Bibliography

Agresti, A. and Finlay, B. (2009). *Statistical Methods for the Social Sciences*. (4th ed.). New Jersey: Pearson Prentice Hall.

Baguley, T. (2009). Standardized or simple effect size: what should be reported? *British Journal of Psychology, 100*, 603–617.

Beal, S. L. (1989). Sample size determination for confidence intervals on the population mean and on the difference between two population means. *Biometrics, 45*, 969–977.

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Englewood Cliffs, NJ: Prentice Hall.

Bloom, H. (1995). Minimum detectable effects: a simple way to report the statistical power of experimental designs. *Evaluation Review, 19*, 547–56.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Bouten, C. V., van Marken Lichtenbelt, W. D., and Westerterp, K. R. (1996). Body mass index and daily physical activity in anorexia nervosa. *Medicine and Science in Sports and Exercise, 28*, 967–973.

Browne, R. H. (1995). On the use of a pilot sample for sample size determination. *Statistics in Medicine, 14*, 1933–1940.

Carlson, D., Borman, G. D., and Robinson, M. (2011). A multi-state district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis, 33*, 378–398.

Chou, Y., Arthur, K., Rosenstein, R., and Owen, D. (1984). New representations of the noncentral chi-square density and cumulative. *Communications in Statistics - Theory and Methods, 13*, 2673–2678.

Cochran, W. (1977). *Sampling Techniques*. (3rd ed.). New York: Wiley.

Cochran, W. (1983). *Planning and Analysis of Observational Studies*. New York: John Wiley and Sons.

Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. (1st ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Cumming, G. and Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532–574.

Dennis, C. L., Hodnett, E., Kenton, L., Weston, J., Zupancic, J., Stewart, D. E., and Kiss, A. (2009). Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. *British Medical Journal, 338*, 280–284.

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. (2nd ed.). Oxford: Oxford University Press.

Educational Testing Service (2007). *Graduate Record Examinations: Guide to the Use of Scores 2007–2008*. http://www.ets.org/Media/Tests/GRE/pdf/994994.pdf.

Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., and Steven, D. D. (1991). Making strategies and self-talk visible: writing instruction in regular and special education classrooms. *American Educational Research Journal, 28*, 337–372.

Finn, J. D. and Achilles, C. M. (1990). Answers and questions about class size: a statewide experiment. *American Educational Research Journal, 27*, 557–577.

Fuller, R. K., Mattson, M. E., Allen, J. P., Randall, C. L., Anton, R. F., and Babor, T. F. (1994). Multisite clinical trials in alcoholism treatment research: organizational, methodological and management issues. *Journal of Studies on Alcohol Suppl, 12,* 30–37.

Gatsonis, C. and Sampson, A. R. (1989). Multiple correlation: exact power and sample size calculations. *Psychological Bulletin, 106,* 516–524.

Graybill, F. (1958). Determining sample size for a specified width confidence interval. *The Annals of Mathematical Statistics, 29,* 282–287.

Grieve, A. P. (1991). Confidence intervals and sample sizes. *Biometrics, 47,* 1597–1603.

International Committee of Medical Journal Editors (1988). Uniform requirements for manuscripts submitted to biomedical journals. *Annals of Internal Medicine, 108,* 258–265.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66,* 373–388.

Hedges, L.V. and Olkin, I. (1985). *Statistical Methods for Meta-analysis.* Orlando: Academic Press.

Hedges, L. V. and Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6,* 203–217.

Hsu, L. (1994). Unbalanced designs to maximize statistical power in psychotherapy efficacy studies. *Psychotherapy Research, 4,* 95–106.

Jiroutek, M. R., Muller, K. E., Kupper, L. L., and Stewart, P. W. (2003). A new method for choosing sample size for confidence interval-based inferences. *Biometrics, 59,* 580–590.

Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions.* (Vol. 2, 2nd ed.). New York: Wiley.

Johnson, R. A. and Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis.* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Kaiser, H. F. and Michael, W. B. (1977). Little jiffy factor scores and domain validities. *Educational and Psychological Measurement, 37,* 363–365.

Kelley, K., Maxwell, S. E., and Rausch, J. R. (2003). Obtaining power or obtaining precision: delineating methods of sample-size planning. *Evaluation and the Health Professions, 26,* 258–287.

Kirk, R. E. (1995). *Experimental Design.* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Kraemer, H. C. and Thiemann, S. (1987). *How Many Subjects? Statistical Power Analysis in Research.* Newbury Park, CA: Sage.

Kupper, L. and Hafner, K. (1989). How appropriate are population sample size formulas? *The American Statistician, 43,* 101–105.

Land, K. C. (1973). Identification, parameter estimation, and hypothesis testing in recursive sociological models. In A. S. Goldberger and O. D. Duncan (eds.). *Structural Equation Models in the Social Sciences.* New York: Seminar.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* New York: Wiley.

Lehmann, E. L. (1993). The Fisher, Neyman–Pearson theories of testing hypotheses: one theory or two? *Journal of American Statistical Association, 88,* 1242–1249.

Lenth, R. (2001). Some practical guidelines for effective sample size determination. *American Statistician, 55,* 187–193.

Lipsey, M. W. (1992). Juvenile delinquency treatment: a meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, and F. Mosteller (eds.). *Meta-analysis for Explanation: A Casebook* (pp. 83–127). NY: Russell Sage Foundation.

Liu, X. (2003). Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *Journal of Educational and Behavioral Statistics, 28,* 231–248.

Liu, X. (2009). Sample size and the width of the confidence interval for mean difference. *British Journal of Mathematical and Statistical Psychology*, *62*, 201–215.

Liu, X. (2010). A note on non-centrality parameters for contrast tests in one-way ANOVA. *Journal of Experimental Education*, *78*, 53–59.

Liu, X. (2011). The effect of a covariate on standard error and confidence interval width. *Communications in Statistics: Theory and Methods*, *40*, 449–456.

Liu, X. (2012). Implications of statistical power for confidence intervals. *British Journal of Mathematical and Statistical Psychology, 65*, 427–437.

Liu, X. (2013a). A unified way to represent and compute the cumulative distribution function for non-central t, F, and chi square. *Communications in Statistics: Simulation and Computation, 42*, 1433–1436.

Liu, X. (2013b). A note on statistical power in multi-site randomized trials with multiple treatments at each site. *British Journal of Mathematical and Statistical Psychology*, DOI: 10.1111/bmsp.12016.

Liu, X. (2013c). Statistical power in three-arm cluster randomized trials, *Evaluation and the Health Professions*, DOI: 10.1177/0163278713498392.

Liu, X. (in press). Orthogonal decomposition of interaction effect in analysis of variance. *Journal of Statistics and Management Systems*.

Liu, X. and Raudenbush, S. (2004). A note on the non-centrality parameter and effect size estimates for the F test in ANOVA. *Journal of Educational and Behavioral Statistics, 29*, 251–255.

MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

McCambridge, J. and Strang, J. (2004). The efficacy of single-session motivational interviewing in reducing drug consumption and perceptions of drug-related risk and harm among young people: results from a multi-site cluster randomized trial. *Addiction, 99*, 39–52.

McGraw, K. O. and Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*, 361–365.

Pan, Z. and Kupper, L. L. (1999). Sample size determination for multiple comparison studies. *Statistics in Medicine, 18*, 1475–1488.

Parker, D. R., Evangelou, E., and Eaton, C. B. (2005). Intraclass correlation coefficients for cluster randomized trials in primary care: the cholesterol education and research trial (CEART). *Contemporary Clinical Trials, 26*, 260–267.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal, 3*, 1243–1246.

Pedhazur, E. (1997). *Multiple Regression in Behavior Research* (3rd ed.). Belmont, CA: Wadsworth Publishing.

Perrin, M. A., Chen, H., Sandberg, D. E., Malaspina, D., and Brown, A. S. (2007). Growth trajectory during early life and risk of adult schizophrenia. *British Journal of Psychiatry, 191*, 512–520.

Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Akerstedt, T., Charles, A., and Bioulac, B. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep, 28*, 1511–1516.

Porter, A. C. and Raudenbush, S. W. (1987). Analysis of covariance: its model and use in psychological research. *Journal of Counseling Psychology, 34*, 383–392.

Potthoff, R. F. and Roy, S. W. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika, 51*, 313–326.

Rao, C. R. (1951). An asymptotic expansion of the distribution of Wilks' criterion. *Bulletin de l'Institut International de Statistique, 33*, 177–180.

Raudenbush, S. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*, 173–185.

Raudenbush, S. and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W. and Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, 5, 199–213.

Raudenbush, S. W. and Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6, 387–401.

Rencher, A. (1995). *Methods of Multivariate Analysis*. New York: John Wiley & Sons.

Rencher, A. C. (1998). *Multivariate Statistical Inference and Applications*. New York: Wiley.

Rothman, R., Malone, R., Bryant, B., Horlen, C., and Pignone, M. (2003). Pharmacist-led, primary care-based disease management improves hemoglobin A1c in high-risk patients with diabetes. *American Journal of Medical Quality*, 18, 51–58.

Satorra, A. and Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, 50, 83–90.

Saxena, K. M. and Alam, K. (1982). Estimation of the non-centrality parameter of a chi squared distribution. *The Annals of Statistics*, 10, 1012–1016.

Searle, S. (1982). *Matrix Algebra Useful for Statistics*. New York: John Wiley & Sons.

Sidorov, J., Gabbay, R., Harris, R., Shull, R. D., Girolami, S., Tomcavage, J., Starkey, R., and Hughes, R. (2000). Disease management for diabetes mellitus: impact on hemoglobin A1c. *American Journal of Managed Care*, 6, 1217–1226.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Education and Behavioral Statistics*, 23, 323–355.

Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Methods for Studying Change and Event Occurrence*. New York: Oxford University Press.

Snijders, T. and Bosker, R. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational and Behavioral Statistics, 18*, 237–259.

Spearman, C. (1904) The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72–101.

Specht, D. A. (1975). On the evaluation of causal models. *Social Science Research, 4*, 113–133.

Stevenson, M. D., Scope, A., Sutcliffe, P. A., Booth, A., Slade, P., Parry, G., Saxon, D., and Kalthenthaler, E. (2010). Group cognitive-behavioural therapy for postnatal depression: a systematic review of clinical and cost effectiveness and value of information analyses. *Health Technology Assessment, 14*, 1–107.

Stricker, L. J. (2002). *TOEFL Research Report: The Performance of Native Speakers of English and ESL Speakers on the Computer-Based TOEFL and GRE General Test*. ETS: Princeton.

Tarasińska, J., (2005). Confidence intervals for the power of Student *t*-test. *Statistics and Probability Letters, 73*, 125–130.

Teerenstra, S., Moerbeek, M., van Achterberg, T., Pelzer, B. J., and Borm, G. F. (2008). Sample size calculations for 3-level cluster randomized trials. *Clinical Trials, 5*, 486–495.

Wilkinson, L. (1999). Statistical methods in psychology journals: guidelines and explanations. *American Psychologist, 54*, 494–604.

Wilks, S. (1932). On the sampling distribution of the multiple correlation coefficient. *Annals of Mathematical Statistics, 3*, 196–203.

Woolhouse, R. S. (1988). *The Empiricists*. Oxford: Oxford University Press.