Human Factors Division
Institute of Aviation

University of Illinois
at Urbana-Champaign
1 Airport Road
Savoy, Illinois 61874

Relative Importance of Conflict Geometry Variables in Influencing Pilots Conflict Detection Using a Cockpit Display of Traffic Information

Xidong Xu, Esa M. Rantanen,
& Yan Huo

Technical Report HFD-06-02

February 2006

Abstract

When independent variables are inter-correlated with each other, ANOVAs cannot determine their relative importance in influencing a dependent measure. Similarly, when predictor variables are inter-correlated, there are inherent problems with traditional multiple regression methods for analyzing the relative importance of the prediction variables in accounting for the variance in a criterion variable. This report describes a method called Dominance Analysis (DA) (Budescu, 1993; Azen & Budescu, 2003) as a better approach than the traditional methods in determining the relative importance of several independent variables in accounting for the variances in dependent variables observed in a pilot conflict detection task with the cockpit display of traffic information (CDTI; Xu, Rantanen, Wickens, 2004). The three inter-correlated variables in question were an intruder aircraft's distance to, and time to, the closest point of approach (CPA) between the pilot's ownship and the intruder aircraft, and relative speed between the two aircraft. Results indicate 1) for absolute miss distance estimate error, distance to CPA was the most important variable than the other two variables; (2) for signed miss distance estimate error, time to CPA and distance to CPA were more important than relative speed; 3) for both absolute and signed time to CPA estimate errors, time to CPA was the most important compared to the other two variable, and (4) for absolute orientation at CPA estimate error, relative speed was the least important variable compared to distance to CPA and time to CPA.

Introduction

Summary of the Experiment

Xu, Rantanen, and Wickens (2004) systematically investigated the effects of 2-D air traffic geometry on pilots' conflict detection performance using a CDTI. The goal of the experiment was to answer two general questions regarding the use of CDTIs for conflict detection unaided by automation. First, we were interested in what geometry features of air traffic make conflict understanding difficult. 'Understanding' here was mainly assessed by estimate errors of miss distance at the closest point of approach (CPA) between the pilot's ownship and an intruder aircraft, time to CPA, and orientation at CPA, where these errors were the differences between the true values of these three parameters and participants' estimation of the respective values. 'Difficulty' was operationally defined by an increase in the absolute estimate errors. We note, however, that an increase in absolute error can result either from an increase in symmetrical variability around the true value or from a systematic bias (i.e., constant error away from the true value). Hence, our second question addressed the systematic biases to over- or underestimate a quantity. Such biases were assessed by measures of signed errors.

A part task simulation was run on a Dell PC and viewed on a 21-in. color display. Participating pilots individually observed the development of conflict scenarios each involving the ownship and the intruder aircraft. The ownship was flying to a designated navigation waypoint located directly ahead of the ownship but outside the view on the display. The ownship and the intruder were flying at the same altitude on straight and converging courses, at constant but not necessarily the same speed. The ownship icon was positioned in the center of the display throughout the whole experiment, appearing stationary to the participant and thus yielding an ownship-centered view of the traffic situation. Participants observed the development of a conflict scenario for 15 s, after which the scenario froze, retaining the depiction of the aircraft icons on the display. Participants were then required to mentally extrapolate the development of the scenario and press a key when they estimated that the CPA was reached (had the scenario not been frozen), thereby providing an estimate of time to CPA. Participants were then required to move the cursor to their best estimate of the location of the CPA and click the left mouse button, thus yielding an estimate of miss distance and orientation at CPA.

The independent variables employed were (1) intruder's distance to CPA at freezing point (1.33, 2.67, and 4.0 nm), (2) intruder's relative speed, which was defined as the speed at which the intruder was moving in the ownship-centered frame of reference and thus determined how rapidly the two aircraft would converge (160, 240, and 480 knots), (3) miss distance (0.67 nm, 2.67 nm, and 4.67 nm), (4) approach side (from the left vs. from the right), and (5) conflict angle, the angular difference between the traffic's relative trajectory and the ownship's heading (45˚, 90˚, and 135˚). See Figure 1 for an illustration of the key components and independent variables in this experiment. Another independent variable manipulated was the aspect ('front' vs. 'behind') on which the intruder passed the ownship when CPA was reached. However, since analysis showed no effect of approach aspect on performance, trials with different aspects were treated as replicates. Note that because intruder's time to CPA was determined by distance to CPA and relative speed (i.e., time to CPA was determined by distance to CPA divided by relative speed), an orthogonal manipulation of time to CPA was unnecessary as well as impossible. Also note that since coupling the longest distance to CPA (4.0 nm) with the slowest speed (160 knots) resulted in a time to CPA of 90 s, which might have been excessive and resulted in participant distraction and impatience, it was excluded from the experiment. Within the above-described distance to CPA and relative speed levels, some pairs of trials had the same time to CPA when the freezing occurred, but different distances to CPA because of different relative speed levels, allowing for the testing of the distance-over-speed bias hypothesis.

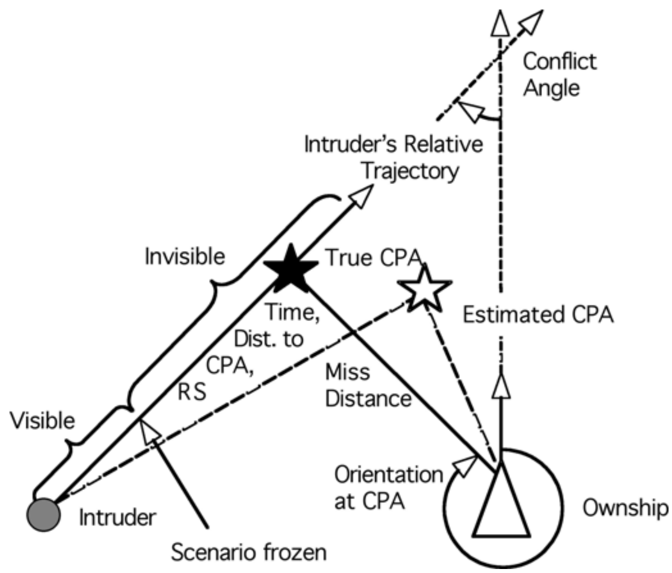

Figure 1. Schematic illustration of key components of the experimental paradigm and independent variables. The ownship icon was stationary to the participant.

This was a mixed design. Distance to CPA was varied between subjects and the other variables varied within subjects. For the 1.33-nm and 2.67-nm distance to CPA groups, crossing the three relative speed levels with the three miss distance levels, the two approach sides, and the three conflict angles yielded 54 conflict geometries. Two replicates (across two approach aspects of 'in front' and 'behind' ownship) of each of the 54 conditions resulted in 108 trials in total for each of the two distances to CPA groups (3 relative speeds x 3 miss distances x 2 approach sides x 3 conflict angles x 2 replicates). For the 4.0-nm distance to CPA group, there were a total of 72 trials (2 faster relative speeds x 3 miss distances x 2 approach sides x 3 conflict angles x 2 replicates).

The primary dependant variables were absolute and signed estimate errors of miss distance and those of time to CPA, derived by subtracting the true values from their corresponding estimated values (i.e., '|estimated values – true values|' and 'estimated values – true values,' respectively). For orientation at CPA, only its absolute estimate error was employed, defined by the absolute difference between the estimated and the true values. Absolute errors would reveal the estimation accuracy, whereas signed errors would reveal the estimation directions (under- or overestimation), an indication of biases.

Results based on ANOVAs indicated: (1) Increased estimation errors with slower speeds, longer times and distances to CPA, and longer miss distances; (2) best performance for conflict angle of 90°, (3) a bias to judge conflicts to be more risky than was actually the case; and (4) a 'distance-over-speed' bias, such that two aircraft farther apart and converging rapidly were perceived as less risky than when they were closer to each other and converging at a slower rate, despite identical time until conflict. These findings have important implications for pilot selection and training, and the design of procedures, displays, and decision support tools for the free flight environment need to take the human performance limitations into account.

Relative Importance of Independent Variables in Accounting for Variances in Dependent Variables

In Xu et al. (2004), because of the interrelations among distance to CPA, relative speed, and time to CPA as noted above, the effect of one variable could well be confounded with the effect(s) of another or the other two. For example, when there was an effect of distance to CPA (as revealed by the ANOVA), it is not clear whether this effect was due to distance to CPA itself, or time to CPA, or a combination of these two variables, because a change in the former variable was always associated with a change in the latter variable when speed was controlled for. Similarly, when an effect of speed was found using ANOVA, it is difficult to tell whether it was caused by speed per se, by time, or by both speed and time, because when distance was controlled for, an increase in speed was always associated with a decrease in time. Yet, understanding the relative importance of each variable in influencing the estimation performance is critical to correct interpretation of results of this research.

Problems with traditional multiple regression analysis methods. Multiple regression is a common tool to examine the relative importance of predictor variables or independent variables that are correlated, and various relative importance measures have been developed such as those based on regression coefficients (e.g., the regression coefficient β_i, associated with predictor X_i), those based on correlation (e.g., simple product-moment correlation, r_{YXi}, or the squared product-moment correlation, r^2_{YXi}, between the criterion and each of the predictors, and squared partial correlation), and those based on a combination of the regression coefficients and the correlations (e.g., the product of the standardized regression coefficient of the predictor and the predictor's correlation with the criterion, $\beta_i r_{YXi}$) (see Azen & Budescu, 2003 for a review). However, according to Azen and Budescu, most of these measures are either correct only in special cases (e.g., r^2_{YXi} is a correct relative importance measure only when predictors are not correlated, but not necessarily so when they are correlated) or lack an intuitive interpretation, being subject to problems such as confusion and misunderstanding (also see Courville & Thompson, 2001).

An alternative—dominance analysis. Instead of relying on conventional multiple regression methods, Budescu (1993) and Azen and Budescu (2003) developed a general yet intuitive method of analyzing relative importance of predictor variables, known as dominance analysis (DA), which has been recognized as a better approach and used by many researchers in recent years (e.g., Behson, 2002; Block, 1995; Johnson, 2000; Nickerson, Schwartz, Diener, & Kahneman, 2003; Suh, Diener, Oishi, & Triandis, 1998; Weinberger, 1995). According to Azen and Budescu (2003), relative importance of predictor variables in predicting the criterion variable, Y, can be analyzed at three levels, namely, complete dominance, conditional dominance, and general dominance, with decreasing strictness of dominance. Variable X_i is said to completely dominate variable X_j if the additional contribution of X_i to each of the subset models for which the comparison between the two variables is meaningful is greater than that of X_j. Here the additional contribution of a predictor variable to a given regression model is defined as the increase in the proportion of variance in Y accounted for by adding that predictor variable to the regression model. However, complete dominance between two variables cannot be established, if one variable's additional contribution is greater than another's for some, but not all, of the subset models. To reduce the dominance indeterminacy at the complete dominance level, the second level of dominance, conditional dominance, is based on the average of the additional contributions to all subset models of a given model size, which is the number of predictor variables in a model. For instance, if X_i's average additional contribution for all the model sizes is greater than that of X_j's, then X_i conditionally dominates X_j. As in the case of complete dominance, conditional dominance cannot be established if one variable's average contribution for some but not for all model sizes is greater than another. The last level of dominance, general dominance, further reduces dominance indeterminacy, by taking the overall average of the average values from all the model sizes used for conditional dominance. X_i is said to be generally dominant over X_j, if X_i's overall average conditional contribution is greater than that of X_j.

Method

For the present experiment, because the three independent variables are inter-correlated, the DA method described by Azen and Budescu (2003) should be an appropriate method for analyzing their relative importance in influencing pilot's conflict detection (DA can also be used with experimental data by specifying fixed rather than random independent variables; personal communications with Razia Azen, September 19, 2005). Tables 1 and 2 show the correlation matrixes for the three predictor variables:

relative speed, distance to CPA, and time to CPA, for DA (A) and (B), respectively. In Xu et al. (2004), because of the incomplete factorial design (eight rather than nine conditions resulting from the dropping of the longest distance to CPA/slowest relative speed condition), hypotheses with respect to the effects of relative speed, distance to CPA, and time to CPA were evaluated with two partly overlapping ANOVAs—(A) on the two shorter distances to CPA and all the three relative speeds, and (B) on all the three distances to CPA at the two faster relative speeds. Similarly, the data for the DA were also split into two partly overlapping sets—(A) and (B) corresponding to the ANOVAs (A) and (B). The DA was conducted separately for (A) and (B) and will be referred to as DA (A) and (B), respectively, and they were expected to reveal similar result patterns. For each of the five dependent variables, the DA was performed at all the three hierarchical levels—complete, conditional, and general dominance.

Table 1

Sample (n = 48) correlation matrix (A) for the three predictor variables: relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA)

Variable	1. RS	2. DCPA	3. TCPA
1. RS	--	.000	
2. DCPA	.000	--	.612**
3. TCPA	-.721**	.612**	--

** Correlation is significant at the .01 level (2-tailed).

Table 2

Sample (n = 48) correlation matrix (B) for the three predictor variables: relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA)

Variable	1. RS	2. DCPA	3. TCPA
1. RS	--	.000	-.612**
2. DCPA	.000	--	.750**
3. TCPA	-.612**	.750**	--

** Correlation is significant at the .01 level (2-tailed).

We are also interested in whether the dominance results obtained can be generalized beyond the particular sample observed from the current experiment. The bootstrap procedure was used to obtain the inference results. 'The bootstrap is a procedure that was developed to empirically estimate the variability of a statistic whose theoretical distribution is unknown. The procedure 'resamples' the data with replacement to generate an empirical estimate of the entire sampling distribution of the statistic of interest (Efron, 1979; Mooney & Duval, 1993). Starting with the n observations in the parent sample (obtained by the researcher), each resample or bootstrap sample is obtained by randomly drawing n observations with replacement from the parent sample. The resulting bootstrap sample thus consists of a set of n observations that are similar, but not (usually) identical, to those in the parent sample and can be treated as a random sample drawn from the same population that generated the parent sample. This process is repeated S times (where S is a large number) to generate a total of S bootstrap samples' (Azen & Budescu, 2003, p. 140). For multiple regression, different bootstrap processes are used for fixed and random predictors. In this study, although the predictors (or independent variables) were fixed (rather than random), we used random re-sampling bootstrap process. This was justified because using random re-sampling bootstrap has some statistical advantages over fixed re-sampling bootstrap even when the predictors are fixed (Fox, 2002).

The DA and the bootstrap procedure were performed using a SAS program called Dominance Probability Macro provided by Azen and Budescu (2003) (see Appendixes for the SAS codes used for the present study, along with the data).

Results

The additional contributions of each predictor to each subset model used in DA are shown in Tables 3, 5, 7, 9, 11, 13, 15, 17, 19, and 21 for the five dependent variables, respectively. The corresponding bootstrap inference results are shown in Tables, 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20, respectively. Following Azen and Budescu (2003), we use the notation R2 to represent the proportion of variance in the criterion variable accounted for by the predictors in each model. In each of Tables 3, 5, 7, 9, 11, 13, 15, 17, 19, and 21, the 23 = 8 models and their corresponding R2 values are shown in the first two columns. For example, in Table 3 regarding DA (A) for absolute miss distance estimate error, R2 = .4522 represents the proportion of variance in absolute miss distance estimate error that is accounted for by the model consisting of relative speed and distance to CPA. The additional contribution of a given predictor is measured by the increase in R2 that results from adding that predictor to the regression model. Thus, the additional contributions of relative speed, for example, are computed as the increase in the proportion of variance accounted for when relative speed is added to each subset of the remaining predictors (i.e., the null subset {.}, {distance to CPA}, {time to CPA}, and {distance to CPA, time to CPA}. For example, in Table 3, the additional contribution of relative speed to the subset model {time to CPA} is defined as R2relative speed, time to CPA - R2time to CPA. This is the difference between the proportion of variance in absolute miss distance estimate error accounted for by both relative speed and time to CPA and the proportion of variance in absolute miss distance estimate error accounted for by time to CPA alone. This value is also the squared semipartial correlation between relative speed and absolute miss distance estimate error partialling out or controlling for the effect of time to CPA on relative speed.

In the following sections, detailed results will be described for DA (A) for absolute miss distance estimate error as an example to illustrate the DA method, whereas for the rest of the analysis, only summaries of the results will be provided.

DA for Absolute Miss Distance Estimate Error

DA (A). In Table 3 the only subset models to which both relative speed and distance to CPA make additional contributions are {.} (the null subset) and {time to CPA}. Thus, the additional contribution of relative speed can be meaningfully compared with the additional contribution of distance to CPA only in the context of these common subset models. The additional contributions made by relative speed and distance to CPA to the null subset are simply .0452 and .4070, respectively. In the body of Table 3 these are the entries in the first row under the RS and DCPA columns, respectively. The additional contributions of relative speed and distance to CPA to the subset model {time to CPA} are .0878 and .1278, respectively. In the body of Table 3 these are the entries in the row labeled TCPA under the RS and DCPA columns, respectively. Because the additional contribution of distance to CPA is larger than that of relative speed to each of these subset models, distance to CPA dominates relative speed (and has a higher relative importance). One predictor is said to completely dominate another if its additional contribution to each of the subset models that form the basis for comparison is greater than that of the other predictor.

Table 3

Dominance analysis (A) for absolute miss distance (MD) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.0452	.4070	.3366
RS	.0452		.4070	.3793
DCPA	.4070	.0452		.0575
TCPA	.3366	.0878	.1278	
k = 1 average		.0665	.2674	.2184
RS, DCPA	.4522			.0125
RS, TCPA	.4245		.0402	
DCPA, TCPA	.4644	.0002		
k = 2 average		.0002	.0402	.0125
RS, DCPA, TCPA	.4647			
Overall average		.0373	.2382	.1892

This analysis can be repeated for each pair of predictors (the additional contribution of time to CPA is also given in Table 3, under the TCPA column). I can be seen from Table 3 that time to CPA completely dominates relative speed because the contribution of the former to each of the subset models that form the basis for comparison is greater than that of the latter. Similarly, distance to CPA completely dominates time to CPA. Summarizing the above we conclude that for absolute MD estimate error, distance to CPA completely dominated time to CPA, which in turn completely dominated relative speed.

To reduce the incidence of undetermined dominance, Azen and Budescu (2003) introduce two weaker levels of dominance. The first of these compares the additional contributions of each predictor to all subset models as before, but the measure ultimately used to compare the predictors is the average of the additional contributions to all subset models of a given model size. The model size is defined as the number of predictors included in the subset model and is denoted by *k*. If the average additional contribution within each model size is greater for one predictor than the other, then that predictor is said to conditionally dominate the other. For example, the average contribution of relative speed to models of size 1 is computed as [.0452 + .0878]/2 = .0665, which is the entry in the '*k* = 1 average' row under the RS column. Similarly, the average contribution of distance to CPA to the models of size 1 is [.4070 + .1278]/2 = .2674. Because the average contribution (within model size) of distance to CPA is greater than that of relative speed for each model size (i.e., .4070 > .0452 for k = 0, .2674 > .0665 for *k* = 1, and .0402 > .0002 for *k* = 2), distance to CPA conditionally dominated relative speed. In addition, distance to CPA conditionally dominates time to CPA, and time to CPA conditionally dominates relative speed.

The last level of dominance summarizes the additional contributions of each predictor to all subset models by averaging all the conditional values. In this example, this consists of averaging the three averaged entries in each column of Table 3. If this overall averaged additional contribution is greater for one predictor than the other, that predictor is said to generally dominate the other. The corresponding values for this measure are shown in the last row of Table 3. For example, the general measure for relative speed is computed as [.0452 + .0665 + .0002]/3 = .0373. In terms of interpretation, the general measure represents the average difference in fit between all subset models (of equal size) that include X_i and those that do not include it. It is always possible to establish a general dominance ordering unless the general measure is identical for a pair of predictors. In this example, distance to CPA generally dominates the other two predictors, and time to CPA generally dominates relative speed (thus the ordering is distance CPA, time to CPA, relative speed).

Table 4

Inference results for dominance analysis (A) for absolute miss distance (MD) estimate error: D_{ij} values in the sample (n = 48) and their means (D_{ij}), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap samples

i	j	Sample *Dij*	AverageDij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
				Complete dominance				
DCPA	RS	1	0.9165	0.187	0.833	0.000	0.167	0.833
DCPA	TCPA	1	0.7260	0.402	0.651	0.199	0.150	0.651
TCPA	RS	1	0.8490	0.230	0.698	0.000	0.302	0.698
				Conditional dominance				
DCPA	RS	1	0.9165	0.187	0.833	0.000	0.167	0.833
DCPA	TCPA	1	0.7260	0.402	0.651	0.199	0.150	0.651
TCPA	RS	1	0.8490	0.230	0.698	0.000	0.302	0.698
				General dominance				
DCPA	RS	1	1.0000	0.000	1.000	0.000	0.000	1.000
DCPA	TCPA	1	0.7410	0.438	0.741	0.259	0.000	0.741
TCPA	RS	1	1.0000	0.000	1.000	0.000	0.000	1.000

For inference results, the total number of bootstrap samples is 1000 for the present study for each of the DAs. For each DA, the data obtained from the experiment can be considered as a sample (n = 48) generated from the population. This sample was then bootstrapped S = 1,000 times. The results for DA (A) for absolute miss distance estimate error are shown in Table 4. The first and second columns are the two variables being compared (e.g., distance to CPA vs. relative speed); the third column is the value of Dij obtained from the sample; the fourth column is the average (D_{ij}) value over the bootstrap samples; and the fifth column is the standard error (SE) of the D_{ij} values over the bootstrap samples, which represents the expected variability of the dominance value over repeated sampling. The next three columns describe the distribution of D_{ij} over the S bootstrap samples; they show the proportion of bootstrap samples in which X_i dominated X_j or $D_{ij} = 1$ (Pij), the proportion of samples in which X_j dominated X_i or $D_{ij} = 0$ (Pji), and the proportion of samples in which dominance between X_i and X_j could not be established or $D_{ij} = 0.5$ (P_{noij}). The last column is the reproducibility of (or proportion of bootstrap samples that agree with) the sample results. The complete dominance results show that, for this particular sample, distance CPA indeed dominates time to CPA, and time to CPA dominates relative speed. The average D_{ij} values reflect these results and refine them by demonstrating the expected value of D_{ij} over repeated sampling. For example, although both distance to CPA and time to CPA dominate relative speed in the sample, the dominance of time to CPA over relative speed is expected to be established less often (average $D_{\text{time to CPA over relative speed}}$ = .8490) than the dominance of distance to CPA over relative speed (average $D_{\text{distance to CPA over relative speed}}$ = .9165). The standard error estimates give the expected amount of variability of the Dij values around their expected values. More details are given by the three probability entries, one of which also serves as the reproducibility measure. Note, for example, that the proportion of undetermined outcomes was much higher between time to CPA and relative speed (P_{no} = .302) than between distance to CPA and relative speed (P_{no} = .167). The reproducibility values show that the sample result D time to CPA over relative speed = 1 was obtained in only 69.8% of all bootstrap samples, whereas the sample result D distance to CPA over relative speed = 1 was obtained in 83.3% of all bootstrap samples.

The inference results of the conditional dominance are exactly the same at those for complete dominance. The general dominance results show a similar pattern while the average D_{ij} values become closer to either 0 or 1 as the level of analysis becomes 'weaker.' For example, in the complete and conditional dominance cases, average $D_{\textit{time to CPA over relative speed}}$ = 0.8490; and in the general dominance case, average $D_{\textit{time to CPA over relative speed}}$ = 1. This result is similarly demonstrated by the fact that the sample reproducibility of $D_{\textit{time to CPA over relative speed}}$ = 1 increases from 69.8% for the complete and conditional dominance to 100% for the general dominance. The conclusion that can be drawn from these results is that the level of confidence in the dominance of time to CPA over relative speed is higher in the general dominance case that uses a weaker standard.

The pattern of results indicates an increase in the reproducibility of $D_{ij} = 1$ as the dominance level is relaxed (from complete and conditional to general). Intuitively, this pattern of results makes sense—in the case of $D_{ij} = 1$, as the level of dominance is relaxed, more cases that satisfy this dominance relationship will occur, leading to higher reproducibility.

DA (B). The results of DA (B) are very similar to those of DA (A), except that complete dominance cannot be established between distance to CPA and time to CPA. According to Azen and Budescu (2003), if one predictor's additional contribution is greater than the other's for some, but not all, of these subset models, complete dominance (and relative importance) cannot be established and is said to be undetermined. The additional contribution of distance to CPA to the subset model {.} (in the null row and DCPA column of Table 5) is .6240, whereas the additional contribution of time to CPA to the subset model {.} (in the null row and TCPA column of Table 5) is .5776, suggesting that in this case distance to CPA dominates time to CPA. However, the additional contribution of distance to CPA to the subset model {relative speed} (in the RS row and DCPA column of Table 5) is .6240 and the additional contribution of time to CPA to the subset model {relative speed} (in the RS row and TCPA column of Table 5) is .6244, suggesting that time to CPA dominates distance to CPA. Thus, complete dominance cannot be established between distance to CPA and time to CPA—their relative importance changes depending on which subset model forms the basis of comparison. The results in Table 5 indicate that although distance to CPA completely dominates relative speed, and time to CPA completely dominates relative speed, complete dominance cannot be established between distance to CPA and time to CPA.

Similarly, Table 5 shows that although at the conditional dominance level distance to CPA dominates relative speed, and time to CPA dominates relative speed, dominance cannot be established between distance to CPA and time to CPA, because distance to CPA conditionally dominates time to CPA for k = 0 and 1, but time to CPA conditionally dominates distance to CPA for k = 2. Finally, at the general dominance level, distance to CPA dominates relative speed and time to CPA, which in turn dominates relative speed.

Table 5

Dominance analysis (B) for absolute miss distance (MD) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.0488	.6240	.5776
RS	.0488		.6240	.6244
DCPA	.6240	.0488		.0642
TCPA	.5776	.0956	.1106	
k = 1 average		.0722	.3673	.3443
RS, DCPA	.6728			.0166
RS, TCPA	.6732		.0162	
DCPA, TCPA	.6882	.0013		
k = 2 average		.0013	.0162	.0166
RS, DCPA, TCPA	.6895			
Overall average		.0408	.3358	.3129

With respect to inference results, although at the complete an conditional levels, both distance to CPA and time to CPA dominate relative speed in the sample, the dominance of distance to CPA over relative speed (average $D_{\text{time to CPA over relative speed}}$ = .9670) is expected to be more often than the dominance of distance to CPA over relative speed (average $D_{\text{distance to CPA over relative speed}}$ = .8805). Their reproducibility values are consistent with the average dominance value: the sample result $D_{\text{time to CPA over relative speed}}$ = 1 was obtained in 93.4% of all bootstrap samples, whereas the sample result $D_{\text{distance to CPA over relative speed}}$ = 1 was obtained only in 76.1%. At the complete and conditional dominance levels, dominance between distance to CPA and time to CPA cannot be established, and its reproducibility value is only 33.9%. The inference results at the general dominance level are similar to those at the complete and conditional levels while reducing the number of indeterminacies, as shown by the average D_{ij} values becoming closer to either 0 or 1 as the level of analysis becomes ‘weaker.’ In fact, in the complete and conditional dominance cases, dominance cannot be established between distance to CPA and time to CPA, but at the general level, distance to CPA became dominant over time to CPA, and average $D_{\text{distance to CPA over time to CPA}}$ = .6225 at the complete and conditional levels has increased to average $D_{\text{distance to CPA over time to CPA}}$ = .7030. This result is also reflected on the reproducibility values: the sample reproducibility value increases from 33.9% in the complete and conditional cases to 70.3% in the general case.

Table 6

Inference results for dominance analysis (B) for absolute miss distance (MD) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap samples

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
				Complete dominance				
DCPA	RS	1.0	0.8805	0.213	0.761	0.000	0.239	0.761
DCPA	TCPA	0.5	0.6225	0.388	0.453	0.208	0.339	0.339
TCPA	RS	1.0	0.9670	0.124	0.934	0.000	0.066	0.934
				Conditional dominance				
DCPA	RS	1.0	0.8805	0.213	0.761	0.000	0.239	0.761
DCPA	TCPA	0.5	0.6225	0.388	0.453	0.208	0.339	0.339
TCPA	RS	1.0	0.9670	0.124	0.934	0.000	0.066	0.934
				General dominance				
DCPA	RS	1.0	1.0000	0.000	1.000	0.000	0.000	1.000
DCPA	TCPA	1.0	0.7030	0.457	0.703	0.297	0.000	0.703
TCPA	RS	1.0	1.0000	0.000	1.000	0.000	0.000	1.000

DA for Signed Miss Distance Estimate Error

DA (A). The results in Table 7 show the following pattern: 1) time to CPA completely dominates relative speed and distance to CPA, but complete dominance cannot be established between relative speed and distance to CPA; 2) time to CPA conditionally dominates relative speed and distance to CPA, but conditional dominance cannot be established between relative speed and distance to CPA; and 3) time to CPA generally dominates distance to CPA, which in turn generally dominates relative speed.

Table 7

Dominance analysis (A) for signed miss distance (MD) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.0132	.0206	.0450
RS	.0132		.0206	.0349
DCPA	.0206	.0132		.0247
TCPA	.0450	.0030	.0003	
k = 1 average		.0081	.0105	.0298
RS, DCPA	.0338			.0163
RS, TCPA	.0480		.0021	
DCPA, TCPA	.0453	.0048		
k = 2 average		.0048	.0021	.0163
RS, DCPA, TCPA	.0501			
Overall average		.0087	.0110	.0304

At the complete and conditional dominance levels, the reproducibility values for sample $D_{\text{distance to CPA over relative speed}} = .5$, $D_{\text{time to CPA over relative speed}} = 1$, and $D_{\text{time to CPA over distance to CPA}} = 1$ are 57.4%, 68.3%, and 56.2%, respectively. At the general dominance level, sample $D_{\text{distance to CPA over relative speed}} = 1$, and its reproducibility value is now 57.3; the reproducibility values for sample $D_{\text{time to CPA over relative speed}} = 1$, and $D_{\text{time to CPA over distance to CPA}} = 1$ are 86.3% and 70.5%, respectively.

Table 8

Inference results for dominance analysis (A) for signed miss distance (MD) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap sample

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
				Complete dominance				
DCPA	RS	0.5	0.564	0.320	0.277	0.149	0.574	0.574
TCPA	RS	1.0	0.801	0.317	0.683	0.081	0.236	0.683
TCPA	DCPA	1.0	0.669	0.410	0.562	0.224	0.214	0.562
				Conditional dominance				
DCPA	RS	0.5	0.564	0.320	0.277	0.149	0.574	0.574
TCPA	RS	1.0	0.801	0.317	0.683	0.081	0.236	0.683
TCPA	DCPA	1.0	0.669	0.410	0.562	0.224	0.214	0.562
				General dominance				
DCPA	RS	1.0	0.573	0.495	0.573	0.427	0.000	0.573
TCPA	RS	1.0	0.863	0.344	0.863	0.137	0.000	0.863
TCPA	DCPA	1.0	0.705	0.456	0.705	0.295	0.000	0.705

DA (B). The results in Table 9 show the following pattern: 1) time to CPA completely dominates relative speed, and distance to CPA completely dominates relative speed, but complete dominance cannot be established between distance to CPA and time to CPA; 2) time to CPA conditionally dominates relative speed, and distance to CPA conditionally dominates relative speed, but conditional dominance cannot be established between distance to CPA and time to CPA; and 3) distance to CPA generally dominates time to CPA, which in turn generally dominates relative speed.

At the complete and conditional dominance levels, the reproducibility values for sample $D_{\text{distance to CPA over relative speed}} = 1$, $D_{\text{distance to CPA over time to CPA}} = .5$, and $D_{\text{time to CPA over relative speed}} = 1$ are 53.1%, 31.7%, and 78.8%, respectively. At the general dominance level, sample $D_{\text{distance to CPA over time to CPA}} = 1$, and its reproducibility value is now 58.4%; the reproducibility values for sample $D_{\text{distance to CPA over relative speed}} = 1$, and $D_{\text{time to CPA over relative speed}} = 1$ are 99.9% and 98.9%, respectively.

Table 9

Dominance analysis (B) for signed miss distance (MD) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.0145	.2750	.2453
RS	.0145		.2750	.2841
DCPA	.2750	.0145		.0237
TCPA	.2453	.0534	.0535	
k = 1 average		.0340	.1643	.1539
RS, DCPA	.2896			.0126
RS, TCPA	.2987		.0035	
DCPA, TCPA	.2988	.0034		
k = 2 average		.0034	.0035	.0126
RS, DCPA, TCPA	.3022			
Overall average		.0173	.1476	.1373

Table 10

Inference results for dominance analysis (B) for signed miss distance (MD) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap samples

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
Complete dominance								
DCPA	RS	1.0	0.7655	0.250	0.531	0.000	0.469	0.531
DCPA	TCPA	0.5	0.5305	0.412	0.372	0.311	0.317	0.317
TCPA	RS	1.0	0.8925	0.209	0.788	0.003	0.209	0.788
Conditional dominance								
DCPA	RS	1.0	0.7655	0.250	0.531	0.000	0.469	0.531
DCPA	TCPA	0.5	0.5305	0.412	0.372	0.311	0.317	0.317
TCPA	RS	1.0	0.8925	0.209	0.788	0.003	0.209	0.788
General dominance								
DCPA	RS	1.0	0.9990	0.032	0.999	0.001	0.000	0.999
DCPA	TCPA	1.0	0.5840	0.493	0.584	0.416	0.000	0.584
TCPA	RS	1.0	0.9890	0.104	0.989	0.011	0.000	0.989

DA for Absolute Time to CPA Estimate Error

DA (A). The results in Table 11 show the following pattern: 1) time to CPA completely dominates distance to CPA and relative speed, but complete dominance cannot be established between relative speed and distance to CPA; 2) time to CPA conditionally dominates distance to CPA and relative speed, but conditional dominance cannot be established between relative speed and distance to CPA; and 3) time to CPA generally dominates relative speed, which in turn generally dominates distance to CPA.

Table 11

Dominance analysis (A) for absolute time to CPA (TCPA) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.2895	.1567	.5643
RS	.2895		.1567	.2748
DCPA	.1567	.2895		.4142
TCPA	.5643	.0000	.0066	
k = 1 average		.1448	.0816	.3445
RS, DCPA	.4462			.1386
RS, TCPA	.5643		.0205	
DCPA, TCPA	.5709	.0139		
k = 2 average		.0139	.0205	.1386
RS, DCPA, TCPA	.5848			
Overall average		.1494	.0863	.3491

Table 12

Inference results for dominance analysis (A) for absolute time to CPA (TCPA) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap sample

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
				Complete dominance				
RS	DCPA	0.5	0.5850	0.211	0.188	0.018	0.794	0.794
TCPA	RS	1.0	0.9970	0.039	0.994	0.000	0.006	0.994
TCPA	DCPA	1.0	0.9995	0.016	0.999	0.000	0.001	0.999
				Conditional dominance				
RS	DCPA	0.5	0.5850	0.211	0.188	0.018	0.794	0.794
TCPA	RS	1.0	0.9970	0.039	0.994	0.000	0.006	0.994
TCPA	DCPA	1.0	0.9995	0.016	0.999	0.000	0.001	0.999
				General dominance				
RS	DCPA	1.0	0.9080	0.289	0.908	0.092	0.000	0.908
TCPA	RS	1.0	0.9980	0.045	0.998	0.002	0.000	0.998
TCPA	DCPA	1.0	1.0000	0.000	1.000	0.000	0.000	1.000

At the complete and conditional dominance levels, the reproducibility values for sample $D_{relative\ speed\ over\ distance\ to\ CPA} = .5$, $D_{time\ to\ CPA\ over\ relative\ speed} = 1$, and $D_{time\ to\ CPA\ over\ distance\ to\ CPA} = 1$ are 79.4%, 99.4%, and 99.9%, respectively. At the general dominance level, sample $D_{relative\ speed\ over\ distance\ to\ CPA} = 1$, and its reproducibility value is now 90.8%; the reproducibility values for sample $D_{time\ to\ CPA\ over\ relative\ speed} = 1$, and $D_{time\ to\ CPA\ over\ distance\ to\ CPA} = 1$ are 99.8% and 100%, respectively.

Table 12

Inference results for dominance analysis (A) for absolute time to CPA (TCPA) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap sample

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
			Complete dominance					
RS	DCPA	0.5	0.5850	0.211	0.188	0.018	0.794	0.794
TCPA	RS	1.0	0.9970	0.039	0.994	0.000	0.006	0.994
TCPA	DCPA	1.0	0.9995	0.016	0.999	0.000	0.001	0.999
			Conditional dominance					
RS	DCPA	0.5	0.5850	0.211	0.188	0.018	0.794	0.794
TCPA	RS	1.0	0.9970	0.039	0.994	0.000	0.006	0.994
TCPA	DCPA	1.0	0.9995	0.016	0.999	0.000	0.001	0.999
			General dominance					
RS	DCPA	1.0	0.9080	0.289	0.908	0.092	0.000	0.908
TCPA	RS	1.0	0.9980	0.045	0.998	0.002	0.000	0.998
TCPA	DCPA	1.0	1.0000	0.000	1.000	0.000	0.000	1.000

DA (B). The results in Table 13 show the following pattern: 1) time to CPA completely dominates distance to CPA, which in turn completely dominates relative speed; 2) time to CPA conditionally dominates distance to CPA, which in turn conditionally dominates relative speed; and 3) time to CPA generally dominates distance to CPA, which in turn generally dominates relative speed.

Table 13

Dominance analysis (B) for absolute time to CPA (TCPA) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.1269	.2640	.4007
RS	.1269		.2640	.2754
DCPA	.2640	.1269		.1402
TCPA	.4007	.0016	.0035	
k = 1 average		.0642	.1337	.2078
RS, DCPA	.3909			.0139
RS, TCPA	.4023		.0025	
DCPA, TCPA	.4042	.0006		
k = 2 average		.0006	.0025	.0139
RS, DCPA, TCPA	.4048			
Overall average		.0639	.1334	.2075

At the complete and conditional dominance levels, the reproducibility values for sample $D_{\textit{distance to CPA over relative speed}} = 1$, $D_{\textit{time to CPA over relative speed}} = 1$, and $D_{\textit{time to CPA over distance to CPA}} = 1$ are 47.8%, 63.0%, and 57.0%, respectively. At the general dominance level, the reproducibility values for sample $D_{\textit{distance to CPA over relative speed}} = 1$, $D_{\textit{time to CPA over relative speed}} = 1$, and $D_{\textit{time to CPA over distance to CPA}} = 1$ are 79.5%, 99.5%, and 89.1%, respectively.

Table 14

Inference results for dominance analysis (B) for absolute time to CPA (TCPA) estimate error: D_{ij} values in the sample (n = 48) and their means (D_{ij}), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap sample

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
				Complete dominance				
DCPA	RS	1	0.677	0.345	0.478	0.124	0.398	0.478
TCPA	RS	1	0.814	0.244	0.630	0.002	0.368	0.630
TCPA	DCPA	1	0.757	0.301	0.570	0.056	0.374	0.570
				Conditional dominance				
DCPA	RS	1	0.677	0.345	0.478	0.124	0.398	0.478
TCPA	RS	1	0.814	0.244	0.630	0.002	0.368	0.630
TCPA	DCPA	1	0.757	0.301	0.570	0.056	0.374	0.570
				General dominance				
DCPA	RS	1	0.795	0.404	0.795	0.205	0.000	0.795
TCPA	RS	1	0.995	0.071	0.995	0.005	0.000	0.995
TCPA	DCPA	1	0.891	0.312	0.891	0.109	0.000	0.891

DA for Signed Time to CPA Estimate Error

DA (A). The results in Table 15 show the following pattern: 1) time to CPA completely dominates distance to CPA and relative speed, but complete dominance cannot be established between relative speed and distance to CPA; 2) time to CPA conditionally dominates distance to CPA and relative speed, but conditional dominance cannot be established between relative speed and distance to CPA; and 3) time to CPA generally dominates relative speed, which in turn generally dominates distance to CPA.

At the complete and conditional dominance levels, the reproducibility values for sample $D_{\text{relative speed over distance to CPA}} = .5$, $D_{\text{time to CPA over relative speed}} = 1$, and $D_{\text{time to CPA over distance to CPA}} = 1$ are 90.8%, 86.8%, and 99.9%, respectively. At the general dominance level, sample $D_{\text{relative speed over distance to CPA}} = 1$, and its reproducibility value is now 90.8%; the reproducibility values for sample $D_{\text{time to CPA over relative speed}} = 1$, and $D_{\text{time to CPA over distance to CPA}} = 1$ are 95.1% and 100%, respectively.

Table 15

Dominance analysis (A) for signed time to CPA (TCPA) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.4281	.0555	.5344
RS	.4281		.0555	.1401
DCPA	.0555	.4281		.5508
TCPA	.5344	.0338	.0719	
k = 1 average		.2310	.0637	.3454
RS, DCPA	.4836			.1256
RS, TCPA	.5682		.0410	
DCPA, TCPA	.6063	.0029		
k = 2 average		.0029	.0410	.1256
RS, DCPA, TCPA	.6092			
Overall average		.2207	.0534	.3351

Table 16

Inference results for dominance analysis (A) for signed time to CPA (TCPA) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap samples

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
				Complete dominance				
RS	DCPA	0.5	0.5420	0.146	0.088	0.004	0.908	0.908
TCPA	RS	1.0	0.9285	0.190	0.868	0.011	0.121	0.868
TCPA	DCPA	1.0	0.9995	0.016	0.999	0.000	0.001	0.999
				Conditional dominance				
RS	DCPA	0.5	0.5420	0.146	0.088	0.004	0.908	0.908
TCPA	RS	1.0	0.9285	0.190	0.868	0.011	0.121	0.868
TCPA	DCPA	1.0	0.9995	0.016	0.999	0.000	0.001	0.999
				General dominance				
RS	DCPA	1.0	0.9940	0.077	0.994	0.006	0.000	0.994
TCPA	RS	1.0	0.9510	0.216	0.951	0.049	0.000	0.951
TCPA	DCPA	1.0	1.0000	0.000	1.000	0.000	0.000	1.000

DA (B). The results in Table 17 show the following pattern: 1) time to CPA completely dominates distance to CPA, but complete dominance cannot be established between relative speed and distance to CPA, and between relative speed and time to CPA; 2) time to CPA completely dominates distance to CPA, but complete dominance cannot be established between relative speed and distance to CPA, and between relative speed and time to CPA; and 3) time to CPA generally dominates relative speed, which in turn dominates distance to CPA.

Table 17

Dominance analysis (B) for signed time to CPA (TCPA) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and k = 0 average	0	.4186	.0232	.3656
RS	.4186		.0232	.0695
DCPA	.0232	.4186		.5500
TCPA	.3656	.1225	.2075	
k = 1 average		.2706	.1153	.3098
RS, DCPA	.4418			.1424
RS, TCPA	.4881		.0960	
DCPA, TCPA	.5731	.0110		
k = 2 average		.0110	.0960	.1424
RS, DCPA, TCPA	.5841			
Overall average		.2334	.0782	.2726

At the complete and conditional dominance levels, the reproducibility values for sample $D_{\textit{relative speed over distance to CPA}} = .5$, $D_{\textit{time to CPA over relative speed}} = .5$, and $D_{\textit{time to CPA over distance to CPA}} = 1$ are 97.6%, 66.2%, and 95.7%, respectively. At the general dominance level, sample $D_{\textit{relative speed over distance to CPA}} = 1$, $D_{\textit{time to CPA over relative speed}} = 1$ and their reproducibility values are now 99.4% and 67.9%; the reproducibility value $D_{\textit{time to CPA over distance to CPA}} = 1$ are 99.7%.

Table 18

Inference results for dominance analysis (B) for signed time to CPA (TCPA) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap samples

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
Complete dominance								
RS	DCPA	0.5	0.508	0.077	0.020	0.004	0.976	0.976
TCPA	RS	0.5	0.663	0.241	0.332	0.006	0.662	0.662
TCPA	DCPA	1.0	0.978	0.105	0.957	0.001	0.042	0.957
Conditional dominance								
RS	DCPA	0.5	0.508	0.077	0.020	0.004	0.976	0.976
TCPA	RS	0.5	0.663	0.241	0.332	0.006	0.662	0.662
TCPA	DCPA	1.0	0.978	0.105	0.957	0.001	0.042	0.957
General dominance								
RS	DCPA	1.0	0.994	0.077	0.994	0.006	0.000	0.994
TCPA	RS	1.0	0.679	0.467	0.679	0.321	0.000	0.679
TCPA	DCPA	1.0	0.997	0.055	0.997	0.003	0.000	0.997

DA for Absolute Orientation at CPA Estimate Error

DA (A). The results in Table 19 show the following pattern: 1) distance to CPA completely dominates time to CPA, which in turn completely dominates relative speed; 2) distance to CPA conditionally dominates time to CPA, which in turn conditionally dominates relative speed; and 3) distance to CPA generally dominates time to CPA, which in turn generally dominates relative speed.

Table 19

Dominance analysis (A) for absolute orientation at CPA (OCPA) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.0372	.2583	.2343
RS	.0372		.2583	.2478
DCPA	.2583	.0372		.0478
TCPA	.2343	.0506	.0718	
k = 1 average		.0439	.1650	.1478
RS, DCPA	.2955			.0109
RS, TCPA	.2850		.0214	
DCPA, TCPA	.3061	.0002		
k = 2 average		.0002	.0214	.0109
RS, DCPA, TCPA	.3064			
Overall average		.0271	.1482	.1310

Table 20

Inference results for dominance analysis (A) for absolute orientation at CPA (OCPA) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap samples

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
Complete dominance								
DCPA	RS	1	0.8500	0.237	0.707	0.007	0.286	0.707
DCPA	TCPA	1	0.5885	0.442	0.495	0.318	0.187	0.495
TCPA	RS	1	0.8425	0.233	0.686	0.001	0.313	0.686
Conditional dominance								
DCPA	RS	1	0.8500	0.237	0.707	0.007	0.286	0.707
DCPA	TCPA	1	0.5885	0.442	0.495	0.318	0.187	0.495
TCPA	RS	1	0.8425	0.233	0.686	0.001	0.313	0.686
General dominance								
DCPA	RS	1	0.9840	0.126	0.984	0.016	0.000	0.984
DCPA	TCPA	1	0.5890	0.492	0.589	0.411	0.000	0.589
TCPA	RS	1	0.9960	0.063	0.996	0.004	0.000	0.996

At the complete and conditional dominance levels, the reproducibility values for sample $D_{\textit{distance to CPA over relative speed}} = 1$, $D_{\textit{distance to CPA over time to CPA}} = 1$, and $D_{\textit{time to CPA over relative speed}} = 1$ are 70.7%, 49.5%, and 68.6%, respectively. At the general dominance level, the reproducibility values for sample $D_{\textit{distance to CPA over relative speed}} = 1$, $D_{\textit{distance to CPA over time to CPA}} = 1$, and $D_{\textit{time to CPA over relative speed}} = 1$ are 98.4%, 58.9%, and 99.6%, respectively.

DA (B). The results in Table 21 show the following pattern: 1) time to CPA completely dominates distance to CPA, which in turn completely dominates relative speed; 2) time to CPA conditionally dominates distance to CPA, which in turn conditionally dominates relative speed; and 3) time to CPA generally dominates distance to CPA, which in turn generally dominates relative speed.

Table 21

Dominance analysis (B) for absolute orientation at CPA (OCPA) estimate error with relative speed (RS), distance to CPA (DCPA), and time to CPA (TCPA) as predictors

		Additional contribution of:		
Subset model	R^2	RS	DCPA	TCPA
Null and *k* = 0 average	0	.0423	.2311	.2669
RS			.2311	.2441
DCPA		.0423		.0557
TCPA		.0196	.0199	
k = 1 average		.0310	.1255	.1499
RS, DCPA	.2735			.0144
RS, TCPA	.2865		.0014	
DCPA, TCPA	.2868	.0011		
k = 2 average		.0011	.0014	.0144
RS, DCPA, TCPA	.2879			
Overall average		.0248	.1194	.1437

At the complete and conditional dominance levels, the reproducibility values for sample $D_{\textit{distance to CPA over relative speed}} = 1$, $D_{\textit{time to CPA over relative speed}} = 1$, and $D_{\textit{time to CPA over distance to CPA}} = 1$ are 50.3%, 79.4%, and 48.5%, respectively. At the general dominance level, the reproducibility values for sample $D_{\textit{distance to CPA over relative speed}} = 1$, $D_{\textit{time to CPA over relative speed}} = 1$, and $D_{\textit{time to CPA over distance to CPA}} = 1$ are 96.1%, 99.8%, and 64.9%, respectively.

Table 22

Inference results for dominance analysis (B) for absolute orientation at CPA (OCPA) estimate error: Dij values in the sample (n = 48) and their means (Dij), standard errors, probabilities, and reproducibility over S = 1,000 bootstrap samples

i	j	Sample *Dij*	Average Dij	SE(Dij)	Pij	Pji	Pnoij	Reproducibility
			Complete dominance					
DCPA	RS	1	0.7380	0.276	0.503	0.027	0.470	0.503
TCPA	RS	1	0.8965	0.204	0.794	0.001	0.205	0.794
TCPA	DCPA	1	0.6380	0.393	0.485	0.209	0.306	0.485
			Conditional dominance					
DCPA	RS	1	0.7380	0.276	0.503	0.027	0.470	0.503
TCPA	RS	1	0.8965	0.204	0.794	0.001	0.205	0.794
TCPA	DCPA	1	0.6380	0.393	0.485	0.209	0.306	0.485
			General dominance					
DCPA	RS	1	0.9610	0.194	0.961	0.039	0.000	0.961
TCPA	RS	1	0.9980	0.045	0.998	0.002	0.000	0.998
TCPA	DCPA	1	0.6490	0.478	0.649	0.351	0.000	0.649

Discussion

Summarizing all the DA results reported above, an obvious pattern we found is that the relative importance of predictor variables for the complete and conditional dominance are identical for each dependent variable for both DA (A) and (B); and the DA results at the general dominance level show a similar pattern to that at the complete and conditional dominance levels while reducing the number of indeterminacy.

For absolute miss distance estimate error, across both DA (A) (Tables 3 and 4) and (B) (Tables 5 and 6), distance to CPA emerged to be the most important variable in influencing pilot's estimate error of this most important measure of conflict detection, followed by time to CPA, with relative speed being the least important. However, the relative importance between distance to CPA and time to CPA is less clear in DA (B) than in DA (A).

For signed miss distance estimate error, the relative importance between distance to CPA and time to CPA is ambiguous in that DA (A) (Tables 7 and 8) reveals that time to CPA is more important than distance CPA, but DA (B) (Tables 9 and 10) suggests a general tendency of distance to CPA's dominance over time to CPA. However, in both DA (A) and (B), both time to CPA and distance to CPA are for the most part more important than relative speed in accounting for the variance in signed miss distance estimate error.

For absolute time to CPA estimate error, it is clearly the case that among the three predictor variables, time to CPA is the most dominant or important in influencing performance, which is true across both DA (A) (Tables 11 and 12) and (B) (Tables 13 and 14). However, the relative importance between the other two variables, distance to CPA and relative speed, is less clear, because the results of the two DAs revealed opposite patterns—dominance of relative speed over distance to CPA in DA (A) and the reverse in DA (B).

For signed time to CPA estimate error, across both DA (A) (Tables 15 and 16) and (B) (Tables 17 and 18), time to CPA in large is the most important factor in influencing its variance. The relative importance between distance to CPA and relative speed is somewhat less clear, because in both DA (A) and (B), although at the general dominance level, relative speed is more important than distance to CPA, their relative importance cannot be established at the complete and conditional dominance levels. This result is different from the distance-over-speed bias reported in Xu et al. (2004), which suggests the dominance of distance to CPA over relative speed. This discrepancy may be due to the fact that the analysis for the distance-over-speed bias considered only a subset of data comparing the pairs of conditions with equal times to CPA, whereas the DA method included the data in all the conditions.

Finally, for absolute orientation at CPA estimate error, the relative importance between distance to CPA and time to CPA revealed by DA (A) (Tables 19 and 20) is the opposite of that by DA (B) (Tables 21 and 22), so no clear conclusion can be drawn. However, in both DA (A) and (B), relative speed is the least important factor compared to the other two factors.

Thus, the following conclusions can be drawn with reasonable confidence regarding the relative importance of the three independent or predictor variables in predicting or influencing pilots' conflict detection performance: 1) for absolute miss distance estimate error, distance to CPA is the most important variable, followed by time to CPA, which is more important than relative speed; 2) for signed miss distance estimate error, time to CPA and distance to CPA are more important than relative speed (i.e., relative speed is the least important), 3) for both absolute and signed time to CPA estimate errors, time to CPA is the most important, and 4) for absolute orientation at CPA estimate error, relative speed is the least important variable compared to distance to CPA and time to CPA in predicting this aspect of pilots' estimation performance.

While all the three variables had some effects (Xu et al., 2004), the data suggest that they contributed differently to the measures of conflict understanding (estimation accuracy of time to CPA and orientation and miss distance at CPA). As the DA results revealed, distance to CPA was a more important factor than time to CPA and relative speed in influencing absolute miss distance estimation error. Since miss distance was derived from the location of the CPA, whatever influenced the estimate accuracy of CPA location would also influence the miss distance estimate accuracy. A conceivable method for the pilot participant to estimate the CPA location was to estimate the interception point between the future trajectory of the intruder icon and a line connecting this trajectory and the ownship icon forming a right angle (90°), both of which themselves needed to be extrapolated (see Figure 1). The estimate error of CPA location (and therefore miss distance estimate error) was subject to the angular extrapolation errors of a) the line linking the intruder's extrapolated trajectory and the ownship icon at 90°, and b) the intruder's extrapolated trajectory. The latter angular error was presumably amplified at greater distance to CPA. Although relative speed (perceived while the intruder icon was visible) and time to CPA might also have affected the stability of this angular estimate, they were apparently not as influential as distance to CPA. By the similar mechanism, distance to CPA was more influential than relative speed when it comes to the estimation of orientation at CPA, because orientation at CPA was also derived from the CPA location. In contrast, the effect of the true time to CPA was more pronounced than that of distance to CPA and relative speed on time to CPA estimation. This is consistent with Peterken et al. (1991), where it was shown that it was the time rather than the distance during which a moving object was not visible that was the more important factor in influencing the time-to-contact estimation accuracy.

Conclusions

This report has described the application of a method known as dominance analysis or DA (Budescu, 1993; Azen & Budescu, 2003) to analyze the relative importance of three inter-correlated independent variables (distance to CPA, time to CPA, and relative speed) in influencing the performance measures of pilot conflict detection using a CDTI. ANOVAs and the traditional multiple regression methods in general do a poor job in tearing apart the relative contributions of independent variables when they are tangled with each other. The DA method seems to offer a viable and better alternative. Although the ANOVA results in Xu et al. (2004) showed that distance to CPA, time to CPA, and relative speed all had effects on the performance measures to a more or less degree, the DA results illustrated that distance to CPA and time to CPA were the most important variable in influencing the estimation accuracy of miss distance at CPA and time to CPA, respectively; and relative speed is the least important than the other two variables in influencing the estimation accuracy of orientation at CPA. Why relative importance varied among different performance measures may be due to the different methods or strategies participants used when different aspects of conflict detection were involved.

Acknowledgement

We would like to thank Dr. David Budescu of the University of Illinois at Urbana-Champaign and Dr. Razia Azen of the University of Wisconsin--Milwaukee for offering their generous and invaluable assistance with the relative importance analysis for this study.

References

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods, 8*(2), 129-148.

Behson, S. J. (2002). Which dominates? The relative importance of work–family organizational support and general organizational context on employee outcomes. *Journal of Vocational Behavior, 61*(1), 53–72.

Block, J. (1995). On the relation between IQ, impulsivity, and delinquency: Remarks on the Lynam, Moffitt, and Stouthamer-Loeber (1993) interpretation. *Journal of Abnormal Psychology, 104*(2), 395–398.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114*(3), 542–551.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement, 61*(2), 229–248.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*, 1–26.

Fox, J. (2002). *An R and S-PLUS companion to applied regression*. Thousand Oaks, CA: Sage.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35*(1), 1-19.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference* (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-095). Newbury Park, CA: Sage.

Nickerson, C., Schwartz, N., Diener, E., & Kahneman, D. (2003). Zeroing on the dark side of the American Dream: A closer look at the negative consequences of the goal for financial success. *Psychological Science, 14*(6), 531-536.

Peterken, C., Brown, B., & Bowman, K. (1991). Predicting the future position of a moving target. *Perception, 20*, 5-16.

Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology, 74*(2), 482–493.

Weinberger, T. E. (1995). Determining the relative importance of compensable factors: The application of dominance analysis to job evaluation. *Compensations & Benefits Management, 11*(1), 17–23.

Xu, X., Rantanen, E. M., & Wickens, C. D. (2004). *Estimation of conflict risk using cockpit displays of traffic information* (AHFD-04-11/FAA-04-4). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Appendix

1) Data for DA (A) and SAS codes to record data and retrieve SAS macro, where ‘rs,’ ‘dcpa,’ ‘tcpa,’ ‘a_tcp,’ ‘s_tcp,’ ‘a_dcp,’ ‘s_dcp,’ and ‘a_ocp’ represent relative speed, distance to CPA, time to CPA, absolute and signed time to CPA estimate errors, absolute and signed miss distance at CPA estimate errors, and absolute orientation at CPA estimate error, respectively.

```
data Xu;
input Obs subject      rs     dcpa   tcpa   a_tcp  s_tcp  a_dcp  s_dcp  a_ocp;
datalines;
    1     1    160  2.67   60   18.14  -17.65   0.34   0.03  10.47
    2     1    240  2.67   40    8.67   -3.32   0.33   0.07  12.05
    3     1    480  2.67   20    6.29    3.04   0.28  -0.01   9.56
    4     2    160  1.33   30   11.83  -11.05   0.21  -0.02   9.91
    5     2    240  1.33   20    6.27   -2.91   0.18   0.02   9.34
    6     2    480  1.33   10    3.76    1.63   0.18  -0.02   7.15
    7     4    160  2.67   60   19.04  -15.15   0.42  -0.18  14.46
    8     4    240  2.67   40   11.14   -4.03   0.42  -0.19  14.63
    9     4    480  2.67   20    6.16    4.38   0.36   0.06  15.99
   10     5    160  1.33   30   13.71    4.71   0.13  -0.02  15.24
   11     5    240  1.33   20   15.95   13.73   0.14  -0.03   8.43
   12     5    480  1.33   10   18.40   17.75   0.12  -0.03   9.63
   13     7    160  1.33   30    6.96    0.07   0.19  -0.08   8.50
   14     7    240  1.33   20    5.20    3.35   0.15  -0.05   4.57
   15     7    480  1.33   10    2.22    0.15   0.12  -0.01   4.57
   16     9    160  2.67   60   20.29  -20.29   0.30  -0.01  17.53
   17     9    240  2.67   40   10.80   -9.97   0.26  -0.03   7.71
   18     9    480  2.67   20    2.92   -1.65   0.19  -0.03   6.12
   19    10    160  2.67   60   16.64  -13.02   0.28   0.10  14.69
   20    10    240  2.67   40    9.82   -0.21   0.35   0.16  12.40
   21    10    480  2.67   20   10.68    8.49   0.17   0.05   7.64
   22    12    160  1.33   30    9.59   -8.90   0.10   0.04   6.81
   23    12    240  1.33   20    3.45   -2.23   0.12  -0.01   7.20
   24    12    480  1.33   10    4.86    4.09   0.11  -0.03   4.23
   25    14    160  2.67   60   17.84  -12.70   0.39   0.02  18.18
   26    14    240  2.67   40   11.52   -4.45   0.44   0.03  19.25
   27    14    480  2.67   20    7.86    4.38   0.22  -0.08   8.95
   28    15    160  1.33   30   17.01  -17.01   0.30  -0.13  12.08
   29    15    240  1.33   20   10.58   -9.99   0.20  -0.09  14.64
   30    15    480  1.33   10    5.26   -4.85   0.26  -0.15  14.64
   31    17    160  2.67   60   24.51  -19.47   0.51   0.02  18.65
   32    17    240  2.67   40   12.53   -9.10   0.53  -0.12  25.44
   33    17    480  2.67   20    9.80    7.39   0.35   0.03  13.76
   34    18    160  1.33   30   10.55   -8.66   0.21  -0.10   8.34
   35    18    240  1.33   20    5.44   -4.04   0.11  -0.01   6.83
   36    18    480  1.33   10    2.31    0.49   0.10  -0.03   6.82
   37    19    160  1.33   30   13.53  -12.91   0.14  -0.05   9.47
   38    19    240  1.33   20    7.87   -0.13   0.18  -0.08   9.60
   39    19    480  1.33   10    5.11    3.07   0.15  -0.08  10.79
   40    21    160  2.67   60   36.76  -35.69   1.05  -0.78  41.20
   41    21    240  2.67   40   19.64  -18.29   0.81  -0.46  35.47
   42    21    480  2.67   20    9.08    0.43   0.68  -0.33  34.79
   43    22    160  1.33   30   10.82  -10.55   0.17  -0.06   7.90
   44    22    240  1.33   20    5.67   -4.02   0.19  -0.11   7.96
   45    22    480  1.33   10    4.70    0.73   0.18  -0.10   8.39
   46    24    160  2.67   60   15.59  -12.30   0.44  -0.32  17.95
   47    24    240  2.67   40    9.62    1.95   0.37  -0.24  13.90
   48    24    480  2.67   20    9.65    8.92   0.20  -0.06  10.16

;
run;

%include 'D:\Xu\bootstrap1.sas';
%dom;

%include 'D:\Xu\bootstrap2.sas';
```

```
%dom;

%include 'D:\Xu\bootstrap3.sas';
%dom;

%include 'D:\Xu\bootstrap4.sas';
%dom;

%include 'D:\Xu\bootstrap5.sas';
%dom;
```

2) Data for DA (B) and SAS codes to record data and retrieve SAS macro, where ‘rs,’ ‘dcpa,’ ‘tcpa,’ ‘a_tcp,’ ‘s_tcp,’ ‘a_dcp,’ ‘s_dcp,’ and ‘a_ocp’ represent relative speed, distance to CPA, time to CPA, absolute and signed time to CPA estimate errors, absolute and signed miss distance at CPA estimate errors, and absolute orientation at CPA estimate error, respectively.

```
data Xu;
input Obs subject      rs     dcpa   tcpa   a_tcp  s_tcp  a_dcp s_dcp   a_ocp;
datalines;
    1     1    240  2.67   40    8.67   -3.32   0.33   0.07  12.05
    2     1    480  2.67   20    6.29    3.04   0.28  -0.01   9.56
    3     2    240  1.33   20    6.27   -2.91   0.18   0.02   9.34
    4     2    480  1.33   10    3.76    1.63   0.18  -0.02   7.15
    5     3    240  4.00   60   13.32   -6.92   0.85  -0.68  13.74
    6     3    480  4.00   30   10.89    9.43   0.56  -0.42   9.58
    7     4    240  2.67   40   11.14   -4.03   0.42  -0.19  14.63
    8     4    480  2.67   20    6.16    4.38   0.36   0.06  15.99
    9     5    240  1.33   20   15.95   13.73   0.14  -0.03   8.43
   10     5    480  1.33   10   18.40   17.75   0.12  -0.03   9.63
   11     6    240  4.00   60   11.42   -2.55   0.72  -0.42  24.32
   12     6    480  4.00   30    5.88   -1.00   0.45  -0.19  17.53
   13     7    240  1.33   20    5.20    3.35   0.15  -0.05   4.57
   14     7    480  1.33   10    2.22    0.15   0.12  -0.01   4.57
   15     8    240  4.00   60   19.36  -19.03   0.70  -0.14  20.62
   16     8    480  4.00   30    7.81    5.42   0.60  -0.22  18.02
   17     9    240  2.67   40   10.80   -9.97   0.26  -0.03   7.71
   18     9    480  2.67   20    2.92   -1.65   0.19  -0.03   6.12
   19    10    240  2.67   40    9.82   -0.21   0.35   0.16  12.40
   20    10    480  2.67   20   10.68    8.49   0.17   0.05   7.64
   21    11    240  4.00   60   12.15   -9.38   0.56  -0.20  10.36
   22    11    480  4.00   30   12.16    9.89   0.46  -0.26  13.55
   23    12    240  1.33   20    3.45   -2.23   0.12  -0.01   7.20
   24    12    480  1.33   10    4.86    4.09   0.11  -0.03   4.23
   25    13    240  4.00   60   16.38  -15.86   0.50  -0.28  18.02
   26    13    480  4.00   30    9.11    6.42   0.30  -0.12   7.15
   27    14    240  2.67   40   11.52   -4.45   0.44   0.03  19.25
   28    14    480  2.67   20    7.86    4.38   0.22  -0.08   8.95
   29    15    240  1.33   20   10.58   -9.99   0.20  -0.09  14.64
   30    15    480  1.33   10    5.26   -4.85   0.26  -0.15  14.64
   31    16    240  4.00   60   11.92   -9.94   0.60  -0.36  31.45
   32    16    480  4.00   30   15.16   13.89   0.62  -0.29  18.55
   33    17    240  2.67   40   12.53   -9.10   0.53  -0.12  25.44
   34    17    480  2.67   20    9.80    7.39   0.35   0.03  13.76
   35    18    240  1.33   20    5.44   -4.04   0.11  -0.01   6.83
   36    18    480  1.33   10    2.31    0.49   0.10  -0.03   6.82
   37    19    240  1.33   20    7.87   -0.13   0.18  -0.08   9.60
   38    19    480  1.33   10    5.11    3.07   0.15  -0.08  10.79
   39    20    240  4.00   60   18.17  -18.00   0.45  -0.08  14.87
   40    20    480  4.00   30    4.65    1.30   0.36  -0.22  10.51
   41    21    240  2.67   40   19.64  -18.29   0.81  -0.46  35.47
   42    21    480  2.67   20    9.08    0.43   0.68  -0.33  34.79
   43    22    240  1.33   20    5.67   -4.02   0.19  -0.11   7.96
   44    22    480  1.33   10    4.70    0.73   0.18  -0.10   8.39
   45    23    240  4.00   60   16.21  -14.76   0.66  -0.18  23.91
   46    23    480  4.00   30   17.59   16.08   0.58  -0.02  24.30
   47    24    240  2.67   40    9.62    1.95   0.37  -0.24  13.90
   48    24    480  2.67   20    9.65    8.92   0.20  -0.06  10.16

;
run;

%include 'D:\Xu\bootstrap1.sas';
```

```
%dom;

%include 'D:\Xu\bootstrap2.sas';
%dom;

%include 'D:\Xu\bootstrap3.sas';
%dom;

%include 'D:\Xu\bootstrap4.sas';
%dom;

%include 'D:\Xu\bootstrap5.sas';
%dom;
```

3) Dominance probability macro. The SAS codes given below are for absolute miss distance at CPA estimate error. For signed miss distance at CPA estimate error, absolute and signed time to CPA estimated errors, and absolute orientation at CPA estimate error, ‘dep=a_dcp’ in the code line highlighted in red below was replaced by ‘dep=s_dcp,’ ‘dep=a_tcp,’ ‘dep=s_tcp,’ and ‘dep=a_ocp,’ respectively.

```
/***************************************************************************
                    DOMINANCE PROBABILITY MACRO

This macro executes dominance analysis as described by Razia Azen & David Budescu
in Psychological Methods, 2003.  Dominance analysis quantifies the importance of
each predictor as its average increment to the model r-squared, across all possible
submodel sizes.
This macro was written by Razia Azen with valuable contributions from Robert
Ceurvorst.

NOTE:  This program is limited to at most 10 predictors!

-------------
INSTRUCTIONS:
-------------
1. In this macro, change the %macro dom line (which appears just below these
instructions) according to the following directions:
%macro dom (data=_last_, dep=Y, indep='list of predictors', p=n_of_predictors,
      noprint=0, short=0, B=n_of_bootstraps, predtype='type_of_predictors',
     seed=random_number);

  If defaults are used, then only p= OR indep= is necessary, e.g., '%dom(p=4)'
  will use the last data set created and operate on variables Y and X1-X4.

Either p= OR indep= is required -- not both.  If both are specified, p is
determined by counting the variables in the indep= list.

data=      SAS dataset to be used.  Default is last dataset created.

dep=       Name of dependent variable.  Default is Y.

indep=     List of predictors in quotes.
 OR
p=         No. of predictors IF and ONLY IF they are named X1-Xp,
           in which case indep= is not required.

noprint=1  Suppresses printing of the input dataset.

short=1    Suppresses listing of each predictor's contributions to individual
           submodels.

B=           The number of bootstrap runs.  Default is 1000.

predtype= 'f' if the predictors are fixed;
          'r' if the predictors are random.  Default is 'r'.

seed=     Six digits.  Default is 0 (uses clock value).

For example, with 4 predictors named X1, X2, X3, X4 and all default values:
(data=_last_, dep=Y, p=4, indep=, noprint=0, short=0, B=1000, predtype='r',
seed=0);
Or, with 4 predictors named A, B, C, D and all default values:
```

```
   (data=_last_, dep=Y, p=, indep='A B C D', noprint=0, short=0, B=100, predtype='r',
seed=0);

   2. Save this macro, and add the following two lines (below) to the
   SAS program in which you've read the data set to be analyzed:

   %include 'k:\sasmacro\uniDAmacro.sas';  *** CHANGE TO PATH WHERE MACRO IS SAVED
***;
   %dom;

   3.  Run the program edited in step 2.

   -- END OF INSTRUCTIONS --
   ***************************************************************************/
   option nosource;
   /********** CHANGE ONLY THE LINE BELOW (See step 1 above)!!! **************/
   %macro dom (data=Xu, dep=a_dcp, p=3, indep='rs    dcpa   tcpa ', noprint=0,
short=0, B=1000, predtype='r', seed=0);
   /************************************************************************/

   %if &indep= and &p= %then %do;
     %put YOU MUST SPECIFY INDEP=list of predictors OR P=no. of predictors (IF THEY
ARE NAMED X1-Xp).;
     %goto done;
   %end;

   %else %if &indep ne %then %do;
     %if %index(&indep,%str(%'))>0 or %index(&indep,%str(%’))>0
         %then %let indep = %substr(&indep,2,%length(&indep)-2);
     %if %index(&indep,-) > 0 %then %expand;
     %let p=1;
     %do %while(%length(%scan(&indep,&p))>0);
         %let x&p = %scan(&indep,&p);
         %let p = %eval(&p+1);
     %end;
     %let p = %eval(&p-1);
   %end;

   %else %if &p > 0 %then %do;
     %let indep=;
     %do i=1 %to &p;
         %let indep=&indep x&i;
         %let x&i=x&i;
     %end;
   %end;

   %if &p>10 %then %do;
     %put  THE MAXIMUM NUMBER OF PREDICTORS IS 10.  YOU HAVE &P..;
     %goto done;
   %end;

   %LET DATANAME=&SYSLAST;
   option nonotes;
   DATA _NULL_;  FILE PRINT LINESLEFT=WITH;  CALL SYMPUT('WITH',WITH);
   DATA _NULL_;
     FILE PRINT NOTITLES LINESLEFT=WITHOUT;
     CALL SYMPUT ('MTITLE',TRIM(LEFT(WITHOUT-&WITH+2)));
     CALL SYMPUT('NMODELS', 2**&p -1);
   RUN;

   data original; set &data;

   %if &noprint=0 %then %do;
   proc print;
   title&mtitle 'Input data set (check to make sure it is correct)'; run;
   %end;

   data labels; set;
     keep &indep;
     length dvlabel $40;
     call label(&dep,dvlabel);
     call symput('dvl',trim(dvlabel));
```

```
  if _n_=1 then stop;
run;


/*********************************************************
     PART 1:  regression/DA of the original data
*********************************************************/

title 'regression results for original data set';
option notes;
proc reg corr data=original outest=onereg (keep=_in_ _rsq_ &indep);
  model &dep=&indep / stb pcorr2 scorr2 ;
  model &dep=&indep / selection=adjrsq best=&nmodels %if &short=1 %then noprint;;
run; quit; option nonotes;


* order all subset models in lexicographical order;
data modelmat;
  set onereg;
  %if %substr(&sysver,1,1)>7 %then if _n_>1;;
  %do i=1 %to &p;
    &&x&i=(&&x&i>.);
  %end;
  if _IN_=. then delete;
  keep &indep _IN_ _RSQ_;
proc sort; by _IN_ %do i=1 %to &p; descending &&x&i %end;; run;
proc print; title 'R-squared values for all subset models'; run;

 /*** %if &short=0 %then %do;
   proc print; id _IN_ _RSQ_; format _RSQ_ 5.4; run;
 %end;  ***/


/**********************************************
          DOMINANCE ANALYSIS TABLE
**********************************************/
option notes; title;
proc IML;
 reset noprint;
 start;
 * read the subset models matrix into DOM;
 use modelmat;
  read all into DOM;
 close modelmat;

* dom contains, for each subset model, 1/0 values for the predictors,
  the number of predictors in (_IN_), and the model r2 (_RSQ_);

   p=ncol(dom)-2;
   ncol=ncol(dom);

* dom is rearranged to contain, for each subset model, the number of predictors
  in and the r2 of the model, followed by 1/0 values for predictors in/out;
  dom1=dom[,1:p];  dom2=dom[,(p+1):ncol];  dom=dom2||dom1;  free dom1 dom2;

* generate table of additional contributions;
   null=J(1,ncol,0);
   dom=null//dom;
   nrow=nrow(dom);
* dom now contains a top row of zeros for the null model;
   full=J(1,p,99);  fullrsq=J(1,1,99);
   reduced=J(1,p,99);  redrsq=J(1,1,99);
   contrib=J(nrow,p,0);  * additional contributions matrix;

    do i=1 to nrow;      * for each model;
     do j=1 to p;        * for each predictor;

       if dom(|i,j+2|)=0 then do; * if predictor is not in subset model;
        reduced=dom[i,3:ncol];    * the 1/0 row represents the reduced model;
        full=reduced;             * the full model is same as reduced model;

         do k=1 to p;
```

```
          if k=j then full(|1,k|)=1;   * add the jth predictor to the full model;
         end;

         do r=1 to nrow;               * for each model;
          comp=dom[r,3:ncol];          * comp is the 1/0 row of dom;
          if comp=full then fullrsq=dom[r,2];    * r2 of row is fullrsq, or;
          if comp=reduced then redrsq=dom[r,2];  * r2 of row is redrsq;
         end;

         contrib(|i,j|)=fullrsq-redrsq;          * contrib is r2 difference;
       end;

       else do;
        contrib(|i,j|)=.;         * if predictor is in model, contib is .;
       end;

     end;
    end;

   contrib=dom||contrib;
   contrib=contrib[1:nrow-1,];

  cols = {IN RSQ %do i=1 %to &p; &&X&i %end; %do i=1 %to &p; CP&i %end;};

  create rsqtable from contrib[colname=cols];
   append from contrib;
  close rsqtable;

finish;
run; quit;

option nonotes;

%if &short=0 %then %do;
 proc print data=rsqtable;  id IN RSQ;  format RSQ CP1-CP&p 5.4;
 title&mtitle 'Dominance Table:  Additional Contributions of Predictors Across All
Subset Regression Models';
 %UNQUOTE(TITLE%EVAL(&MTITLE+1)) 'CPi indicates the additional contribution of
predictor i to the model r-square';
%end;

proc summary nway; var CP1--CP&p; class in;
 output out=avgcont (drop=_type_) mean=&indep;
run;
proc means noprint; var &indep;
 output out=meanc mean=;
run;
data avgcont; set labels meanc (in=y) avgcont;
  if y then in=999;
proc print double;  id IN;  var &indep;
  format &indep 5.4  IN %if &p<11 %then 1.0; %else 2.0;;
title&mtitle 'Dominance Analysis:  Overall Average Contributions of Predictors
(First Row)';
%UNQUOTE(TITLE%EVAL(&MTITLE+1)) 'And Average Contributions to Models of Each Size
(Remaining Rows)'; run;

proc transpose prefix=size out=meanc (rename=(SIZE999=OVERALL _NAME_=VAR)); id IN;
var &indep; run;
proc sort; by descending overall;
title&mtitle 'Dominance Analysis:  Average Predictor Contributions Overall and to
Models of Each Size';
proc print; id _character_; format _numeric_ 5.4; run;


TITLE&MTITLE;
%DONE: option notes _last_=&syslast;
run;


/***********************************************
          DOMINANCE MATRICES (Sample)
***********************************************/
```

```
Proc IML;
start;
use modelmat;
 read all into damat;
close modelmat;
Dcsample=J(&p,&p,0); Dasample=J(&p,&p,0); Dgsample=J(&p,&p,0);
RSQ=damat[,ncol(damat)];

/*****************************************************/
/*****           Complete dominance                ****/
/*****************************************************/
do i=1 to &p-1;
 do j=i+1 to &p;

 * define matrix of constants ('comp') to determine complete dominance between
 each pair of predictors;

        comp=J(2**(&p-2),2**&p-1,0);

        * Xh is any subset model that does not include i and j;
        * Xh contains the columns of damat that do not involve i or j;
        Xh=J(nrow(damat),&p-2,99);
          Xhcol=0;                   * index column number in Xh;
          do h=1 to &p;
           if (i ^= h & j ^= h) then do;    * find non i,j columns;
            Xhcol=Xhcol+1;                        * update column number in Xh;
            Xh[,Xhcol] = damat[,h];         * assign column to Xh;
           end;
          end;

          * contrast rows (subsets) representing XiXh and XjXh;
        comprow=1;
        do r=1 to 2**&p-2;                * for each pair of rows;
         do s=r+1 to 2**&p-1;
          if Xh[r,]=Xh[s,] then do;      * if the rows of Xh are the same;
                                         * and if i and j are contrasted;
           if (damat[r,i]=1 & damat[r,j]=0 & damat[s,i]=0 &
                damat[s,j]=1) then do;
                 comp[comprow,r]=1;
                 comp[comprow,s]=-1;
                 comprow=comprow+1;
           end;
          end;

         end;   * do r loop;
        end;    * do s loop;


*** Determine the complete Dij value ***;

  cdiffij=comp*RSQ;

  zero=J(nrow(cdiffij), ncol(cdiffij),0);
  ijdom=nrow(cdiffij);

  * obtain complete dominance matrices;
     * undetermined case (all differences are zero);
     if cdiffij=zero then do;
        Dcsample[i,j]=0.5;
        Dcsample[j,i]=Dcsample[i,j];
     end;
     * else, check signs of difference elements;
     if cdiffij ^= zero then do;
      nonneg=0; nonpos=0;
      do k=1 to nrow(cdiffij);
       if cdiffij[k,]>=0 then nonneg=nonneg+1;
       if cdiffij[k,]<=0 then nonpos=nonpos+1;
      end;
      * dominance case;
      if nonneg=ijdom then Dcsample[i,j]=1; else Dcsample[i,j]=0;
      if nonpos=ijdom then Dcsample[j,i]=1; else Dcsample[j,i]=0;
```

```
         * undetermined case (differences have different signs);
         if (Dcsample[i,j]=0 & Dcsample[j,i]=0) then do;
           Dcsample[i,j]=0.5;
           Dcsample[j,i]=Dcsample[i,j];
         end;
        end;


/***************************************************/
/*****     Conditional dominance                ****/
/***************************************************/
* define matrix of constants ('avg') to determine average (within model
size) dominance between each pair of predictors;

subsets=damat;
* subsets is a matrix that consists of p+2 columns (and
the null model in the top row): 1/0 for x1 - xp, in, Rsq;

  avg=J(&p,2**&p-1,0);
  do c=1 to &p;         * for each model size;
   num=0;               * index number of subsets per size;
   do m=1 to 2**&p-1;   * for each model, determine number of models
                                   of size is c-1 to which the ith predictor
makes
                                   an additional contribution;
    if (subsets[m,&p+1]=c-1 & subsets[m,i]=0) then num=num+1;
   end;
   if num=0 then num=1; * for the case of the null model (size=0);
   do r=1 to 2**&p-1;    * determine subsets to contrast;
    * consider additional contributions for model size (c);
    if (subsets[r,&p+1]=c-1 | subsets[r,&p+1]=c) then do;
     * average contribution of models that inlcude i but not j;
     if (subsets[r,i]=1 & subsets[r,j]=0) then avg[c,r]=1/num;
     * average contribution of models that inlcude j but not i;
     if (subsets[r,i]=0 & subsets[r,j]=1) then avg[c,r]=-1/num;
    end;
   end;
  end;

*** Determine the average Dij value ***;

  adiffij=avg*RSQ;

  zero=J(nrow(adiffij), ncol(adiffij),0);
  ijdom=nrow(adiffij);

  * obtain average dominance matrices;
     * undetermined case (all differences are zero);
     if adiffij=zero then do;
        Dasample[i,j]=0.5;
        Dasample[j,i]=Dasample[i,j];
     end;
     * else, check signs of difference elements;
     if adiffij ^= zero then do;
      nonneg=0; nonpos=0;
      do k=1 to nrow(adiffij);
       if adiffij[k,]>=0 then nonneg=nonneg+1;
       if adiffij[k,]<=0 then nonpos=nonpos+1;
      end;
      * dominance case;
      if nonneg=ijdom then Dasample[i,j]=1; else Dasample[i,j]=0;
      if nonpos=ijdom then Dasample[j,i]=1; else Dasample[j,i]=0;
      * undetermined case (differences have different signs);
      if (Dasample[i,j]=0 & Dasample[j,i]=0) then do;
        Dasample[i,j]=0.5;
        Dasample[j,i]=Dasample[i,j];
      end;
     end;


/***************************************************/
/*****         General dominance                ****/
```

```
/**************************************************/
* define matrix of constants ('glob') to determine global (overall average)
dominance between each pair of predictors;

  glob=J(1,2**&p-1,99);
  do g=1 to 2**&p-1;
   glob[,g]=avg[+,g]/&p;
  end;

*** Determine the global Dij value ***;

  gdiffij=glob*RSQ;

  zero=J(nrow(gdiffij), ncol(gdiffij),0);
  ijdom=nrow(gdiffij);

  * obtain global dominance matrices;
     * undetermined case (all differences are zero);
     if gdiffij=zero then do;
        Dgsample[i,j]=0.5;
        Dgsample[j,i]=Dgsample[i,j];
     end;
     * else, check signs of difference elements;
     if gdiffij ^= zero then do;
      nonneg=0; nonpos=0;
      do k=1 to nrow(gdiffij);
       if gdiffij[k,]>=0 then nonneg=nonneg+1;
       if gdiffij[k,]<=0 then nonpos=nonpos+1;
      end;
      * dominance case;
      if nonneg=ijdom then Dgsample[i,j]=1; else Dgsample[i,j]=0;
      if nonpos=ijdom then Dgsample[j,i]=1; else Dgsample[j,i]=0;
      * undetermined case (differences have different signs);
      if (Dgsample[i,j]=0 & Dgsample[j,i]=0) then do;
        Dgsample[i,j]=0.5;
        Dgsample[j,i]=Dgsample[i,j];
      end;
     end;


end; *(i loop);
end; *(j loop);


create Dcsample from Dcsample;
 append from Dcsample;
close Dcsample;
create Dasample from Dasample;
 append from Dasample;
close Dasample;
create Dgsample from Dgsample;
 append from Dgsample;
close Dgsample;

* create a file with sample dominance values;
* variables: i, j, cell value;
Sample_dc = J(&p*(&p-1), 3, 0); Sample_da = J(&p*(&p-1), 3, 0);
Sample_dg = J(&p*(&p-1), 3, 0); row=0;
do i=1 to nrow(Dcsample);
 do j=1 to ncol(Dcsample);
  if i ^= j then do;
   row=row+1;
   Sample_dc[row,1]=i; Sample_dc[row,2]=j; Sample_dc[row,3]=Dcsample[i,j];
   Sample_da[row,1]=i; Sample_da[row,2]=j; Sample_da[row,3]=Dasample[i,j];
   Sample_dg[row,1]=i; Sample_dg[row,2]=j; Sample_dg[row,3]=Dgsample[i,j];
  end;
 end;
end;
create sample_dc from Sample_dc[colname={i j dij}];
 append from Sample_dc;
close sample_dc;
create sample_da from Sample_da[colname={i j dij}];
```

```
 append from Sample_da;
close sample_da;
create sample_dg from Sample_dg[colname={i j dij}];
 append from Sample_dg;
close sample_dg;

finish;
run;

/*************************************************************************
*         PART 2:  Bootstrapping Loop
*************************************************************************/
* input:  data set containing Y, data set containing X;
data depvar;
 set original;
 keep &dep;
run;

data indepvar;
 set original;
 keep &indep;
run;


* steps:  1. resample, 2. compute r-squared vector, 3. repeat B times;
option ls=70 nonotes nosource nosource2;
proc IML;
reset noprint; * nolog;
start;

*A. Generate a vector of B seeds (seedvec);
/***********************************************************************/
/* The vector seedvec contains B random seeds from a uniform          */
/* distribution, generated using the initial value (seed) specified    */
/* above.  The vector is then saved to an external file, seeds,        */
/* which is subsequently read into the SAS dataset (seeds).            */
/***********************************************************************/

seedvec=J(&B,1,0);
 do i=1 to &B;
   seedvec(|i,|) = uniform(&seed)*100000;
 end;

create seeds from seedvec;
 append from seedvec;
close seeds;

finish;
run; quit;


/*====================================================================*/
/*     start 1 to B loop                                              */
/*====================================================================*/

%do r = 1 %to &B;

proc IML;
reset noprint; * nolog;
start;
* obtain Aij and Dij for each ij pair;

* read in the original data set (into DATA);
 use depvar;
  read all into YData;
 close depvar;
 use indepvar;
  read all into XData;
 close indepvar;
DATA = YData || XData;
Y = YData;
X = XData;
```

```
* read in the seeds vector (into seedvec);
use seeds;
 read all into seedvec;
close seeds;

 n=nrow(DATA);
 r=ncol(DATA);

*** check input;
if &r=1 then do;
 print 'Check that the data were read properly!!!';
 first5Y = J(1,5,0); first5Y=Y[1:5,];
 first5X = J(&p,5,0); first5X=X[1:5,];
 print 'The number of observations used is:' n;
 print 'The total number of variables used is:' r;
 print 'The dependent values for the first five cases are:' first5Y;
 print 'The predictor values for the first five cases are:' first5X;
 print 'You requested' &B 'bootstrap samples';
 print 'The bootstrap procedure is now running... PLEASE WAIT...';
end;


/*******************************************************************/
/* For bootstrapping run i, the i-th entry of the vector seedvec      */
/* (resample) is used to generate a vector (integers) whose entries    */
/* consist of the integers 1 to n sampled randomly, with replacement,  */
/* from a uniform distribution.                                        */
/*******************************************************************/
 integers=J(n,1,0);

 * Specify seed for resampling (resample);
 resample=J(1,1,0);
 resample=seedvec(|&r,|);

 * generate a vector (integers) containing the integers 1 to n sampled
 randomly, with replacement, from a uniform distribution whose initial seed is
 resample;
 Do i=1 to n;
  integers(|i|)=int(uniform(resample)*(n-1)+1);
 end;
*print &r integers;

%if &predtype='r' %then %do;
/*=================================================================*/
/***                  PAIR RESAMPLING BOOTSTRAP                     ***/
/*=================================================================*/
/*******************************************************************/
/* The i-th entry in integers is an integer (j) between 1 and n which  */
/* indicates the entry (row) number of the data matrix to be taken     */
/* as the i-th entry (row) in the matrix obsboot.  This is repeated    */
/* for i = 1 to n.                                              */
/*******************************************************************/
 * obsboot is a matrix of randomly sampled observations;
 obsboot=J(n,r,0);
 Do i=1 to n;
  j=integers(|i, |);
  obsboot(|i, |)=DATA(|j, |);
 end;

 * convert bootstrapped data to external file;
 **************************************************;
create bootdata from obsboot[colname={&dep &indep}];
 append from obsboot;
close bootdata;


%end;
/*==================== END of Pair Resampling ======================*/
```

```
%if &predtype='f' %then %do;
/*=====================================================================*/
/***                  RESIDUAL RESAMPLING BOOTSTRAP                  ***/
/*=====================================================================*/
/**********************************************************************/
/* From the matrix DATA, whose first column represents Y and the      */
/* remaining columns represent X (as specified above), the vector of   */
/* Least-squares predicted values, YHat, and the vector of the        */
/* Least Squares residuals, Res, are computed.                        */
/**********************************************************************/

* fit a model with an intercept to obtain YHat and Res;
 ONE=J(n,1,1);
 X1=ONE||X;
 YHat=X1*INV(X1`*X1)*X1`*Y;
 Res=YHat-Y;


/**********************************************************************/
/* The i-th entry in integers is an integer between 1 and n (j) which  */
/* indicates the entry (row) number of the residuals vector, Res,     */
/* to be added to the i-th predicted value, Y-Hat.  This is repeated   */
/* for i = 1 to n.  The vector Yboot consists of the initial predicted */
/* values (YHat) with the randomly selected residuals added.  The      */
/* matrix Fixdata consists of Yboot and the original X matrix.         */
/**********************************************************************/

 * Resboot is a vector of randomly sampled residuals;
 Resboot=J(n,1,0);
 Do i=1 to n;
  j=integers(|i, |);
  Resboot(|i, |)=Res(|j, |);
 end;

 * add random residuals to Y-hat;
 Yboot=YHat+Resboot;

 * an n by r matrix of bootstrapped Y and fixed Xs;
 Fixdata=Yboot||X;

 * convert bootstrapped data to external file;
 ****************************************************;
create bootdata from obsboot[colname={&dep &indep}];
 append from obsboot;
close bootdata;

%end;
/*=================== END of Residual Resampling =====================*/

finish;
run; quit;


* produce bootstrap data sets;


/* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%*/
/*        Computing Dij                     */
/* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%*/

* regression of bootstrapped data;
 ***********************************;

data resample;
set bootdata;
run;

proc reg corr data=resample noprint outest=regfile (keep=_in_ _rsq_ &indep);
  model &dep=&indep / selection=adjrsq;
run; quit; option nonotes;

* order all subset models in lexicographical order;
```

```
 data bootmat;
  set regfile;
  *%if %substr(&sysver,1,1)>7 %then if _n_>1;;
  %do i=1 %to &p;
    &&x&i=(&&x&i>.);
  %end;
  if _IN_=. then delete;
  keep &indep _IN_ _RSQ_;
 run;
proc sort; by _IN_ %do i=1 %to &p; descending &&x&i %end;; run;


* obtain Aij and Dij for each ij pair;
proc IML;
reset noprint; * nolog;
start;
/***********************************************************************
*         COMPUTE DOMINANCE MATRICES (resample)
***********************************************************************/
use bootmat;
 read all into damat;
close bootmat;
* damat is a matrix that consists of p+2 columns: 1/0 for x1 – xp, in, Rsq;

use Dcsample;
 read all into sampleDc;
close Dcsample;
use Dasample;
 read all into sampleDa;
close Dasample;
use Dgsample;
 read all into sampleDg;
close Dgsample;

* at first run, initialize frequency matrices and sample pattern counter;
if &r =1 then do;
 freqDc1=J(&p, &p, 0); freqDc0=J(&p, &p, 0); freqDcn=J(&p, &p, 0);
 freqDa1=J(&p, &p, 0); freqDa0=J(&p, &p, 0); freqDan=J(&p, &p, 0);
 freqDg1=J(&p, &p, 0); freqDg0=J(&p, &p, 0); freqDgn=J(&p, &p, 0);
 patcount=J(1,3,0); *columns correspond to dominance types;
end;

* at subsequent runs retrieve files as sum matrices;
if &r ^= 1 then do;
use sumDc1;
 read all into freqDc1;
close sumDc1;
use sumDc0;
 read all into freqDc0;
close sumDc0;
use sumDcn;
 read all into freqDcn;
close sumDcn;

use sumDa1;
 read all into freqDa1;
close sumDa1;
use sumDa0;
 read all into freqDa0;
close sumDa0;
use sumDan;
 read all into freqDan;
close sumDan;

use sumDg1;
 read all into freqDg1;
close sumDg1;
use sumDg0;
 read all into freqDg0;
close sumDg0;
use sumDgn;
 read all into freqDgn;
```

```
close sumDgn;

use patfreq;
 read all into patcount;
close patfreq;

end;

* r-squared vector is the last column of damat (without null model);
RSQ=damat[,ncol(damat)];
*print &r RSQ;

Dc=J(&p,&p,0); Da=J(&p,&p,0); Dg=J(&p,&p,0);



/**************************************************/
/*****         Complete dominance           ****/
/**************************************************/
do i=1 to &p-1;
 do j=i+1 to &p;

 * define matrix of constants ('comp') to determine complete dominance between
 each pair of predictors;

        comp=J(2**(&p-2),2**&p-1,0);

        * Xh is any subset model that does not include i and j;
        * Xh contains the columns of damat that do not involve i or j;
        Xh=J(nrow(damat),&p-2,99);
          Xhcol=0;                    * index column number in Xh;
          do h=1 to &p;
           if (i ^= h & j ^= h) then do;    * find non i,j columns;
            Xhcol=Xhcol+1;                          * update column number in Xh;
            Xh[,Xhcol] = damat[,h];         * assign column to Xh;
           end;
          end;

          * contrast rows (subsets) representing XiXh and XjXh;
        comprow=1;
        do r=1 to 2**&p-2;                * for each pair of rows;
         do s=r+1 to 2**&p-1;
          if Xh[r,]=Xh[s,] then do;      * if the rows of Xh are the same;
                                         * and if i and j are contrasted;
           if (damat[r,i]=1 & damat[r,j]=0 & damat[s,i]=0 &
                damat[s,j]=1) then do;
                 comp[comprow,r]=1;
                 comp[comprow,s]=-1;
                 comprow=comprow+1;
           end;
          end;

         end;   * do r loop;
        end;    * do s loop;


*** Determine the complete Dij value ***;

  cdiffij=comp*RSQ;

  zero=J(nrow(cdiffij), ncol(cdiffij),0);
  ijdom=nrow(cdiffij);

  * obtain complete dominance matrices;
     * undetermined case (all differences are zero);
     if cdiffij=zero then do;
        Dc[i,j]=0.5;
        Dc[j,i]=Dc[i,j];
     end;
     * else, check signs of difference elements;
     if cdiffij ^= zero then do;
      nonneg=0; nonpos=0;
```

```
        do k=1 to nrow(cdiffij);
         if cdiffij[k,]>=0 then nonneg=nonneg+1;
         if cdiffij[k,]<=0 then nonpos=nonpos+1;
        end;
        * dominance case;
        if nonneg=ijdom then Dc[i,j]=1; else Dc[i,j]=0;
        if nonpos=ijdom then Dc[j,i]=1; else Dc[j,i]=0;
        * undetermined case (differences have different signs);
        if (Dc[i,j]=0 & Dc[j,i]=0) then do;
          Dc[i,j]=0.5;
          Dc[j,i]=Dc[i,j];
        end;
       end;



/**************************************************/
/*****         Conditional dominance            ****/
/**************************************************/
* define matrix of constants ('avg') to determine average (within model
size) dominance between each pair of predictors;

subsets=damat;
* subsets is a matrix that consists of p+2 columns (and
the null model in the top row): 1/0 for x1 - xp, in, Rsq;

  avg=J(&p,2**&p-1,0);
  do c=1 to &p;         * for each model size;
   num=0;               * index number of subsets per size;
   do m=1 to 2**&p-1;   * for each model, determine number of models
                                   of size is c-1 to which the ith predictor
makes
                                   an additional contribution;
    if (subsets[m,&p+1]=c-1 & subsets[m,i]=0) then num=num+1;
   end;
   if num=0 then num=1; * for the case of the null model (size=0);
   do r=1 to 2**&p-1;    * determine subsets to contrast;
    * consider additional contributions for model size (c);
    if (subsets[r,&p+1]=c-1 | subsets[r,&p+1]=c) then do;
     * average contribution of models that inlcude i but not j;
     if (subsets[r,i]=1 & subsets[r,j]=0) then avg[c,r]=1/num;
     * average contribution of models that inlcude j but not i;
     if (subsets[r,i]=0 & subsets[r,j]=1) then avg[c,r]=-1/num;
    end;
   end;
  end;

*** Determine the average Dij value ***;

  adiffij=avg*RSQ;

  zero=J(nrow(adiffij), ncol(adiffij),0);
  ijdom=nrow(adiffij);

  * obtain average dominance matrices;
     * undetermined case (all differences are zero);
     if adiffij=zero then do;
        Da[i,j]=0.5;
        Da[j,i]=Da[i,j];
     end;
     * else, check signs of difference elements;
     if adiffij ^= zero then do;
      nonneg=0; nonpos=0;
      do k=1 to nrow(adiffij);
       if adiffij[k,]>=0 then nonneg=nonneg+1;
       if adiffij[k,]<=0 then nonpos=nonpos+1;
      end;
      * dominance case;
      if nonneg=ijdom then Da[i,j]=1; else Da[i,j]=0;
      if nonpos=ijdom then Da[j,i]=1; else Da[j,i]=0;
      * undetermined case (differences have different signs);
      if (Da[i,j]=0 & Da[j,i]=0) then do;
```

```
        Da[i,j]=0.5;
        Da[j,i]=Da[i,j];
      end;
     end;


/*****************************************************/
/*****         General dominance                ****/
/*****************************************************/
* define matrix of constants ('glob') to determine global (overall average)
dominance between each pair of predictors;

  glob=J(1,2**&p-1,99);
  do g=1 to 2**&p-1;
   glob[,g]=avg[+,g]/&p;
  end;

*** Determine the global Dij value ***;

  gdiffij=glob*RSQ;
  *print 'global dominance comparisons';
  *print i j gdiffij;

  zero=J(nrow(gdiffij), ncol(gdiffij),0);
  ijdom=nrow(gdiffij);

  * obtain global dominance matrices;
     * undetermined case (all differences are zero);
     if gdiffij=zero then do;
        Dg[i,j]=0.5;
        Dg[j,i]=Dg[i,j];
     end;
     * else, check signs of difference elements;
     if gdiffij ^= zero then do;
      nonneg=0; nonpos=0;
      do k=1 to nrow(gdiffij);
       if gdiffij[k,]>=0 then nonneg=nonneg+1;
       if gdiffij[k,]<=0 then nonpos=nonpos+1;
      end;
      * dominance case;
      if nonneg=ijdom then Dg[i,j]=1; else Dg[i,j]=0;
      if nonpos=ijdom then Dg[j,i]=1; else Dg[j,i]=0;
      * undetermined case (differences have different signs);
      if (Dg[i,j]=0 & Dg[j,i]=0) then do;
        Dg[i,j]=0.5;
        Dg[j,i]=Dg[i,j];
      end;
     end;


end; *(i loop);
end; *(j loop);



**** compare current D matrices to sample D matrices ****;
if Dc=sampleDc then patcount[,1]=patcount[,1]+1;
if Da=sampleDa then patcount[,2]=patcount[,2]+1;
if Dg=sampleDg then patcount[,3]=patcount[,3]+1;
create patfreq from patcount;
 append from patcount;
close patfreq;


thisDc1=J(&p,&p,0); thisDc0=J(&p,&p,0); thisDcn=J(&p,&p,0);
thisDa1=J(&p,&p,0); thisDa0=J(&p,&p,0); thisDan=J(&p,&p,0);
thisDg1=J(&p,&p,0); thisDg0=J(&p,&p,0); thisDgn=J(&p,&p,0);
do i=1 to &p;
 do j=1 to &p;
  if i ^=j then do;
   if Dc[i,j]=1 then thisDc1[i,j]=1;
   if Dc[i,j]=0 then thisDc0[i,j]=1;
```

```
   if Dc[i,j]=0.5 then thisDcn[i,j]=1;
   if Da[i,j]=1 then thisDa1[i,j]=1;
   if Da[i,j]=0 then thisDa0[i,j]=1;
   if Da[i,j]=0.5 then thisDan[i,j]=1;
   if Dg[i,j]=1 then thisDg1[i,j]=1;
   if Dg[i,j]=0 then thisDg0[i,j]=1;
   if Dg[i,j]=0.5 then thisDgn[i,j]=1;
  end;
 end;
end;

**** add current D matrices to previous D matrices ****;
freqDc1 = freqDc1+thisDc1;
freqDa1 = freqDa1+thisDa1; freqDg1 = freqDg1+thisDg1;
freqDc0 = freqDc0+thisDc0;
freqDa0 = freqDa0+thisDa0; freqDg0 = freqDg0+thisDg0;
freqDcn = freqDcn+thisDcn;
freqDan = freqDan+thisDan; freqDgn = freqDgn+thisDgn;

create sumDc1 from freqDc1;
 append from freqDc1;
close sumDc1;
create sumDc0 from freqDc0;
 append from freqDc0;
close sumDc0;
create sumDcn from freqDcn;
 append from freqDcn;
close sumDcn;

create sumDa1 from freqDa1;
 append from freqDa1;
close sumDa1;
create sumDa0 from freqDa0;
 append from freqDa0;
close sumDa0;
create sumDan from freqDan;
 append from freqDan;
close sumDan;

create sumDg1 from freqDg1;
 append from freqDg1;
close sumDg1;
create sumDg0 from freqDg0;
 append from freqDg0;
close sumDg0;
create sumDgn from freqDgn;
 append from freqDgn;
close sumDgn;

*at last run create summary tables containing the following columns:
i, j, f=1, f=0, f=0.5 pij pji;
if &r=&B then do;
 row=1;
 summary_c=J(&p*(&p-1), 8, -9);
 summary_a=J(&p*(&p-1), 8, -9);
 summary_g=J(&p*(&p-1), 8, -9);
 do i=1 to &p;
  do j=1 to &p;
   if i^=j then do;
    summary_c[row,1]=i; summary_c[row,2]=j;
    summary_c[row,3]=freqDc1[i,j];
    summary_c[row,4]=freqDc0[i,j]; summary_c[row,5]=freqDcn[i,j];
    summary_c[row,6]=freqDc1[i,j]/&B; summary_c[row,7]=freqDc0[i,j]/&B;
    summary_c[row,8]=freqDcn[i,j]/&B;
    summary_a[row,1]=i; summary_a[row,2]=j;
    summary_a[row,3]=freqDa1[i,j];
    summary_a[row,4]=freqDa0[i,j]; summary_a[row,5]=freqDan[i,j];
    summary_a[row,6]=freqDa1[i,j]/&B; summary_a[row,7]=freqDa0[i,j]/&B;
    summary_a[row,8]=freqDan[i,j]/&B;
    summary_g[row,1]=i; summary_g[row,2]=j;
    summary_g[row,3]=freqDg1[i,j];
    summary_g[row,4]=freqDg0[i,j]; summary_g[row,5]=freqDgn[i,j];
```

```
   summary_g[row,6]=freqDg1[i,j]/&B; summary_g[row,7]=freqDg0[i,j]/&B;
    summary_g[row,8]=freqDgn[i,j]/&B;
    row=row+1;
   end;
  end;
 end;
 create csummary from summary_c[colname={i j f1 f0 fn Pij Pji pno}];
  append from summary_c;
 close csummary;
 create asummary from summary_a[colname={i j f1 f0 fn Pij Pji pno}];
  append from summary_a;
 close asummary;
 create gsummary from summary_g[colname={i j f1 f0 fn Pij Pji pno}];
  append from summary_g;
 close gsummary;
 create patsum from patcount[colname={complete conditional general}];
  append from patcount;
 close patsum;
end;

finish;
run; quit;

%end; ***** end the 1 to B loop ****;


/*********************************************************************/
/*         PART 3:  Probabilities and reproducibilities
*/
/*********************************************************************/
* input:  nine p by p matrices of COUNTS;
* stpes:  convert to probabilities, compute test statistics / CIs, plots;
* output:  tables;
data average;
 set asummary;
 dominance='conditional';
 pair=i*10+j;
run;
data complete;
 set csummary;
 dominance='complete';
 pair=i*10+j;
run;
data global;
 set gsummary;
 dominance='general';
 pair=i*10+j;
run;

data merged;
 set complete average global;
run;
/* proc print data=merged; title 'results'; run; */
data results; set merged;
 array f[3] f1 f0 fn;
  do domtype=1 to 3;
   domfreq = f[domtype];
   output;
  end;
run;
data results; set results;
  if domtype=1 then Dij=1;
  if domtype=2 then Dij=0;
  if domtype=3 then Dij=0.5;
run;
proc sort data=results; by dominance pair; run;
proc means data=results noprint;
 var Dij; freq domfreq;
 by dominance pair;
 output out=meansout mean=Dbar STD=SE;
run;
```

```
data samplea;
 set Sample_da; pair=i*10+j; dominance='conditional';
run;
data samplec;
 set Sample_dc; pair=i*10+j; dominance='complete';
run;
data sampleg;
 set Sample_dg; pair=i*10+j; dominance='general';
run;
data sample;
 set samplec samplea sampleg;
run;


******* SUMMARY OUTPUT *******;
* variables: dominance-type, i, j, p(iDj), p(jDi), p(noD), Dbar, SE,
sample-value, pattern-prob;
*** merge sample, results and meansout data sets by dominance and pair;
proc sort data=sample; by dominance pair; run;
data results; set results; if domtype=1; run;
proc sort data=results; by dominance pair; run;
proc sort data=meansout; by dominance pair; run;
data printout;
 merge results sample meansout;
  by dominance pair;
run;

data table;
 set printout(drop=domtype domfreq _TYPE_ _FREQ_);
/* i_dom_j=pij; j_dom_i=pji; no_dom=pno; */
 Pijno=pno;
 Dij_mean = dbar;
 Dij_se = se;
 if dij=1 then reprod = pij;
 if dij=0 then reprod = pji;
 if dij=0.5 then reprod = pijno;
 format Dij_se F6.3;
run;
data table; set table(drop = f1 f0 fn pno dbar); run;
proc sort; by dominance pair; run;
proc print data=table;
 title 'summary of results: all pairs';
 title2 'Pij, Pji, Pijno are dominance probabilities and';
 title3 'reprod are reproducibility values';
 var dominance i j Dij Dij_mean dij_SE Pij Pji Pijno reprod;
run;
data plots; set table;
 if Pij ge Pji;
 format Dij_se F6.3;
run;
proc print data=plots;
 title 'summary of results: pair arranged by dominant predictor';
 title2 'Pij, Pji, Pijno are dominance probabilities and';
 title3 'reprod are reproducibility values';
 var dominance i j Dij Dij_mean dij_SE Pij Pji Pijno reprod;
run;

%mend dom;



/**********************************************************
   The following macros expand a variable list containing
   hyphens into a list specifying each individual variable.
**********************************************************/

%MACRO EXPAND;
  %LET LNGTH = %LENGTH(&INDEP);  %LET TEMP=;
  %DO _INDEX_ = 1 %TO &LNGTH;
    %LET ITEM = %SCAN(&INDEP,&_INDEX_,%QUOTE( ));
```

```
    %IF %LENGTH(&ITEM) EQ 0 %THEN %GOTO DONE;
    %IF %INDEX(&ITEM,-) > 0 %THEN %EXPANDED;
    %LET TEMP = &TEMP &ITEM;
  %END;  %DONE:  %LET INDEP = &TEMP;
%MEND EXPAND;

%MACRO EXPANDED;
  %LET DASH = %INDEX(&ITEM,-);
  %DO I = %EVAL(&DASH-1) %TO 1 %BY -1;
    %LET ALPHANUM = %SUBSTR(&ITEM,&I,1);
    %DO II = 0 %TO 9;  %IF &ALPHANUM EQ &II %THEN %GOTO FOUND;  %END;
    %GOTO DONE;
  %FOUND:  %END;
  %DONE:  %LET PREFX = %SUBSTR(&ITEM,1,&I);
  %LET LOWER = %SUBSTR(&ITEM,%EVAL(&I+1),%EVAL(&DASH-&I-1));
  %LET UPPER = %SUBSTR(&ITEM,%EVAL(&DASH+&I+1));
  %LET ITEM=; %DO II = &LOWER %TO &UPPER; %LET ITEM = &ITEM &PREFX.&II;
  %END;
%MEND EXPANDED;
```