Hypothesis testing with confidence intervals and P values

Ned Kock

December 2015

Hypothesis testing with confidence intervals and P values

Ned Kock

Abstract

E-collaboration researchers usually employ P values for hypothesis testing, a common practice in a variety of other fields. This is also customary in many methodological contexts, such as analyses of path models with or without latent variables, as well as simpler tests that can be seen as special cases of these (e.g., comparisons of means). We discuss here how a researcher can use another major approach for hypothesis testing, the one building on confidence intervals. We contrast this approach with the one employing P values through the analysis of a simulated dataset, created based on a model grounded on past theory and empirical research. The model refers to social networking site use at work and its impact on job performance. The results of our analyses suggest that tests employing confidence intervals and P values are likely to lead to very similar outcomes in terms of acceptance or rejection of hypotheses.

Introduction

There has been steady growth in the use of structural equation modeling (SEM) employing the partial least squares (PLS) method (Kock, 2014; Kock & Lynn, 2012), referred to here as PLS-SEM, in the field of e-collaboration (Kock, 2005; Kock & Nosek, 2005) as well as in many other fields where multivariate statistics are employed.

This is in part due to the emergence of powerful yet user-friendly software tools such as WarpPLS (Kock, 2010; 2015c), which implement the PLS-SEM method. Another reason for this growth is that, from a conceptual perspective, many statistical tests can be seen as special cases of PLS-SEM analyses. Examples are robust path analyses (Kock & Gaskins, 2014), nonparametric comparisons of means (Kock & Chatelain-Jardón, 2011), and even nonparametric tests where only one group and one condition are available (Kock, 2013). All of these tests can be easily conducted with WarpPLS. A third reason for the growth in the use of PLS-SEM is the emergence of factor-based algorithms, which bring together the advantages of algorithms employed in traditional PLS-SEM with the precision of covariance-based SEM under common factor model conditions (Kock, 2015a).

Hypothesis testing in the context of PLS-SEM is usually conducted through the calculation of a P value for each path coefficient, where the P value may be one-tailed or two-tailed depending on the researcher's prior knowledge about the direction of the path and the sign of its associated coefficient (Kock, 2015b). One question that frequently arises in this context is: How can one use confidence intervals for hypothesis testing, in addition to or instead of P values? This paper is an answer to this question.

Illustrative model and data

Figure 1 shows an illustrative model that is used in our discussion about using confidence intervals and P values in hypothesis testing. This model contains five latent variables: internal social networking tool use (SN), job satisfaction (JS), organizational commitment (OC), job innovativeness (JI), and job performance (JP).

Internal social networking tool use (SN) measures the degree to which an employee uses a social networking tool (e.g., Facebook) made available internally in the employee's organization to facilitate employee-employee socialization. This is the only exogenous (or independent) latent variable in the model. Job performance (JP), the main endogenous (or dependent) latent variable in the model measures the employee's overall performance at work.

Three latent variables mediated the indirect relationship between SN and JP. Job satisfaction (JS) measures the degree to which the employee is satisfied with the organization. Organizational commitment (OC) measures the degree to which the employee is personally committed to the organization. Job innovativeness (JI) measures the degree to which the employee engages in innovative or creative behavior at work.

The figure has been created with WarpPLS, and thus employs the software's notation for summarized latent variable description: the alphanumeric combination under each latent variable's label (e.g., "JP") in the model describes the measurement approach used for that latent variable and the number of indicators. For example "(R)5i" means reflective measurement with 5 indicators.

We employed a Monte Carlo simulation (Robert & Casella, 2005) to create sample data based on this model and on assumptions based on past empirical research. The sample we created had 160 cases, or rows in the data table. The number of columns was 21, the total number of

indicators. We assumed that the indirect relationship between SN and JP is fully mediated by the latent variables JS, OC and JI. That is, we assumed a neutral direct effect SN>JP at the population level.

Figure 1: Illustrative model representation in WarpPLS
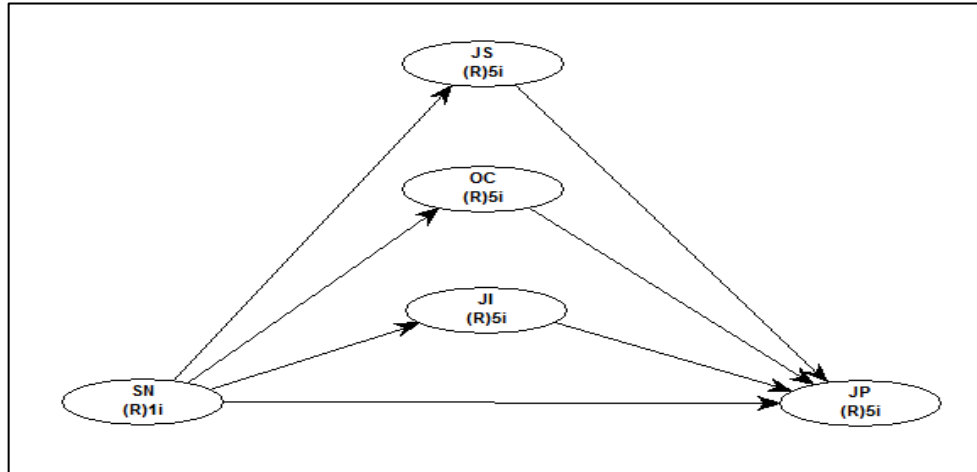
Notes: SN = internal social networking tool use; JS = job satisfaction; OC = organizational commitment; JI = job innovativeness; JP = job performance; notation under latent variable acronym describes measurement approach and number of indicators, e.g., (R)5i = reflective measurement with 5 indicators.

Because of the above, the direct path SN>JP was expected in the subsequent analysis to be associated with a non-significant effect. Nevertheless this path must be included in the model to avoid bias, as its coefficient is only reduced to zero due to the multivariate analysis performed by WarpPLS controlling for the effects of the three latent variable mediators. As a result, the link SN>JP appropriately influences the estimation of the coefficients associated with the links JS>JP, OC>JP and JI>JP. An analysis without the link SN>JP would lead to biased estimates for those links. Our data creation process also forced the link OC>JP to be associated with a weak effect, which we expected to be non-significant at the small sample size we used (*N*=160).

## Hypothesis testing with P values and confidence intervals

Figure 2 illustrates the relationship between P values and confidence intervals. From an intuitive conceptual perspective, P values and confidence intervals are closely related. More technical views grounded on mathematical statistics, however, reveal that they are profoundly different. In our discussion here we highlight the intuitive conceptual perspective, as our focus is primarily on how to contrast the use of P values and confidence intervals for hypothesis testing.

Let us assume that we estimated a path coefficient β through $\hat{\beta}$, and that we also estimated the standard error σ associated with this path coefficient through $\hat{\sigma}$. The standard error estimate $\hat{\sigma}$ is in fact an estimate of the standard deviation of the distribution of estimated path coefficients in a set of samples, of a given size, taken randomly from a population. How do we employ these estimates for hypothesis testing?

P value test. To conduct a test of the hypothesis that $\hat{\beta} > 0$, at the 0.05 significance level (i.e., 1-95%), we calculate the one-tailed P value associated with the path coefficient. Generally speaking, this quantity could be interpreted as the probability that $\hat{\beta}$ belongs to a distribution

4

with mean of zero and standard deviation of $\hat{\sigma}$. It can be calculated as the area under the curve shown on the left side of the figure, considering the total area to be 1. If $P \leq 0.05$ the hypothesis is accepted, otherwise it is rejected. The **T ratio test** can be seen as a variation of this test, where the T ratio (a.k.a. T value, T statistic, and *t*-statistic), calculated as $\hat{\beta}/\hat{\sigma}$, is used instead of the corresponding P value for comparison against a threshold such as 1.64 or 1.96.

Figure 2: P values versus confidence intervals
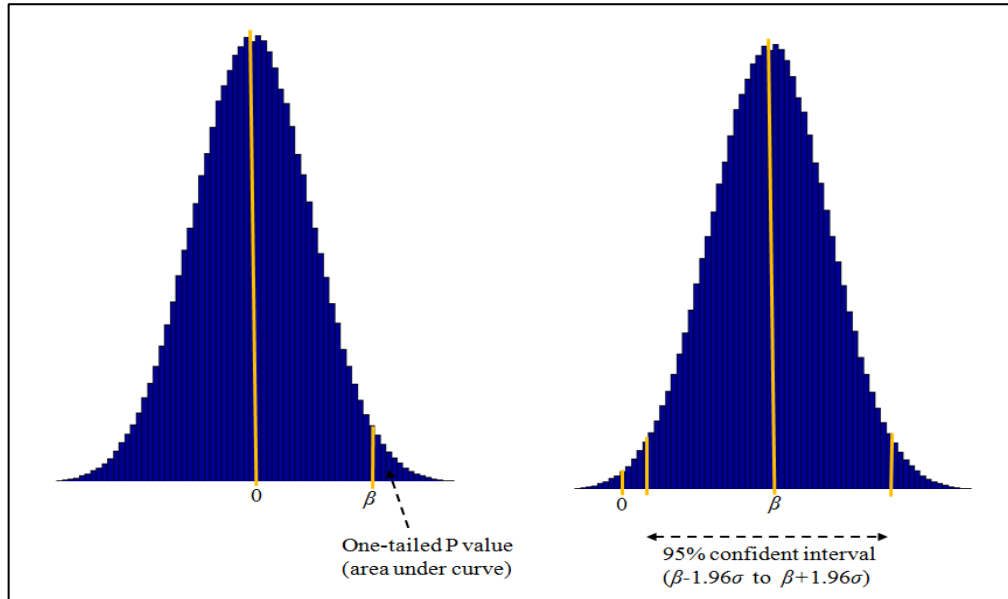
Confidence interval test. To conduct the same test using a 95% confidence interval, we calculate the lower and upper limits of the confidence interval. These are given respectively by $\hat{\beta} - 1.96\hat{\sigma}$ and $\hat{\beta} + 1.96\hat{\sigma}$. If the value 0 (zero) does not fall within this interval (i.e., $0 \notin CI$) the hypothesis is accepted, otherwise (i.e., $0 \in CI$) it is rejected. Looking at the right side of the figure, we see that the value 0 (zero) falls outside the theoretical confidence interval, being located to the left of the interval. Thus, in this theoretical example, the hypothesis that $\hat{\beta} > 0$ would be accepted.

Hypothesis testing results based on illustrative model and data

Table 1 shows the hypothesis testing results based on the illustrative model discussed earlier and the data we created for it. The path coefficients, standard errors, and P values were calculated with WarpPLS employing the following settings: "PLS Mode A" was selected as the outer model analysis algorithm, "linear" was selected as the inner model analysis algorithm for all paths, and "Stable3" was selected as the "resample" method (or the method used for the calculation of standard errors and P values).

As we can see the hypothesis testing results were the same for both tests, employing P values and confidence intervals. With P values, frequently the criterion $P < 0.05$ is used instead $P \leq 0.05$ for accepting a hypothesis. In practice, using either criterion tends to have the same

effect, since it is rare for a P value to be exactly 0.05. This is the case in our example; either criterion leads to the same results.

Table 1: Hypotheses testing with P values and confidence intervals

Path	Path coefficient	Standard error	T ratio	P value	Supported?	Why?	Confidence interval (95%)		Supported?	Why?
SN>JS	0.301	0.074	4.068	<0.001	Yes	P≤0.05	0.156	0.446	Yes	0 ∉ CI
SN>OC	0.281	0.074	3.797	<0.001	Yes	P≤0.05	0.136	0.426	Yes	0 ∉ CI
SN>JI	0.198	0.076	2.605	0.005	Yes	P≤0.05	0.049	0.347	Yes	0 ∉ CI
SN>JP	0.027	0.079	0.342	0.367	No	P>0.05	-0.128	0.182	No	0 ∈ CI
JS>JP	0.265	0.075	3.533	<0.001	Yes	P≤0.05	0.118	0.412	Yes	0 ∉ CI
OC>JP	0.068	0.078	0.872	0.192	No	P>0.05	-0.085	0.221	No	0 ∈ CI
JI>JP	0.318	0.074	4.297	<0.001	Yes	P≤0.05	0.173	0.463	Yes	0 ∉ CI

It is noteworthy that the hypothesis testing results were the same for both tests, because a test based on a one-tailed P value is perceived by many as opening the door for type I errors (false positives). In this example the one-tailed P value tests are apparently as effective in terms of avoiding type I errors as the equivalent tests based on confidence intervals. We know this because we created the data with the expectation that the links SN>JP and OC>JP would be non-significant at the sample size used (N=160).

Discussion and concluding remarks

PLS-SEM has experienced steady growth in the field of e-collaboration and in many other fields where multivariate statistics are employed. One of the reasons for this is the emergence of powerful and user-friendly software tools for PLS-SEM, such as WarpPLS. Another reason is that many statistical tests can be conceptually seen as special cases of PLS-SEM analyses; e.g., nonparametric comparisons of means tests (Kock, 2013).

Researchers usually employ P values for hypothesis testing in PLS-SEM, where each hypothesis refers to a path in a model. P values may be one-tailed or two-tailed, depending on the prior knowledge of the researcher about the path's direction and the sign of its associated coefficient (Kock, 2015a). We discussed here how a researcher can use confidence intervals for hypothesis testing, contrasting this approach with the one employing P values.

We used an illustrative model as a basis for our discussion. This illustrative model contained five latent variables: internal social networking tool use (SN), job satisfaction (JS), organizational commitment (OC), job innovativeness (JI), and job performance (JP). We employed a Monte Carlo simulation to create sample data based on this model and on assumptions based on past empirical research.

The results of our analyses suggested that employing confidence intervals and P values are likely to lead to very similar outcomes in terms of acceptance or rejection of hypotheses. In fact, the outcomes were identical in our model. The results of our analyses also suggested that, in our model, employing tests based on one-tailed P values at the 0.05 significance level has approximately the same effect in terms of avoidance of false positives as employing tests based on 95% confidence intervals.

Acknowledgments

References

Kock, N. (2005). What is e-collaboration. *International Journal of e-Collaboration*, 1(1), 1-7.

Kock, N. (2010). Using WarpPLS in e-collaboration studies: An overview of five main analysis steps. *International Journal of e-Collaboration*, 6(4), 1-11.

Kock, N. (2013). Using WarpPLS in e-collaboration studies: What if I have only one group and one condition? International Journal of e-Collaboration, 9(3), 1-12.

Kock, N. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration*, 10(3), 1-13.

Kock, N. (2015a). A note on how to conduct a factor-based PLS-SEM analysis. *International Journal of e-Collaboration*, 11(3), 1-9.

Kock, N. (2015b). One-tailed or two-tailed P values in PLS-SEM? *International Journal of e-Collaboration*, 11(2), 1-7.

Kock, N. (2015c). *WarpPLS 5.0 User Manual*. Laredo, TX: ScriptWarp Systems.

Kock, N., & Chatelain-Jardón, R. (2011). Four guiding principles for research on evolved information processing traits and technology-mediated task performance. *Journal of the Association for Information Systems*, 12(10), 684-713.

Kock, N., & Gaskins, L. (2014). The mediating role of voice and accountability in the relationship between Internet diffusion and government corruption in Latin America and Sub-Saharan Africa. *Information Technology for Development*, 20(1), 23-43.

Kock, N., & Lynn, G.S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, 13(7), 546-580.

Kock, N., & Nosek, J. (2005). Expanding the boundaries of e-collaboration. *IEEE Transactions on Professional Communication*, 48(1), 1-9.

Robert, C.P., & Casella, G. (2005). *Monte Carlo statistical methods*. New York, NY: Springer.