



UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Statistical Models for the Precision of Categorical Measurement

van Wieringen, W.N.

Publication date
2003

[Link to publication](#)

Citation for published version (APA):

van Wieringen, W. N. (2003). *Statistical Models for the Precision of Categorical Measurement*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

References

- Agresti, A.A. (1988), "A Model for Agreement Between Ratings on an Ordinal Scale," *Biometrics*, 44, pp. 539-548.
- AIAG (2002), *Measurement System Analysis; Reference Manual*, 3rd edition, Detroit, MI: Automotive Industry Action Group.
- Allen, M.J. and Yen, W.M. (1979), *Introduction to Measurement Theory*, Monterey, California: Brooks / Cole.
- Bartholomew, D.J. and Knott, M. (1999), *Latent Variable Models and Factor Analysis*, London: Arnold.
- Bartko, J.J. and Carpenter, W.T. (1976), "On the Methods and Theory of Reliability," *Journal of Nervous and Mental Disease*, 163, pp. 307-317.
- Bickel, P.J. and Doksum, K.A. (2001), *Mathematical Statistics: Basic Ideas and Selected Topics, Vol. I*, New Jersey: Prentice-Hall.
- Bisgaard, S. (1999), "Proposals: A Mechanism for Achieving Better Experiments," in *Quality Engineering*, 11, 645-649.
- Blishcke, W.R. (1962), "Moment Estimators for the Parameters of a Mixture of two Binomial Distributions," *Annals of Mathematical Statistics*, 33, pp. 444-454.
- Blishcke, W.R. (1964), "Estimating the Parameters of Mixtures of Binomial Distributions," *Journal of the American Statistical Association*, 59, pp. 510-528.
- Box, G.E.P, Hunter, W.G. and Hunter, J.S. (1978), *Statistics for Experimenters*, New York: Wiley.
- Boyles, R.A. (2001), "Gauge Capability for Pass-Fail Inspection," *Technometrics*, 43, pp. 223-229.
- Brennan, R.L. and Prediger, D.J. (1981), "Coefficient Kappa: Some Uses, Misuses, and Alternatives," *Educational and Psychological Measurement*, 41, pp. 687-699.
- Chatfield, C. (1995), *Problem Solving: A Statistician's Guide* (2nd ed.), London: Chapman and Hall.
- Cicchetti, D.V. and Feinstein, A.R. (1990), "High Agreement but Low Kappa: II. Resolving the Paradoxes," *Journal of Clinical Epidemiology*, 43, pp. 551-558.
- Cohen, J. (1960), "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, 20, pp. 37-46.
- Cohen, J. (1968), "Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit," *Psychological Bulletin*, 70, pp. 213-220.

- Conger, A.J. (1980), "Integration and Generalization of Kappas for Multiple Raters," *Psychological Bulletin*, 88, pp. 322–328.
- Cox, D.R. and Snell, E.J. (1981), *Applied Statistics: Principles and Examples*, London: Chapman and Hall.
- Cramér, H. (1974), *Mathematical Methods of Statistics*, Almqvist and Wiksells, Uppsala, Sweden.
- De Mast, J. and Van Wieringen, W.N. (2003), "Measurement System Analysis for Bounded Ordinal Data," submitted for publication in *Technometrics*.
- Dunn, G. (1989), *Design and Analysis of Reliability Studies*, London: Edward Arnold.
- Elffers, H. (2001), "How High Should Cohen's Kappa Be When Dealing With Unbalanced Categories?," (In Dutch) *Kwantitatieve Methoden*, 66, pp. 33–41.
- Engel, J. and De Vries, B. (1997), "Evaluating a Well-Known Criterion for Measurement Precision," *Journal of Quality Technology*, 29, pp. 469–476.
- Everitt, B.S. (1968), "Moments of the Statistics Kappa and Weighted Kappa," *The British Journal of Mathematical and Statistical Psychology*, 21, pp. 97–103.
- Feinstein, A.R. and Cicchetti, D.V. (1990), "High Agreement but Low Kappa: I. The Problems of Two Paradoxes," *Journal of Clinical Epidemiology*, 43, pp. 543–549.
- Feldstein, F.L. and Davis, H.T. (1984), "Poisson Models for Assessing Rater Agreement in Discrete Response Studies," *British Journal of Mathematical and Statistical Psychology*, 37, pp. 49–61.
- Fleiss, J.L. (1965), "Estimating the Accuracy of Dichotomous Judgments," *Psychometrika*, 30, pp. 469–479.
- Fleiss, J.L. (1971), "Measuring Nominal Scale Agreement among Many Raters," *Psychological Bulletin*, 76, pp. 378–382.
- Futrell, D. (1995), "When Quality Is a Matter of Taste, Use Reliability Indexes," *Quality Progress*, May, pp. 81–86.
- Goodman, L.A. and Kruskal, W.H. (1954), "Measures of Association for Cross Classifications," *Journal of American Statistical Association*, 49, pp. 732–764.
- Goodman, L.A. (1974), "Exploratory Latent Structure Analysis using both Identifiable and Unidentifiable Models," *Biometrika*, 61, pp. 215–231.
- Hubert, L.H. (1977), "Kappa Revisited," *Psychological Bulletin*, 84, pp. 289–297.
- ISO (1993), *Guide to the Expression of Measurement Uncertainty*, 1st edition, Switzerland: International Organization for Standardization.
- Kendall, M. and Gibbons, J.D. (1990), *Rank Correlation Methods* (5th ed.), London: Edward Arnold.
- Kendall, M. and Stuart, A. (1977), *The Advanced Theory of Statistics* (Vol. 1, 4th ed.), London: Charles Griffin.
- Kerlinger, F.N. and Lee, H.B. (2000), *Foundations of Behavioral Research* (4th ed.), New York: Harcourt College Publisher.

- Krippendorff, K. (1970), "Bivariate Agreement Coefficients for Reliability of Data," in *Sociological Methodology*, ed. E.F. Borgatta, San Francisco: Jossey-Bass, pp. 139–150.
- Kruskal, W.H. (1954), "Ordinal Measures of Association," *Journal of the American Association*, 54, pp. 814–861.
- Landis, J.R. and Koch, G.G. (1975a), "A Review of the Statistical Methods in the Analysis of Data Arising from Observer Reliability Studies (Part I)," *Statistica Neerlandica*, 29, pp. 101–123.
- Landis, J.R. and Koch, G.G. (1975b), "A Review of the Statistical Methods in the Analysis of Data Arising from Observer Reliability Studies (Part II)," *Statistica Neerlandica*, 29, pp. 151–161.
- Landis, J.R. and Koch, G.G. (1977), "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, 33, pp. 159–174.
- Lehmann, E.L. (1983), *Theory of Point Estimation*, New York: John Wiley.
- Lord, F.M. and Novick, M.R. (1968), *Statistical Theories of Mental Test Scores*, London: Addison-Wesley.
- MacKay, R.J. and Oldford, R.W. (2000), "Scientific Method, Statistical Method, and the Speed of Light," *Statistical Science*, 15, pp. 254–278.
- McHugh, R.B. (1956), "Efficient Estimation and Local Identifiability in Latent Class Analysis," *Psychometrika*, 21, pp. 331–347.
- McLachlan, G.J. and T. Krishnan (1997), *The EM Algorithm and Extensions*, New York: Wiley.
- Montgomery, D.C. and Runger, G.C. (1993a), "Gauge Capability and Designed Experiments. Part I: Basic Methods," *Quality Engineering*, 6, pp. 115–135.
- Montgomery, D.C. and Runger, G.C. (1993b), "Gauge Capability and Designed Experiments. Part II: Experimental Design Methods and Variance Component Estimation," *Quality Engineering*, 6, pp. 289–305.
- Read, T.R.C. and Cressie, N.A.C. (1988), *Goodness-of-Fit Statistics for Discrete Multivariate Data*, Springer-Verlag, New York.
- Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.
- Shewhart, W.A. (1931), *Economic Control of Quality of Manufactured Product*, Princeton: Van Nostrand Reinhold.
- Shrout, P.E. and Fleiss, J.L. (1979), "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psychological Bulletin*, 86, pp. 420–428.
- Stein, P. (2002), "Statistical Issues in Measurement," *ASQ Statistics Division Newsletter*, Special Publication 2002, Milwaukee, WI: American Society for Quality.
- Stromberg, K.R. (1981), *An Introduction to Classical Real Analysis*, Wadsworth, California.
- Tanner, M.A. and Young, M.A. (1985), "Modeling Agreement among Raters," *Journal of the American Statistical Association*, 80, pp. 175–180.
- Uebersax, J.S. (1988), "Validity Inferences from Interobserver Agreement," *Psychological Bulletin*, 104, pp. 405–416.

- Uebersax, J.S. and Grove, W.M. (1993), "A Latent Trait Finite Mixture Model for the Analysis of Rating Agreement," *Biometrics*, 49, pp. 823-835.
- Van den Heuvel, E.R. (2000), "Gage R&R Studies for Nonstandard Situations," *ASQ's 54th Annual Quality Congress Progress*, Milwaukee, WI: American Society for Quality.
- Van den Heuvel, E.R. and Trip, A. (2003), "Evaluation of Measurement Systems with a Small Number of Observers," *Quality Engineering*, 15, pp. 323-331.
- Vanleeuwen, D.M. and Mandabach, K.H. (2002), "A Note on the Reliability of Ranked Items," *Sociological Methods & Research*, 31, pp. 87-105.
- Van Wieringen, W.N. and Van den Heuvel, E.R. (2003), "A Comparison of Methods for the Evaluation of Binary Measurement Systems," accepted for publication in *Quality Engineering*.
- Vardeman, S.B. and VanValkenburg, E.S. (1999), "Two-Way Random Effects Analyses and Gauge R&R Studies," *Technometrics*, 41, pp. 202-211.
- Wallsten, T.S. (1988), "Measurement Theory," *Encyclopedia of Statistical Sciences*, 8th ed., Kotz, S. and Johnson, N.L. (eds.), Vol. 5, Wiley, New York.
- Wheeler, D.J. and Lyday, R.W. (1989), *Evaluating the Measurement Process*, Knoxville, Tennessee: SPC Press.
- Wheeler, D.J. (1992), "Problems with Gauge R&R Studies," in *ASQC Quality Congress Transactions*, 46, pp. 179-185.
- Wheeler, D.J. (2003), "Good Data, Bad Data, and Process Behavior Charts," preprint, downloadable from: <http://www.asq.org>.