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.

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:

$$C_{X_v}^i = \tag{2}$$

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 coefficient 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 ≈ 7.69 for observation one and for observation two is $e^{.2+.5*8+-.8*-.2}$ or ≈ 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 ≈ 12.68 and observation two's predicted value is $e^{.2+.5*9+-.8*-.2}$ or ≈ 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 ≈ 5 whereas observation two's change was ≈ 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 CDMs 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 the PR and NBR models.

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 and NBR models.

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, a Poisson, and a negative Binomial DV in the section below. These three 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 and NBR.

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

Data Generation The 100,000 simulated observations in this manuscript was generated from a data structure with four IVs that have differing relationships with

the DV, differing relationships with one another, and differing variances. The four IVs created were based on the correlation matrix in Table X.

	IV1	IV2	IV3	IV4
Parameter				
IV1	1.00000000	0.67082039	0.41833001	0.25000000
IV2	0.67082039	1.80000000	0.28062430	0.67082039
IV3	0.41833001	0.28062430	2.80000000	1.67332005
IV4	0.25000000	0.67082039	1.67332005	4.00000000
Sampled				
IV1	0.13580061	0.08195510	-0.16607185	0.09640778
IV2	0.08195510	0.64261656	-0.04569956	-1.11854957
IV3	-0.16607185	-0.04569956	2.05240296	-1.27216522
IV4	0.09640778	-1.11854957	-1.27216522	3.88941540

The correlation matrix in Table X was designed to produce a realistic set of IV intercorrelations that could plausibly be uncovered in real data. Table X's intercorrelations was also designed balance extent of redundancy between the four IVs. That is, the correlations in Table X do not result in one or two IVs being more redundant than other IVs overall; all IVs have equivalent total redundancies as other IVs. Stated differently, all the IVs overlap with other IVs to the same extent and thus will not have differential extents of multicollinearity.

The means for all four IVs were set to 0. The variances for each IV differed. V_1 was set to 1, V_2 was set to 1.8, V_3 was set to 2.8, and V_4 was set to 4. The differences between the variances were in equal intervals on the square root scale between 1 and 2 (i.e., $1.8 \approx 1.33^2$ and $2.8 \approx 1.67^2$). The differences between IV variances were designed to show the effect of IV variance on dominance statistic computation.

The simulated IV values were used to generate the three DVs. The Gaussian/Normally distributed DV was generated first. The equation to generate the continuous, Normally distributed DV, Y_{Normal}, was: $Y_{Normal} = V_1 * .2 + V_2 * .17 + V_3 * .14 + V_4 * .1$. The true coefficients ascribed to each IV were in equal intervals between .04 and .01 on the exponential scale with a power of two (i.e., $.14 \approx \sqrt{.02}$ and $.17 \approx \sqrt{.03}$). A residual variance for Y_{Normal} was generated for each dataset as $1 - \sigma_{Y_{Normal}}$; hence, Y_{Normal} had a variance of 1 and a mean of 0.

Y_{Normal} was used as a basis for computing the Poisson distributed DV, $Y_{Poisson}$ to ensure that they were as similar and comparable as possible.

... ended here - note the update to the corr matrix and coefficient generation - ...

A Poisson distributed version of Y_{cont}, Y_{pois}, was also generated by obtaining the cumulative probability of each observation's Y_{cont} 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.

A negative Binomial version, Y_{nb}, was created similarly by translating Y_{cont}'s cumulative probability for each observation into a count variable using the negative Binomial quantile function assuming a Gamma shape parameter of .5 and scale parameter of .33; these factors result in a mean of 1 and variance of 3 for the negative Binomial distribution. The negative Binomial parameters obtaining a mean of 1 and variance of 3 were chosen to maintain similarities to the Poisson version of the variable (i.e., their means are identical) but to emphasize the unique feature of the negative Binomial distribution (i.e., the capability to model over-dispersion).

... offets ...

... ZIP ...

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.

Base Relative Importance Results

The data generated above by data structure and DV distribution were submitted to separate DAs to produce a total of twelve different DAs. The four DAs with each Normally distributed DV was based on linear regression. The four DAs with each Poisson distributed DV was based on Poisson regression/generalized linear model with a log link and Poisson family. The four DAs with each negative Binomial distributed DV was based on negative Binomial regression/generalized linear model with a log link and negative Binomial family. The model parameter estimates for each of the twelve models DAs were also estimated but are not reported in this section as they are not a focus; these results are reported in Appendix X. The general dominance statistic results from each DA are presented in table X.

Linear Model The results from the linear model-based DA is, as noted above, most useful in this table as a comparison point for the other results computed from the CRMs. The linear model is used as a comparison point as it is a better known model than most CRMs and is easier to work with mathematically. The EU data structure is also intended to be a comparison point data structure representing idealized, if unrealistic, data conditions for estimation. As such, I begin the interpretation of the results with the EU data. The idealized data structure represented by the EU condition produced results that were consistent with those which would be theoretically expected with identically distributed, uncorrelated IVs. Specifically, when there are no intercorrelations and the variance of the variable is 1, the expected R^2 value is the square of the coefficient (cite?). The linear model results in Table X for the EU data structure follow this pattern well and show that general dominance statistic results, when there are no intercorrelations and equal variances, recover the theoretically expected R^2 values. Indeed, as is well known (cite?), DA is not a necessary procedure when IVs are not intercorrelated as the incremental R^2 values for each IV are tantamount to the general dominance statistics.

The UU results vary from those obtained by the EU in that V_4, despite having a known coefficient value that is half that of V_1 obtained nearly the same general

dominance statistic value. As with the EU structure, UU's results are theoretically known. For variables with unequal variances with no intercorrelations, the expected R^2 value is the square of the coefficient times the IV variance (cite?). Hence, the expected general dominance statistics for V_1 and V_4 in the UU data structure were theoretically the same (i.e., $C_{V_1} = .2^2 * 1 = .1^2 * 4 = C_{V_4}$) however as realized in data as simulated in this manuscript, V_1 gained a slight edge and generally dominates V_4. The results from the UU condition are intended to reinforce the idea that there are two 'routes' to importance with DA, the magnitude of the relationship between the IV and the DV as well as the variance of the IV. Both relationship and variance contribute to the magnitude of the fit metric.

The EC results for the linear model also vary from those obtained from the EU model by both increasing the total R^2 values obtained for each IV, in particular V_3 and V_4, but also by decreasing the relative differences between the IVs. The increases in the values obtained by V_3 and V_4, despite their variance and known effect on Y_{cont} being identical to EU, occur as a result of the intercorrelations between each IV. The introduction of the non-zero intercorrelations makes between IVs results in ambiguity about how the R^2 should be decomposed as the large effect V_1 is known to have on Y_{cont} is now associated with the effects of the other three IVs when considering predicted values. That is, a change in V_1 is usually also associated with a change in V_2 (as they are correlated at .5 or share 25% of their variance) and, to a lesser extent, also V_3 and V_4. As a result of this correlation, predicted values on Y_{cont} tend to move up or down together. This shared change in predicted values makes it more difficult to tease apart to which variable should the change in predicted values, and thus improvement to explained variability when prediction improves, be attributed. Previously less important IVs are then ascribed higher shares of the R^2 as the DA algorithm tends to resolve the ascription ambiguity by ascribing R^2 value to IVs based on weighted a weighted average of their ΔR^2 across numbers of IVs in a sub-model (see equation ...). Ultimately, *some* of the effect of an IV with a stronger effect on Y_{cont} will be ascribed to those with less strong effects as is observed in Table X's EC results by contrast with EU given the ambiguity in ascription. In addition, the total/model R^2 has increased given the intercorrelations ... reasons - artifact of way I made these? ... Although the intercorrelations have increased the share of the R^2 ascribed to V_3 and V_4, it is important to note that, in this case, the rank order of the IVs does not change relative to the rank order in EU. The results from the EC condition then emphasize that the intercorrelations between IVs increase the difficulty in ascribing R^2 to IVs which tends to benefit IVs with weaker relationships with the DV and penalize IVs with stronger relationships with the DV.

Finally, the UC results combine both the difficulty in ascribing R^2 to IVs of the EC condition with the unequal variances of the UU condition to produce a result that both substantially increased the total R^2 values attributed to each IV but also changed the rank order of the IVs such that V_4 generally dominates each other IV. Overall, the results from the UC condition are similar to those obtained from the UU condition in that V_1 and V_4 are most important overall but, like the results from EC, the relative differences between the IVs' general dominance statistics have shrunk. The UC condition is arguably the most realistic condition of the four data structures as variance differences and non-zero IV intercorrelations are common in collected data.

The UC condition is thus intended as a condition to show reflect the effect of realistic, but non-ideal, data conditions on DA results.

The sections below transition to evaluating these same four data structures among DA results for the Poisson CRM with the recommended deviance R^2. A focus of the discussion will be on similarities and differences across models that are based on the same underlying data structures and DV predictive model.

Poisson Model Before discussing the results from the Poisson model/Y_{pois} DV, it is important to recall that Y_{pois} is based on the save data structure as all the results based on Y_{cont} and, in addition, is a direct translation of Y_{cont} into a Poisson distributed variable. Thus, the results are intended to be as identical as possible across models to assist in comparability. It is then not surprising to note that the Poisson model DA results show a similar pattern to those of the linear model of which they are an extension.

...

can be compared to the linear model results and reveal the same pattern

Table X shows the effect of the intended differences across data structures as realized in the DA results.

Table X shows stark differences between the data structures in terms of the dominance statistics obtained by each of the IVs. The EU data structure

The effect of the data structure is apparent in these results. The uncorrelated, equal variance results generally recovered the correct amount of explained variance for each IV. When introducing correlations to the model, ...

Whereas there were clear differences between the data structures in terms of importance determination, the results across fit metrics within a model were quite similar. Table X shows that the proportional contribution to the fit metric each IV was ascribed remained stable across fit metrics within a model. For instance, IV_1s proportional contribution for the Poisson model... Moreover, the dominance determinations did not differ across metrics within a model. IV_1 completely dominated all other...

Negative Binomial Model

Section Summary The base relative importance section reported on the results of the linear, Poisson, and negative Binomial

In the section below the focus moves to determining relative importance among CRMs with specific attention to cases where the DVs have unequal exposures to the count generating process.

Exposure and Offset Terms

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

Poisson Regression: The Most Basic Case

The most conceptually simple CDV model is the Poisson regression. The Poisson regression's log likelihood is a sum of three components as is shown below in (can I reference this equation?).

$$\ln L = \sum_{i=1}^{N} xb_i y_j - (\ln(y_i!) + e^{xb_i}) \tag{4}$$

Where xb_i refers to a respondent's untransformed/linear predicted value from the Poisson regression. Of note with this log-likelihood is the division into two separate components. On the one hand, the log-likelihood increases, with the $xb_i y_j$ term. Hence, the product of the observed CDV value and the predicted value for a respondent contributes directly to the log-likelihood and indicates better fit to the data. On the other hand, the second set of terms, $\ln(y_i!) + e^{xb_i}$ decrease the log-likelihood. Thus, has the

log factorial of the CDV increases and as the exponential function of the predicted values increase without a concomitant increase in the $xb_i y_j$.

For example consider a respondent with a CDV value of 5. The best fit to this value would be a transformed predicted value of 5 or an untransformed/linear predicted value of $e^5 = -1.740$. When applied to the log-likelihood function, the value obtained is $5 * -1.740 - \ln(5) + e^{-1.740} = -1.740$ or untransformed/linear predicted value. By contrast, two hypothetical less ideal predicted values might be 4 and 6. In both cases, the log-likelihood values diverge from the best fitting one and increase, indicating a lower likelihood. Specifically, 4 gives $5 * \ln(4) - \ln(5) + 4 = -1.856$ and 6 gives $5 * \ln(6) - \ln(4) + 6 = 1.829...$ This also shows an important point about the Poisson log likelihood–that it is not symmetric around the best fitting value. By contrast, ordinary least squares regression would penalize both 4 and 6 as squared deviations of 1 from the line of best fit (footnote that logged DV can do this too - but the loglik is still symmetrical and the model is now fitting the mean of a transformed DV as opposed to the original DV). That the model fit to the data is asymmetrical has implications for assessing fit to the model that will be discussed in greater detail later.

Negative Binomial: A More Complex and Flexible Case

A constraint with the Poisson model is that it has only a single parameter estimated from the data, it's mean. The variance of the Poisson distribution is, by assumption, identical to the mean.

One way to extend on the Poisson to make it more flexible is to conceptualize the mean of the Poisson distribution as a random variable. If the mean of the Poisson is conceptualized as Gamma distributed, the hybrid Poisson-Gamma distribution can be reduced to a unique distribution called the negative Binomial.

The negative Binomial model is a more complex generalization of the Poisson that relaxes the assumption that the variance and mean are identical

...negative binomial...

$$\ln L = \sum_{i=1}^{N} \ln(\Gamma(\frac{1}{\alpha} + y_i)) + m - (\ln(y_i!) + e^{xb_i}) \tag{5}$$

In applying DA to CDVs, many readers might question whether there is a need to use methods other than the standard linear regression with the explained variance R^2 metric.

Although CDV models share a number of similarities with continuous DVs, CDVs and the models designed to work with them are a distinct subset of generalized linear models and behave differently than linear regression in ways that *(cite Blevins and summarize somewhere around here)*.

One similarity across count and continuous DVs is that, as the mean of a CDV variable increases, it grows increasingly good at being approximated with a Normal distribution *(cite!)*. This points to an important difference when applied to CDVs with rare events in that the distributions for count and continuous DVs diverge notably when the mean is nearer 0 (example here?). Such divergence results in important differences to model-to-data fit that not only affects the model fit metric's value overall but also

how specific IVs explain variation in the CDV. In particular, CDVs are discrete, non-negative integers and, accordingly, CDV models are discrete probability distributions that accommodate non-negative integers. Normal distributions are continuous

Applying DA to CDVs

A useful first step toward defining recommended practice for applying DA to CDVs is to understand how CDV models differ from linear models. Differences between CDV and linear models affect model estimation and how relative importance among IVs should be determined. One substantial difference between CDV models such as the Poisson from linear models is in the functional form of the model. In a linear model, the magnitude of the regression coefficient for an IV reflects the expected change in the DV given one unit of change in the IV. By contrast, CDV models most often use a log-linear linking function. In a log-linear link model like the Poisson, the coefficients are estimated using from the data as though they were transformed using a natural logarithm. This implied transformation, or linking function, results in the CDV effectively ranging over all real numbers just like a continuous DV is expected to in linear regression.

$$ll = \frac{1}{1} \tag{6}$$

Although the CDV is implied to range over all real numbers in the estimation algorithm, the observed CDV is not changed and the predicted values from the CDV model are in log-linear units as opposed to those of the CDV (i.e., counts of the event). In order to produce meaningful predicted values, CDV models need to back-translate their predicted values to the metric of the CDV. The back-translation applies an anti-log or exponential function to the predicted values.

the natural logarithm linking function used to translate the predicted values model from a linear model back to the original count metric results in a one unit change in the IV producing a different magnitude of change to the dependent variable depending on where on the continuum of the dependent variable the change is located.

The log-linear nature of the coefficients produced by CDV models make the difficult to interpret directly. Typically, CDV coefficients are translated using an exponential function to produce *Incidence Rate Ratios* or *IRRs*. that naturally produce multiplicative effects across the range of each IV.

The naturally multiplicative functional form of CDV models makes the explained-variance R^2 metric less useful for DA. This is because CDVs are not guaranteed to produce an increase to the explained-variance R^2 as more IVs are added to the model ().

There are pseudo-R^2s that are better able to characterize model fit for CDVs.

$$\ln y = \sum \beta_{x_i} \tag{7}$$

Parameter Estimation Results

The data analysis begins where all relative importance analysis should, with parameter estimation using the statistical model appropriate for each DV. Relative importance analysis is an extension of parameter estimation (cite!) that provides additional, useful

information about the model and data. As such, relative importance analysis results should be interpreted in light of the parameter estimates produced by the model that is being dominance analyzed. The goal of this section is to highlight similarities and differences between the models' results across the four data structures with attention to each data structure's effect parameter estimate magnitudes.

Base Model Results This subsection discusses the parameter estimates obtained from the Normal, Poisson, and negative Binomial distributed DVs across all four data structures. Note that descriptive statistics for the IVs and DVs are reported in Appendix X. The results from the four models estimated on the Normal DVs are listed below in Table X.

The effects of data structure on estimation for Normally distributed variables with linear regression are well-known and the results in Table X are, as a result, consistent with extant research in showing that these data tended to recover the correct parameter estimates.

...

data generated were a rather large population size (100,000) and our model was correctly specified.

Hence, it is not a surprise that the known model parameters were recovered very well from the data across all four data structures and models.

Recall that the true parameters were .2 for X_1, .15 for X_2, .125 for X_3, and .1 for X_4. Also recall that these linear regression equations underlie the generation of both the Poisson and negative Binomial DVs. I expect then that the linear regression results will serve as a useful comparison point for the CRM parameter estimation results.

Numerically, the Poisson results look rather similar to the linear regression results as was intended. Indeed, the linear regression prediction equation (with error) was used to generate the underlying Poisson distributed variable and both had the same underlying variance of 1. A further similarity between the linear and Poisson results is that the coefficients obtained across data structures were nearly identical with the exception of the intercept.

The negative Binomial regression results, by contrast to the Poisson, were numerically less similar to the linear regression and more noticeably sensitive to data conditions during model estimation. For instance, the negative Binomial results tended to produce larger coefficients when the variables were both correlated and had unequal variances. The next section shifts to focus on the effect of the model, data structure, and fit metric chosen have on the determination of relative importance.

Offset Models

Zero-inflated Models

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.