**APPENDIX D**

**REFERENCES**

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66*(6), 423–437.

Becker, W., & Kennedy, P. (1992). A lesson in least squares and R squared. *The American Statistician, 46*(4), 282–283.

Bem, S. L. (1977). *Bem Sex-Role Inventory Professional Manual*. California: Consulting Psychologists Press, Inc.

Bloom, H. (2009). *Modern Regression Discontinuity Analysis.* An MDRC publication.

 www.mdrc.org/publication/**modern**-**regression**-**discontinuity**-***analysis***

Burrill, G., & Hopensperger, P. (1993). *Exploring Statistics with the T1-81*. Reading, Massachusetts: Addison-Wesley.

Campbell, Donald T. (1969). Reforms as experiments. *American Psychologist*, 24, 409-

 429.

Card, David. (1995). Using Geographic Variation in College Proximity to Estimate the

 Return to Schooling. In *Aspects of Labour Economics: Essays in Honour of John*

 *Vanderkamp.* Edited by Louis Christofides, E. Kenneth Grant and Robert Swindinsky.

 University of Toronto Press.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48*(3), 378–398.

[Charig](http://en.wikipedia.org/w/index.php?title=C._R._Charig&action=edit), C.R., [D. R. Webb](http://en.wikipedia.org/w/index.php?title=D._R._Webb&action=edit), D.R., [Payne](http://en.wikipedia.org/w/index.php?title=S._R._Payne&action=edit), S. R., [Wickham](http://en.wikipedia.org/w/index.php?title=O._E._Wickham&action=edit), O. E. (March 1986). "[Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy](http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=3083922)". [*British Medical Journal (Clin Res Ed)*](http://en.wikipedia.org/w/index.php?title=Br_Med_J_%28Clin_Res_Ed%29&action=edit)*,* **292** (6524): 879–882

Cohen, J. (1994). The earth is round (*p* < .05). *American Psychologist, 49*(12), 997–1003.

(1990). Things I have learned (so far). *American Psychologist, 45*, 1304–1312.

(1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edition.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd Edition*. NJ: Lawrence Erlbaum Associates.

Cohen, R.J., & Swerdlik, M. (2005). *Psychological Testing and Assessment: An Introduction to Tests and Measurement, 6th Edition*. Boston, MA: McGraw Hill.

Cohen, R. J., Swerdlik, M., & Sturman, E. (2012). *Psychological Testing*

 *and Assessment: An Introduction to Tests and Measurement*.McGraw-Hill Education.

Cook, T. D., &Campbell, D. T. (1979). *Quasi-Experimentation: Design & Analysis*

 *Issues for Field Settings*. Houghton-Mifflin.

Darlington, R. B. (1990). *Regression and linear models.* New York: McGraw Hill.

de Moivre, A. (1756). *The Doctrine of Chances, 3rd edition*. London: Millar.

Elmore, J. et. al. (1998). Ten-year risk of false positive screening mammograms and clinical breast examinations. *The New England Journal of Medicine, 338(16)*, 1089–1096.

Efron, B. 1979. Bootstrap methods: Another look at the jackknife. Annals of Statistics 7: 1–26.

Efron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia:

 Society for Industrial and Applied Mathematics.

Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard. The amazing persistence of a probabilistic misconception. *Theory and Psychology, 5*(1), 75–98.

Fisher, R.A. (1934). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd. 319 pp.

Fisher, R. A. (1959). *Statistical methods and scientific inference* (2nd ed.). Edinburgh, Scotland: Oliver and Boyd.

Galton, F. (1886). "Regression towards mediocrity in hereditary stature". *The Journal of the Anthropological Institute of Great Britain and Ireland* **15**: 246–263.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*, published by Cambridge University Press.

Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance, *Review of Educational Research 42*(3), pp. 237-288.

Gough, P. (1996). How children learn to read and why they fail. *Annals of Dyslexia*,

 *46*(*1*), 1-20.

Groves, R.M., F. J. Fowler, Jr. F.J., Couper, M.P., Lepkowski, J.M., Singer, E., &

 Tourangeau, R. (2004). *Survey Methodology*. New Jersey: John Wiley & Sons.

Hagen, R. L. (1997). In praise of the null hypothesis statistical test. *American Psychologist, 52*, 15–24.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (eds.) (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates.

Hays, W. L. (1973). *Statistics for the social sciences*. 2nd ed. New York: Holt, Rinehart, and Winston.

­­­­­­­­­­Holland, P. W. & Rubin, D. B. (1982). *On Lord’s Paradox*. ETS RR-82-36, Princeton,

 NJ.

Horn, L., Hafner, & Owings (1992). A profile of American eighth-grade mathematics and science instruction. National Education Longitudinal Study of 1988. Statistical Analysis Report. ERIC Document (ERIC Document Reproduction Service No. ED337094).

Imbens, G.W. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to

 Practice. *Journal of Economics*, (142): 615-635.

[Julious](http://en.wikipedia.org/w/index.php?title=Steven_A._Julious&action=edit), S. A. and [Mullee](http://en.wikipedia.org/w/index.php?title=Mark_A._Mullee&action=edit), M.A. (1994). [Confounding and Simpson's paradox](http://bmj.bmjjournals.com/cgi/content/full/309/6967/1480).

 [*British Medical J*](http://en.wikipedia.org/wiki/BMJ)*ournal,* **309** (6967): 1480–1481.

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. San Francisco: Chandler Publishing Company.

Keppel, G. (1991). *Design and analysis: A researcher’s handbook*. Englewood, New Jersey: Prentice Hall.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and psychological measurement, 56*(5), 746–759.

Kivimaki, M., Ferrie, J.E., Brunner, E., Head, J., Shipley, M.J., Vahtera, J., & Marmot,

 M.G. (October 24, 2005). Justice at Work and Reduced Risk of Coronary Heart

 Disease Among Employees. *Archives of Internal Medicine*, *pp*. 2245-2251.

Kruskal, W.H. & Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association 47*(260), 583 – 621.

Lord, F.M. (1967) “A paradox in the interpretation of group comparisons”, *Psychological Bulletin*, 68, 304-305.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151–159.

Mann, H.B. & Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*, *pp*. 50-60.

Marascuilo, L.A. and Busk, P. L. (1987). Loglinear Models: A Way to Study Main Effects and Interactions for Multidimensional Contingency Tables with Categorical Data. *Journal of Counseling Psychology*,

Marascuilo, L. A. and Serlin, R. C. (1988). *Statistical Methods for the Behavioral Sciences.* W. H. Freeman.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective*. (2nd edition). Mahwah, N.J.: Lawrence Erlbaum Associates.

Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*, 103–115.

Mitchell, Joel. (2005). *Measurement in Psychology*. Cambridge: Cambridge University Press.

Nagourney, Eric. (November 1, 2005). Injustices at work may harm men’s hearts. Vital

 Signs Column, *The New York Times*.

Pearson, K. (1900). On the criterion that a given system of deviations from the probably in the case of a correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling. *Phil. Mag. (5)* 50, 157 – 175. Reprinted in K Pearson (1956), pp. 339 – 357.

Riley, R. (1998). The state of mathematics education: building a strong foundation for the 21st century. *Notices of the AMS, 45*(4), 487.

Rosenbaum, Paul R.; Rubin, Donald B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects". *[Biometrika](http://en.wikipedia.org/wiki/Biometrika%22%20%5Co%20%22Biometrika)* **70** (1): 41–55.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44*, 1276–1284.

Rozeboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin, 57*, 416–428.

Salsburg, David. (2001). The Lady Tasting Tea: How statistics revolutionized science in the twentieth century. A.W.H. Freeman/Owl Book Edition.

Satterthwaite, F. W. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 100–144.

Shadish, W. R., Cook, T.D. & Campbell, D. T. (2002). *Experimental and Quasi-*

 *Experimental Designs for Generalized Causal Inference*. Houghton Mifflin,

 Boston.

Shadish, W. R., Galindo, R., Wong, V.C., Steiner, P. M., & Cook, T. M. (2011). A

 Randomized Experiment Comparing Random and Cutoff-Based Assignment.

 *Psychological Methods*.

Siegel, Sidney, & Castellan, Jr., N. J. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Stanton, J. (2001). Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors. *Journal of Statistics Education,* vol. 9, no. 3.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 1946, *103*, 677-680.

Stevens, S. S. (1951). Mathematics, Measurement, and Psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: John Wiley & Sons.

Stigler, Stephen M (1997). ["Regression toward the mean, historically considered"](http://smm.sagepub.com/content/6/2/103.abstract).  *Statistical Methods in Medical Research* **6** (2): 103-114.

Stigler, S. M. (1986). *The History of Statistics: The measurement of uncertainty before 1900*. Cambridge: Harvard University Press.

Thompson, B. (1994). The concept of statistical significance testing. *ERIC/AE Digest*. ERIC Document (ERIC Document Reproduction Service No. ED366654).

(1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*, 26–30.

Tomasi, S., & Weinberg, S. L. (1999). Classifying children as learning disabled: An analysis of current practice in an urban setting. *Learning Disability Quarterly, 22*, 31–42.

Tukey, J. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Wainer, H. (2011). *Uneducated Guesses: Using evidence to uncover misguided education policies.* Princeton University Press.

Weinberg, S. L., Carroll, J. D. & Cohen, H. S. (1984). Confidence Regions for

 lNDSCAL Using The Jackknife and Bootstrap Techniques. *Psychometrika, 49,*

475-491.

Weinberg, S. L., & Goldberg, K. P. (1990). *Statistics for the behavioral sciences*. Cambridge: Cambridge University Press.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.

Willerman, L., Schultz, R., Rutledge, J. N., and Bigler, E. (1991). In vivo brain size and intelligence. *Intelligence 15*, 223–228.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*, *pp.* 80-83.

*The World Almanac and Book of Facts 2006.* (2006). New Jersey: World Almanac Books.

Yates, F. (1943). Contingency tests involving small numbers and the 2 test. *Journal of the Royal Statistical Society Supplement, 1*, 217 – 235.