

## Chapter 158

# Tests for the Matched-Pair Difference of Two Proportions in a Cluster-Randomized Design

## Introduction

Cluster-randomized designs are those in which whole clusters of subjects (classes, hospitals, communities, etc.) are sampled, rather than individual subjects. This sample size and power procedure is used for the case where the subject responses are binary (proportion outcome). To reduce the variation (and thus increase power), clusters are matched, with one cluster of each pair assigned to the control group, and the other assigned the treatment group. This procedure gives the number of pairs needed for the desired power requirement.

The formulas used here are based on Hayes and Bennett (1999) and Hayes and Moulton (2009). The methods are discussed in Donner and Klar (2000) and briefly in Campbell and Walters (2014).

## Technical Details

### Definition of Terms

The following table presents the various terms that are used.

<b>Group</b>	<b>1 (Control)</b>	<b>2 (Treatment)</b>
<b>Sample size:</b>	$K$	$K$
<b>Proportions:</b>	$P_1$	$P_2$

Coefficient of Variation:  $CV_M$  (within-pair coefficient of variation between clusters in the absence of intervention)

Number of cluster pairs:  $K$  ( $2K$  is the total number of clusters)

Cluster size:  $M$  (Average number of individuals per cluster)

## Hypotheses

The null and alternative hypotheses are

$$H_0: P_{Diff} = 0 \quad \text{vs.} \quad H_1: P_{Diff} \neq 0$$

Corresponding one-sided hypotheses may also be used. These hypotheses may be tested based on an appropriate paired difference test.

## Sample Size and Power Calculations

### Sample Size Calculation

The sample size calculation for the number of cluster pairs, as given in Hayes and Bennett (1999) and Hayes and Moulton (2009), is

$$K = 2 + (z_{\alpha/2} + z_{\beta})^2 \frac{P_1(1 - P_1)/M + P_2(1 - P_2)/M + CV_M^2(P_1^2 + P_2^2)}{(P_1 - P_2)^2}$$

### Estimating $CV_M$

Hayes and Bennett (1999) suggest that  $CV_M$  may be estimated based on the following, if prior data is available:

Let  $P_{ij}$  represent the true proportion in the  $j^{\text{th}}$  cluster ( $j = 1, 2$ ) of the  $i^{\text{th}}$  pair ( $i = 1, \dots, M$ ) and  $p_{ij}$  represent the corresponding observed proportion. If the empirical variance of the  $i^{\text{th}}$  pair is called  $s_i^2$  then define

$$s_m^2 = \sum s_i^2 / M$$

as the average of the within-pair variances.

$CV_M$  may be estimated from

$$CV_M^2 = \frac{s_m^2 - Av(p_i(1 - p_i)/n_{ij})}{Av(p_i^2)}$$

where  $n_{ij}$  is the number of individuals in the  $i^{\text{th}}$  pair of the  $j^{\text{th}}$  cluster, and  $Av()$  indicates the mean over all  $M$  clusters.

According to Hayes and Bennett (1999), "If only unmatched data are available, a conservative approach is to use [the coefficient of variation (SD/Mean) between clusters within each group] as an upper limit for  $CV_M$ ."

If no data is available, a series of plausible values, usually between 0 and 0.5, should be considered.

## Power Calculation

The corresponding power calculation to the sample size calculation above is

$$Power = 1 - \Phi \left( \sqrt{\frac{(K-2)(P_1 - P_2)^2}{P_1(1-P_1)/M + P_2(1-P_2)/M + CV_M^2(P_1^2 + P_2^2)}} - z_{\alpha/2} \right)$$

where  $z_x = \Phi(x)$  is the standard normal distribution function.

## Example 1 – Calculating Sample Size

One difficulty in calculating sample size for a matched-pair difference of two proportions in a cluster-randomized design is obtaining a value for  $CV_M$ , the within-pair coefficient of variation between clusters. This example shows how to enter a range of values to determine the effect of  $CV_M$  on sample size. Suppose that a cluster randomized study is to be conducted in which the control proportion is assumed to be  $P_1 = 0.70$ , and the treatment proportion is  $P_2 = 0.58$ . Each cluster is assumed to have about 70 individuals. A range of values between 0.05 and 0.5 for  $CV_M$  will be examined. The desired power and alpha are 0.9 and 0.05, respectively. The test will be 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 .....	<b>K (Number of Cluster Pairs)</b>
Alternative Hypothesis .....	<b>Two-Sided</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
M (Cluster Size) .....	<b>80</b>
P1 (Proportion for Group 1) .....	<b>0.80</b>
Enter P2, Diff, or Ratio for Group 2.....	<b>P2 (Proportion for Group 2)</b>
P2 (Proportion for Group 2) .....	<b>0.58</b>
CVM (Within-Pair Coefficient of Variation).....	<b>0.05 to 0.50 by 0.05</b>

## Tests for the Matched-Pair Difference of Two Proportions in a Cluster-Randomized Design

## Output

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

## Numeric Results

Solve For: K (Number of Cluster Pairs)  
 Groups: 1 = Control, 2 = Treatment  
 Alternative Hypothesis: Two-Sided

Power	Number of Cluster Pairs K	Total Number of Clusters 2K	Cluster Size M	Total Sample Size N	Proportion				Within-Pair Coefficient of Variation CVM	Alpha
					Group 1 P1	Group 2 P2	Difference P2 - P1	Ratio P2 / P1		
0.9491	4	8	80	640	0.8	0.58	-0.22	0.73	0.05	0.05
0.9511	6	12	80	960	0.8	0.58	-0.22	0.73	0.10	0.05
0.9064	8	16	80	1280	0.8	0.58	-0.22	0.73	0.15	0.05
0.9119	12	24	80	1920	0.8	0.58	-0.22	0.73	0.20	0.05
0.9123	17	34	80	2720	0.8	0.58	-0.22	0.73	0.25	0.05
0.9111	23	46	80	3680	0.8	0.58	-0.22	0.73	0.30	0.05
0.9094	30	60	80	4800	0.8	0.58	-0.22	0.73	0.35	0.05
0.9078	38	76	80	6080	0.8	0.58	-0.22	0.73	0.40	0.05
0.9062	47	94	80	7520	0.8	0.58	-0.22	0.73	0.45	0.05
0.9047	57	114	80	9120	0.8	0.58	-0.22	0.73	0.50	0.05

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.  
 K Represents the number of cluster pairs needed in the design.  
 2K The total number of clusters in the design, 2 clusters per pair.  
 M The average number of individuals or items in each cluster.  
 N The total number of individuals or items in the design.  $N = 2KM$ .  
 P1 The proportion for the control group. This is the baseline proportion.  
 P2 The proportion for the treatment group.  
 P2 - P1 The difference between the treatment proportion and the control proportion.  
 P2 / P1 The ratio of the treatment proportion to the control proportion.  
 CVM The within-pair coefficient of variation between clusters in the absence of intervention. See the documentation or references for suggestions for estimating CVM.  
 Alpha The probability of rejecting a true null hypothesis, that is, rejecting  $H_0$  when the proportions are actually equal.

## Summary Statements

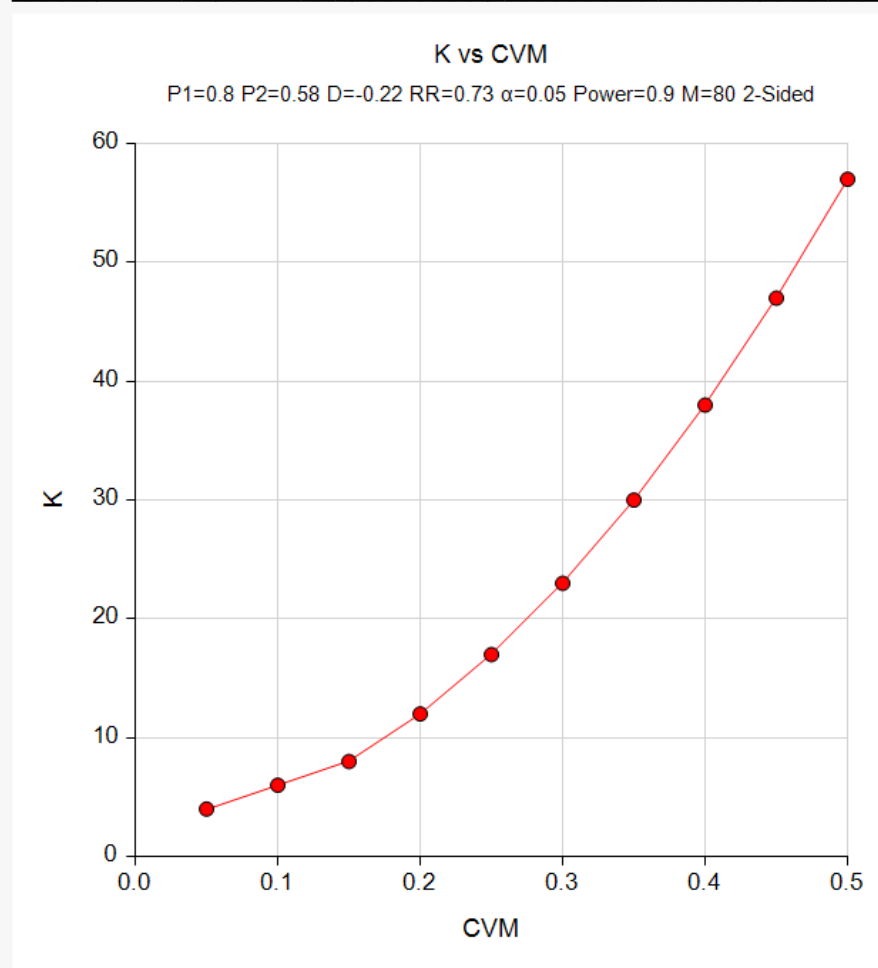
A matched-cluster-pair cluster-randomized design will be used to test whether there is a difference in proportions ( $P_2 - P_1 = P_{\text{treatment}} - P_{\text{control}}$ ). The comparison will be made using a two-sided, paired-difference Z-test, with a Type I error rate ( $\alpha$ ) of 0.05. The within-pair coefficient of variation between clusters in the absence of intervention is assumed to be 0.05. To detect a proportion difference of -0.22 ( $P_2 = 0.58$ ,  $P_1 = 0.8$ ) with 80 subjects per cluster and 90% power, the number of needed cluster pairs is 4, or 8 total clusters (640 total subjects).

## References

Hayes, R.J. and Bennett, S. 1999. 'Simple sample size calculation for cluster-randomized trials'. International Journal of Epidemiology. Vol 28, pages 319-326.  
 Hayes, R.J. and Moulton, L.H. 2009. Cluster Randomised Trials. CRC Press. New York.  
 Campbell, M.J. and Walters, S.J. 2014. How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research. Wiley. New York.

Tests for the Matched-Pair Difference of Two Proportions in a Cluster-Randomized Design

Plots



This report shows the needed number of cluster pairs for each of the coefficient of variation values.

## Example 2 – Validation using Hayes and Moulton (2009)

Hayes and Bennett (1999) and Hayes and Moulton (2009) give an example of calculating the required number of cluster pairs for a two-sided matched-pair difference of proportions in a cluster-randomized design. The group 1 proportion was 0.02 and the group 2 proportion was 0.01. The cluster size was 1,000 adults per cluster. A  $CV_M$  value of 0.25 was used. The power and alpha were 80% and 0.05, respectively. The calculated number of cluster pairs needed is 6.8 (or 7 with proper rounding).

### 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 ..... **K (Number of Cluster Pairs)**  
 Alternative Hypothesis ..... **Two-Sided**  
 Power..... **0.80**  
 Alpha..... **0.05**  
 M (Cluster Size) ..... **1000**  
 P1 (Proportion for Group 1) ..... **0.02**  
 Enter P2, Diff, or Ratio for Group 2..... **P2 (Proportion for Group 2)**  
 P2 (Proportion for Group 2) ..... **0.01**  
 CVM (Within-Pair Coefficient of Variation)..... **0.25**

### Output

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

**Numeric Results**

---

Solve For: **K (Number of Cluster Pairs)**  
 Groups: 1 = Control, 2 = Treatment  
 Alternative Hypothesis: Two-Sided

---

Power	Number of Cluster Pairs K	Total Number of Clusters 2K	Cluster Size M	Total Sample Size N	Proportion				Within-Pair Coefficient of Variation CVM	Alpha
					Group 1 P1	Group 2 P2	Difference P2 - P1	Ratio P2 / P1		
0.8183	7	14	1000	14000	0.02	0.01	-0.01	0.5	0.25	0.05

**PASS** calculates the number of cluster pairs needed to be 7, which matches the rounded-up calculation value of Hayes and Moulton (2009).