# Regression-Discontinuity with Nonparametric Bootstrap

**Randall E. Schumacker**
University of North Texas

**Robert E. Mount**
Dallas Independent School District

The regression-discontinuity design (RD) is a powerful methodological alternative to the quasi-experimental design when conducting evaluations. RD designs involve testing post-test mean treatment differences between the experimental and comparison group regression lines at the centered cutoff point for statistical significance. This study simulated a RD treatment effect of 7 points in simulated normal and non-normal data distributions. The bootstrap technique was then used to estimate stability of estimates. Evaluation data oftentimes is non-normal, so understanding whether this impacts the RD design analysis is important. The bias (difference between the observed treatment effect and bootstrap estimate) and the confidence intervals are reported. Bootstrap estimates are useful in understanding whether the treatment effect is stable and the amount of estimation error present in RD given underlying normal and non-normal distributions. Results indicated that estimates of RD treatment effects are not severely impacted by non-normal, positive skewed, distributions. Consequently, robust estimation methods and/or data transformations such as probit are most likely sufficient to provide accurate stable estimates of treatment effects when concerned about meeting assumptions in regression analyses.

The regression-discontinuity (RD) design is different from a quasi-experimental design in that the assignment status is determined on the basis of a cutoff score on the pre-test measure and the pre-test measure can be different than the post-test measure (Cappelleri & Trochim, 2000). The statistical analysis of data in an RD design involves testing post-test mean differences between the group regression lines at the centered cutoff point for statistical significance, i.e., treatment effect. The RD design is therefore very useful when researching programs, procedures, or treatments given on the basis of need or merit. The basic regression-discontinuity equation can be expressed as: $Y_{Post} = b_0 + b_1 X_{\mathrm{Pre}} + b_2 Z + e$, where $\mathrm{Y}_{\mathrm{Post}}$ = post-test measure, Z = assignment status, $\mathrm{X}_{\mathrm{Pre}}$ = pre-test measure, e = residual error, and the b's are estimated sample regression weights. The regression weight for Z, the treatment effect variable, indicates the amount of gain or loss in the post-test assessment measure, i.e., a positive sign indicates gain, while a negative sign indicates loss. This study will explore RD with nonparametric bootstrap under normal and non-normal distributions. The purpose of this study will be to determine if the application of bootstrap is helpful in obtaining more stable estimates of treatment effect under non-normal data conditions, especially since RD is used in program evaluation where non-normal data is commonly encountered. The comparison of RD treatment effects using normal and non-normal data under simulated conditions will provide an understanding of how results may be affected.

Development of the regression-discontinuity design began in 1958 with a problem faced by Donald T. Campbell and his colleagues (Trochim, 1984). They were trying to assess differences between National Merit Scholarship Program participants and non-participants. In this situation, experimental design randomization was not possible, rather students above a cutoff point on the exam would receive a scholarship and those below the cutoff point would not. The regression discontinuity approach was developed because educational researchers quite often encountered this type of evaluation design. The researchers' dilemma was compounded by the fact that around the cutoff point, some students were awarded scholarships and some were not, depending on variables chosen for the analysis. Inconsistent assignment to groups based on the pre-test measure, around the cutoff point, was later named “fuzzy” group assignment. Specific techniques were developed for use in the situations where assignments of subjects with borderline pre-test scores were no longer made solely on the basis of the pretest score.

The regression-discontinuity design (RD) is a powerful methodological alternative to the quasi-experimental design when conducting evaluations. Thistlethwaite and Campbell (1960) proposed the RD approach to avoid problems inherent in ex-post facto designs that required matching of subjects. Bottenberg and Ward (1963) described the RD design as a special type of regression analysis involving two mutually exclusive groups that didn't necessarily require random assignment of subjects to an experimental and control group. The basic RD design requires a pre-test measure, a post-test measure, and an assignment status, i.e., received treatment or did not receive treatment. RD analyzes the treatment difference between the regression lines of the two groups at the cutoff score that was used to assign group membership (Trochim, Internet article). If the regression lines differ, then there is a *discontinuity*, the magnitude of which can be statistically interpreted. Regression-discontinuity designs do not randomly assign subjects nor control for extraneous variables. Unlike quasi-experimental designs, the regression-discontinuity design allows for some control over group assignment through use of a cut-off score.

Following its initial conceptualization, regression-discontinuity was not widely used until 1974 with the passage of Public Law 93-380, which attempted to standardize evaluation of the programs that various school districts were implementing to make use of Title I funds. Three evaluation models that school districts could use to evaluate various compensatory programs were created. RD was one of the models (Trochim 1984, McNeil 1984). Unfortunately, many school districts did not adopt the RD design, possibly because it was a more complex statistical analysis technique (Trochim & Davis, Internet article). In the decades that followed, the regression-discontinuity design has remained an underutilized technique in educational evaluation.

One of many problems educational researchers face when using regression-discontinuity is that data must meet all assumptions in regression analysis. Real-world data often has the potential to violate the assumptions of normally distributed data, equal variance, normally distributed residuals, and linear relationship between the predictor and outcome variables. Regression-discontinuity analysis in circumstances where such violations exist may produce biased estimates of the treatment effect. The bootstrap technique will be undertaken in order to assess the impact on RD of non-normal data. Not everyone is in agreement about using bootstrap techniques in least-squares regression to resolve problems related to these violations of assumptions. For example, Venables and Ripley (2001) view bootstrap as having little use in least-squares regression because if residual errors are close to being normally distributed, the standard theory applies. If not, robust regression estimation methods are available (p. 175) (Schumacker, Monahan, & Mount, 2002). A data transformation approach using a profile likelihood function is also suggested (p. 182). We have found that non-normal data can be transformed using probit to yield a more normal distribution prior to statistical analyses.

## Method and Procedures

Nonparametric bootstrap is a resampling procedure with replacement (Searle, Internet article). In brief, bootstrap involves using the sample data to construct a theoretical pseudo-population, composed of repeated random samplings from the original data set. Each additional sample can be equal to the number in the original sample. If the original sample included a full range of the values that exist in the actual population, then each additional sample can be thought of as representative of the true population distribution. When the desired statistic is computed for the original sample and for all additional bootstrap samples, a reasonable idea of the population distribution and the error distribution of the test statistic may be obtained. In this study, the non-parametric bootstrap technique in S-PLUS was utilized (S-PLUS, 2005).

*Data Sources*

The normal and non-normal data distributions for the study were simulated using two S-PLUS script programs written by the author and run in S-PLUS (S-PLUS, 2005). The normal distribution used the function, *rnorm*, and the non-normal distribution used the function, *rexp* (Johnson & Kotz, 1970).

The normal distribution true score was based on a sample size of 500, mean of 14, and standard deviation of 1. Random residual error was added to the true score based on a sample size of 500, mean of 0, and standard deviation of 1. The pre-test (X) mean value of 14 was used to determine group assignment (z), i.e., z = 1 if x <= 14, else z = 0. The z = 1 denoted the experimental group and Z = 0 the comparison group. A 7-point post-test gain was added to values for members in the group assigned z = 1, i.e., a 7-point treatment effect was introduced (post-test – pre-test = 7 point gain). These results were centered prior to RD regression analysis (XC = X – 14) (Cohen, Cohen, West & Aiken, 2003). The resulting display of RD data for the experimental and comparison group are in Figure 1.

The non-normal distribution true score was based on a sample size of 500 and a rate of 2 (inverse of mean to produce skew) that yielded a skewed distribution with mean of 14 and a positively skewed long right tailed distribution. Residual error was similarly added to the true score, but used a rate of 4. The pre-test (X) mean value of 14 was once again used to determine group assignment (z), i.e., z = 1 if x <= 14, else z = 0. A 7-point post-test gain was therefore added to values for members in the group assigned z = 1, i.e., a 7-point treatment effect was introduced (post-test – pre-test = 7 point gain). These results were also centered prior to RD regression analysis (XC = X – 14). The *Moment* formula for skewness and kurtosis should be used rather than the *Fisher* formula when resampling using bootstrap or jackknife procedures (S-PLUS, 2005). Figure 2 displays of RD data for the experimental and comparison groups.

The pre-test mean was 14, the post-test mean was 21, and therefore a known treatment effect of 7.0 was specified in both the normal and non-normal distributions. The non-normal distribution however was created to be positively skewed (Figure 2). For both types of distributions, the resulting treatment effect mean bootstrap estimate was compared to the pseudo population treatment effect mean with confidence intervals based on 500 bootstrap samples with a bootstrap sample size of 500. A comparison of the normal and non-normal treatment means as well as a comparison of each to the known treatment effect was conducted using an independent t-test at the .05 level of significance.

## Results

In this study, the experimental and comparison group regression lines were compared at a centered cutoff point for differences in treatment effect. The cutoff point was set to maximize the magnitude of the "discontinuity" between groups observed at the cutoff point. Schumacker (1992) has pointed out the importance of carefully considering the cutoff score in actual RD designs and discussed methods for locating the most useful cutoff score. Trochim (1984) had earlier suggested situations involving multiple comparison groups in which it might be helpful to use more than one cutoff score. Multiple comparison groups and cut-off scores however were not employed in this study. The comparative results for the normal versus non-normal distributions using simulated data are presented next.

The mean, standard deviation, median, and skewness values for the pre-test scores in the normal and non-normal data distributions are in Table 1. The normal distribution indicates the same mean and median values with skewness close to zero, as expected. The non-normal data indicates a median value that is lower than the mean and a skewness value indicating of a positively skewed distribution. Both distribution types were simulated to have the same pre-test mean. In the RD design, the pre-test scores are used to determine group assignment, i.e., experimental (treatment) versus comparison (non-treatment) groups. If the pre-test mean was equal to or less than 14, a person was assigned to the treatment group, else assigned to the non-treatment group. Results of this group assignment are in Table 2. In the normal distribution, we would expect the same number in each group (50/50); however, some sampling error is present and expected. In a non-normal distribution, we would expect a larger number in the treatment group (60/40) due to the positively skewed distribution (The median value is less than the mean value in a positively skewed distribution). Centering at the cut-off score was accomplished by taking the pretest score minus 14 (XC = X – 14). The expected mean for this cut-off value in both distributions is zero (normal distribution mean = 0.009; non-normal distribution mean = 0.001).

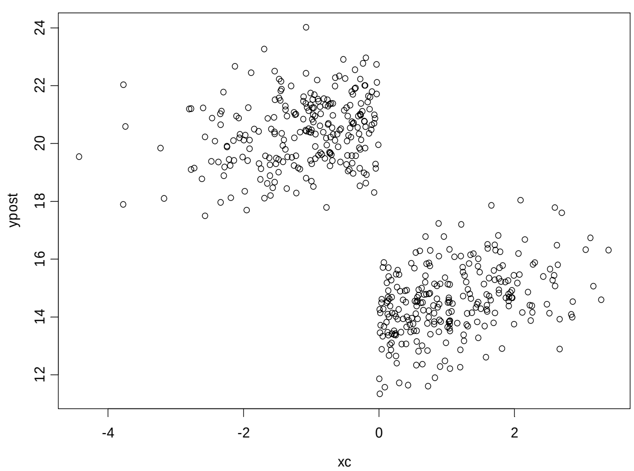


**Figure 1**. RD experimental and control group data at cut-off for normal distribution

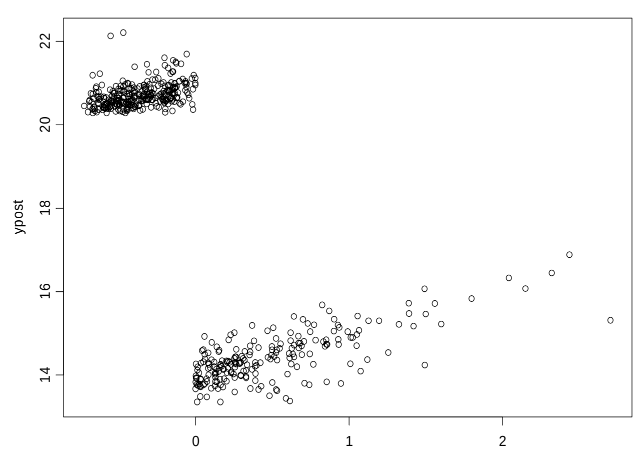


**Figure 2**. RD experimental and control group data at cut-off for non-normal distribution

**Table 1**. Descriptive Statistics

| Distribution | Mean | SD | Median | Skewness |
|---|---|---|---|---|
| Normal | 14 | 1.37 | 14 | -.08 |
| Non-Normal | 14 | 0.54 | 13.86 | 1.42 |

**Table 2**. Group Assignment using Pre-test score

| Distribution | Group | *N* | Percent |
|---|---|---|---|
| Normal | Non-treatment (0) | 256 | 0.51 |
| | Treatment (1) | 244 | 0.49 |
| Non-Normal | Non-treatment (0) | 211 | 0.42 |
| | Treatment (1) | 289 | 0.58 |

**Table 3**. RD Regression analyses (n=500)

| Distribution | Intercept | Slope | Treatment |
|---|---|---|---|
| Normal | 14.01 | .46 | 6.89 |
| Non-Normal | 14.00 | .86 | 7.02 |

**Table 4**. RD Bootstrap Estimates

| Distribution | Observed | Bootstrap | Bias | SE |
|---|---|---|---|---|
| Normal | 7.167 | 7.167 | 0 | .002 |
| Non-Normal | 7.294 | 7.294 | 0 | .001 |

**Table 5**. Independent t-test between distribution means and known treatment mean

| Distribution | Mean | Difference | *SE* | *t* |
|---|---|---|---|---|
| Normal | 6.89 | -.13 | .11 | -1.18 |
| Non-Normal | 7.02 | | | |
| Normal | 6.89 | -.11 | .17 | -.65 |
| Population | 7.00 | | | |
| Non-Normal | 7.02 | .02 | .05 | .40 |
| Population | 7.00 | | | |

* critical $t$ = 1.96, $p$<.05

The RD regression analyses for simulated data from both the normal and non-normal distributions are in Table 3. The RD regression equation is:

$$\hat{Y}_{Post} = b_0 + b_1 X_{\mathrm{Pre}} + b_2 Z$$

where a post test score (Y) is predicted using pretest score (X) and group assignment (Z), experimental versus comparison. The regression weights refer to $b_o$ = intercept, $b_1$ = slope, and $b_2$ = treatment effect (positive value is gain; negative value is loss). An intercept of 14 is expected for both types of distributions, however, the treatment effects ($b_2$) are expected to differ if skewness affects the RD design analysis. The treatment effects were similar (normal = 6.895; non-normal = 7.02), with any differences due to sampling error.

A non-parametric bootstrap was applied to the 500 simulated data set results for both the normal and non-normal distributions. Results are presented in Table 4. The observed mean treatment effect departed only slightly from the known specified treatment effect of 7.0. The difference in the normal distribution can be attributed to sampling error. The difference in the non-normal distribution can be contributed to sampling error and skewness. Bootstrap for both the normal and non-normal distributions reproduced the observed mean values, thus no difference in the expected outcome, i.e., bias = 0. Consequently, very little standard error was present in the bootstrap estimates. The 5% and 95% confidence intervals for the non-normal distribution results were: (7.292; 7.296) or 7.294 +/- .002. The 5% and 95% confidence intervals for the normal distribution were: (7.163; 7.171) or 7.167 +/- .004. The 5% and 95% confidence intervals around the bootstrap estimate contain 2 standard errors (SE).

An independent $t$-test was used to test whether the observed means were different from the known treatment mean and between themselves. These results are in Table 5. Results indicated that the treatment means were not statistically significantly different nor were each different from the known specified population treatment mean.

## Conclusions

Results indicated similar treatment effects whether normal or non-normal distributions were used with a centered cut-off score. Given a know treatment effect of 7.0 with random sampling error, bootstrap estimates were similar to observed estimates from pseudo-populations (bootstrap samples). No bias was reported and the findings indicated stable estimates in the presence of non-normality. The treatment effect estimated in RD using normal and non-normal simulated data distributions indicated that the RD design is not severely impacted by skewed data distributions commonly found in program evaluation. Robust estimation methods and/or data transformations such as probit are most likely sufficient to provide accurate stable estimates of treatment effects when concerned about meeting assumptions in regression analyses.

Regression discontinuity is appropriate for evaluation data and should be used more often in lieu of not meeting assumptions in quasi-experimental designs, especially in analysis of covariance. RD analyses can explore treatment effect differences at different cutoff points, use different pre-test measures than post-test measures, does not require matching of subjects, and can use multiple comparison groups with different cut-off scores. Educational researchers should therefore make increased use of the regression-discontinuity technique for program evaluation.

## References

Bottenberg, R.A. and Ward, J.H. (1963). *Applied Multiple Linear Regression*. Lackland Air Force Base, San Antonio, TX: Aerospace Medical Division, No. AD413128.

Cappelleri, J.C. and Trochim, W. (2000). Cutoff Designs. In Chow, Shein-Chung (Ed.) *Encyclopedia of Biopharmaceutical Statistics*, 149-156. Marcel Dekker: New York, NY.

Chambers, J. M., Mallows, C. L. and Stuck, B. W. (1976). A Method for Simulating Stable Random Variables. *Journal of the American Statistical Association*, *71*, 340-344.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Lawrence Erlbaum, Mahwah: NJ.

Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions, vol. 1 and 2*. Houghton-Mifflin, Boston.

McNeil, K. (April, 1984). *Random Thoughts on Why the Regression Discontinuity Design Is Not Widely Used*. Paper presented at the American Educational Research Association Annual Meeting. New Orleans, LA.

Searle, W. *Bootstrap Confidence Intervals*. Internet article located at: http://www.pitt.edu/~wpilib/statfaq/bootfaq.html

Schumacker, Randall E. (April, 1992). *Factors Affecting Regression-Discontinuity*. Paper presented at the American Educational Research Association Annual Meeting. San Francisco, CA.

Schumacker, R.E., Monahan, M.P., & Mount, R.E. (2002). A Comparison of OLS and Robust Regression using S-PLUS. *Multiple Linear Regresssion Viewpoints*, *28*(*2*), 10 - 13.

S-PLUS (2005). *S-PLUS 6 User’s Guide for Windows*. Insightful, Inc., Seattle, WA.

Thistlethwaite, D.L. & Cambell, D.T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Educational Psychology*, *51*(*6*), 309-317.

Trochim, William M. K. (1984). *Research Design for Program Evaluation, the Regression Discontinuity Approach*. Sage Publications: Beverly Hills, CA.

Trochim, William and Davis, Sarita. *Computer Simulations for Research Design*. Internet article located at: http://www.socialresearchmethods.net/simul/simul.htm

Trochim, W. *The Regression-Discontinuity Design*. Internet article located at: http://www.socialresearchmethods.net/kb/quasird.htm

Venables, W.N. & Ripley, B.D. (2001). *Modern Applied Statistics with S-PLUS* (3rd Edition). Springer Verlag: Statistics and Computing Series, New York: NY.

Send correspondence to: Randall E. Schumacker, Ph.D.
University of North Texas
Email: RSchumacker@unt.edu