

## Chapter 244

# Tests for the Difference Between Two Poisson Rates in a Cluster-Randomized Design with Adjustment for Varying Cluster Sizes

---

## Introduction

This procedure calculates power and sample size for tests of two means in a cluster-randomized design in which the outcome variable is a count. It uses the work of Wang, Zhang, and Ahn (2018) which give the power for the case of varying cluster sizes. The analysis is of a simple z-test comparing the two means.

---

## Technical Details

The following discussion summarizes the results in Wang, Zhang, and Ahn (2018).

Suppose you are interested in comparing the means of two groups (treatment and control). Further suppose that the response is known to be related to other covariates (such as age, race, or gender) and so their impact needs to be adjusted for. This may be accomplished by stratifying on the covariates and forming hypotheses about a common mean difference across all clusters and strata. Often, the stratification is based on cluster size, but this is not required.

Let  $Y_{jki}$  be the count outcome of the  $i$ th ( $i = 1, \dots, n_{jk}$ ) subject in the  $k$ th ( $k = 1, \dots, K_j$ ) cluster of the  $j$ th ( $j = 1, 2$ ) group. Assuming that  $Y_{jki}$  follows a Poisson model with

$$\lambda_j = E(Y_{jki}) = \text{Var}(Y_{jki})$$

and a common intracluster correlation coefficient (ICC)  $\rho = \text{corr}(Y_{jki}, Y_{jki'})$  for  $i \neq i'$ .

### Test Statistic

An unbiased estimator of  $\lambda_j$  is

$$\hat{\lambda}_j = \frac{\sum_{k=1}^{K_j} \sum_{i=1}^{n_{jk}} Y_{jki}}{\sum_{k=1}^{K_j} n_{jk}}$$

## Tests for the Difference Between Two Poisson Rates in a C-R Design with Adjustment for Varying Cluster Sizes

with estimated variance

$$s_j^2 = \frac{\hat{\lambda}_j \sum_{k=1}^{K_j} n_{jk} [1 + (n_{jk} - 1)\hat{\rho}]}{\left(\sum_{k=1}^{K_j} n_{jk}\right)^2}$$

Using a two-sided hypothesis  $H_0: \lambda_1 = \lambda_2$  versus  $H_1: \lambda_1 \neq \lambda_2$ ,  $H_0$  is rejected if

$$\frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\sqrt{s_1^2 + s_2^2}} > z_{1-\alpha/2}$$

## Sample Size and Power

Wang, Zhang, and Ahn (2018) provide the following formula for estimating  $K_2$  (the number of clusters in the control group) when there are an unequal number of clusters in each group.

$$K_2 = \frac{\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2 \left(\frac{\lambda_1}{r} + \lambda_2\right)}{(\lambda_1 - \lambda_2)^2} \left[ (1 - \rho) \frac{1}{\theta} + \rho + \rho\gamma^2 \right]$$

where  $r = K_1/K_2$ ,  $\theta = E(n_{jk})$ ,  $\tau^2 = \text{Var}(n_{jk})$ , and  $\gamma = \tau/\theta$ . Hence  $\theta$  is the average cluster size of all clusters in the study and  $\gamma$  is the coefficient of variation of the cluster sizes. Note that  $z_x = \Phi(x)$  is the standard normal distribution function.

This equation can easily be rearranged to provide a formula for power  $(1 - \beta)$ .

## Example 1 – Finding the Number of Clusters

A study is being planned to investigate whether a new intervention will decrease a certain response variable. This response is a count. For a number of reasons, the researchers decide to administer the intervention to whole clusters (clinics) rather than randomize the treatment to individuals within the cluster. The number of clinics receiving each treatment will be balanced.

The average number of subjects per clinic is 21. The coefficient of variation of the cluster sizes is 0.42.

Prior studies have obtained a base-line mean score of 8.4 and an ICC of 0.31. The researchers want to compare the necessary numbers of clusters when the change in the mean score is -1, -2, and -3.

The two-sided significance level is set to 0.05 and the power is set to 0.9.

### 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>K1 (Number of Clusters in Group 1)</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\delta \neq 0</math>)</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
K2 (Clusters in Group 2) .....	<b>K1</b>
M (Average of Cluster Sizes) .....	<b>21</b>
CV (Coeff. of Variation of Cluster Sizes).....	<b>0.42</b>
Enter $\lambda_1$ or $\delta$ .....	<b><math>\delta</math> (Difference = <math>\lambda_1 - \lambda_2</math>)</b>
$\delta$ (Difference = $\lambda_1 - \lambda_2$ ) .....	<b>-1 -2 -3</b>
$\lambda_2$ (Mean of Group 2).....	<b>8.4</b>
ICC (Intracluster Correlation) .....	<b>0.31</b>

## Output

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

## Numeric Reports

### Numeric Results

Solve For: K1 (Number of Clusters in Group 1)  
 Groups: 1 = Treatment, 2 = Control  
 Hypotheses: H0:  $\delta = 0$  vs. H1:  $\delta \neq 0$

Power	Sample Size N	Number of Clusters			Allocation Ratio K1/K2	Cluster Size		Mean Event Rate per Subject			Intracluster Correlation Coefficient	
		Group 1 K1	Group 2 K2	Total K		Average M	CV	Group 1 $\lambda_1$	Group 2 $\lambda_2$	Difference $\delta$	ICC	Alpha
0.9235	294	7	7	14	1	21	0.42	5.4	8.4	-3	0.31	0.05
0.9096	672	16	16	32	1	21	0.42	6.4	8.4	-2	0.31	0.05
0.9000	2772	66	66	132	1	21	0.42	7.4	8.4	-1	0.31	0.05

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The total sample size, i.e., the total number of subjects from all clusters.
K1	The number of clusters assigned to group 1, the treatment group.
K2	The number of clusters assigned to group 2, the control group.
K	The total cluster count. It is the total number of clusters.
K1/K2	The allocation ratio. This is the number of group 1 clusters divided by the number of group 2 clusters.
M	The average cluster size. It is the average number of subjects per cluster based on all cluster sizes.
CV	The coefficient of variation of the cluster sizes. It is the standard deviation of cluster sizes divided by the average cluster size.
$\lambda_1$	The mean number of events per subject in group 1, the treatment group.
$\lambda_2$	The mean number of events per subject in group 2, the control group.
$\delta$	The mean difference assumed by the alternative hypothesis. This is the difference at which the power is computed. $\delta = \lambda_1 - \lambda_2$ .
ICC	The intracluster correlation coefficient. This is the correlation between any two subjects within a particular cluster.
Alpha	The probability of rejecting a true null hypothesis.

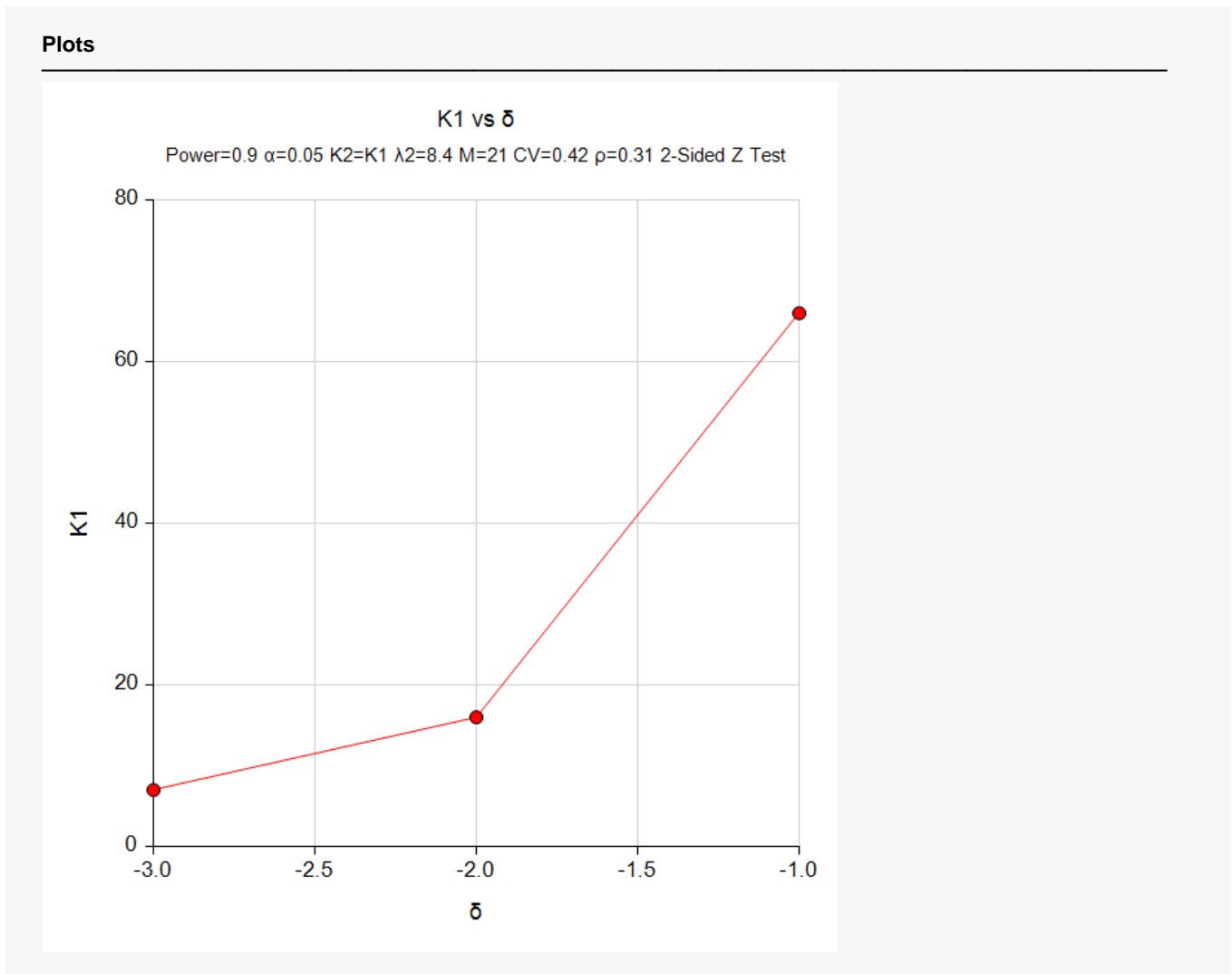
### Summary Statements

A parallel two-group cluster-randomized design will be used to test whether the Group 1 (treatment) event rate ( $\lambda_1$ ) is different from the Group 2 (control) event rate ( $\lambda_2$ ) (H0:  $\delta = 0$  versus H1:  $\delta \neq 0$ ,  $\delta = \lambda_1 - \lambda_2$ ). The comparison will be made using a two-sided rate-difference Z-test with a Type I error rate ( $\alpha$ ) of 0.05. The intracluster correlation coefficient is assumed to be 0.31. The control group event rate (mean) is assumed to be 8.4. The individual cluster sizes (the number of subjects per cluster) are assumed to vary according to a discrete distribution with mean 21 and coefficient of variation 0.42. To detect an event rate difference ( $\lambda_1 - \lambda_2$ ) of -3 (or treatment group event rate of 5.4) with 90% power, with an average of 21 subjects per cluster, the number of clusters needed will be 7 in Group 1 (treatment) and 7 in Group 2 (control) (for an overall total of 294 subjects).

### References

Wang, J., Zhang, S., and Ahn, C. 2018. 'Sample size calculation for count outcomes in cluster randomization trials with varying cluster sizes.' Communications in Statistics--Theory and Methods, DOI: 10.1080/03610926.2018.1532004.

## Plots Section



The values from the Numerical Results report are displayed in this plot.

## Example 2 – Validation using Wang, Zhang, and Ahn (2018)

Wang, Zhang, and Ahn (2018) provide an example of sizing a cluster-randomized study. We will use their results to validate this procedure.

The average number of subjects per cluster was expected to be 50. Three coefficients of variation were investigated: 0, 0.12111, and 0.29439. These corresponded to discrete uniform distributions of DU(50,50), DU(40,60), and DU(25,75). Note that DU(50,50) is a case in which all cluster sizes were equal to 50 and the coefficient of variation was 0.

The treatment mean was set at 4.35 and the control mean was set at 3.63. The ICC was set at 0.32. A two-sided test at a significance level of 0.05 is planned. The necessary number of clusters per group was calculated as 54, 55, and 59.

### 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 .....	<b>K1 (Number of Clusters in Group 1)</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\delta \neq 0</math>)</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
K2 (Clusters in Group 2) .....	<b>K1</b>
M (Average of Cluster Sizes) .....	<b>50</b>
CV (Coeff. of Variation of Cluster Sizes).....	<b>0 0.12111 0.29439</b>
Enter $\lambda_1$ or $\delta$ .....	<b><math>\lambda_1</math> (Mean Count of Group 1)</b>
$\lambda_1$ (Mean of Group 1).....	<b>4.35</b>
$\lambda_2$ (Mean of Group 2).....	<b>3.63</b>
ICC (Intracluster Correlation).....	<b>0.32</b>

## Tests for the Difference Between Two Poisson Rates in a C-R Design with Adjustment for Varying Cluster Sizes

## Output

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

### Numeric Results

Solve For:  $K_1$  (Number of Clusters in Group 1)

Groups: 1 = Treatment, 2 = Control

Hypotheses:  $H_0: \delta = 0$  vs.  $H_1: \delta \neq 0$

Power	Sample Size N	Number of Clusters			Allocation Ratio $K_1/K_2$	Cluster Size		Mean Event Rate per Subject			Intracluster Correlation Coefficient ICC	Alpha
		Group 1 $K_1$	Group 2 $K_2$	Total K		Average M	CV	Group 1 $\lambda_1$	Group 2 $\lambda_2$	Difference $\delta$		
0.9002	5400	54	54	108	1	50	0.000	4.35	3.63	0.72	0.32	0.05
0.9015	5500	55	55	110	1	50	0.121	4.35	3.63	0.72	0.32	0.05
0.9027	5900	59	59	118	1	50	0.294	4.35	3.63	0.72	0.32	0.05

PASS has obtained  $K_1$  values of 54, 55, and 59.