The Stata Journal (2022)
22, Number 1, pp. 134–157 DOI: 10.1177/1536867X221083901

Effect sizes for contrasts of estimated marginal effects

Brian P. Shaw
Indiana University
Bloomington, IN
bpshaw@indiana.edu

Abstract. The statistical literature is replete with calls to report standardized measures of effect size alongside traditional p-values and null hypothesis tests. While effect-size measures such as Cohen's d and Hedges's g are straightforward to calculate for t tests, this is not the case for parameters in more complex linear models, where traditional effect-size measures such as η^2 and ω^2 face limitations. After a review of effect sizes and their implementation in Stata, I introduce the community-contributed command `mces`. This postestimation command reports standardized effect-size statistics for dichotomous comparisons of marginal-effect contrasts obtained from `margins` and `mimrgns`, including with complex samples, for continuous outcome variables. `mces` provides Stata users the ability to report straightforward estimates of effect size in many modeling applications.

Keywords: st0667, mces, svysd, effect size, margins, esize, marginal effects, contrasts of marginal effects

1 Introduction

Classical frequentist statistical inference involves calculating p-values, or the probability that a null hypothesis would be observed in the target population given the data. It is well known that p-values are often misinterpreted as the probability that the null hypothesis is true (Cohen 1994). This widespread misunderstanding, combined with a raft of criticism admonishing both researchers and consumers that statistical significance does not imply clinical or practical significance, has led many voices in the field of statistics to encourage a move toward reporting standardized measures of effect size alongside or in place of traditional null hypothesis significance tests (for example, Kline [2013]; Trafimow and Marks [2015]; Wasserstein and Lazar [2016]; and Ziliak and McCloskey [2008]). This article begins with a brief primer on standardized effect sizes and illustrates ways that they are traditionally estimated in Stata. Because the estimation of marginal effects is a core Stata capability, I review various ways of manually calculating effect sizes for `margins` results. These manual calculations are cumbersome, and in some cases impossible, using existing methods. Accordingly, I introduce a new command, `mces`, that facilitates computing contrasts of postestimation marginal effects for continuous outcome variables using `margins`. I then demonstrate its use with some of the same examples.

2 Overview of effect-size measures

While null hypothesis significance testing concerns whether "no effect is unlikely", measures of effect size report whether an observed effect is "large in magnitude". Because of differences in ways that effects are estimated, there can be no single way that effect sizes are calculated or reported. When researchers refer to "effect sizes", they are almost always referring to standardized measures of effect size, which permit unit-neutral comparisons across studies and are a central tool of meta-analysis (Vacha-Haase and Thompson 2004). The literature frequently describes three "families" of effect sizes with similar properties, depending on the nature of the variables in question and the estimation procedure (Ellis 2010). While effect sizes are not infallible (for example, Cheung and Slavin [2016]), they are still preferable to reporting p-values alone because they report fundamentally different information (Kelley and Preacher 2012). While the terms "treatment group" and "control group" are the language of experimentation and are used in this article for exposition, the logic is the same for any binary demarcation of group membership, such as `i.male` or `i.collgrad`.

2.1 The "d" family of effect sizes

The "d" family reports magnitudes in terms of group mean differences. For continuous outcomes, these measures are variations on the generic formula $(M_T - M_C)/\text{SD}$: the mean difference between the treatment group and the control group, standardized by dividing by the standard deviation. The most familiar of these may be Cohen's d (Cohen 1988), which involves dividing the differences in means by the pooled standard deviation (1):

$$d = \frac{M_T - M_C}{\sqrt{\frac{\sum(X_T - \bar{X}_T)^2 + \sum(X_C - \bar{X}_C)^2}{n_T + n_C - 2}}} \tag{1}$$

While Cohen's d is likely familiar to readers, it is not the only statistic for standardized mean difference. Glass's Δ (Glass, McGaw, and Smith 1981) reports the effect size using the standard deviation for the control group on the theory that this estimates average treatment effects for future untreated populations (2). It is also useful for small samples, when estimates of the standard deviation could be unstable.

$$\Delta = \frac{M_T - M_C}{\text{SD}_C} \tag{2}$$

Hedges's g (Hedges 1981) uses a pooled standard deviation that is weighted by the relative sample sizes of the two groups (3). Hedges's g is similar to Cohen's d, but Cohen's d has been shown to be positively biased in small samples.

$$g = \frac{M_T - M_C}{\sqrt{\frac{\text{SD}_T^2 \cdot (n_T - 1) + \text{SD}_C^2 \cdot (n_C - 1)}{n_T + n_C - 2}}} \tag{3}$$

Hedges's g and Cohen's d are equivalent in large samples and will be similar to Glass's Δ when the two groups have similar standard deviations.

2.2 The "r" family of effect sizes

In contrast to the "d" family's emphasis on mean differences, the "r" family of effect-size measures revolves around the "proportion of variance accounted for". The most familiar of these may be the squared multiple correlation coefficient (the "coefficient of determination"), or R^2, and its "corrected" corollary, the adjusted R^2, which incorporates information about sample size and number of predictors. These range in value from 0 to +1. While software traditionally reports R^2 in regression analyses, with ANOVA "correlation index" values η^2 and ω^2 (sometimes ϵ^2) are more common, even though in linear relationships they are functionally equivalent to R^2 and adjusted R^2 (Cohen et al. 2003). Because they are measures of the proportion of variance accounted for, both R^2 and η^2 are figured by dividing the variance explained by the model (which may be as simple as a one-factor ANOVA or as complex as a structural equation model) by the total variance observed, as in (4).

$$R^2 \equiv \eta^2 = \frac{\mathrm{SS}_{\mathrm{Model}}}{\mathrm{SS}_{\mathrm{Total}}} = \frac{\text{Variance explained}}{\text{Total variance}} \tag{4}$$

Partial η^2 and ω^2, on the other hand, have a slightly different formula (5), which is comparable with (4) in a one-way ANOVA. However, in complex models, they can differ widely, and there is a great deal of published literature that appears to conflate the two (Levine and Hullett 2002). Readers may be expecting η^2 and ω^2 statistics to report the proportion of the "total" variance explained, as calculated in (4). Levine and Hullett recommended that partial η^2 be reported, but other authors recommend the opposite (for example, Tabachnick and Fidell [2019] and Olejnik and Algina [2003]).

$$\text{Partial } \eta^2 = \frac{\mathrm{SS}_{\mathrm{Explained}}}{\mathrm{SS}_{\mathrm{Explained}} + \mathrm{SS}_{\mathrm{Error}}} \tag{5}$$

Use of measures such as η^2 to report relative-effect magnitude has been criticized in the literature on regression (for example, Pedhazur [1997]). The standardized regression coefficient β is also occasionally advocated as an analogue to effect size, but it is not an ideal method to convey the magnitude of an effect across studies (Greenland et al. 1986; Pedhazur 1997).

Effect sizes for categorical outcomes are also members of the "r" family. While they are related to r, more common measures of effect size for contingency tables (that is, categorical data) are coefficient ϕ, Cramér's V, Kendall's τ, and Cohen's w. Equation 6 demonstrates the formula for these measures in a 2×2 table. While Cramér's V can be calculated for multiway tables, ϕ is only estimated for two dichotomous variables, and Pearson's r is only equivalent to V and ϕ in that case.

$$V = \sqrt{\frac{\chi^2}{n}} = \phi \equiv r \tag{6}$$

Kendall's τ is a nonparametric measure of association that does not use the χ^2 statistic but rather ordinal "concordances" and "discordances". There are three formulas for Kendall's τ, according to whether the table is square and whether to account for ties.

Stata reports Kendall's τ_b, which is determined by the number of concordances (C), the number of discordances (D), the number of ties (T), and the number of observations (n), shown in (7).

$$\tau_b = \frac{C - D}{\sqrt{\{n(n-1)/(2 - T_X)\}\{n(n-1)/(2 - T_Y)\}}} \tag{7}$$

Because the "d" and the "r" families are both undergirded by the same general linear model, they imply the same meaning, and researchers can use formulas to convert not only within families but also from one family to another (Vacha-Haase and Thompson 2004).

2.3 The "OR" family of effect sizes

Applied most commonly to categorical data and particularly generalized linear models, the odds ratio reports the odds of an outcome given a treatment or condition, relative to the odds of the outcome in the absence of that treatment or condition. Odds are figured from probabilities according to the formula $\pi/(1 - \pi)$, where π is the probability of a "yes" result for a dichotomous outcome variable.[1] Odds are accordingly the expected number of "yes" results for every "no".

Odds ratios report magnitudes of association as a multiplier for the increase or decrease in odds for a one-unit change in a continuous predictor or, for categorical variables, membership in one category relative to another. For example, an odds ratio of 2 for a binary regression predictor variable `i.urban` implies that the odds of "yes" are twice as high for cities coded as urban as they are for those not coded as urban. Relative risk is an analogous standardized effect-size statistic measured in raw probabilities rather than odds. However, because relative risk has skewed sampling distributions, odds ratios are preferred in many fields.

3 Effect sizes in Stata

Stata has methods for estimating each family of effect-size measures. For an additional overview, see Huber (2013).

3.1 The "d" family in Stata

Stata's base `esize` command reports "d" family effect sizes. The `unequal` option, which is generally recommended, requests that Stata use a pooled standard deviation rather than making the (strong) assumption that both groups have equal variances.

1. Models such as ordinal and nominal logistic regression, which involve categorical outcome variables with more than two levels, still estimate odds and odds ratios relative to one of the other levels.

```
. sysuse auto
(1978 automobile data)
. esize twosample mpg, by(foreign) unequal cohensd hedgesg glass
Effect size based on mean comparison, unequal variances
                                Obs per group:
                                     Domestic =         52
                                      Foreign =         22
------------------------------------------------------------------------------
        Effect size |   Estimate     [95% conf. interval]
--------------------+---------------------------------------------------------
          Cohen's d |  -.9234449    -1.466436   -.3679697
        Hedges's g  |  -.9137865    -1.451098    -.364121
    Glass's Delta 1 |  -1.042693    -1.576083   -.5004845
    Glass's Delta 2 |  -.7480963    -1.287558   -.1938573
------------------------------------------------------------------------------
            Satterthwaite's degrees of freedom =  30.5463
```

`esize` reports two values of Glass's Δ, one using the standard deviation from the first group ("Glass's Delta 1", which is `Domestic` in this output) and one using the standard deviation from the second group ("Glass's Delta 2", which here is `Foreign`). In this example, Cohen's d and Hedges's g values are similar,[2] while Glass's Δ values are quite different, even though they all use the same $M_C - M_T$ numerator. A closer examination of the sample statistics is instructive.

```
. table foreign, statistic(count mpg) statistic(sd mpg) nototals
------------------------------------------------------------------------------
             | Number of nonmissing values   Standard deviation
-------------+----------------------------------------------------------------
Car origin   |
  Domestic   |                          52                4.743297
  Foreign    |                          22                6.611187
------------------------------------------------------------------------------
```

The standard deviations and sample sizes for the two groups are both different. After some algebra, we determine that the pooled unweighted standard deviation used for Cohen's d is 5.36, while the pooled weighted standard deviation used for Hedges's g is similar at 5.41. Both are closer to the `Domestic` standard deviation because the `Domestic` n is larger. The difference between $\text{SD}_{\texttt{Domestic}}$ and $\text{SD}_{\texttt{Foreign}}$ is responsible for the discrepancy between the two values of Glass's Δ. In practice, the control group standard deviation will typically be closer to the population standard deviation, so the appropriate value of Δ is almost always the control group—in this case, "Glass's Delta 1".

3.2 The "r" family in Stata

Stata provides several ways to estimate the "r" family of effect sizes. The `correlate` command and the related `pwcorr` command report the Pearson correlation coefficient,

2. According to [R] **esize**, Stata uses a different formula for d than the formula found on page 44 of Cohen's (1988) seminal text. The formula for d in the [R] **esize** documentation is described as the formula for Hedges's g elsewhere (for example, Ellis [2010] and Durlak [2009]).

which is the archetypal r. Coefficient r is equivalent to coefficient ϕ in a 2×2 table but is less often reported for contingency tables.

```
. webuse nhanes2
. correlate highbp sex
(obs=10,351)
             |   highbp      sex
-------------+------------------
      highbp |   1.0000
         sex |  -0.0886   1.0000
```

Coefficient ϕ, Cramér's V, and Kendall's τ are obtained in Stata with the `tabulate twoway` command. As expected, the value of r from the `correlate` command equals the estimated Cramér's V because both variables are dichotomous.

```
. tabulate highbp sex, chi V taub
High blood |          Sex
  pressure |      Male     Female |     Total
-----------+----------------------+----------
         0 |     2,611      3,364 |     5,975
         1 |     2,304      2,072 |     4,376
-----------+----------------------+----------
     Total |     4,915      5,436 |    10,351

          Pearson chi2(1) =  81.1787   Pr = 0.000
               Cramér's V =  -0.0886
          Kendall's tau-b =  -0.0886  ASE = 0.010
```

The χ^2 test indicates statistical significance, but a small p-value is not an indication of how strongly these variables are associated with one another. The values of Cramér's V and Kendall's τ_b report that, on a scale from 0 (independence) to 1.0 (perfect association), sex's association with the incidence of high blood pressure is less than 0.1. Coefficient ϕ is not reported, but in a 2×2, table it is equal to Cramér's V. Here is another example, this time with a 2×5 table:

```
. tabulate highbp agegrp, chi V taub
High blood |                    Age group
  pressure |     20-29      30-39      40-49      50-59      60-69 |     Total
-----------+-----------------------------------------------------+----------
         0 |     1,928      1,167        770        590      1,193 |     5,975
         1 |       392        455        502        701      1,667 |     4,376
-----------+-----------------------------------------------------+----------
     Total |     2,320      1,622      1,272      1,291      2,860 |    10,351

High blood | Age group
  pressure |       70+ |     Total
-----------+-----------+----------
         0 |       327 |     5,975
         1 |       659 |     4,376
-----------+-----------+----------
     Total |       986 |    10,351

          Pearson chi2(5) =  1.4e+03   Pr = 0.000
               Cramér's V =  0.3640
          Kendall's tau-b =  0.3185  ASE = 0.008
```

The p-value again leads us to reject the null hypothesis of no differences between the groups. Even though the tests for sex and age report p-values of 0.000, the estimated effect sizes show that age has a much stronger association with blood pressure than sex does. Taken together, these results underscore the importance of effect sizes in the interpretation of statistical test results, because without the effect sizes, the analyst could miss the critical differences in magnitude between the two associations.

Next we will consider the "r" family effect sizes in the context of linear models such as regression and analysis of variance. We begin with a simple `regress` specification (although typing `anova hgb sex hsizgp` is equivalent) followed by `estat esize` with and without the `omega` option.

```
. regress hgb sex hsizgp
```

Source	SS	df	MS	Number of obs	=	10,351
				F(2, 10348)	=	2327.79
Model	6157.66008	2	3078.83004	Prob > F	=	0.0000
Residual	13686.7097	10,348	1.32264299	R-squared	=	0.3103
				Adj R-squared	=	0.3102
Total	19844.3698	10,350	1.91733041	Root MSE	=	1.1501

hgb	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
sex	-1.545825	.0226569	-68.23	0.000	-1.590237	-1.501413
hsizgp	-.0309753	.0084939	-3.65	0.000	-.047625	-.0143256
_cons	16.70453	.0441952	377.97	0.000	16.6179	16.79116

The value of R^2 reported in the regression table is equivalent to η^2, and the adjusted R^2 is equivalent to ω^2. In this instance with few predictors, η^2 and ω^2 (and R^2 and adjusted R^2) are similar. They will diverge with the addition of more predictors.

```
. estat esize
Effect sizes for linear models
```

Source	Eta-squared	df	[95% conf.	interval]
Model	.3102976	2	.2965246	.3237334
sex	.3102711	1	.2965655	.323773
hsizgp	.0012835	1	.0002746	.0030286

```
Note: Eta-squared values for individual model terms are partial.
. estat esize, omega
Effect sizes for linear models
```

Source	Omega-squared	df
Model	.3101436	2
sex	.3101838	1
hsizgp	.0011869	1

```
Note: Omega-squared values for individual
      model terms are partial.
```

The `regress` and `anova` commands report coefficients, t-values, and significance tests for each coefficient, along with R^2 and adjusted R^2, but only the postestimation `estat size` command reports η^2 or ω^2. These statistics can also be informative (compare Pedhazur [1997]). While both the `sex` and `hsizgrp` variables have p-values of less than 0.001 and are "statistically significant", the partial ω^2 shows that `sex` is much more strongly associated with hemoglobin levels when controlling for house size. In this instance, the t-values also suggest a difference in strength of association, but this is not always the case. The community-contributed command `pcorr2` (Williams 2003) also reports partial correlation coefficients. In simple models, squared partial correlation coefficients from `pcorr2` are equivalent to η^2. Another community-contributed command, `esizereg` (Linden 2019), reports Cohen's d effect sizes for a single regression coefficient.

3.3 The "OR" family in Stata

Odds ratios are most often calculated for logistic and ordered logistic regression models using `logit` or `ologit`. Because the default coefficients of these models are uninterpretable log odds, the `or` option requests exponentiated odds ratios instead, which facilitates interpretation. (The `logistic` command requests odds ratios by default, the same as `logit, or`.) Multinomial (also called "polytomous") logistic regression models using `mlogit` with the `rrr` option report the similarly interpreted "relative-risk ratio", and models for count outcomes, such as `poisson` and `nbreg`, calculate the "incidence-rate ratio" with the `irr` option. As an example, consider a logistic regression analysis modeling risk factors for diabetes. The research question might be, "Do sex, age, or body mass index predict the likelihood of a person being diagnosed with diabetes?"

```
. logistic diabetes i.sex bmi i.agegrp
Logistic regression                                     Number of obs = 10,349
                                                        LR chi2(7)    = 411.88
                                                        Prob > chi2   = 0.0000
Log likelihood = -1793.8169                             Pseudo R2     = 0.1030

------------------------------------------------------------------------------
    diabetes | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         sex |
      Female |   1.080252   .1031392     0.81   0.419     .8958901    1.302552
         bmi |    1.07604   .0089067     8.85   0.000     1.058724    1.093639
             |
      agegrp |
       30-39 |   1.770952   .6029793     1.68   0.093     .9086303    3.451649
       40-49 |   4.459514   1.355222     4.92   0.000      2.45817    8.090271
       50-59 |   7.403023   2.127383     6.97   0.000     4.215043    13.00218
       60-69 |   11.81136   3.171803     9.19   0.000      6.97782     19.9931
         70+ |   16.80083   4.683804    10.12   0.000      9.72813    29.01564
             |
       _cons |   .0010223   .0003422   -20.57   0.000     .0005305    .0019702
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

The model reports that females in the dataset have odds of being diagnosed with diabetes that are 1.08 times higher than males after controlling for the effects of the other

predictors—not a huge difference. Critically, this is not the same as the "probability" of a diabetes diagnosis being 1.08 times higher. Odds use the formula $\pi/(1-\pi)$, and they are not linearly related to predicted probabilities. To obtain predicted probabilities, we use the `margins` command. Measures of effect size for regression models are elaborated further in the next section.

The coefficient for body mass index (BMI) also rounds to 1.08, but because BMI is a continuous predictor, the interpretation is that for each ceteris paribus one-unit increase in BMI, the odds of a diabetes diagnosis are expected to increase by a factor of 1.08, which would mean that a four-unit increase in BMI should predict odds of a diabetes diagnosis that are 1.36 times higher. And, relative to the base category of 20–29-year-olds, those aged 50–59 have odds of diabetes that are approximately 7.4 times higher. More information on odds ratios and their interpretations in Stata are available in Long and Freese (2014) and Mitchell (2021).

4 Effect sizes for marginal effects

The interpretation of regression coefficients is less straightforward when models become complex. When transformations, interactions, and polynomials are specified and combined, individual model coefficients can lose their clear, substantive meaning. Fortunately, as experienced Stata users know, the ability to easily calculate postestimation predicted values from even the most complicated models using the `margins` command is one of Stata's core capabilities. Just as interpretation of regression coefficients becomes more difficult with increasing model complexity, so does the interpretation of the "r" family of effect sizes. While η^2 and ω^2 have straightforward interpretations in simpler models, they offer less clarity on the magnitude of a variable's effects on the outcome when models become complex, especially without careful centering and hand calculations of "simple slopes" (Aiken and West 1991). η^2 and ω^2 values also do not leverage the flexible specifications of `margins`. The `pwcompare` option for `margins` can be used to produce the $M_T - M_C$ component of the formula for the "d" family of effect sizes with many types of regression models. However, as we shall see, calculating a valid standard deviation can be a challenge.

4.1 Marginal effect sizes for categorical regression predictors

Suppose we are interested in whether systolic blood pressure is higher for females after controlling for BMI, race, and hemoglobin levels. We might specify the following regression model:

```
. webuse nhanes2, clear
. generate log_bmi = log(bmi)
. regress bpsystol c.log_bmi##c.log_bmi i.race i.sex##c.hgb
```

Source	SS	df	MS		
Model	764657.969	7	109236.853	Number of obs	= 10,351
Residual	4870012.06	10,343	470.851016	F(7, 10343)	= 232.00
				Prob > F	= 0.0000
				R-squared	= 0.1357
				Adj R-squared	= 0.1351
Total	5634670.03	10,350	544.412563	Root MSE	= 21.699

bpsystol	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
log_bmi	-10.11869	27.98112	-0.36	0.718	-64.96711	44.72972
c.log_bmi# c.log_bmi	8.344867	4.280715	1.95	0.051	-.0461622	16.7359
race						
Black	1.657852	.7227593	2.29	0.022	.241104	3.0746
Other	.6872607	1.553159	0.44	0.658	-2.357231	3.731752
sex						
Female	-44.75682	5.333441	-8.39	0.000	-55.2114	-34.30224
hgb	-1.114247	.2693264	-4.14	0.000	-1.642178	-.5863149
sex#c.hgb						
Female	2.912048	.3710501	7.85	0.000	2.184718	3.639378
_cons	95.04859	45.79675	2.08	0.038	5.278105	184.8191

The regression table does not provide a simple answer to the question of whether males or females are predicted to have higher systolic blood pressure. The command `margins, pwcompare` uses all available information, including all model terms and the proportions of the sample with each distribution of the covariates.

```
. margins sex, asobserved pwcompare
Pairwise comparisons of predictive margins              Number of obs = 10,351
Model VCE: OLS
Expression: Linear prediction, predict()
```

	Contrast	Delta-method std. err.	Unadjusted [95% conf.	interval]
sex				
Female vs Male	-3.229743	.5258416	-4.260494	-2.198992

Accounting for all model predictors, `margins` reports that females in the sample are predicted to have a systolic blood pressure 3.23 points lower than that of males. This difference is clearly statistically significant, but by default `margins` does not report an effect size, and very small p-values are not an indication of a large effect. It is possible to calculate the effect size by hand using Stata's base `esizei` command. First, we need to store the estimated contrast in the scalar we choose to name `diff` and then use `summarize` to store the within-group means and standard deviations.

```
. scalar diff = el(r(b_vs),1,1)
. quietly summarize bpsystol if sex==1
. scalar sd1 = r(sd)
. scalar n1 = r(N)
. quietly summarize bpsystol if sex==2
. scalar sd2 = r(sd)
. scalar n2 = r(N)
```

Now that the necessary summary statistics are in memory, `esizei` will report the effect size for the marginal comparison between males' and females' values on the outcome after adjusting for all the predictors in the regression equation, using the values for each group from the `margins` command. We use the stored coefficient from `margins, pwcompare` for the first group and a zero for the other.

```
. esizei `=n1' `=diff' `=sd1' `=n2' 0 `=sd2', unequal
Effect size based on mean comparison, unequal variances
                                Obs per group:
                                      Group 1 =      4,915
                                      Group 2 =      5,436
------------------------------------------------------------------------------
        Effect size |   Estimate     [95% conf. interval]
--------------------+---------------------------------------------------------
          Cohen's d |  -.1388798    -.1775009    -.100252
        Hedges's g  |  -.1388698    -.1774881   -.1002447
------------------------------------------------------------------------------
           Satterthwaite's degrees of freedom =  1.0e+04
```

The regression-adjusted difference between males' and females' systolic blood pressure is approximately 0.14 standard deviations. Field-specific context informs a judgment about whether this is a large or small difference. Nevertheless, if we are confident in our `margins` specification, we can be confident in the interpretation of the estimated "d" family effect size.

4.2 Marginal effect sizes for continuous regression predictors

The approach in section 4.1 applies only to dichotomous variables. For continuous predictors, the analyst can report ω^2 or η^2 statistics or use `margins` to create a dichotomous comparison. Here we analyze the differences between females with high and low levels of hemoglobin.

```
. quietly summarize hgb if sex==2
. scalar fem_hihgb = `r(mean)' + `r(sd)'
. scalar fem_lowhgb = `r(mean)' - `r(sd)'
. margins, at(hgb=(`=fem_hihgb' `=fem_lowhgb') sex=2) asobserved pwcompare(effects)
Pairwise comparisons of predictive margins              Number of obs = 10,351
Model VCE: OLS

Expression: Linear prediction, predict()
1._at: sex =        2
       hgb = 14.66133
2._at: sex =        2
       hgb = 12.39485
```

	Contrast	Delta-method std. err.	t	P>|t|	Unadjusted [95% conf.	interval]
_at						
2 vs 1	-4.074689	.6022449	-6.77	0.000	-5.255206	-2.894173

The predicted difference in systolic blood pressure between female subjects with hemoglobin levels 1-standard deviation higher than the mean and 1-standard deviation lower than the mean is 4.07 points. This difference is statistically significant at $\alpha = 0.05$. Determining practical significance using measures of effect size necessitates figuring the standard deviations. This is complicated by the fact that the approach used to calculate the standard deviation in section 4.1 is not directly available when the predictor in question is continuous. Furthermore, it usually does not make sense to estimate the standard deviation only for cases with a very specific hemoglobin level. For example, in this dataset containing over 10,000 cases, none have a rounded value of `hgb` that equals the grand mean of 14.3.

There is no single accepted method to define the standard deviations for calculating effect sizes when predictors are continuous. One possibility is to simply use the standard deviation of the outcome variable for both the high and the low values of the continuous predictor (following Cohen et al. 2003). This is analogous to Glass's Δ approach, so the `delta` option is specified.

```
. margins, at(hgb=(`=fem_hihgb' `=fem_lowhgb') sex=2) asobserved pwcompare(effects)
  (output omitted)
. scalar diff = el(r(b_vs),1,1)
. quietly summarize bpsystol if sex==2
. scalar sd1 = r(sd)
. scalar sd2 = r(sd)
. scalar n1 = r(N)
. scalar n2 = r(N)
```

```
. esizei `=n1' `=diff' `=sd1' `=n2' 0 `=sd2', glass
Effect size based on mean comparison
                                Obs per group:
                                      Group 1 =      5,436
                                      Group 2 =      5,436
------------------------------------------------------------------------------
        Effect size |   Estimate     [95% conf. interval]
--------------------+---------------------------------------------------------
    Glass's Delta 1 |  -.1621648    -.1998752   -.1244396
    Glass's Delta 2 |  -.1621648    -.1998752   -.1244396
------------------------------------------------------------------------------
```

The estimated difference of 0.16 standard deviations helps the analyst understand the magnitude of the effect, which the small p-value does not.

4.3 Marginal effect sizes for recoded continuous regression predictors

There is an even simpler and more straightforward approach to find the denominator: divide cases into a small number of groups based on their value of the continuous predictor, and then substitute the new categorical variable in the regression equation. The advantage of this approach is that there are clearly defined groups for comparison and for calculating the standard deviation. If theory suggests logical thresholds, then Stata's `recode` and `generate` commands are useful for creating the groups. If not, then splitting the variable into quantiles with `xtile` would also suffice. In this example, there is theoretical guidance for establishing a threshold: hemoglobin levels for men are regarded as elevated when they are above 17 grams per deciliter, while the standard for women is 15 grams per deciliter (Cleveland Clinic 2018). We can create indicator variables for a high hemoglobin level and then reestimate the regression using the indicator variable in place of the continuous measure of hemoglobin. Moving from a continuous measure to a categorical measure results in some loss of efficiency. Still, if a marginal comparison is central to the analysis, the ease of calculating an effect size may justify small reductions in R^2 values.

```
. webuse nhanes2, clear
. svyset, clear
. keep if sex==2
(4,915 observations deleted)
. generate log_bmi = log(bmi)
. generate fem_hihgb = 0
. replace fem_hihgb = 1 if hgb > 15
(452 real changes made)
```

```
. regress bpsystol c.log_bmi##c.log_bmi i.race##i.fem_hihgb
```

Source	SS	df	MS	Number of obs	=	5,436
				F(7, 5428)	=	149.29
Model	553998.315	7	79142.6164	Prob > F	=	0.0000
Residual	2877433.64	5,428	530.109366	R-squared	=	0.1614
				Adj R-squared	=	0.1604
Total	3431431.95	5,435	631.358225	Root MSE	=	23.024

bpsystol	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
log_bmi	39.0894	35.34388	1.11	0.269	-30.19879	108.3776
c.log_bmi# c.log_bmi	1.302717	5.385973	0.24	0.809	-9.25595	11.86138
race						
Black	1.370566	1.038928	1.32	0.187	-.6661497	3.407282
Other	-.5987116	2.466003	-0.24	0.808	-5.433066	4.235643
1.fem_hihgb	6.186013	1.175418	5.26	0.000	3.881723	8.490304
race#fem_hihgb						
Black#1	-9.243625	5.951811	-1.55	0.120	-20.91156	2.424311
Other#1	3.525941	8.579992	0.41	0.681	-13.29429	20.34617
_cons	-10.96456	57.8521	-0.19	0.850	-124.3779	102.4488

```
. margins fem_hihgb, pwcompare(effects)
```

```
Pairwise comparisons of predictive margins                Number of obs = 5,436
Model VCE: OLS

Expression: Linear prediction, predict()
```

	Contrast	Delta-method std. err.	t	Unadjusted P>|t|	Unadjusted [95% conf.	interval]
fem_hihgb 1 vs 0	5.252469	1.216172	4.32	0.000	2.868284	7.636654

```
. scalar diff = el(r(b_vs),1,1)
. quietly summarize bpsystol if fem_hihgb == 1
. scalar sd1 = r(sd)
. scalar n1 = r(N)
. quietly summarize bpsystol if fem_hihgb == 0
. scalar sd2 = r(sd)
. scalar n2 = r(N)
```

```
. esizei `=n1' `=diff' `=sd1' `=n2' 0 `=sd2', hedges
Effect size based on mean comparison
                                Obs per group:
                                      Group 1 =        452
                                      Group 2 =      4,984
------------------------------------------------------------------------------
        Effect size |   Estimate     [95% conf. interval]
--------------------+---------------------------------------------------------
        Hedges's g  |   .2101564     .1138004     .306493
------------------------------------------------------------------------------
```

These calculations show that females with clinically elevated hemoglobin levels are predicted to have a systolic blood pressure that is 0.21 standard deviations higher than those without high hemoglobin levels. However, estimating the standardized effect size for a postestimation contrast is cumbersome, particularly with multiply imputed data. Furthermore, `esizei` does not work with data that are `svyset` or `mi svyset`.

5 The mces and svysd commands

The `mces` command calculates one of three "d"-family effect-size measures for between-group contrasts of marginal effects obtained after `margins, pwcompare post`. The default effect size, which I am calling the root mean squared error (RMSE)-based Δ, uses the RMSE of the regression as the denominator in a calculation similar to (1)–(3). Hedges's g and Cohen's d can be requested as well by specifying a binary grouping variable used to calculate the pooled standard deviations for the same equations. The community-contributed `mimrgns` command[3] (Klein 2016) for marginal effects with multiply imputed data is also supported, as are complex survey designs specified with `svyset` or `mi svyset`.

Alternatively, if an estimate of only the survey-adjusted standard deviation is desired, the `svysd` command can be used independently of any `margins` results.

5.1 Syntax

The syntax after `margins, pwcompare post` or `mimrgns, pwcompare post` is

`mces` [`,` `sdbyvar(`*varname*`)` `hedgesg` `cohensd` `sdupdate` `nowarning` `force`]

The syntax to calculate a survey-adjusted standard deviation only is

`svysd` *depvar*`,` `sdbyvar(`*varname*`)` [`unweighted` `nowarning` `force`]

The `mces` command should work with most models based on linear regression that store coefficients from `margins, pwcompare post` and `mimrgns, pwcompare post` in

3. `mimrgns` is available from the Statistical Software Components Archive.

the macro `e(b_vs)`, such as `regress`, `truncreg`, `sem` and `gsem`, and `tobit`. `mces` is not appropriate for multilevel models, because it does not account for intraclass correlation (see Lorah [2018]) nor for categorical outcomes or generalized linear models that do not have standard deviations or RMSEs.

A further note about `mces` is that it estimates the standard deviations based upon all cases in the dataset and not only those used in the estimation. When requesting Hedges's g or Cohen's d, users may wish to run `keep if e(sample)` prior to `mces` if out-of-sample cases should not be used.

5.2 Options

`sdbyvar(`*varname*`)` specifies a dichotomous variable defining the comparison groups. `sdbyvar()` is required with `svysd`.

`hedgesg` (`mces` only) requests Hedges's g instead of the default RMSE-based Δ.

`cohensd` (`mces` only) requests Cohen's d instead of the default RMSE-based Δ.

`sdupdate` (`mces` only) requests a recalculation of the standard deviation, which is useful if the dataset has changed since the standard deviation was last calculated.

`unweighted` (`svysd` only) requests the unweighted pooled standard deviation used for Cohen's d instead of the default weighted pooled standard deviation used for Hedges's g.

`nowarning` suppresses warning messages about applicability of the standard deviation to the estimated pairwise comparisons.

`force` bypasses a check of whether the outcome variable is continuous.

5.3 Stored Results

`mces` and `svysd` store the following in `r()`:

Scalars		
	`r(RMSE)`	the estimated RMSE used for RMSE-based Δ
	`r(sdstar)`	sd*, the pooled weighted standard deviation used for Hedges's g
	`r(pooledsd)`	the unweighted pooled standard deviation used for Cohen's d
	`r(n_sdbyvar_at_#)`	the sample size in the group `sdbyvar(#)`
	`r(sd_sdbyvar_at_#)`	the standard deviation for the group `sdbyvar(#)`
Macros		
	`r(depvar)`	the outcome variable
	`r(sdbyvar)`	the `margins` variable used to mark groups for the standard deviation
Matrices		
	`r(Delta)`	the estimated RMSE-based Δ values
	`r(g)`	the estimated Hedges's g-values
	`r(d)`	the estimated Cohen's d-values

5.4 Example

`mces` can be used to streamline the process outlined in section 4.3. The reported Hedges's g from `mces` is equal to the value computed by hand in that example.

```
. webuse nhanes2, clear
. svyset, clear
. keep if sex==2
(4,915 observations deleted)
. generate log_bmi = log(bmi)
. generate fem_hihgb = 0
. replace fem_hihgb = 1 if hgb > 15
(452 real changes made)
. regress bpsystol c.log_bmi##c.log_bmi i.race##i.fem_hihgb
```

Source	SS	df	MS		
Model	553998.315	7	79142.6164	Number of obs	= 5,436
Residual	2877433.64	5,428	530.109366	F(7, 5428)	= 149.29
				Prob > F	= 0.0000
Total	3431431.95	5,435	631.358225	R-squared	= 0.1614
				Adj R-squared	= 0.1604
				Root MSE	= 23.024

bpsystol	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
log_bmi	39.0894	35.34388	1.11	0.269	-30.19879	108.3776
c.log_bmi# c.log_bmi	1.302717	5.385973	0.24	0.809	-9.25595	11.86138
race						
Black	1.370566	1.038928	1.32	0.187	-.6661497	3.407282
Other	-.5987116	2.466003	-0.24	0.808	-5.433066	4.235643
1.fem_hihgb	6.186013	1.175418	5.26	0.000	3.881723	8.490304
race#fem_hihgb						
Black#1	-9.243625	5.951811	-1.55	0.120	-20.91156	2.424311
Other#1	3.525941	8.579992	0.41	0.681	-13.29429	20.34617
_cons	-10.96456	57.8521	-0.19	0.850	-124.3779	102.4488

```
. margins fem_hihgb, asobserved pwcompare(effects) post
Pairwise comparisons of predictive margins                Number of obs = 5,436
Model VCE: OLS
Expression: Linear prediction, predict()
```

	Contrast	Delta-method std. err.	Unadjusted t	P>|t|	Unadjusted [95% conf.	interval]
fem_hihgb 1 vs 0	5.252469	1.216172	4.32	0.000	2.868284	7.636654

```
. mces, hedgesg sdbyvar(fem_hihgb)
Calculating values of Hedges's g...
```

	Contrast	Weighted SD*	Hedges's g
1 vs 0	5.25	24.99	0.21

The next example demonstrates an application of `mces` with data that are multiply imputed and have a complex sampling design. The first step is to set up and fit a regression model.

```
. webuse nhanes2, clear
. mi set mlong
. mi register imputed diabetes
(2 m=0 obs now marked as incomplete)
. mi impute chained (logit) diabetes = bpsystol female race age bmi, rseed(1111)
> add(5)
  (output omitted)
. mi svyset [pw=finalwgt], psu(psu) strata(strata) singleunit(centered)
  (output omitted)
```

```
. mi estimate: svy: regress bpsystol i.female##i.diabetes i.race##i.female age
Multiple-imputation estimates                   Imputations       =            5
Survey: Linear regression                       Number of obs     =       10,351
Number of strata  =        31                   Population size   =  117,157,513
Number of PSUs    =        62
                                                Average RVI       =       0.0000
                                                Largest FMI       =       0.0000
                                                Complete DF       =           31
DF adjustment:   Small sample                   DF:     min       =        29.18
                                                        avg       =        29.18
                                                        max       =        29.18
Model F test:       Equal FMI                   F(   8,   29.2)   =       314.32
Within VCE type:   Linearized                   Prob > F          =       0.0000
```

bpsystol	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
female						
Female	-6.85529	.5038991	-13.60	0.000	-7.885608	-5.824971
diabetes						
Diabetic	5.795767	2.012836	2.88	0.007	1.680137	9.911396
female# diabetes						
Female # Diabetic	3.745692	2.732983	1.37	0.181	-1.842418	9.333802
race						
Black	1.268734	.9392677	1.35	0.187	-.65178	3.189247
Other	-4.29781	1.349159	-3.19	0.003	-7.056424	-1.539196
race#female						
Black#Female	3.532286	1.268824	2.78	0.009	.9379297	6.126641
Other#Female	5.756211	1.991604	2.89	0.007	1.683994	9.828429
age	.6327208	.0165571	38.21	0.000	.5988667	.6665749
_cons	103.2413	.80308	128.56	0.000	101.5992	104.8833

Because the data are `mi set`, we use the `mimrgns` command to produce the pairwise comparisons. This example uses a more complex `at()` statement, which `mces` supports. By default, the RMSE-based Δ effect size is requested.

```
. mimrgns female, at(diabetes=(0 1) (median) age) pwcompare post
note: option predict() not specified; predict(xb) assumed
Multiple-imputation estimates                Imputations       =            5
Pairwise comparisons of predictive margins   Number of obs     =       10,351
Number of strata  =        31                Population size   =  117,157,513
Number of PSUs    =        62
                                             Average RVI       =       0.0000
                                             Largest FMI       =       0.0000
                                             Complete DF       =           31
DF adjustment:   Small sample                DF:     min       =        29.18
                                                     avg       =        29.18
Within VCE type: Delta-method                        max       =        29.18
Expression   : Linear prediction, predict(xb)
1._at        : diabetes        =           0
               age             =          40 (median)
2._at        : diabetes        =           1
               age             =          40 (median)
```

	Contrast	Std. err.	[95% conf.	interval]
_at#female				
(1#Female) vs (1#Male)	-6.372075	.4855671	-7.323769	-5.420381
(2#Male) vs (1#Male)	5.795767	2.012836	1.850681	9.740852
(2#Female) vs (1#Male)	3.169383	1.674325	-.1122333	6.451
(2#Male) vs (1#Female)	12.16784	2.034996	8.179323	16.15636
(2#Female) vs (1#Female)	9.541459	1.721093	6.168178	12.91474
(2#Female) vs (2#Male)	-2.626383	2.676626	-7.872473	2.619707

```
. mces
The estimated RMSE from this regression is 18.61616698876345
Calculating values of RMSE-based Δ...
WARNING: Results from mces only apply to ceteris paribus comparisons between groups
defined by a single dichotomous variable. Otherwise, the results are invalid.
Ensure that this condition applies to each line of the margins results.
You may want to run margins, pwcompare post followed by mces once per dichotomous
comparison.
```

	Contrast	RMSE	RMSE-based Δ
(_at=1 # Female) vs (1 # Male)	-6.37	18.62	-0.34
(_at=2 # Male) vs (1 # Male)	5.80	18.62	0.31
(_at=2 # Female) vs (1 # Male)	3.17	18.62	0.17
(_at=2 # Male) vs (1 # Female)	12.17	18.62	0.65
(_at=2 # Female) vs (1 # Female)	9.54	18.62	0.51
(_at=2 # Female) vs (2 # Male)	-2.63	18.62	-0.14

The results include a warning message that warrants explanation. `mces` does not attempt to second-guess the analysis and uses the estimated RMSE or standard deviation to calculate an effect-size statistic for each line in the `margins` output. However, `margins, pwcompare` estimates all possible pairwise comparisons between the variables, regardless of their properties. The analyst must therefore be careful to ensure that the dichotomous grouping variable (the `sdbyvar(`*varname*`)` option, if specified) is the only difference in each comparison. Otherwise, the results are invalid and should not be

considered. For a typical six-comparison output from `margins, pwcompare` with two binary grouping variables (such as one `at()` and one `over()`), the first and last lines will be the only two that meet these conditions. For clarity and to avoid errors, separate `margins` or `mimrgns` statements can be used to formally specify the desired comparison.

```
. quietly mi estimate: svy: regress bpsystol i.female##i.diabetes
> i.race##i.female age
. quietly mimrgns female, at(diabetes=(0) (median) age) pwcompare post
. mces
The estimated RMSE from this regression is 18.61616698876345
Calculating values of RMSE-based Δ...
```

	Contrast	RMSE	RMSE-based Δ
Female vs 0	-6.37	18.62	-0.34

Because the `post` option in the `margins, pwcompare` command has cleared the stored regression results, the model needs to be refit for additional comparisons using `mces`. `estimates store` can help to save run time if the model is computationally intensive.

```
. quietly mi estimate: svy: regress bpsystol i.female##i.diabetes
> i.race##i.female age
. quietly mimrgns female, at(diabetes=(1) (median) age) pwcompare post
. mces
The estimated RMSE from this regression is 18.61616698876345
Calculating values of RMSE-based Δ...
```

	Contrast	RMSE	RMSE-based Δ
Female vs 0	-2.63	18.62	-0.14

6 Conclusion

Both regression-based modeling and standardized effect-size measures are increasingly prevalent in applied quantitative research, yet existing effect sizes for complex regression models have been unsatisfying. `mces` offers an additional avenue for estimating effect sizes with linear models.

Because `mces`'s functionality is limited to models with continuous outcomes, this is an area that offers possible avenues for future development. Applications to additional types of models, such as generalized linear models, multilevel models, and longitudinal models, would extend researchers' abilities to report standardized effect sizes to further complement or replace null hypothesis significance testing in more contexts.

7 Acknowledgments

I thank Miguel Dorta, Daniel Klein, Chris Cheng, and the anonymous reviewers for helpful contributions to the code and to the manuscript.

8 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 22-1
. net install st0667     (to install program files, if available)
. net get st0667         (to install ancillary files, if available)
```

9 References

Aiken, L. S., and S. G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: SAGE.

Cheung, A. C. K., and R. E. Slavin. 2016. How methodological features affect effect sizes in education. *Educational Researcher* 45: 283–292. https://doi.org/10.3102/0013189X16656615.

Cleveland Clinic. 2018. High Hemoglobin Count. Cleveland Clinic Health Library, Disease and Conditions. https://my.clevelandclinic.org/health/diseases/17789-high-hemoglobin-count/.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

———. 1994. The earth is round ($p < .05$). *American Psychologist* 49: 997–1003. https://doi.org/10.1037/0003-066X.49.12.997.

Cohen, J., P. Cohen, S. G. West, and L. S. Aiken. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd ed. Mahwah, NJ: Lawrence Erlbaum.

Durlak, J. A. 2009. How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology* 34: 917–928. https://doi.org/10.1093/jpepsy/jsp004.

Ellis, P. D. 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press.

Glass, G. V., B. McGaw, and M. L. Smith. 1981. *Meta-Analysis in Social Research*. Beverly Hills, CA: SAGE.

Greenland, S., J. J. Schlesselman, and M. H. Criqui. 1986. The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology* 123: 203–208. https://doi.org/10.1093/oxfordjournals.aje.a114229.

Hedges, L. V. 1981. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6: 107–128. https://doi.org/10.3102/10769986006002107.

Huber, C. 2013. Measures of effect size in Stata 13. The Stata Blog: Not Elsewhere Classified. http://blog.stata.com/2013/09/05/measures-of-effect-size-in-stata-13/.

Kelley, K., and K. J. Preacher. 2012. On effect size. *Psychological Methods* 17: 137–152. https://doi.org/10.1037/a0028086.

Klein, D. 2016. Marginal effects in multiply imputed datasets. 14th German Stata Users Group meeting proceedings. http://www.stata.com/meeting/germany16/slides/de16_klein.pdf.

Kline, R. B. 2013. *Beyond Significance Testing: Statistics Reform in the Behavioral Sciences*. 2nd ed. Washington, DC: American Psychological Association.

Levine, T. R., and C. R. Hullett. 2002. Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research* 28: 612–625. https://doi.org/10.1111/j.1468-2958.2002.tb00828.x.

Linden, A. 2019. esizereg: Stata module for computing the effect size based on a linear regression coefficient. Statistical Software Components S458607, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s458607.html.

Long, J. S., and J. Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. 3rd ed. College Station, TX: Stata Press.

Lorah, J. 2018. Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large-scale Assessments in Education* 6(8). https://doi.org/10.1186/s40536-018-0061-2.

Mitchell, M. N. 2021. *Interpreting and Visualizing Regression Models Using Stata*. 2nd ed. College Station, TX: Stata Press.

Olejnik, S., and J. Algina. 2003. Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods* 8: 434–447. https://psycnet.apa.org/doi/10.1037/1082-989X.8.4.434.

Pedhazur, E. J. 1997. *Multiple Regression in Behavioral Research: Explanation and Prediction*. 3rd ed. San Diego: Harcourt.

Tabachnick, B. G., and L. S. Fidell. 2019. *Using Multivariate Statistics*. 7th ed. New York: Pearson.

Trafimow, D., and M. Marks. 2015. Editorial. *Basic and Applied Social Psychology* 37: 1–2. https://doi.org/10.1080/01973533.2015.1012991.

Vacha-Haase, T., and B. Thompson. 2004. How to estimate and interpret various effect sizes. *Journal of Counseling Psychology* 51: 473–481. https://psycnet.apa.org/doi/10.1037/0022-0167.51.4.473.

Wasserstein, R. L., and N. A. Lazar. 2016. The ASA statement on p-values: Context, process, and purpose. *American Statistician* 70: 129–133. https://doi.org/10.1080/00031305.2016.1154108.

Williams, R. A. 2003. pcorr2: Stata module to display partial and semipartial correlation coefficients. Statistical Software Components S436203, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s436203.html.

Ziliak, S. T., and D. N. McCloskey. 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor, MI: University of Michigan Press.

About the author

Brian Shaw is Assistant Professor of Music (Music Education) in the Jacobs School of Music at Indiana University–Bloomington.