

Last revised: July 2006

## Statistical Inference Bibliography 1920-Present

1. Pearson, K. (1920) "The Fundamental Problem in Practical Statistics." *Biometrika*, 13(1): 1-16.
2. Edgeworth, F.Y. (1921) "Molecular Statistics." *Journal of the Royal Statistical Society*, 84(1): 71-89.
3. Fisher, R. A. (1922) "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character*, 222: 309-268.
4. Neyman, J. and E. S. Pearson. (1928) "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I." *Biometrika*, 20A(1/2): 175-240.
5. Fisher, R. A. (1933) "The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters." *Proceedings of the Royal Society of London, Series A, Containing Papers of Mathematical and Physical Character*, 139(838): 343-348.
6. Fisher, R. A. (1935) "The Logic of Inductive Inference." *Journal of the Royal Statistical Society*, 98(1): 39-82.
7. Fisher, R. A. (1936) "Uncertain inference." *Proceedings of the American Academy of Arts and Sciences*, 71: 245-258.
8. Berkson, J. (1942) "Tests of Significance Considered as Evidence." *Journal of the American Statistical Association*, 37(219): 325-335.
9. Barnard, G. A. (1949) "Statistical Inference." *Journal of the Royal Statistical Society, Series B (Methodological)*, 11(2): 115-149.
10. Fisher, R. (1955) "Statistical Methods and Scientific Induction." *Journal of the Royal Statistical Society, Series B (Methodological)*, 17(1): 69-78.
11. Pearson, E. S. (1955) "Statistical Concepts in their Relation to Reality." *Journal of the Royal Statistical Society, Series B (Methodological)*, 17(2): 204-207.
12. Yates, F. (1955) "Discussion on the Paper by Dr. Box and Dr. Anderson." *Statistical Inference, Robustness, and Modeling Strategy, JRSS-B*, 17(1): 31.
13. Cox, D. R. (1958) "Some Problems Connected with Statistical Inference." *Annals of Mathematical Statistics*, 29(2): 357-372.
14. Good, I. J. (1958) "Significance Tests in Parallel and In Series." *Journal of the American Statistical Association*, 53: 799-813.

15. Natrella, M. G. (1960) "The Relation Between Confidence Intervals and Tests of Significance." *The American Statistician*, 14: 20-22 & back cover.
16. Rozeboom, W. W. (1960) "The Fallacy of the Null-Hypothesis Significance Test." *Psychological Bulletin*, 57(5): 416-428.
17. Neyman, J. (1961) "Silver Jubilee of My Dispute with Fisher." *Journal of the Operations Research Society of Japan*, 3(4): 145-154.
18. Pratt, J. W. (1961) "Testing Statistical Hypotheses." *Journal of the American Statistical Association*, 56(293): 163-167.
19. Barnard, G. A., G. M. Jenkins, & C. B. Winsten. (1962) "Likelihood Inference and Time Series." *Journal of the Royal Statistical Society, Series A (General)*, 125(3): 321-372.
20. Birnbaum, A. (1962) "On the Foundations of Statistical Inference." *Journal of the American Statistical Association*, 57(298): 269-306.
21. Pearson, E. S. (1962) "Some Thoughts on Statistical Inference." *Annals of Mathematical Statistics*, 33(2): 394-403.
22. Fraser, D. A. S. (1963) "On the Sufficiency and Likelihood Principles." *Journal of the American Statistical Association*, 58(303): 641-647.
23. Kendall, M. G. (1963) "Ronald Aylmer Fisher, 1890-1962." *Biometrika*, 50(1/2):1-15.
24. Platt, J. R. (1964) "Strong Inference." *Science*, 146(3642): 347-353.
25. Dempster, A. P. and M. Schatzoff. (1965) "Expected Significance Level as a Sensitivity Index for Test Statistics." *Journal of the American Statistical Association*, 60(310): 420-436.
26. Cornfield, J. (1966) "Sequential Trials, Sequential Analysis and the Likelihood Principle." *The American Statistician*, 20: 18-23.
27. Cutler, S. J., et al. (1966) "The Role of Hypothesis Testing in Clinical Trials." *Journal of Chronic Disease*, 19: 857-882.
28. Selvin, H. C. and Stuart, A. (1966) "Data-dredging Procedures in Survey Analysis." *The American Statistician*, 20:20-23.
29. Edwards, A. W. F. (1969) "Statistical Methods in Scientific Inference." *Nature*, 222(June): 1233-1237.
30. Edwards, A. W. F. (1970) "Likelihood." *Nature*, 227(July): 92.

31. Durbin, J. (1970) "On Birnbaum's Theorem on the Relation Between Sufficiency, Conditionality and Likelihood." *Journal of the American Statistical Association*, 65(329): 395-398.
32. Leamer, E. E. (1974) "False Models and Post-Data Model Construction." *Journal of the American Statistical Association*, 69(345): 122-131.
33. Kempthorne, O. (1975) "Inference from Experiments and Randomization." In *A Survey of Statistical Design and Linear Models*, J. N. Srivastava, ed., North-Holland Publishing Company. Pages 303-331.
34. Cox, D. R. (1977) "The Role of Significance Tests." *Scandinavian Journal of Statistics*, 4: 49-70.
35. Guttman, L. (1977) "What is Not What in Statistics." *The Statistician*, 26(2): 81-107.
36. Carver, R. P. (1978) "The Case Against Statistical Significance Testing." *Harvard Educational Review*, 48(3): 378-398.
37. Good, I. J. (1980) "The diminishing significance of a p-value as the sample size decreases." *Journal of Statistical Computation & Simulation*, 11: 307-313.
38. Dolby, G. R. (1982) "The Role of Statistics in the Methodology of the Life Sciences." *Biometrics*, 38: 1069-1083.
39. Good, I. J. (1982) "Standardized tail-area probabilities." *Journal of Statistical Computation and Simulation*, 16: 65-75.
40. Schweder, T. and E. Spjøtvoll. (1982). "Plots of P-values to Evaluate Many Tests Simultaneously." *Biometrics*, 69(3): 493-502.
41. Leamer, E. E. (1983) "Let's Take the Con out of Econometrics." *The American Economic Review*, 73(1): 31-43.
42. Leamer, E. and H. Leonard. (1983) "Reporting the Fragility of Regression Estimates." *The Review of Economics and Statistics*, 65(2): 306-317.
43. Good, I. J. (1984) "How Should Tail-Area Probabilities be Standardized for Sample Size in Unpaired Comparisons?" C191 in *Journal of Statistical Computation and Simulation*, 19: 174.
44. Berger, J. O. (1986) "Are P-Values Reasonable Measures of Accuracy?" In *Pacific Statistical Congress*, I. S. Francis et al., eds., Elsevier Science Publishers, the Netherlands. Pages 21-27.

45. Cox, D. R. (1986) "Some General Aspects of the Theory of Statistics." *International Statistical Review*, 54(2): 117-126.
46. Fleiss, J. L. (1986) "Significance Tests Have a Role in Epidemiologic Research: Reactions to A. M. Walker." *American Journal of Public Health*, 76(5): 559-560.
47. Fleiss, J. L. (1986) "Letters to the Editor: Confidence Intervals vs Significance Tests: Quantitative Interpretation." *American Journal of Public Health*, 76(5): 587.
48. Hall, P. and B. Selinger. (1986) "Statistical Significance: Balancing Evidence Against Doubt." *Australian Journal of Statistics*, 28(3): 354-370.
49. Royall, R. M. (1986) "The Effect of Sample Size on the Meaning of Significance Tests." *The American Statistician*, 40(4): 313-315. Also: Bailey, K. R. (1987) "Comment on Royall (1986)." *The American Statistician*, 41(3): 245-246.
50. Walker, A. M. (1986) "Reporting the Results of Epidemiologic Studies." *American Journal of Public Health*, 76(5): 556-558.
51. Berger, J. O. and M. Delampady. (1987) "Testing Precise Hypotheses." *Statistical Science*, 2(3): 317-352.
52. Berger, J. O. and T. Sellke. (1987) "Testing a Point Null Hypothesis: The Irreconcilability of  $P$  Values and Evidence." *Journal of the American Statistical Association*, 82(397): 112-139.
53. Casella, G. and R. L. Berger. (1987) "Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem." *Journal of the American Statistical Association*, 82(397): 106-135.
54. Hill, B. M. (1987) "The validity of the likelihood principle." *The American Statistician* 41(2) 95-100.
55. Poole, C. (1987) "Beyond the Confidence Interval." *American Journal of Public Health*, 77(2): 195-199.
56. Thompson, W. D. (1987) "Statistical Criteria in the Interpretation of Epidemiologic Data." *American Journal of Public Health*, 77(2): 191-194.
57. Berger, J. O. and D. A. Berry. (1988) "Statistical Analysis and the Illusion of Objectivity." *American Scientist*, 76(2): 159-165.
58. Goodman, S. N. and R. Royall. (1988) "Evidence and Scientific Research." *American Journal of Public Health*, 78(12): 1568-1574.

59. Schweder, T. (1988) "A Significance Version of the Basic Neyman-Pearson Theory for Scientific Hypothesis Testing." *Scandinavian Journal of Statistics*, 15: 225-242.
60. Sorić, B. (1989) "Statistical "Discoveries" and Effect-Size Estimation." *Journal of the American Statistical Association*, 84(406): 608-610.
61. Vuong, Q. H. (1989) "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*, 57(2): 307-333.
62. Anscombe, F. J. (1990) "The Summarizing of Clinical Experiments by Significance Levels." *Statistics in Medicine*, 9: 703-708.
63. Barnard, G. A. (1990) "Must Clinical Trials Be Large? The Interpretation of  $P$ -Values and the Combination of Test Results." *Statistics in Medicine*, 9: 601-614.
64. Begg, C. B. (1990) "On Inferences from Wei's Biased Coin Design for Clinical Trials." *Biometrika*, 77(3): 467-84.
65. Cohen, J. (1990) "Things I Have Learned (So Far)." *American Psychologist*, 45(12): 1304-1312.
66. Peterman, R. M. (1990) "The Importance of Reporting Statistical Power: The Forest Decline and Acidic Deposition Example." *Ecology*, 71(5): 2024-2027.
67. Rice, W. R. (1990) "A Consensus Combined  $P$ -Value Test and the Family-Wide Significance of Component Tests." *Biometrics*, 46: 303-308.
68. Salsburg, D. (1990) "Hypothesis Versus Significance Testing for Controlled Clinical Trials: A Dialogue." *Statistics in Medicine*, 9: 201-211.
69. Besag, J. and P. Clifford. (1991) "Sequential Monte Carlo  $p$ -values." *Biometrika*, 78(2): 301-304.
70. Yoccoz, N. G. (1991) "Commentary: Use, Overuse, and Misuse of Significance Tests in Evolutionary Biology and Ecology." *Bulletin of the Ecological Society of America*, 72(2): 106-111.
71. Goodman, S. N. (1992) "A Comment on Replication,  $P$ -Values and Evidence." *Statistics in Medicine*, 11: 875-879.
72. Wright, S. P. (1992) "Adjusted  $P$ -Values for Simultaneous Inference." *Biometrics*, 48: 1005-1013.
73. Freeman, P. R. (1993) "The Role of  $P$ -Values in Analysing Trial Results." *Statistics in Medicine*, 12: 1443-1452.

74. Lee, Y. J. and H. Quan. (1993) “P-Values After Repeated Significance Testing: A Simple Approximation Method.” *Statistics in Medicine*, 12: 675-684.
75. Lehmann, E. L. (1993) “The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?” *Journal of the American Statistical Association*, 88(424): 1242-1249.
76. McBride, G., J. C. Loftis, and N. C. Adkins. (1993) “What Do Significance Tests Really Tell Us About the Environment?” *Environmental Management*, 17(4): 423-432.
77. Wang, C. (1993) *Sense and Nonsense of Statistical Inference: Controversy, Misuse, and Subtlety*. Marcel Dekker, New York.
78. Goodman, S. N. and J. A. Berlin. (1994) “The Use of Predicted Confidence Intervals When Planning Experiments and the Misuse of Power When Interpreting Results.” *Annals of Internal Medicine*, 121(3): 200-206.
79. Edwards, A. W. F. (1995) “XVIIIth Fisher Memorial Lecture Delivered at the Natural History Museum, London, on Thursday, 20<sup>th</sup> October, 1994, Fiducial Inference and the Fundamental Theorem of Natural Selection.” *Biometrics*, 51(3): 799-809.
80. Goutis, C. and G. Casella. (1995) “Frequentist Post-Data Inference.” *International Statistical Review*, 63(3): 325-344.
81. Keuzenkamp, H. A. and J. R. Magnus. (1995) “On Tests and Significance in Econometrics.” *Journal of Econometrics*, 67: 5-24.
82. Keuzenkamp, H. A. and M. McAleer. (1995) “Simplicity, Scientific Inference, and Econometric Modelling.” *The Economic Journal*, 105: 1-21.
83. Sagan, C (1995) *The Demon-Haunted World: Science as a Candle in the Dark*. (see page 113 for the quote “absence of evidence is not evidence of absence.”)
84. Tsou, T. and R. M. Royall. (1995) “Robust Likelihoods.” *Journal of the American Statistical Association*, 90(429): 316-320.
85. Dollinger, M. et al. (1996) “When is a p-Value a Good Measure of Evidence?” In *Robust Statistics, Data Analysis, and Computer Intensive Methods*, H. Rieder, ed., pages 119-134.
86. Mislevy, R. J. (1996) “Evidence and Inference in Educational Assessment.” *CSE Technical Report 414*, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Graduate School of Education and Information Studies, The Regents of the University of California, Los Angeles.
87. Nester, M. R. (1996) “An Applied Statistician’s Creed.” *Applied Statistics*, 45(4): 401-410.

88. Schervish, M. J. (1996) "P Values: What They Are and What They Are Not." *The American Statistician*, 50(3): 203-206.
89. Bower, B. (1997) "Psychology's Statistical Status Quo Draws Fire." *Science News*, 151: 356-357.
90. Hayes, J. P. and R. J. Steidl. (1997) "Statistical Power Analysis and Amphibian Population Trends." *Conservation Biology*, 11(1): 273-275.
91. Hung, H. M. J., et al. (1997) "The Behavior of the P-Value When the Alternative Hypothesis is True." *Biometrics*, 53: 11-22.
92. Pruzek, R. M. (1997) "An Introduction to Bayesian Inference and its Applications." In *What if There Were No Significance Tests?*, L. L. Harlow, et al., eds. Lawrence Erlbaum Associates, Publishers: Mahwah, New Jersey, & London, pages 287-318.
93. Rindskopf, D. M. (1997) "Testing "Small," Not Null, Hypotheses: Classical and Bayesian Approaches." In *What if There Were No Significance Tests?*, L. L. Harlow, et al., eds. Lawrence Erlbaum Associates, Publishers: Mahwah, New Jersey, & London, pages 319-332.
94. Royall, R. (1997) *Statistical Evidence: A Likelihood Paradigm*. Chapman&Hall/CRC.
95. Thomas, L. (1997) "Retrospective Power Analysis." *Conservation Biology*, 11(1): 276-280.
96. Cherry, S. (1998) "Statistical Tests in Publications of The Wildlife Society." *Wildlife Society Bulletin*, 26(4): 947-953.
97. Efron, B. (1998) "R. A. Fisher in the 21<sup>st</sup> Century: Invited Paper Presented at the 1996 R. A. Fisher Lecture." *Statistical Science*, 13(2): 95-122.
98. Gerard, P. D., et al. (1998) "Limits of Retrospective Power Analysis." *Journal of Wildlife Management*, 62(2): 801-807.
99. Shen, W. and T.A. Louis. (1998) "Triple-goal estimates in two-stage hierarchical models." *Jornal of the Royal Statistical Society B*, 60(2): 455-471.
100. Thompson, J. R. (1998) "Invited Commentary: Re: "Multiple Comparisons and Related Issues in the Interpretation of Epidemiologic Data." *American Journal of Epidemiology*, 147(9): 801-806.
101. Vieland, V. J. and S. E. Hodge. (1998) "Book Reviews: Statistical Evidence: A Likelihood Paradigm, by Richard Royall." *American Journal of Human Genetics*, 63: 283-289.
102. Zumbo, B. D. and A. M. Hubley. (1998) "A note on misconceptions concerning prospective and retrospective power." *The Statistician*, 47(2): 385-388.

103. Donahue, R. M. J. (1999) "A Note on Information Seldom Reported Via the  $P$  Value." *The American Statistician*, 53(4): 303-306.
104. Goodman, S. N. (1999) "Toward Evidence-Based Medical Statistics. 1: The  $P$  Value Fallacy." *Annals of Internal Medicine*, 130(12): 995-1004.
105. Goodman, S. N. (1999) "Toward Evidence-Based Medical Statistics. 1: The Bayes Factor." *Annals of Internal Medicine*, 130(12): 1005-1013.
106. Johnson, D. H. (1999) "The Insignificance of Statistical Significance Testing." *Journal of Wildlife Mangement*, 63(3): 763-772.
107. Lindsey, J. K. (1999) "Some Statistical Heresies." *The Statistician*, 48(1): 1-40.
108. Perlman, M. D. and L. Wu. (1999) "The Emperor's New Tests." *Statistical Science*, 14(4): 355-381.
109. Sackrowitz, H. and E. Samuel-Cahn. (1999) " $P$  Values as Random Variables—Expected  $P$  Values." *The American Statistician*, 53(4): 326-331.
110. Stockmarr, A. (1999) "Likelihood Ratios for Evaluating DNA Evidence When the Suspect is Found Through a Database Search." *Biometrics*, 55: 671-677.
111. Bayarri, M. J. and J. O. Berger. (2000) " $P$  Values for Composite Null Models." *Journal of the American Statistical Association*, 95(452): 1127-1172.
112. Robinson, A. (2000) "Slides from A. Robinson's talk on *A Jaundiced View of Hypothesis and Significance Testing*." University of Idaho.
113. Royall, R. (2000) "On the Probability of Observing Misleading Statistical Evidence." *Journal of the American Statistical Association*, 95(451): 760-773.
114. Barker, L., H. Rolka, D. Rolka, C. Brown. (2001) "Equivalence Testing for Binomial Random Variables: Which Test to Use?" *The American Statistician*, 55(4): 279.
115. Dennis, B. (2001) "Statistics and the Scientific Method in Ecology." Draft for *The Nature of Scientific Evidence*, M. L. Taper and S. R. Lele, eds., The University of Chicago Press. 33 pages.
116. Gregoire, T. G. (2001) "Biometry in the 21<sup>st</sup> Century: Whither Statistical Inference?" Keynote address presented at *The Conference on Forest Biometry and Information Science* (<http://cms1.gre.ac.uk/conferences/iufro/proceedings/>), 25-29 June 2001, The University of Greenwich, London, U.K.

117. Hoenig, J. M. and D. M. Heisey. (2001) "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis." *The American Statistician*, 55(1): 1-6.
118. Lenth, R. V. (2001) "Some Practical Guidelines for Effective Sample Size Determination." *The American Statistician*, 55(3): 187-193.
119. Pace, M. L. (2001) "Prediction and the Aquatic Sciences." *Canadian Journal of Fisheries and Aquatic Sciences*, 58: 63-72.
120. Salsburg, D. (2001) *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. A. W. H. Freeman: New York.
121. Schenker, N. and J. F. Gentleman. (2001) "On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals." *The American Statistician*, 55(3): 182-186.
122. Schnute, J. T. and L. J. Richards. (2001) "Use and Abuse of Fishery Models." *Canadian Journal of Fisheries and Aquatic Science*, 58: 10-17.
123. Schweder, T. and N. L. Hjort. (2001) "Confidence and Likelihood." *Statistical Research Report*, Department of Mathematics, University of Oslo. [ISBN: 82-553-1278-1]
124. Selke, T., M. J. Bayarri, and J. O. Berger. (2001) "Calibration of  $p$  values for testing precise null hypotheses." *The American Statistician* 55(1) 62-71.
125. Senn, S. (2001) "Statistical Issues in Bioequivalence." *Statistics in Medicine*, 20: 2785-2799.
126. Stern, J. A. C. and G. D. Smith. (2001) "Sifting the Evidence—What's Wrong With Significance Tests?" *British Medical Journal*, 322: 226-231.
127. Berger, J. O. (2002) "Could Fisher, Jeffreys, and Neyman Have Agreed on Testing?" Paper based on *the Fisher lecture*, given at the 2001 Joint Statistical Meetings by the author, Duke University.
128. Blume, J. D. (2002) "Likelihood methods for measuring statistical evidence." *Statistics in Medicine* 21: 2563-2599.
129. Farrant, T. (2002) "To  $p$  or not to  $p$ ." *Royal Statistical Society News*, 29(10): 21.
130. Goodman, S. N. (2002) "Author's Reply." *Statistics in Medicine*, 21: 2445-2447.
131. Knapp, T. R. (2002) "Some reflections on significance testing" *Journal of Modern Applied Statistical Methods* 1(2) 240-242.

132. Senn, S. (2002) "Letter to the Editor: A comment on replication,  $p$ -values and evidence." *Statistics in Medicine*, 21: 2437-2444.
133. Browner, W. S. (2003) "The reliability of P values". *Science*, 301, 167-168.
134. Dass, S. C. and J. O. Berger. (2003) "Unified Conditional Frequentist and Bayesian Testing of Composite Hypotheses." *Scandinavian Journal of Statistics*, 30: 193-210.
135. Edwards, A.W.F. (2003) "R.A. Fisher—twice Professor of Genetics: London and Cambridge, or 'A fairly well-known geneticists'." *The Statistician*, 52(3): 311-318.
136. Green, P. J. (2003) "Notes on the life and work of R.A. Fisher." *The Statistician*, 52(3): 299-301.
137. Healy, M. J. R. (2003) "R. A. Fisher the statistician." *The Statistician*, 52(3): 303-310.
138. Hubbard, R. and M. J. Bayarri. (2003) "Confusion Over Measure of Evidence ( $p$ 's) Versus Errors ( $\alpha$ 's) in Classical Statistical Testing." *The American Statistician*, 57(3): 171-178.
139. Lele, S. (2003). Various work and correspondence. Includes: Subhash, L. and A. Das. (2003) "Elicited data and incorporation of expert opinion for statistical inference in spatial studies." *Mathematical Geology*, 32(4): 465-467.
140. Onwuegbuzie, A.J. and J.R. Levin. (2003) "Without supporting statistical evidence, where would reported measures of substantive importance lead? To no good effect." *Journal of Modern Applied Statistical Methods*, 2(1): 133-151.
141. Senn, S. (2003) "Foreword: A blue plaque for Fisher." *The Statistician*, 52(3) 297-298.
142. Christensen, R. (2005) Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician*, 59(2)121-126.