Sawtooth Software

RESEARCH PAPER SERIES

Driver Analysis: How to Do It Badly and Well

Keith Chrzan Sawtooth Software, Inc.

© Copyright 2022, Sawtooth Software, Inc. 3210 N. Canyon Rd., Provo, Utah +1 801 477 4700 www.sawtoothsoftware.com

Driver Analysis: How to Do It Badly and Well

Keith Chrzan, Sawtooth Software

1.0 Introduction

"Driver analysis" (AKA derived importance analysis) is a staple deliverable for customer satisfaction, brand image, concept testing and loyalty research. With driver analysis, a marketer seeks to quantify how much each of several predictor or "independent" variables influences or contributes to some outcome or "dependent" variable. In applied marketing research studies, the dependent variable is usually some overall rating (satisfaction with the brand, liking or perceived excellence of the brand, intent to purchase the brand, etc.) and predictors are usually rating scale measures of the performance of the brand on each of several attributes (here "brand" means any entity one might want to study, e.g. a bank branch, an individual provider of home building services, a non-profit charity, a city park, a therapy for sarcoidosis, a brand name product or service). Analysis produces a set of scores that quantifies the importance of each of the predictors.

The dependent variable is usually some kind of rating scale, so for the predictive model we typically use a member of the regression analysis family (some sticklers in academia note that rating scales aren't continuous measures, but this is how it's done in practice – and by academics in the social sciences, too). Years ago, we used correlation or multiple regression analysis. As shown below, however, a pervasive data problem called collinearity afflicts and ruins these methods as vehicles for importance measurement. Fortunately, some newer methods (a family of relative importance measures and random forests) address the data problem and allow us to do driver analysis well.

Section 2 below describes traditional derived importance methods, using a case study and some visualizations of how the methods measure importance. Section 3 defines, illustrates, and quantifies collinearity. Section 4 introduces the newer methods for driver analysis and illustrates their success in the face of collinearity. Finally, section 5 extends the analysis to three new case studies to show that the results of the first case study generalize. Three appendices expand on some tangential topics related to driver analysis.

2.0 Doing Driver Analysis Badly: Traditional Methods

When I first started running derived importance models back in the 80s, the tools we had were regression analysis and correlation analysis. To illustrate these, let's use a casual dining customer satisfaction study: 1,284 respondents rated their overall satisfaction with their most recent visit to a casual dining restaurant and they rated the restaurant's performance on 10 attributes.

Correlation looks at how each attribute relates in turn to the overall satisfaction measure, independently of each other attribute. The correlations, ranging from 0.387 to 0.543, do not vary much:

Attribute	Correlation
Prompt Greeting	0.391
Cleanliness	0.390
Comfortable Environment	0.394
Attentive Server	0.486
Friendly Server	0.435
Appropriate Pace	0.436
Food Taste	0.543
Food Temperature	0.444
Timely Check	0.387
Reasonably Priced	0.400

Correlation over-counts the variation that the predictors share with the dependent variable, to the extent the predictors themselves overlap. This becomes clear when we square the correlations, so that each one represents the percent of the variance in the dependent variable it explains:

Attribute	Correlation Squared
Prompt Greeting	0.153
Cleanliness	0.152
Comfortable Environment	0.155
Attentive Server	0.236
Friendly Server	0.189
Appropriate Pace	0.190
Food Taste	0.295
Food Temperature	0.197
Timely Check	0.150
Reasonably Priced	0.160

The squared correlations sum to 188%, meaning they explain more than 100% of the variance in the dependent variable.

Regression analysis looks simultaneously at the relation of all 10 attributes to overall satisfaction. These regression coefficients result:

Attribute	Correlation Squared	Regression
Prompt Greeting	0.153	0.051
Cleanliness	0.152	-0.051
Comfortable Environment	0.155	0.004
Attentive Server	0.236	0.222
Friendly Server	0.189	0.006
Appropriate Pace	0.190	0.086
Food Taste	0.295	0.389
Food Temperature	0.197	0.031
Timely Check	0.150	0.010
Reasonably Priced	0.160	0.119

The regression importances seem to be more discriminating, but they include a negative importance weight for cleanliness, as if respondents prefer a dirty restaurant over a clean one. Moreover, only 3 of the 10 predictors are significant in the regression model, versus all 10 for the correlations.

3.0 Collinearity

The underlying data problem, collinearity (or sometimes "multicollinearity") happens in a multivariate model when the predictor variables (the 10 restaurant attributes in this case) are correlated with each other. Both correlation and regression would produce equivalent (and accurate) importances if the 10 predictors were not correlated among themselves. Unfortunately, in survey research, predictor variables measured in attitudinal ratings are almost always correlated with one another. Thorndyke (1920) discovered an explanation for this and named it the "Halo Effect." In a nutshell, respondents liking a brand tend to give it high ratings on all attributes, while respondents disliking a brand tend to give it low marks across the board. In Thorndyke's memorable phrase, the halo effect causes all the correlations to be "too high and too even."

The halo effect clearly affects the casual dining study, as all 10 predictors are highly correlated with each other:

	Prompt Greeting	Cleanliness	Comfortable Environment	Attentive Server	Friendly Server	Appropriate Pace	Food Taste	Food Temperature	Timely Check	Reasonably Priced
Prompt Greeting	1.00									
Cleanliness	0.58	1.00								
Comfortable Environment	0.52	0.63	1.00							
Attentive Server	0.61	0.55	0.56	1.00						
Friendly Server	0.65	0.61	0.60	0.71	1.00					
Appropriate Pace	0.54	0.58	0.54	0.59	0.57	1.00				
Food Taste	0.48	0.59	0.55	0.55	0.56	0.55	1.00			
Food Temperature	0.54	0.66	0.57	0.61	0.62	0.62	0.63	1.00		
Timely Check	0.52	0.53	0.52	0.64	0.58	0.62	0.46	0.54	1.00	
Reasonably Priced	0.40	0.40	0.48	0.48	0.45	0.49	0.47	0.46	0.45	1.00

It turns out we can quantify the amount of collinearity present in a data set, using what's called a condition index. Condition indices over 10 suggest a likely problem with collinearity and condition indices over 30 suggest a severe collinearity problem (Belsley, Kuh and Welsch 1980). The condition index for the casual dining restaurant data is 17.6; even though it's not above 30, we still have collinearity severe enough to keep us from interpreting the regression importances.

Collinearity means that the importances from correlation may only be interpreted ordinally (higher means more important) and the regression results cannot be interpreted as importances at all. In the presence of collinearity, neither correlation nor regression analysis works the way we'd like for measuring attribute importance.

4.0 Doing Driver Analysis Well: Some Newer Methods

Three newer methods, developed with collinearity in mind, handle driver analysis well.

4.1 Averaging over orderings (AOO)

Think of running a regression analysis where we enter the variables in order. We attribute to the first variable to enter the model all the variance the predictor shares with the dependent variable. To the second variable we attribute only the variance in the dependent variable the unique variance NOT shared by the first variable. This goes on until we get to the last variable, and we attribute to it only the unique variance it shares with the dependent variable that's not shared by ANY of the previously entered predictors. In effect, correlation treats each predictor as if it enters model first, while regression treats each variable as if it enters last (see Appendix 2 for more details).

Not knowing which ordering of variables to choose, with AOO we run the regression analysis in every possible order (e.g. for 10 predictors we run 10! = 3.6 million). In each of those many orderings, we keep track of how much variance a given predictor explains, given its place in that ordering and then we average that contribution over all the orderings. Normalizing those averaged contributions (dividing them by the 36% total explained variance) they sum to 100% and give us a set of importances:

Attribute	Correlation	Regression	<i>AOO</i>
	Squared		
Prompt Greeting	0.153	0.051	0.061
Cleanliness	0.152	-0.051	0.056
Comfortable Environment	0.155	0.004	0.060
Attentive Server	0.236	0.222	0.142
Friendly Server	0.189	0.006	0.078
Appropriate Pace	0.190	0.086	0.087
Food Taste	0.295	0.389	0.279
Food Temperature	0.197	0.031	0.087
Timely Check	0.150	0.010	0.058
Reasonably Priced	0.160	0.119	0.091

In addition to the theoretical advantage of slicing up and apportioning the shared variance, the AOO importances have a couple of other desirable properties for practitioners: they sum to 100% and they have a ratio-level interpretation (e.g. we can say that Food Taste, with an importance of 27.9%, has about three times the importance of being Reasonably Priced, with its importance of 9.1%). Neither correlation nor regression has this ratio level interpretation when collinearity is present.

In 1980 Lindeman, Merenda and Gold first suggested averaging the incremental contribution to R^2 over orderings in their general textbook about multivariate methods. Cox (1985) noted that AOO was equivalent to the Shapley value. Unaware of the earlier work that looked at the incremental R^2 each variable adds, Kruskal (1987) suggested averaging a different entity, squared partial correlations and Theil and Chung (1988) suggested averaging an information-theoretic entity called entropy over orderings. Budescu (1993) invented something he called dominance analysis, which also turned out to be equivalent to Lindeman, Merenda and Gold's averaging of R^2 measure. While the analyst has several variants of this "relative importance analysis" to choose from, and while each variant has its proponents, in practice all of these AOO methods seem to return very nearly the same attribute importances.

The huge number of regressions run as part of an AOO analysis means that some problems can take a very long time to run, time that increases geometrically with the number of predictors (with 21 predictors, AOO involves more regression models than there are grains of sand on the earth). In a recent study with 23 predictors, AOO took about 20 minutes to run on my computer.

4.2 Johnson's relative importance weights (e)

Johnson (2000) came at the idea of identifying the importance of predictors using some straightforward maxrix algebra (interesting trivia here: Johnson drew on earlier work that Rich Johnson, founder of Sawtooth Software, had done translating correlated attributes into uncorrelated variables). Johnson's relative importance algorithm differs from those used by the AOO methods, and it runs quickly for pretty much any number of predictors. Johnson's relative weight scores, which he calls e, turn out to be very similar to those that come from the AOO methods:

Attribute	Correlation Squared	Regression	A00	8
Prompt Greeting	0.153	0.051	0.061	0.066
Cleanliness	0.152	-0.051	0.056	0.054
Comfortable Environment	0.155	0.004	0.060	0.061
Attentive Server	0.236	0.222	0.142	0.137
Friendly Server	0.189	0.006	0.078	0.075
Appropriate Pace	0.190	0.086	0.087	0.089
Food Taste	0.295	0.389	0.279	0.269
Food Temperature	0.197	0.031	0.087	0.087
Timely Check	0.150	0.010	0.058	0.060
Reasonably Priced	0.160	0.119	0.091	0.103

Like the AOO methods, Johnson's e scores sum to 100% have they have a ratio-level interpretation.

4.3 Random Forests (RF)

Proposed by Brieman (2001), random forests extends the idea of running tree-based regression: Instead of building a single regression tree, however, RF builds a forest of trees, where each tree involves two randomization steps that serve to "decorrelate" the variables (and hence to combat collinearity). Prediction happens by way of a vote that each tree makes, but for our purposes, RF also produces a couple of importance measures. One of these, called "increase in node purity," works a little better in practice because its scores are always positive. Like AOO and e, the RF importances have a ratio level interpretation and they can be normalized to sum to 100%:

Attribute	Correlation Squared	Regression	<i>AOO</i>	8	RF
Prompt Greeting	0.153	0.051	0.061	0.066	0.061
Cleanliness	0.152	-0.051	0.056	0.054	0.067
Comfortable Environment	0.155	0.004	0.060	0.061	0.071
Attentive Server	0.236	0.222	0.142	0.137	0.143
Friendly Server	0.189	0.006	0.078	0.075	0.087
Appropriate Pace	0.190	0.086	0.087	0.089	0.085
Food Taste	0.295	0.389	0.279	0.269	0.255
Food Temperature	0.197	0.031	0.087	0.087	0.084
Timely Check	0.150	0.010	0.058	0.060	0.065
Reasonably Priced	0.160	0.119	0.091	0.103	0.082

Interestingly, despite the fact that the RF algorithm is different from both the AOO and from Johnson's e, it produces a very similar vector of importance scores. In fact, if we compare the results of these five methods, we can see that RF, AOO and e importances are very highly

correlated with one another, and less so with either correlations or regression coefficients (which also have lower correlation with one another):

	Correlation Squared	Regression	A00	ε	RF
Correlation Squared	1.000				
Regression	0.896	1.000			
AOO	0.949	0.952	1.000		
ε	0.933	0.963	0.997	1.000	
RF	0.956	0.936	0.995	0.987	1.000

So similar are the importances from AOO, e and RF that one could use any of them and have a high degree of confidence in the answer. In practice, I use AOO when I have fewer than about 20 rating scale variables, Johnson's e when I have more than 20 or so (or when I have missing data, because e handles missing data well), and RF when I have a mix of rating scale and categorical predictors.

5.0 It's Not a Fluke: Three Additional Case Studies

Of course, you should convince yourself of all of the above by running the analyses above on any of your driver analysis studies. Just for fun I ran them on three more data sets. The condition indices for two of these were similar to the casual dining study above, in the territory Belsley, Kun and Welsch would call likely problematic: an airline satisfaction study with condition index of 17.0 and an auto service satisfaction study with a condition index of 19.6. The third, from a burger joint brand image study, is well into the "serious problem" range with a condition index of 33.8. All studies had at least several hundred respondents and they had 9-13 predictor variables. Each time the correlations exhibited a great amount of overcounting of shared variance and each time most of the regression coefficients were nonsignificant, some with incorrect sign reversals. Again, the three new importance metrics produce very similar importance scores, less similar to those from regression and correlation analysis:

	Correlation Squared	Regression	A00	ε	RF
Correlation Squared	1.000				
Regression	0.763	1.000			
AOO	0.937	0.913	1.000		
ε	0.898	0.945	0.990	1	
RF	0.921	0.859	0.960	0.935	1

6.0 Where to Find Driver Analysis Programs

All of the methods discussed are available in R. The package relaimpo contains several AOO programs, while party includes RF. You can run Johnson's relative weights as part of the iopsych package or in the dedicated rwa package. An online calculator also runs Johnson's

relative weights: <u>https://relativeimportance.davidson.edu/multipleregression.html</u>. I'm not aware of any publicly available version of the Theil's information-theoretic AOO method, but Joe Retzer wrote a proprietary program for it, so if you ever need it run, he's your guy.

7.0 Conclusion

Traditional derived importance methods have serious shortcomings due to a serious data problem endemic to survey research: collinearity. Three new methods designed to avoid the ill effects of collinearity do so while producing usefully scaled outputs. Moreover, though the three methods come at the problem from very different directions, their results agree so much as to suggest their convergent validity.

References

Belsley, D.A., E. Kuh and R.E. Welsch (1980) *Regression diagnostics: Identifying influential data and sources of collinearity.* Hoboken, NJ: Wiley.

Breiman, L. (2001) "Random Forests," *Machine Learning*, **45**: 5–32.

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

Chrzan, K,., J. Retzer and J. Busbice (2003) "The Predictive Validity of Kruskal's Relative Importance Algorithm," *Sawtooth Software Conference Proceedings*, 77-85.

Cox, L.A. Jr., (1985) "A new measure of attributable risk for public health applications," *Management Science*, **31**:800-813.

Feldman, B. (2005) "Relative importance and value," Manuscript version 1.1, 2005-03-19.

Gibson, W.A. (1962) "Orthogonal predictors: a possible resolution of the Hoffman-Ward controversy," *Psychological Reports*, 11: 32-34.

Gromping, U. (2006) "Relative importance for linear regression in R: The package relaimpo," *Journal of Statistical Software*, **17**: 1-27.

Gromping, U. (2007) "Estimators of relative importance in linear regression based on variance decomposition," *The American Statistican*, **61**: 139-147.

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

Johnson, J.W. and J.M. LeBreton (2004) "History and use of relative importance indices in organizational research," *Organizational Research Methods*, **7**: 238-257.

Johnson, R.M. (1966) "The minimal transformation to orthonormality," *Psychometrica*, **31**: 61-66.

Kruskal, W. (1987) "Relative importance by averaging over orderings," *The American Statistician*, **41**, 6-10.

Lindeman, R.H., P.F. Merenda and R.Z. Gold (1980) *Introduction to Bivariate and Multivariate Analysis*. Glenview, IL: Scott, Foresman.

Lipovetsky, S. and M. Conklin (2001) "Analysis of regression in game theory approach," *Applied Stochastic Models in Business and Industry*, **17**: 319-330.

Soofi, E.S., J. Retzer and M. Yasai-Ardekani (2000) "A framework for measuring the importance of variables with applications to management research and decision models," *Decision Sciences*, **31**: 1-31.

Tabachnick, B. and L.S. Fidell (1983) Using Multivariate Statistics. Cambridge: Harper & Row.

Theil, H. and C-F.Chung (1988) "Information-theoretic measures of fit for univariate and multivariate linear regression," *The American Statistician*, **42**: 249-252.

Theil, H. (1987) "How many bits of information does an independent variable yield in a multiple regression?" *Statistics & Probability Letters*, **6**, 107-108.

Thorndyke, E.L. (1920) "A constant error in psychological ratings," *Journal of Applied Psychology*, **4**, 25-29.

Tonidandel, S. and J.M LeBreton (2011) "Relative importance analysis: a useful supplement to multiple regression analyses," *Journal of Business and Psychology*, **26**: 1-9.

Tonidandel, S., J.M. LeBreton & J.W. Johnson (2009) "Determining the statistical significance of relative weights," *Psychological Methods*, **14**: 387-399.

Appendix 1: What About Factor Analysis?

For some time I thought the best I could to do combat collinearity was to run factor analysis. Factor analysis groups a set of intercorrelated attributes into a smaller set of uncorrelated factors. Originally developed for testing psychometric theories, factor analysis can assess the validity of conceptual models (if a construct is real, the items purported to measure it should load together onto a factor) and the validity of measurement models (if an item is a good measure of a construct it should load highly on the one factor representing that construct and not on any other factor). One can imagine using factor analysis to bypass collinearity by using these uncorrelated factors as the inputs to driver analysis.

Most marketing research surveys, however, use attribute lists not constructed with any psychometric theory in mind. Researchers build haphazard assortments of items into their attribute lists, rarely with any thought about any theorized factor structure. Factor analysis of such attribute lists often produce cross-loaders (important items whose impact is diluted because they load on multiple factors) or unique items that don't load onto any factors because they have low correlations with the other items. In a nutshell, factor analyzing attribute rating scales prior to regression analysis can obscure important predictors.

Even when factor analysis works well, researchers' clients often struggle with their meaning: a marketing manager may care about attributes like "ease of use," "quality" and "timely delivery" but may become frustrated if they load on the same factor without separately quantified importances.

In short, I love factor analysis for the kind of psychometric testing for which it was developed, but not as a panacea that turns hodge-podge attribute lists into driver analysis gold.

Appendix 2 – Visualizing Correlation, Regression and Collinearity

Assume we have two predictors, x1 and x2, and that they each relate to some extent to a dependent variable dv. The circles below represent the variance in the measures of x1, x2 and dv. Correlation counts all the variance that a predictor shares with the dv as the importance of that predictor, i.e. the purple shaded area below measures the correlation of x1 with dv.



Figure 1: What Correlation Counts

Similarly, the correlation of x2 and dv would be the overlap, or intersection, of the x2 and dv circles (part of which is already shaded purple because to some extent x1, x2 and dv all overlap). So correlation over-counts the variation that the predictors share with the dependent variable, to the extent the predictors themselves overlap. With 10 predictors, correlations can end up explaining more than 100% of the variance in the dependent variable!

dv

Compare this to what regression analysis does. Regression credits only the unique variance that a predictor shares with the dependent variable (i.e. only the overlap that x1 has with dv that is not also shared with x2:

Figure 2: What Regression Counts



Just as correlation can double (or triple, etc.) count shared variance, regression analysis undercounts it, because regression ignores all the variance that multiple predictors share with the dependent variable (in the diagram above it would count the purple shaded area plus a similar area that x2 shares with dv but not with x2). The variance shared by all three of x1, x2 and dv gets ignored.

Collinearity happens when we have lots of predictors sharing lots of variance. For example, here x1 and x2 overlap almost completely, so the unique variance x1 shares with dv is just the shaded sliver:





Here the high correlation between x1 and x2 ignores most of the variance either of them shares with dv. Small changes to the position of the circles can cause large fluctuations in the size of the shaded sliver, meaning collinearity can cause instability in regression models (it inflates the variance in estimates of the regression coefficients, leading to another mathematical measure of collinearity called the variance inflation factor or VIF).

Appendix 3: Technical Bits Saved to the End

Making a slight improvement to regression analysis

Just as we were able to make correlation a little more useful by squaring the raw correlation coefficients, we can improve regression a little bit. We use regression coefficients for making predictions but for apportioning importance to the individual predictors a related entity called squared semipartial correlation (sr_i^2) may make more sense (Tabachnick and Fidell 1983). The squared semipartial correlations measure only the unique variance a predictor shares with the dependent variable; we can normalize them so they sum to 100% by dividing by their sum, which is 8% in the casual dining study. Normalizing makes them more comparable in scale to correlations:

	Correlation	Regression
Attribute	Squared	(sr _i ²)
Prompt Greeting	0.153	0.010
Cleanliness	0.152	0.007
Comfortable Environment	0.155	0.000
Attentive Server	0.236	0.157
Friendly Server	0.189	0.000
Appropriate Pace	0.190	0.028
Food Taste	0.295	0.708
Food Temperature	0.197	0.003
Timely Check	0.150	0.000
Reasonably Priced	0.160	0.087

It also turns out that the squared semipartial correlations are (slightly) more correlated with the importances from AOO, from Johnson's e and from RF.

Extending AOO methods

Soofi, Retzer and Yasai-Ardekani (2000) illustrate that the entropy measure that Theil used for his AOO method allows the AOO methodology to be extended to other statistical models, like ANOVA and logistic regression.

At least two ways of allowing AOO methods to operate with large numbers of predictors have been developed. To support the analyses done in Chrzan, Retzer and Busbice (2003), John Busbice developed a variation of AOO which, instead of running all possible orderings, starts with a large number of random orderings and then runs further orderings only until importances converge. Kevin Lattery developed a method that uses AOO among different random subsets of a given number of predictors (e.g. 5) to estimate importances. I don't think Kevin ever presented this method outside of Maritz Research, because at the time we considered it a proprietary product offered only to our customer satisfaction clients.

More detail on the history of AOO methods

Useful summaries of the origin and history of AOO methods appear in Gromping (2007), Johnson and LeBreton (2004), Tonidandel and LeBreton (2011), and Tonidandel, LeBreton and Johnson (2009).