
Relative importance analysis for count regression models

Journal Title
XX(X):1–18
©The Author(s) 0000
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/

SAGE

Joseph N. Luchman¹

Abstract

Determining independent variable relative importance is a highly useful practice in organizational science. Whereas techniques to determine independent variable importance are available for normally distributed and binary dependent variable models, such techniques have not been extended to count dependent variables (CDVs). The current work extends previous research on binary and multi-category dependent variable relative importance analysis to provide a methodology for conducting relative importance analysis on CDV models using dominance analysis (DA). Moreover, the current work provides a set of comprehensive data analytic examples that demonstrate how and when to use CDV models in a DA and the advantages general DA statistics offer in interpreting CDV model results. Moreover, the current work outlines best practices for determining independent variable relative importance for CDVs using replaceable examples on data from the publicly available National Longitudinal Survey of Youth 1979 cohort. The present work then contributes to the literature by using in-depth data analytic examples to outline best practices in conducting relative importance analysis for CDV models and by highlighting unique information DA results provide about CDV models.

Keywords

Dominance Analysis, Relative Importance, Poisson Regression, Negative Binomial Regression, R-square

Introduction

Discrete, infrequent events are common dependent variables in Organizational Science (Bettinazzi & Feldman, 2021; Naumovska, Zajac & Lee, 2021; Soda, Mannucci & Burt, 2021, e.g.,) and are modeled using *count regression models/CRMs* such as

¹Fors Marsh

Poisson regression/PR or negative Binomial regression/NBR (Blevins, Tsang & Spain, 2015).

CRMs differ from the commonly used linear regression model/LRM as their functional forms are exponential or log-linear (i.e., have the form $e^{\sum \beta x}$) to accommodate the necessarily truncated (i.e., non-negative) range of the dependent variable/DV. The log-linear form of CRMs requires the use of new metrics to accurately interpret model results and changes the language that can be used to describe model predictions. For example, log-linear CRMs' model coefficients are often interpreted in their exponential form as *incidence rate ratios/IRRs* which describes percentage changes in counts for 1 unit increments to the independent variable/IV. In addition, the explained variance R^2 does not apply in a straightforward way to log-linear CRMs for describing the quality of model predictions' fit to the data (Cameron & Windmeijer, 1996).

The way in which count data are constructed also tends to differ from how many continuous, Gaussian distributed variables are constructed that have implications for modeling. Count variables are often constructed as aggregated events over a specific time period such as the number of organizations adopting a specific practice in a week (Naumovska, Zajac & Lee, 2021) or number of divestitures in a year (Bettinazzi & Feldman, 2021). For some data, the period of aggregation may differ across observations in the data. Such differences might be observed when some observations report outcomes for one year/12 months but some report on outcomes for 9 months. These differences in aggregation window can result in different "exposure" to the count generating process and can result in a confound. In such circumstances, an additional *offset* variable reflecting the amount of exposure the observation had to the count generating process is introduced to re-establish parity across observations (Glerum, Joseph, McKenny & Fritzsche, 2021, see, for an example). In other cases, count data aggregations include observations that "opt out" of the count generating process such as firms choosing not to outsource their patent submissions to a law firm in a year producing 0 patents (Somaya, Williamson & Lorinkova, 2008). Such opting out results in *zero inflation*; a situation in which there are more 0 values than would be expected from a Poisson or negative Binomial distribution. Zero inflation can be accommodated by using a specialized zero-inflation model (Glerum, Joseph, McKenny & Fritzsche, 2021, again see, for an example).

The complexities associated with the use of CRMs as well as those associated with processes used to aggregated count data also extend to considering the use of CRM-based postestimation methods such as relative importance analysis. Relative importance analysis is a widely used model postestimation method that is applied to assist in the interpretation of model results. Specifically, relative importance analysis adds detail to the estimation of parameter estimates and can provide practically useful information about IV predictive utility (Tonidandel & LeBreton, 2011). To date, published methodological work on relative importance analysis has discussed how to apply the method to similarly complex models such as binary (Azen & Traxel, 2009), ordered, and multinomial logit models (Luchman, 2014) but has not provided an extensive discussion relative importance analysis with CRMs. As was discussed above, there are multiple estimation and data generation complexities with CRMs that are unique and were not discussed in past work on LRM or logit models. Thus, extending

the literature on relative importance analysis to the unique features of CRMs is an important step toward applying the method to these commonly estimated models in Organizational Science.

The purpose of this work is to provide an extensive discussion of the application of relative importance analysis methods to CRMs. First, this manuscript discusses how to use CRMs in (Azen & Budescu, 2003, dominance analysis/DA), a relative importance method with a strong conceptual foundation (Grömping, 2007) and flexible implementation in terms of extensibility (Luchman, 2021, see). A focus of this discussion will be to provide recommendations on the model fit metric to use as well as providing a full data analytic example of how to implement the method with CRMs focusing on PR and NBR. Second, this paper provides a detailed discussion of the concepts of exposure and zero inflation with particular attention to issues these two CRM-relevant data generation facets have for computing dominance statistics and determining importance. Finally, this paper extends on the work of Blevins, Tsang, and Spain's (Blevins, Tsang & Spain, 2015) work by discussing how to apply model postestimation methods to assist in adding context and detail to the model results, using their paper's work to choose the correct model given the structure of the data to be analyzed.

In the section below, I begin this work with a detailed discussion of the conceptual background for DA. Below I discuss what DA is, why it applies to the results for CRMs, and how it is used to infer the importance of IVs in a predictive model.

Dominance Analysis

DA is a method that evaluates IV relative importance based on unique contributions to a model fit metric. DA is then a methodology that uses empirical results, in particular those related to expected/predicted versus observed differences, to evaluate IV importance. The use of predicted versus observed differences is a form of variance-based importance and has a long history in Organizational and Behavioral Science as a method for inferring importance (Johnson & LeBreton, 2004, see). What makes DA unique among variance-based importance methods is DA's conceptual foundation as an extension of Shapley value decomposition from Cooperative Game Theory ().

Shapley Value Decomposition

Cooperative games can be thought of as a structured form of interaction in which the interacting parties are required to work together toward a common goal and share information with one another (cite). This form of interaction is not unlike a team task where the different team members have different pieces of information or different tools and must cooperate to accomplish the task. The Shapley value decomposition methodology then uses the structure of the game to determine the unique value ascribed to each player independent of other players irrespective of potentially overlapping player contributions. Computationally, the Shapley value computation is the average increment to the obtained cooperative value a player obtains across all possible permutations of coalitions. Formally, player A's (Pl_a) Shapley value would be:

$$SV_{Pl_a} = \frac{\sum_{i=1}^P V_{O_i \cup Pl_a} - V_{O_i}}{P} \quad (1)$$

Where P refers to the total number of permutations of the p players and O_i is some distinct ordered set (i.e., where the order of inclusion matters) of players not including Pl_a that can include the null set of no players.

Predictive models work in a way similar to cooperative games in that IVs jointly enter into a predictive equation (i.e., must interact) and are adjusted for redundancy in terms of prediction of the DV (i.e., share information). Thus, Shapley value decomposition can be applied in a straightforward way to predictive models if the IVs are thought of as players and the fit statistic is thought of as the cooperative goal value. Indeed, the Shapley value methodology has received a great deal of attention recently in the machine learning literature as a general, model agnostic method for understanding complex model predictions (Lundberg, Erion, Chen, DeGrave, Prutkin, Nair, Katz, Himmelfarb, Bansal & Lee, 2020, e.g.,). Moreover, the generality of the Shapley value decomposition methodology means that the approach extends in a natural way to the decomposition of CRM values. Shapley values are a general method for decomposing values but its extension to DA is the methodology that most clearly defined methods for determining the relative importance of predictive models.

General Dominance Statistics

The *General Dominance Statistic* in the DA approach to relative importance is directly related to the Shapley value computation but is focused on decomposing a model fit metric and simplifies the Shapley value computation given the features of fit metrics. The simplification general dominance statistics apply is based on the acknowledgment that, for a predictive model, the ordering of IV input is usually irrelevant. That is, the order of inclusion for fit metrics in predictive models tends to produce equal fit statistics (i.e., $R_{IV_x IV_z}^2 = R_{IV_z IV_x}^2$). Computationally, the general dominance statistic for an IV (IV_x) is:

$$GDS_{IV_x} = \sum_{j=1}^{TCb} \frac{F_{U_j \cup IV_x} - F_{U_j}}{(C_{U_j \cup IV_x})k} \quad (2)$$

Where TCb refers to the total number of unique combinations not including IV_x (i.e., 2^{k-1}) of the k IVs, U_j is some distinct unordered set (i.e., where the order of inclusion doesn't matter) of IVs not including IV_x that can include the null set of no IVs, and $C_{U_j \cup IV_x}$ is the number of distinct combinations of IVs with all the IVs in U_j as well as IV_x included in the model. The general dominance statistics are then generated as a weighted-average increment to the fit metric across all combinations of IVs to which the focal IV is included. The weighted averaging component of the denominator of Equation 2 reflects the idea that there are redundancies in the Shapley value computation that can be rescaled such that they can be reflected by a weighted average.

The general dominance statistics are used to determine importance among the IVs by comparing their values. For example, if IV_x has a larger general dominance statistic than IV_z , IV_x is said to *generally dominate*, and is thus more important than, IV_z .

DA extends on Shapley value composition further in the next section that discusses two other importance designations used in DA to make stronger importance determinations than that which general dominance statistics are capable.

Other Dominance Computations and Designations

DA also computes statistics known as a *conditional dominance statistic* that is closely related to the general dominance statistic. The conditional differs from the general dominance statistic as it is computed by a specified number of IVs included in the model. Thus, each IV has k conditional dominance statistics. The conditional dominance statistic for an IV (IV_x) with a set number of IVs in the model (s) is computed:

$$CDS_{IV_x}^s = \sum_{l=1}^{C_{U_l \cup IV_x}} \frac{F_{U_l \cup IV_x} - F_{U_l}}{C_{U_l \cup IV_x}} \quad (3)$$

Where U_l is some distinct unordered set of $s - 1$ IVs not including IV_x that can include the null set of no IVs, and $C_{U_l \cup IV_x}$ is the number of distinct combinations of the s IVs with all the IVs in U_l as well as IV_x included in the model.

Conditional dominance statistics computed using Equation 3 are a subset of the computations in the general dominance statistic in Equation 2 and can be averaged to obtain the general dominance statistic. That is, $GDS_{IV_x} = \sum_{g=1}^k \frac{CDS_{IV_x}^g}{k}$.

The conditional dominance statistics generated for each IV are also used to determine importance among the IVs but the procedure for doing so requires a different approach than that applied to general dominance statistics. For conditional dominance statistics, if, within a number of IVs in the model, all of IV_x 's values are larger than IV_z , IV_x is said to *conditionally dominate*, and is thus more important than, IV_z . Another way of thinking about conditional dominance is that, if the conditional dominance statistics for an IV were graphed as a line by the number of IVs in the model, IV_x 's value is above IV_z value for the entirety of the trend. Because conditional dominance statistics are a 'less averaged' version of the general dominance statistics, conditional dominance of one IV over another is a stricter criterion than general dominance and implies a stronger form of IV relative importance.

DA also uses a third type of comparison called *complete dominance* that does not averaging at all. In complete dominance, each increment to the fit metric associated with IV_x is compared to an corresponding increment for IV_z . More formally, for all sets of IVs U_q that do not include IV_x or IV_z and can include the null set of no IVs, $F_{U_q \cup IV_x} - F_{U_q}$ is always larger than $F_{U_q \cup IV_z} - F_{U_q}$. In such situations, IV_x is said to *completely dominate* IV_z . Because there is no averaging involved in the comparisons that produce complete dominance, it is the most stringent importance comparison between two IVs.

This section has outlined why Shapley value decomposition and the extension into DA applies to CRMs. This section has also provided a discussion of how DA statistics and designations are computed and determined. In the section below, I transition from a broad outline of DA to a more targeted discussion of the application of DA to CRMs.

The discussion of CRM-based DA begins by considering which fit metric should be used and proceeds to an extensive example applying DA to both PR and NBR.

Applying Dominance Analysis to Count Regression Models

Conceptually, DA can be applied to CRMs as its basis in Shapley values allows the application of the method to any procedure that is similar to a cooperative game. A complication of applying DA to CRMs is that most of the literature on DA has focused on its application to LRM with the residual R^2 (i.e., R_{RES}^2). CRMs and LRM have many similarities and, given the numerical nature of count DVs and that the output from CRMs is typically in the form of a predicted count/mean value, the R_{RES}^2 could be applied to CRM-based DA.

Although R_{RES}^2 could be applied, there are good conceptual reasons to choose another fit metric. The section below provides a detailed rationale for the choice of a different metric, the deviance R^2 (i.e., R_{DEV}^2), that better reflects how CRMs fit to data.

Count Regression Fit Metric: Deviance R^2

CRMs are not only log-linear models but also follow probability distributions that differ from the LRM that has been traditionally been the focus of DA. This is important as statistical models are fit using information about the data as applied to a probability distribution to find their most likely parameter values. It is thus important to choose a fit metric that matches the probability distribution's underlying fitting criterion.

First consider LRM. LRM is based on a Normal or Gaussian probability distribution that can be simplified to the least-squares criterion which seeks to minimize $\sum (y - \hat{y})^2$ or the sums of squares between the predicted values from the LRM and the observed DV. The R_{RES}^2 that has been traditionally applied in LRM-based DA is one less than the ratio of the sum of squares of the observed DV and predicted value to the sum of squares of the observed DV and the mean of the observed DV (i.e., $\sum (y - \bar{y})^2$). Because the R_{RES}^2 uses the LRM minimization goal's in computing differences between the observed data and the predictions, the R_{RES}^2 is a useful model fit metric for LRM.

The $(y - \hat{y})^2$ computation used in the computation of the R_{RES}^2 is also known as the *unit deviance* for the Normal distribution as it describes how the model's prediction "deviates" from observed values (McCullagh & Nelder, 2019). The value in the numerator of the R_{RES}^2 is then a deviance for a Normal distribution. The denominator can be similarly cast as a deviance if considering a null LRM that estimates only an intercept and thus returns the mean. Thus, the R_{RES}^2 can also be thought of as a deviance R^2 .

Cameron and Windmeijer (Cameron & Windmeijer, 1996) first formalize the concept of extending the R_{RES}^2 to a version of the metric based on deviance. The deviance R^2 is computed as:

$$R_{DEV}^2 = 1 - \frac{D_{model}}{D_{null}} \quad (4)$$

Where D_{null} is the model deviance associated with a null model describing an intercept- or mean-only model. Thus, when the LRM's deviance computations are substituted in for the D_{model} and D_{null} values, the R_{DEV}^2 is equivalent to R_{RES}^2 .

An advantage of constructing the R_{DEV}^2 is that the deviance concept has been generalized from LRM to generalized linear models such as PR and NBR as well. For example, the unit deviance computation for PR and a special case of the NBR* is $2(y \ln \frac{y}{\hat{y}} - (y - \hat{y}))$. This PR-focused deviance value differs notably from the Normal distribution deviance in that it tends to penalize underprediction more than overprediction. The extra penalties assigned to underprediction are consistent with the PR's log-linear nature in that the log distance between 1 and 2 (.69) is larger than between 2 and 3 (.41); on average, lower log-transformed differences get penalized more than higher log-transformed differences. By contrast, Normal distribution deviance penalizes discrepancies from observed values equally in either direction.

The R_{DEV}^2 is then a useful generalization of the R_{RES}^2 that allows the researcher to apply an residual computation that makes sense for the model in question, as opposed to applying the LRM's deviance criterion to a model that is not based on LRM (Cameron & Windmeijer, 1996, .e.g.,). The R_{DEV}^2 is then the fit metric recommended for use in conducting DA with CRMs.

This section has provided a rationale for the choice of a specific fit metric to use when applying DA to CRMs. In the sections to follow, I transition to the description of the methodology for applying DA to CRMs. This discussion briefly outlines the data generated for this purpose, describes the models estimated from the data, and also describes how DA statistics and designations were determined with the results.

Count Regression Model-based Dominance Analysis: An Analytic Example

In this section I describe how I generated data for a both a PR and NBR model. These models are estimated in such a way as they would be were they based on "real" data and follow a procedure similar to that that would be applied to answering a "real" research question with the data.

The next section provides a high-level overview of how the data were generated and descriptive statistics about the data. Note that a full outline of the data generation and analysis process is available in the online supplement.

Data Generation The data for this analysis were generated as opposed to obtained empirically to maintain control over their usefulness in illustrating the key concepts for this manuscript. A goal in generating these data were to make them "realistic" when possible. That is, to reduce the elements of the data that made them simulated in order to focus on the concepts

The data simulated in this section were intended to describe the generation of garments from a tailors in a fortnight or two-week period. Data from a total of 6,780

*This special case is the NBR estimated using a quasi-likelihood method. Maximum likelihood methods require a more complex form given the estimation of the α parameter.

simulated tailors were generated over this period across two different garments/DVs and for four garment generation factors/IVs.

One dependent variable, the number of stock sport coats of a set size made by the tailors, is Poisson distributed. The second dependent variable, number of made-to-measure sport coats with custom elements made by the tailors, is negative Binomial distributed. The four garment generation factors include a) equipment reliability, b) assistant staffing levels, c) tailor skill-level, and d) tailor work experience. Each of the four garment generation factors were designed to be multivariate Normally distributed with pre-specified relationships with the two forms of garment count DVs. The pre-specified relationships between each of the garment generation factors with both garment count DVs were identical. The only factor that differed between garment generation factors was the distribution of the DV. Both garment count DVs were designed to have mean values of 1. The negative Binomial, made-to-order sport coat count variable was also designed to have a variance of 3.

The means, standard deviations, and correlations between all four IVs and two DVs is reported below in Table 1.

Table 1: Descriptive Statistics

Variable	Mean	Standard Deviation	Correlations					
			ER	AS	SL	WE	SJ	MJ
ER	-0.0355	1.1853	1.0000	0.4168	0.1446	0.2180	0.4419	0.3820
AS	-0.0261	1.4494	0.4168	1.0000	0.1958	0.2850	0.4025	0.3488
SL	-0.0056	1.5480	0.1446	0.1958	1.0000	0.3295	0.3393	0.2935
WE	0.0144	1.9211	0.2180	0.2850	0.3295	1.0000	0.3973	0.3487
SJ	0.9920	1.0040	0.4419	0.4025	0.3393	0.3973	1.0000	0.8773
MJ	0.9971	1.7478	0.3820	0.3488	0.2935	0.3487	0.8773	1.0000

Table 1 shows that the means of each of the IVs are 0 and the means of the DVs are 1. Table 1 also shows variability between the IVs in terms of their variances with ER having the lowest variability and WE the highest. The results also show modest to moderate-sized correlations between IVs with SL having some of the lowest overlap with other IVs and AS having the highest. The results for the DVs also show distinctly different standard deviations with SJ following the Poisson distribution requirement that the mean equal the variance and MJ showing a much higher variation than the mean as is typical of negative Binomial distributed variables.

The next section moves on to using these data to estimate a PR with the SJ variable as well as an NBR with the MJ variable predicted by the four IVs.

Regression Results The PR results predicting SJ are reported in Table 2.

Table 2: Poisson Regression

Coef	SE	Confidence Interval		z	p	IRR	Std
		Low	High				

Inter	−0.1805	0.0142	−0.2084	−0.1529	−12.7404	0.0000	0.8349	0.0000
ER	0.2463	0.0114	0.2239	0.2687	21.5862	0.0000	1.2793	0.2919
AS	0.1320	0.0095	0.1133	0.1506	13.8945	0.0000	1.1411	0.1913
SL	0.1264	0.0083	0.1100	0.1427	15.1533	0.0000	1.1347	0.1956
WE	0.1167	0.0069	0.1032	0.1302	16.9536	0.0000	1.1238	0.2243

Table 2 shows that each of the IVs has a statistically significant, positive effect at the $p < .05$ level. In terms of effect size magnitude, ER had the largest unstandardized coefficient/IRR value with a 1 unit change in equipment reliability leading to a 27.9% increase in the number of stock jackets produced in a fortnight. By contrast, WE obtained the smallest unstandardized coefficient/IRR value with a 1 unit change in equipment reliability leading to a 12.4% increase in the number of stock jackets produced in a fortnight. AS (14.1% increase) and SL (13.5% increase) fell between ER and WE in terms of unstandardized results.

The standardized coefficients reported in Table 2 were computed by first standardizing each IV to have a variance of 1 and then re-running the model. These standardized results show a different ordering of the IVs' effect sizes as compared to the unstandardized results. The most noteworthy change is related to the change to the ordering of WE, SL, and AS. In the standardized results, WE obtains the largest coefficient among the IVs followed by SL and then AS. This kind of change is a common finding when comparing coefficients using unstandardized and standardized coefficients. This kind of change is also one reason the use of coefficients for determining importance is not recommended in the literature (Johnson & LeBreton, 2004, see).

The NBR results predicting MJ are reported below in Table 3.

Table 3: Negative Binomial Regression

	Coef	SE	Confidence Interval		z	p	IRR	Std
			Low	High				
Inter	−0.4608	0.0207	−0.5015	−0.4205	−22.2364	0.0000	0.6308	0.0000
ER	0.4138	0.0169	0.3802	0.4477	24.4244	0.0000	1.5126	0.4905
AS	0.2249	0.0139	0.1977	0.2523	16.1768	0.0000	1.2522	0.3260
SL	0.2171	0.0123	0.1930	0.2413	17.7168	0.0000	1.2425	0.3361
WE	0.1979	0.0102	0.1780	0.2180	19.4610	0.0000	1.2189	0.3803

Theta parameter is 1.371.

Table 3's unstandardized/IRR results show many similarities to the PR modeling of SJ reported Table 2 as each IV is positive and statistically significant at $p < .05$. The magnitude of the unstandardized coefficients/IRR values obtained by the NBR tended to be larger than the PR but showed the same rank order among the IVs. Specifically, ER remained largest with a 1 unit change in equipment reliability leading to a 51.3% increase in the number of made to order jackets produced in a fortnight. WE again obtained the smallest unstandardized coefficient/IRR value with a 1 unit change in equipment reliability leading to a 21.9% increase in the number of made to

order jackets produced in a fortnight. AS (25.2% increase) and SL (24.3% increase) fell between ER and WE in terms of unstandardized results. Finally, the reported θ parameter value was small. For the NBR, the θ shows the extent to which the model had "excess" variation over and above a Poisson as the variance of the negative Binomial distribution is taken to be $\mu + \frac{\mu^2}{\theta}$. Hence, large θ values shrink the variance toward the Poisson (i.e., where μ describes both the mean and variance) and small values allow for variance over and above what would be expected of a Poisson.

The standardized coefficients for the NBR reported in Table 3 showed the same pattern as those results obtained by the PR model in Table 2. In both models ER resulted in the largest value followed by WE, then SL, and then AS.

The PR and NBR modeling results reported in Tables 2 and 3 provide the key inferential and predictive model results needed to interpret the simulated data. In the section below I proceed to using both models in a DA. For each model, I collect all possible R^2_{DEV} values, report all dominance statistics, and make all dominance designations. In addition, I compute a selection of the reported dominance statistics in the course of describing each models' results.

Dominance Analysis Results DA, as well as Shapley value decomposition, requires that R^2 values from all possible combinations of the IVs as included and excluded from the model are collected[†]. The four IVs used in this manuscript result in $2^4 = 16$ combinations. Because one combination includes the condition where no IVs are included, which will produce a 0 R^2_{DEV} value, it is omitted below in Table 4.

Table 4: All Subsets Deviance R-square Results Across Models

	Poisson/SJ	Negative Binomial/MJ
ER	0.1704	0.1845
AS	0.1418	0.1519
SL	0.1008	0.1051
WE	0.1384	0.1483
ER + AS	0.2209	0.2455
ER + SL	0.2388	0.2665
ER + WE	0.2530	0.2847
AS + SL	0.2047	0.2285
AS + WE	0.2200	0.2454
SL + WE	0.1816	0.2007
ER + AS + SL	0.2748	0.3127
ER + AS + WE	0.2821	0.3212
ER + SL + WE	0.2868	0.3272
AS + SL + WE	0.2524	0.2866
ER + AS + SL + WE	0.3113	0.3584

Table 4 shows that, on average, the R^2_{DEV} values for the NBR model, like their coefficient values, were larger than for the PR model. The last entry for each model in

[†]Several implementations of Shapley value decomposition in machine learning approximate this process and do not collect results from all combinations for computational feasibility.

Table 4 is the model R^2_{DEV} for the PR in Table 2 (i.e., .3113) and NBR in Table 3 (i.e., .3584).

These series of 15 values are applied using Equation 2 to obtain general dominance statistic values for each IV. The general dominance statistic values for each model is reported in Table 5.

Table 5: General Dominance Statistics

	Poisson/SJ	Negative Binomial/MJ
ER	0.1048	0.1206
AS	0.0726	0.0833
SL	0.0570	0.0657
WE	0.0769	0.0888
Overall	0.3113	0.3584

The results reported in Table 5 show that the rank order of/the general dominance designations among the IVs is identical across models. Both models show that ER generally dominates each other IV, followed by WE generally dominating SL and AS, and AS generally dominating SL. This pattern of results is similar to but does not match those from the standardized coefficients reported in Tables 2 and 3. Specifically, the standardized coefficients overestimate SL's impact relative to AS and produce results, as will be seen below, that are more similar to each IV's conditional dominance statistics at subset size 4.

General dominance statistic computation involves applying Equation 2 to the values in Table 4. Equation 2 is, however, abstract and the process of translating individual R^2_{DEV} values into general dominance statistics can be hard to follow. Below I provide an example of how the general dominance statistic for the ER variable is computed. The below is a conceptual schematic of the ER general dominance statistic and could be applied to either the PR or NBR values in Table 4. The schematic is the following combination of R^2_{DEV} values:

$$\begin{aligned}
 GDS_{ER} = & \frac{R^2_{DEV_{ER}}}{1 * 4} + \frac{R^2_{DEV_{ER+AS}} - R^2_{DEV_{AS}}}{3 * 4} + \frac{R^2_{DEV_{ER+SL}} - R^2_{DEV_{SL}}}{3 * 4} + \\
 & \frac{R^2_{DEV_{ER+WE}} - R^2_{DEV_{WE}}}{3 * 4} + \frac{R^2_{DEV_{ER+AS+SL}} - R^2_{DEV_{AS+SL}}}{3 * 4} + \\
 & \frac{R^2_{DEV_{ER+AS+WE}} - R^2_{DEV_{AS+WE}}}{3 * 4} + \frac{R^2_{DEV_{ER+SL+WE}} - R^2_{DEV_{SL+WE}}}{3 * 4} + \\
 & \frac{R^2_{DEV_{ER+AS+SL+WE}} - R^2_{DEV_{AS+SL+WE}}}{1 * 4}
 \end{aligned}$$

These 8 terms, when summed, produce the .1048 value for the PR when its values are applied or .1206 value for the NBR when its values are applied. To get general dominance statistics for other IVs, the computations are re-arranged such that the R^2_{dev} values in the minuend (i.e., fist value in the difference) for each computation

contains the focal IV and the R_{dev}^2 values in the subtrahend (i.e., the second value in the difference) contains all the other IVs in the model.

One point of note about this computation is that the general dominance statistic is a weighted average of all the different subsets' R_{DEV}^2 values from either model. Note also that these terms are weighted such that their contribution is balanced by subset size. That is, the average is such that values at subset size 1 do not contribute less to the statistic than those values at subset size 2 despite there being more terms from subset size 2 than 1. Because there are three subset size 2 terms to put into the statistic, each of the subset size 2 terms are "downweighted" such that the average of the values of subset size 2, as opposed to each individual value, is used in the general dominance statistic. As will be discussed further below, these subset size averages are conditional dominance statistics.

The weighted averaging scheme imposed by DA to produce general dominance statistics derives directly from Shapley value decomposition. The weighting scheme used by DA is a simplification of Shapley value decomposition's scheme that focuses on the key difference between the methods: Shapley values are order dependent whereas general dominance statistics are not. Shapley values differentiate between all possible *permutations* of IV inclusion orders when computing IV contributions to the model and apply an arithmetic average of these contributions to compute the values for that IV. Thus, in Shapley value decomposition computing the value for ER would require averaging the increment ER makes to the R_{DEV}^2 over $4! = 4 * 3 * 2 * 1 = 24$ model IV inclusion sequences.

The Shapley value decomposition method, when applied to statistical models PR and NBR, makes the strong assumption that the strict order of inclusion of the IVs matters over and above the unordered combination of IVs included at any step. That is, Shapley values assume that it is useful to distinguish between the order of inclusion ER, AS, SL, and WE from the order of inclusion ER, AS, WE, and SL for the contribution ER makes to the model. Although Shapley values differentiate between these two orderings, ER is the first IV in the model for both and will have an identical increment to prediction across orderings. In fact, in all 6 IV inclusion order permutations where ER is the first IV to enter the model the value will be identical. Shapley value decomposition then effectively re-uses information about the model during computation.

Table 4 shows that there are really only 15 unique R_{DEV}^2 that can be obtained from the model and thus only 15 models, as opposed to 24, that need to be estimated from the data. Because Shapley values re-use information about the model, the Shapley value decomposition computation in Equation 1 can be written as the average of 24 computations where different elements of the model are re-used a specific number of times. When applied to computing ER's Shapley value decomposition across the 24 inclusion permutations, 6 permutations put ER as the first IV included and 6 permutations put ER as the last IV included. The six other combinations are possible in the data, both represented twice in the 24 permutations. These 8 total terms mirror those used in computing the general dominance statistics and can be written as:

$$SV_{ER} = 24^{-1} \{ 6(R_{DEV_{ER}}^2) + 2(R_{DEV_{ER+AS}}^2 - R_{DEV_{AS}}^2) + 2(R_{DEV_{ER+SL}}^2 - R_{DEV_{SL}}^2) + 2(R_{DEV_{ER+WE}}^2 - R_{DEV_{WE}}^2) + 2(R_{DEV_{ER+AS+SL}}^2 - R_{DEV_{AS+SL}}^2) + \dots \}$$

$$2(R_{DEV_{ER+AS+WE}}^2 - R_{DEV_{AS+WE}}^2) + 2(R_{DEV_{ER+SL+WE}}^2 - R_{DEV_{SL+WE}}^2) + 6(R_{DEV_{ER+AS+SL+WE}}^2 - R_{DEV_{AS+SL+WE}}^2)\}$$

When multiplying through by $24^{-1} = \frac{1}{24}$, the increments repeated 6 times have a $\frac{6}{24} = \frac{1}{4}$ weight where the increments repeated 2 times have a $\frac{2}{24} = \frac{1}{12}$ weight matching the general dominance statistic computation. Thus, the weighting scheme used by general dominance statistics is a linear transformation of Shapley values that equally weights different IV inclusion permutations.

The general dominance statistics are an arithmetic average of the conditional dominance statistics. Those conditional dominance statistics are also computed in this section using Table 4's values with Equation 3. Conditional dominance statistic values are reported, in graphical format, in Figure 1. The graphic format is useful with conditional dominance statistics as their values are less useful to interpret directly than general dominance statistics that are interpretable as parts of the R_{DEV}^2 . Rather conditional dominance statistics are useful to interpret as a *trend* across subset sizes. When an IV conditionally dominates another IV, its conditional dominance statistic trendline will always be above the other IV across the entirety of the graph. If an IV's trendline ever meets or slips below the other IV's trendline, conditional dominance between the two IVs cannot be determined. The graphical format simplifies thus simplifies conditional dominance designation among IVs. For this graph, the y-axis is also on a log scale to improve legibility nearer subsets sizes of 4.

Figure 1: Conditional Dominance Statistics

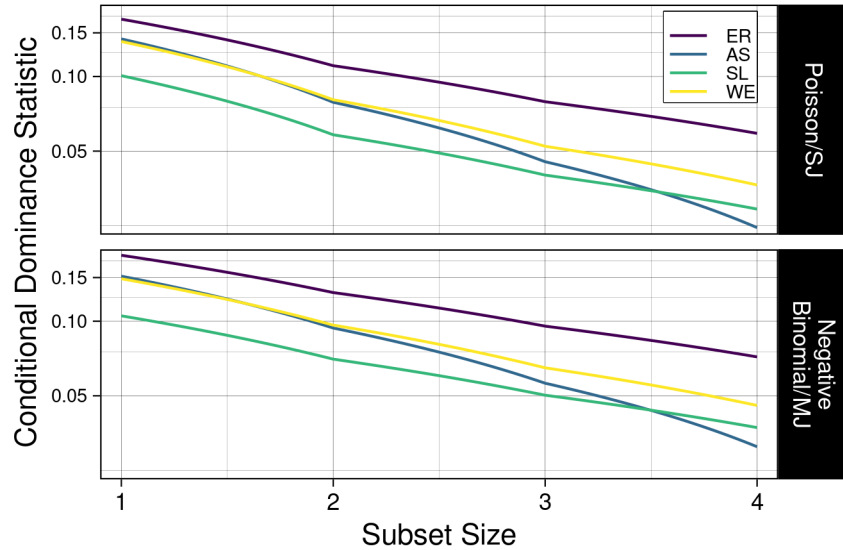


Figure 1's results showed that a few conditional dominance designations were possible for both PR and NBR models. ER conditionally dominated AS, SL and WE. In addition, WE conditionally dominated SL. By contrast, no conditional dominance relationships were possible with AS across both PR and NBR models as, for both

models, AS starts with the second largest value at subset size of 1 and ends with the smallest value at subset size of 4. In part, this complex relationship has to do with AS's generally high correlations with other IVs. At lower subset sizes, AS shows a larger value as its coefficient value is relatively large and its overlap with other IVs has not been taken into account. At higher subset sizes, AS's larger correlation with other IVs makes the increment it adds to prediction smaller as it overlaps with other IVs to a large extent. Thus, in combination, AS's pattern of intercorrelations produces a complex conditional dominance statistic trend.

The example computation using ER above is extended here to compute conditional dominance statistics. The four conditional dominance statistics for ER are computed as:

$$\begin{aligned}
 CDS_{ER}^1 &= R_{DEV_{ER}}^2 \\
 CDS_{ER}^2 &= \frac{R_{DEV_{ER+AS}}^2 - R_{DEV_{AS}}^2}{3} + \frac{R_{DEV_{ER+SL}}^2 - R_{DEV_{SL}}^2}{3} + \\
 &\quad \frac{R_{DEV_{ER+WE}}^2 - R_{DEV_{WE}}^2}{3} \\
 CDS_{ER}^3 &= \frac{R_{DEV_{ER+AS+SL}}^2 - R_{DEV_{AS+SL}}^2}{3} + \frac{R_{DEV_{ER+AS+WE}}^2 - R_{DEV_{AS+WE}}^2}{3} + \\
 &\quad \frac{R_{DEV_{ER+SL+WE}}^2 - R_{DEV_{SL+WE}}^2}{3} \\
 CDS_{ER}^4 &= R_{DEV_{ER+AS+SL+WE}}^2 - R_{DEV_{AS+SL+WE}}^2
 \end{aligned}$$

Each of the conditional dominance statistics are, again, components of the general dominance statistic computed above.

Conditional dominance is a pre-requisite for the final designation between two IVs, complete dominance. Hence, it is only necessary to attempt to determine complete dominance for IVs where there is a conditional dominance designation. Recall that there were 4 conditional dominance designations above: WE conditionally dominated SL and ER conditionally dominated all other IVs. The results in Table 4 show that all four IV comparisons produce complete dominance designations for the PR and NBR models.

Complete dominance is determined by comparing R_{DEV}^2 from individual, comparable models. For the ER to AS determination this involved $R_{DEV_{AS}}^2$ compared to $R_{DEV_{ER}}^2$, $R_{DEV_{AS+SL}}^2$ compared to $R_{DEV_{ER+SL}}^2$, $R_{DEV_{AS+WE}}^2$ compared to $R_{DEV_{ER+WE}}^2$, and $R_{DEV_{AS+SL+WE}}^2$ compared to $R_{DEV_{ER+SL+WE}}^2$. For each of these four comparisons, the model containing ER was greater than the model containing AS.

In conclusion, the DA results provide a set of dominance designations between the IVs indicating different levels of importance between them. AS was, overall, the most important variable in terms of explaining the DV as reflected in the E_{DEV}^2 across PR and NBR models. ER completely dominated AS, SL and WE. WE completely dominated SL, and generally dominated AS. Finally, AS generally dominated SL.

Section Summary This section was intended to provide an extensive analytic of how to apply the recommended DA methodology using the R^2_{DEV} to PR and NBR models. For these examples, estimated and interpreted model coefficients, their DA statistics, as well as provided schematics related to the computation of DA statistics.

The data used in the modeling was simulated but was generated to be realistic and interpreted as though it was obtained empirically. As such, the data analytic examples above provided a valuable overview of recommended practice for computing DA statistics and determining DA designations with count DVs.

The DA methodology discussed in the sections above grows more complicated when the count data meet specific conditions. One condition discussed in the introduction is the condition where the observations in the data are "exposed" to the count generating mechanism differently. The section below shifts to discussing the conceptual underpinnings of these different exposure times for modeling and proceeds to discuss how such exposure times affect DA statistics and designations using an extension of the simulated data.

Additional Consideration for CRMs

CRMs necessarily differ from other linear models as they have an implied level of aggregation that is associated with them. In fact, one interpretation of a Poisson distribution is as accumulated repeated sampling from a Binomial distribution with a set number of trials (cite). In this interpretation of the Poisson, the Poisson mean parameter μ is the product of the number of trials n for the Binomial distribution times the probability p of the event occurring in the Binomial distribution. Given the Binomial interpretation of the Poisson, it is clear that CRMs assume that the counts of events realized in the data are derived from Binomial processes that share the same number of trials and probabilities of occurrence. That is, CRMs assume that the realized counts stem from the same underlying Binomial processes.

CRMs are able to accommodate adjustments where different Binomial processes are observed. Below is a conceptual discussion of what these adjustments are, why they are important to apply, and how they affect DA statistics.

Exposure Variables and Offset Terms

The concept of exposure to a count generating process, as well as its importance to generating observed values, is easier to explain in the context of an example. Consider the following research question using a count DV. Does additional customer service training increase the number of positive reviews for a business? When thinking about research questions that are answered using CRMs, the Binomial interpretation of the Poisson is a convenient way to describe the parameter estimate results and how they are interpreted. A key interpretive question for this research is whether we believe that the training will affect the probability of receiving a positive review, the number of customers who can provide a review, or both factors. For this research question, the customer service training is likely to focus on the probability of receiving a positive review. Conceptually, the customer service training should, with a fixed set of customers, increase the count as the probability that any one customer is happy and

thus provides a positive review is increased. These questions, focused on probability of the event occurring, are common in application to CRMs and are the focus of this discussion.

Probability-oriented count DV questions are complicated by the nature of the data generation process as the same probability can produce different numbers of observed events with different numbers of trials or different levels of "exposure" to the probability. For example, two business locations might have 10 positive reviews each but one may have had 1,000 customers and the other 100 customers. Both observations yield the same count value but imply starkly different probabilities of observing an event (i.e., $\frac{10}{1000} = .01$ and $\frac{10}{100} = .1$). In this case, the number of customers is a nuisance factor that obscures the underlying probability of occurrence of the sought after event.

CRMs can control for such nuisance/unequal exposure to the count generating mechanism using what is known as an *offset* term. The offset term is usually assumed to be natural log transformed version of the exposure variable and enters into the CRM with a coefficient of 1. The coefficient of 1 results in the count DV being transformed into a rate dependent on values of the exposure variable. How the count DV is transformed into a rate with a coefficient of 1 follows from the following algebraic manipulations. Imagine a simple case where an intercept only model is fit with an offset. In a log-linear CRM, this produces: $y = e^{offset + \beta}$. Recall that the offset term is the natural log of a variable, say, o . Applying the natural log and substituting $\ln o$ for *offset*, the prior equation can be written as $\ln y = \ln o + \beta$. Re-arranging the $\ln o$ term results in $\ln y - \ln o = \beta$ which, given the properties of logarithms, is tantamount to $\ln \frac{y}{o} = \beta$. Therefore, the inclusion of a natural log transformed exposure variable turns the count DV into a rate that depends on exposure as reflected by the offset term.

The re-scaling by the offset term adjusts for differences in trials between observations and better reflects the underlying probability which is desirable for probability-based research questions with CRMs.

...ended here...

The key role of the offset term is to re-scale the intercept of the CRM to be a rate value given the offset term as opposed to a mean count value. The re-scaling process reduces the extent of unexplainable variation, or deviance, in the count DV as observations with high counts and high exposure values are pulled nearer those with low counts and low exposure values. As a result, the offset term will affect R^2_{DEV} values and DA statistics on which they are based. An additional important condition to consider is a situation where the exposure variable is causally related to IVs in the model.

The sections below extend on the modeling above by describing the generation of two new variables that have different patterns of exposure to the count generating process. The modeling process above is then repeated on these new variables to demonstrate the effect that exposure variables have on both modeling as well as DA results.

Dominance Analysis with Offset Terms: Analytic Examples

The data generation ...

Multiple Equations

Finally, a common model applied to CDVs are *zero inflated* models that are recommended for use in modeling CDVs in many situations (see Figure 3 of (Blevins, Tsang & Spain, 2015)). Zero inflated models offer a great deal of flexibility in evaluating the processes

The flexibility of zero inflated models comes at the cost of greater complexity when considering how to evaluate the contributions IVs have to prediction as there are two predictive equations, the count-producing model and the zero-producing model, which need not have the same set of predictors. As such, it may be necessary to examine parameter estimate relative importance (PERI; (Luchman, Lei & Kaplan, 2020)) as opposed to independent variable relative importance (IVRI) when examining

Discussion

The purpose of this manuscript was to ...

References

- Azen, R. & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological methods*, 8(2), 129.
- Azen, R. & Traxel, N. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics*, 34(3), 319–347.
- Bettinazzi, E. L. & Feldman, E. R. (2021). Stakeholder orientation and divestiture activity. *Academy of Management Journal*, 64(4), 1078–1096.
- Blevins, D. P., Tsang, E. W. & Spain, S. M. (2015). Count-based research in management: Suggestions for improvement. *Organizational Research Methods*, 18(1), 47–69.
- Cameron, A. C. & Windmeijer, F. A. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics*, 14(2), 209–220.
- Glerum, D. R., Joseph, D. L., McKenny, A. F. & Fritzsche, B. A. (2021). The trainer matters: Cross-classified models of trainee reactions. *Journal of Applied Psychology*, 106(2), 281.
- Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, 61(2), 139–147.
- Johnson, J. W. & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational research methods*, 7(3), 238–257.
- Luchman, J., Lei, X. & Kaplan, S. (2020). Relative importance analysis with multivariate models: Shifting the focus from independent variables to parameter estimates. *Journal of Applied Structural Equation Modeling*, 4(2), 1–20.
- Luchman, J. N. (2014). Relative importance analysis with multicategory dependent variables: An extension and review of best practices. *Organizational Research Methods*, 17(4), 452–471.
- Luchman, J. N. (2021). Determining relative importance in stata using dominance analysis: domin and domme. *The Stata Journal*, 21(2), 510–538.

- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. & Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1), 56–67.
- McCullagh, P. & Nelder, J. A. (2019). *Generalized linear models*. Routledge.
- Naumovska, I., Zajac, E. J. & Lee, P. M. (2021). Strength and weakness in numbers? unpacking the role of prevalence in the diffusion of reverse mergers. *Academy of Management Journal*, 64(2), 409–434.
- Soda, G., Mannucci, P. V. & Burt, R. S. (2021). Networks, creativity, and time: Staying creative through brokerage and network rejuvenation. *Academy of Management Journal*, 64(4), 1164–1190.
- Somaya, D., Williamson, I. O. & Lorinkova, N. (2008). Gone but not lost: The different performance impacts of employee mobility between cooperators versus competitors. *Academy of Management Journal*, 51(5), 936–953.
- Tonidandel, S. & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology*, 26(1), 1–9.