

“What is the annotated output for Principal Components Analysis in SPSS?”

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). “*What is the annotated output for Principal Components Analysis in SPSS?*”. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160720>

The annotated output for Principal Components Analysis (PCA) in SPSS provides a comprehensive summary of the results from the analysis. It includes a detailed report of the components extracted, the amount of variance explained by each component, and the factor loadings for each variable. Additionally, the output also presents graphical representations such as a scree plot and a component plot, which aid in interpreting the results. The annotated output also includes statistical values, such as eigenvalues and communalities, that assist in determining the significance and contribution of each component. Overall, the annotated output for PCA in SPSS serves as a useful tool for understanding and interpreting the complex results of the analysis.

Principal Components Analysis | SPSS Annotated Output

This page shows an example of a principal components analysis with footnotes explaining the output. The data used in this example were collected by Professor James Sidanius, who has generously shared them with us. You can download the data set here: [m255.sav](#).

Overview: The "what" and "why" of principal components analysis

Principal components analysis is a method of data reduction. Suppose that you have a dozen variables that are correlated. You might use principal components analysis to reduce your 12

measures to a few principal components. In this example, you may be most interested in obtaining the component scores (which are variables that are added to your data set) and/or to look at the dimensionality of the data. For example, if two components are extracted and those two components accounted for 68% of the total variance, then we would say that two dimensions in the component space account for 68% of the variance. Unlike factor analysis, principal components analysis is not usually used to identify underlying latent variables. Hence, the loadings onto the components are not interpreted as factors in a factor analysis would be. Principal components analysis, like factor analysis, can be performed on raw data, as shown in this example, or on a correlation or a covariance matrix. If raw data are used, the procedure will create the original correlation matrix or covariance matrix, as specified by

the user. If the correlation matrix is used, the variables are standardized and the total variance will equal the number of variables used in the analysis (because each standardized variable has a variance equal to 1). If the covariance matrix is used, the variables will remain in their original metric. However, one must take care to use variables whose variances and scales are similar. Unlike factor analysis, which analyzes the common variance, the original matrix in a principal components analysis analyzes the total variance. Also, principal components analysis assumes that each original measure is collected without measurement error.

Principal components analysis is a technique that requires a large sample size. Principal components analysis is based on the correlation matrix of the variables involved, and correlations usually need a

large sample size before they stabilize. Tabachnick and Fidell (2001, page 588) cite Comrey and Lee's (1992) advise regarding sample size: 50 cases is very poor, 100 is poor, 200 is fair, 300 is good, 500 is very good, and 1000 or more is excellent.

As a rule of thumb, a bare minimum of 10 observations per variable is necessary to avoid computational difficulties.

In this example we have included many options, including the original and reproduced correlation matrix and the scree plot.

While you may not wish to use all of these options, we have included them here to aid in the explanation of the analysis. We have also created a page of annotated output for a factor analysis that parallels this analysis. For general information regarding the similarities and differences between principal components analysis and factor analysis, see Tabachnick and Fidell (2001), for example.

get file "D:dataM255.sav".

factor

**/variables item13 item14 item15 item16 item17 item18
item19 item20 item21 item22 item23 item24**

**/print initial correlation det kmo repr extraction
univariate**

/format blank(.30)

/plot eigen

/extraction pc

/method = correlate.

ARABPSYCHOLOGY.COM

Descriptive Statistics

	Mean ^a	Std. Deviation ^b	Analysis N ^c
INSTRUC WELL PREPARED	4.46	.729	1365
INSTRUC SCHOLARLY GRASP	4.53	.700	1365
INSTRUCTOR CONFIDENCE	4.45	.732	1365
INSTRUCTOR FOCUS LECTURES	4.28	.829	1365
INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	4.17	.895	1365
INSTRUCTOR SENSITIVE TO STUDENTS	3.93	1.035	1365
INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	4.08	.964	1365
INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	3.78	.909	1365
INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	3.77	.984	1365
I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	3.61	1.116	1365
COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	3.81	.957	1365
COMPARED TO OTHER COURSES THIS COURSE WAS	3.67	.926	1365

The table above is output because we used the univariate option on the /print subcommand. Please note that the only way to see how many cases were actually used in the principal components analysis is to include the univariate

option on the /print subcommand. The number of cases used in the analysis will be less than the total number of cases in the data file if there are missing values on any of the variables used in the principal components analysis, because, by default, SPSS does a listwise deletion of incomplete cases. If the principal components analysis is being conducted on the correlations (as opposed to the covariances), it is not much of a concern that the variables have very different means and/or standard deviations (which is often the case when variables are measured on different scales).

a. Mean - These are the means of the variables used in the factor analysis.

b. Std. Deviation - These are the standard deviations of the variables used in the factor analysis.

c. Analysis N - This is the number of cases used in the factor analysis.

Correlation Matrix^a

	item13 INSTRUC WELL PREPARED	item14 INSTRUC SCHOLARLY GRASP	item15 INSTRUC TOR CONFIDE NCE	item16 INSTRUC OR FOCUS LECTURES	item17 INSTRUC OR USES CLEAR RELEVANT EXAMPLES	item18 INSTRUCTOR SENSITIVE TO STUDENTS	item19 INSTRUC TOR ALLOWS ME TO ASK QUESTIONS	item20 INSTRUC TOR IS ACCESSIB LE TO STUDENTS OUTSIDE CLASS	item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTAN DING	item22 I AM SATISFIED WITH STUDENT PERFORMAN CE EVALUATION	item23 COMPARED TO OTHER INSTRUC TORS, THIS INSTRUC TOR IS	item24 COMPARED TO OTHER COURSES THIS COURSE WAS
Correlation	1.000	.661	.600	.566	.577	.409	.286	.304	.476	.333	.564	.454
item13 INSTRUC WELL PREPARED												
item14 INSTRUC SCHOLARLY GRASP	.661	1.000	.635	.500	.552	.433	.320	.315	.449	.333	.565	.443
item15 INSTRUC TOR CONFIDENCE	.600	.635	1.000	.505	.587	.457	.359	.356	.509	.369	.582	.435
item16 INSTRUCTOR FOCUS LECTURES	.566	.500	.505	1.000	.586	.405	.335	.317	.452	.363	.459	.430
item17 INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	.577	.552	.587	.586	1.000	.555	.449	.417	.595	.450	.613	.521
item18 INSTRUCTOR SENSITIVE TO STUDENTS	.409	.433	.457	.405	.555	1.000	.627	.521	.554	.536	.569	.474
item19 INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	.286	.320	.359	.335	.449	.627	1.000	.446	.499	.484	.444	.374
item20 INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	.304	.315	.356	.317	.417	.521	.446	1.000	.425	.383	.410	.357
item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	.476	.449	.509	.452	.595	.554	.499	.425	1.000	.507	.598	.500
item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	.333	.333	.369	.363	.450	.536	.484	.383	.507	1.000	.493	.444
item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	.564	.565	.582	.459	.613	.569	.444	.410	.598	.493	1.000	.705
item24 COMPARED TO OTHER COURSES THIS COURSE WAS	.454	.443	.435	.430	.521	.474	.374	.357	.500	.444	.705	1.000

^a Determinant = .002

The table above was included in the output because we included the keyword **correlation** on the **/print** subcommand. This table gives the correlations between the original variables (which are specified on the **/variables** subcommand). Before conducting a principal components analysis, you want to check the correlations between the variables. If any of the correlations are too high (say above .9), you may need to remove one of the variables from the analysis, as the two variables

seem to be measuring the same thing. Another alternative would be to combine the variables in some way (perhaps by taking the average). If the correlations are too low, say below .1, then one or more of the variables might load only onto one principal component (in other words, make its own principal component). This is not helpful, as the whole point of the analysis is to reduce the number of items (variables).

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy. ^a		.934
Bartlett's Test of Sphericity ^b	Approx. Chi-Square	8676.712
	df	66
	Sig.	.000

a. Kaiser-Meyer-Olkin Measure of Sampling Adequacy -

This measure varies between 0 and 1, and values closer to 1 are better. A value of .6 is a suggested minimum.

b. Bartlett's Test of Sphericity - This tests the null hypothesis that the correlation matrix is an identity matrix. An identity matrix is matrix in which all of the diagonal elements are 1 and all off diagonal elements are 0.

You want to reject this null hypothesis.

Taken together, these tests provide a minimum standard which should be passed before a principal components analysis (or a factor analysis) should be conducted.

Communalities^a

	Initial ^b	Extraction ^c
item13 INSTRUC WELL PREPARED	1.000	.731
item14 INSTRUC SCHOLARLY GRASP	1.000	.690
item15 INSTRUCTOR CONFIDENCE	1.000	.652
item16 INSTRUCTOR FOCUS LECTURES	1.000	.549
item17 INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	1.000	.661
item18 INSTRUCTOR SENSITIVE TO STUDENTS	1.000	.704
item19 INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	1.000	.658
item20 INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	1.000	.494
item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	1.000	.601
item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	1.000	.557
item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	1.000	.673
item24 COMPARED TO OTHER COURSES THIS COURSE WAS	1.000	.509

Extraction Method: Principal Component Analysis.

a. Communalities - This is the proportion of each variable's variance that can be explained by the principal components (e.g., the underlying latent continua). It is also noted as h^2 and can be defined as the sum

of squared factor loadings.

b. Initial - By definition, the initial value of the communality in a principal components analysis is 1.

c. Extraction - The values in this column indicate the proportion of each variable's variance that can be explained by the principal components.

Variables with high values are well represented in the common factor space, while variables with low values are not well represented. (In this example, we don't have any particularly low values.) They are the reproduced variances from the number of components that you have saved. You can find these values on the diagonal of the reproduced correlation matrix.

Total Variance Explained

Component ^a	Initial Eigenvalues ^b			Extraction Sums of Squared Loadings ^f		
	Total ^c	% of Variance ^d	Cumulative % ^e	Total	% of Variance	Cumulative %
1	6.249	52.076	52.076	6.249	52.076	52.076
2	1.229	10.246	62.322	1.229	10.246	62.322
3	.719	5.992	68.313			
4	.613	5.109	73.423			
5	.561	4.676	78.099			
6	.503	4.192	82.291			
7	.471	3.927	86.218			
8	.389	3.240	89.458			
9	.368	3.066	92.524			
10	.328	2.735	95.259			
11	.317	2.645	97.904			
12	.252	2.096	100.000			

Extraction Method: Principal Component Analysis.

a. Component - There are as many components extracted during a principal components analysis as there are variables that are put into it.

In our example, we used 12 variables (item13 through item24), so we have 12 components.

b. Initial Eigenvalues - Eigenvalues are the variances of the principal components. Because we conducted our principal components analysis on the correlation matrix, the variables are standardized, which means that the each

variable has a variance of 1, and the total variance is equal to the number of variables used in the analysis, in this case, 12.

c. **Total** - This column contains the eigenvalues. The first component will always account for the most variance (and hence have the highest eigenvalue), and the next component will account for as much of the left over variance as it can, and so on. Hence, each successive component will account for less and less variance.

d. **% of Variance** - This column contains the percent of variance accounted for by each principal component.

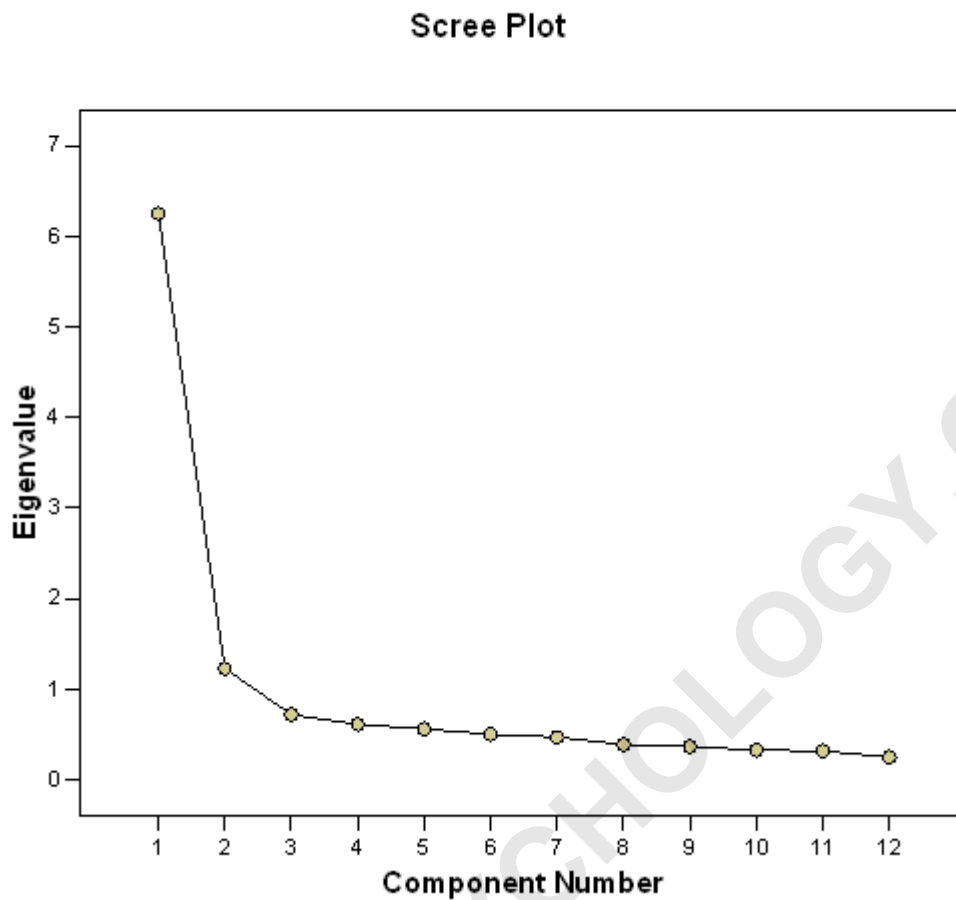
e. **Cumulative %** - This column contains the cumulative percentage of variance accounted for by the current and all preceding principal components.

For example, the third row shows a value of 68.313. This means that the first three components together account for 68.313% of

the total variance.

(Remember that because this is principal components analysis, all variance is considered to be true and common variance. In other words, the variables are assumed to be measured without error, so there is no error variance.)

f. Extraction Sums of Squared Loadings - The three columns of this half of the table exactly reproduce the values given on the same row on the left side of the table. The number of rows reproduced on the right side of the table is determined by the number of principal components whose eigenvalues are 1 or greater.



The scree plot graphs the eigenvalue against the component number. You can see these values in the first two columns of the table immediately above.

From the third component on, you can see that the line is almost flat, meaning the each successive component is accounting for smaller and smaller amounts of the total variance. In general, we are interested in

keeping only those principal components whose eigenvalues are greater than 1. Components with an eigenvalue of less than 1 account for less variance than did the original variable (which had a variance of 1), and so are of little use. Hence, you can see that the point of principal components analysis is to redistribute the variance in the correlation matrix (using the method of eigenvalue decomposition) to redistribute the variance to first components extracted.

Component Matrix^a

	Component ^c	
	1	2
item13 INSTRUC WELL PREPARED	.727	-.449
item14 INSTRUC SCHOLARLY GRASP	.724	-.408
item15 INSTRUCTOR CONFIDENCE	.746	-.308
item16 INSTRUCTOR FOCUS LECTURES	.685	
item17 INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	.806	
item18 INSTRUCTOR SENSITIVE TO STUDENTS	.755	.366
item19 INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	.641	.497
item20 INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	.593	.378
item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	.763	
item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	.651	.364
item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	.819	
item24 COMPARED TO OTHER COURSES THIS COURSE WAS	.714	

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

b. Component Matrix - This table contains component loadings, which are the correlations between the variable and the component. Because these are correlations, possible values range from -1 to +1. On the

/format

subcommand, we used the option blank(.30), which tells SPSS not to print any of the correlations that are .3 or less. This makes the output easier to read by removing the clutter of low correlations that are probably not meaningful anyway.

c. Component - The columns under this heading are the principal components that have been extracted. As you can see by the footnote provided by SPSS (a.), two components were extracted (the two components that had an eigenvalue greater than 1). You usually do not try to interpret the components the way that you would factors that have been extracted from a factor analysis. Rather, most people are interested in the component scores, which are used for data reduction (as opposed to factor analysis where you are looking for underlying latent continua). You can save the

component scores to your data set for use in other analyses using the /save subcommand.

Reproduced Correlations^c

	Item13 INSTRUC WELL PREPARED	Item14 INSTRUC SCHOLARLY GRASP	Item15 INSTRUC TOR CONFIDENCE	Item16 INSTRUC OR FOCUS LECTURES	Item17 INSTRUC OR USES CLEAR RELEVANT EXAMPLES	Item18 INSTRUC TOR SENSITIVE TO STUDENTS	Item19 INSTRUC TOR ALLOWS ME TO ASK QUESTIONS	Item20 INSTRUC OR IS ACCESSIB LE TO STUDENTS OUTSIDE CLASS	Item21 INSTRUC TOR AWARE OF STUDENTS UNDERSTAN DING	Item22 I AM SATISFIED WITH STUDENT PERFORMAN CE EVALUATION	Item23 COMPARED TO OTHER INSTRUC TORS, THIS INSTRUC TOR IS	Item24 COMPARED TO OTHER COURSES THIS COURSE WAS	
Reproduced Correlation ^d	Item13 INSTRUC WELL PREPARED	.731 ^b	.710	.681	.625	.634	.385	.243	.261	.495	.310	.614	.517
	Item14 INSTRUC SCHOLARLY GRASP	.710	.690 ^b	.666	.611	.627	.397	.261	.275	.498	.323	.609	.515
	Item15 INSTRUC TOR CONFIDENCE	.681	.666	.652 ^b	.598	.634	.451	.325	.326	.528	.374	.624	.531
	Item16 INSTRUC TOR FOCUS LECTURES	.625	.611	.598	.549 ^b	.592	.414	.299	.300	.485	.344	.573	.488
	Item17 INSTRUC TOR USES CLEAR RELEVANT EXAMPLES	.634	.627	.634	.582	.681 ^b	.570	.465	.438	.601	.487	.665	.575
	Item18 INSTRUC TOR SENSITIVE TO STUDENTS	.385	.397	.451	.414	.570	.704 ^b	.666	.586	.626	.625	.604	.541
	Item19 INSTRUC TOR ALLOWS ME TO ASK QUESTIONS	.243	.261	.325	.299	.465	.666	.658 ^b	.568	.556	.599	.505	.460
	Item20 INSTRUC TOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	.261	.275	.326	.300	.438	.586	.588	.494 ^b	.503	.524	.470	.425
	Item21 INSTRUC TOR AWARE OF STUDENTS UNDERSTANDING	.495	.498	.528	.485	.601	.626	.556	.503	.601 ^b	.546	.620	.545
	Item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	.310	.323	.374	.344	.487	.625	.599	.524	.546	.557 ^b	.519	.467
	Item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	.614	.609	.624	.573	.665	.604	.505	.470	.620	.519	.673 ^b	.584
	Item24 COMPARED TO OTHER COURSES THIS COURSE WAS	.517	.515	.531	.488	.575	.541	.460	.425	.545	.467	.584	.509 ^b
Residual ^{a,e}	Item13 INSTRUC WELL PREPARED		-.048	-.081	-.058	-.057	.024	.043	.043	-.019	.023	-.050	-.063
	Item14 INSTRUC SCHOLARLY GRASP	-.048		-.031	-.111	-.075	.036	.059	.040	-.049	.010	-.045	-.072
	Item15 INSTRUC TOR CONFIDENCE	-.081	-.031		-.093	-.048	.006	.034	.030	-.019	-.005	-.041	-.096
	Item16 INSTRUC TOR FOCUS LECTURES	-.058	-.111	-.093		.004	-.010	.036	.017	-.033	.019	-.114	-.058
	Item17 INSTRUC TOR USES CLEAR RELEVANT EXAMPLES	-.057	-.075	-.048	.004		-.016	-.015	-.021	-.006	-.037	-.052	-.055
	Item18 INSTRUC TOR SENSITIVE TO STUDENTS	.024	.036	.006	-.010	-.016		-.039	-.065	-.071	-.089	-.034	-.067
	Item19 INSTRUC TOR ALLOWS ME TO ASK QUESTIONS	.043	.059	.034	.036	-.015	-.039		-.121	-.057	-.115	-.061	-.086
	Item20 INSTRUC TOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	.043	.040	.030	.017	-.021	-.065	-.121		-.078	-.141	-.061	-.068
	Item21 INSTRUC TOR AWARE OF STUDENTS UNDERSTANDING	-.019	-.049	-.019	-.033	-.006	-.071	-.057	-.078		-.040	-.022	-.046
	Item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	.023	.010	-.005	.019	-.037	-.089	-.115	-.141	-.040		-.026	-.022
	Item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	-.050	-.045	-.041	-.114	-.052	-.034	-.061	-.061	-.022	-.026		.120
	Item24 COMPARED TO OTHER COURSES THIS COURSE WAS	-.063	-.072	-.096	-.058	-.055	-.067	-.086	-.068	-.046	-.022	.120	

Extraction Method: Principal Component Analysis.

a. Residuals are computed between observed and reproduced correlations. There are 28 (42.0%) nonredundant residuals with absolute values greater than 0.05.

b. Reproduced communalities

c. Reproduced Correlations - This table contains two tables, the

reproduced correlations in the top part of the table, and the residuals in the bottom part of the table.

d. **Reproduced Correlation** - The reproduced correlation matrix is the correlation matrix based on the extracted components. You want the values in the reproduced matrix to be as close to the values in the original correlation matrix as possible. This means that you want the residual matrix, which contains the differences between the original and the reproduced matrix, to be close to zero. If the reproduced matrix is very similar to the original correlation matrix, then you know that the components that were extracted accounted for a great deal of the variance in the original correlation matrix, and these few components do a good job of representing the original data. The numbers on the diagonal of the reproduced correlation matrix are presented

in the **Communalities** table in the column labeled **Extracted**.

e. **Residual** - As noted in the first footnote provided by SPSS (a.), the values in this part of the table represent the differences between original correlations (shown in the correlation table at the beginning of the output) and the reproduced correlations, which are shown in the top part of this table.

For example, the original correlation between item13 and item14 is .661, and the reproduced correlation between these two variables is .710. The residual is $-.048 = .661 - .710$ (with some rounding error).