

Bibliography

- Agresti, A. (1988), “A model agreement between ratings on an ordinal scale.” *Biometrics*, 44, 539–548.
- Axelson, R.D. and C.D. Kreiter (2009), “Reliability.” In *Assessment in Health Professions Education* (Steven M. Downing and Rachel Yudkowsky, eds.), 57–74, Taylor and Francis.
- Bartko, J.J. (1976), “On various intraclass correlation reliability coefficients.” *Psychological Bulletin*, 83, 762–765.
- Bland, M.J. and D.G. Altman (1986), “Statistical methods for assessing agreement between two methods of clinical measurement.” *Lancet*, 1, 307–310.
- Bland, M.J. and D.G. Altman (1996), “Statistics notes: Measurement error.” *British Medical Journal*, 312, 1654.
- Bowman, F. Dubois, Ying Guo, and Gordana Derado (2007), “Statistical approaches to functional neuroimaging data.” *Neuroimaging Clinics of North America*, 17, 441–458.
- Carletta, J. (1996), “Assessing agreement on classification tasks: the kappa statistic.” *Computational Linguistics*, 22, 1–6.
- Casella, G. (1992), “Conditional inference from confidence sets.” In *Current Issues in Statistical Inference: Essays in Honor of D. Basu* (M. Ghosh and P.K. Pathak, eds.), volume 17 of *IMS Lecture Notes-Monograph Series*, 1–12, Institute of Mathematical Statistics.
- Chen, Gang, Paul A. Taylor, Simone P. Haller, Katharina Kircanski, Joel Stoddard, Daniel S. Pine, Ellen Leibenluft, Melissa A. Brotman, and Robert W. Cox (2017), “Intraclass correlation: Improved modeling approaches and applications for neuroimaging.” *Human Brain Mapping*, 39, 1187–1206.
- Cohen, J (1960), “A coefficient of agreement for nominal scales.” *Educational and Psychological Measurement*, 20, 37–46.

- Doros, G. and R. Lew (2010), “Design based on intr-class correlation coefficients.” *American Journal of Biostatistics*, 1, 1–8.
- Eckes, T. (2011), *Introduction to Many-Facet Rasch Measurement*. Peter Lang, Internationaler Verlag der Wissenschaften.
- Efron, B. (1979), “Bootstrap methods: another look at the jackknife.” *Annals of statistics*, 7, 1–26.
- Finn, R.H. (1970), “A note on estimating the reliability of categorical data.” *Educational and Psychological Measurement*, 30, 71–76.
- Fleiss, J. L. (1971), “Measuring nominal scale agreement among many raters.” *Psychological Bulletin*, 76, 378–382.
- Friedman, M. (1937), “The use of ranks to avoid the assumption of normality implicit in the analysis of variance.” *Journal of the American Statistical Association*, 32, 675–701.
- Giraudeau, B. and J.Y. Mary (2001), “Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient.” *Statistics in Medicine*, 20, 3205–3214.
- Gwet, K. L. (2008), “Computing inter-rater reliability and its variance in the presence of high agreement.” *British Journal of Mathematical and Statistical Psychology*, 61, 29–48.
- Gwet, Kilem L. (2020), *Beginner’s Guide to Principal Components: Applications with Excel*, 1 edition. AgreeStat Analytics, Maryland, USA.
- Gwet, Kilem L. (2021), *Handbook of Inter-Rater Reliability - Volume 1: Chance-corrected Agreement Coefficients for Categorical Ratings*, 5 edition, volume 1. AgreeStat Analytics, Maryland, USA.
- Henderson, C.R. (1953), “Estimation of variance and covariance components.” *Biometrics*, 9, 226–252.
- James, Lawrence R., Robert G. Demaree, and Gerrit Wolf (1984), “Estimating within-group interrater reliability with and without response bias.” *Journal of Applied Psychology*, 69, 85–98.
- Jolliffe, I.T. (2002), *Principal Component Analysis*. Springer, New York.
- Jung, H. W. (2003), “Evaluating interrater agreement in spice-based assessments.” *Computer Standards & Interfaces*, 25, 477–499.
-

-
- Kaiser, H.F (1960), "The application of electronic computers to factor analysis." *Educational and Psychological Measurement*, 20, 141–151.
- Kendall, M.G. (1938), "A new measure of rank correlation." *Biometrika*, 30, 81–93.
- Kendall, M.G. and B. Babington-Smith (1939), "The problem of m rankings." *Annals of Mathematical Statistics*, 10, 275–287.
- Kolmogorov, A.N. (1999), "The theory of probability." In *Mathematics - Its Contents, Methods and Meaning* (A.D. Aleksandrov, A.N. Kolmogorov, and M.A. Lavrent'ev, eds.), chapter Chapter XI, 229–264, Dover Publications - Dover Books on Mathematics.
- Koo, Terry K. and Mae Y. Li (2016), "A guideline of selecting and reporting intraclass correlation coefficients for reliability research." *Journal of Chiropractic Medicine*, 15, 155–163.
- Kottner, Jan and David L. Streiner (2011), "The difference between reliability and agreement." *Journal of Clinical Epidemiology*, 64, 701–702.
- Krippendorff, K. (2012), *Content Analysis: An Introduction to Its Methodology*, 3 edition. SAGE Publications, Inc., Thousand Oaks, CA.
- Krippendorff, Klaus (2011), "Agreement and information in the reliability of coding." *Communication Methods and Measures*, 5, 93–112.
- Kruskal, W.H. (1952), "A nonparametric test for the several sample problem." *Annals of Mathematical Statistics*, 23, 525–540.
- Kruskal, W.H. and W.A. Wallis (1952), "Use of ranks in one-criterion variance analysis." *Journal of the American Statistical Association*, 47, 583–621.
- Leone, M.A., P. Gaviani, and G. Ciccone (2006), "Inter-coder agreement for icd-9-cm coding of stroke." *Neurological Sciences*, 27, 445–448.
- Lin, L., A.S. Hedayat, and W. Wu (2012), *Statistical Tools for Measuring Agreement*. Springer, New York.
- Lin, L.I. (1989), "A concordance correlation coefficient to evaluate reproducibility." *Biometrics*, 45, 255–268.
- Lin, L.I. (1992), "Assay validation using the concordance correlation coefficient." *Biometrics*, 48, 599–604.
- Lin, L.I. (2000), "A note on the concordance correlation coefficient." *Biometrics*, 56, 324–325.
-

- Lin, L.I., A.S. Hedayat, B. Sinha, and M. Yang (2002), "Statistical methods in assessing agreement." *Journal of the American Statistical Association*, 97, 257–270.
- Lindeman, R.H., P.F. Meranda, and R.Z. Gold (1980), *Introduction to Bivariate and Multivariate Analysis*. Scott, Foresman and Company, Glenview, IL.
- Liu, Shiming, Murali Srinivasan, Roland Mörzinger, Marcel Lancelle, Thabo Beeler, Markus Gross, Barbara Solenthaler, Vincent Fehmer, and Irena Sailer (2019), "Reliability of a three-dimensional facial camera for dental and medical applications: A pilot study." *The Journal of Prosthetic Dentistry*, 122, 282 – 287.
- Mann, H. and D. Whitney (1947), "On a test of whether one of two random variables is stochastically larger than the other." *Annals of Mathematical Statistics*, 18, 50–60.
- Marascuilo, L.A. and L. McSweeney (1977), *Nonparametric and Distribution-free Methods for the Social Sciences*. Brooks/Cole Publishing Company, Monterey, CA.
- McCarthy, P. J. (1966), "Replication: An approach to the analysis of data from complex surveys." *National Center for Health Statistics, Washington, D.C., Series*, 2, 14.
- McGraw, K.O. and S.P. Wong (1996), "Forming inferences about some intraclass correlation coefficients." *Psychological Methods*, 1, 30–46.
- Mehta, S., R.F. Bastero-Caballero, Y. Sun, R. Zhu, D.K. Murphy, B. Hardas, and G. Koch (2018), "Performance of intraclass correlation coefficient (icc) as a reliability index under various distributions in scale reliability studies." *Statistics in Medicine*, 37, 2734–2752.
- O'Neill, Thomas A. (2017), "An overview of interrater agreement on likert scales for researchers and practitioners." *Frontiers in Psychology*, 8, 1–15, URL <https://doi.org/10.3389/fpsyg.2017.00777>.
- Pearson, K. (1896), "Mathematical contributions to the theory of evolution - iii. regression, heredity and panmixia." *Philosophical Transactions of the Royal Society of London, Series A*, 187, 253–318.
- Pearson, K. (1900), "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen in a random sampling." *Philosophical Magazine*, 5, 157–175.
- Quenouille, M. H. (1949), "Approximate tests of correlation in time series." *Journal of The Royal Statistical Society, Series B*, 11, 68–84.
-

-
- Quenouille, M. H. (1956), "Notes on bias in estimation." *Biometrika*, 61, 353–360.
- Satterthwaite, F.E. (1946), "An approximate distribution of estimates of variance components." *Biometrics*, 2, 110–114.
- Schuster, C. and A. von Eye (2001), "Models for ordinal agreement data." *Biometrical Journal*, 43, 795–808.
- Scott, W.A. (1955), "Reliability of content analysis: the case of nominal scale coding." *Public Opinion Quarterly*, XIX, 321–325.
- Searle, S.R. (1971), *Linear Models*. John Wiley & Sons.
- Searle, S.R. (1997), *Linear Models (Wiley Classics Library)*. Wiley-Interscience: John Wiley & Sons, Inc.
- Sheskin, D.J. (2004), *Handbook of Parametric and Nonparametric Statistical Procedures*, 3 edition. Chapman & Hall/CRC.
- Shou, H., A. Eloyan, S. Lee, V. Zipounnikov, A.N. Crainiceanu, M.B. Nebel, B. Caffo, M.A. Lindquist, and C.M. Crainiceanu (2013), "Quantifying the reliability of image replication studies: the image intra-class correlation coefficient (i2c2)." *Cognitive, Affective, & Behavioral Neuroscience*, 13, 714–724.
- Shrout, P.E. and J.L. Fleiss (1979), "Intraclass correlations: Uses in assessing rater reliability." *Psychological Bulletin*, 86, 420–428.
- Siegel, S. and N.J. Castellan Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*, 2 edition. McGraw-Hill Book Company, New York.
- Smith-Crowe, K., M. J. Burke, A. Cohen, and E. Doveh (2014), "Statistical significance criteria for the r_{WG} and average deviation interrater agreement indices." *Journal of Applied Psychology*, 99, 239–261.
- Spearman, C. (1904), "The proof and measurement of association between two things." *American Journal of Psychology*, 15, 72–101.
- Tanner, M.A. and M.A. Young (1985), "Modeling agreement among raters." *Journal of American Statistical Association*, 80, 175–180.
- Tinsley, Howard E. A. and David J. Weiss (1975), "Interrater reliability and agreement of subjective judgments." *Journal of Counselling Psychology*, 22, 358–376.
- Tinsley, Howard E. A. and David J. Weiss (2000), "Journal of counselling psychology." In *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (S. D. Tinsley, H. E. A. & Brown, ed.), 94–124, Academic Press, New York.
-

- Tukey, J. W. (1958), "Bias and confidence in not quite large samples (abstract)." *Annals of Mathematical Statistics*, 29, 614.
- Wallis, W.A. (1939), "The correlation ratio for ranked data." *Journal of the American Statistical Association*, 34, 533–538.
- Wilcoxon, F. (1949), "Some rapid approximate statistical procedures." Technical report, American Cyanamid Corporation, Stamford, CT: Stamford Research Laboratories.
- Wu, J., Y. Zhang, Y. Gao, and T. Luo (2019), "Assessment the reliability of ultrasonography in the imaging of the plantar fascia: a comparative study." *BMC Medical Imaging*, 19, Published online 2019 Aug 7. doi: 10.1186/s12880-019-0361-1.
- Yue, Chen, Shaojie Chen, Haris I. Sair, Raag Airan, and Brian S. Caffo (2015), "Estimating a graphical intra-class correlation coefficient (gicc) using multivariate probit-linear mixed models." *Computational Statistics & Data Analysis*, 89, 126–133.
- Zhao, X., J.S. Liu, and K. Deng (2013), "Assumptions behind intercoder reliability indices." In *Communication Yearbook* (C.T. Salmon, ed.), 419–480, Routledge.
-

List of Notations

- a , 135, 141, 166, 167, 172–174, 193, 195
 A , 141
 A_1 , 248
 A_2 , 248
 A^* , 174
 α , 84, 88
- b , 135, 141, 166, 167, 172–174, 193, 195
 β , 150
- c , 135, 141
 χ_r^2 , 258
 C_L , 84, 88
 C_U , 84, 88
- δ_r , 127
 δ_s , 127
- e_{ijk} , 71, 77, 122, 164, 180
- F , 86, 194, 199
 f , 136
 F_0 , 86
 F_1 , 134, 167, 192, 193, 195
 f_1 , 172–174
 F_2 , 134, 167, 192, 193, 195
 f_2 , 172–174
 F_L , 88, 194
 F_{obs} , 88, 142, 167, 198
 $F_{observed}$, 138
FSE, 174
FSI, 173
FSR, 172–174
- FSS, 172–174, 223
 F^* , 172, 173
 F_U , 88, 194
- γ , 78, 91, 132, 140, 141, 184
 γ_0 , 91, 141, 167, 191
- h_6 , 190, 220
- I2C2, 277
ICC, 71, 120
ICC(1A, 1), 73
ICC(1B, 1), 79, 80
ICC(2, 1), 126, 130, 131, 165
ICC(3, 1), 185, 186, 188, 190
ICC_a(2, 1), 132, 133, 165
ICC_a(3, 1), 186, 189
- k_1 , 80, 127
 k'_1 , 127
 k'_2 , 127, 165
 k_3 , 127, 165
 k_4 , 127, 165, 189
 k'_5 , 127
KCC, 254
 k_0 , 73
- λ_1 , 165
 λ_2 , 165
 λ_0 , 127, 189
LC3, 259
LCB, 83, 88, 134, 141, 166, 172–174,
192, 194, 195, 223
- M , 72, 127, 189

- MDD, 92, 202
 $m_{i.}$, 72, 80, 127, 192
 m_{ij} , 72, 127, 189
 $m_{.j}$, 72, 80, 127, 189
MSE, 74, 79, 83, 88, 130, 135, 185–187, 193, 194
MSI, 130, 135, 185, 186, 192
MSR, 79, 88, 89, 130, 135
MSS, 74, 83, 130, 134, 185–187, 192, 194
MSV, 232
MSV, 234
 μ , 71, 77, 122, 164, 180
- PCA, 17
- r , 244, 245
 \bar{R} , 255
 R_i , 255
 ρ , 71, 125, 181, 190
 ρ_0 , 86, 167, 191, 197
 R_{ij} , 255
 r_j , 77, 122, 164, 180
 r_s , 244, 247
RSS, 190, 220
 r_s^* , 248
- S, 255
 S^* , 255
 s_i , 122, 164, 180
 σ_e^2 , 71, 77, 123
 $\hat{\sigma}_e^2$, 73, 80, 126, 133, 165, 189, 190
 $\hat{\sigma}_r^2$, 80, 126, 133, 165
 $\hat{\sigma}_{rs}^2$, 126
 $\hat{\sigma}_s^2$, 73, 126, 133, 165, 189, 190
 $\hat{\sigma}_{sr}^2$, 133, 189
- σ_r^2 , 77, 123
 σ_s^2 , 71, 73, 123, 180
 σ_{sr}^2 , 123, 181
 $(sr)_{ij}$, 122, 180
- t , 246
 $T_{2\mu}$, 165
 T_{2r} , 80, 127, 165, 190
 T_{2s} , 72, 127, 165
 T_{2sr} , 127, 190
 T_{2y} , 72, 80, 127, 165, 190
 τ , 249
 τ^* , 252
T, 256
 T_1 , 249
 T_2 , 249
 T_y , 72, 80, 127
 T_y' , 127
- UCB, 83, 88, 134, 141, 166, 172–174, 192, 194, 195, 223
- v , 135, 141, 166, 167, 172–174, 193, 195
- W, 256, 258
WDF, 35
- $\bar{y}_{i..}$, 192
 $\bar{y}_{ij.}$, 193
 $\bar{y}_{.j.}$, 79
 $y_{i..}$, 127
 $y_{i..}^2$, 72
 $y_{ij.}$, 190
 y_{ijk} , 71, 77, 122, 164, 180
 $y_{.j.}$, 127, 190
- z , 253
-

Author Index

- Agresti, A., 9
Airan, R., 51, 59, 277
Altman, D.G., 54–56, 124, 128, 168, 271
Axelson, R.D., 19

Babington-Smith, B., 257
Bartko, J.J., 182
Bland, M.J., 54–56, 124, 128, 168, 271

Caffo, B., 51, 59, 277, 278
Caffo, B.S., 51, 59, 277
Carletta, J., 7
Casella, G., 236
Castellan Jr., N.J., 258
Chen, S., 51, 59, 277
Ciccone, G., 8
Cohen, J., 233
Crainiceanu, A.N., 51, 59, 277, 278
Crainiceanu, C.M., 51, 59, 277, 278

Demaree, R.G., 54, 229
Deng, K., 7
Doros, G., 99, 144, 208

Eckes, T., 7–9
Eloyan, A., 51, 59, 277, 278

Finn, R.H., 51, 54, 229, 232, 235
Fleiss, J.L., 52, 54, 72–74, 78, 126, 132, 181, 185, 233
Friedman, M., 258
Gaviani, P., 8

Giraudeau, B., 99, 144, 208
Gold, R.Z., 249
Gwet, K.L., 5, 27, 233, 246, 281, 282

Hedayat, A.S., 259
Henderson, C.R., 166, 176

James, L.R., 54, 229
Jung, H.W., 8

Kaiser, H.F., 291
Kendall, M.G., 249, 257
Kolmogorov, A.N., 9
Koo, T.K., 268, 269
Kottner, J., 5
Kreiter, C.D., 19
Krippendorff, K., 5, 7
Kruskal, W.H., 184

Lee, S., 51, 59, 277, 278
Leone, M.A., 8
Lew, R., 99, 144, 208
Li, M.Y., 268, 269
Lin, L.I., 67, 259, 261
Lindeman, R.H., 249
Lindquist, M.A., 51, 59, 277, 278
Liu, J.S., 7

Marascuilo, L.A., 258
Mary, J.Y., 99, 144, 208
McGraw, K.O., 54, 72, 74, 78, 126, 181, 182, 185
McSweeney, L., 258
Meranda, P.F., 249

- Nebel, M.B., 51, 59
Nebel, N.B., 277, 278
- O'Neill, T.A., 54
- Pearson, K., 245
- Sair, H.I., 51, 59, 277
Satterthwaite, F.E., 137, 194
Schuster C., 9
Scott, W.A., 7
Searle, S.R., 72, 79, 126, 132, 176, 188, 220
Sheskin, D.J., 249
Shou, H., 51, 59, 277, 278
Shrout, P.E., 52, 54, 72–74, 78, 126, 132, 181, 185
Siegel, S., 258
Sinha, B., 259
Spearman, C., 247
- Streiner, D.L., 5
- Tanner, M.A., 9
Tinsley, H.E.A., 4
- von Eye, A., 9
- Wallis, W.A., 184, 257
Weiss, D.J., 4
Wilcoxon, F., 184
Wolf, G.J., 54
Wong, S.P., 54, 72, 74, 78, 126, 181, 182, 185
Wu, W., 259
- Yang, M., 259
Young, M.A., 9
Yue, C., 51, 59, 277
- Zhao, X., 7
Zipounnikov, V., 51, 59, 277, 278
-

Subject Index

- Benchmark Scale, 268
- Benchmarking, 268
- Bland-Altman Plot, 54, 128, 168
- Bootstrap, 236, 237
- Bootstrapping, 237

- Chance Agreement, 12
- Coefficient of Variation, 144
- Confidence
 - Interval, 82
 - Level, 83, 134
- Consistency
 - Measures of, 182
- Correlation
 - Coefficient, 71
 - Pearson's Product-Moment, 244
 - Spearman, 247
 - Spearman's Rank-Order, 244
- Covariance Matrix, 61

- Error Effect, 230
- Estimand, 27
- Estimator, 27

- F-Distribution, 193
- Factorial Design, 53, 120
 - Mixed, 53, 121
 - Random, 53, 121
- Finn's Coefficient, 228

- Gao, Y., 276

- Henderson
 - Method 1, 74
 - Method I, 126, 166

- Inference, 26, 82, 191
 - Statistical, 26, 82
- Influence Analysis, 268
- Inter-Annotator Reliability, 7
- Inter-Rater Reliability
 - Applications, 6
 - Calculation, 185
 - Definition, 4, 181
 - Parameter, 27
 - Sample Size Calculation, 144, 149, 152, 202, 207, 208
- Interaction Effect, 56, 120, 121
- Intercoder Reliability, 7
- Intra-Rater Reliability, 52
 - Calculation, 78, 132, 185
 - Confidence Interval, 140, 166, 194
 - Definition, 75, 77, 131, 184
 - p-value, 141, 199
 - Sample Size Calculation, 160, 205, 214
 - Statistical Inference, 140
- Intraclass Correlation, 50
 - Definition, 71
 - Inference, 191
 - Mixed Factorial Design, 180
 - One-Factor Studies, 69
 - Random Factorial Design, 122

- Jackknife, 289
- Jolliffe's Rule, 285

- Kappa, 118, 177
 - Kendall
 - Coefficient of Concordance, 254
 - Tau, 249
 - Kruskal-Wallis Test, 184
 - LCB, 83, 88
 - Limits of Agreement, 54
 - Lin's concordance correlation coefficient, xii, 242, 258
 - linear correlation coefficient, 245
 - Loadings, 60
 - Luo, T., 276
 - Mann-Whitney Test, 184
 - Mean of Squares for Errors, 79
 - Mean of Squares for Raters, 79
 - Mean Subject Variance, 232
 - Minimum Detectable Difference, 92, 202, 263
 - Mixed Factorial Design, 53, 179, 180
 - Intra-Rater Reliability, 184
 - Model 1A, 71
 - Model 1B, 76
 - Model 2, 124
 - Model 3, 181
 - p-value, 83, 86, 91, 134, 138, 141, 167, 191, 197, 199, 237
 - PCA, 280
 - Pearson correlation coefficient, 244
 - Post Hoc, 145
 - Power Analysis, 151
 - Post-Hoc, 151
 - Principal
 - Component, 59, 280
 - Component Analysis, 17, 280
 - Component Scores, 281
 - Components, 280
 - Q-Q Plot, 237
 - Rater Effect, 230
 - Rater Sampling, 18
 - Reference Set, 236
 - Reliability, 4
 - Inter-Annotator, 7
 - Inter-Rater, 4
 - Intercoder, 7
 - Test-Retest, 76
 - Repeatability Plot, 130, 169
 - Replicate Sample, 289
 - Reproducibility, 185
 - Research Hypothesis, 92, 143
 - Sample Size
 - Mixed Factorial Model, 201
 - Model 1A/Confidence Interval Approach, 99
 - Model 1A/Statistical Power Approach, 93
 - Model 1B, 108
 - Random Factorial Model, 143
 - Sampling
 - Plan, 19
 - Rater Population, 18
 - Subject Population, 18
 - Scoring Rubric, 21
 - Significance Level, 141, 166
 - Spearman Correlation
 - Definition, 247
 - Tie-Adjusted, 248
 - Spearman's correlation, 244
 - Standard Error, 17
 - Statistical Power, 109
 - Statistically Significant, 236
 - Subject Effect, 230
 - Subject Sample, 18
 - Subject Sampling, 18
 - Test
 - Kruskal-Wallis, 184
 - Mann-Whitney, 184
 - Wilcoxon Rank Sum, 184
 - Test of Hypothesis, 82
-

Test-Retest Reliability, [76](#)

Two-Way Mixed Factorial Design, [53](#)

Type II Error, [150](#)

UCB, [83](#), [88](#)

Variable, [21](#)

Variance Matrix, [61](#)

Variance-Covariance Matrix, [61](#)

Vector, [60](#)

Wilcoxon Rank Sum Test, [184](#)

Within-Group Inter-Rater Reliability,
[229](#)

Wolf, G., [229](#)

Wu, J., [276](#)

Zhang, Y., [276](#)

Handbook of Inter-Rater Reliability, 5th Ed. Vol 2: Analysis of Quantitative Ratings

Inter-rater reliability assessment has become an essential component in the process of evaluating the quality of experimental data in almost all fields of research. The 4th edition of the *Handbook of Inter-Rater Reliability* covered the analysis of categorical and quantitative ratings in a single volume. In response to comments on previous editions, the current 5th edition is released in 2 volumes. Volume 1 is devoted to the analysis of categorical ratings, whereas the current volume 2 focuses on the analysis of quantitative ratings.

Here is the link to the webpage of volume 2, where you can find, a link to the errata page, some example workbooks, and other datasets:

www.agreestat.com/books/icc5/

Here are a few topics that are new to the 5th edition:

- Chapter 2 describes various methods for setting up your dataset of ratings before analysis.
- Chapter 7 covers Finn's coefficient, which is recommended when the classical intraclass correlation coefficient is inapplicable.
- Lin's concordance correlation coefficient is covered in an updated chapter 8.
- A benchmarking method for qualifying the strength of agreement measured by the intraclass correlation coefficient is described in chapter 9. This chapter also covers the multivariate intraclass correlation used in the multivariate analysis of correlated variables.

About the Author

Kilem L. Gwet, Ph.D.

Statistical consultant, mathematical statistician, researcher, and instructor. Over 20 years of experience in various industries, and several publications in peer-reviewed journals.

AgreeStat Analytics

PO Box 2696

Gaithersburg, Maryland 20886-2696 – USA

