

Vote-Counting Method to Obtain an Estimate and Confidence Interval for the Population Correlation Coefficient Using SAS/AF*

Frank C. Lieble III, SAS Institute Inc., Orlando, Florida
Morgan C. Wang, University of Central Florida, Orlando, Florida

Abstract

In today's information age, scientific studies are being conducted at an exponential rate in areas such as psychology, pharmaceuticals, and education. The meta-analyst wants to combine these studies which investigate a common phenomenon into useful information. However, missing values pose a particularly difficult problem in meta-analysis. This paper will provide a solution, vote-counting procedures, to estimate the population correlation coefficient when some of the studies do not provide the information about the sample correlation coefficient. SAS/AF was used in the development of an interactive application system for calculating and displaying the confidence interval for the population correlation coefficient.

Introduction

A meta-study consists of a sample of independently related studies, for instance, SAT scores, drug effectiveness, public health, or human behavior. Meta-analysis is a quantitative approach to organize and integrate independent, related studies so that generalized conclusions can be made. Meta-analysis uses a variety of statistical techniques, specifically designed for summarizing, integrating, and testing the results of scientifically related studies (Mullen, 1989). This paper will illustrate one of the meta-analytic procedures, vote-counting, for estimating the population correlation coefficient. Due to the complexity of unequal sample sizes within vote-counting procedures, equal sample sizes have been assumed in this application.

Vote-Counting Procedures

Vote-counting procedures are methods used for estimating the confidence interval of the population correlation coefficient and are influenced by the level of significance and sample sizes.

Methods

- **proportion of positive correlations** - the number of positively correlated sample studies divided by the total number of sample studies in the meta-study.

- **proportion of significant positive correlations** - the number of statistically significant positively correlated sample studies divided by the total number of sample studies in the meta-study.
- **proportion of positive and negative significant correlations** - the number of statistically significant positively correlated sample studies divided by the sum of negative and positive statistically significant sample studies in the meta-study.

Level of Significance

The level of significance is the statistical metric chosen by the meta-analyst in determining whether or not the meta-study is statistically significant.

Sample Sizes

The single largest concern of sample sizes is whether or not the sample studies included in the meta-study are equal/similar or vastly different. This paper addresses methods with equal/similar sample sizes. Vastly different sample sizes use the maximum likelihood estimation which is out of the scope of this paper.

Sample studies that have equal/similar sample sizes can be averaged using a general class of means suggested by Hedges and Olkin (1985).

$$\bar{n}_m = \left(\frac{1}{k} \sum_{i=1}^k n_i^{1/m} \right)^m, \quad m = 0, 1, 2, \dots$$

The arithmetic mean is derived when $m=1$. If $m=2$ then the square mean root is derived. Gibbons, Olkin, and Sobel (1977) recommend the square mean root which will provide a more conservative average than the arithmetic mean.

Incomplete Sample Study Data

Many sample studies have the problem of missing data. The removal of these studies will bias the calculation of the population correlation coefficient by reducing the denominator of the proportion of correlation. Instead of

removing these studies, the vote-counting procedure will provide an effective way for estimating the population correlation coefficient confidence interval (Wang and Silver, 1994; Bushman and Wang, 1994).

Meta-Study Application

A customized application was developed using SAS/AF FRAME entries which allows the meta-analyst to manage and execute meta-studies as well as displaying the correlation coefficient confidence intervals.

When the application is invoked, the Meta-Study Management screen (Figure 1) will appear. This screen will provide the ability to create, modify, or delete a meta-study, by name, which is restricted to eight characters. Help is readily available throughout the application by pressing the **Help** button.

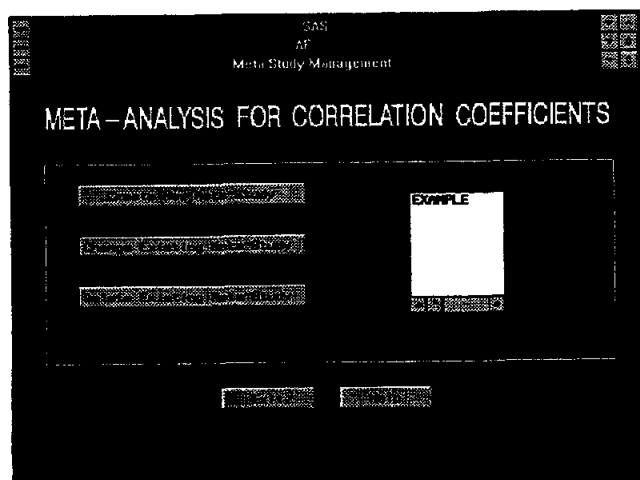


Figure 1

The Meta-Study Attributes screen (Figure 2) is used to enter meta-study attributes interactively. The Meta-Study Description field will provide a detailed description of the Meta-Study Name.

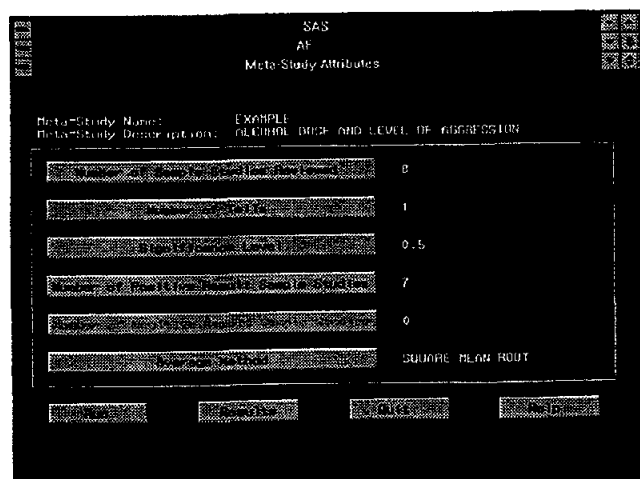


Figure 2

The remaining buttons will invoke pop-up dialog screens, in which the meta-analyst can enter the appropriate meta-study attribute.

The attribute buttons are defined as follows:

Number of Sample Studies Reviewed - Defines the number of independently related sample studies in the meta-study. By the use of SAS/FSP's FSVIEW edit procedure (Figure 3), the meta-analyst can enter the sample size of each sample study as an observation.

OBS	SIZE
1	10
2	4
3	5
4	6
5	4
6	3
7	4
8	6

Figure 3

Number of Tails - Define one tail when samples are based on positive or positive significant meta-studies, or two-tails for significant positive and negative meta-studies.

Significance Level - Define a significance level between 0.0 and 0.5 inclusive.

Number of Positive Result Sample Studies - Defines the number of positive or positive significant sample studies. The number of positive result sample studies cannot exceed the total number of sample studies in the meta-study.

Number of Negative Result Sample Studies - This button is active only when the two-tail option has been selected. Enter the number of negative significant sample studies. The number of negative result sample studies cannot exceed the total number of sample studies in the meta-study or the difference between the total number of sample studies and positive result sample studies.

Average Method - Defines the type of average method to apply to the meta-study. The available methods are:

- Minimum of sample sizes
- Arithmetic average
- Square mean root
- Harmonic mean

Run - Invoke SAS® software to calculate population correlation coefficient confidence intervals.

Results - This button will activate after the first execution of a meta-study. The Meta-Analysis Results screen (Figure 4) will display confidence intervals for the population correlation coefficients. The screen includes meta-study attributes, average sample study size, method of estimation, 90, 95, 99 percent confidence interval levels and corresponding lower/upper intervals.

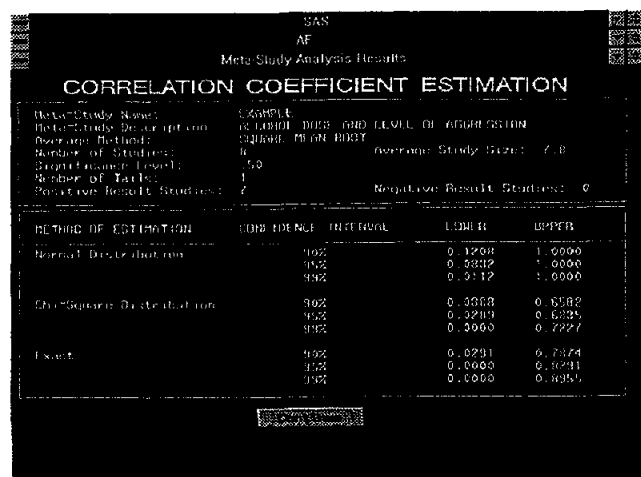


Figure 4

Quit - Return to the Meta-Study Management screen.

Help - On-line help.

Example

For hundreds of years, it has been assumed that people behave more aggressively while "under the influence" of alcohol. Recent meta-analytic reviews offer strong support for this idea (Bushman, in press; Bushman & Cooper, 1990). Table 1 - Correlation Between Alcohol Dose and Human Aggression in Adult Males, contains the results from eight independent sample studies from seven research reports. The research studies used randomized block designs to examine the relation between alcohol dose and level of aggression. Note that the meta-analyst has decided the sample sizes are equal/similar.

The remainder of this paper will present an example of each vote-counting method for calculating the confidence interval for the population correlation coefficient.

Confidence Intervals Based on the Proportion of Positive Correlations

Using the alcohol dose example, the meta-analyst determines the predictive direction of each sample study assigning a plus or minus sign for the direction of the correlation. It was determined that seven of the eight sample studies have positive direction (Table 1, direction column). Because the meta-study is based on positive correlation, the number of tails will be one and the level of significance will be 0.5 which represents the probability of getting a positive direction. The square

Table 1 - Correlation Between Alcohol Dose and Human Aggression in Adult Males

Sample Study	Alcohol dose g/kg	Aggression measure	n	r	Direction	Statistically Significant
1	0, 0.26, 0.53, 0.79	shock	16	.43	+	*
2	0, 0.10, 0.20, 0.40, 0.79, 0.99	money loss	4	.55	+	*
3	0, 0.12, 0.23, 0.46	money loss	8	.84	+	*
4	0, 0.12, 0.23, 0.46	noise	8	.05	+	ns
5	0, 0.12, 0.24, 0.36	money loss	4	.09	+	ns
6	0, 0.26, 0.52, 0.79	shock	8	-.20	-	ns
7	0, 0.24, 0.48, 0.71	money loss	4	.72	+	*
8	0, 0.12, 0.24, 0.48	money loss	6	.28	+	ns

Note. g/kg = grams 100% alcohol per kilogram body weight, 0 g/kg = placebo, n = sample size, r = correlation coefficient, + = positive result, - = negative result, * p<0.05, ns = statistically non significant at the 0.05 level.

mean root will be selected as the average method for the meta-study. These are the attributes required to execute the application.

After the meta-study attributes have been identified, the meta-analyst is ready to enter them into the application. A new meta-study is created by selecting the **Create New Meta Study** button on the Meta-Study Management screen (Figure 1). The meta-analyst enters the meta-study attributes in the Meta-Study Attributes screen (Figure 2). Because the number of tails is one, the **Number of Negative Result Sample Studies** button is inactive. The sample study sizes are entered into a SAS/FSP FSVIEW edit screen (Figure 3) after the **Number of Sample Studies Reviewed** button is selected. It is important to enter the exact number of sample studies defined in the meta-study into the FSVIEW edit screen to ensure data integrity.

The meta-analyst can execute the correlation coefficient calculation by pressing the **Run** button. When the execution process is completed, the results can be reviewed by selecting the **Results** button.

The Meta-Analysis Results screen (Figure 4) shows the 90%, 95%, and 99% confidence intervals for three methods of estimation: normal distribution, chi-square distribution and exact. The meta-analyst should use the method that gives the narrowest confidence interval. Using Figure 4 and the exact method of estimation, the 90% confidence interval does not include the value zero. The meta-analyst has enough statistical evidence to conclude that the level of aggression increases as the alcohol dose increases at the 90% confidence level.

Confidence Intervals Based on the Proportion of Positive Significant Correlations

The difference between the proportion of positive correlations and the positive significant correlations are the number positive results studies and the level of significance. The number of positive results studies are based on the number of sample studies that tested statistically significant at the $\alpha=0.05$. It was determined that four of the eight sample studies have positive significance (Table 1, statistically significant column). The attributes are entered into the Meta-Study Attributes screen (Figure 5).

Figure 5

Figure 6 displays the confidence interval based on the proportion of positive significant corrections. Note that the confidence intervals are smaller than those of the first method which implies that this method produces stronger statistical confidence interval for the population correlation coefficient. The meta-analyst has enough evidence to conclude that the level of aggression increases as the alcohol dose increases for all three confidence intervals using the exact and chi-square distribution estimation methods.

METHOD OF ESTIMATION	CONFIDENCE INTERVAL	LOWER	UPPER
Normal Distribution	90%	0.2509	0.8052
	95%	0.1348	0.8434
	99%	0.0000	1.0000
Chi-Square Distribution	90%	0.3891	0.7448
	95%	0.3501	0.7436
	99%	0.2781	0.7948
Exact	90%	0.3212	0.7768
	95%	0.2681	0.7987
	99%	0.1568	0.8569

Figure 6

Confidence Intervals Based on the Proportion of Positive and Negative Significant Correlations

This vote-counting method is used to examine the extent of biases in the samples. Sample studies with significant results are more likely to be included in a meta-study than are sample studies with non significant results. This will result in a biased subsample of the total collection of studies in the population. The problem can be minimized if the meta-analyst counts both positive and negative

significant results (Hedges and Olkin, 1980). Referring to correlation column r in Table 1, the positive and negative significant results can be determined by counting the number of positive and negative correlation values. There are seven samples with positive correlation and one sample with negative correlation. The meta-analyst enters the counts into the Meta-Study Attributes screen (Figure 7), specify a level of significance of 0.05, and set the number of tails to two.

Figure 7

Figure 8 shows confidence intervals for the population correlation coefficient with biases minimized. Note the confidence intervals does not include the value zero. Again, the meta-analyst has enough evidence to conclude that the level of aggression increases as the alcohol dose increases.

METHOD OF ESTIMATION	CONFIDENCE INTERVAL	LOWER	UPPER
Normal Distribution	95%	0.0421	1.0000
	90%	0.0303	1.0000
	99%	0.0094	1.0000
Chi-Square Distribution	95%	0.0321	0.1252
	90%	0.0112	0.1263
	99%	0.0000	0.1276
Exact	95%	0.0113	0.1292
	90%	0.0000	0.1297
	99%	0.0000	0.1302

Figure 8

Conclusion

Meta-analysis is a quantitative approach to organize and integrate independent, related scientific studies so that generalized conclusions can be made. A meta-study is a

sample of independently related studies from areas such as pharmaceuticals, psychology, and education. With vote-counting methods the meta-analyst is able to calculate the population correlation coefficient using a meta-study. An interactive application was developed using SAS/AF to provide the meta-analyst the capability to enter meta-study attributes for the calculation of the population correlation coefficient.

References

- Bushman, B. J. (In Press). *Human Aggression While Under the Influence of Alcohol and Other Drugs: A Meta-Analysis*. Current Directions in Psychological Science.
- Bushman, B. J. and Cooper, H. M. (1990). *Effects of Alcohol on Human Aggression: An Integrated Research Review*. Psychological Bulletin, 107, 341-354.
- Bushman, B. J. and Wang, M. C. (1994). *Vote-Counting Procedures for Obtaining a Confidence Interval for the Population Correlation Coefficients*. Psychological Bulletin (In Press).
- Gibbons, J. D., Olkin, I. and Sobel, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology*. New York, Wiley.
- Hedges, L. V., and Olkin, I. (1985). *Statistical Methods For Meta-Analysis*. New York, Academic Press.
- Mullen, Brian (1989). *Advance BASIC Meta-Analysis*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Wang, M. C. and Silver, N. C. (1994). *A Microsoft FORTRAN Program for Determining the Confidence Interval Around the Estimate of the Population Correlation Coefficient for the Vote-Counting Method*. Educational and Psychological Measurement (In Press)
- SAS, SAS/AF, and SAS/FSP, are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Contact

A copy of the application may be obtained at no charge by sending a self addressed stamped mailer to:

POSTERS

Morgan C. Wang
University of Central Florida
Department of Statistics
Orlando, FL 32816-2370
407-823-2818

Frank C. Lieble III
SAS Institute Inc.
2250 Lucien Way Suite 100
Maitland, FL 32751
407-661-1711