


Visualization strategies for regression estimates with randomization inference

Marshall A. Taylor
Department of Sociology
New Mexico State University
Las Cruces, NM
mtaylor2@nmsu.edu

 <https://orcid.org/0000-0002-1825-0097>

Abstract. Coefficient plots are a popular tool for visualizing regression estimates. The appeal of these plots is that they visualize confidence intervals around the estimates and generally center the plot around zero, meaning that any estimate that crosses zero is statistically nonsignificant at least at the alpha level around which the confidence intervals are constructed. For models with statistical significance levels determined via randomization models of inference and for which there is no standard error or confidence intervals for the estimate itself, these plots appear less useful. In this article, I illustrate a variant of the coefficient plot for regression models with p -values constructed using permutation tests. These visualizations plot each estimate's p -value and its associated confidence interval in relation to a specified alpha level. These plots can help the analyst interpret and report the statistical and substantive significances of their models. I illustrate using a nonprobability sample of activists and participants at a 1962 anticommunism school.

Keywords: gr0083, permutation tests, visualization, coefficient plots, p -values

1 Introduction

Visual alternatives to traditional estimate tables have become increasingly popular in the social sciences, especially with the growing interest in data science and computational social science (Healy 2019; Wickham and Golemund 2017). One visualization that is supplanting more traditional regression tables is the coefficient plot (Gelman et al. 2018; Jann 2014; Lander 2018). Coefficient plots can provide a more parsimonious method for reporting regression coefficients, especially when many independent variables (IVs), models, or both make tables difficult to read or present. Thus, these plots have become quite popular in sociological research, ranging in application from cultural sociology (for example, Lizardo and Skiles [2016a,b]; Stewart, Edgell, and Delehanty [2018]) and social networks (Bähr and Abraham 2016) to the sociology of education (Schühler, Carbonaro, and Grodsky 2016) and the sociology of work and occupations (Friedman and Laurison 2017; Laurison and Friedman 2016).

Coefficient plots visualize the confidence intervals (CIs) and their corresponding regression estimates. The graphs are generally centered around zero—meaning that any estimate that crosses zero is statistically nonsignificant at least at the alpha level around

which the CIs are constructed. Consider, for example, the coefficient plot in figure 1 below, which was generated from a basic linear-additive multiple ordinary least-squares (OLS) model regressing systolic blood pressure readings on a vector of covariates using the National Health and Nutrition Examination Survey data available from StataCorp. The plot is interpreted like a regular regression table: black individuals, for example, have systolic blood pressure readings that, on average, are about 2.333 mm/Hg. higher than white individuals' readings, net of their weight, height, sex, age, and self-reported health. The plot also shows that this difference is statistically significant with at least an $\alpha = 0.001$ because the thinnest band does not cross zero (the vertical line).

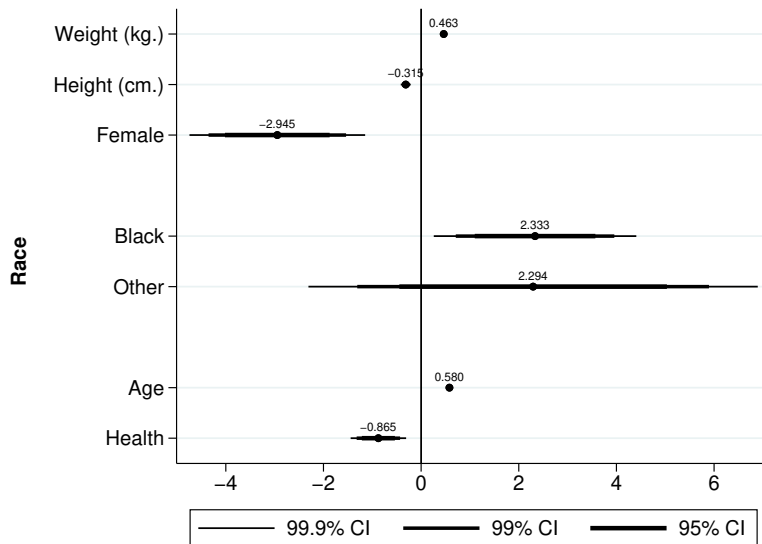


Figure 1. Coefficient plot of OLS estimated effects on systolic blood pressure. NOTE: Constant = 127.028 ($p < 0.001$). $n = 10335$. Race reference category = White. Unstandardized coefficients are reported. CIs are two sided.

For all their utility, coefficient plots still come with assumptions about the inference statistics accompanying the estimates—namely, coefficient standard errors (SEs) and their CIs. This is, of course, a valid assumption for the majority of analysts using probability samples and relying on the asymptotic properties of random sampling distributions. However, analysts using nonprobability samples must adopt a different inference model. Such an alternative—the “randomization model” (Ernst 2004; Ludbrook and Dudley 1998) or what might be termed the “process inference” model (Darlington and Hayes 2017, 513–514)—allows for sample-specific inferences by using Monte Carlo permutation tests to derive empirical p -values. These p -values indicate the proportion of test statistics—say, a specific coefficient—from the models computed from the randomly permuted samples that are greater than or equal to the observed test statistic (in the case of a right-tailed test).

The randomization model poses problems for the standard coefficient plot. Specifically, unlike the conceptually similar bootstrap technique, SEs derived from Monte Carlo permutation tests are with respect to the p -value—not the coefficient itself. Thus, there are no coefficient-specific CIs to report and no way to visualize the effect’s proximity to zero without explicit reference to the effect’s p -value.

In this article, I illustrate a variant of the coefficient plot for regression models with p -values constructed using permutation tests and discuss its utility in interpreting and presenting research. These visualizations, which I refer to as permutation p -value (PPV) plots, plot each estimate’s p -value and its associated CI in relation to an alpha level of the analyst’s choice. These plots can help the analyst interpret and report the statistical and substantive significances of their models with permutation-based inference without recourse to tables. This article unfolds as follows: First, I overview the logic behind Monte Carlo permutation tests and how they can be used to construct sample-specific inferences, noting especially how they introduce complications for how regression estimates and their associated predicted values are visualized. Second, I introduce PPV plots, illustrate how they can be generated, and demonstrate their utility for visualizing regression estimates and predicted values. PPV plot illustrations are made using a series of models predicting heightened perceptions of communist threat among a nonprobability convenience sample of people who participated in the CACC’s San Francisco Bay Region School of Anti-Communism in January and February of 1962 (Wolfinger 1992). I then conclude with an outline of how PPV plots can be further expanded for data visualization.

2 Randomization inference

2.1 Using Monte Carlo permutation tests to generate p -values

The traditional “population model” (Ernst 2004; Ludbrook and Dudley 1998) of statistical inference with which most social scientists are accustomed is not always appropriate when the sample is constructed using nonprobabilistic procedures or when the sample suffers from nonrandom systematic error that cannot be mitigated with bias-correcting statistical adjustments. When such “prestatistical” methodological constraints are present, a different model must be adopted if the analyst wishes to use the language of inferential statistics. One option is the “randomization” (Ernst 2004; Ludbrook and Dudley 1998; Manly 2007, 1–4) or “process inference” (Darlington and Hayes 2017, 513–514) model—a strategy that has, unfortunately, been slow to gain ground in sociological research using observational data,¹ save for work in social network analysis.²

1. The lack of a probability sample can often lead researchers to avoid the language of statistical inference altogether. In reviewing work using qualitative comparative analysis (QCA), for example, Thomann and Maggetti (Forthcoming, 12) note that “[f]ew QCA applications resort to statistical inference, partly because its use on full or purposive samples yields nonsensical inference to a nonexistent population.”

2. Permutation-derived p -values have been a common characteristic in statistical models for networks, because the interdependencies inherent in relational data necessarily violate the observational independence assumptions behind asymptotic test statistics (see, for example, Butts [2007]; Krackhardt [1988]).

The key distinction between the population model and the randomization model is the “thing” to which statistical inferences are being made. In both cases, the p -value, its associated CI (for the coefficient in the population model and the p -value itself in the randomization model), or both are used to assess the confidence in the inference.³

Consider the following simple OLS estimation equation:

$$\widehat{Y}_i = \widehat{a} + \widehat{\beta}X_i$$

On one hand, if p is the p -value attached to some estimated coefficient $\widehat{\beta}$, then under the population model and assuming a two-tailed test, p can be expressed as

$$p_{\widehat{\beta}} = p \left(|\beta| \geq |\widehat{\beta}| \mid H_0: \beta = 0 \right) \quad (1)$$

where β is the true but unobserved estimate of the effect of X on Y . This p -value expresses the probability that (if the coefficient was truly generated from a random sample from the target population and with no systematic error) an effect of at least the absolute size of $\widehat{\beta}$ would be observed from a sample of that population if the true effect $\beta = 0$.

On the other hand, under the randomization model with 1,000 Monte Carlo permutation tests and again assuming a two-tailed test, \widehat{p} indicates the proportion of $\widehat{\beta}^*$ estimates from the 1,000 permutations with an absolute value greater than or equal to the absolute value of $\widehat{\beta}$ observed prior to randomly permuting the data (see also Baker and Hubert [1981, 347]).⁴ (The hat over the p will be explained in the next section.) In other words, this p -value can be written as

$$\widehat{p}_{\widehat{\beta}} = \frac{\sum_{i=1}^k I \left(|\widehat{\beta}_i^*| \geq |\widehat{\beta}| \right)}{k}$$

3. Randomization inference is often (if not usually) discussed within the context of randomized experiments (for example, Aronow [2012]; Edgington and Onghena [2007]; Ludbrook and Dudley [1998]), where random assignment of participants into treatment and control groups is assumed but not necessarily random sampling. Nonetheless, as Manly (2007, 180) notes, justification for using permutation tests for establishing statistical inference in multivariate cases can come in three forms: 1) in the case of random sampling, when the H_0 is that $Y \perp \mathbf{X}$; 2) in the case of random assignment to treatment and control conditions (but not necessarily in conjunction with random sampling); and 3) when, with observational data that are neither from a random sample nor after an experimental design, “the mechanism generating the data makes any of the Y values equally likely to occur with any of the sets of X values.” This article adopts the “data-generating mechanism” justification for using permutation tests, given the prevalence of observational data relative to experimental data in sociology.

4. The variable that is permuted may vary. In experimental settings, the treatment variable, X , is often permuted while keeping the outcome vector, Y , the same. Others promote incorporating some form of the model residuals into the permutation procedure (for example, Freedman and Lane [1983]; O’gorman [2006]). Still others suggest simply permuting the Y vector (Manly 2007; Stoltz and Taylor 2017). For an overview of different permutation techniques, see Anderson and Robinson (2001).

where k is the number of permutations, $|\widehat{\beta}_i^*|$ denotes the i th estimate derived from one of the k permutations that is greater than or equal in absolute value to $|\widehat{\beta}|$, and $I(\cdot)$ is an indicator function that tallies the total number of times $|\widehat{\beta}_i^*| \geq |\widehat{\beta}|$ (Ernst 2004, 679; Stoltz and Taylor 2017, 33). Making generalizations to some population is, of course, not possible under the randomization model (at least on statistical grounds) because the asymptotic properties that come with assuming a probability sample free of nonrandom error is a luxury not enjoyed for most analysts who adopt the randomization model over the population model. Instead, working from a null hypothesis that $\widehat{\beta}$ can be accounted for by randomness, a \widehat{p} under some alpha level would indicate that, whatever the data-generating process that created $\widehat{\beta}$ might be, it is likely not a purely random process. In other words, we reject the null hypothesis when randomness as a mechanism for generating the observed effects can be ruled out at some level of confidence. In this case, we test “randomness” by creating it artificially: that is, by using the permutation procedure to randomly shuffle the observed Y_k s across the X_k s and therefore “forcing” the association between Y_k and X_k to be stochastic (Manly 2007, 2). When the \widehat{p} associated with the observed $\widehat{\beta}$ is under a specified α , then $\widehat{\beta}$ was likely generated from a systematic process because its absolute size is larger than or equal to at least $1 - \alpha\%$ of the total $|\widehat{\beta}^*|$, which were generated via an artificial random data-generating process. The emphasis on quantifying the likelihood of a random versus systematic data-generating process as opposed to estimating the probability of observing a given observed effect of at least that size in some population under the condition that there is not one is the defining feature that differentiates the process model from the population model of inference.

2.2 Complicating the coefficient plot

Regression estimates generated under the randomization model pose unique problems for visualization. For the popular coefficient plot, the problem is quite fundamental: there are no SES or CIs for the coefficient, $\widehat{\beta}$, so a visualization such as the one in figure 1 above is not possible. SES and CIs are instead computed for the coefficient p -value, $\widehat{p}_{\widehat{\beta}}$. Importantly, though $SE(\widehat{\beta})$ and $CI(\beta)$ cannot be calculated without again assuming asymptotic distributional properties of $\widehat{\beta}$, $SE(\widehat{p}_{\widehat{\beta}})$ and $CI(p_{\widehat{\beta}})$ can be computed because they are with respect to k and not n (the observation sample size). Assuming that k is a random subset of the total number of possible permutations K , then $SE(\widehat{p}_{\widehat{\beta}})$ indicates the average deviation of any randomly sampled p -value $\widehat{p}_{\widehat{\beta}}$ from the true but unobserved $p_{\widehat{\beta}}$ if every possible permutation was considered.⁵ $CI(p_{\widehat{\beta}})$ then provides the

5. Of course, $p_{\widehat{\beta}}$ could be observed if every possible permutation can be accounted for.

range of p -values within which the true $p_{\hat{\beta}}$ would be located with a $1 - \alpha$ degree of confidence.⁶

The fact that SEs and CIs for p -values are still valid provides leverage for constructing a new type of coefficient plot for visualizing regression estimates and their subsequent predicted values under the randomization model. This new series of plots, which I refer to collectively as PPV plots, is outlined in the following section.

3 The PPV plot toolkit

A PPV plot can be thought of as a coefficient plot for regression models with significance levels determined via permutation tests. Its structure is quite simple: instead of arraying model coefficients within a variable-by- Y -unit space, the p -values for each model coefficient are arrayed within a variable-by- $[0, 1]$ space. Unlike a traditional coefficient plot centered around 0 to indicate no statistically significant effect, the PPV plot is centered around an analyst-determined alpha level. In this article, I use a combination of Stata's `permute` (see [R] `permute`) command and a modified version of Jann's (2014) `coefplot` command to construct PPV plots. The code to reproduce all tables and visualization is available as supplementary material.

I illustrate how PPV plots can be used to visualize regression coefficients and postestimation predicted values. Each strategy is addressed in turn below. First, however, I provide a brief overview of the case that I use for illustration: perceptions of communist threat among a nonprobability convenience sample of far-right activists who participated in the 1962 San Francisco Bay Region School of Anti-Communism (Wolfinger 1992).

3.1 The case: An anticommunism school

The Christian Anti-Communism Crusade (CACC) was founded in 1953 by Fred Schwarz. Schwarz, an Australian originally trained in medicine, spearheaded the “second wave of religious anti-Communism” in the United States (Herzog 2011, 207) and started CACC in an attempt to illuminate and teach on the insidious activities of communism—which he believed to be “the very mask of Satan” (The Schwartz Report 2018). CACC generated considerable funds by “promoting the belief that no bilateral negotiation could exist with communism” (Aiello 2005), largely through their tax-exempt and tuition-based “anti-Communism schools” that the organization would put together. In these schools,

6. $SE(\hat{p}_{\hat{\beta}})$ can be found with the SE formula for a proportion because $p_{\hat{\beta}}$ is necessarily bound between 0 and 1:

$$SE(\hat{p}_{\hat{\beta}}) = \sqrt{\frac{\hat{p}_{\hat{\beta}}(1 - \hat{p}_{\hat{\beta}})}{k}}$$

CI($p_{\hat{\beta}}$) is then found with

$$CI(p_{\hat{\beta}}) = \hat{p}_{\hat{\beta}} \pm \{z \times SE(\hat{p}_{\hat{\beta}})\}$$

Schwarz and other organizers would teach on the un-American perils of communism and put on fundraising activities (Wolfinger et al. 1964, 22).

The fleeting nature of these meetings posed problems for academic researchers wishing to survey the attendees at these “schools”. No clearly definable sampling frame of participants existed—and, even if it did, it was unlikely that such a list would have been provided to researchers given the high skepticism of colleges and universities amid the McCarthyism of the time. Nonetheless, Wolfinger and colleagues wished to know just who these school attendees were and what they believed. To this end, they adopted purposive sampling techniques to gather data from attendees at the San Francisco Bay Region School of Anti-Communism, held from January 29, 1962, to February 2, 1962, at the Oakland Auditorium in Oakland, California. Wolfinger and colleagues describe their data-collection process as follows:

We were able to conduct a study of the personal characteristics and political attitudes and activities of several hundred people who attended this Crusade school. Under our direction two dozen Stanford undergraduates interviewed members of the audience and distributed questionnaires (similar to the interview forms) which could be completed later and mailed back to Stanford. (Wolfinger et al. 1964, 23)

As the excerpt suggests, respondents constituted a convenience (and clearly non-probabilistic) sample. Given that participants were not known beforehand and were there for activities completely unrelated to data-collection efforts, data had to be gathered in a way that was not optimally systematic. Using the population model to draw statistical inferences with this dataset would be inappropriate because the sample was not a random “draw” from a defined target population of American anticommunists (or even the “population” of anticommunism school attendees), meaning that asymptotic distributional properties of probability sampling cannot be assumed. The randomization model, however, can be used to draw sample-specific inferences on the likelihood that statistical associations and relationships could arise from a random data-generating process. These data (Wolfinger 1992) are used in the remainder of this article to illustrate the utilities of the PPV plot toolkit. More specific information on the variables used are discussed below.

3.2 PPV plots and regression coefficients

Consider the following pseudo-research question: Among the participants at this school, what were some of the sociodemographic factors that predicted a heightened perception of impending communist threat (given, of course, an already heightened baseline perception of communist threat relative to the general population because they self-selected

into an anticommunism school)?⁷ To address this question, I constructed factor scores from a series of dummy variables related to different perceptions of communist dangers (for example, communism in academia and communism in the Democratic Party). Details on how these variables were constructed can be found in appendix A; higher scores indicate higher perceptions of communist threat. This variable was then regressed on the following variables: degree of political engagement, party identification, level of educational attainment, union member status, and whether the respondent was strictly self-employed.⁸ All predictor variables were observed variables, save for the political engagement score, which was also constructed using factor analysis on a series of dummy variables. (Details on how these factor scores were constructed are in appendix A.)

The results from this OLS model could be arranged within a table, as shown in table 1. This table is similar to a traditional regression table, with the only consequential difference being that the coefficient p -values rather than SEs are reported in the parentheses.

7. This research question is “pseudo” in the sense that, from the tradition of quantitative sociology, it identifies only one half of what would be expected from a proper question: the outcome, that is, heightened perceptions of communist threat. The “missing piece” here would be the specific IV (or IVs) that we theorize will account for variation in that outcome. For illustrative purposes, however, this will suffice.

8. I qualify self-employment with “strictly” because, in addition to the “self-employed” category, the survey also included an option for those who were self-employed and employed by someone else. I collapsed this “dual employment” category with those who were employed but not self-employed, retired, unemployed, or students.

Table 1. OLS model predicting heightened perceptions of communist threat (std) with permutation-derived p -values

IV	Model
Political engagement (std)	-0.236 (0.091)
Self-employed	-0.175 (0.547)
Republican	-0.525 (0.377)
Other	-0.670 (0.328)
Union member	0.268 (0.469)
Grammar school	0.155 (0.839)
College	-0.580 (0.133)
Constant	1.021 (0.114)
n	56
R^2 (adj.)	0.042
Root mean squared error	0.979
# of permutations of DV	1,000

NOTE: p -values in parentheses. Reference categories for the party and schooling variables are Democrat and high school, respectively. All schooling categories are “some/complete” for the corresponding level.

Alternatively, the model could be represented graphically as a simple PPV plot, as illustrated in figure 2. Like figure 1, the y axis lists the predictor variables. Unlike figure 2, the x axis can vary within the $[0, 1]$ interval and is used to plot the coefficient p -values and their associated CIs. The markers are the p -values, and the numbers above them are the coefficients themselves. The vertical line represents $\alpha = 0.05$ —meaning that any coefficient with a p -value to the left of the line is statistically significant at least at the 0.05 level. For instance, consider the self-employment coefficient, $\hat{\beta}_{SE} = -0.175$, which has a $\hat{p} = 0.547$. This p -value indicates that after randomly permuting the perception of communist threat variable across observations 1,000 times and recomputing the difference in intercept between those who were strictly self-employed and those who were not each time, 54.7% of the permutation estimates had absolute sizes that were greater than or equal to the absolute size of the observed estimate of -0.175 ; thus, one could say that at an alpha level of at least 0.05, it is highly possible that randomness

can account for the association between self-employment and perceptions of communist threat in this sample. Further, the p -value's 95% CI is far to the right of the vertical line—suggesting the p -value itself is not statistically significant (that is, it is likely greater than 0.05 if all permutations were accounted for and not just 1,000 of them).

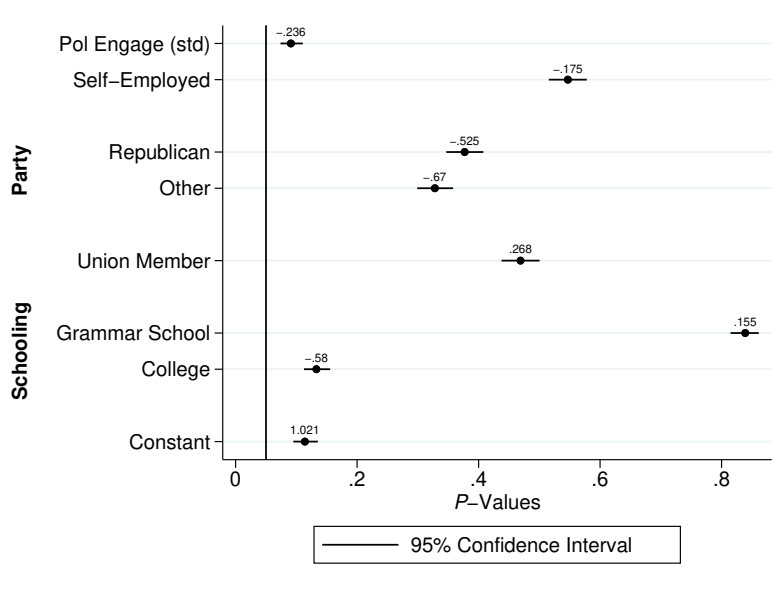


Figure 2. PPV plot of OLS estimated effects on perception of communist threat. NOTE: $n = 56$. Party reference category = Democrat; Schooling reference category = high school. All schooling categories are “some/complete” for the corresponding level. CIs are two sided and based on 1,000 permutation tests. Vertical line = 0.05. Dependent variable (DV) is standardized.

Conceptually, a PPV plot such as figure 2 is constructed in the following simple steps. First, the regression model is fit. Assume, for instance, the following linear-additive estimation model,

$$\widehat{Y}_i = \widehat{a} + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_m X_{mi}$$

Second, the data (in the examples in this article, the DV) are randomly permuted k number of times, with the p -value for each $\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_m$ determined following (1) above and the CIs following the equation in footnote 6. These statistics are then used to construct an m -by-3 matrix, \mathbf{D} ,

$$\mathbf{D} = \begin{bmatrix} \widehat{p}_{\widehat{\beta}_1} & \text{LL} \left(p_{\widehat{\beta}_1} \right) & \text{UL} \left(p_{\widehat{\beta}_1} \right) \\ \widehat{p}_{\widehat{\beta}_2} & \text{LL} \left(p_{\widehat{\beta}_2} \right) & \text{UL} \left(p_{\widehat{\beta}_2} \right) \\ \vdots & \vdots & \vdots \\ \widehat{p}_{\widehat{\beta}_m} & \text{LL} \left(p_{\widehat{\beta}_m} \right) & \text{UL} \left(p_{\widehat{\beta}_m} \right) \end{bmatrix}$$

where $\text{UL}(p_{\widehat{\beta}_m})$ is the upper CI limit $[\widehat{p}_{\widehat{\beta}} + \{z \times \text{SE}(\widehat{p}_{\widehat{\beta}})\}]$ for $p_{\widehat{\beta}_m}$ and $\text{LL}(p_{\widehat{\beta}_m})$ is the lower limit $[\widehat{p}_{\widehat{\beta}} - \{z \times \text{SE}(\widehat{p}_{\widehat{\beta}})\}]$.

Lastly, any coefficient plot function can be used to plot \mathbf{D} —such as the `coefplot` command in Stata (Jann 2014) or R (Lander 2018) or the `arm` package in R (Gelman et al. 2018). The goal is to simply map each p -value and its CI into a variable-by- $[0, 1]$ space and with a clear demarcation of the alpha level (on whichever axis represents the $[0, 1]$ interval) around which the CIs were constructed.

The fact that \mathbf{D} is treated like any other coefficient matrix means that the analyst can take advantage of virtually any visualization strategies that are available with coefficient plots. Nested models, for instance, can also be reported in the same PPV plot. If we treat the regression in figure 2 as nested within a larger model that also conditions perceptions of communist threat on sex, age, whether the respondent was Protestant, income level, and industry of employment, then we can combine the plots, as shown in figure 3. This PPV plot illustrates that, after controlling on this new set of control variables, none of the effect p -values from model 1 goes to the left of the 0.05 line—meaning that none of the estimates reaches statistical significance under the randomization model at $\alpha = 0.05$. However, the dashed vertical line—indicating $\alpha = 0.1$ —shows that, while a respondent’s level of political engagement is not associated with perceptions of communist threat beyond what we would expect under the condition of a random data-generating process before conditioning on the variables added in model 2, the observed estimate $\widehat{\beta}_{\text{PE}_2} = -0.281$ in model 2 is less likely to be due to random association when we relax our alpha level to 0.1.

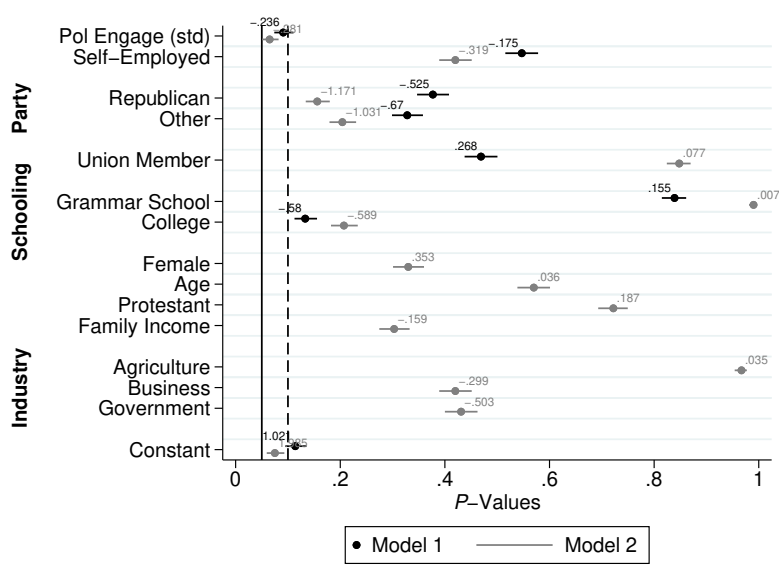


Figure 3. Nested PPV plot of OLS estimated effects on perception of communist threat. NOTE: $n = 56$ for both models. Party reference category = Democrat; Schooling reference category = high school; Industry reference category = service; reference category for Protestant is “Other”. All schooling categories are “some/complete” for the corresponding level. CIs are two sided and based on 1,000 permutation tests. Solid vertical line = 0.05; dashed vertical line = 0.1. DV is standardized.

Similarly, the analyst can reconfigure the matrix of p -values and CIs into other plot types. For example, figure 3 can be displayed as a bar chart (figure 4), where a p -value below 0.05 (now a horizontal solid line) indicates statistical significance with at least $\alpha = 0.05$. In this plot, spikes above the 0.05 reference line indicate an observed coefficient ($\hat{\beta}$) that was smaller in absolute size than at least 5% of the coefficients ($\hat{\beta}^*$) derived from rerunning the model after we randomly permuted the perceptions of communist threat values across observations 1,000 times.

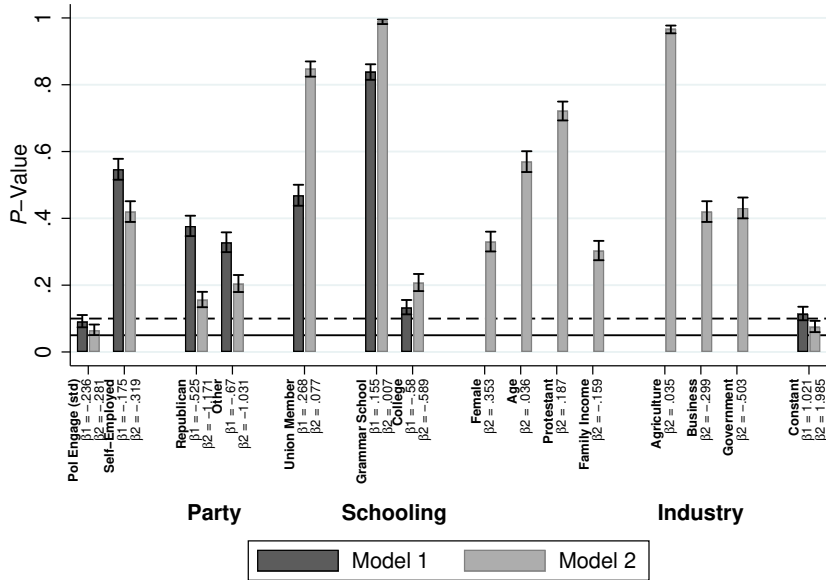


Figure 4. PPV plot reconfigured as a bar chart.
 NOTE: Solid horizontal line = 0.05; dashed horizontal line = 0.1. All other interpretation notes are the same as those for figure 3.

The substantive significance of the estimates can also be incorporated into PPV plots using standard graphical options. For example, assume an analyst is interested in the gap in perceptions of communist threat between those who were strictly self-employed and those who were not. They include an interaction between the political engagement scale and the dichotomous self-employment variable. They use 1,000 permutation tests and construct the PPV plot in the top panel of figure 5, which shows a statistically significant interaction ($\hat{\beta}_{PE \times SE} = 0.862$; $\hat{p}_{PE \times SE} = 0.007$, where $0.003 \leq p_{PE \times SE} \leq 0.014$). This PPV plot can also be reconfigured so that the marker shape indicates whether the effect is positive or negative and the color intensity represents the effect size of the standardized coefficient. The bottom panel of figure 5 introduces effect-size visualizations in lieu of the estimates themselves, with squares indicating negative estimates, circles indicating positive estimates, and the hue of the marker and CIs corresponding to the size of the standardized effect size. As the plot suggests, the interaction between political engagement and self-employment status was not only the only statistically significant predictor of the extent to which an attendee perceived heightened communist threats, but also one of the substantively strongest predictors (as indicated by the darker markers for the political engagement main effect and the engagement \times employment cross-product term).

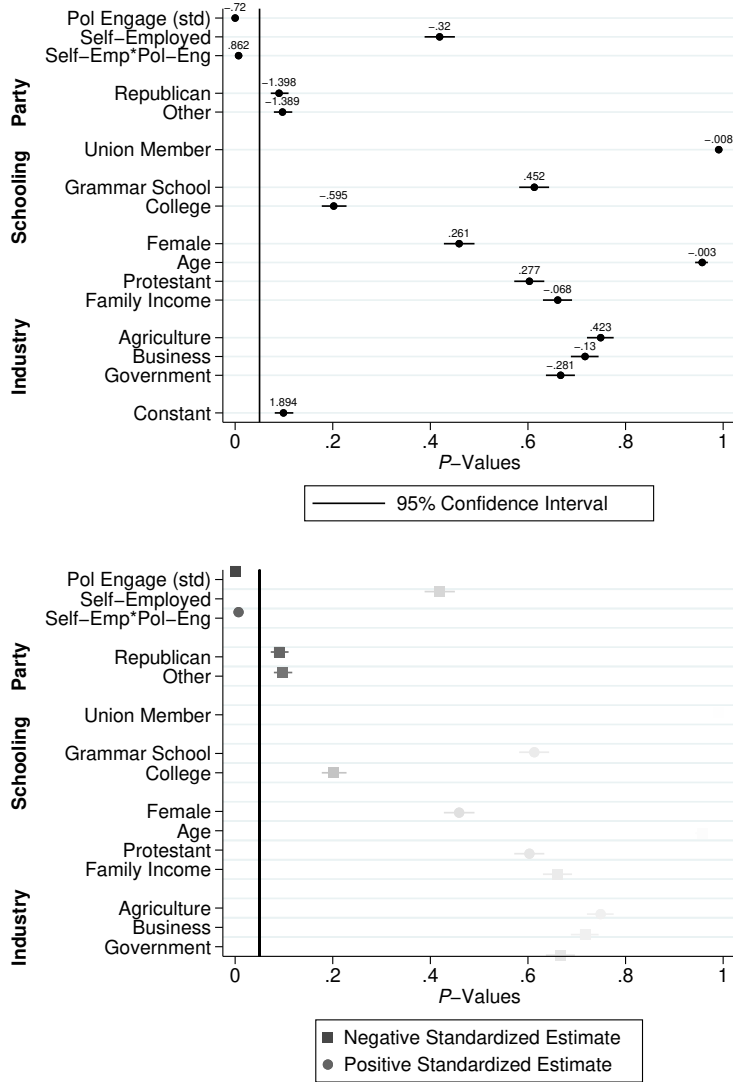


Figure 5. Multiplicative model with and without effect-size visualization.
 NOTE: $n = 56$. Party reference category = Democrat; Schooling reference category = high school; Industry reference category = service; reference category for Protestant is “Other”. All schooling categories are “some/complete” for the corresponding level. CIs are two sided and based on 1,000 permutation tests. Vertical line = 0.05. Bottom panel hues based on effect size of standardized estimates; however, estimates reported in the top graph are unstandardized. Darker markers in the bottom panel indicate larger absolute effect sizes.

3.3 PPV plots and predicted values

p -values for postestimation predicted values may be of little use under the population model of inference when zero is not a meaningful value in the units of Y . This is because the prediction p -values are typically derived from one-sample t - or z -tests of the null hypothesis that the “true” value for the hypothetical case in question is zero. The (two-tailed) prediction p -value, then, is just the probability of generating a predicted value of at least that absolute size if the true value for a case with that profile of IV characteristics were zero—again assuming that the sample used to derive that prediction is a random draw from the target population. This means that for DVs without a meaningful zero, most empirically plausible predictions with reasonable SEs will generate large test statistics and therefore small p -values. Further, because a statistically nonsignificant predicted value under the population model simply indicates a $1 - \alpha$ or greater probability of observing that predicted value for that case when the “true” value is 0, a predicted value for a DV with a meaningful zero (such as a standardized variable) could easily have $p < 0.05$ but still indicate a “nonzero” prediction.

With that said, it might be important for scholars basing statistical significance on permutation-based p -values to report the p -values for predictions. This is because permutation-based p -values are based on the differences between the observed prediction and the predictions from the permuted models and not on the difference of the prediction from zero. Thus, a two-tailed permutation-based p -value for a single observed prediction following a regression model indicates the proportion of predictions from the regressions on the permuted samples that are greater than or equal in absolute size to that observed prediction.⁹ This p -value, then, provides an estimate of the extent to which the observed prediction deviates from the predictions generated from models where the association between the DVs and IVs are random by design.

PPV plots can also be used to visualize these permutation-based prediction p -values. One can accomplish this by first writing a program that adds the calculation of predicted values as an extra step to each of the 1,000 permutation tests. Consider again, for instance, the multiplicative model used to construct the PPV plots in figure 5, which

9. Unless the predictions from the permuted samples take signs in the opposite direction of the observed prediction, then, for permutation-based p -values for predictions, the two-tailed and one-tailed tests will have the following equivalencies. If the observed prediction is positive and all of its k predictions from the permuted samples are positive, then

$$\left\{ \frac{I\left(\left|\widehat{\beta}_t^*\right| \geq \left|\widehat{\beta}\right|\right)}{k} \mid \left(\forall \widehat{\beta}_t^* \in \beta\right) \right\} \equiv \frac{I\left(\widehat{\beta}_t^* \geq \widehat{\beta}\right)}{k}$$

And, if the observed prediction is negative and all of its k predictions from the permuted samples are negative, then

$$\left\{ \frac{I\left(\left|-\widehat{\beta}_t^*\right| \geq \left|-\widehat{\beta}\right|\right)}{k} \mid \left(\forall -\widehat{\beta}_t^* \in \beta\right) \right\} \equiv 1 - \frac{I\left(-\widehat{\beta}_t^* \leq -\widehat{\beta}\right)}{k}$$

In other words, the two-tailed test will be equivalent to the one-tailed test to the right for positively valued predictions and the inverse of the one-tailed test to the left for negatively valued predictions.

showed a statistically significant interaction between an attendee’s level of political engagement and their self-employment status when it comes to heightened perceptions of communist threat. The analyst can now add one additional step to his or her permutation procedure for obtaining permutation-derived p -values for postestimation predictions: after the analyst randomly permutes the DV k times and fits the regression model, a series of predictions can be generated for each permutation and compared with the predictions from the observed regression model. This analysis could result in something akin to table 2, which presents predicted perceptions of communist threat (on a standardized scale) for both the self-employed and those who are not, within a standard deviation (SD) of the sample mean political engagement scale.

Table 2. Predicted standardized perceptions of communist threat (std) from OLS model

IV	Model
Not S-E’s political engagement (std)	
–1 SD	0.771** (0.002)
Mean	0.051 (0.789)
+1 SD	–0.669* (0.033)
S-E’s political engagement (std)	
–1 SD	–0.411 (0.224)
Mean	–0.269 (0.224)
+1 SD	–0.127 (0.672)
<hr/>	
n	56
# of permutations of DV	1,000

NOTE: p -values in parentheses. S-E is “self-employed”.

* $p < 0.05$; ** $p < 0.01$ (two-tailed tests)

Alternatively, these predictions could be displayed as a PPV plot that allows for a visual comparison of the relative statistical significance levels, as shown in figure 6. These predicted-value PPV plots can also be easily reconfigured into other plot types— for example, as a bar chart (figure 7) similar to the coefficient PPV bar chart in figure 4. Or, alternatively, the prediction p -values can be displayed as a line graph for continuous variables. Figure 8 provides an example, with separate political engagement p -value “slopes” for hypothetical respondents of each employment category.

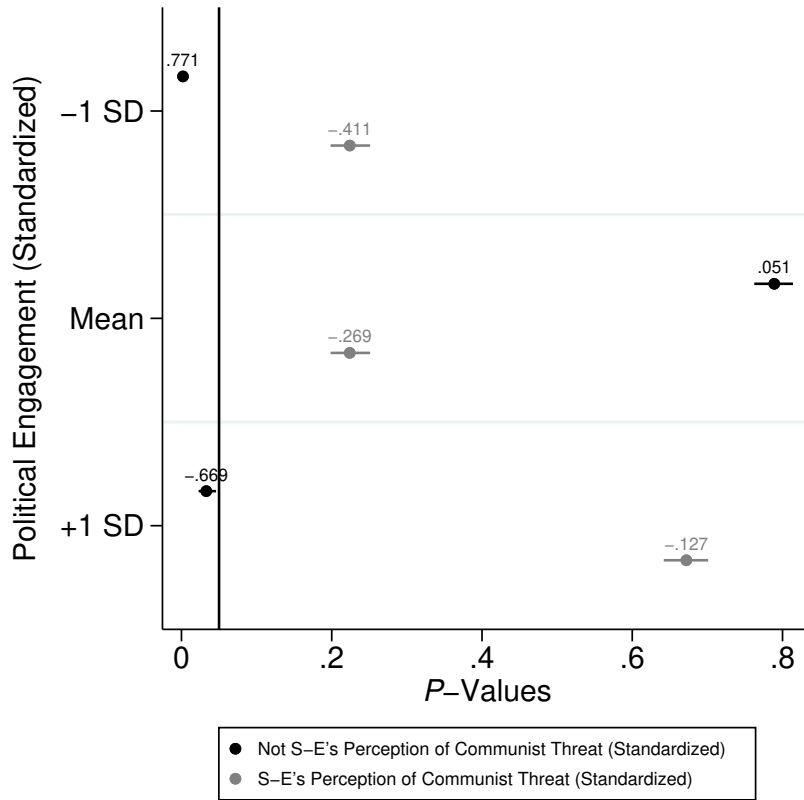


Figure 6. PPV plot of predicted standardized perceptions of communist threat by self-employment and level of political engagement

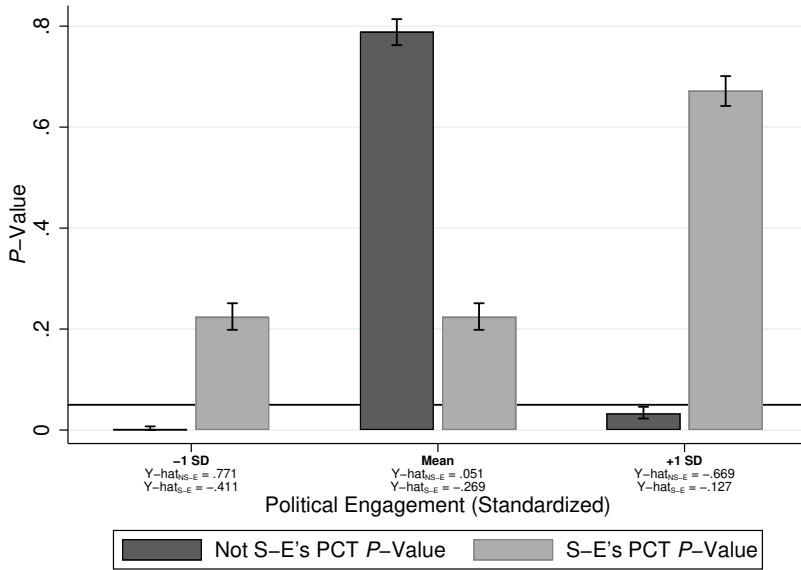


Figure 7. Predicted value PPV plot reconfigured as a bar chart

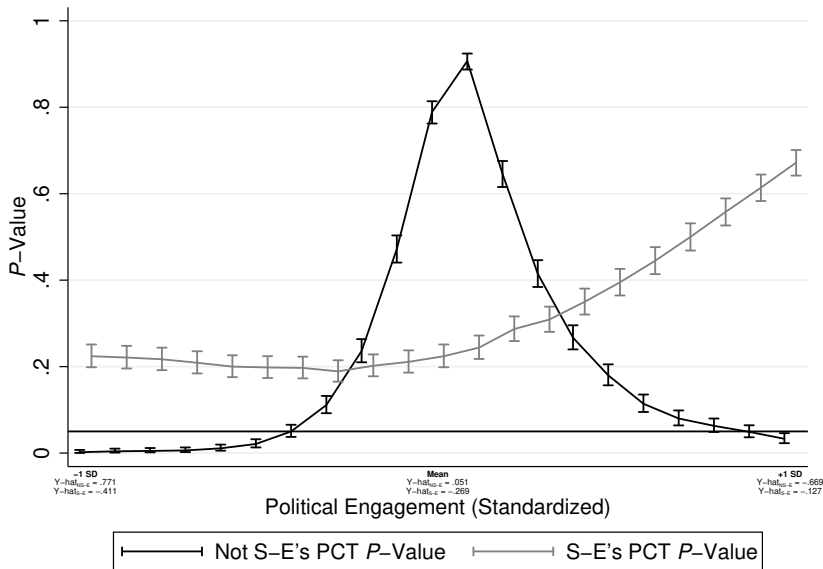


Figure 8. Predicted value PPV plot reconfigured as a line graph

Figures 6, 7, and 8 indicate that higher levels of political engagement are negatively associated with higher perceptions of communist threat for those who are not strictly self-employed (according to the switch in signs for marker labels in figure 6 and the x -axis labels in figures 7 and 8 for those that are not self-employed), while the estimate remains negative for the self-employed but loses some of its effect size (as indicated by the smaller negative effect sizes for those that are self-employed). However, these figures suggest that this interaction may be asymmetric. Among these 56 school attendees, only those who are not strictly self-employed with lower than average and higher than average levels of political engagement have predicted levels of perceptions of communist threat that deviate from what would be expected under the condition of a random data-generating process. Non-self-employed (NSE) respondents with lower than average levels of political engagement exhibit slightly above average perceptions of threat ($\hat{Y}_{\text{NSE}} = 0.771$; $0.000 \leq p_{\hat{Y}_{\text{NSE}}} \leq 0.007$), and those with higher than average levels of political engagement exhibit slightly below average perceptions ($\hat{Y}_{\text{NSE}} = -0.669$; $0.023 \leq p_{\hat{Y}_{\text{NSE}}} \leq 0.046$).

The NSE with average levels of political engagement and the self-employed at any level of political engagement have predicted threat perceptions that never deviate from what we would expect if the predictions were computed from a model of random associations. If this hypothetical analyst were hoping to draw substantive conclusions from these results—and assuming a properly specified regression model (which these models, because they are for illustrative purposes, may not be)—he or she may conclude something to the following effect: Among these 56 school attendees, those who were not strictly self-employed but who were less politically engaged than the average participant were more likely to have higher than average perceptions of a communist threat—but, when they were more politically engaged than the average participant, they tended to have some of the weakest perceptions of communist threat.

The asymmetry of this interaction would likely have been more difficult to pick up without adding prediction calculations to the permutation tests, and PPV plots certainly make this task of comparing effects across groups that much easier.

4 Summary and conclusions

Recent interest in data analytics and computational social science has sparked a flurry of activity in data visualization among social scientists (Healy 2019; Wickham and Grolemund 2017). One popular graph type for displaying regression models is the coefficient plot. While coefficient plots are useful for simultaneously visualizing both the statistical and substantive significance of the estimates, their applicability assumes that statistical significance is determined under the population model of inference—that is, that the coefficients have SEs and CIs, derived either with reference to theoretical sampling distributions or to some asymptotically based resampling technique such as the bootstrap or jackknife (Shao and Tu 1995). The PPV plot “toolkit” introduced in this article offers a way of leveraging the intuitiveness, parsimony, and aesthetic simplicity of the coefficient plot for regression models with statistical significance levels derived via the randomization model of inference, where SEs and CIs are reported for the coefficient

p -values rather than the coefficients themselves and where a plot centered at zero is less useful.

In this article, I illustrated the utility of PPV plots for visualizing permutation-based regression estimates and postestimation predictions using the case of heightened perceptions of threat among 56 anticommunists at an anticommunism school. In both cases, PPV plots leverage the same benefits of coefficient plots in that they visualize the direction and size of the estimates while concomitantly incorporating estimate uncertainty into the graph. For coefficient plots, the uncertainty is captured with the CIs around the estimates: the wider the interval, the more uncertainty there is about where the true population parameter may be. For PPV plots, the uncertainty is captured with the CI around the p -value. Specifically, the wider the interval, the more uncertainty there is about where the true p -value would be had every possible permutation been accounted for. Accounting for this uncertainty is particularly important for boundary estimates: that is, when a p -value indicates a statistically significant estimate with a CI that spans across alpha levels or a nonsignificant estimate that spans to the left of the alpha level necessary to establish statistical significance. PPV plots can be particularly helpful in the case of a model with many predictors because they can simultaneously display the number of boundary estimates. Importantly, because the size of the p -value CIs is a function of the number of permutations performed, PPV plots in this case can also provide a general sense of how much closer the analyst should approximate an exact test (that is, account for every possible permutation) to shrink the CIs so that the number of boundary estimates is minimized.

There will undoubtedly be the question of how much traction visualization strategies for randomization inference will get in sociology, where this type of statistical inference is not the disciplinary norm. Indeed, if the population model is far and away the more common mode of inference, then what real leverage do PPV plots provide? This concern is itself symptomatic of a fundamental question about quantitative inferential reasoning: namely, is it really inference if there is no population to which one can infer things (compare Ludbrook and Dudley [1998, 128])? Such a perspective is likely a partial function of the fact that the inferential statistics curricula for most statistics courses in sociology are predicated on the logic that inference is synonymous with generalizing to a population. Ironically, though, it is the other type of inference that seems to predominate the actual practice of sociological research, because, as network analysts have observed for some time, “most social science aims at studying mechanisms rather than describing specific populations” (Snijders 2011, 135). Perhaps to no surprise, this implicit reliance on mechanistic reasoning seems particularly important in subfields where datasets constructed using rigorous probabilistic sampling procedures are by far the exception rather than the norm—for example, cultural sociology, social movement studies, and organizational sociology. Following this, I conclude this article with a more general call for quantitative sociologists to reflect critically on what sort of statistical inferences are appropriate given their sample design and, by extension, on how they choose to “see” this significance.

5 Supplemental materials

A replication repository for this article can be found at <https://github.com/Marshall-Soc/PPV>.

6 Acknowledgments

I thank the editorial team at the *Stata Journal* and an anonymous reviewer for their helpful feedback. I also thank Omar Lizardo and Dustin S. Stoltz for their thoughts on earlier versions of this article.

7 References

- Aiello, T. 2005. Constructing “godless communism”: Religion, politics, and popular culture, 1954–1960. *Americana: The Journal of American Popular Culture* 4(1).
- Anderson, M. J., and J. Robinson. 2001. Permutation tests for linear models. *Australian & New Zealand Journal of Statistics* 43: 75–88. <https://doi.org/10.1111/1467-842X.00156>.
- Aronow, P. M. 2012. A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research* 41: 3–16. <https://doi.org/10.1177/0049124112437535>.
- Bähr, S., and M. Abraham. 2016. The role of social capital in the job-related regional mobility decisions of unemployed individuals. *Social Networks* 46: 44–59. <https://doi.org/10.1016/j.socnet.2015.12.004>.
- Baker, F. B., and L. J. Hubert. 1981. The analysis of social interaction data: A nonparametric technique. *Sociological Methods & Research* 9: 339–361. <https://doi.org/10.1177/004912418100900305>.
- Butts, C. T. 2007. Permutation models for relational data. *Sociological Methodology* 37: 257–281. <https://doi.org/10.1111/j.1467-9531.2007.00183.x>.
- Darlington, R. B., and A. F. Hayes. 2017. *Regression Analysis and Linear Models: Concepts, Applications, and Implementation*. New York: Guilford Press.
- Edgington, E. S., and P. Onghena. 2007. *Randomization Tests*. 4th ed. Boca Raton, FL: Chapman & Hall/CRC.
- Ernst, M. D. 2004. Permutation methods: A basis for exact inference. *Statistical Science* 19: 676–685. <https://doi.org/10.1214/088342304000000396>.
- Freedman, D., and D. Lane. 1983. A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics* 1: 292–298. <https://doi.org/10.2307/1391660>.

- Friedman, S., and D. Laurison. 2017. Mind the gap: Financial London and the regional class pay gap. *British Journal of Sociology* 68: 474–511. <https://doi.org/10.1111/1468-4446.12269>.
- Gadermann, A. M., M. Guhn, and B. D. Zumbo. 2012. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation* 17(3).
- Gelman, A., Y.-S. Su, M. Yajima, J. Hill, M. G. Pittau, J. Kerman, T. Zheng, and V. Dorie. 2018. *arm: Data analysis using regression and multilevel/hierarchical models*. R package version 1.10-1. <https://cran.r-project.org/web/packages/arm/>.
- Healy, K. 2019. *Data Visualization: A Practical Introduction*. Princeton, NJ: Princeton University Press.
- Herzog, J. P. 2011. *The Spiritual-Industrial Complex: America's Religious Battle against Communism in the Early Cold War*. New York: Oxford University Press.
- Jann, B. 2014. Plotting regression coefficients and other estimates. *Stata Journal* 14: 708–737. <https://doi.org/10.1177/1536867X1401400402>.
- Krackhardt, D. 1988. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks* 10: 359–381. [https://doi.org/10.1016/0378-8733\(88\)90004-4](https://doi.org/10.1016/0378-8733(88)90004-4).
- Kubinger, K. D. 2003. On artificial results due to using factor analysis for dichotomous variables. *Psychology Science* 45: 106–110.
- Lander, J. P. 2018. *coefplot: Plots coefficients from fitted models*. R package version 1.2.6. <https://cran.r-project.org/web/packages/coefplot/>.
- Laurison, D., and S. Friedman. 2016. The class pay gap in higher professional and managerial occupations. *American Sociological Review* 81: 668–695. <https://doi.org/10.1177/0003122416653602>.
- Lizardo, O., and S. Skiles. 2016a. Cultural objects as prisms: Perceived audience composition of musical genres as a resource for symbolic exclusion. *Socius: Sociological Research for a Dynamic World* 2: 1–17. <https://doi.org/10.1177/2378023116641695>.
- . 2016b. The end of symbolic exclusion? The rise of “categorical tolerance” in the musical tastes of Americans: 1993–2012. *Sociological Science* 3: 85–108. <https://doi.org/10.15195/v3.a5>.
- Ludbrook, J., and H. Dudley. 1998. Why permutation tests are superior to t and F tests in biomedical research. *American Statistician* 52: 127–132. <https://doi.org/10.1080/00031305.1998.10480551>.
- Manly, B. F. J. 2007. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC.

- O'gorman, T. W. 2006. An adaptive test of a subset of regression coefficients using permutations of residuals. *Journal of Statistical Computation and Simulation* 76: 1095–1103. <https://doi.org/10.1080/10629360600565376>.
- Schührrer, S., W. Carbonaro, and E. Grodsky. 2016. Reproduction of inequality in educational attainment through curricular differentiation in secondary school—A case study of the USA. In *Models of Secondary Education and Social Inequality: An International Comparison*, ed. H.-P. Blossfeld, S. Buchholz, J. Skopek, and M. Triventi, 249–267, 249–267. Northampton, MA: Edward Elgar Publishing.
- Shao, J., and D. Tu. 1995. *The Jackknife and Bootstrap*. New York: Springer.
- Snijders, T. A. B. 2011. Statistical models for social networks. *Annual Review of Sociology* 37: 131–153. <https://doi.org/10.1146/annurev.soc.012809.102709>.
- Stewart, E., P. Edgell, and J. Delehanty. 2018. The politics of religious prejudice and tolerance for cultural others. *Sociological Quarterly* 59: 17–39. <https://doi.org/10.1080/00380253.2017.1383144>.
- Stoltz, D. S., and M. A. Taylor. 2017. Paying with change: The purposeful enunciation of material culture. *Poetics* 64: 26–39. <https://doi.org/10.1016/j.poetic.2017.07.003>.
- The Schwartz Report. 2018. About us. Manitou Springs, CO: CACC & The Schwartz Report. <https://www.schwarzreport.org/about>.
- Thomann, E., and M. Maggetti. Forthcoming. Designing research with qualitative comparative analysis (QCA): Approaches, challenges, and tools. *Sociological Methods & Research*. <https://doi.org/10.1177/0049124117729700>.
- Thomson, G. H. 1951. *The Factorial Analysis of Human Ability*. 5th ed. London: University of London Press.
- Wickham, H., and G. Grolemund. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol, CA: O'Reilly.
- Wolfinger, R. E. 1992. America's radical right, 1962 [cumulative file] (ICPSR 07273). Ann Arbor, MI. Inter-university Consortium for Political and Social Research. <https://doi.org/10.3886/ICPSR07273.v1>.
- Wolfinger, R. E., B. K. Wolfinger, K. Prewitt, and S. Rosenhack. 1964. Crusaders for the right. *Trans-action* 1: 22–27. <https://doi.org/10.1007/BF02813614>.

About the author

Marshall A. Taylor is an assistant professor of sociology at New Mexico State University. His areas of research interest include cultural and cognitive sociology, political sociology, and computational social science. His published work has appeared in such venues as *Sociological Theory*, *Poetics*, *Journal of Computational Social Science*, *American Journal of Cultural Sociology*, *Socius*, and the *Stata Journal*, among others.

A Principal factor analyses

Six variables were used to construct a “perception of communist threat” variable. The variables were all dichotomous; thus, tetrachoric correlations were decomposed to extract the principal factors (Gadermann, Guhn, and Zumbo 2012; Kubinger 2003). “Communist danger” indicates whether the respondent sees American communists as a danger to the country at that moment, where 0 = no/some/little danger and 1 = a great or very great deal of danger. “Communist professors” indicates whether the respondent believes communist professors have a lot of influence in U.S. higher education, where 0 = no and 1 = yes. “Communist Democrats” indicates whether the respondent sees communists as having a lot of influence in the U.S. Democratic Party, where 0 = no and 1 = yes. “Communist speech” indicates whether the respondent believes the U.S. Communist Party should be allowed to speak on radio or TV, where 0 = no and 1 = yes. “Communist neighbor” indicates whether the respondent believes there are communists living in their neighborhood, where 0 = no and 1 = yes. Table 3 provides the interitem tetrachoric correlations.

Table 3. Tetrachoric correlations between “perceptions of communist threat” variables

	C danger	C professors	C Democrats	C speech	C neighbor	Local C
C danger	1					
C professors	0.560	1				
C Democrats	0.580	0.739	1			
C speech	-0.415	-0.213	-0.011	1		
C neighbor	0.235	0.446	0.256	-0.056	1	
Local C	0.752	0.426	0.497	-0.079	0.559	1

Table 4 contains the loadings for the first principal factor after orthogonal varimax rotation. All variables load in the expected direction (positive) for a “perceptions of communist threat” latent variable. This factor explains as much variance as approximately 1.821 of the raw indicator variables (eigenvector, $\lambda = 1.821$)—that is, about 45.67% of the total variance. Factor weights were derived using the regression method (Thomson 1951). Higher values indicate heightened perceptions that communists pose a clear and imminent threat.

Table 4. Factor loadings for table 3

Variable	Loadings
Communist danger	0.367
Communist professors	0.600
Communist Democrats	0.594
Communist speech	0.168
Communist neighbor	0.674
Local communism	0.707
λ	1.821
% variance explained	45.67%
n	116

NOTE: Principal factor loadings are with orthogonal varimax rotation.

Six dichotomous variables were also used to create a “political engagement” variable, again using tetrachoric correlations and principal factor analysis. The interitem correlations are presented in table 5. “Attend meetings” indicates whether the respondent goes to political rallies, meetings, and dinners, where 0 = no and 1 = yes. “Write political letters” indicates whether the respondent, in the past year before being surveyed, ever wrote letters or sent telegrams to the president, congressperson, or state politician, where 0 = no and 1 = yes. “Organizational member” indicates whether the respondent was a member of a political organization, where 0 = no and 1 = yes. “Influence voters” indicates whether the respondent ever tried to convince people to vote one way or another, where 0 = no and 1 = yes. “Give money/buy tickets” indicates whether the respondent ever gave money or bought tickets to try to help a candidate or his or her campaign or party. Finally, “work for the candidate” indicates whether the respondent did any other work for a candidate or party, where 0 = no and 1 = yes.

Table 5. Tetrachoric correlations between “political engagement” variables

	Attend	Letters	Org member	Voters	Money/ tickets	Work
Attend	1					
Letters	0.466	1				
Member	0.658	0.120	1			
Voters	0.318	0.145	0.442	1		
Money/tickets	0.698	0.329	0.630	0.347	1	
Work	0.554	0.222	0.802	0.423	0.582	1

NOTE: Correlations based on 190 valid cases after listwise deletion. All variables are dichotomous.

Table 6 lists the loadings for these variables on the first principal factor after orthogonal varimax rotation. All variables load in the direction consistent with a “political engagement” latent variable (that is, they are all positive). This “political engagement” factor explains as much variance as approximately 2.297 raw indicator variables (about 72.55% of the total variance). The regression method was used to compute the factor weights. Higher scores indicate that the respondent is more politically active.

Table 6. Factor loadings for table 5

Variable	Loadings
Attend meetings	0.541
Write political letters	0.071
Organizational member	0.888
Influence voters	0.459
Give money/buy tickets	0.588
Work for candidate	0.808
λ	2.297
% variance explained	72.55%
n	190

NOTE: Principal factor loadings are with orthogonal varimax rotation.