

Paper 208-31

**Like Wine, the TTEST Procedure Improves with Age**

John Castelloe, SAS Institute Inc. Cary, NC

Randy Tobias, SAS Institute Inc. Cary, NC

**ABSTRACT**

PROC TTEST is an old reliable among SAS/STAT procedures, testing hypotheses about means of one and two normal samples. While this is a narrow class of analyses, it's an important class, with key applications in all areas of science, including clinical trials. Recent updates to the TTEST procedure make it an even more powerful tool for the statistician. The enhancements include:

- Equivalence tests, which answer not "Are the samples significantly different?" but rather "Are they significantly the same?"
- AB/BA crossover studies, a common repeated measures design for clinical trials

Perhaps the most useful new feature is the new plotting capability of PROC TTEST. It now produces univariate and comparative versions of histograms, box plots, and densities.

This paper describes these new features of PROC TTEST in SAS 9.2 and illustrates their use in a number of examples.

**INTRODUCTION****TTEST Nouveau**

The TTEST procedure is an old reliable in the SAS wine cellar, a bottle that satisfies even if it doesn't surprise. In versions of SAS previous to 9.2, PROC TTEST performs  $t$  tests and computes confidence intervals for one sample, two samples, and paired samples of normal data.

The TTEST procedure has matured in SAS 9.2, with graphics on the palate and new analytics in the nose. The most noticeable addition to PROC TTEST is the use of ODS Graphics to supplement the numerical results with visual displays. The graphs are automatically generated, including univariate and comparative versions of histograms, box plots, and densities. You can also customize them.

An intriguing new note is the new SIDES= option for one-sided versions of tests and confidence intervals. This gives you higher statistical power when the directionality can be assumed in advance.

The big, bold new TOST= option produces equivalence tests and intervals, popular in such situations as bioequivalence, where the typical goal is to establish whether the effects of two drugs are significantly *similar* rather than different. In this case, standard  $p$ -values from  $t$  tests are not appropriate for assessing similarity, since insignificant results may stem from low power rather than from the lack of a meaningful effect.

Finally, in addition to one-sample, two-sample, and paired designs, PROC TTEST now gets new legs with the CROSSOVER= option. AB/BA crossover designs are a favorite in clinical trials and many other arenas. They reduce variability and increase the power of treatment comparisons by administering each of two treatments to each subject.

Table 1 summarizes the major enhancements for PROC TTEST in SAS 9.2 and the required syntax.

**Table 1.** Summary of 9.2 TTEST Enhancements

Enhancement	New Syntax
ODS Graphics	<b>ods graphics on;</b> PROC TTEST <b>PLOTS</b> < <i>plot-options</i> > <b>ods graphics off;</b>
1-sided <i>p</i> -values	PROC TTEST <b>SIDES= L   U   2</b>
TOST equivalence tests	PROC TTEST <b>TOST</b> (< <i>lower</i> , > <i>upper</i> )
AB/BA crossover design	VAR var1 var2 / <b>CROSSOVER</b> (trtvar1 trtvar2) <b>IGNOREPERIOD</b>

The remainder of this paper discusses these 9.2 enhancements in detail, with a section devoted to each. Each section starts with an explanation of the statistical issues and syntax description and proceeds to an example showing input and output.

#### Example Apertif

The examples all follow a story line in which you play the role of a statistical consultant for the fictional Merry-a-Lot chain of hotels, whose recent marketing experiment involves offering a free half-carafe of house wine to any group of patrons ordering dinner. The inventor of the experiment, advertising executive Shara Glass, believes that the free wine will encourage patrons to stay longer at the restaurant, creating a more inviting atmosphere for other potential patrons in addition to spending more money themselves. Rather than studying the effect on the income from each recipient, which she considers too narrow in scope, or on the total income of the restaurant, which she considers too variable, she wants to study the impact of this “free-wine (and fee-dine)” strategy on the length of time spent by patrons in the restaurant.

#### GRAPHICS

With ODS Graphics incorporated in PROC TTEST in SAS 9.2, you can request a variety of plots using the new PLOTS= option in the PROC TTEST statement. You can use suboptions such as HISTOGRAM, BOX, and ALL to choose which plots to display or the UNPACKPANEL suboption to decant paneled plots into separate displays. For a thorough discussion of ODS Graphics for SAS/STAT procedures, see Rodriguez and Balan (2006).

The graphics for PROC TTEST are still under development at the time of this writing, but you can refer to later updates of this document or the documentation released with SAS 9.2 for more thorough treatment. The examples in this section demonstrate some summary plots you can expect to see produced by default in PROC TTEST analyses.

#### Univariate Displays

Continuing now with our consulting challenge introduced in “Example Apertif,” suppose that Shara Glass asks you to analyze the average times for patron visits reported by 42 Merry-a-Lot hotels participating in the marketing experiment (offering the free half-carafe of house wine with dinner). She would like to compare the mean for these hotel restaurants to the benchmark of 88 minutes, as determined from a previous study of locations not offering free wine. A visual summary of the data would also be nice.

The data set Wine1 is created with the following SAS statements:

```
data Wine1;
  input VisitLength @@; * in minutes;
```

```

    datalines;
  93 74 98 87 106 115 76 89 103 91 121
143 81 119 97 31 107 80 112 109 67 95
  71 107 72 116 105 66 99 104 91 132 141
  79 113 69 78 107 101 92 120 87
;
run;

```

The following SAS statements compute  $t$  tests, confidence intervals, and default plots for the Wine1 data:

```

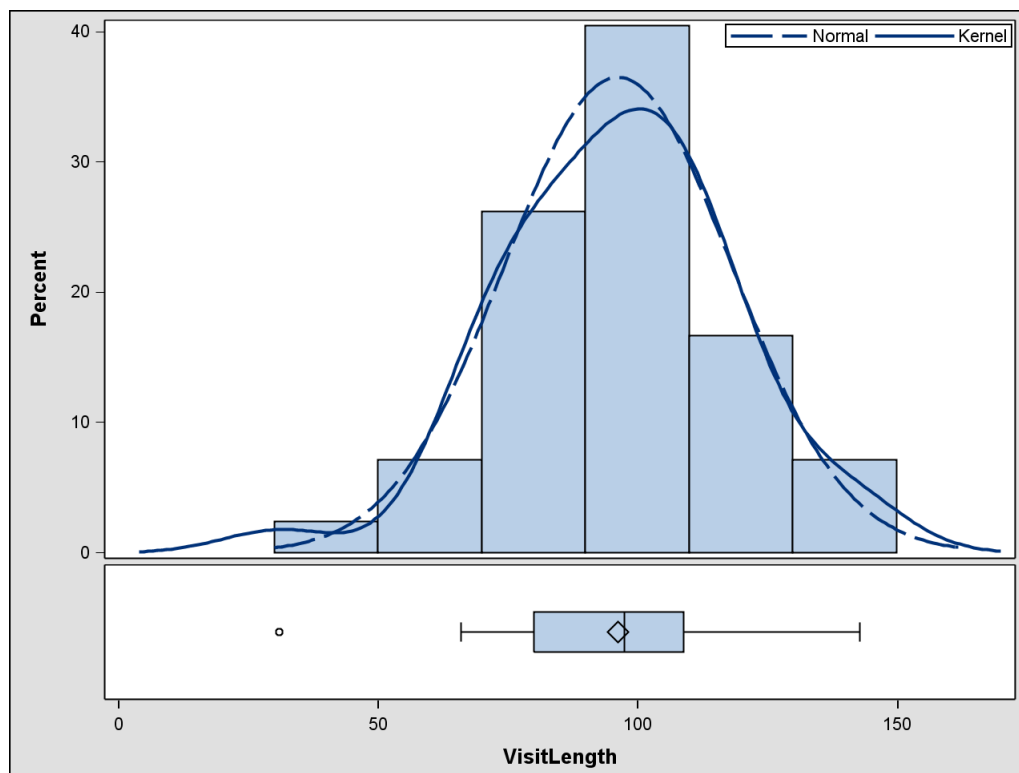
ods graphics on;
proc ttest data=Wine1 h0=88;
  var VisitLength;
run;
ods graphics off;

```

The ODS GRAPHICS ON statement enables the graphical results. With the PLOTS option omitted, the default plots are produced. Figure 1 and Figure 2 show the numerical output and a summary graph.

The TTEST Procedure						
Variable: VisitLength						
N	Mean	Std Dev	Std Err	Minimum	Maximum	
42	96.2857	21.8580	3.3728	31.0000	143.0	
Mean	95% CL Mean	Std Dev	95% CL Std Dev			
96.2857	89.4743	103.1	21.8580	17.9848	27.8725	
DF	t Value	Pr >  t				
41	2.46	0.0183				

Figure 1.  $t$  Test Analysis of Patron Visit Lengths



**Figure 2.** Histogram with Overlaid Densities and Box Plot of Patron Visit Lengths

Figure 1 reveals a significant difference between the average visit of 96.3 minutes and the null value of 88 minutes ( $p = 0.0183$ ). Overlaid on the histogram in Figure 2 are two fitted density curves, a normal density (with mean and standard deviation estimated from the data set) and a nonparametric kernel density. The box plot underneath marks the mean of 96.3 with a diamond symbol, the median with a vertical line inside the box, and the 25th and 75th percentiles as the left and right ends of the box. The “whiskers” extending from the box denote the minimum and maximum data points excluding any classified as outliers. In this case there is one observation (31 minutes) identified as an outlier by a small circle to the left of the leftmost whisker.

### Comparative Displays

A few weeks after reporting your findings to Shara, you receive a bouquet of flowers in appreciation. Attached is a small note:

I discovered that each restaurant only stocks one variety of house wine at a time. I wonder if it matters whether it's red or white. Perhaps you'd consider extending your consulting services by checking this out for me. I figured out the types of wine used at those same restaurants you analyzed before. Here is the new DATA step code.

```
data Wine2;
  input WineType $ VisitLength @@;
  datalines;
white 80 white 98 white 115 white 89 white 103
white 91 white 119 white 31 white 109 white 95
white 71 white 105 white 66 white 141 white 79
white 113 white 69 white 120 white 87
red 93 red 87 red 106 red 76 red 121 red 143
```

```

red 81 red 97 red 74 red 107 red 112 red 67
red 107 red 72 red 116 red 99 red 104 red 91
red 132 red 78 red 107 red 101 red 92
;
run;

```

You perform a two-sample analysis on this new data set with the following statements:

```

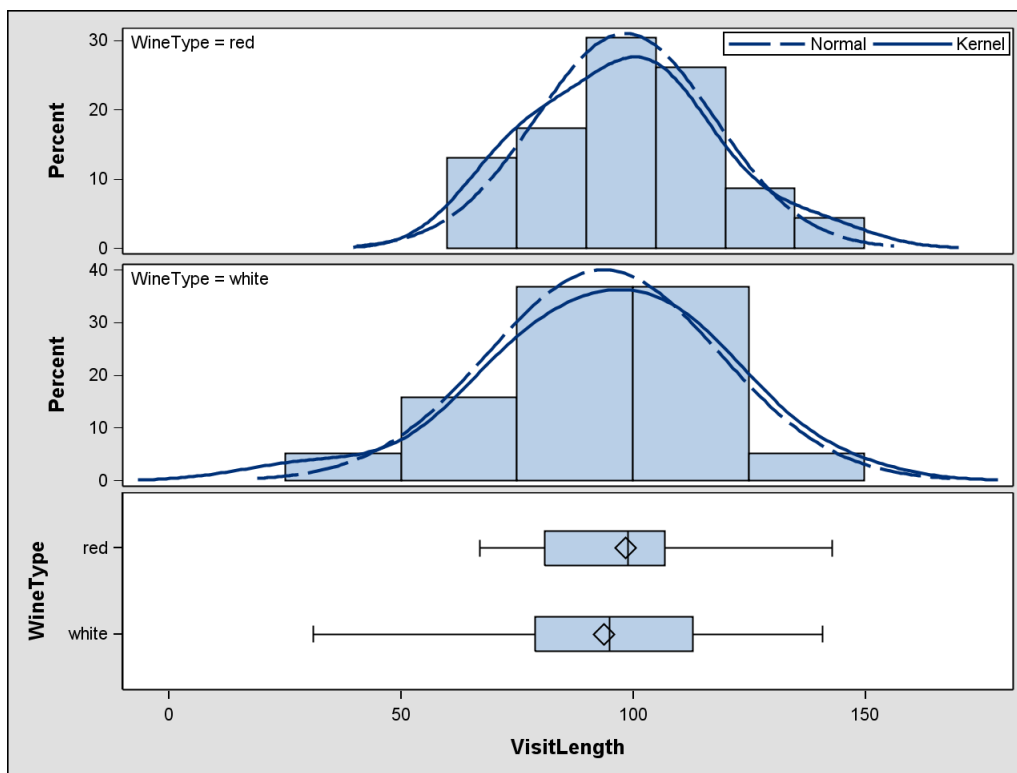
ods graphics on;
proc ttest data=Wine2;
  class WineType;
  var VisitLength;
run;
ods graphics off;

```

As with the one-sample case, the default plots are produced. Figure 3 and Figure 4 show the numerical output and a summary graph.

The TTEST Procedure						
Variable: VisitLength						
WineType	N	Mean	Std Dev	Std Err	Minimum	Maximum
red	23	98.3913	19.3032	4.0250	67.0000	143.0
white	19	93.7368	24.9061	5.7139	31.0000	141.0
Diff (1-2)		4.6545	22.0018	6.8209		
WineType	Method	Mean	95% CL Mean		Std Dev	
red		98.3913	90.0440	106.7	19.3032	
white		93.7368	81.7325	105.7	24.9061	
Diff (1-2)	Pooled	4.6545	-9.1311	18.4400	22.0018	
Diff (1-2)	Satterthwaite	4.6545	-9.5565	18.8654		
WineType	Method	95% CL		Std Dev		
red		14.9290		27.3208		
white		18.8194		36.8318		
Diff (1-2)	Pooled	18.0638		28.1514		
Diff (1-2)	Satterthwaite					
Method	Variances		DF	t Value	Pr >  t	
Pooled	Equal		40	0.68	0.4989	
Satterthwaite	Unequal		33.539	0.67	0.5100	
Equality of Variances						
Method	Num DF	Den DF	F Value	Pr > F		
Folded F	18	22	1.66	0.2549		

Figure 3. *t* Test Analysis of Red Vs. White Wine



**Figure 4.** Comparative Histograms, Densities, and Box Plots

The average visit length is about five minutes longer for red wine, but the difference between red and white is insignificant with a  $p$ -value of about 0.5 for either the usual or Satterthwaite  $t$  test. As a side note, the confidence interval using the Satterthwaite  $t$  statistic has been added in SAS 9.2.

The histograms, densities, and box plots in Figure 4 are all aligned to facilitate comparisons between distributions, means, medians, quartiles, and ranges for red wine vs. white wine.

### ONE-SIDED TESTS AND CONFIDENCE INTERVALS

Two-sided tests are generally considered more conservative than their one-sided counterparts because they do not assume that the effect is above or below the null value. But in cases where it is reasonable to assume the relative positions of the effect and the null in advance, you can use a one-sided test to achieve higher statistical power. Also, if you are only interested in a lower bound (or only in an upper bound) for the mean, then a one-sided confidence interval is more appropriate.

You could compute one-sided tests and intervals indirectly using earlier versions of PROC TTEST by utilizing the symmetry of the  $t$  distribution: double the input ALPHA= value while using only the lower computed confidence limit, and halve the computed  $p$ -value. But this indirect approach suffers from a few pitfalls, making it difficult to represent the actual error rate and to round correctly after halving odd decimal values.

In SAS 9.2 you can perform upper one-sided  $t$  tests, where “upper” denotes assumption of an effect larger than the null value, directly using the new SIDES=U option. This also produces upper one-sided confidence intervals ranging from a finite lower bound to infinity. Similarly, the SIDES=L option computes lower one-sided  $t$  tests and intervals.

You find a cordial letter from Shara Glass in your mailbox a few days later:

I figured I'd present a one-sided version of that original analysis at our next executive meeting. I don't think anyone will question the assumption that the free wine would *increase* the average

patron visit time if it differed at all. Since I can report a higher lower bound with a one-sided confidence interval, can you please produce one for me? And I figure the new  $p$ -value will be 0.0091 or 0.0092, but I can't tell which by just looking at my printout from last time. (Neal from Decision Support might care about that fourth decimal place.)

The simple addition of the SIDES=U option to the original PROC TTEST analysis is sufficient for the task:

```
proc ttest data=Wine1 h0=88 sides=U;
  var VisitLength;
run;
```

A look at the output in Figure 5 shows a one-sided confidence interval of  $[90.6098, \infty)$  minutes, where the lower bound has increased from 89.4743 in the original two-sided analysis. Note also that the  $p$ -value is more highly significant, at 0.0092.

The TTEST Procedure					
Variable: VisitLength					
N	Mean	Std Dev	Std Err	Minimum	Maximum
42	96.2857	21.8580	3.3728	31.0000	143.0
Mean	95% CL Mean	Std Dev	95% CL Std Dev		
96.2857	90.6098	Infty	21.8580	17.9848	27.8725
DF	t Value	Pr > t			
41	2.46	0.0092			

Figure 5. One-Sided  $t$  Test Analysis of Patron Visit Lengths

## EQUIVALENCE TESTS

Suppose that instead of testing for a *difference* between two means, you want to test whether the means are significantly *similar*. This situation is quite common in bioequivalence testing, which aims to demonstrate lack of clinically meaningful differences between the side effects or efficacy of competing drugs. Insignificant  $p$ -values in usual  $t$  tests do *not* imply equivalence, because it could be that the sample size is too low to declare an inherent difference as significant.

A popular statistical method for such an *equivalence analysis* is Schuirman's (1987) TOST ("Two One-Sided Test") method, which involves conducting two one-sided  $t$  tests for the mean or mean difference. For a thorough discussion of the TOST and related bioequivalence and bioavailability methods, see Chow and Liu (2000).

To conduct a TOST equivalence analysis, you must specify *equivalence bounds* to define the null and alternative hypotheses. You can request an equivalence test and specify these bounds using the new TOST option in the PROC TTEST statement.

Once again you receive a gift from Shara Glass, a basket of fruit. A polite little note is attached:

The folks at the meeting were interested in that red/white comparison from before, but not for the reasons I had guessed. They're encouraged by the overall increase in patron visit lengths, but they're hoping for a *lack* of a distinction between red and white wines. Apparently the restaurant

managers prefer to choose their wine type based on changing trends and supplier discounts. So folks are looking for a reason to downplay the importance of wine type when it comes to getting patrons to stay a little longer. I was about to reassure them by reminding them how insignificant that  $p$ -value was (around 0.5), but then I realized that it doesn't really prove anything about the similarity of red and white wines...it only demonstrates that we didn't have enough evidence to conclude a significant *difference*! Neal from Decision Support proposed a toast when I told him that, but I'm not sure why, since we weren't drinking anything at the time.

You chuckle at the last sentence, realizing that Neal was on the right track. So you write back to Shara proposing a "TOST" and ask her to decide what length of patron visit time she and her colleagues would consider a minimally meaningful difference in comparing red and white wines. She settles on a criterion of a 10-minute difference (in either direction). So, you should use  $-10$  and  $10$  as equivalence bounds, defining the null and alternative hypotheses as follows:

$$H_0: \mu_{\text{diff}} < -20 \quad \text{or} \quad \mu_{\text{diff}} > 20$$

$$H_1: \quad \quad -20 \leq \mu_{\text{diff}} \leq 20$$

where  $\mu_{\text{diff}}$  is the mean visit length with red wine minus the mean visit length with white wine.

You can perform the TOST equivalence analysis by specifying TOST( $-10, 10$ ):

```
proc ttest data=Wine2 tost(-10,10);
  class WineType;
  var VisitLength;
run;
```

In this case, you could alternatively use the syntax shorthand TOST(10) since the lower and upper bounds are equidistant from the null value of 0.

The two parts of the TOST equivalence test are

1. an upper 1-sided  $t$  test with a null mean difference of  $-10$
2. a lower 1-sided  $t$  test with a null mean difference of  $10$

The overall  $p$ -value of the TOST is the maximum  $p$ -value from these two individual tests. The next-to-last table in Figure 6 reveals an overall  $p$ -value of 0.2189, which is insignificant at the default  $\alpha=0.05$  level, although more significant than that 0.5 from before. Therefore, with  $\alpha=0.05$ , there is not enough evidence in favor of the equivalence of the red and white wines.



The TTEST Procedure						
Variable: VisitLength						
WineType	N	Mean	Std Dev	Std Err	Minimum	Maximum
red	23	98.3913	19.3032	4.0250	67.0000	143.0
white	19	93.7368	24.9061	5.7139	31.0000	141.0
Diff (1-2)		4.6545	22.0018	6.8209		
WineType	Method	Mean	95% CL Mean		Std Dev	
red		98.3913	90.0440	106.7	19.3032	
white		93.7368	81.7325	105.7	24.9061	
Diff (1-2)	Pooled	4.6545	-9.1311	18.4400	22.0018	
Diff (1-2)	Satterthwaite	4.6545	-9.5565	18.8654		
WineType	Method	95% CL		Std Dev		
red		14.9290	27.3208			
white		18.8194	36.8318			
Diff (1-2)	Pooled	18.0638	28.1514			
Diff (1-2)	Satterthwaite					
WineType	Method	Mean				
Diff (1-2)	Pooled	4.6545				
Diff (1-2)	Satterthwaite	4.6545				
WineType	Method	Lower Bound	90% CL Mean		Upper Bound	Assessment
Diff (1-2)	Pooled	-10 <	-6.8309	16.1399	> 10	Not equivalent
Diff (1-2)	Satterthwaite	-10 <	-7.1683	16.4772	> 10	Not equivalent
Method	Variances	Test	Null	DF	t Value	Pr > t
Pooled	Equal	Upper	-10	40	2.15	0.0189
Pooled	Equal	Lower	10	40	-0.78	0.2189
Pooled	Equal	Overall				0.2189
Satterthwaite	Unequal	Upper	-10	33.539	2.10	0.0218
Satterthwaite	Unequal	Lower	10	33.539	-0.76	0.2249
Satterthwaite	Unequal	Overall				0.2249
Equality of Variances						
Method	Num DF	Den DF	F Value	Pr > F		
Folded F	18	22	1.66	0.2549		

**Figure 6.** TOST Equivalence Analysis of Red and White Wines

Since rejection of the TOST at level  $\alpha$  is the same as capture of the  $100(1-2\alpha)\%$  confidence interval completely within the equivalence bounds, the  $100(1-2\alpha)\% = 90\%$  confidence interval of  $[-6.8309, 16.1399]$  for the mean difference is also shown. Comparing this interval to the equivalence bounds of -10 and 10, you can see that the upper confidence limit falls outside the bounds, resulting in the "Not equivalent" assessment.

Note that the analogous results using the Satterthwaite  $t$  test, which allows unequal variances for the two

groups, agree with those based on the usual  $t$  test.

### AB/BA CROSSOVER DESIGNS

Crossover trials are experiments in which each subject is given a sequence of different treatments. They are especially common in clinical trials for medical studies. The reduction in variability from taking multiple measurements on a subject enables more precise treatment comparisons. The simplest such design is the AB/BA crossover, in which each subject receives each of two treatments in a randomized order.

The washout period between the two measurement periods in an AB/BA crossover design helps to reduce the chances of nontrivial carryover effects and interaction effects among subjects, treatments, and periods. If such effects are assumed negligible, then three primary tests of interest emerge:

- test of treatment effect assuming negligible period effect
- test of treatment effect allowing for possible period effect
- test of period effect

All three of these tests reduce to two-sample or paired  $t$  tests on simple transformations of the data (hence the choice of PROC TTEST as a new home for these analyses).

- If the period effect is allowed for, then the data are split into two groups according to the two sequences, and both treatment and period effects can be tested:
  - A two-sample  $t$  test on halved period differences between the sequences assesses the treatment effect.
  - A two-sample  $t$  test on halved crossover differences, defined as period differences but with those for the second sequence negated, between sequences assesses the period effect.
- If the period effect is assumed negligible, then a paired  $t$  test comparing the A and B measurements, regardless of sequence, for each subject assesses the treatment effect.

See Senn (2002, Chapter 3) for a more detailed discussion.

It is possible to analyze an AB/BA crossover design by using earlier versions of PROC TTEST after computing the appropriate transformations in the DATA step or by using linear models procedures such as MIXED, GLM, or ANOVA. However, explicit coverage of this design in PROC TTEST allows the inclusion of specialized output and graphs that have been developed specifically for crossover designs.

You can perform an analysis of the treatment and period effects, assuming a model including terms for both effects, by using the new Crossover option in the VAR statement of the TTEST procedure. To ignore the period effect (that is, assume it is zero) and gain one more degree of freedom in testing the treatment difference, use the IGNOREPERIOD option as well.

Continuing with our wine example, a large package arrives at your door a month later. You discover that Shara has sent you the oakiest gift yet, a tree. There is, of course, a note attached:

I hope you know I'm not harboring any tannins about the insignificant results in both of the red/white analyses you did for me. I realize the sample size was probably too low to achieve enough power for those tests. I'd have to be insane to blame a statistician who wasn't even consulted during the design phase!

We've done a confirmatory study on the free-wine strategy. This time we asked 30 restaurants to offer the free half-carafe for a period of two months, then spend a month offering it only to patrons who happened to ask about it, then discontinue the practice for the following two months. Another 30 restaurants did the same thing but with the order reversed. They all reported the average patron visit lengths during each of the two-month periods.

Hopefully this has reduced the variability from the previous study. And now we have a control group to improve the legitimacy of the results too. Please see the new DATA step code on the other side of this note.

```

data Wine3;
  input Free1 $ Free2 $ Visit1 Visit2 @@;
  datalines;
yes no 96 110 yes no 109 105 yes no 104 98
yes no 97 64 yes no 92 59 yes no 90 77
yes no 74 75 yes no 136 98 yes no 104 98
yes no 94 84 yes no 95 88 yes no 107 100
yes no 79 86 yes no 112 76 yes no 94 101
yes no 86 70 yes no 103 82 yes no 86 106
yes no 90 76 yes no 98 88 yes no 86 90
yes no 109 93 yes no 98 71 yes no 89 65
yes no 104 75 yes no 127 110 yes no 91 105
yes no 114 80 yes no 81 103 yes no 98 75
no yes 80 99 no yes 44 116 no yes 81 104
no yes 98 101 no yes 104 108 no yes 93 69
no yes 57 83 no yes 95 92 no yes 106 98
no yes 81 113 no yes 44 68 no yes 100 86
no yes 81 110 no yes 70 90 no yes 96 84
no yes 77 83 no yes 112 83 no yes 112 90
no yes 71 71 no yes 52 83 no yes 64 84
no yes 97 94 no yes 75 78 no yes 70 103
no yes 80 108 no yes 79 113 no yes 72 124
no yes 86 103 no yes 94 102 no yes 91 94
;
run;

```

You instantly recognize this study as an AB/BA crossover design with the variables Visit1 and Visit2 representing the average visit lengths for each restaurant in periods 1 and 2. The variables Free1 and Free2 represent the “treatment” (free wine, yes or no) applied in each period. You write the following SAS statements to analyze these data:

```

proc ttest data=Wine3;
  var Visit1 Visit2 / crossover= (Free1 Free2);
run;

```

The TTEST Procedure							
Response Variables: Visit1 Visit2							
Crossover Variable Information							
		Period	Response	Treatment			
		1	Visit1	Free1			
		2	Visit2	Free2			
Treatment	Period	N	Mean	Std Dev	Std Err	Minimum	Maximum
no	1	30	82.0667	18.3189	3.3446	44.0000	112.0
no	2	30	86.9333	14.7366	2.6905	59.0000	110.0
yes	1	30	98.1000	13.3787	2.4426	74.0000	136.0
yes	2	30	94.4667	14.4240	2.6334	68.0000	124.0
Diff (no-yes)			-11.7833	10.0042	2.5831		
	Diff (1-2)		-0.6167	10.0042	2.5831		
Treatment	Period	Method	Mean	95% CL Mean	Std Dev		
no	1		82.0667	75.2263 88.9071	18.3189		
no	2		86.9333	81.4306 92.4361	14.7366		
yes	1		98.1000	93.1043 103.1	13.3787		
yes	2		94.4667	89.0807 99.8527	14.4240		
Diff (no-yes)		Pooled	-11.7833	-16.9539 -6.6128	10.0042		
Diff (no-yes)		Satterthwaite	-11.7833	-16.9624 -6.6043			
	Diff (1-2)	Pooled	-0.6167	-5.7872 4.5539	10.0042		
	Diff (1-2)	Satterthwaite	-0.6167	-5.7957 4.5624			
Treatment	Period	Method	95% CL	Std Dev			
no	1		14.5893	24.6264			
no	2		11.7363	19.8107			
yes	1		10.6549	17.9852			
yes	2		11.4873	19.3903			
Diff (no-yes)		Pooled	8.4689	12.2246			
Diff (no-yes)		Satterthwaite					
	Diff (1-2)	Pooled	8.4689	12.2246			
	Diff (1-2)	Satterthwaite					
Treatment	Period	Method	Variances	DF	t Value	Pr >  t	
Diff (no-yes)		Pooled	Equal	58	-4.56	<.0001	
Diff (no-yes)		Satterthwaite	Unequal	53.863	-4.56	<.0001	
	Diff (1-2)	Pooled	Equal	58	-0.24	0.8122	
	Diff (1-2)	Satterthwaite	Unequal	53.863	-0.24	0.8122	
Equality of Variances							
Method	Num DF	Den DF	F Value	Pr > F			
Folded F	29	29	1.77	0.1312			

Figure 7. Analysis of Treatment and Period Effects in Crossover Design

The results in Figure 7 indicate an even more optimistic conclusion about the effectiveness of the free wine strategy than the original analysis in "Graphics" showed. The estimated "treatment" difference between the mean patron visit times, shown in the second table, is 11.7833, compared to 8.3 previously, with a 95% confidence interval [6.6128, 16.9539]. The next-to-last table reveals a highly significant  $p < 0.001$  for the treatment effect using the usual  $t$  test or the Satterthwaite  $t$  test.

The period effect is not significant at the  $\alpha=0.05$  level, with  $p=0.8122$  using the usual  $t$  test or the Satterthwaite  $t$  test. In other words, it does not seem to matter whether the restaurants offered the free wine in the first period or the second.

The question of equal variances, which determines the choice of usual vs. Satterthwaite  $t$  test and characterizes the Equality of Variances test, is the same for both the treatment and period effect tests, as both of these tests are two-sample  $t$  tests using the treatment sequence as the group.

Adding the IGNOREPERIOD option requests the analysis assuming no period effect:

```
proc ttest data=Wine3;
  var Visit1 Visit2 / crossover= (Free1 Free2) ignoreperiod;
run;
```

The TTEST Procedure							
Response Variables: Visit1 Visit2							
Crossover Variable Information							
		Period	Response	Treatment			
		1	Visit1	Free1			
		2	Visit2	Free2			
Treatment	Period	N	Mean	Std Dev	Std Err	Minimum	Maximum
no	1	30	82.0667	18.3189	3.3446	44.0000	112.0
no	2	30	86.9333	14.7366	2.6905	59.0000	110.0
yes	1	30	98.1000	13.3787	2.4426	74.0000	136.0
yes	2	30	94.4667	14.4240	2.6334	68.0000	124.0
Diff (no-yes)		60	-11.7833	19.8478	2.5623	-72.0000	29.0000
Treatment	Period	Mean	95% CL Mean	Std Dev	95% CL Std Dev		
no	1	82.0667	75.2263 88.9071	18.3189	14.5893 24.6264		
no	2	86.9333	81.4306 92.4361	14.7366	11.7363 19.8107		
yes	1	98.1000	93.1043 103.1	13.3787	10.6549 17.9852		
yes	2	94.4667	89.0807 99.8527	14.4240	11.4873 19.3903		
Diff (no-yes)		-11.7833	-16.9106 -6.6561	19.8478	16.8237 24.2076		
		Treatment	DF	t Value	Pr >  t		
		Diff (no-yes)	59	-4.60	<.0001		

**Figure 8.** Analysis of Treatment Ignoring Period in Crossover Design

The test of treatment difference ignoring period, shown in Figure 8, is also highly significant. The treatment estimate is the same as in Figure 7 because the design is balanced, but in general the estimates might differ slightly for unbalanced designs (that is, unequal numbers of observations for the two sequences no-yes and yes-no). The standard deviation, and hence confidence intervals and  $p$ -values, in general differs regardless of whether the design is balanced or not.

It is also worth noting that you can analyze multiple pairs of response variables in a crossover design. For example, if you also measured the income from patrons, and the variables Income1 and Income2 represented these measurements, you could append a crossover analysis of income with the following statements:

```
proc ttest data=Wine3;  
  var Visit1 Visit2 Income1 Income2 / crossover= (Free1 Free2);  
run;
```

Recapping our consulting project for Shara Glass and Merry-a-Lot, the two-sample  $t$  test and TOST equivalence test revealed a lack of evidence for either a strong difference or strong similarity between the effects of red and white wines. But the one-sided test using SIDES=U and the follow-up AB/BA crossover analysis demonstrated the overall potent effect of the free half-carafe, encouraging patrons to stay around 8 to 12 minutes longer at dinnertime. Clearly, restaurants should offer free wine, whether red or white!

## CONCLUSION

The appreciably matured TTEST procedure in SAS 9.2 offers a blending of new graphs (univariate and comparative), analyses (one-sided alternatives and TOST equivalence tests), and designs (the AB/BA crossover). Further 9.2 development work beyond the scope of this paper includes more graphics (QQ plots, subject-profile plots for paired and crossover data, and several other graphs for AB/BA to visualize treatment and period effects) and support for lognormal data and mean ratios.

Please visit [www.sas.com/statistics](http://www.sas.com/statistics) for updated versions of this paper.

## REFERENCES

- Chow, S. and Liu, J. (2000), *Design and Analysis of Bioavailability and Bioequivalence Studies, Second Edition*, New York: Marcel Dekker, Inc.
- Jones, B. and Kenward, M.G. (2003), *Design and Analysis of Cross-Over Trials, Second Edition*, Washington, D.C.: Chapman & Hall/CRC.
- Rodriguez, R.N. and Balan, T.E. (2006), "Creating Statistical Graphics in SAS® 9.2: What Every Statistical User Should Know," *Proceedings of the Thirty-first Annual SAS Users Group International Conference*, Paper 192-31, Cary, NC: SAS Institute.
- Schuurmann, D.J. (1987), "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability," *Journal of Pharmacokinetics and Biopharmaceutics*, 15, 657–680.
- Senn, S. (2002), *Cross-over Trials in Clinical Research, Second Edition*, New York: John Wiley & Sons, Inc.

SAS and SAS/STAT are registered trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.

## CONTACT INFORMATION

John Castelloe, SAS Institute Inc.,  
SAS Campus Drive, Cary, NC 27513  
Email: [john.castelloe@sas.com](mailto:john.castelloe@sas.com)

Version 1.0