Tutorials in Quantitative Methods for Psychology
2008, Vol. 4 (2), p. 46-50.

Using the Chow Test to Analyze Regression Discontinuities

Howard B. Lee
Department of Psychology
California State University, Northridge

The Chow Test (Chow, 1960) is a method well known in econometrics. It was originally designed to analyze the same variables obtained in two different data sets to determine if they were similar enough to be pooled together. Regression discontinuity design is a variation of the two-group pre-test-post-test design. The usual method of data analysis for data collected using this design is multiple regression with one dummy coded variable representing the cut-off value. This article discusses the use of the Chow Test on data obtained in a regression discontinuity study.

An infrequently used and perhaps underutilized quasi-experimental research design that is covered in a number of publications on research methodology such as Christensen (2006) and Kerlinger & Lee (2000) is the regression discontinuity design. The first article describing this design found in a PsycINFO search was by Thistlewaite & Campbell in 1960. In the 47 years that followed this seminal article, only 41 citations are listed in the PsycINFO search. This design starts with a selection criterion that separates entities (people) into two groups on the basis of some measurement such as intelligence. This design then tests whether some intervention or program change alters the relation between the selection criteria and the outcome measure. Thistlewaite & Campbell (1960) originally proposed this design as an alternate to the ex-post-facto design using a pretest and post-test with a control and experimental group. A major characteristic of the regression discontinuity design is how the participants are assigned to groups. A cutoff score is used for assignment purposes instead of random assignment. That is, all participants who score below a cutoff value are assigned to one group while those who have a score at or above the value are assigned to the other group. The two groups are usually designated as experimental or intervention group and the other as the control-group.

One of the primary strengths in using the regression discontinuity design is its robustness to threats to internal validity from time-related situations. These time-related situations would include history, maturation and regression to the mean. Unlike the regular pre-post test designs, the regression discontinuity design can accommodate polynomial and curvilinear relationships between the variables. Additionally, a sensitivity analysis can be conducted on the cutoff value.

Recent articles written by Zuckerman, Lee, et al (2006); Gormley, Gayer, & Phillips (2005); Braden and Bryant (1990) show the utility of the regression-discontinuity method in pharmaceutical health services research; studying the effects of preschool on cognitive development and effects of gifted students' program in school psychology research. Moss & Yeaton (2006) used the regression discontinuity design to explore the effectiveness of a developmental English program in a large, multicampus community college. Cook and Campbell (1979) and Shadish, Cook, & Campbell, (2001) demonstrated its possible usefulness in field research that does not have the constraints of laboratory research. Braden & Bryant (1990) used both real and fictitious data from the regression-discontinuity design that could be analyzed with either multiple regression or analysis of covariance with the use of dummy variables. Braden and Bryant (1990) provides an example of a two-dimensional application where the independent variable is group membership (i.e., gifted, not gifted), the dependent variable was achievement scores and the covariarte was IQ scores. Lesik (2006) describe how regression-discontinuity design can be used to make causal inferences when random assignment may not be practical. cut-off All of these studies used some form of multiple regression where the separation or was a dummy-coded variable in the equation.

The author proposes a method of analysis that is a variation of the analysis of covariance that does not use dummy variables directly in the computations of the test statistic. Plus this proposed method is general enough that it could be easily used in cases where there are more than one independent variable, more than one covariate and nonlinear (e.g. quadratic, curvilinear) predictor variables. This method was originally developed by G.C. Chow (1960) and later explored, expanded and enhanced by Fisher (1970), Toyoda (1974), Schmidt & Sickles (1977). It has appeared in textbooks such as Johnston (1972), Johnston & Dinardo (1996) and Kempthorne (1952). This method is well-known in economics and econometrics. The Chow Test as it is called, was created to statistically determine if two sets of observations could be regarded as belonging to the same regression model. This test allows one to test whether *m* additional observations are from the same regression as the first sample of *n* observations. Trochim (1984) briefly mentioned this method in analyzing data from regression discontinuity studies. Trochim gave a few references where the Chow Test was discussed in more detail. However these references were either unpublished dissertations or difficult to obtain reports. It is the intention of the author of this paper to show in greater detail where the Chow Test can be applied to regression discontinuity problems with distinct classifications.

The original paper by Chow (1960) considered observations (financial data) that were taken from two different time periods. Chow showed that his method could aid the researcher in determining whether or not newly collected data exhibit the same relationship between dependent and independent variables as the previous data. However, this method could be applied to situations where observations are from two different samples, such as those found in regression-discontinuity designs to determine if a change has occurred between the two samples collected over time.

According to Chow (1960), to test the equality between sets of coefficients in two linear regressions, one starts with the assumption that both are equal. A regression equation is fitted to the combined set of observations, i.e., excluded and selected, and the residual sum-of-squares is computed. Next, a regression equation is fitted to the data without assuming the sets are equal. Likewise, the residual sum-of-squares is obtained. Chow shows that the ratio of the difference between these two sums to the latter sum, adjusted for the corresponding degrees of freedom, will be distributed as an F-ratio under the null hypothesis. Chow presents two variations of his method.

Depending on which situation, a different F-ratio is computed. One situation occurs when one sample has more observations than regression parameters or weights estimated ($n > p$) but the second sample does not have enough observations ($m < p$) to compute a regression equation. Here, the sum of squares is computed for the sample of n observations where $n > p$ (number of regression weights). Another regression equation is computed using the combination of first and second sample. The Chow Test can be computed using the following steps:

Step 1: For the first n observations, fit the least squares equation:

$$Y_1 = X_1\beta_1 + e_1$$

Step 2: Compute the residual sum of squares, $e_1'e_1$

Step 3: pool the $n + m$ sample observations to give **Y** and **X** and fit the least squares regression:

$$Y = X\beta + e$$

Step 4: compute the residual sum of squares, $e'e$.

Step 5: The test of the null hypothesis that the m additional observations obey the same relation as the first is given by:

$$F = \frac{(e'e - e_1'e_1)/m}{e_1'e_1/(n-p)}$$

that is distributed as F with m and $n - p$ degrees of freedom.

The other situation is when both samples have enough observations to compute a regression equation. That is, the number of observations exceeds the number of regression parameters estimated. For this situation, the following steps of the Chow Test would be

Step 1: To the first n observations, fit the least squares equation:

$$Y_1 = X_1\beta_1 + e_1$$

Step 2: Compute the residual sum of squares, $e_1'e_1$.

Step 3: To the second m observations, fit the least squares equation:

$$Y_2 = X_2\beta_2 + e_2$$

Step 4: Compute the residual sum of squares, $e_2'e_2$.

Step 5: Pool the $n + m$ sample observations to give **Y** and **X** and fit the least squares regression:

$$Y = X\beta + e$$

Step 6: Compute the residual sum of squares, $e'e$.

Step 7: The test of the null hypothesis that the m additional observations obey the same relation as the first is given by:

$$F = \frac{(e'e - e_1'e_1 - e_2'e_2)/p}{(e_1'e_1 + e_2'e_2)/(n+m-2p)}$$

which is distributed as F with p and $n + m - 2p$ degrees of freedom.

Examples

Using the data from Braden & Bryant (1990), one regression equation would be for the selected group, another would be for the excluded group and a third equation is fitted to a combination of both groups. This can be done because the number of observations in both excluded and selected groups are greater than the number of regression weights to be estimated. Braden & Bryant (1990), used regression discontinuity to study children who were placed or not placed in gifted educational programs and the effect it had on achievement.

Without the actual data used by Braden & Bryant (1990), the author visually estimated each data point from the graphs presented in Braden & Bryant. There were two data sets. The first Braden & Bryant data set consisted of 60 data points. The second data set consisted of 90 data points. The second data set consisted of the data points from the first data set plus 30 fictitious data points to create a significant regression discontinuity. However, in the visual estimation process, only 88 data points were distinguishable and were used in illustrating the method in this paper. The results obtained in using the Chow Test were essentially the same. The null hypothesis could not be rejected at the α = .05 level for the first data set but was rejected for the second data set.

The analyses were done using SPSS for Windows with three separate executions of the regression subprogram. From the output, the residual sum of squares presented in each of the three ANOVA summary table were used in the computations.

Braden & Bryant Example 1

$n = 30$; $m = 30$, $p = 2$, $n + m = 60$

$$e_1'e_1 = 2575.563$$
$$e_2'e_2 = 1408.964$$
$$e'e = 4029.183$$

$$\begin{aligned} F &= \frac{(e'e - e_1'e_1 - e_2'e_2)/p}{(e_1'e_1 + e_2'e_2)/(n+m-2p)} \\ &= \frac{(4029.183 - 2575.563 - 1408.964)/2}{(4029.183 - 2575.563 - 1408.964)/(60-4)} \\ &= \frac{44.656/2}{3984.527/56} \\ &= \frac{22.328}{153.251} = 0.146 \end{aligned}$$

Critical F value (α = .05, df = 2, 56) = 3.17.

Since F = 0.146 < 3.17, the hypothesis of equality is not rejected. There is insufficient evidence that being in a gifted program led to different (higher) achievement than those who were not in a gifted program. Figure 1 shows the regression lines for the two groups.

Braden & Bryant Example 2

$n = 58$; $m = 30$, $p = 2$, $n + m = 88$

$$e_1'e_1 = 4937.961$$
$$e_2'e_2 = 1408.964$$
$$e'e = 6851.371$$

$$\begin{aligned} F &= \frac{(e'e - e_1'e_1 - e_2'e_2)/p}{(e_1'e_1 + e_2'e_2)/(n+m-2p)} \\ &= \frac{(6851.371 - 4937.961 - 1408.964)/2}{(4937.961 + 1408.964)/(88-4)} \\ &= \frac{504.446/2}{6346.925/78} = \frac{252.223}{75.559} = 3.338 \end{aligned}$$

Critical F value (α = .05, df = 2, 84) = 3.13.

Since F = 3.338 > 3.13, the hypothesis of equality is rejected. There is evidence that being in a gifted program led to different (higher) achievement than those who were not in a gifted program. Figure 2 shows a plot of the two regression lines.

Another Example.

Seaver & Quarton (1976) used regression discontinuity to

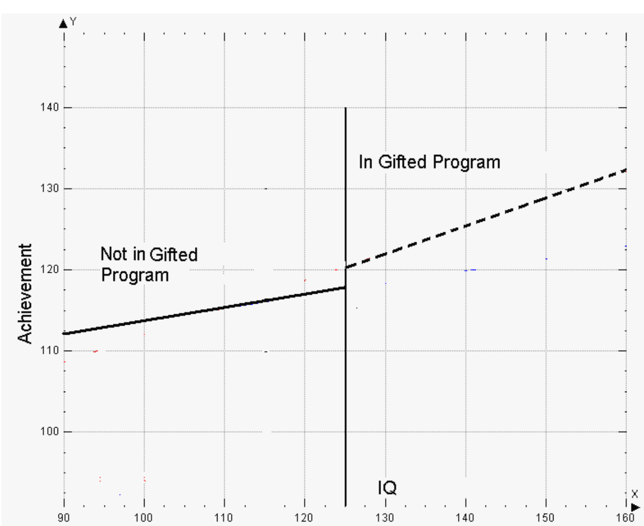

Figure 1. Regression lines for two groups using Braden & Bryant Data N = 60.

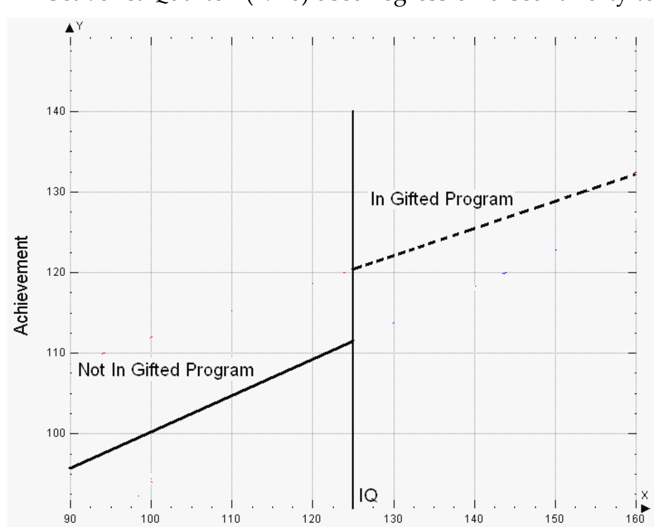

Figure 2. Regression lines for two groups using Braden & Bryant Data N = 88.

analyze the effects of a student being on the Dean's honor list. Two groups were identified and tracked over three academic terms. The first group consisted of students who achieve a grade-point average that allowed them to be on the Dean's list. The other group consisted of students who did not achieve grades to qualify them to be on the Dean's list. Seaver & Quarton (1976) found that membership in the Dean's list had a positive effect on grade point average in subsequent terms. They concluded that early membership on the Dean's list helps maintain the quality of academic work.

For purposes of this paper, Seaver & Quarton's (1976) study was partially replicated. Similar data were obtained on 1845 psychology undergraduate students at this author's university (California State University, Northridge) for the academic year of 2005-2006. The number of students that made the Dean's list was 632. The number of students excluded was 1213. To qualify for the Dean's List, a student had to carry 12 semester units or more and earns a grade-point average of 3.40 or higher.

Using SPSS regression analysis subprogram and formulas for the Chow Test, the results were similar to those found by Seaver & Quarton (1976). The numbers are given below.

n = 632; m = 1213, p = 2, $n + m$ = 1845

$$e_1'e_1 = 715.615$$

$$e_2'e_2 = 173.817$$

$$e'e = 901.796$$

$$F = \frac{(e'e - e_1'e_1 - e_2'e_2)/p}{(e_1'e_1 + e_2'e_2)/(n+m-2p)}$$

$$= \frac{(901.796 - 715.615 - 173.817)/2}{(715.615 + 173.817)/(1845 - 4)}$$

$$= \frac{12.364/2}{892.770/1841} = \frac{6.182}{.485} = 12.748$$

Critical F value (α = .01, df = 2, 1841) = 4.60.

Since F = 12.748 > 4.60, the hypothesis of equality is rejected. This tells us that being on the Dean's List yield different (positive) results than not being on the Dean's list. Figure 3 shows the two regression lines.

Discussion

This paper introduced the Chow test and its possibility in analyzing data from a regression discontinuity study. The formulas used in the Chow Test are not difficult to compute. Obtaining the values for the Chow Test is straightforward and can be obtained by using canned statistical packages such as SPSS. The Chow Test gives the researcher using this type of research design an alternative method of analysis from the traditional simple regression or analysis of covariance approaches. All possibilities of the Chow Test were not explored. Future research would need to be done to see how well it would do if the researcher decided to fit multiple regression models where there are more than two regression coefficients. Other interesting studies would be to see how well the Chow Test fares against other methods where a curvilinear or quadratic relationship exists between the independent and dependent variables.

The Chow Test was shown to be useful in the situations demonstrated in this paper. With the Chow Test, perhaps researchers will be more willing to develop a greater number of suitable studies where regression discontinuity would be used.

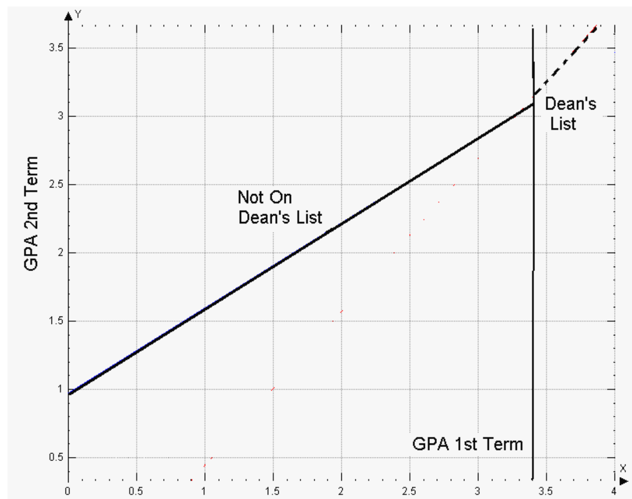

Figure 3. Regression lines for two groups using Dean's List Data N = 1845.

References

Braden, J. P. & Bryant, T. J. (1990). Regression discontinuity designs: Applications for school psychology. *School Psychology Review*. *19*(2), 232-239.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regression. *Econometrica*. *26*(3), 591-605.

Christensen, L. B. (2006). *Experimental methodology*, 10th Ed. Allyn & Bacon.

Cook, T.D. & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand McNally.

Fisher, F. M. (1970). Tests of equality between two sets of coefficients in two linear regression: An expository note. *Econometrica*. *38*(2), 361-366.

Gormley, W.T. Jr., Gayer, T. & Phillips, D. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, *41*(6), 872-884.

Johnston, J. (1972). *Econometric methods*. 2nd, Ed. New York: McGraw-Hill.

Johnston, J. & Dinardo, J. (1996). *Econometric methods*. New York: McGraw-Hill.

Kempthorne, O. (1952). *The design and analysis of* *experiments*. New York: Wiley.

Kerlinger, F. N. & Lee, H.B. (2000). *Foundations of behavioral research, 4th Ed*. Fort Worth, TX: Harcourt.

Lesik, S. A. (2006). Applying the regression-discontinuity design to infer causality with non-random assignment. *Review of Higher Education: Journal of the Association for the Study of Higher Education*, *30*(1), 1-19.

Moss, B. G. & Yeaton, W. H. (2006). Shaping policies related to developmental education: An evaluation using the regression-discontinuity design. *Educational Evaluation and Policy Analysis*, *28*(3), 215-229.

Schmidt, P. & Sickles, R. (1977). Some further evidence on the use of the Chow Test under heteroskedasticity. *Econometrica*. *45*(5), 1293-1298.

Seaver, W. B. & Quarton, R. J. (1976). Regression discontinuity analysis of dean's list effects. *Journal of Educational Psychology*, *68*(4), 459-465.

Shadish, W. P., Cook, T.D. & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin Company.

Thistlewaite, D. & Campbell, D.E. (1960). Regression discontinuity analysis: An alternative to the ex-post-factor experiment. *Journal of Educational Psychology*, *51*, 309-317.

Toyoda, T. (1974). Use of the Chow Test under heteroskedasticity. *Econometrica*. *42*, 601-608.

Trochim, W. M. K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Newbury Park, CA: Sage.

Zuckerman, I. H.; Lee, E., Wutoh, A. K. , Xue, Z. & Stuart, B. (2006). Application of regression-discontinuity analysis in pharmaceutical health services research. *Health Services Research*, *41*(2), 550-563.

Manuscript received March 10th, 2008
Revisions received March 19th, 2008