Relative importance analysis for count regression models

Journal Title
XX(X):1–18

DOI: 10.1177/ToBeAssigned
www.sagepub.com/

SAGE

Joseph N. Luchman[1]

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

[1]Fors 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 *incidince 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, $R^2_{IV_x IV_z} = R^2_{IV_z IV_x}$). 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 an 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 a statistic known as a *conditional dominance statistic* that is closely related to the general dominance statistic. The conditional dominance statistic for an IV (IV_x) is:

... ended here ... describe the CDS - confirm the # of combos/computation

$$CDS_{IV_x} = \sum_{l=1}^{C_{U_l}} \frac{F_{U_l \cup IV_x} - F_{U_l}}{C_{U_l}} \tag{3}$$

In fact, DA produces three different results that it uses to compare the contribution each IV makes in the predictive model against the contributions attributed to each other IV. The use of these three results to compare IVs is the reason DA is an extension of Shapley value decomposition.

Complete dominance between two IVs is designated by:

$$X_v D X_z \; if \; 2^{p-2} = \Sigma 2 \tag{4}$$

Where X_v and X_z are two IVs, S_j is a distinct set of the other IVs in the model not including X_v and X_z which can include the null set (...) with no other IVs, and F is a model fit statistic. Conceptually, this computation implies that when all 2^{p-2} comparisons show that X_v is greater than X_z, then X_v completely dominates X_z.

In the section below, I transition to discussing some of the nuances of CDVs for the application of DA.

Applying Dominance Analysis to Count Regression Models

As a variant of Shapley Value Decomposition, DA is a model agnostic method for decomposing a fit statistic from a statistical model into components that are attributable to IVs. Although DA may be conceptually applicable to CRMs, at current there is no guidance in the literature as to how to implement DA using CRMs. A discussion of how to apply DA to CRMs is useful as CRMs more complex than LRM given their log-linear form.

Log-linear Form and Multiplicative Effects

The effect of the log-linear link function to the underlying linear model has the effect of resulting in a multiplicative relationship between an IV and the DV such that a one unit change in the IV results in a change in the DV that is the anti-log or exponential function (i.e., e^{β}) of the coefficient. The effect of the log-linear form is that predicted counts from a CRM produced by a change in the IV are sensitive to initial conditions—a the magnitude of change realized from a CRM's coefficient depends on the location of the DV prior to the change.

The effect the log-linear model has on changes to predicted values are clearer in the context of an example. Consider first a LRM with the following form: $.2 + .5 * x_1 + -.8 * x_2$. Thus, the intercept is .2, x_1's coeffiecient is .5 and x_2's coeffiecient is -.8. Two observations in the data have values of 4 and 8 for x_1 and .2 and -.2 for x_2 respectively. The predicted values for observation one is then $.2 + .5 * 4 + -.8 * .2$ or 2.04 and for observation 2 is $.2 + .5 * 8 + -.8 * -.2$ or 4.36. If x_1 changes by a value of 1 for each observation, the LRM model will produce predicted values of $.2 + .5 * 5 + -.8 * .2$ or 2.54 for observation 1 is and $.2 + .5 * 9 + -.8 * -.2$ or 4.86 for observation two. Hence, given the coefficient of .5, a one unit change for both observations on x_1 results in .5 change in the DV for both observations.

Now consider what occurs if the same equation was applied to a log-linear model. Before the one unit change on x_1, the log-linear model would produce predicted values of $e^{.2+.5*4+-.8*.2}$ or $\approx$ 7.69 for observation one and for observation two is $e^{.2+.5*8+-.8*-.2}$ or $\approx$ 78.26. Already the log-linear results produce very different results compared to the LRM even with the same underlying predictive equation. The differences also extend to changes in IVs. Assume again a one unit change to x_1. With the log-linear formulation observation one's predicted value is $e^{.2+.5*5+-.8*.2}$ or $\approx$ 12.68 and observation two's predicted value is $e^{.2+.5*9+-.8*-.2}$ or $\approx$ 129.02. As can be seen, the change to predicted values was not uniform as it was with the LRM. Observation one's change was $\approx$ 5 whereas observation two's change was $\approx$ 50.

As was mentioned at the outset of this section, log-linear models like CRMs have multiplicative effects and the expected magnitude of the predicted value is the anti-log of the prediction equation and, by extension, each coefficient. The expected effect of the .5 coefficient associated with x_1 then $e^{.5} \approx 1.65$. This value, known as the *incidence rate ratio*, describes the percent by which the effect is multiplied for each one unit increase in the IV. This value can be recovered by comparing the rates of change for both observation one ($\frac{12.68}{7.69} \approx 1.65$) and observation two ($\frac{129.02}{78.26} \approx 1.65$).

In addition to being more complex to interpret, the multiplicative effects produced by CRMs complicate fit metric computation. The explained variance R^2 (i.e., R^2_{EXP}) commonly used with LRM assumes that the underlying model seeks to minimize the sum of the squared residuals from the model. As will be discussed below, CRMs do not which can make the most common model fit metric used in DA, the R^2_{EXP}, a poor fit metric choice for determining importance in CRMs.

Choosing a Fit Metric for Count Regression Model Relative Importance

A key consideration for implementing DA in any model is choosing an appropriate fit metric that accurately reflects the models' predictive capability and the form of the DV being predicted. The model fit metric that has received the most attention in past research on DA is R^2_{EXP}. The R^2_{EXP} is the squared Pearson correlation between the DV and the predicted values from the model. Pearson correlations are based on the sums of squared deviations between the two variables' values in the data. Similarly, the LRM is based on a least-squares criterion that seeks to minimize $\sum(y - \hat{y})^2$ or the residual sums of squares between the predicted values from the LRM and the observed

DV. The similarity in minimization goal of LRM and computation of the R^2_{EXP} then makes R^2_{EXP} a useful model fit metric for LRM.

By contrast to LRM, the minimization criterion for PR is:

$$\ln L = \sum \hat{y}y - \ln(y!) - e^{\hat{y}}$$

When considering parameter estimation, the equation can be simplified to $\sum(\hat{y} - y \ln \hat{y})$ (Cameron & Windmeijer, 1996) which shows that PR depends on the sum of deviations from the product of the predicted value from the natural log of the predicted value and the DV. This minimizing criterion can deviate from the least-squares criterion, especially in small samples, and although an extensive discussion is not presented here, NBR shows similar, if not more, substantial deviations from the least-squares criterion as it is a direct extension of the Poisson (i.e., a mixture of Poisson and Gamma distributions (Blevins, Tsang & Spain, 2015)).

The recommended model fit metric to use for DA of CRMs borrows from the comprehensive review and critique of different fit metrics for CRMs offered by Cameron and Windmeijer (Cameron & Windmeijer, 1996). Cameron and Windmeijer conclude that the R^2 that best reflects fit to the data for CRMs is the deviance R^2 or R^2_{DEV}. The deviance R^2 is computed as:

$$R^2_{DEV} = \frac{D_{model}}{D_{null}}$$

Where D is the GLM deviance residual computation.

It is also worth noting that the R^2_{DEV} can be applied to other models for which a deviance function can be computed such as the logistic regression model and LRM. When applied to the logistic regression model, the way the model is scaled is such that it is tantamount to the McFadden R^2 recommended by Axen and Traxel (Azen & Traxel, 2009) in previous work on DA. Moreover, the R^2_{DEV}, when computed on a Gaussian generalized linear model, is tantamount to the R^2_{EXP}. Thus, the R^2_{EXP} is a choice of model fit metric for the DA of CRMs that comes the closest to both the well-known R^2_{EXP} and recommendations from allied work focusing on other generalized linear model distribution families.

The next section extends on the discussion above by providing extensive data analytic examples of DA as applied to CRMs.

Data Analytic Examples of Dominance Analysis

Throughout this manuscript I have highlighted similarities and differences between LRM and CRMs in terms of parameter estimation, predicted value generation, and model interpretation. Despite their estimation and interpretation differences, the process for determining importance using DA for LRM and CRMs is identical. This section provides an extensive and illustrative example of the process for determining importance for both the familiar and well-examined LRM and for two extensions to novel DA computations using PR. An example using NBR models is not discussed directly in the manuscript but is included in supplemental materials. I have chosen to not include NBR models as considerations for those models in terms of how to apply DA are identical to PR models. In addition, negative Binomial distributed data is more

complex to generate and describe. The focus of all data analytic examples will then be focused on Poisson distributed variables.

The intention of showing how to determine importance for each model is to highlight their similarities in terms of the underlying computational process and to provide an example of how to interpret DA with the recommended CRM R^2_{DEV} fit metric. For these examples, I generate data for a Gaussian and a Poisson DV in the section below. These two DVs are generated in such a way as to make them as similar as possible to one another save for their distribution form. In making these three DVs maximally similar to one another, the reader will more clearly be able see how DA with the LRM corresponds with DA as applied to an analogous version of the analysis in PR.

The data generation process for this data analytic example is described in the next section.

Data Generation One hundred thousand simulated observations in this manuscript was generated from a data structure with four IVs and a single relationship to a DV. The simulated data for the four IVs was generated using a single set of draws from a random data generators with pre-specified population parameters. In developing a dataset for this illustration, I approached the specification of population parameters in a way that did not intentionally provide advantages to any single IV. Any advantage in terms of explaining variance was intended to be introduced by the random number generator. Using the random number generator to introduce advantages for different IVs is intended to make the data used in this illustration more like real data and yet reproducible and with a known structure.

The population parameter set used to generate the 100,000 simulated observations produced a set of IV variance and IV-DV regression slope parameters that produced the same R^2 value assuming the IVs were not correlated. The four sets of variance-slope values were: $\sigma^2_{IV_1} = 1$, $\beta_{IV_1} = .2$, $\sigma^2_{IV_2} = 1.8$, $\beta_{IV_2} = .15$, $\sigma^2_{IV_3} = 2.8$, $\beta_{IV_3} = .12$, and $\sigma^2_{IV_4} = 4$, $\beta_{IV_4} = .1$. Given that the variance explained for a LRM with no intercorrrelaions can be computed as $R^2_{IV} = \sigma^2_{IV}\beta^2_{IV}$ or the product of the IV's variance and the squared slope, all four IVs have an R^2 of .04.

Correlations were also added between the IVs as there is little advantage to using DA in cases where IVs are uncorrelated and most data collected in real studies are intercorrelated to some degree. The population parameter correlations added between the IVs intended, again, to not unduly give advantage to any one IV. All IV interrelationships took one of three values .5, .25, and .125 and all IV pairs had one value. The six different correlations specified were: $r_{IV_1IV_2} = .5$, $r_{IV_1IV_3} = .25$, $r_{IV_1IV_4} = .125$, $r_{IV_2IV_3} = .125$, $r_{IV_2IV_4} = .25$, and $r_{IV_3IV_4} = .5$. These six intercorrelations are balanced such that no one IV overlaps more or less with other IVs.

The combination of variances and correlations were combined into a single matrix denoted as the *Parameter* matrix in Table X. The Parameter matrix was used, along with a degree of freedom parameter of 7, as a scale matrix in a draw from a random inverse Wishart distribution. A value of 7 was chosen as the degrees of freedom for the sample covariance matrix to allow for a rather wide sampling of covariances values to make the process less deterministic but somewhat less wildly varying than the maximum variability (a value of 4 or the dimension of the scale matrix). As the degrees of

freedom increases from the minimum of 4, the values sampled from an inverse Wishart distribution converge more closely with the scale matrix between draws. The sampled covariance matrix obtained is reported in Table X1 denoted as *Sampled*.

	IV1	IV2	IV3	IV4
Parameter				
IV1	1.000000000	0.67082039	0.41833001	0.250000000
IV2	0.670820393	1.80000000	0.28062430	0.670820393
IV3	0.418330013	0.28062430	2.80000000	1.673320053
IV4	0.250000000	0.67082039	1.67332005	4.000000000
Sampled				
IV1	0.091590095	0.06153029	-0.03506662	0.002963387
IV2	0.061530289	0.28517891	-0.12537445	-0.356761781
IV3	-0.035066620	-0.12537445	0.71795109	-0.095787710
IV4	0.002963387	-0.35676178	-0.09578771	1.311862465

The sampled covariance matrix in Table X1 was used as the covariance matrix, with a vector of 0 means, in a multivariate Normal distribution obtaining 100,000 draws.

The simulated IV values were used to generate the three DVs. The Gaussian/Normally distributed DV was generated first. The coefficient vector implied above was entered as means, into a Normal distribution with a standard deviation of .2 from which a single draw from each mean was obtained. The standard deviation of .2 was intended to make the process of generating coefficients for the DVs less deterministic. The vector of coefficient draws resulted in the following set of values: $\beta_{IV_1} = 0.259$, $\beta_{IV_2} = 0.042$, $\beta_{IV_3} = 0.112$, $\beta_{IV_4} = 0.293$. These four values were multiplied by their IV to generate an initial set of Normally distributed DV values. The Normally distributed DV Y_{Normal} was completed by adding a draw from a Normal distribution with a mean of 0 and a standard deviation that was computed as $\sqrt{1 - \sigma^2_{Y_{NormalNoError}}}$ using the initial Normally distributed variable without error.

A Poisson distributed version of Y_{Normal}, $Y_{Poisson}$, was generated by obtaining the cumulative probability of each observation's Y_{Normal} score given the standard Gaussian/Normal distribution (i.e., $N(0, 1)$) and translating that value into a count value using the Poisson quantile function assuming an underlying mean and variance of 1. The mean of 1 value was chosen to make the Poisson distributed result as close to the Normally distributed result as possible as both have variances of 1.

The means, standard deviations, and correlations between all four IVs and two DVs is reported below in Table X2.

	Means	SDs	IV1	IV2	IV3	IV4	Y_Norm
IV1	0.001330068	0.3039554	1.000000000	0.3809086	-0.13690020	0.005635913	0.078031
IV2	-0.001447403	0.5344685	0.380908600	1.0000000	-0.27832116	-0.585105979	-0.168605
IV3	0.003514622	0.8471215	-0.136900200	-0.2783212	1.00000000	-0.094324027	0.048287

IV4	0.007954359	1.1458255	0.005635913	-0.5851060	-0.09432403	1.000000000	0.313719
Y_Normal	0.002088013	0.9984605	0.078031061	-0.1686056	0.04828792	0.313719350	1.000000
Y_Poisson	1.000140000	0.9999350	0.070678555	-0.1533697	0.04471122	0.285672719	0.912707

There are a few points of note about the descriptive results from Table X2. Note the wide variability in the sampled intercorrelations among the four IVs which, in a number of cases for IV_3 and IV_4 are negative as sampled despite being based on positive correlations in the population. The standard deviations for the four IVs also show variability from their population values though generally maintain a pattern of $\sigma_{IV_4} > \sigma_{IV_3} > \sigma_{IV_2} > \sigma_{IV_1}$. In addition, the magnitude and direction of several of the correlations may be surprising to the reader given the known coefficient values that produced them. For example, the correlations between IV_2 and all three DVs is negative despite IV_2's known positive coefficient. The negative correlation is due to IV_2's strong, negative correlations with IV_3 and IV_4 which both have stronger relationships with the DVs. Correlations reflect a form of *total effect* of an IV on a DV (cite??) and, as will be seen, the biased effect of the correlation does not persist when the model is estimated on the data.

The data for use in the DA illustrations are now generated and ready for use in analysis. DA is a method for extracting additional information from a statistical model such as an LRM. As such, it should be interpreted in light of the model on which it is based. The section below estimates the LRM, as well as PR and NBR, CRMs on their respective DVs using all four IVs.

Regression Results The LRM results predicting Y_{Normal} are reported in Table X3.

	Coefficient	SE	t	p	CI_low	CI_high	Std_Coefficient
$Intercept$	-0.001	0.003	-0.315	0.752	-0.007	0.005	0.000
IV1	0.262	0.011	23.485	0.000	0.240	0.284	0.080
IV2	0.049	0.008	5.726	0.000	0.032	0.065	0.026
IV3	0.116	0.004	29.837	0.000	0.108	0.123	0.098
IV4	0.294	0.004	82.493	0.000	0.287	0.301	0.338

Table X3 shows that the most substantial slope size obtained by any of the IVs is for IV_4 with a value of .294. A one unit change in IV_4 then produces a 0.294 change in Y_{Normal}. Compare IV_4's results with those obtained from IV_2. IV_2 obtained the smallest value for its coefficient with 0.049. A one unit change in IV_2 then produces a 0.049 change in Y_{Normal}. The combination of four coefficients presented in Table X3 resulted in an explained variance R^2 of .1125.

The values obtained by the IVs in Table X3 were, effectively, the population values reported on earlier in the manuscript. This result occurred as the LRM is correctly specified and, as such, the correct, known parameter estimates for each of the IVs is recovered. Thus, as noted above, the potentially surprising results observed of the IV-DV correlations is an artifact of the IV interrelationships and does not carry over to the parameter estimates.

The coefficients in Table X3 are unstandardized which makes their interpretation comparable only when their variances are equal. This is because one unit changes

in IVs differ in their 'typical'-ness and how well they reflect change in an IV. For instance, the standard deviation of IV_4 (1.15) is closest to a value of 1 of all the IVs. The unstandardized coefficient for this IV then describes this variable's effect best in terms of typical changes. By contrast, IV_1 has a standard deviation of 0.30. Thus, a 1 unit change is far from typical and actually describes a nearly 3 standard deviation change in that IV. The comparisons between the coefficients can be improved by standardization where the typical changes are harmonized such that they all reflect one standard deviation of change. When standardized, the results from the LRM show slightly different trends. IV_4 remains strongest with a standardized coefficient of 0.338. For every one standard deviation change in IV_4 there is a 0.338 standard deviation change in Y_{Normal}. The standardized value for IV_4 remained similar to it's value when unstandardized as its variance was near 1. Note however the substantial change to IV_1 which had an unstandardized value very close to IV_4's value but a standardized value (0.080) that is around 30% of its unstandardized value. This is because, as noted above, the typical change for IV_1 is around 0.30; 30% of the value of 1 assumed with an unstandardized coefficient.

The standardized as compared to the unstandardized results in Table X3 show that there are interpretive challenges for this model given the differing variances and coefficient values and attempting to determine which IVs are most important. The results in Table 3 will be revisited below when considering the DA results for the LRM. The CRM results predicting $Y_{Poisson}$ are presented next.

The CRM results for $Y_{Poisson}$ are reported in Table X4.

	Coefficient	SE	z	p	CI_low	CI_high	Std_Coefficient
$Intercept$	-0.049	0.003	-14.889	0.000	-0.056	-0.043	0.000
IV1	0.237	0.012	19.994	0.000	0.213	0.260	0.072
IV2	0.046	0.009	5.141	0.000	0.029	0.064	0.025
IV3	0.107	0.004	25.926	0.000	0.099	0.115	0.091
IV4	0.269	0.004	70.957	0.000	0.262	0.277	0.308

Table X4, like the results from X3, shows that the most substantial slope size obtained by any of the IVs is for IV_4 with a value of 0.269. By comparison with the LRM results, the PR results are not as readily interpretable in their linearized form as it indicates that a one unit change in IV_4 then produces a 0.269 change in the natural logarithm of $Y_{Poisson}$. A more useful way to interpret IV_4's result is in IRR/exponentated form as $e^{.269} \approx 1.309$. The IRR for IV_4 then indicates that a one unit change in IV_4 results in a 30.9% increase in the value of $Y_{Poisson}$. As is discussed above, CRMs are multiplicative models and their results increase in percentages as opposed to strictly additively like the LRM. Again, IV_2 obtained the smallest value for its coefficient with 0.046 or an IRR of $e^{0.046} \approx 1.047$. A one unit change in IV_2 then produces a 4.7% increase in $Y_{Poisson}$. The combination of four IVs in Table X4 produced a R^2_{DEV} of .0814.

The CRM results suffer, like the LRM results, from interpretive challenges due to the variances of the IVs and can be similarly standardized*. The standardized CRM results remain multiplicative but are interpreted in standard deviation change terms. IV_4's value of 0.308 can be interpreted as a one standard deviation change in IV_4 is associated with a $e^{0.308} \approx 1.361$ or 36.1% increase in $Y_{Poisson}$.

Before moving on, as the reader has no doubt observed, $Y_{Poisson}$'s CRM results are, as intended, numerically similar to those obtained from the results for Y_{Normal} LRM. Again, these results were intended to show the strong similarities across models in many aspects that, as I intend to show, carry over to the determination of importance using DA. The most noteworthy difference between the CRM and LRMs is the non-zero intercept value for the PR model which reflect the baseline incidence rate of $e^{-0.049} \approx 0.952$. Because the baseline incidence rate is the incidence rate when each of the IVs is 0, it reflects the value at the means of each of the IVs; the value of nearly 1 indicates that the known mean of $Y_{Poisson}$ is achieved, as expected, when all IVs are at their means.

The LRM and CRM's unstandardized and standardized coefficients have been discussed at length in the preceeding section. The next section transitions to discussing how the DA process is implemented with both the LRM and CRM and interpreting their DA statistics.

Dominance Analysis Results The DA results for both the LRM and CRM are based on the series of R^2 values each associated with different subsets of IVs reported in Table X5.

formula	$R^2_{Y_{Normal}}$	$R^2_{Y_{Poisson}}$
Y ~ IV1	0.0061	0.0044
Y ~ IV2	0.0284	0.0205
Y ~ IV1 + IV2	0.0521	0.0375
Y ~ IV3	0.0023	0.0017
Y ~ IV1 + IV3	0.0096	0.0070
Y ~ IV2 + IV3	0.0284	0.0205
Y ~ IV1 + IV2 + IV3	0.0521	0.0375
Y ~ IV4	0.0984	0.0711
Y ~ IV1 + IV4	0.1042	0.0753
Y ~ IV2 + IV4	0.0988	0.0714
Y ~ IV1 + IV2 + IV4	0.1046	0.0756
Y ~ IV3 + IV4	0.1045	0.0757
Y ~ IV1 + IV3 + IV4	0.1122	0.0812
Y ~ IV2 + IV3 + IV4	0.1076	0.0780
Y ~ IV1 + IV2 + IV3 + IV4	0.1125	0.0814

The R^2 values associated with each model are reported together in this section to, again, highlight the similarities across models in terms of how the IVs explain

*Standardized coefficients for the Poisson model were generated by standardizing the IVs and ...

variability or reduce deviance in the DV. The final value for the LRM and CRM were reported above and reflect the value of the R^2 to be decomposed by the DA procedure for importance determination.

The results in Table X5, when applied to Equation 3?, produce the General Dominance Statistics for the LRM and CRM reported on in Table X6.

	General_Dominance	Percent	Ranks
Normal			
IV1	0.0089	0.0793	3
IV2	0.0171	0.1515	2
IV3	0.0048	0.0424	4
IV4	0.0818	0.7268	1
Poisson			
IV1	0.0064	0.0785	3
IV2	0.0123	0.1509	2
IV3	0.0035	0.0435	4
IV4	0.0592	0.7270	1

There are a series of points of note from the results presented in Table X6. First, IV_4 generally dominated all three other IVs and by a wide margin for both the LRM and CRM results. Moreover, IV_4 has been ascribed nearly 73% of all the explained variance/reduced deviance across both models. This result is not surprising given IV_4's coefficient results and variance which show both a high amount of variability coupled with a large regression coefficient.

Second, one perhaps surprising result is that IV_2, the IV with the smallest coefficient and second smallest variance, obtained the second largest general dominance statistic, generally dominating both IV_1 and IV_3. This result is related directly to the pattern of intercorrelations observed in Table X1. IV_2 has the second largest correlation with the DVs and the strongest observed correlations with the other IVs. As a result of IV_2's overlap with other IVs, it is more difficult to distinguish how much explained variability in each DV *should* be ascribed to IV_2. In fact, the strong correlation IV_2 has with other IVs allows it to 'borrow' other IVs' predictive usefulness as it is unclear how to parse the predicted values produced by IV_2 relative to other IVs as they appear to correspond so strongly in the data. The Shapley Value decomposition method underlying DA tends to split the difference between IVs when it is unclear as to how to subdivide fit statistic's value. It is worth noting that the next set of results, focusing on the conditional dominance statistics, are insightful for showing how IV_2 resulted in being second most important across subsets of models.

Finally, as has been implied in the discussion of the general dominance results, the LRM and CRM based general dominance statistics are similar; in fact the Percentage of the R^2 values explained by each IV are identical to two decimal places. The general dominance statistic results then support my contention that CRM-based DA is a straightforward extension of LRM-based DA using an R^2 that is computed slightly differently but is conceptually identical to the explained variance R^2.

The next section discusses the more conditional dominance statistics; which as sub-components of the general dominance statistics.

The results for the LRM and CRM-based DA conditional dominance statistics are reported in Figure X1. The graphic format is useful with conditional dominance statistics as it makes the trends across number of IVs included in the model, as well as conditional dominance determination, easier to see.

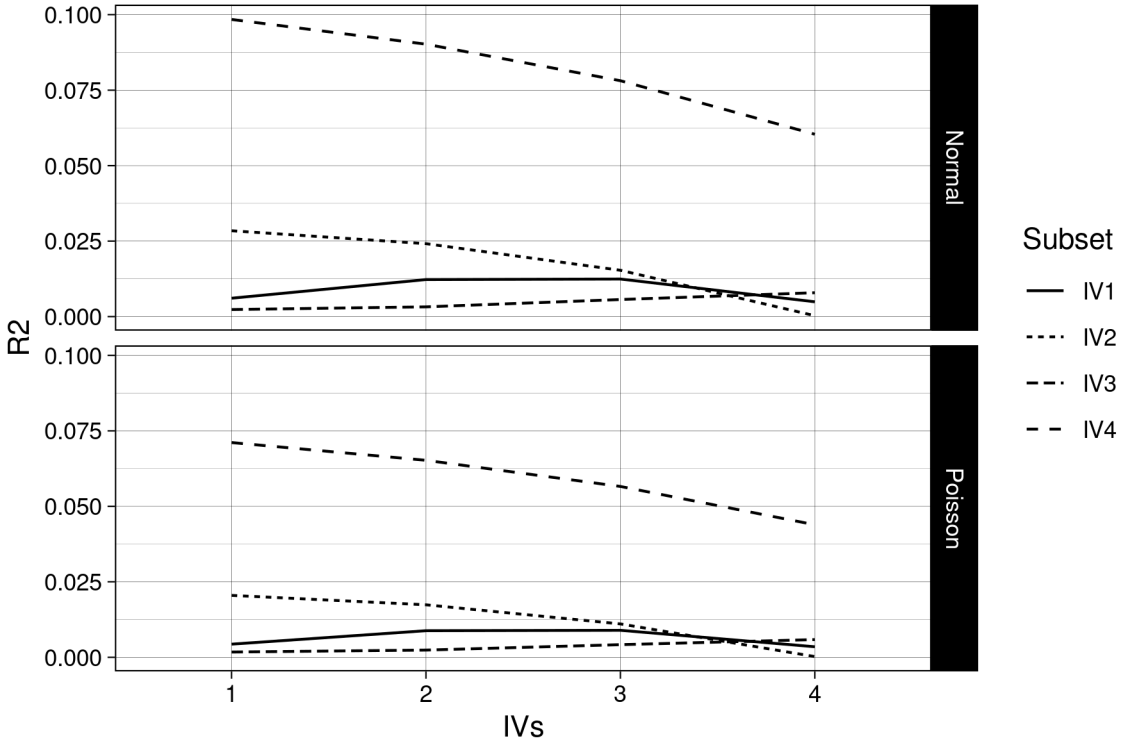

Figure X1 shows that IV_4 conditionally dominates all three of the other IVs. By contrast, the three other IVs have no conditional dominance relationships. This is because the three lines describing IV_1, IV_2, and IV_3's conditional dominance statistics cross over one another between the three IV and all 4 IV in the model.

It is worth noting that the ordering of the conditional dominance statistics at 4 IVs in the model for the LRM and CRM mirror the order of the standardized coefficient magnitudes ... whereas for 3 IVs or less in the model the ordering match the order of the correlations with the DV...

The conditional dominance statistics reinforce the similarity across LRM and CRM DA results as the trends for each IV across numbers of IVs in the model are nearly identical. Again, the difference between LRM-based DA and CRM-based DA is primarily in the choice of fit metric and underlying predictive model; implementation between both is identical and, when the results are designed to be similar, the DA statistics produced are similar.

The R^2 values in Table X5 can also be used to determine complete dominance of IVs over one another. Because IV_1, IV_2, and IV_3 have no conditional dominance designations over one another, they cannot have complete dominance designations. IV_4 does conditionally dominate each of IV_1, IV_2, and IV_3; thus, IV_4 could possible also completely dominate each other IV. The pattern of R^2 values in Table X5 shows that for both the LRM and CRM results, IV_4 completely dominates IV_1, IV_2, and IV_3.

Section Summary This section intended to provide a detailed example of the estimation of both a LRM and CRM as well as DAs based on those models. The examples provided were based on generated data that were produced with the intention of being (reasonably) realistic as well as being as comparable as possible across LRM and CRM results. The data analytic examples of DA presented provide empirical evidence for the theoretical expectation that a LRM and CRM that are based on the same underlying causal structure should produce similar DA results. This demonstration thus serves to illustrate the straightforward generalization of DA from LRM to CRMs.

Although DA with both LRM and CRMs can produce similar results when sharing an underlying causal structure, there are additional complexities to CRM estimation that do not have similar analogs with LRM given the data generating mechanisms underlying count DVs. The discussion of DA with CRMs turns, in next section, to a more extensive discussion of this additional complexity and when it can impact DA statistics and determinations.

The data for the analytic examples were generated using R (R Core Team, 2022) using a combination of base R's `stats` package as well as the `MASS` package (Venables & Ripley, 2002). Other details of data generation for the reproducibility of the analysis can be found in Appendix X. The section below transitions from discussing data generation processes to data analysis.

Additional Consideration for CRMs: Exposure

CRMs, as models of counts of discrete events, implicitly assume that the events realized in the data are derived from a data generating mechanism that had the same number of chances to observe events, or same number of trials, across observations. In CRMs, an observation with a count of 10 events is assumed to have had the same chance to get that count of 10 events as any other observation with 10 events. The problem that arises when the number of chances to observe events is different across observations is more straightforward to discuss in the context of an example. Imagine that one of the observations with 10 events discussed above had 100 chances to observe an event but the other had 1,000 chances. These two counts of 10 then imply starkly different probabilities of observing an event (i.e., $\frac{10}{100} = .1$ and $\frac{10}{1000} = .01$). In data, such equifinality in counts from different underlying probabilities can arise from observations that are units reflecting populations of different sizes or units reflecting different time periods in which events could occur (i.e., unequal employment tenures).

...

CDMs can control for such unequal *exposure* to the count generating conceptual phenomenon using what is known as an *offset* term. The offset term is usually assumed

to be natural log transformed for a CDM and enters into the CDM with a coefficient of 1. The coefficient of 1 results in the count DV being transformed into a rate out of the offset variable. How the count DV is transformed into a rate with a coefficient of 1 follows from the following algebraic manipulations. First, consider a simple case where an intercept only model is fit with an offset such as $y = e^{offset+\beta}$. Recall that the offset term is the natural log of a variable, say, o. Applying the natural log, 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 offset variable results in the count DV transforming into a rate. This offset transformation corrects for the issue discussed above as the 10 counts result in the correct underlying rate of .1 versus .01 for each observation.

In large part, an offset adjustment serves to re-scale specific observations' predictions and has the biggest effect on CDM's intercept value; adjusting it back to reflect the correct average rate across observations. The offset can, however, affect the magnitude of estimated coefficients if the offset is correlated with an IV. When the offset is correlated with an IV, then the IV may be both related to exposure to the count generating phenomenon as well as the rate of count generation broadly. Not including the offset can then bias the CDM and can have a notable impact on IV relative importance. How offsets affect CRM modeling and importance determination will also be explored further in the empirical examples below.

A second consideration relevant to CDV models is the use of model offsets. An offset is a variable that is intended to reflect *exposure* or differences between observations in the capability for that observation to produce a count. Offset variables are usually factors such as population sizes (cite) or exposure time windows (cite) that will affect the observations' counts and are known about different observations beforehand. Offsets are included into the model with a coefficient of 1 and serve to make the CDV a rate as they adjust the count such that $e^{y_i - offset_i} =$

In the sections below, each of the three complications regarding CDVs is discussed in greater detail with a focus on how each can affect the determination of relative importance.

Multiple Equations

... Poisson vs alternatives ...

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

Modeling with Exposure

...

Modeling with Zero-inflation

...

Poisson and linear regression share many similarities and, as a result of these similarities, often give similar answers when a count DV is applied to either model. Indeed the DA results for the Poisson model with the deviance R^2 are strikingly similar to those obtained from the Gaussian DV DA both in terms of the magnitudes of the deviance R2. This result is as expected in that the Poisson results are an extension of the Gaussian DV-based results and the deviance R^2 is a direct generalization of the explained variance R^2 to generalized linear models.

When applying

...

CDV models are complex, inherently multiplicative (note that it is possible to estimate them as non-multiplicative - but this is rarely done) models that require the use of estimation techniques such as maximum likelihood to obtain parameter estimates and sampling variances. It is helpful to discuss some of the nuances of how these models are estimated to understanding the implications of these complexities for DA. Hence, in the sections to come, I discuss aspects

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.

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.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth Ed.). New York: Springer. ISBN 0-387-95457-0.