Chapter 43
The MULTTEST Procedure

Chapter Table of Contents

OVERVIEW 2313

GETTING STARTED 2314
Drug Example 2314

SYNTAX 2317
PROC MULTTEST Statement 2317
BY Statement 2321
CLASS Statement 2321
CONTRAST Statement 2322
FREQ Statement 2323
STRATA Statement 2323
TEST Statement 2324

DETAILS 2325
Statistical Tests 2325
p-Value Adjustments 2332
Missing Values 2336
Computational Resources 2337
Output Data Sets 2337
Displayed Output 2339
ODS Table Names 2339

EXAMPLES 2340
Example 43.1 Cochran-Armitage Test with Permutation Resampling 2340
Example 43.2 Freeman-Tukey and *t*-Tests with Bootstrap Resampling 2343
Example 43.3 Peto Test 2347
Example 43.4 Fisher Test with Permutation Resampling 2349
Example 43.5 Inputting Raw *p*-Values 2353

REFERENCES 2354

Chapter 43
The MULTTEST Procedure

Overview

The MULTTEST procedure addresses the multiple testing problem. This problem arises when you perform many hypothesis tests on the same data set. Carrying out multiple tests is often reasonable because of the cost of obtaining data, the discovery of new aspects of the data, and the many alternative statistical methods. However, a negative feature of multiple testing is the greatly increased probability of declaring false significances.

For example, suppose you carry out 10 hypothesis tests at the 5% level, and you assume that the distributions of the p-values from these tests are uniform and independent. Then, the probability of declaring a particular test significant under its null hypothesis is 0.05, but the probability of declaring at least 1 of the 10 tests significant is 0.401. If you perform 20 hypothesis tests, the latter probability increases to 0.642. These high chances illustrate the danger in multiple testing.

PROC MULTTEST approaches the multiple testing problem by adjusting the p-values from a family of hypothesis tests. An adjusted p-value is defined as the smallest significance level for which the given hypothesis would be rejected, when the entire family of tests is considered. The decision rule is to reject the null hypothesis when the adjusted p-value is less then α; in most cases, this procedure controls the *familywise error rate* at or below the α level. PROC MULTTEST offers the following p-value adjustments:

- Bonferroni
- Sidak
- stepdown
- Hochberg's method
- Hochberg and Benjamini's method
- bootstrap
- permutation

The Bonferroni and Sidak adjustments are simple functions of the raw p-values. They are computationally quick, but they can be too conservative. Stepdown methods remove some conservativeness, as do the step-up methods of Hochberg (1988). The bootstrap and permutation adjustments resample the data with and without replacement, respectively, to approximate the distribution of the minimum p-value of all tests. This distribution is then used to adjust the individual raw p-values. The bootstrap and permutation methods are computationally intensive but appealing in that,

unlike the other methods, correlations and distributional characteristics are incorporated into the adjustments (Westfall and Young 1989, 1993).

PROC MULTTEST handles data arising from a multivariate one-way ANOVA model, possibly stratified, with continuous and discrete response variables; it can also accept raw p-values as input data. You can perform a *t*-test for the mean for continuous data and the following statistical tests for discrete data:

- Cochran-Armitage (CA) linear trend test
- Freeman-Tukey (FT) double arcsine test
- Peto (PETO) mortality-prevalence (log-rank) test
- Fisher (FISHER) exact test

The CA and PETO tests have exact versions that use permutation distributions and asymptotic versions that use an optional continuity correction. Also, with the exception of the FISHER test, you can use a stratification variable to construct Mantel-Haenszel type tests. All of the previously mentioned tests can be one- or two-sided.

As in the GLM procedure, you can specify linear contrasts that compare means or proportions of the treated groups. The output contains summary statistics and regular and multiplicity-adjusted p-values. You can create output data sets containing raw and adjusted p-values, test statistics and other intermediate calculations, permutation distributions, and resampling information.

Getting Started

Drug Example

Suppose you conduct a small study to test the effect of a drug on 15 subjects. You randomly divide the subjects into three balanced groups receiving 0 mg, 1 mg, and 2 mg of the drug, respectively. You carry out the experiment and record the presence or absence of 10 side effects for each subject. Your data set is as follows:

```
data Drug;
   input Dose$ SideEff1-SideEff10;
   datalines;
0MG  0   0   1   0   0   1   0   0   0   0
0MG  0   0   0   0   0   0   0   0   0   1
0MG  0   0   0   0   0   0   0   0   1   0
0MG  0   0   0   0   0   0   0   0   0   0
0MG  0   1   0   0   0   0   0   0   0   0
1MG  1   0   0   1   0   1   0   0   1   0
1MG  0   0   0   1   1   0   0   1   0   1
1MG  0   1   0   0   0   0   1   0   0   0
1MG  0   0   1   0   0   0   0   0   0   1
1MG  1   0   1   0   0   0   0   1   0   0
2MG  0   1   1   1   0   1   1   1   0   1
2MG  1   1   1   1   1   1   0   1   1   0
```

```
2MG  1   0   0   1   0   1   1   0   1   0
2MG  0   1   1   1   1   0   1   1   1   1
2MG  1   0   1   0   1   1   1   0   0   1
;
```

The increasing incidence of 1s for higher dosages in the preceding data set provides an initial visual indication that the drug has an effect. To explore this statistically, you decide to perform an analysis in which the possibility of side effects increases linearly with drug level. You can analyze the data for each side effect separately, but you are concerned that, with so many tests, there may be a high probability of incorrectly declaring some drug effects significant. You want to correct for this multiplicity problem in a way that accounts for the discreteness of the data and for the correlations between observations on the same unit.

PROC MULTTEST addresses these concerns by processing all of the data simultaneously and adjusting the p-values. The following statements perform a typical analysis:

```
proc multtest bootstrap nsample=20000 seed=41287 notables pvals;
   class Dose;
   test ca(SideEff1-SideEff10);
   contrast 'Trend' 0 1 2;
run;
```

This analysis uses the BOOTSTRAP option to adjust the p-values. The NSAMPLE= option requests 20,000 samples for the bootstrap analysis, and the starting seed for the random number generator is 41287. The NOTABLES option suppresses the display of summary statistics for each side effect and drug level combination.

The CLASS statement is used to specify the grouping variable, Dose. The CA(SIDEEFF1-SIDEEFF10) specification in the TEST statement requests a Cochran-Armitage linear trend test for all 10 characteristics. The CONTRAST statement gives the coefficients for the linear trend test.

The results from this analysis are as follows.

The Multtest Procedure

Model Information

Description	Value
Test for discrete variables:	Cochran-Armitage
Z-score approximation used:	Everywhere
Continuity correction:	0
Tails for discrete tests:	Two-tailed
Strata adjustment?	No
P-value adjustment:	Bootstrap
Number of resamples:	20000
Seed:	41287

Figure 43.1. Output Summary for the MULTTEST Procedure

Figure 43.1 describes the statistical tests performed by PROC MULTTEST. For this example, PROC MULTTEST carries out a two-tailed Cochran-Armitage linear trend test with no continuity correction or strata adjustment. This test is performed on the raw data and on 20,000 bootstrap samples.

The Multtest Procedure

Contrast Coefficients

	Dose		
Contrast	0MG	1MG	2MG
Trend	0	1	2

Figure 43.2. Coefficients Used in the MULTTEST Procedure

Figure 43.2 displays the coefficients for the Cochran-Armitage test. They are 0, 1, and 2, as specified in the CONTRAST statement.

The Multtest Procedure

p-Values

Variable	Contrast	Raw	Bootstrap
SideEff1	Trend	0.0519	0.3471
SideEff2	Trend	0.1949	0.8388
SideEff3	Trend	0.0662	0.5232
SideEff4	Trend	0.0126	0.0937
SideEff5	Trend	0.0382	0.2438
SideEff6	Trend	0.0614	0.4455
SideEff7	Trend	0.0095	0.0540
SideEff8	Trend	0.0519	0.3471
SideEff9	Trend	0.1949	0.8388
SideEff10	Trend	0.2123	0.9002

Figure 43.3. Summary of p-values for the MULTTEST Procedure

Figure 43.3 lists the p-values for the drug example. The Raw column lists the p-values for the Cochran-Armitage test on the original data, and the Bootstrap column provides the bootstrap adjustment of the raw p-values.

Note that the raw p-values lead you to reject the null hypothesis of no linear trend for 3 of the 10 characteristics at the 5% level and for 7 of the 10 characteristics at the 10% level. The bootstrap p-values, however, lead to this conclusion for 0 of the 10 characteristics at the 5% level and only 2 of the 10 characteristics at the 10% level. The bootstrap adjustment gives the probability of observing a p-value as extreme as each given p-value, considering all ten tests simultaneously. This adjustment incorporates the correlation of the raw p-values, the discreteness of the data, and the multiple testing problem. Failure to account for these issues can certainly lead to misleading inferences for these data.

Syntax

The following statements are available in PROC MULTTEST.

PROC MULTTEST < *options* > **;**
BY *variables* **;**
CLASS *variable* **;**
CONTRAST *'label' values* **;**
FREQ *variable* **;**
STRATA *variable* **;**
TEST *name (variables* < */ options* > *)* **;**

Items within angle brackets (< >) are optional, and statements following the PROC MULTTEST statement can appear in any order. The CLASS and TEST statements are required. The syntax of each statement is described in the following section in alphabetical order after the description of the PROC MULTTEST statement.

PROC MULTTEST Statement

PROC MULTTEST < *options* > **;**

You can specify the following options in the PROC MULTTEST statement.

BONFERRONI
BON
specifies that the Bonferroni adjustments (number of tests $\times$ p-value) be computed for each test. These adjustments can be extremely conservative and should be viewed with caution. When exact tests are specified via the PERMUTATION= option in the TEST statement, the actual permutation distributions are used, resulting in a much less conservative version of this procedure (Westfall and Wolfinger 1997).

BOOTSTRAP
BOOT
specifies that the p-values be adjusted using the bootstrap method to resample vectors (Westfall and Young 1993). Resampling is performed with replacement and independently within levels of the STRATA variable. Continuous variables are mean-centered by default prior to resampling. The BOOTSTRAP option is not allowed with the PETO test for theoretical reasons.

If the PERMUTATION= option is used with the CA test, the exact permutation distribution is recomputed for each bootstrap sample. **Caution:** This can be very time-consuming. It is preferable to use permutation resampling when permutation base tests are used.

CENTER
requests that continuous variables be mean-centered prior to resampling. The default action is to mean-center for bootstrap resampling and not to mean-center for permutation resampling.

DATA=*SAS-data-set*
names the input SAS data set to be used by PROC MULTTEST. The default is to use the most recently created data set.

FDR
requests adjusted p-values using the method of Benjamini and Hochberg (1995). These p-values do not control the familywise error rate, but they do control the false discovery rate in some cases.

HOC
requests adjusted p-values using Hochberg's (1988) step-up Bonferroni method.

HOLM
is an alias for the STEPBON adjustment.

NOCENTER
requests that continuous variables not be mean-centered prior to resampling. The default action is to mean-center for bootstrap resampling and not to mean-center for permutation resampling.

NOPRINT
suppresses the normal display of results. Note that this option temporarily disables the Output Delivery System (ODS); see Chapter 15, "The Output Delivery System," for more information.

NOTABLES
suppresses display of the "Discrete Variable Tabulations" and "Continuous Variable Tabulations" tables.

NOZEROS
suppresses display of tables having zero occurrences for all CLASS levels.

NSAMPLE= *number*
N= *number*
specifies the number of resamples for use with the BOOTSTRAP and PERMUTATION options; it is assumed to be 20,000 by default. Large values of *number* (20,000 or more) are usually recommended for accuracy, but long execution times may result, particularly with large data sets.

ORDER=DATA | FORMATTED | FREQ | INTERNAL
specifies the sorting order for the levels of the CLASS variable. This ordering determines which parameters in the model correspond to each level in the data, so the ORDER= option may be useful when you use CONTRAST statements.

The default is ORDER=FORMATTED, and its behavior has been modified for Version 8. Now, for numeric variables for which you have supplied no explicit format (that is, for which there is no corresponding FORMAT statement in the current PROC MULTTEST run or in the DATA step that created the data set), the levels are ordered

by their internal (numeric) value. In releases previous to Version 8, numeric class levels with no explicit format were ordered by their BEST12. formatted values. In order to revert to the previous method, you can specify this format explicitly for the CLASS variable. The change was implemented because the former default behavior for ORDER=FORMATTED often resulted in levels not being ordered numerically and required you to use an explicit format or to specify ORDER=INTERNAL to get the more natural ordering.

The following table shows how PROC MULTTEST interprets values of the ORDER= option.

Value of ORDER=	Levels Sorted By
DATA	order of appearance in the input data set
FORMATTED	external formatted value, except for numeric variables with no explicit format, which are sorted by their unformatted (internal) value
FREQ	descending frequency count; levels with the most observations come first in the order
INTERNAL	unformatted value

For FORMATTED and INTERNAL, the sort order is machine dependent. For more information on sorting order, see the chapter on the SORT procedure in the *SAS Procedures Guide* and the discussion of BY-group processing in *SAS Language Reference: Concepts*.

OUT=*SAS-data-set*
names the output SAS data set containing variable names, contrast names, intermediate calculations, and all associated p-values.

OUTPERM=*SAS-data-set*
names the output SAS data set containing entire permutation distributions (upper-tail probabilities) for all tests when the PERMUTATION= option is used. **Caution:** This data set can be very large.

OUTSAMP=*SAS-data-set*
names the output SAS data set containing information from the resampled data sets when resampling is performed. **Caution:** This data set can be very large.

PDATA=*SAS-data-set*
names an input SAS data set containing the variable raw_p with observations that consist of raw p-values. The MULTTEST procedure adjusts the collection of raw p-values for multiplicity. The resampling-based adjustments cannot be performed using this type of data input, but all other adjustments can be performed. Output from PROC MULTTEST is contained in the OUT= data set when you specify the PDATA= input form, so you must use the OUT= option to obtain the results in this case.

PERMUTATION
PERM
specifies adjusted p-values in identical fashion as the BOOTSTRAP option, with the exception that PROC MULTTEST resamples without replacement rather than with replacement. Resampling is performed independently within levels of the STRATA variable. Continuous variables are not mean-centered prior to resampling. The PERMUTATION option is not allowed with the PETO test for theoretical reasons.

PVALS
requests that a summary table of raw and adjusted p-values be included.

SEED= *number*
S= *number*
specifies the initial seed for the random number generator used for resampling. The value for *number* must be a positive integer; the computer clock time is the default. For more details about seed values, refer to *SAS Language Reference: Concepts*.

SIDAK
SID
specifies that the Sidak adjustments be computed for each test. These adjustments take the form

$$1 - (1 - p)^n$$

where p is the raw p-value and n is the number of tests. These are slightly less conservative than the Bonferroni adjustments, but they still should be viewed with caution. When exact tests are specified via the PERMUTATION= option in the TEST statement, the actual permutation distributions are used, resulting in a much less conservative version of this procedure (Westfall and Wolfinger 1997).

STEPBON
requests adjusted p-values using the stepdown Bonferroni method of Holm (1979).

STEPBOOT
requests that adjusted p-values be computed using bootstrap resampling as described under the BOOTSTRAP option, but in stepdown fashion.

STEPPERM
requests that adjusted p-values be computed using permutation resampling as described under the PERMUTATION option, but in stepdown fashion.

STEPSID
requests adjusted p-values using the Sidak method as described in the SIDAK option, but in stepdown fashion.

BY Statement

BY *variables* **;**

You can specify a BY statement with PROC MULTTEST 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. The *variables* are one or more variables in the input data set.

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

- Sort the data using the SORT procedure with a similar BY statement.
- Specify the BY statement option NOTSORTED or DESCENDING in the BY statement for the MIXED procedure. The NOTSORTED option does not mean that the data are unsorted but rather 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.
- Create an index on the BY variables using the DATASETS procedure (in base SAS software).

Since sorting the data changes the order in which PROC MULTTEST reads observations, this can affect the sorting order for the levels of the CLASS variable if you have specified ORDER=DATA in the PROC MULTTEST statement. This, in turn, affects specifications in the CONTRAST statements.

For more information on the BY statement, refer to the discussion in *SAS Language Reference: Concepts*. For more information on the DATASETS procedure, refer to the discussion in the *SAS Procedures Guide*.

CLASS Statement

CLASS *variable***;**

The CLASS statement is required. It declares a single variable (character or numeric) used to identify the groups for the analysis. For example, if the variable Treatment defines different levels of a treatment, then the statement is

```
class Treatment;
```

The order of the CLASS levels used by PROC MULTTEST correspond to their formatted values; this order can be changed with the ORDER= option in the PROC MULTTEST statement. You need to be aware of this feature when using the CONTRAST statement. You should check the “Contrast Coefficients” table to verify that the appropriate order is used.

CONTRAST Statement

CONTRAST *'label' values* **;**

This statement is used to identify tests between the levels of the CLASS variable. The *label* is a string naming the contrast; it contains a maximum of 21 characters. The *values* are scoring coefficients across the CLASS variable levels.

You can specify multiple CONTRAST statements, thereby specifying multiple contrasts for each variable. Multiplicity adjustments are computed for all contrasts and all variables simultaneously. The coefficients are applied in the order of the CLASS variables; this order can be changed with the ORDER= option in the PROC MULTTEST statement. For example, consider a four-group experiment with CLASS variable levels A1, A2, B1, and B2 denoting two levels of two treatments. The following statements produce three linear trend tests for each variable identified in the TEST statement. PROC MULTTEST computes the multiplicity adjustments over the entire collection of tests, which is three times the number of variables.

```
contrast 'a vs b'    -1 -1  1  1;
contrast 'a linear'  -1  1  0  0;
contrast 'b linear'   0  0 -1  1;
```

As another example, consider an animal carcinogenicity experiment with dose levels 0, 4, 8, 16, and 50. You might consider trend tests defined using the following statement:

```
contrast 'arithmetic trend' 0 4 8 16 50;
```

This statement produces a trend test using the indicated scoring coefficients. Multiplicity-adjusted p-values are then computed over the collection of variables identified in the TEST statement. Refer to Lagakos and Louis (1985) for guidelines on the selection of contrast-scoring values.

When a Fisher test is specified in the TEST statement, the CONTRAST statement coefficients are used to group the CLASS variable's levels. Groups with a -1 scoring coefficient are combined and compared with groups with a 1 scoring coefficient for each test, and groups with a 0 coefficient are not included in the contrast. For example,

```
contrast 'c vs all' -1 1 1 1;
contrast 'c vs t1'  -1 1 0 0;
contrast 'c vs t3'  -1 0 0 1;
```

compute Fisher exact tests for (a) control versus the combined treatment groups, (b) control versus the first treatment group, and (c) control versus the third treatment group. Multiplicity adjustments are then computed over the entire collection of tests

and variables. If the FISHER option is specified and no CONTRAST statement is specified, then all contrasts of control versus treatment are automatically generated, with the first level of the CLASS variable deemed to be the control. Only -1, 1, and 0 are acceptable CONTRAST scoring coefficients when the Fisher test is specified; PROC MULTTEST ignores the CONTRAST statement if any other coefficients appear.

CONTRAST values are 0,1,2,... by default, except for the Fisher test described in the previous paragraph.

For continuous data (and for the FT tests), the contrast coefficients are centered to have mean 0. The resulting centered scoring coefficients are then applied to the sample means (or to the double-arcsine-transformed proportions in the case of the FT tests).

FREQ Statement

FREQ *variable* **;**

The FREQ statement names a variable that provides frequencies for each observation in the DATA= data set. Specifically, if n is the value of the FREQ variable for a given observation, then that observation is used n times.

If the value of the FREQ variable is missing or is less than 1, the observation is not used in the analysis. If the value is not an integer, only the integer portion is used.

STRATA Statement

STRATA *variable* **;**

The STRATA statement identifies a single variable to use as a stratification variable in the analysis. This yields tests similar to those discussed in Mantel and Haenszel (1959) and Hoel and Walburg (1972) for binary data and pooled-means tests for continuous data. For example, when you test for prevalence in a carcinogenicity study, it is common to stratify on intervals of the time at death; the first level of the stratification variable may represent weeks $0-52$, the second weeks $53-80$, and so on. In multicenter clinical studies, each level of the stratification variable may represent a particular center.

TEST Statement

TEST *name (variables* < */ options* >*)* **;**

The TEST statement is required. It identifies statistical tests to be performed and the discrete and continuous variables to be tested. The following tests are permitted as *name* in the TEST statement.

CA requests the Cochran-Armitage linear trend tests for group comparisons. The test variables should take the value 0 for a failure and 1 for a success. The PERMUTATION= option can be used to request an exact permutation test; otherwise, a Z-score approximation is used. The CONTINUITY= option can be used to specify a continuity correction for the Z-score approximation.

FISHER requests Fisher exact tests for comparing two treatment groups. The test variables should take the value 0 for a failure and 1 for a success.

FT requests Z-score CA tests based upon the Freeman-Tukey double arcsine transformation of the frequencies. The test variables should take the value 0 for a failure and 1 for a success.

MEAN requests the t-test for the mean. The test variables can take on any numeric values.

PETO requests the Peto mortality-prevalence test. The test variables should take the value 0 for a failure, 1 for a success, and 2 for a fatality. The TIME= option should be used with the PETO test to specify a variable giving the age at death. The CONTINUITY= option can be used to specify a continuity correction for the test.

If the value of a TEST variable is invalid, the observation is not used in the analysis. You can specify two tests only if one of them is MEAN. For example, the following statement is valid

```
test ca(d1-d2) mean(c1-c2);
```

but the statement

```
test ca(d1-d2) ft(d1-d2);
```

is invalid.

You can specify the following options in the TEST statement (some apply to only one test).

BINOMIAL
specifies that the binomial variance estimate be used for CA and PETO tests in their asymptotic normal approximations. The default is to use the hypergeometric variance.

CONTINUITY= *number*
C= *number*
specifies *number* as a particular continuity correction for the Z-score approximation in the CA and PETO tests. The default is 0.

LOWERTAILED
LOWER
is used to make all tests lower-tailed. All tests are two-tailed by default.

PERMUTATION= *number*
PERM= *number*
specifies that p-values for the CA and PETO tests be computed using exact permutation distributions when marginal success or failure totals within a stratum are *number* or less. For values greater than *number* (or when the PERMUTATION= option is omitted), PROC MULTTEST uses standard normal approximations with a continuity correction chosen to approximate the permutation distribution. PROC MULTTEST computes the appropriate convolution distributions when you use the STRATA statement along with the PERMUTATION= option.

TIME= *variable*
identifies the PETO test variable containing the age at death, which is assumed to be integer valued. If the TIME= option is omitted, all ages are assumed to equal 1.

UPPERTAILED
UPPER
is used to make all tests upper-tailed. All tests are two-tailed by default.

Details

Statistical Tests

The following section discusses the statistical tests performed in the MULTTEST procedure. For continuous data, a t-test for the mean is available. For discrete variables, available tests are the Cochran-Armitage (CA) linear trend test, the Freeman-Tukey (FT) double arcsine test, the Peto mortality-prevalence test, and the Fisher exact test.

Throughout this section, the discrete and continuous variables are denoted by S_{vgsr} and X_{vgsr}, respectively, where v is the variable, g is the treatment group, s is the stratum, and r is the replication. A plus sign ($+$) subscript denotes summation over an index.

Cochran-Armitage Linear Trend Test

The Cochran-Armitage linear trend test (Cochran 1954; Armitage 1955; Agresti 1990) is implemented using a Z-score approximation, an exact permutation distribution, or a combination of both.

Z-Score Approximation

Let m_{vgs} denote the sample size for a binary variable v within group g and stratum s. The pooled probability estimate for variable v and stratum s is

$$p_{vs} = \frac{S_{v+s+}}{m_{v+s}}$$

The expected value (under constant within-stratum treatment probabilities) for variable v, group g, and stratum s is

$$E_{vgs} = m_{vgs}p_{vs}$$

The test statistic for variable v has numerator

$$N_v = \sum_s \sum_g t_g(S_{vgs+} - E_{vgs})$$

where t_g denotes a trend coefficient (specified by the CONTRAST statement). The binomial variance estimate for this statistic is

$$V_v = \sum_s p_{vs}(1 - p_{vs}) \sum_g m_{vgs}(t_s - \bar{t}_{vs})^2$$

where

$$\bar{t}_{vs} = \sum_g \frac{m_{vgs}t_g}{m_{v+s}}$$

The hypergeometric variance estimate (the default) is

$$V_v = \sum_s \{m_{v+s}/(m_{v+s} - 1)\}p_{vs}(1 - p_{vs}) \sum_g m_{vgs}(t_s - \bar{t}_{vs})^2$$

For any strata s with $m_{v+s} \leq 1$, the contribution to the variance is taken to be zero.

PROC MULTTEST computes the Z-score statistic

$$Z_v = \frac{N_v}{\sqrt{V_v}}$$

The p-value for this statistic comes from the standard normal distribution. Whenever a 0 is computed for the denominator, the p-value is set to 1. This p-value approximates the probability obtained from the exact permutation distribution, discussed in the following text.

The Z-score statistic can be continuity-corrected to better approximate the permutation distribution. With continuity correction c, the upper-tailed p-value is computed from

$$Z_v = \frac{N_v - c}{\sqrt{V_v}}$$

For two-tailed, noncontinuity-corrected tests, PROC MULTTEST reports the p-value as $2\min(p, 1-p)$, where p is the upper-tailed p-value. The same formula holds for the

continuity-corrected test, with the exception that when the noncontinuity-corrected Z and the continuity-corrected Z have opposite signs, the two-tailed p-value is 1.

When the PERMUTATION= option is specified and no STRATA variable is specified, PROC MULTTEST uses a continuity correction selected to optimally approximate the upper-tail probability of permutation distributions with smaller marginal totals (Westfall and Lin 1988). Otherwise, the continuity correction is specified using the CONTINUITY= option in the TEST statement.

The CA Z-score statistic is the Hoel-Walburg (Mantel-Haenszel) statistic reported by Dinse (1985).

Exact Permutation Test

When you use the PERMUTATION= option for CA in the TEST statement, PROC MULTTEST computes the exact permutation distribution of the trend score

$$T_v = \sum_s \sum_g t_g S_{vgs+}$$

and then compares the observed value of this trend with the permutation distribution to obtain the p-value

$$p_v = \Pr(X \geq \text{ observed } T_v)$$

where X is a random variable from the permutation distribution and where upper-tailed tests are requested. This probability can be viewed as a binomial probability, where the within-stratum probabilities are constant and where the probability is conditional with respect to the marginal totals S_{v+s+}. It also can be considered a rerandomization probability.

Because the computations can be quite time-consuming with large data sets, specifying the PERMUTATION=*number* option in the TEST statement limits the situations where PROC MULTTEST computes the exact permutation distribution. When marginal total success or total failure frequencies exceed *number* for a particular stratum, the permutation distribution is approximated using a continuity-corrected normal distribution. You should be cautious in using the PERMUTATION= option in conjunction with bootstrap resampling because the permutation distribution is recomputed for each bootstrap sample. This recomputation is not necessary with permutation resampling.

The permutation distribution is computed in two steps:

1. The permutation distributions of the trend scores are computed within each stratum.
2. The distributions are convolved to obtain the distribution of the total trend.

As long as the total success or failure frequency does not exceed *number* for any stratum, the computed distributions are exact. In other words, if $S_{v+s+} \leq$ *number* or $(m_{v+s} - S_{v+s+}) \leq$ *number* for all s, then the permutation trend distribution for variable v is computed exactly.

In step 1, the distribution of the within-stratum trend

$$\sum_g t_g S_{vgs+}$$

is computed using the multivariate hypergeometric distribution of the S_{vgs+}, provided *number* is not exceeded. This distribution can be written as

$$\Pr(S_{v1s+}, S_{v2s+}, \ldots, S_{vGs+}) = \prod_{g=1}^{G} \frac{\left(\begin{array}{c} m_{vgs} \\ S_{vgs+} \end{array} \right)}{\left(\begin{array}{c} m_{v+s} \\ S_{v+s+} \end{array} \right)}$$

The distribution of the within-stratum trend is then computed by summing these probabilities over appropriate configurations. For further information on this technique, refer to Bickis and Krewski (1986) and Westfall and Lin (1988). In step 2, the exact convolution distribution is obtained for the trend statistic summed over all strata having totals that meet the threshold criterion. This distribution is obtained by applying the fast Fourier transform to the exact within-stratum distributions. A description of this general method can be found in Pagano and Tritchler (1983) and Good (1987).

The convolution distribution of the overall trend is then computed by convolving the exact distribution with the distribution of the continuity-corrected standard normal approximation. To be more specific, let S_1 denote the subset of stratum indices that satisfy the threshold criterion, and let S_2 denote the subset of indices that do not satisfy the criterion. Let T_{v1} denote the combined trend statistic from the set S_1, which has an exact distribution obtained using Fourier analysis as previously outlined, and let T_{v1} denote the combined trend statistic from the set S_2. Then the distribution of the overall trend $T_v = T_{v1} + T_{v2}$ is obtained by convolving the analytic distribution of T_{v1} with the continuity-corrected normal approximation for T_{v2}. Using the notation from the "Z-Score Approximation" section on page 2325, this convolution can be written as

$$\begin{aligned} \Pr(T_{v1} + T_{v2} \geq u) &= \sum_{u1} \Pr(T_{v1} + T_{v2} \geq u \mid T_{v1} = u1) \Pr(T_{v1} = u1) \\ &\approx \sum_{u1} \Pr(Z \geq z) \Pr(T_{v1} = u1) \end{aligned}$$

where Z is a standard normal random variable, and

$$z = \frac{1}{\sqrt{V_v}} \left(u - u1 - \sum_{S_2} p_{vs} \sum_g t_g m_{vgs} - c \right)$$

In this expression, the summation of s in V_v is over S_2, and c is the continuity correction discussed under the Z-score approximation.

When a two-tailed test is requested, the expected trend

$$E_v = \sum_s \sum_g t_g E_{vgs}$$

is computed, and the two-tailed p-value is reported as the permutation tail probability for the observed trend T_v plus the permutation tail probability for $2E_v - T_v$, the reflected trend.

Freeman-Tukey Double Arcsine Test

For this test, the trend scores $t_1, \ldots, t_G$ are centered to the values $c_1, \ldots, c_G$, where $c_g = t_g - \bar{t}$, $\bar{t} = \sum_g t_g / G$, and G is the number of groups. The numerator of this test statistic is

$$N_v = \sum_s m_{v+s} \sum_g c_g f(S_{vgs+}, m_{vgs})$$

and is weighted by the within-strata sample size (m_{v+s}) to ensure comparability with the ordinary C-A trend statistic.

The function $f(r, n)$ is the double arcsine transformation:

$$f(r,n) = \arcsin\left(\sqrt{\frac{r}{n+1}}\right) + \arcsin\left(\sqrt{\frac{r+1}{n+1}}\right)$$

The variance estimate is

$$V_v = \sum_s m_{v+s}^2 \sum_g \frac{c_g^2}{m_{vgs} + \frac{1}{2}}$$

and the test statistic is

$$Z_v = \frac{N_v}{\sqrt{V_v}}$$

The Freeman-Tukey transformation and its variance are described by Freeman and Tukey (1950) and Miller (1978). Since its variance is not weighted by the pooled probabilities, as is the CA test, the FT test can be more useful than the CA test for tests involving only a subset of the groups.

Peto Mortality-Prevalence Test

The Peto test is a modified Cochran-Armitage procedure incorporating mortality and prevalence information. It represents a special case in PROC MULTTEST because the data structure requirements are different, and the resampling methods used for adjusting p-values are not valid. The TIME= option variable is required to specify "death" times or, more generally, time of occurrence. In addition, the test variables must assume one of the following three values.

- 0 = no occurrence
- 1 = incidental occurrence
- 2 = fatal occurrence

Use the TIME= option variable to define the mortality strata, and use the STRATA statement variable to define the prevalence strata.

The Peto test is computed like two Cochran-Armitage Z-score approximations, one for prevalence and one for mortality.

In the following notation, the subscript v represents the variable, g represents the treatment group, s represents the stratum, and t represents the time. Recall that a plus sign $(+)$ in a subscript location denotes summation over that subscript.

Let S^P_{vgs} be the number of incidental occurrences, and let m^P_{vgs} be the total sample size for variable v in group g, stratum s, excluding fatal tumors.

Let S^F_{vgt} be the number of fatal occurrences in time period t, and let m^F_{vgt} be the number alive at the end of time $t-1$.

The pooled probability estimates are

$$\begin{aligned} p^P_{vs} &= \frac{S^P_{v+s}}{m^P_{v+s}} \\ p^F_{vt} &= \frac{S^F_{v+t}}{m^F_{v+t}} \end{aligned}$$

The expected values are

$$\begin{aligned} E^P_{vgs} &= m^P_{vgs} p^P_{vs} \\ E^F_{vgt} &= m^F_{vgt} p^F_{vt} \end{aligned}$$

Define the numerator terms:

$$\begin{aligned} N^P_v &= \sum_s \sum_g t_g \left(S^P_{vgs} - E^P_{vgs} \right) \\ N^F_v &= \sum_t \sum_g t_g \left(S^P_{vgt} - E^P_{vgt} \right) \end{aligned}$$

where t_g denotes a trend coefficient. Define the denominator variance terms (using the binomial variance) :

$$\begin{aligned} V^P_v &= \sum_s p^P_{vs} \left(1 - p^P_{vs}\right) \left[\left(\sum_g m^P_{vgs} {t_g}^2 \right) - \frac{1}{m^P_{v+s}} \left(\sum_g m^P_{vgs} t_g \right)^2 \right] \\ V^F_v &= \sum_s p^F_{vt} \left(1 - p^F_{vt}\right) \left[\left(\sum_g m^F_{vgt} {t_g}^2 \right) - \frac{1}{m^F_{v+t}} \left(\sum_g m^F_{vgt} t_g \right)^2 \right] \end{aligned}$$

The hypergeometric variances (the default) are calculated by weighting the within-strata variances as discussed in the "Z-Score Approximation" section on page 2325.

The Peto statistic is computed as

$$Z_v = \frac{N_v^P + N_v^F - c}{\sqrt{V_v^P + V_v^F}}$$

where c is a continuity correction. The p-value is determined from the standard normal distribution unless the PERMUTATION=*number* option is used. When you use the PERMUTATION= option for PETO in the TEST statement, PROC MULTTEST computes the "discrete approximation" permutation distribution described by Mantel (1980) and Soper and Tonkonoh (1993). Specifically, the permutation distribution of

$$\sum_s \sum_g t_g S_{vgs}^P + \sum_t \sum_g t_g S_{vgt}^F$$

is computed, assuming that $\{\sum_g t_g S_{vgs}^P\}$ and $\{\sum_g t_g S_{vgt}^F\}$ are independent over all s and t. The p-values are exact under this independence assumption. However, the independence assumption is valid only asymptotically, which is why these p-values are called "approximate."

An exact permutation distribution is available only under the assumption of equal risk of censoring in all treatment groups; even then, computing this distribution can be cumbersome. Soper and Tonkonoh (1993) describe situations where the discrete approximation distribution closely fits the exact permutation distribution.

Fisher Exact Test

The CONTRAST statement in PROC MULTTEST enables you to compute Fisher exact tests for two-group comparisons. No stratification variable is allowed for this test. Note, however, that the FISHER exact test is a special case of the exact permutation tests performed by PROC MULTTEST and that these permutation tests allow a stratification variable. Recall that contrast coefficients can be -1, 0, or 1 for the Fisher test. The frequencies and sample sizes of the groups scored as -1 are combined, as are the frequencies and sample sizes of the groups scored as 1. Groups scored as 0 are excluded. The -1 group is then compared with the 1 group using the Fisher exact test.

Letting x and m denote the frequency and sample size of the 1 group, and y and n denote those of the -1 group, the p-value is calculated as

$$\Pr(X \geq x \mid X + Y = x + y) = \sum_{i=x}^{m} \frac{\binom{m}{i}\binom{n}{x+y-i}}{\binom{m+n}{x+y}}$$

where X and Y are independent binomially distributed random variables with sample sizes m and n and common probability parameters. The hypergeometric distribution is used to determine the stated probability; Yates (1984) discusses this technique. PROC MULTTEST computes the two-tailed p-values by adding probabilities from

both tails of the hypergeometric distribution. The first tail is from the observed x and y, and the other tail is chosen so that the resulting probability is as large as possible without exceeding the probability from the first tail.

t-Test for the Mean

For continuous variables, PROC MULTTEST automatically centers the trend coefficients, as in the Freeman-Tukey test. These centered coefficients c_g are then used to form a t-statistic contrasting the within-group means. Let n_{vgs} denote the sample size within group g and stratum s; it depends on variable v only when there are missing values. Define

$$\bar{X}_{vgs+} = \frac{1}{n_{vgs}} \sum_r X_{vgsr}$$

as the sample mean within a group-and-stratum combination, and define

$$s_v^2 = \frac{\sum_s \sum_g \sum_r \left(X_{vgsr} - \bar{X}_{vgs+}\right)^2}{\sum_s \sum_g \left(n_{vgs} - 1\right)}$$

as the pooled sample variance. Assume constant variance for all group-and-stratum combinations. Then the t-statistic for the mean is

$$M_v = \frac{\sum_s n_{v+s} \sum_g c_g \bar{X}_{vgs+}}{\sqrt{s_v^2 \left(\sum_s n_{v+s}^2 \sum_g \frac{c_g^2}{n_{vgs}} \right)}}$$

and is weighted by the within-strata sample size (n_{v+s}) to ensure comparability with the C-A trend and Freeman-Tukey statistics.

Let μ_{vgs} denote the treatment means. Then under the null hypothesis that

$$\sum_s n_{v+s} \sum_g c_g \mu_{vgs} = 0$$

and assuming normality, independence, and homoscedasticity, M_v follows a t-distribution with $\sum_s \sum_g \left(n_{vgs} - 1\right)$ degrees of freedom.

Whenever a denominator of 0 is computed, the p-value is set to 1. When missing data force $n_{vgs} = 0$, then the contribution to the denominator of the pooled variance is 0 and not -1. This is also true for degrees of freedom.

p-Value Adjustments

PROC MULTTEST offers p-value adjustments using Bonferroni, Sidak, Bootstrap resampling, and Permutation resampling, all with single-step or stepdown versions.

In addition, Hochberg's (1988) and Benjamini and Hochberg's (1995) step-up methods are offered. The Bonferroni and Sidak methods are calculated from the permutation distributions when exact permutation tests are used with CA or PETO tests.

All methods but the resampling methods are calculated using simple functions of the raw p-values or marginal permutation distributions; the permutation and bootstrap adjustments require the raw data. Because the resampling techniques incorporate distributional and correlational structures, they tend to be less conservative than the other methods.

When a resampling (bootstrap or permutation) method is used with only one test, the adjusted p-value is the bootstrap or permutation p-value for that test, with no adjustment for multiplicity, as described by Westfall and Soper (1994).

Bonferroni

Suppose that PROC MULTTEST performs R statistical tests, yielding p-values $p_1, \ldots, p_R$. Then the Bonferroni p-value for test r is simply Rp_r. If the adjusted p-value exceeds 1, it is set to 1.

If the unadjusted p-values are computed using exact permutation distributions, then the Bonferroni adjustment for p_r is $p_1^* + \cdots + p_R^*$, where p_j^* is the largest p-value from the permutation distribution of test j satisfying $p_j^* \leq p_r$, or 0 if all permutational p-values of test j are greater than p_r. These adjustments are much less conservative than the ordinary Bonferroni adjustments because they incorporate the discrete distributional characteristics. However, they remain conservative in that they do not incorporate correlation structures between multiple contrasts and multiple variables (Westfall and Wolfinger 1997).

Sidak

A technique slightly less conservative than Bonferroni is the Sidak p-value (Sidak 1967), which is $1 - (1 - p_r)^R$. It is exact when all of the p-values are uniformly distributed and independent, and it is conservative when the test statistics satisfy the positive orthant dependence condition (Holland and Copenhaver 1987).

If the unadjusted p-values are computed using exact permutation distributions, then the Sidak adjustment for p_r is $1 - (1 - p_1^*) \cdots (1 - p_R^*)$, where the p_j^* are as described previously. These adjustments are less conservative than the corresponding Bonferroni adjustments, but they do not incorporate correlation structures between multiple contrasts and multiple variables (Westfall and Wolfinger 1997).

Bootstrap

The bootstrap method creates pseudo-data sets by sampling observations with replacement from each within-stratum pool of observations. An entire data set is thus created, and p-values for all tests are computed on this pseudo-data set. A counter records whether the minimum p-value from the pseudo-data set is less than or equal to the actual p-value for each base test. (If there are R tests, then there are R such counters.) This process is repeated a large number of times, and the proportion of resampled data sets where the minimum pseudo-p-value is less than or equal to an actual p-value is the adjusted p-value reported by PROC MULTTEST. The algorithms are described by Westfall and Young (1993).

In the case of continuous data, the pooling of the groups is not likely to recreate the shape of the null hypothesis distribution, since the pooled data are likely to be multimodal. For this reason, PROC MULTTEST automatically mean-centers all continuous variables prior to resampling. Such mean-centering is akin to resampling residuals in a regression analysis, as discussed by Freedman (1981). You can specify the NOCENTER option if you do not want to center the data. (In most situations, it does not seem to make much difference whether or not you center the data.)

The bootstrap method explicitly incorporates all sources of correlation, from both the multiple contrasts and the multivariate structure. The adjusted p-values incorporate all correlations and distributional characteristics.

Permutation

The permutation-style adjusted p-values are computed in identical fashion as the bootstrap adjusted p-values, with the exception that the within-stratum resampling is performed without replacement instead of with replacement. This produces a rerandomization analysis such as in Brown and Fears (1981) and Heyse and Rom (1988). In the spirit of rerandomization analyses, the continuous variables are not centered prior to resampling. This default can be overridden by using the CENTER option.

The permutation method explicitly incorporates all sources of correlation, from both the multiple contrasts and the multivariate structure. The adjusted p-values incorporate all correlations and distributional characteristics.

Stepdown Methods

Stepdown testing is available for the Bonferroni, Sidak, bootstrap, and permutation methods. The benefit of using stepdown methods is that the tests are made more powerful (smaller adjusted p-values) while, in most cases, maintaining strong control of the familywise error rate. The stepdown method was pioneered by Holm (1979) and further developed by Shaffer (1986), Holland and Copenhaver (1987), and Hochberg and Tamhane (1987).

Suppose the base test p-values are ordered as $p_1 < p_2 < \cdots < p_R$. The Bonferroni stepdown p-values $s_1, \ldots, s_R$ are obtained from

$$\begin{aligned} s_1 &= Rp_1 \\ s_2 &= \max\left(s_1, (R-1)p_2\right) \\ s_3 &= \max\left(s_2, (R-2)p_3\right) \\ &\vdots \end{aligned}$$

As always, if any adjusted p-value exceeds 1, it is set to 1. The Sidak stepdown p-values are determined similarly:

$$\begin{aligned} s_1 &= 1-(1-p_1)^R \\ s_2 &= \max\left(s_1, 1-(1-p_2)^{R-1}\right) \\ s_3 &= \max\left(s_2, 1-(1-p_3)^{R-2}\right) \\ &\vdots \end{aligned}$$

Stepdown Bonferroni adjustments using exact tests are defined as

$$\begin{aligned} s_1 &= p_1^* + \cdots + p_R^* \\ s_2 &= \max\left(s_1, p_2^* + \cdots + p_R^*, \right) \\ s_3 &= \max\left(s_2, p_3^* + \cdots + p_R^*, \right) \\ &\vdots \end{aligned}$$

where the p_j^* are defined as before. Note that p_j^* is taken from the permutation distribution corresponding to the jth smallest unadjusted p-value. Also, any s_j greater than 1.0 is truncated to 1.0.

Stepdown Sidak adjustments for exact tests are defined analogously by substituting $1 - (1 - p_j^*) \cdots (1 - p_R^*)$ for $p_j^* + \cdots + p_R^*$.

The resampling-style stepdown method is analogous to the preceding stepdown methods; the most extreme p-value is adjusted according to all R tests, the second-most extreme p-value is adjusted according to $(R - 1)$ tests, and so on. The difference is that all correlational and distributional characteristics are incorporated when you use resampling methods. More specifically, assuming the same ordering of p-values as discussed previously, the resampling-style stepdown adjusted p-value for test r is the probability that the minimum pseudo-p-value of tests $r, \ldots, R$ is less than or equal to p_r.

This probability is evaluated using Monte Carlo, as are the previously described resampling-style adjusted p-values. In fact, the computations for stepdown adjusted p-values are essentially no more time-consuming than the computations for the non-stepdown adjusted p-values. After Monte Carlo, the stepdown adjusted p-values are corrected to ensure monotonicity; this correction leaves the first adjusted p-values alone, then corrects the remaining ones as needed. The stepdown method approximately controls the familywise error rate, and it is described in more detail by Westfall and Young (1993).

Hochberg

Assuming p-values are independent and uniformly distributed under their respective null hypotheses, Hochberg (1988) demonstrated that Holm's stepdown adjustments control the familywise error rate even when calculated in *step-up* fashion. Since the adjusted p-values are uniformly smaller for Hochberg's method than for Holm's method, the Hochberg method is more powerful. However, this improved power comes at the cost of having to make the assumption of independence.

The Hochberg adjusted p-values are defined in reverse order as the stepdown Bonferroni:

$$\begin{aligned} s_R &= p_R \\ s_{(R-1)} &= \min\left(s_R, 2p_{(R-1)}\right) \\ s_{(R-2)} &= \min\left(s_{(R-1)}, 3p_{(R-2)}\right) \\ &\vdots \end{aligned}$$

False Discovery Rate

The FDR option requests p-values that control the "false discovery rate," described by Benjamini and Hochberg (1995). These adjustments are potentially much less conservative than the Hochberg adjustments; however, they do not necessarily control the familywise error rate. Furthermore, they are guaranteed to control the false discovery rate only with independent p-values that are uniformly distributed under their respective null hypotheses.

The FDR adjusted p-values are defined in step-up fashion, like the Hochberg adjustments, but with less conservative multipliers:

$$\begin{aligned} s_R &= p_R \\ s_{(R-1)} &= \min\left(s_R, [R/(R-1)]p_{(R-1)}\right) \\ s_{(R-2)} &= \min\left(s_{(R-1)}, [R/(R-2)]p_{(R-2)}\right) \\ &\vdots \end{aligned}$$

Missing Values

If a CLASS or STRATA variable has a missing value, then PROC MULTTEST removes that observation from the analysis.

When there are missing values for test variables, the within group-and-stratum sample sizes may differ from variable to variable. In most cases this is not a problem; however, it is possible for all data to be missing for a particular group within a particular stratum. For continuous variables and Freeman-Tukey tests, PROC MULTTEST recenters the trend scores within strata where all data for a particular group are missing. The Cochran-Armitage and Peto tests are unaffected by this situation.

PROC MULTTEST uses missing values for resampling if they exist in the original data set. If all variables have missing values for any observation, then PROC MULTTEST removes it prior to resampling. Otherwise, PROC MULTTEST treats all missing values as ordinary observations in the resampling. This means that different resampled data sets can have different group sizes. In some cases it means that a resampled data set can have all missing values for a particular variable in a particular group/stratum combination, even when values exist for that combination in the original data. For this reason, PROC MULTTEST recomputes all quantities within each pseudo-data set, including such items as centered scoring coefficients and degrees of freedom for p-values.

While PROC MULTTEST does provide analyses in missing value cases, you should not feel that it completely solves the missing value problem. If you are concerned about the adverse effects of missing data on a particular analysis, you should consider using imputation and sensitivity analyses to assess the effects of the missing data.

Computational Resources

PROC MULTTEST keeps all of the data in memory to expedite resampling. A large portion of the memory requirement is thus 8*NOBS*NVAR bytes, where NOBS is the number of observations in the data set, and NVAR is the number of variables analyzed, including CLASS, FREQ, and STRATA variables.

If you specify PERMUTATION=*number* (for exact permutation distributions), then PROC MULTTEST requires additional memory. This requirement is approximately 4*NTEST*NSTRATA*CMAX**number**(*number*+1) bytes, where NTEST is the number of contrasts, NSTRATA is the number of STRATA levels, and CMAX is the maximum contrast coefficient.

The execution time is linear in the number of resamples.

Output Data Sets

OUT= Data Set

The OUT= data set contains contrast names (_test_), variable names (_var_), the contrast label (_contrast_), raw p-values (raw_p), and all requested adjusted p-values (bon_p, sid_p, stpbon_p, stpsid_p, boot_p, perm_p, stpbootp, stppermp, hoc_p, or fdr_p).

If a resampling-based adjusted p-value is requested, then the simulation standard error is included as either sim_se or stpsimse, depending upon whether single-step or stepdown adjustments are requested. The simulation standard errors are used to bound the true resampling-based adjusted p-value. For example, if the resampling-based estimate is 0.0312 and the simulation standard error is 0.00123, then a 95% confidence interval for the true adjusted p-value is $0.0312 \pm 1.96(0.00123)$, or 0.0288 to 0.0336.

Intermediate statistics used to calculate the p-values are also written to the OUT= data set. The statistics are separated by the _strat_ level. When _strat_ is reported as missing, then the statistics refer to the pooled analysis over all _strat_ levels. The p-values are provided only for the pooled analyses and are therefore reported as missing for the strata-specific statistics.

For the PETO test, an additional variable, _tstrat_, is included to indicate whether the stratum is an incidental occurrence stratum (_tstrat_=0) or a fatal occurrence stratum (_tstrat_=1).

The statistic _value_ is the per-strata contribution to the numerator of the overall test statistic. In the case of the MEAN test, this is the contrast function of the sample means multiplied by the total number of observations within the stratum. For the FT test, _value_ is the contrast function of the double-arcsine transformed proportions, again multiplied by the total number of observations within the stratum. For the CA and PETO tests, _value_ is the observed value of the trend statistic within that stratum.

When either PETO or CA is requested, the variable _exp_ is included; this variable contains the expected value of the trend statistic for the given stratum.

The statistic _se_ is the square root of the variance of the per-strata _value_ value for any of the tests.

For MEAN tests, the variable _nval_ is included. When reported with an individual stratum level (that is, when the _strat_ value is nonmissing), the value _nval_ refers to the within-stratum sample size. For the combined analysis (that is, the value of the _strat_ is missing), the value _nval_ contains degrees of freedom of the t-distribution used to compute the unadjusted p-value.

When the FISHER test is requested, the OUT= data set contains variables _xval_, _mval_, _yval_, and _nval_, which define observations and sample sizes in the two groups defined by the CONTRAST statement.

For example, the OUT= data set from the drug example in the "Getting Started" section on page 2314 is displayed in Figure 43.4.

Obs	_test_	_var_	_contrast_	_value_	_exp_	_se_	raw_p	boot_p	sim_se
1	CA	SideEff1	Trend	8	5	1.54303	0.05187	0.34705	.003366053
2	CA	SideEff2	Trend	7	5	1.54303	0.19492	0.83880	.002600140
3	CA	SideEff3	Trend	10	7	1.63299	0.06619	0.52315	.003531742
4	CA	SideEff4	Trend	10	6	1.60357	0.01262	0.09370	.002060586
5	CA	SideEff5	Trend	7	4	1.44749	0.03821	0.24380	.003036129
6	CA	SideEff6	Trend	9	6	1.60357	0.06137	0.44545	.003514430
7	CA	SideEff7	Trend	9	5	1.54303	0.00953	0.05400	.001598186
8	CA	SideEff8	Trend	8	5	1.54303	0.05187	0.34705	.003366053
9	CA	SideEff9	Trend	7	5	1.54303	0.19492	0.83880	.002600140
10	CA	SideEff10	Trend	8	6	1.60357	0.21232	0.90020	.002119433

Figure 43.4. Output Data for the MULTTEST Procedure

OUTPERM= Data Set

The OUTPERM= data set contains contrast names (_contrast_), variable names (_var_), and the associated permutation distributions (_value_ and upper_p). PROC MULTTEST computes the permutation distributions when you use the PERMUTATION= option with the CA test. The _value_ variable represents the support of the distributions, and upper_p represents their cumulative upper-tail probabilities. The size of this data set depends on the number of variables and the support of their permutation distributions. For information on how this distribution is computed, see the "Exact Permutation Test" section on page 2327. For an illustration, see Example 43.1 on page 2340.

OUTSAMP= Data Set

The OUTSAMP= data set contains the data sets used in the resampling analysis, if such an analysis is requested. The variable _sample_ indicates the number of the resampled data set. This variable ranges from 1 to NSAMPLE. For each value of the _sample_ variable, an entire resampled data set is included, with _strat_, _class_, and all other variables in the original data set. The values of the original variables are mean-centered for the mean test, if requested. The variable _obs_ indicates the observation's position in the original data set.

Each new data set is randomly drawn from the original data set, either with (bootstrap) or without (permutation) replacement. The size of this data set is, thus, the number of observations in the original data set times the number of samples.

Displayed Output

The output produced by PROC MULTTEST is divided into several tables:

- The "Model Information" table provides a list of the options and settings used for that particular invocation of the procedure. Included in this list are the following items:
  - statistical tests
  - support of the exact permutation distribution for the CA and PETO tests
  - continuity corrections used for the CA test
  - test tails
  - strata adjustment
  - p-value adjustments
  - centering of continuous variables
  - number of samples and seed
- The "Contrast Coefficients" table lists the coefficients used in constructing the statistical tests. These coefficients are either specified in CONTRAST statements or generated by default. The coefficients apply to the levels of the CLASS statement variable.
- The "Variable Tabulations" tables provide summary statistics for each variable listed in the TEST statement. Included for discrete variables are the count, sample size, and percentage of occurrences. For continuous variables, the mean, sample standard deviation, and sample size are displayed. All of the previously mentioned statistics are computed for distinct combinations of the CLASS and STRATA statement variables.
- The "p-Values" table is a collection of the raw and adjusted p-values from this run of PROC MULTTEST. The p-values are listed by variable and test.

ODS Table Names

PROC MULTTEST assigns a name to each table it creates, and you must use this name to reference the table when using the Table Delivery System (ODS). These names are listed in the following table. For more information on ODS, see Chapter 15, "Using the Output Delivery System."

Table 43.1. ODS Tables Created by the MULTTEST Procedure

ODS Table Name	Description	Statement
Continuous	Continuous variable tabulations	TEST with MEAN
Contrasts	Contrast coefficients	default
Discrete	Discrete variable tabulations	TEST with CA, FT, PETO,

Table 43.1. (continued)

ODS Table Name	Description	Statement
		or FISHER
ModelInfo	Model information	default
pValues	p-values from the tests	default

Examples

Example 43.1. Cochran-Armitage Test with Permutation Resampling

This example, from Keith Soper at Merck, illustrates the exact permutation Cochran-Armitage test carried out on permutation resamples. In the following data set, the 0s represent failures and the 1s represent successes. Note that the binary variables S1 and S2 have perfect negative association. The grouping variable is Dose.

```
data a;
   input S1 S2 Dose @@;
   datalines;
0 1 1   1 0 1   0 1 1   0 1 1
0 1 1   1 0 1   1 0 2   1 0 2
0 1 2   1 0 2   0 1 2   1 0 2
1 0 3   1 0 3   1 0 3   0 1 3
0 1 3   1 0 3
;

proc multtest data=a permutation nsample=10000
              seed=36607 outperm=pmt pvals;
   test ca(S1 S2 / permutation=10 uppertailed);
   class Dose;
   contrast 'CA Linear Trend' 0 1 2;
run;

proc print data=pmt;
run;
```

The PROC MULTTEST statement requests 10,000 permutation resamples. The OUTPERM=PMT option requests an output SAS data set for the exact permutation distribution computed for the CA test.

The TEST statement specifies an upper-tailed Cochran-Armitage linear trend test for S1 and S2. The cutoff for exact permutation calculations is 10, as specified with the PERMUTATION= option in the TEST statement. Since S1 and S2 have ten and eight successes, respectively, PROC MULTTEST uses exact permutation distributions to compute the p-values for both variables.

The groups for the CA test are the levels of Dose from the CLASS statement. The coefficients applied to these groups are 0, 1, and 2, respectively, as specified in the CONTRAST statement.

Finally, the invocation of PROC PRINT displays the SAS data set containing the permutation distributions.

The results from this analysis are listed in Output 43.1.1.

Output 43.1.1. Cochran-Armitage Test with Permutation Resampling

The Multtest Procedure

Model Information

Description	Value
Test for discrete variables:	Cochran-Armitage
Exact permutation distribution used:	Everywhere
Tails for discrete tests:	Upper-tailed
Strata adjustment?	No
P-value adjustment:	Permutation
Number of resamples:	10000
Seed:	36607

You should check the preceding table to verify that the analysis specifications are correct.

The Multtest Procedure

Contrast Coefficients

	Dose		
Contrast	1	2	3
CA Linear Trend	0	1	2

The preceding table lists the label and coefficients from the CONTRAST statement.

The Multtest Procedure

Discrete Variable Tabulations

Variable	Dose	Count	NumObs	Percent
S1	1	2	6	33.33
S1	2	4	6	66.67
S1	3	4	6	66.67
S2	1	4	6	66.67
S2	2	2	6	33.33
S2	3	2	6	33.33

The preceding table contains summary statistics for the two test variables, S1 and S2. The Count column lists the number of successes for each level of the class variable, Dose. The NumObs column is the sample size, and the Percent column is the percentage of successes in the sample.

The Multtest Procedure

p-Values

Variable	Contrast	Raw	Permutation
S1	CA Linear Trend	0.1993	0.3979
S2	CA Linear Trend	0.9220	1.0000

The Raw column in the preceding "p-Values" table contains the p-values from the CA test, and the Permutation column contains the permutation-adjusted p-values.

This table shows that, for S1, the adjusted p-value is almost twice the raw p-value. In fact, from theoretical considerations, the permutation-adjusted p-value for S1 should be $2 \times 0.1993 = 0.3986$. For S2, the raw p-value is 0.9220, and the adjusted p-value equals 1, as you would expect from theoretical considerations. The permutation p-values for S1 and S2 also happen to be the Bonferroni-adjusted p-values for this example.

Obs	_contrast_	_var_	_value_	upper_p
1	CA Linear Trend	S1	0	1.00000
2	CA Linear Trend	S1	1	1.00000
3	CA Linear Trend	S1	2	1.00000
4	CA Linear Trend	S1	3	1.00000
5	CA Linear Trend	S1	4	1.00000
6	CA Linear Trend	S1	5	0.99966
7	CA Linear Trend	S1	6	0.99609
8	CA Linear Trend	S1	7	0.97827
9	CA Linear Trend	S1	8	0.92205
10	CA Linear Trend	S1	9	0.80070
11	CA Linear Trend	S1	10	0.61011
12	CA Linear Trend	S1	11	0.38989
13	CA Linear Trend	S1	12	0.19930
14	CA Linear Trend	S1	13	0.07795
15	CA Linear Trend	S1	14	0.02173
16	CA Linear Trend	S1	15	0.00391
17	CA Linear Trend	S1	16	0.00034
18	CA Linear Trend	S1	17	0.00000
19	CA Linear Trend	S1	18	0.00000
20	CA Linear Trend	S1	19	0.00000
21	CA Linear Trend	S1	20	0.00000
22	CA Linear Trend	S2	0	1.00000
23	CA Linear Trend	S2	1	1.00000
24	CA Linear Trend	S2	2	1.00000
25	CA Linear Trend	S2	3	0.99966
26	CA Linear Trend	S2	4	0.99609
27	CA Linear Trend	S2	5	0.97827
28	CA Linear Trend	S2	6	0.92205
29	CA Linear Trend	S2	7	0.80070
30	CA Linear Trend	S2	8	0.61011
31	CA Linear Trend	S2	9	0.38989
32	CA Linear Trend	S2	10	0.19930
33	CA Linear Trend	S2	11	0.07795
34	CA Linear Trend	S2	12	0.02173
35	CA Linear Trend	S2	13	0.00391
36	CA Linear Trend	S2	14	0.00034
37	CA Linear Trend	S2	15	0.00000
38	CA Linear Trend	S2	16	0.00000

The preceding table lists the OUTPERM= data set, which contains the exact permutation distributions for S1 and S2 in terms of cumulative probabilities.

Example 43.2. Freeman-Tukey and *t*-Tests with Bootstrap Resampling

The data for the following example are the same as for Example 43.1, except that a continuous variable T has been added.

```
data a;
   input S1 S2 T Dose;
   datalines;
0  1  104  1
1  0   80  1
0  1  104  1
0  1  104  1
0  1  100  1
1  0  104  1
```

```
1  0   85  2
1  0   60  2
0  1   89  2
1  0   96  2
0  1   96  2
1  0   99  2
1  0   60  3
1  0   50  3
1  0   80  3
0  1   98  3
0  1   99  3
1  0   50  3
;

proc multtest data=a bootstrap nsample=10000
              pvals seed=37081 outsamp=res;
   test ft(S1 S2 / lowertailed) mean(T / lowertailed);
   class Dose;
   contrast 'Linear Trend' 0 1 2;
run;

proc print data=res(obs=36);
run;
```

The BOOTSTRAP option in the PROC MULTTEST statement requests bootstrap resampling, and NSAMPLE=10000 requests 10,000 bootstrap samples. The seed for the random number generation is 37081. The OUTSAMP=RES option requests an output SAS data set containing the 10,000 bootstrap samples.

The TEST statement specifies the Freeman-Tukey test for S1 and S2 and specifies the *t*-test for T. Both tests are lower-tailed. The grouping variable in the CLASS statement is Dose, and the coefficients across the levels of Dose are 0, 1, and 2, as specified in the CONTRAST statement. PROC PRINT displays the first 36 observations of the Res data set containing the bootstrap samples.

The results from this analysis are listed in Output 43.2.1.

Output 43.2.1. FT and *t*-tests with Bootstrap Resampling

The Multtest Procedure

Model Information

Description	Value
Test for discrete variables:	Freeman-Tukey
Test for continuous variables:	Mean t-test
Tails for discrete tests:	Lower-tailed
Tails for continuous tests:	Lower-tailed
Strata adjustment?	No
P-value adjustment:	Bootstrap
Center continuous variables?	Yes
Number of resamples:	10000
Seed:	37081

The information in the preceding table corresponds to the specifications in the invocation of PROC MULTTEST.

The Multtest Procedure

Contrast Coefficients

	Dose		
Contrast	**1**	**2**	**3**
Linear Trend	**0**	**1**	**2**

The preceding table shows the coefficients from the CONTRAST statement, and they model a linear trend.

The Multtest Procedure

Discrete Variable Tabulations

Variable	Dose	Count	NumObs	Percent
S1	1	2	6	33.33
S1	2	4	6	66.67
S1	3	4	6	66.67
S2	1	4	6	66.67
S2	2	2	6	33.33
S2	3	2	6	33.33

Continuous Variable Tabulations

Variable	Dose	NumObs	Mean	Standard Deviation
T	1	6	99.3333	9.6056
T	2	6	87.5000	14.4326
T	3	6	72.8333	22.7017

The summary statistics in the preceding table for S1 and S2 are the same as those from Example 43.1. The variables S1 and S2 are discrete, and T is a continuous variable. The mean, standard deviation, and sample size for each level of Dose is listed in the table for T. The p-values for S1 and S2 are from the Freeman-Tukey test, and the p-values for T are from the t-test.

The Multtest Procedure

p-Values

Variable	Contrast	Raw	Bootstrap
S1	Linear Trend	0.8547	1.0000
S2	Linear Trend	0.1453	0.4471
T	Linear Trend	0.0070	0.0253

The p-values are listed in the preceding table. The Raw column contains the results from the tests on the original data, and the Bootstrap column contains the bootstrap resampled adjustment to raw_p. Note that the adjusted p-values are larger than the raw p-values for all three variables. The adjusted p-values more accurately reflect the correlation of the raw p-values, the small size of the data, and the multiple testing.

Obs	_sample_	_class_	_obs_	S1	S2	T
1	1	1	11	0	1	8.5000
2	1	1	16	0	1	25.1667
3	1	1	16	0	1	25.1667
4	1	1	14	1	0	-22.8333
5	1	1	18	1	0	-22.8333
6	1	1	14	1	0	-22.8333
7	1	2	4	0	1	4.6667
8	1	2	12	1	0	11.5000
9	1	2	8	1	0	-27.5000
10	1	2	7	1	0	-2.5000
11	1	2	3	0	1	4.6667
12	1	2	12	1	0	11.5000
13	1	3	13	1	0	-12.8333
14	1	3	5	0	1	0.6667
15	1	3	8	1	0	-27.5000
16	1	3	5	0	1	0.6667
17	1	3	13	1	0	-12.8333
18	1	3	6	1	0	4.6667
19	2	1	8	1	0	-27.5000
20	2	1	3	0	1	4.6667
21	2	1	9	0	1	1.5000
22	2	1	13	1	0	-12.8333
23	2	1	14	1	0	-22.8333
24	2	1	12	1	0	11.5000
25	2	2	14	1	0	-22.8333
26	2	2	18	1	0	-22.8333
27	2	2	15	1	0	7.1667
28	2	2	6	1	0	4.6667
29	2	2	13	1	0	-12.8333
30	2	2	1	0	1	4.6667
31	2	3	7	1	0	-2.5000
32	2	3	7	1	0	-2.5000
33	2	3	6	1	0	4.6667
34	2	3	13	1	0	-12.8333
35	2	3	4	0	1	4.6667
36	2	3	6	1	0	4.6667

The preceding table lists the first 36 observations of the SAS data set resulting from the OUTSAMP=RES option in the PROC MULTTEST statement. The entire data set has 180,000 observations. The _sample_ variable is the sample indicator and _class_ indicates the resampling group, that is, the level of the CLASS variable assigned to the new observation. The number of the observation in the original data set is represented by _obs_. Also listed are the values of the original test variables, S1 and S2, and the mean-centered values of T.

Example 43.3. Peto Test

This example illustrates the use of the Peto mortality-prevalence test. In the data set, each observation represents an animal, and S1–S3 are three tumor types. A 0 in the data set indicates a nonoccurrence of the tumor, 1 indicates an incidental (nonlethal) occurrence, and 2 indicates a lethal occurrence. The time variable T indicates the time of death of the animal. A strata variable B is constructed from T, and a grouping variable Dose is drug dosage.

```
data a;
   input S1-S3 T Dose;
   if T<=90 then B=1; else B=2;
   datalines;
0  0  0  104   0
2  0  1   80   0
0  0  1  104   0
0  0  0  104   0
0  2  0  100   0
1  0  0  104   0
2  0  0   85   1
2  1  0   60   1
0  1  0   89   1
2  0  1   96   1
0  0  0   96   1
2  0  1   99   1
2  1  1   60   2
2  0  0   50   2
2  0  1   80   2
0  0  2   98   2
0  0  1   99   2
2  1  1   50   2
;

proc multtest data=a notables out=p stepsid;
   test peto(S1-S3 / permutation=20 time=T uppertailed);
   class Dose;
   strata B;
   contrast 'mort-prev' 0  1  2;
run;

proc print data=p;
run;
```

The NOTABLES option in the PROC MULTTEST statement suppresses the display of the summary statistics for each variable. The OUT=P option requests an output SAS data set containing all p-values and intermediate statistics. The STEPSID option is used to adjust the p-values.

The TEST statement specifies an upper-tailed Peto test for S1–S3, and TIME=T indicates the variable with values that are death times. The CLASS statement contains the grouping variable Dose, and the STRATA statement contains the blocking

variable B. The CONTRAST statement lists linear trend coefficients. PROC PRINT displays the requested p-value data set.

The results from this analysis are listed in Output 43.3.1.

Output 43.3.1. Peto Test

The Multtest Procedure

Model Information

Description	Value
Test for discrete variables:	Peto
Exact permutation distribution used:	Everywhere
Tails for discrete tests:	Upper-tailed
Strata adjustment?	Yes
P-value adjustment:	Stepdown Sidak

The preceding information corresponds to the PROC MULTTEST invocation. In this case the totals for all prevalence and fatality strata are less than 20, so exact permutation tests are used everywhere, and the STEPSID adjustments are computed from these permutation distributions.

The Multtest Procedure

Contrast Coefficients

	Dose		
Contrast	0	1	2
mort-prev	0	1	2

The trend coefficients are listed in the preceding table, and they happen to be the same as the levels of the Dose variable.

The Multtest Procedure

p-Values

Variable	Contrast	Raw	Stepdown Sidak
S1	mort-prev	0.0681	0.0814
S2	mort-prev	0.5000	0.5000
S3	mort-prev	0.0363	0.0781

In the preceding "p-Values" table, the p-values for the Peto tests are listed in the Raw column, and the stepdown Sidak adjusted p-values are in the Stepdown Sidak column.

The raw Peto test is significant at the 5% level for S3, but the adjusted S3 test is no longer significant at 5%. The raw and adjusted p-values for S2 are the same because of the stepdown technique. Significant p-values support the claim that higher dosage levels promote higher mortality and prevalence.

Obs	_test_	_var_	_contrast_	_strat_	_tstrat_	_value_	_exp_	_se_	raw_p	stpsid_p
1	PETO	S1	mort-prev	1	0	0	0.00000	0.00000	.	.
2	PETO	S1	mort-prev	2	0	0	0.62500	0.85696	.	.
3	PETO	S1	mort-prev	50	1	4	2.00000	1.12022	.	.
4	PETO	S1	mort-prev	60	1	3	1.75000	1.06654	.	.
5	PETO	S1	mort-prev	80	1	2	1.57143	1.04978	.	.
6	PETO	S1	mort-prev	85	1	1	0.75000	0.72169	.	.
7	PETO	S1	mort-prev	96	1	1	0.70000	0.78102	.	.
8	PETO	S1	mort-prev	98	1	0	0.00000	0.00000	.	.
9	PETO	S1	mort-prev	99	1	1	0.42857	0.72843	.	.
10	PETO	S1	mort-prev	100	1	0	0.00000	0.00000	.	.
11	PETO	S2	mort-prev	1	0	6	5.50000	1.05221	.	.
12	PETO	S2	mort-prev	2	0	0	0.00000	0.00000	.	.
13	PETO	S2	mort-prev	50	1	0	0.00000	0.00000	.	.
14	PETO	S2	mort-prev	60	1	0	0.00000	0.00000	.	.
15	PETO	S2	mort-prev	80	1	0	0.00000	0.00000	.	.
16	PETO	S2	mort-prev	85	1	0	0.00000	0.00000	.	.
17	PETO	S2	mort-prev	96	1	0	0.00000	0.00000	.	.
18	PETO	S2	mort-prev	98	1	0	0.00000	0.00000	.	.
19	PETO	S2	mort-prev	99	1	0	0.00000	0.00000	.	.
20	PETO	S2	mort-prev	100	1	0	0.00000	0.00000	.	.
21	PETO	S3	mort-prev	1	0	6	5.50000	1.05221	.	.
22	PETO	S3	mort-prev	2	0	4	2.22222	1.08298	.	.
23	PETO	S3	mort-prev	50	1	0	0.00000	0.00000	.	.
24	PETO	S3	mort-prev	60	1	0	0.00000	0.00000	.	.
25	PETO	S3	mort-prev	80	1	0	0.00000	0.00000	.	.
26	PETO	S3	mort-prev	85	1	0	0.00000	0.00000	.	.
27	PETO	S3	mort-prev	96	1	0	0.00000	0.00000	.	.
28	PETO	S3	mort-prev	98	1	2	0.62500	0.85696	.	.
29	PETO	S3	mort-prev	99	1	0	0.00000	0.00000	.	.
30	PETO	S3	mort-prev	100	1	0	0.00000	0.00000	.	.
31	PETO	S1	mort-prev	.	.	12	7.82500	2.42699	0.06808	0.08140
32	PETO	S2	mort-prev	.	.	6	5.50000	1.05221	0.50000	0.50000
33	PETO	S3	mort-prev	.	.	12	8.34722	1.73619	0.03627	0.07811

The preceding table lists the OUT= data set. The first 30 observations correspond to intermediate statistics used to compute the Peto p-values. The _test_ variable lists the name of the test, the _var_ variable lists the name of the TEST variables, and the _contrast_ variable lists the CONTRAST label. The _strat_ variable lists the level of the STRATA variable, and the _tstrat_ variable indicates whether or not the stratum corresponds to values of the TIME= variable. The _value_ variable is the observed contrast for a stratum and the _exp_ variable is its expected value. The variable _se_ contains the square root of the variance terms summed to form the denominator of the Peto statistics.

The final three observations correspond to the three Peto tests, with their p-values listed under the raw_p variable. The stpsid_p variable contains the stepdown Sidak adjusted p-values.

Example 43.4. Fisher Test with Permutation Resampling

These data, from Brown and Fears (1981), are the results from an 80-week carcinogenesis bioassay with female mice. Six tissue sites are examined at necropsy; 1 indicates the presence of a tumor and 0 the absence. A frequency variable Freq is included. A control and four different doses of a drug (in parts per milliliter) make up the levels of the grouping variable Dose.

```
data a;
   input Liver Lung Lymph Cardio Pitui Ovary Freq Dose$;
   datalines;
1   0   0   0   0   0   8    CTRL
0   1   0   0   0   0   7    CTRL
0   0   1   0   0   0   6    CTRL
0   0   0   1   0   0   1    CTRL
0   0   0   0   0   1   2    CTRL
1   1   0   0   0   0   4    CTRL
1   0   1   0   0   0   1    CTRL
1   0   0   0   0   1   1    CTRL
0   1   1   0   0   0   1    CTRL
0   0   0   0   0   0   18   CTRL
1   0   0   0   0   0   9    4PPM
0   1   0   0   0   0   4    4PPM
0   0   1   0   0   0   7    4PPM
0   0   0   1   0   0   1    4PPM
0   0   0   0   1   0   2    4PPM
0   0   0   0   0   1   1    4PPM
1   1   0   0   0   0   4    4PPM
1   0   1   0   0   0   3    4PPM
1   0   0   0   1   0   1    4PPM
0   1   1   0   0   0   1    4PPM
0   1   0   1   0   0   1    4PPM
1   0   1   1   0   0   1    4PPM
0   0   0   0   0   0   15   4PPM
1   0   0   0   0   0   8    8PPM
0   1   0   0   0   0   3    8PPM
0   0   1   0   0   0   6    8PPM
0   0   0   1   0   0   3    8PPM
1   1   0   0   0   0   1    8PPM
1   0   1   0   0   0   2    8PPM
1   0   0   1   0   0   1    8PPM
1   0   0   0   1   0   1    8PPM
1   1   0   1   0   0   2    8PPM
1   1   0   0   0   1   2    8PPM
0   0   0   0   0   0   19   8PPM
1   0   0   0   0   0   4    16PPM
0   1   0   0   0   0   2    16PPM
0   0   1   0   0   0   9    16PPM
0   0   0   0   1   0   1    16PPM
0   0   0   0   0   1   1    16PPM
1   1   0   0   0   0   4    16PPM
1   0   1   0   0   0   1    16PPM
0   1   1   0   0   0   1    16PPM
0   1   0   1   0   0   1    16PPM
0   1   0   0   0   1   1    16PPM
0   0   1   1   0   0   1    16PPM
0   0   1   0   1   0   1    16PPM
1   1   1   0   0   0   2    16PPM
0   0   0   0   0   0   14   16PPM
1   0   0   0   0   0   8    50PPM
0   1   0   0   0   0   4    50PPM
0   0   1   0   0   0   8    50PPM
```

```
0   0   0   1   0   0   1    50PPM
0   0   0   0   0   1   4    50PPM
1   1   0   0   0   0   3    50PPM
1   0   1   0   0   0   1    50PPM
0   1   1   0   0   0   1    50PPM
0   1   0   0   1   1   1    50PPM
0   0   0   0   0   0   19   50PPM
;

proc multtest data=a order=data notables out=p
              permutation nsample=1000 seed=764511;
   test fisher(Liver Lung Lymph Cardio Pitui Ovary /
               lowertailed);
   class Dose;
   freq Freq;
run;

proc print data=p;
run;
```

In the PROC MULTTEST statement, the ORDER=DATA option is required to keep the levels of Dose in the order in which they appear in the data set. Without this option, the levels are sorted by their formatted value, resulting in an alphabetic ordering. The NOTABLES option suppresses the display of summary statistics, and the OUT=P option requests an output data set containing p-values. The PERMUTATION option specifies permutation resampling, NSAMPLE=1000 requests 1000 samples, and SEED=764511 provides a starting value for the random number generator. You should specify a seed if you need to duplicate resampling results.

The TEST statement requests a lower-tailed Fisher exact test for the six tissue sites. The Fisher test is appropriate for comparing a treatment and a control, but multiple testing can be a problem. Brown and Fears (1981) use a multivariate permutation to evaluate the entire collection of tests. PROC MULTTEST adjusts the p-values by simulation.

The treatments make up the levels of the grouping variable Dose, listed in the CLASS statement. Since no CONTRAST statement is specified, PROC MULTTEST uses the default pairwise contrasts with the first level of Dose. The FREQ statement is used since this is summary data containing frequency counts of occurrences.

The results from this analysis are listed in Output 43.4.1.

Output 43.4.1. Fisher Test with Permutation Resampling

```
                    The Multtest Procedure

                      Model Information

   Description                              Value

   Test for discrete variables:             Fisher
   Tails for discrete tests:                Lower-tailed
   Strata adjustment?                       No
   P-value adjustment:                      Permutation
   Number of resamples:                     1000
   Seed:                                    764511
```

The preceding table lists the PROC MULTTEST specifications.

The Multtest Procedure

Contrast Coefficients

	Dose				
Contrast	CTRL	4PPM	8PPM	16PPM	50PPM
CTRL vs. 4PPM	1	-1	0	0	0
CTRL vs. 8PPM	1	0	-1	0	0
CTRL vs. 16PPM	1	0	0	-1	0
CTRL vs. 50PPM	1	0	0	0	-1

The preceding table lists the default contrasts for the Fisher test. Note that each dose is compared with the control.

The Multtest Procedure

p-Values

Variable	Contrast	Raw	Permutation
Liver	CTRL vs. 4PPM	0.2828	0.9640
Liver	CTRL vs. 8PPM	0.3069	0.9770
Liver	CTRL vs. 16PPM	0.7102	1.0000
Liver	CTRL vs. 50PPM	0.7718	1.0000
Lung	CTRL vs. 4PPM	0.7818	1.0000
Lung	CTRL vs. 8PPM	0.8858	1.0000
Lung	CTRL vs. 16PPM	0.5469	1.0000
Lung	CTRL vs. 50PPM	0.8498	1.0000
Lymph	CTRL vs. 4PPM	0.2423	0.9320
Lymph	CTRL vs. 8PPM	0.5898	1.0000
Lymph	CTRL vs. 16PPM	0.0350	0.2690
Lymph	CTRL vs. 50PPM	0.4161	0.9940
Cardio	CTRL vs. 4PPM	0.3163	0.9780
Cardio	CTRL vs. 8PPM	0.0525	0.3630
Cardio	CTRL vs. 16PPM	0.4506	0.9990
Cardio	CTRL vs. 50PPM	0.7576	1.0000
Pitui	CTRL vs. 4PPM	0.1250	0.7300
Pitui	CTRL vs. 8PPM	0.4948	1.0000
Pitui	CTRL vs. 16PPM	0.2157	0.9130
Pitui	CTRL vs. 50PPM	0.5051	1.0000
Ovary	CTRL vs. 4PPM	0.9437	1.0000
Ovary	CTRL vs. 8PPM	0.8126	1.0000
Ovary	CTRL vs. 16PPM	0.7760	1.0000
Ovary	CTRL vs. 50PPM	0.3689	0.9940

The preceding "p-Values" table lists p-values for the Fisher exact tests and their permutation-based adjustments. As noted by Brown and Fears, only one of the twenty-four tests is significant at the 5% level (Lymph, CTRL vs. 16PPM). Brown and Fears report a 12% chance of observing at least one significant raw p-value for 16PPM and a 9% chance of observing at least one significant raw p-value for Lymph (both at the 5% level). Adjusted p-values exhibit much lower chances of false significances. For this example, none of the adjusted p-values are close to significant.

Obs	_test_	_var_	_contrast_	_xval_	_mval_	_yval_	_nval_	raw_p	perm_p	sim_se
1	FISHER	Liver	CTRL vs. 4PPM	14	49	18	50	0.28282	0.964	0.005891
2	FISHER	Liver	CTRL vs. 8PPM	14	49	17	48	0.30688	0.977	0.004740
3	FISHER	Liver	CTRL vs. 16PPM	14	49	11	43	0.71022	1.000	0.000000
4	FISHER	Liver	CTRL vs. 50PPM	14	49	12	50	0.77175	1.000	0.000000
5	FISHER	Lung	CTRL vs. 4PPM	12	49	10	50	0.78180	1.000	0.000000
6	FISHER	Lung	CTRL vs. 8PPM	12	49	8	48	0.88581	1.000	0.000000
7	FISHER	Lung	CTRL vs. 16PPM	12	49	11	43	0.54685	1.000	0.000000
8	FISHER	Lung	CTRL vs. 50PPM	12	49	9	50	0.84978	1.000	0.000000
9	FISHER	Lymph	CTRL vs. 4PPM	8	49	12	50	0.24228	0.932	0.007961
10	FISHER	Lymph	CTRL vs. 8PPM	8	49	8	48	0.58977	1.000	0.000000
11	FISHER	Lymph	CTRL vs. 16PPM	8	49	15	43	0.03498	0.269	0.014023
12	FISHER	Lymph	CTRL vs. 50PPM	8	49	10	50	0.41607	0.994	0.002442
13	FISHER	Cardio	CTRL vs. 4PPM	1	49	3	50	0.31631	0.978	0.004639
14	FISHER	Cardio	CTRL vs. 8PPM	1	49	6	48	0.05254	0.363	0.015206
15	FISHER	Cardio	CTRL vs. 16PPM	1	49	2	43	0.45061	0.999	0.000999
16	FISHER	Cardio	CTRL vs. 50PPM	1	49	1	50	0.75758	1.000	0.000000
17	FISHER	Pitui	CTRL vs. 4PPM	0	49	3	50	0.12496	0.730	0.014039
18	FISHER	Pitui	CTRL vs. 8PPM	0	49	1	48	0.49485	1.000	0.000000
19	FISHER	Pitui	CTRL vs. 16PPM	0	49	2	43	0.21572	0.913	0.008912
20	FISHER	Pitui	CTRL vs. 50PPM	0	49	1	50	0.50505	1.000	0.000000
21	FISHER	Ovary	CTRL vs. 4PPM	3	49	1	50	0.94372	1.000	0.000000
22	FISHER	Ovary	CTRL vs. 8PPM	3	49	2	48	0.81260	1.000	0.000000
23	FISHER	Ovary	CTRL vs. 16PPM	3	49	2	43	0.77596	1.000	0.000000
24	FISHER	Ovary	CTRL vs. 50PPM	3	49	5	50	0.36889	0.994	0.002442

The preceding table lists the OUT= data set. The _test_, _var_, and _contrast_ variables provide the TEST name, TEST variable, and CONTRAST label, respectively. The _xval_, _mval_, _yval_, and _nval_ variables contain the components used to compute the Fisher exact tests from the hypergeometric distribution. The raw_p variable contains the p-values from the Fisher exact tests, and the perm_p variable contains their permutation-based adjustments. The variable sim_se is the simulation standard error from the permutation resampling.

Example 43.5. Inputting Raw *p*-Values

This example illustrates how to use PROC MULTTEST to multiplicity-adjust a collection of raw p-values obtained from some other source. This is a valuable option for those cases where PROC MULTTEST cannot compute the raw p-values directly.

```
data a;
    input Test$ Raw_P;
    datalines;
test1 .09108
test2 .69122
test3 .00177
test4 .57181
test5 .03121
test6 .01413
;

proc multtest pdata=a holm hoc fdr out=new;
run;

proc print data=new;
run;
```

Note that there are no statements other than the PROC MULTTEST statement using the p-value input mode. In this example, the raw p-values are adjusted using the Holm, Hochberg, and Benjamini and Hocherg (FDR) methods. The OUT= data set specification is required. PROC MULTTEST produces no output other than this output data set in this case, and resampling-based adjusted p-values cannot be computed.

The OUT= data set from this analysis is listed in Output 43.5.1.

Output 43.5.1. Inputting Raw *p*-Values

Obs	Test	Raw_P	stpbon_p	hoc_p	fdr_p
1	test1	0.09108	0.27324	0.27324	0.13662
2	test2	0.69122	1.00000	0.69122	0.69122
3	test3	0.00177	0.01062	0.01062	0.01062
4	test4	0.57181	1.00000	0.69122	0.68617
5	test5	0.03121	0.12484	0.12484	0.06242
6	test6	0.01413	0.07065	0.07065	0.04239

Note that the adjusted p-values for the Hochberg method (hoc_p) are less than or equal to those for the Holm method (stpbon_p). In turn, the adjusted p-values for the Benjamini and Hochberg method (fdr_p) are less than or equal to those for the Hochberg method. These comparisons hold generally for all p-value configurations. The FDR method controls the false discovery rate and not the familywise error rate. The Hochberg method controls the familywise error rate under independence. The Holm method controls the familywise error rate without assuming independence.

References

Agresti, A. (1990), *Categorical Data Analysis,* New York: John Wiley & Sons, Inc.

Armitage, P. (1955), "Tests for Linear Trend in Proportions and Frequencies," *Biometrics,* 11, 375–386.

Benjamini, Y. and Hochberg, Y. (1995), "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. B,* 57, 289–300.

Bickis, M. and Krewski, D. (1986), "Statistical Issues in the Analysis of the Long Term Carcinogenicity Bioassay in Small Rodents: an Empirical Evaluation of Statistical Decision Rules," *Environmental Health Directorate,* Health Protection Branch, Health and Welfare Canada, Ottawa, Ontario, Canada.

Brown, C.C. and Fears, T.R. (1981), "Exact Significance Levels for Multiple Binomial Testing with Application to Carcinogenicity Screens," *Biometrics,* 37, 763–774.

Cochran, W. (1954), "Some Methods for Strengthening the Common χ^2 Tests," *Biometrics,* 10, 417–451.

Dinse, G.E. (1985), "Testing for Trend in Tumor Prevalence Rates: I. Nonlethal Tumors," *Biometrics,* 41, 751–770.

Freedman, D.A. (1981), "Bootstrapping Regression Models," *Annals of Statistics,* 9, 1218–1228.

Freeman, M.F. and Tukey, J.W. (1950), "Transformations Related to the Angular and the Square Root," *Annals of Mathematical Statistics,* 21, 607–611.

Good, I.J. (1987), "A Survey of the Use of the Fast Fourier Transform for Computing Distributions," *Journal of Statistical Computation and Simulation,* 28, 87–93.

Heyse, J. and Rom, D. (1988), "Adjusting for Multiplicity of Statistical Tests in the Analysis of Carcinogenicity Studies," *Biometrical Journal,* 30, 883–896.

Hochberg, Y. (1988), "A Sharper Bonferroni Procedure for Multiple Significance Testing," *Biometrika,* 75, 800–803.

Hochberg, Y. and Tamhane, A.C. (1987), *Multiple Comparison Procedures,* New York: John Wiley & Sons, Inc.

Hoel, D.G. and Walburg, H.E. (1972), "Statistical Analysis of Survival Experiments," *Journal of the National Cancer Institute,* 49, 361–372.

Holland, B.S. and Copenhaver, M.D. (1987), "An Improved Sequentially Rejective Bonferroni Test Procedure," *Biometrics,* 43, 417–424.

Holm, S. (1979), "A Simple Sequentially Rejective Bonferroni Test Procedure," *Scandinavian Journal of Statistics,* 6, 65–70.

Lagakos, S.W. and Louis, T.A. (1985), "The Statistical Analysis of Rodent Tumorigenicity Experiments," in *Toxicological Risk Assessment,* eds. Clayson, D.B., Krewski, D., and Munro, I., Boca Raton, FL: CRC Press, 1, 144–163.

Mantel, N.(1980), "Assessing Laboratory Evidence for Neoplastic Activity," *Biometrics,* 36, 381–399.

Mantel, N. and Haenszel, W. (1959), "Statistical Aspects of Analysis of Data from Retrospective Studies of Disease," *Journal of the National Cancer Institute,* 22, 719–748.

Miller, J.J. (1978), "The Inverse of the Freeman-Tukey Double Arcsine Transformation," *The American Statistician,* 32, 138.

Pagano, M. and Tritchler, D. (1983), "On Obtaining Permutation Distributions in Polynomial Time," *Journal of the American Statistical Association,* 78, 435–440.

Peto, R., Pike, M.C., and Day, N.E. (1980), "Guidelines for Simple, Sensitive Significance Tests for Carcinogenic Effects in Long-Term Animal Experiments," *Long-term and Short-term Screening Assays for Carcinogens: A Critical Appraisal,* IARC Monographs, Annex to Supplement 2, 311–426. Lyon: International Agency for Research on Cancer.

Shaffer, J.P. (1986), "Modified Sequentially Rejective Multiple Test Procedures," *Journal of the American Statistical Association,* 81, 826–831.

Sidak, Z. (1967), "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions," *Journal of the American Statistical Association,* 62, 626–633.

Soper, K.A. and Tonkonoh, N. (1993), "The Discrete Distribution Used for the Log-Rank Test Can Be Inaccurate," *Biometrical Journal*, 35, 291–298.

Westfall, P.H. and Lin, Y. (1988), "Estimating Optimal Continuity Corrections in Run Time," *Proceedings of the Statistical Computing Section*, New Orleans: American Statistical Association, 297–298.

Westfall, P.H. and Soper, K.A.(1994), "Nonstandard Uses of PROC MULTTEST: Permutational Peto Tests; Permutational and Unconditional t and Binomial Tests," Proceedings of the Nineteenth Annual SAS User's Group International Conference, 986–989.

Westfall, P.H. and Wolfinger, R.D. (1997), "Multiple Tests with Discrete Distributions," *The American Statistician*, 51, 3–8.

Westfall, P.H. and Young, S.S. (1989), "P-value Adjustments for Multiple Tests in Multivariate Binomial Models," *Journal of the American Statistical Association*, 84, 780–786.

Westfall, P.H. and Young, S.S. (1993), *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, New York: John Wiley & Sons, Inc.

Yates, F. (1984), "Tests of Significance for 2×2 Contingency Tables," *Journal of the Royal Statistical Society*, Series A, 147, 426–463.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/STAT® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999.

SAS/STAT® User's Guide, Version 8

ISBN 1–58025–494–2

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999