Journal Title
XX(X):1–16

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Dominance analysis for count dependent variables

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

Organizational scientists conduct research on work-related problems across many topics including job performance, employee wellness, and effective work staffing. Quantitative research on these topics often requires that researchers use data that are in the form of discrete, sometimes infrequent, events such as number of contracts won in

[1]Fors Marsh Group LLC

a year, number of complaints received in a month, or number of days absent for illness in a business quarter. Discrete, infrequent events are modeled using *count regression models/CRMs* and are commonly applied in organizational science but can present additional complexities for the interpretation of analysis results. The most significant complication is that CRMs diverge in important ways from the statistical assumptions made of the Normal or Gaussian distributed linear regression model/LRM—by far the most common predictive model applied in organizational science ().

CRMs diverge from LRM as the discrete events used as dependent variables have truncated distributions that cannot range lower than 0 and tend to increase and decrease non-linearly. For example, *Poisson regression/PR* as well as *negative Binomial regression/NBR* are two commonly used CRMs in the organizational sciences (). Both PR and NBR are log-linear models (i.e., have the form e^{β}) to adapt their predictions to the non-negative range of real numbers and follow probability distributions that require non-negative integer values. PR and NBR are then both conceptually more applicable to count data but also are far more complex to interpret than LRM as they are log-linear. In addition, count data are often generated as aggregations of events over a specific time period such as a firm over a month's time or across a collective of conceptually related sub-units such as individual team members in a team. CRMs used to model these aggregated events then require the application of additional considerations to establish parity across observations when time windows or collective sizes differ—an issue that tends to be less applicable in LRM.

The complexities of CRM estimation and interpretation is further compounded when considering how to determine the relative importance of independent variables/IVs from CRMs. Relative importance analysis is a useful interpretation tool that adds detail to the estimation of parameter estimates (Tonidandel & LeBreton, 2011) and can provide practically useful information about IV predictive utility (). Among relative importance analysis methods, *dominance analysis/DA* () is most often recommended as a method that has the strongest conceptual foundation () and most flexible implementation in terms of extensibility (Luchman, 2021). Published methodological work on DA has discussed multiple intrinsically non-linear models including binary (Azen & Traxel, 2009), ordered, and multinomial logit (Luchman, 2014) models but has not provided an extensive discussion of how to implement and interpret DA with CRMs. As a consequence, despite the conceptual applicability of DA to CRMs, there have been no published research that applies DA to CRMs to the knowledge of the author.

This manuscript extends the literature on CRMs and DA in three ways. First, this work provides a discussion of and recommendations for DA as applied to CRMs commonly used in organizational science. In discussing how to extend DA do CRMs, a focus will be on recommending the use of a model fit metric that is comparable to those available for LRM. Additionally, I will focus on highlighting the similarities between DA results with the well-understood LRM and those obtained with CRMs given an identical structure of the underlying data. The similarities between these results are intended to empirically demonstrate the idea that DA can be extended to CRMs and reinforce that the DA methodology is a model agnostic-model free decomposition and importance approach. Second, this work focuses on the use of DA for the interpretation of CRMs. In this way, this paper can also be seen as an extension of Blevins, Tsang,

and Spain's (Blevins, Tsang & Spain, 2015) article focused on a conceptual discussion of CRMs and which to choose, to this paper's goal of recommending post-modeling methods to assist in adding context and detail to the model results. Finally, this work discusses several issues that are unique to CRMs and that could influence the results of DA: the concept of observation exposure and the concept of zero-inflation.

The paper to come is organized into ... sections. The first section will review DA... The second section will segue into recommendations for fit metrics as well as an The third section The fourth ...

Dominance Analysis

DA is an extension of Shapley value decomposition from Cooperative Game Theory () which seeks to find a solution to the problem of how to subdivide payoffs to players in a cooperative game based on their relative contributions.

The Shapley value decomposition method views the predictive model as a cooperative game where the different independent variables work together to predict the dependent variable. The payoff from the predictive model is the value of the model fit statistic; usually this payoff is an R^2.

This methodology can be applied to predictive modeling in a conceptually straightforward way. Predictive models are, in a sense, a game in which independent variables cooperate to produce a payoff in the form of predicting the dependent variable. The component of the decomposition/the proportion of the payoff ascribed to each independent variables can then be interpreted as the IVs importance in the context of the model as that is the contribution it makes to predicting the dependent variable.

In application, DA determines the relative importance of IVs in a predictive model based on each IV's contribution to an overall model fit statistic—a value that describes the entire model's predictions on a dataset at once. DA's goal extends beyond just the decomposition of the focal model fit statistic. 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:

$$X_v D X_z \; if \; 2^{p-2} = \Sigma 2 \tag{1}$$

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.

Conditional dominance statistics are computed as:

$$C^i_{X_v} = \tag{2}$$

Where S_i is a subset of IVs not including X_v and [p-1i-1] is the number of distinct combinations produced choosing the number of elements in the bottom value (i-1) given the number of elements in the top value (p-1; i.e., the value produced by choose(p-1, i-1)).

In effect, the formula above amounts to an average of the differences between each model containing X_v from the comparable model not containing it by the number of IVs in the model total.

General dominance is computed as:

$$C_{X_v} = \frac{\sum_p^i C_{X_v}^i}{p} \tag{3}$$

Where, $C_{X_v}^i$ are the conditional dominance statistics for X_v with i IVs. Hence, the general dominance statistics are the arithmetic average of all the conditional dominance statistics for an IV.

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

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

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 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.

The next section discusses the more stringent conditional 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 in terms of determining conditional dominance easier to see.

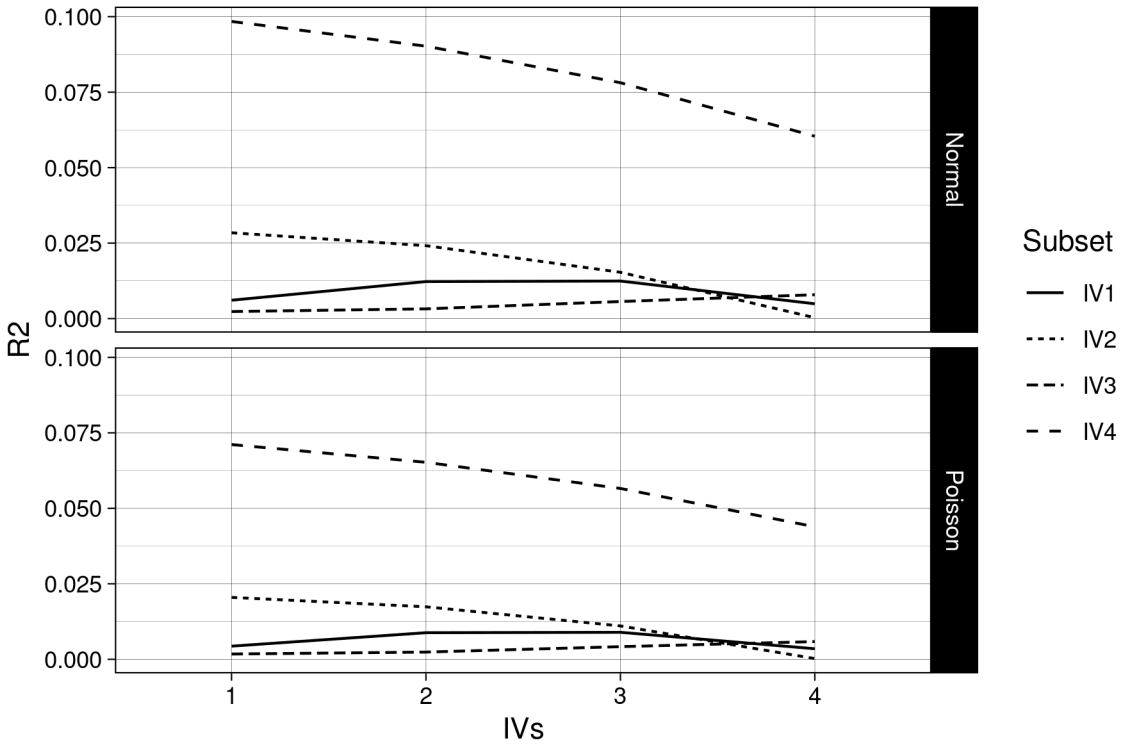

Figure X1 shows that IV_4 conditionally dominates all three of the other IVs. By contrast, the three other IVs have no conditional dominance relationships. This is because the three lines describing IV_1, IV_2, and IV_4's conditional dominance statistics cross over one another between the three IV and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

... note the similarity between the lines across models again...

... only iv4 can be completely dominant - note that and that it is. ...

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

CDMs, as models of counts of discrete events, implicitly assume that the events being counted are comparable. An observation with a count of 10 is assumed to have had the same chance to get that count of 10 events as any other observation with 10 events. That is, that the probability of observing events for both observations is equal. By contrast, it is possible two observations with an observed count of 10 arrived at their count with different underlying probabilities of observing events and number of times the events could be realized into a count. For example, two 10 counts could have resulted from an observation with an event probability of .1 over 100 'trials'. Another could

have arisen from an observation with an event probability of .01 over 1000 'trials'. 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. & Traxel, N. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics*, *34(3)*, 319–347.

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.

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.

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

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.