Chapter 254

Stratified Wilcoxon-Mann-Whitney (van Elteren) Test

Introduction

This module computes power and sample size for stratified Wilcoxon-Mann-Whitney (WMW) tests. This test is often referred to as the *van Elteren* (vE) test. It is used to compare two populations (groups) when the underlying data distributions are continuous (possibly non-normal) with no ties.

Technical Details

The power and sample size formulae presented here are given in Zhao (2006), Zhao et al. (2006), Guo (2012), and Lehmann (1975).

Van Elteren Test Statistic

Suppose variables Y_1 (group 1) and Y_2 (group 2) each have similar types of data. For example, both variables may be continuous, ordinal, or a combination of continuous and ordinal (such as when a continuous variable is combined with a point-mass at zero). In the following, we assume that higher values are “better”.

The difference between the distributions of Y_1 and Y_2 is measured by the competing probability

$$P1 = \Pr(Y_1 > Y_2) + \frac{1}{2}\Pr(Y_1 = Y_2)$$

The null hypothesis of no difference between the two distributions is given by H_0: $P1 = \frac{1}{2}$.

Suppose to give more precision, the *N* data values in the study are separated into *H* strata (or blocks). If we let N_h be the sample size of stratum *h* and N_{1h} and N_{2h} be the corresponding sample sizes of groups 1 and 2 in stratum *h*, then we have

$$N = \sum_{h=1}^{H} N_h = \sum_{h=1}^{H} (N_{1h} + N_{2h})$$

The van Elteren test statistic for testing the null hypothesis of no difference may be computed by first computing the Wilcoxon rank-sum statistic, W_h, for the data in each stratum, and then forming a weighted sum of the W_h as follows.

First, the combined data from both groups within a particular stratum are ranked (ties use the average rank) and W_h is computed as the sum of the ranks associated with group 1 responses. Thus,

$$W_h = \sum_{i=1}^{N_{1h}} \text{rank}(Y_{1hi})$$

The van Elteren test statistic is given by

$$U = \sum_{h=1}^{H} c_h W_h$$

where $c_h = (N_{1h} + N_{2h} + 1)^{-1}$.

Van Elteren Test Power and Sample Size

The large-sample mean and variance of *U* under H_0 is given by

$$E(U|H_0) = \sum_{h=1}^{H} \frac{1}{2} N_{1h} = \mu_0$$

$$V(U|H_0) = \sum_{h=1}^{H} \frac{1}{12} c_h N_{1h} N_{2h} = v_0$$

The large-sample mean and variance of *U* under H_1 is given by

$$E(U|H_1) = \sum_{h=1}^{H} c_h \left[\frac{N_{1h}(N_{1h}+1)}{2} + N_{1h} N_{2h} \pi_{h1} \right] = \mu_1$$

$$V(U|H_1) = (complicated\ expression) = v_1$$

where $P_{1h} = Pr(Y_{1h} \geq Y_{2h})$.

To simplify the calculations, the assumption is made that $v_1 = v_0$ so that the complicated expression for v_1, which involves some difficult quantities to estimate during study planning, can be avoided.

The final formula, which can be arranged to solve either for sample size or power, is

$$N = \frac{v_0}{u^2}\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2$$

where $z_{1-\frac{\alpha}{2}}$ and $z_{1-\beta}$ are lower quantiles of a standard normal distribution and

$$u = \sum_{h=1}^{H} \frac{N_{1h}}{N_h}\left(1 - \frac{N_{1h}}{N_h}\right)\frac{N_h}{N}\left(P_{1h} - \frac{1}{2}\right)$$

$$v_0 = \frac{1}{12}\sum_{h=1}^{H} \frac{N_{1h}}{N_h}\left(1 - \frac{N_{1h}}{N_h}\right)\frac{N_h}{N}$$

Planning Estimates of P_{1h}

We see that in order to compute an estimate of the required sample size, we need planning estimates of the P_{1h}. These can be obtained from previous studies, from pilot studies, or from a form of simulation.

The simulation would proceed as follows. Generate random samples of Y_1 and Y_2 that adhere to the sample size structure defined by the $N, N_{1h}, \text{and } N_{2h},\ h = 1, \ldots, H$. Y_1 and Y_2 are generated so that the variances are the same and the means are different by the amount that the experiment is to detect. From this simulated sample, estimate P_{1h} using

$$\hat{P}_{1h} = \frac{1}{N_{1h}N_{2h}}\sum_{i=1}^{N_{1h}}\sum_{j=1}^{N_{2h}} \delta\left(Y_{1i} - Y_{2j}\right)$$

where

$$\delta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0.5 & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Repeat this simulation several times until a reasonable set of planning estimates is obtained. Use the means of the simulated P_{1h} as the planning estimates for the sample size calculation.

Another way to estimate the P_{1h} is to use the suggestion of Lehmann (1975) which is to use

$$\hat{P}_{1h} = \frac{1}{2} + \frac{\mu_{1h} - \mu_{2h}}{2\sigma_h\sqrt{\pi}}$$

Unfortunately, simulation studies conducted by Zhao (2006) showed that this estimate is only accurate when Y_1 and Y_2 both follow a normal distribution. Otherwise, the resulting sample size is severely inflated. This is disappointing because if the data are truly normally distributed, the parametric ANOVA test is more accurate.

Example 1 – Finding the Sample Size

Suppose a clinical trial is planned to compare the response to certain treatment (group 1) to a control (group 2). The subjects will be divided three strata. In each stratum, the subjects will be further separated into two groups: one receiving the treatment of interest and the other receiving the control. A continuous, non-normal response with no ties is expected. Previous studies have shown that P_{1h} is constant across strata and ranges between 0.55 and 0.65.

They want to look at the sample size requirements to achieve a power of 0.80. They want to set alpha to 0.05 and analyze the results with a two-sided test.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For	**Sample Size**
Alternative Hypothesis	**Two-Sided**
Power	**0.8**
Alpha	**0.05**
Strata Allocation	**All Equal (Search for Common Stratum Sample Size, N0)**
Percent of Nh in Group 1	**All Equal**
Percent of Nh in Group 1 for All Strata	**50**
P1h (Probability Y1 ≥ Y2)	**All Equal**
P1h (Probability Y1 ≥ Y2) for All Strata	**0.55 0.60 0.65**
Set 1 Number of Strata	**1**
Set 2 Number of Strata	**1**
Set 3 Number of Strata	**1**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Reports

Numeric Results

Solve For:	Sample Size
Number of Strata:	3
Alternative Hypothesis:	Two-Sided
Strata Allocation:	Equal (Search for N0)

	Sample Size						
Power	Total N	Group 1 N1	Group 2 N2	Stratum N0	Percent of Nh in Group 1 %N1h	Effect Size P(Y1 ≥ Y2) P1h	Alpha
0.80018	1047	525	522	349	50	0.55	0.05
0.80353	264	132	132	88	50	0.60	0.05
0.80216	117	60	57	39	50	0.65	0.05

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The total number of subjects.
N1	The total sample size of group 1.
N2	The total sample size of group 2.
N0	The common sample size of all strata.
%N1h	The percent of the stratum sample size, Nh, that is in group 1.
P1h	The common probability that P(Y1 ≥ Y2) in all stratum. Yi (i = 1, 2) is a random response from group i. This is the effect size.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A two-group stratified design with 3 strata will be used to test the null hypothesis of no difference between distributions. The comparison will be made using a two-sided Wilcoxon-Mann-Whitney (van Elteren) test, with a Type I error rate (α) of 0.05. The effect size of the test is based on the P1h values for each of the strata, where P1h is the probability that an observation from Group 1 exceeds an observation from Group 2, within stratum h. The P1h values for the 3 strata are assumed to be (0.55 0.55 0.55). To detect P1h values of (0.55 0.55 0.55), with 20% power, the number of subjects needed will be 525 in Group 1, with Group 1 stratum sample sizes of (175 175 175), and 522 in Group 2, with Group 2 stratum sample sizes of (174 174 174).

Strata-Detail Report for Row 1

Stratum ID h	Stratum Allocation Rh	Stratum Allocation Prop	Stratum Sample Size Total Nh	Stratum Sample Size Group 1 N1h	Stratum Sample Size Group 2 N2h	Percent of Nh in Group 1 %N1h	Effect Size P(Y1 ≥ Y2) P1h
1	1	0.33333	349	175	174	50	0.55
2	1	0.33333	349	175	174	50	0.55
3	1	0.33333	349	175	174	50	0.55

h An arbitrary stratum identification number.
Rh The stratum allocation value for each stratum.
Prop The proportion of the sample size allocated to stratum h.
Nh The sample size of stratum h.
N1h The sample size of group 1 of stratum h.
N2h The sample size of group 2 of stratum h.
%N1h The percentage of Nh in group 1.
P1h The probability that Y1 ≥ Y2, where these are randomly selected individuals from groups 1 and 2.

Strata-Detail Report for Row 2

Stratum ID h	Stratum Allocation Rh	Stratum Allocation Prop	Stratum Sample Size Total Nh	Stratum Sample Size Group 1 N1h	Stratum Sample Size Group 2 N2h	Percent of Nh in Group 1 %N1h	Effect Size P(Y1 ≥ Y2) P1h
1	1	0.33333	88	44	44	50	0.6
2	1	0.33333	88	44	44	50	0.6
3	1	0.33333	88	44	44	50	0.6

Strata-Detail Report for Row 3

Stratum ID h	Stratum Allocation Rh	Stratum Allocation Prop	Stratum Sample Size Total Nh	Stratum Sample Size Group 1 N1h	Stratum Sample Size Group 2 N2h	Percent of Nh in Group 1 %N1h	Effect Size P(Y1 ≥ Y2) P1h
1	1	0.33333	39	20	19	50	0.65
2	1	0.33333	39	20	19	50	0.65
3	1	0.33333	39	20	19	50	0.65

References

Zhao, Yan D. 2006. 'Sample size estimation for the van Elteren test--a stratified Wilcoxon-Mann-Whitney test.' Statistics in Medicine, 25, pages 2675-2687.

Zhao, Y.D., Qu, Y., Rahardja, D. 2006. 'Power Approximation for the van Elteren test based on location-scale family of distributions.' Journal of Biopharmaceutical Statistics, 16, pages 803-815.

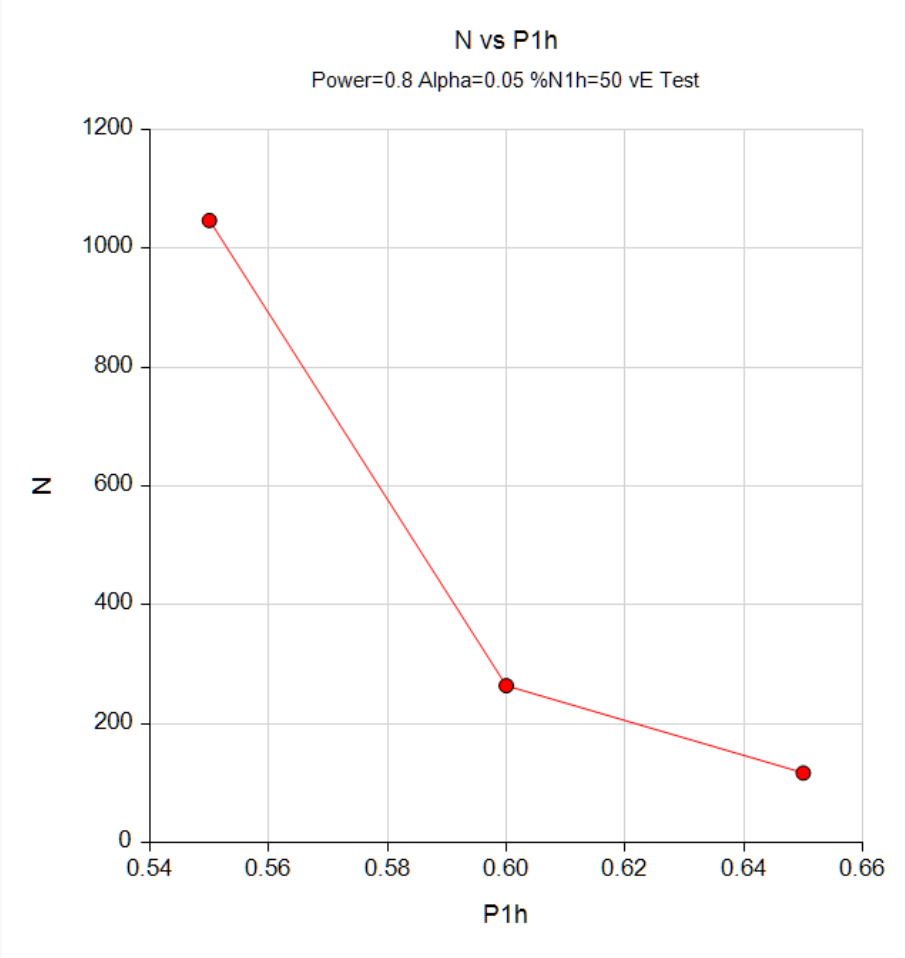
Guo, J.H. 2012. 'Optimal sample size planning for the Wilcoxon-Mann-Whitney and van Elteren tests under cost constraints.' Journal of Applied Statistics, 39(10), pages 2153-2164.

Lehmann, E.L. 1975. 'Nonparametrics - Statistical Methods Based on Ranks.' Holden-Day, Inc. Oakland, California.

This report shows the numeric results of this sample size study. The definitions of the items on the report are given in the Report Definitions section.

Plots Section

Plots



This plot gives a visual presentation to the results in the Numeric Report.

Example 2 – Validation using Zhao et al. (2006)

Zhao *et al.* (2006) pages 807 - 809 present an example in which the sample size found to be 258 when alpha is 0.05, power is 80%, both group and stratum allocation is equal, and P_{1h} is 0.602 in the first stratum and 0.600 in the second. The test is two-sided.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For	**Sample Size**
Alternative Hypothesis	**Two-Sided**
Power	**0.8**
Alpha	**0.05**
Strata Allocation	**All Equal (Search for Common Stratum Sample Size, N0)**
Percent of Nh in Group 1	**All Equal**
Percent of Nh in Group 1 for All Strata	**50**
P1h (Probability Y1 ≥ Y2)	**Custom**
Set 1 Number of Strata	**1**
Set 1 P1h (Probability Y1 ≥ Y2)	**0.602**
Set 2 Number of Strata	**1**
Set 2 P1h (Probability Y1 ≥ Y2)	**0.600**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For:	Sample Size
Number of Strata:	2
Alternative Hypothesis:	Two-Sided
Strata Allocation:	Equal (Search for N0)

	Sample Size						
Power	Total N	Group 1 N1	Group 2 N2	Stratum N0	Percent of Nh in Group 1 %N1h	Effect Size P(Y1 ≥ Y2) P1h*	Alpha
0.8023	258	130	128	129	50	Custom	0.05

* Custom P1h = 0.602 0.6

PASS also calculated the total required sample size to be 258. Thus, the procedure is validated.