Calculating Standard Errors for
Linear Regression in Randomized Evaluations

The Office of Evaluation Sciences often analyzes the results of RCTs by estimating a parametric statistical model — typically an ordinary least squares (OLS) regression — where one of the parameters represents the effect of a treatment or intervention. In order to infer whether this treatment effect is statistically significant, we typically depend on an estimate of the standard error associated with this parameter. This document describes OES's preferred method for calculating parametric standard errors for OLS regressions — in particular, the reasons for using HC2 standard errors — and how to calculate them in R and Stata.

When to Apply This Guidance

One reason for preferring HC2 standard errors is that they tend to represent variation resulting from random assignment of a set of units to different experimental conditions (for example, treatment and control). This guidance thus applies most directly to RCT's. HC2 standard errors may or may not be the most appropriate for other types of analyses, including quasi-experimental and other observational analyses (which we often conduct to establish background conditions during the design phase of a project).

Why Do We Prefer HC2 Standard Errors?

The usual standard errors calculated for OLS regression are not grounded in randomization. Since we tend to analyze randomized experiments with fixed samples, we prefer standard errors that are design-based — that reflect the variation we expect due to assigning units from a fixed sample to one or more experimental conditions.

The usual standard errors in OLS regression are also not consistent under heteroskedasticity. That is, when the variance in the outcome changes (with treatment and/or with covariates), larger samples do not lead the usual estimates of the standard error to collapse to a spike at the true value of the standard error. The "HC" in "HC2" indicates heteroskedasticity-consistent standard errors. HC standard errors are examples of "sandwich" standard errors.

Conceptually, both the usual OLS standard errors and the HC2 standard errors come from the same description of the covariate space (the "hat" or "projection" matrix), but the HC2 standard errors give less weight in their calculation to high-leverage points. That is, points farther from the center of the data are not allowed to unduly shrink the estimate of the standard error around the regression slope.

Generally speaking, the HC2 standard error strikes a balance between unbiasedness, power, heteroskedasticity-consistency, accounting for leverage, and validity.

How Does This Apply to Cluster Randomization?

For cluster-randomized designs (in which groups rather than individual units are assigned to experimental conditions), we prefer CR2 standard errors. The calculation of CR2 standard errors mirrors that of HC2 standard errors, but accounts for the design's clustering. In typical clustered designs with equal-sized clusters, even with few clusters, CR2 standard errors will perform well in terms of coverage, bias, and power.

Calculating HC2 and CR2 Standard Errors and Associated Confidence Intervals in R and Stata

Examples in R

Here is a brief example using `lm_robust()`, where HC2 is the default standard error estimator. If we have a dataset `df` with outcome `Y` and treatment indicator `Z`, we attach the `estimatr` package and estimate the linear model:

```
library(estimatr)
est <- lm_robust(Y ~ Z, data = df)
```

Printing the `est` object will display the HC2 standard errors and associate confidence intervals.

Here is a brief example using the `sandwich` and `lmtest` packages:

```
library(sandwich)
library(lmtest)
est <- lm(Y ~ Z, data = df)
estAndSE <- coeftest(est, vcov. = vcovHC(est, type = "HC2"))
estAndCI <- coefci(est, vcov. = vcovHC(est, type = "HC2"))
```

> The object `est` stores the linear model output; `estAndSE` includes the HC2 standard error and null hypothesis test results; `estAndCI` provides confidence intervals using the HC2 standard error.

For a clustered design, with cluster membership stored as a factor variable `clust_factor`, the `lm_robust()` function defaults to CR2 standard errors:

```
est_cluster <- lm_robust(Y ~ Z, data = df, clusters = clust_factor)
```

Brief Examples in Stata

In Stata, HC2 standard errors can be specified in the standard 'regress' command:

      regress Y Z, vce(hc2)

For Further Information

Below are some references on this approach to calculating standard errors when we analyze RCTs with OLS regression. Please also consult with the OES Methods Team for further information.

- This paper shows the value of using HC2 standard errors when analyzing experiments (and of including treatment-by-covariate interactions in linear models): Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. Annals of Applied Statistics, 7(1), 295-318. doi:10.1214/12-AOAS583

- This paper defines the "HC" standard error estimators, and demonstrates the superiority of alternative HC estimators to the original: MacKinnon, James G. and Halbert White (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3), September, 305-325. https://doi.org/10.1016/0304-4076(85)90158-7.

- For clustered designs, see demonstrations of the value and robustness of the CR2 standard error estimator, even when there are very few clusters, in "How misleading are clustered SEs in designs with few clusters?" at the DeclareDesign blog here.