in PROC TTEST:

```
PROC TTEST <options>; 
   CLASS variable; 
   BY variables; 
   VAR variables; 
```

required statements
optional statements

No statement may be used more than once. There is no restriction on the order of the statements after the PROC statement. The CLASS statement is required.

PROC TTEST Statement

```
PROC TTEST <COCHRAN> <DATA=SAS-data-set>;
```

The following options can appear in the PROC TTEST statement:

- **COCHRAN**
  requests the Cochran and Cox (1950) approximation of the probability level of the approximate t statistic for the unequal variances situation.

- **DATA=SAS-data-set**
  names the SAS data set for the procedure to use. By default, PROC TTEST uses the most recently created SAS data set.

BY Statement

```
BY variables;
```

A BY statement can be used with PROC TTEST to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Use the SORT procedure with a similar BY statement to sort the data.
- Use the BY statement options NOTSORTED or DESCENDING in the BY statement for the TTEST procedure. As a cautionary note, the NOTSORTED option does not mean that the data are unsorted, but rather means that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Use the DATASETS procedure (in base SAS software) to create an index on the BY variables.

For more information on the BY statement, see the discussion in SAS Language: Reference, Version 6, First Edition. For more information on the DATASETS procedure, see the discussion in the SAS Procedures Guide, Version 6, Third Edition.
CLASS Statement

CLASS variable;

A CLASS statement giving the name of the grouping variable must accompany the PROC TTEST statement. The grouping variable must have two, and only two, levels. PROC TTEST divides the observations into the two groups for the t-test using the levels of this variable.

You can use either a numeric or a character variable in the CLASS statement. If you use a character variable longer than 16 characters, the values are truncated and a warning message is issued.

Class levels are determined from the formatted values of the CLASS variable. Thus, you can use formats to define group levels. See the discussion of the FORMAT procedure, the FORMAT statement, and "SAS Informats and Formats" in SAS Language: Reference, Version 6, First Edition.

VAR Statement

VAR variables;

The VAR statement names the variables whose means are to be compared. If the VAR statement is omitted, all numeric variables in the input data set (except a numeric variable appearing in the CLASS statement) are included in the analysis.

DETAILS

Missing Values

An observation is omitted from the calculations if it has a missing value for either the CLASS variable or for the variable to be tested.

If more than one variable is listed in the VAR statement, an observation is included in calculations for all variables for which the observation has nonmissing values. In other words, if an observation has a missing value for one of the variables, the observation is omitted only from the calculations for that variable. The observation is still included in calculations for other variables.

Computational Method

The t Statistic

The $t$ statistic for testing the equality of means from two independent samples with $n_1$ and $n_2$ observations is

$$ t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{s^2(1/n_1 + 1/n_2)}} $$

where $s^2$ is the pooled variance

$$ s^2 = \frac{[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]}{(n_1 + n_2 - 2)} $$

and where $s_1^2$ and $s_2^2$ are the sample variances of the two groups. The use of this $t$ statistic depends on the assumption that $\sigma_1^2 = \sigma_2^2$, where $\sigma_1^2$ and $\sigma_2^2$ are the population variances of the two groups.