Chapter 536

GEE Tests for Multiple Poisson Rates in a Cluster-Randomized Design

Introduction

This module calculates the power and sample size for testing for differences among the group means of count data from a cluster-randomized design that are analyzed using a GEE model.

GEE is different from mixed models in that it does not require the full specification of the joint distribution of the measurements, as long as the marginal mean model is correctly specified. Estimation consistency is achieved even if the correlation matrix is incorrect. For clustered designs such as those discussed here, GEE assumes a *compound symmetric* (CS) correlation structure.

The outcomes are averaged at the cluster level. The precision of the experiment is increased substantially by increasing the number of clusters and slightly by increasing the number of subjects per cluster.

Missing Values

This procedure allows you to specify the proportion of subjects that are missing at the end of the study.

Technical Details

Theory and Notation

Technical details are given in Ahn, Heo, and Zhang (2015), chapter 4, section 4.8.3, pages 132-134, and in Zhang and Ahn (2013).

Suppose we want to compare the means of G groups. Further suppose we have K_g (g = 1, ..., G) clusters in each of the groups, each with an average of M subjects. Let Y_{gki} be the count response of subject i of cluster k in group g. The response is modeled by Poisson distribution

$$f(y_{gki}) = \frac{e^{-\mu_{gki}}\mu_{gki}^{y_{gki}}}{y_{gki}!}$$

The mean of Y_{gki} is modeled by

$$\log(\mu_{gki}) = \beta_g$$

while y_{gki} is modelled as

$$y_{gki} = \mu_{gki} + \epsilon_{gki}$$

where

y_{gki} is the response from subject i in cluster k in group g,

μ_{gki} is group-specific treatment effect (g = 1, …, G),

ϵ_{gki} is a zero-mean error term with variance also given by μ_{gki}.

In this procedure, the primary interest is to test that a specific contrast based on the coefficients $C = c_1, …, c_G$ is zero, that is, that $H_0: \sum_{g=1}^{G} \beta_g c_g = 0$ against the alternative that it is non-zero.

GEE is used to estimate the β_g's and test this hypothesis. The test statistic is

$$T = \frac{C'b}{\sqrt{\frac{1}{K}(C'WA^{-1}VA^{-1}WC)}}$$

where

b_g is the GEE estimate of $\beta_g, g = 1, …, G$

r_g is the proportion of clusters in group g,

C is the vector of contrast coefficients

K is the total number of clusters

W is a diagonal matrix of elements $1/\sqrt{r_g}$

A is a diagonal matrix of elements $M/K_g \sum_{i=1}^{K_g} \exp(b_g)$

V is a diagonal matrix of elements $1/K_g \sum_{g=1}^{K_g} \left[\sum_{i=1}^{M} \hat{\epsilon}_{gki}\right]^2$, where $\hat{\epsilon}_{gki} = y_{gki} - \mu_{gki}(\underline{b})$

H_0 is rejected with a type I error α if $|Z| > z_{1-\alpha/2}$ where $z_{1-\alpha/2}$ is the 100(1 – $\alpha/2$)th percentile of a standard normal distribution.

Correlation Patterns

In a cluster-randomized design with K clusters, each consisting of M subjects, observations from a single cluster are correlated. The resulting correlation matrix is assumed to have a *compound symmetric* pattern with a common correlation coefficient ρ. That is, the correlation matrix within a cluster is

$$[\rho_{jj'}] = \begin{bmatrix} 1 & \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \rho & \cdots & 1 \end{bmatrix}_{M \times M}$$

Missing Data

The problem of missing data occurs for several reasons. In these designs, it is assumed that the responses of some proportion, *P*, of the subjects will be missing.

Sample Size Calculations

The details of the calculation of sample size and power is given in Wang, Zhang, and Ahn (2018). The formula for the cluster count (K) is

$$
K = \frac{C'WA^{-1}VA^{-1}WC\left(z_{1-\frac{\alpha}{2}} + z_{1-\gamma}\right)^2}{\left(C'\underline{\beta}\right)^2}
$$

where

γ	equals 1 – power.
α	is the significance level.
$z_{1-\alpha/2}$	is the 100(1 – *α/2*)th percentile of a standard normal distribution.
$\underline{\beta}$	is a vector of anticipated GEE regression coefficients, $\beta_1, \dots, \beta_G$.
C	is the vector of contrast coefficients.
V	is a diagonal matrix of elements $\{h \exp(\beta_g)\}$.
h	is $\sum_{j=1}^{M}\sum_{j'=1}^{M}\phi_{jj'}\,\rho_{jj'}$.
$\phi_{jj'}$	is the probability that both subjects j and j' are observed. By definition, this is set to one minus the proportion missing.
$\rho_{jj'}$	is the intracluster correlation coefficient between any two subjects in the same cluster.
r_g	is the proportion of subjects in group g.

The above formula is easily rearranged to obtain a formula for power.

Example 1 – Determining the Power

Researchers are planning a study comparing medications: a standard drug and two experimental drugs. The experimental drugs appear to have about the same impact. All patients within a cluster will receive the same drug. The clusters available for study will be randomly assigned to one of the three groups.

The researchers want a sample size large enough to detect a time-averaged response difference of 5 between the highest and lowest time-averaged heart rates. They will use the means of 65, 60, and 60 to represent this difference. The contrast coefficients that they will use are -2, 1, 1.

To begin, the researchers want to determine the power for Ki between 10 and 50. They will assume an average cluster size of 10.

Similar studies had an autocorrelation between subjects within a cluster of between 0.6 and 0.8, so they want to try values in that range. The test will be conducted at the 0.05 significance level. The subjects will be divided equally among the three groups.

At this stage of planning, the researchers want to ignore the possibility that some subjects will drop out.

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 ....**Power**
Alpha..**0.05**
G (Number of Groups) **3**
Group Allocation Input Type ....**Equal (K1 = K2 = ··· = KG)**
Ki (Clusters Per Group)..................................**10 20 30 40 50**
M (Average Cluster Size)................................**10**
μi's Input Type ....**μ1, μ2, ..., μG**
μ1, μ2, ..., μG...**65 60 60**
Contrast Input Type ....**List of Contrast Coefficients**
Contrast Coefficients.....................................**-2 1 1**
ρ (Intracluster Correlation, ICC)....................**0.6 0.7 0.8**
Missing Input Type..**Constant = 0**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Number of Groups: 3

Power	Number of Subjects N	Number of Clusters K	Clusters Per Group Ki	Average Cluster Size M	Group Means μ_i	Contrast Coefficients Ci	Linear Combination of Ci and μ_i $|C_i'\mu_i|$	ICC ρ	Alpha	Missing Proportion
0.5376	300	30	10	10	$\mu_i(1)$	C(1)	10	0.6	0.05	0
0.4855	300	30	10	10	$\mu_i(1)$	C(1)	10	0.7	0.05	0
0.4424	300	30	10	10	$\mu_i(1)$	C(1)	10	0.8	0.05	0
0.8278	600	60	20	10	$\mu_i(1)$	C(1)	10	0.6	0.05	0
0.7765	600	60	20	10	$\mu_i(1)$	C(1)	10	0.7	0.05	0
0.7280	600	60	20	10	$\mu_i(1)$	C(1)	10	0.8	0.05	0
0.9450	900	90	30	10	$\mu_i(1)$	C(1)	10	0.6	0.05	0
0.9149	900	90	30	10	$\mu_i(1)$	C(1)	10	0.7	0.05	0
0.8817	900	90	30	10	$\mu_i(1)$	C(1)	10	0.8	0.05	0
0.9842	1200	120	40	10	$\mu_i(1)$	C(1)	10	0.6	0.05	0
0.9704	1200	120	40	10	$\mu_i(1)$	C(1)	10	0.7	0.05	0
0.9525	1200	120	40	10	$\mu_i(1)$	C(1)	10	0.8	0.05	0
0.9958	1500	150	50	10	$\mu_i(1)$	C(1)	10	0.6	0.05	0
0.9904	1500	150	50	10	$\mu_i(1)$	C(1)	10	0.7	0.05	0
0.9821	1500	150	50	10	$\mu_i(1)$	C(1)	10	0.8	0.05	0

Item	Values
$\mu_i(1)$	65, 60, 60
C(1)	-2, 1, 1

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The total number of subjects in the study.
K	The total number of clusters in the study.
Ki	The number of clusters per group.
M	The average cluster size (number of subjects per cluster) used for all clusters.
μ_i	Group Means. Gives the name and number of the set containing the means for each group.
Ci	Contrast Coefficients. Gives the name of the set containing the contrast coefficients that are combined with the group log means.
$|C_i'\mu_i|$	Linear Combination of Ci and μ_i. Gives the linear combination of the contrast coefficients and the group means.
ρ	The intracluster correlation coefficient (ICC) used for all clusters. This is the correlation between any two subjects within a cluster.
Alpha	The probability of rejecting a true null hypothesis.
Missing Proportion	The proportion of each cluster that is assumed to be missing at the end of the study.

Summary Statements

A parallel, 3-group cluster-randomized design will be used to test the difference among the 3 Poisson rates (means) that is defined by the contrast coefficients -2, 1, 1. The comparison will be made using a generalized estimating equation (GEE) Poisson model Z-test with a Type I error rate (α) of 0.05. The autocorrelation matrix of the responses within a cluster is assumed to be compound symmetric with an intraclass correlation coefficient (ICC) of 0.6. Missing values are assumed to occur completely at random (MCAR), and the anticipated proportion missing is 0. To detect the group Poisson rates (means) 65, 60, 60, with contrast coefficients -2, 1, 1, with a total of 30 clusters (allocated to the 3 groups as 10, 10, 10), with an average cluster size of 10 subjects per cluster (for a total sample size of 300 subjects), the power is 0.5376.

References
_____

Ahn, C., Heo, M., and Zhang, S. 2015. Sample Size Calculations for Clustered and Longitudinal Outcomes in
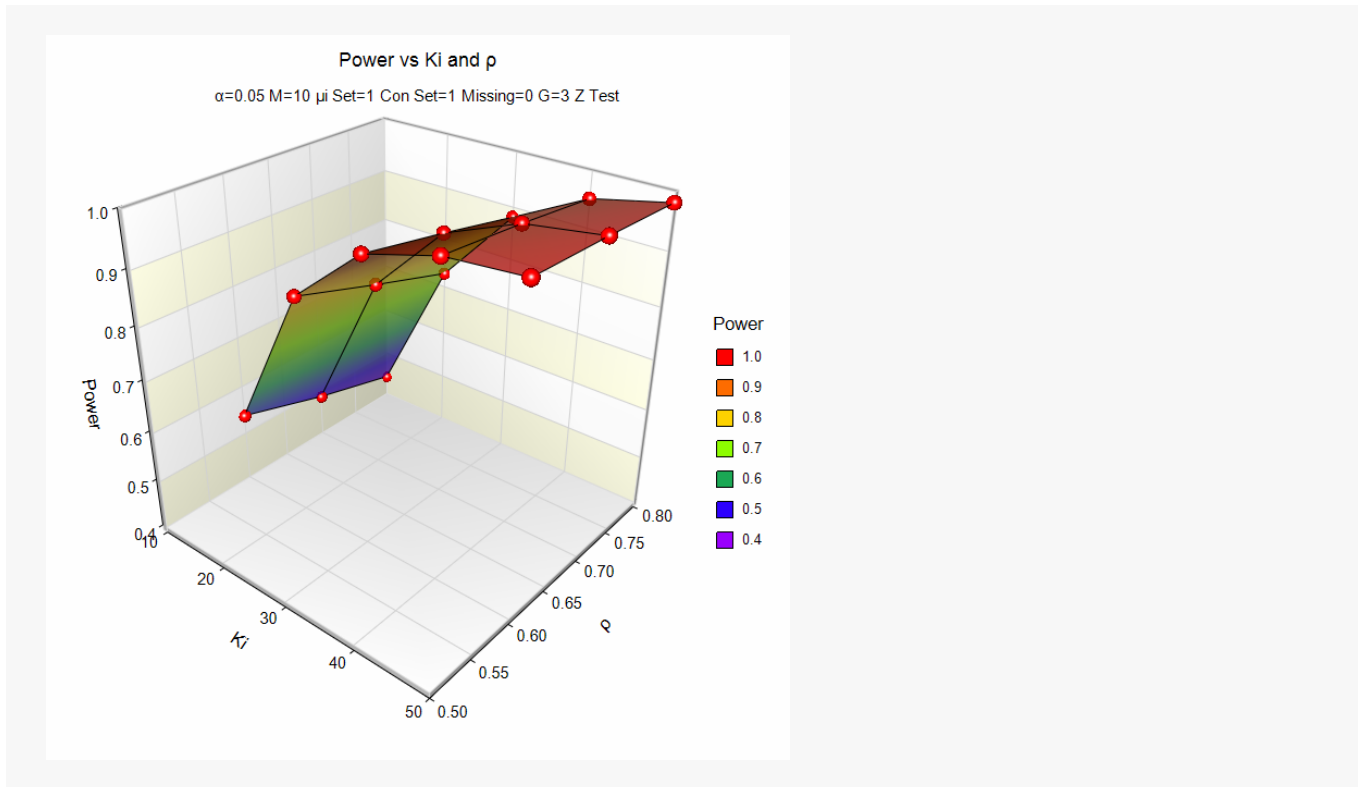   Clinical Research. CRC Press. New York.
_____

This report gives the power for various values of the other parameters. The definitions of each of the columns in the report are shown in the Report Definitions section. Note that the details of the μ_i and the C_i sets are shown below the table.

Plots Section

Plots
_____
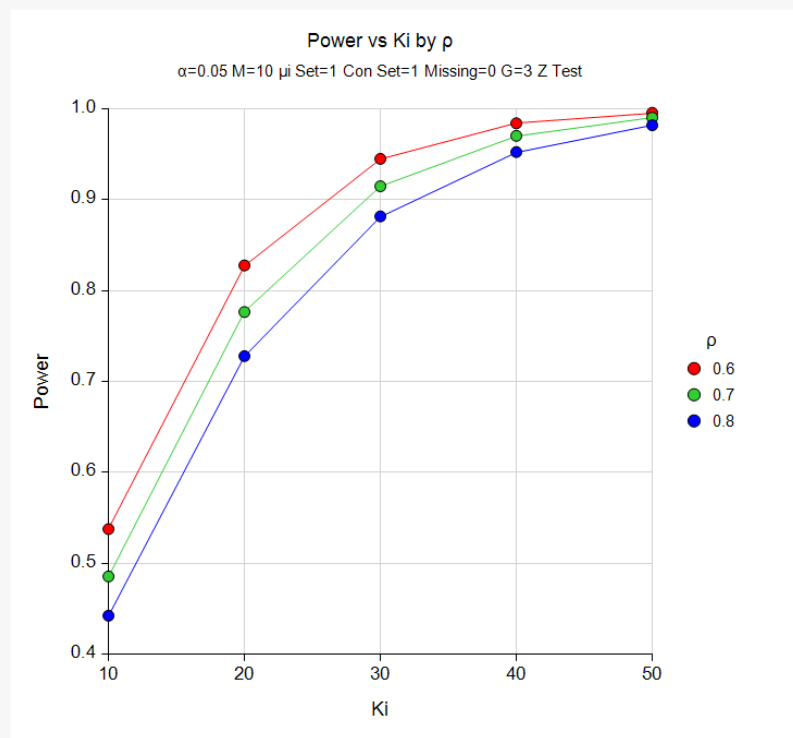
These plots show the relationship among the design parameters.

Example 2 – Finding the Sample Size

Continuing with Example 1, we want to give an example that uses the spreadsheet. This allows us to compare sample size requirements for various cluster allocation patterns. The power is set to 0.9.

This example will use all the settings of Example 1, except that three cluster allocation patterns will be compared. The following cluster allocation patterns are entered on the spreadsheet.

C1	**C2**	**C3**
2	1	1
2	1	2
2	4	3

Also note that each column sums to 6.

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 Clusters)**
Power...**0.90**
Alpha...**0.05**
G (Number of Groups) **3**
Group Allocation Input Type ....**Enter columns of allocation patterns**
Columns of Group Allocation Patterns...........**C1-C3**
M (Average Cluster Size)..............................**10**
μ_i's Input Type ....**μ_1, μ_2, ..., μ_G**
μ_1, μ_2, ..., μ_G..**65 60 60**
Contrast Input Type ....**List of Contrast Coefficients**
Contrast Coefficients....................................**-2 1 1**
ρ (Intracluster Correlation, ICC)....................**0.6 0.7 0.8**
Missing Input Type.......................................**Constant = 0**

Input Spreadsheet Data

Row	C1	C2	C3
1	2	1	1
2	2	1	2
3	2	4	3

Output

Numeric Reports

Numeric Results

Solve For: K (Number of Clusters)
Number of Groups: 3

Power	Number of Subjects N	Number of Clusters K	Group Cluster Allocation ri	Average Cluster Size M	Group Means μi	Contrast Coefficients Ci	Linear Combination of Ci and μi |Ci'μi|	ICC ρ	Alpha	Missing Proportion
0.9012	750	75	C1(1)	10	μi(1)	C(1)	10	0.6	0.05	0
0.9059	870	87	C1(1)	10	μi(1)	C(1)	10	0.7	0.05	0
0.9009	960	96	C1(1)	10	μi(1)	C(1)	10	0.8	0.05	0
0.9050	1320	132	C2(2)	10	μi(1)	C(1)	10	0.6	0.05	0
0.9039	1500	150	C2(2)	10	μi(1)	C(1)	10	0.7	0.05	0
0.9031	1680	168	C2(2)	10	μi(1)	C(1)	10	0.8	0.05	0
0.9029	1200	120	C3(3)	10	μi(1)	C(1)	10	0.6	0.05	0
0.9052	1380	138	C3(3)	10	μi(1)	C(1)	10	0.7	0.05	0
0.9070	1560	156	C3(3)	10	μi(1)	C(1)	10	0.8	0.05	0

Item	Values
C1(1)	0.333, 0.333, 0.333
C2(2)	0.167, 0.167, 0.667
C3(3)	0.167, 0.333, 0.5
μi(1)	65, 60, 60
C(1)	-2, 1, 1

This report gives the number of clusters for various values of the other parameters. Note that the details of the cluster allocation columns and the proportions are shown in the Set footnote below the numeric results.

Plots Section

Plots
_____
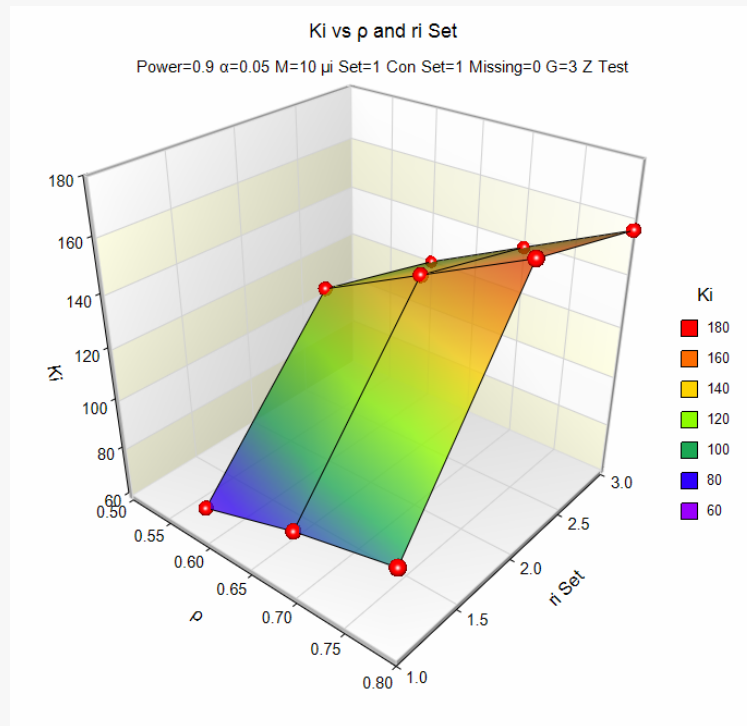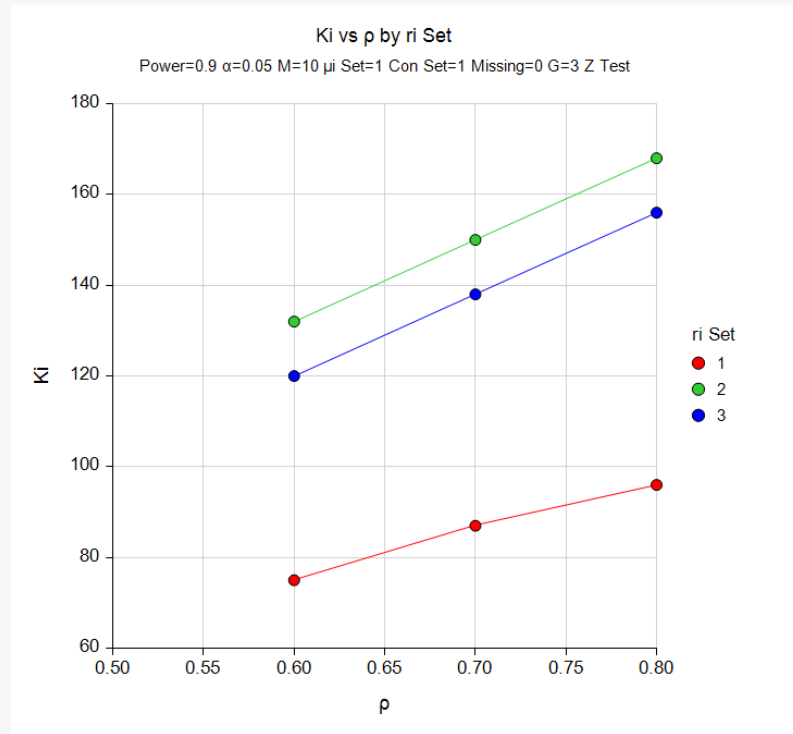
The plot shows the sample size requirements for the three cluster allocation patterns that were used. Note that the equal allocation pattern, C1, requires the smallest number of clusters.

Example 3 – Validation of Sample Size Calculation

We could not find a validation example in the literature. However, we will use results from a previously validated procedure to validate this procedure. That procedure is *GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)* which is procedure 472.

In the other procedure, if we set power = 0.8, alpha = 0.05, G = 4, Group Allocation Input Type = Equal (n1 = n2 = ... = nG = n), Measurement Time Input Type = Equally Spaced measurement Times, M = 6, P1...PG = 0.5 0.62245933, Contrast Coefficients = -3 1 1 1, compound symmetry correlation pattern, ρ = 0.3, and no missing data, the required sample size (N) is calculated as 44.

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 3** 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 Clusters)**
Power..**0.8**
Alpha..**0.05**
G (Number of Groups) **4**
Group Allocation Input Type ....**Equal (K1 = K2 = ··· = KG)**
M (Average Cluster Size).................................**6**
μi's Input Type ....**μ1, μ2, ..., μG**
μ1, μ2, ..., μG...**65 60 60 60**
Contrast Input Type ....**List of Contrast Coefficients**
Contrast Coefficients.....................................**-3 1 1 1**
ρ (Intracluster Correlation, ICC)....................**0.3**
Missing Input Type..**Constant = 0**

Output

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

Numeric Results
───
Solve For: K (Number of Clusters)
Number of Groups: 4
───

Power	Number of Subjects N	Number of Clusters K	Group Cluster Allocation ri	Average Cluster Size M	Group Means μi	Contrast Coefficients Ci	Linear Combination of Ci and μi |Ci'μi|	ICC ρ	Alpha	Missing Proportion
0.8111	264	44	ri(1)	6	μi(1)	C(1)	15	0.3	0.05	0

─────────────────────────────────
Item	Values
ri(1)	0.25, 0.25, 0.25, 0.25
μi(1)	65, 60, 60, 60
C(1)	-3, 1, 1, 1
─────────────────────────────────

The number of clusters, K, of 44 matches the value found in the other procedure. Thus, the procedure is validated.