

If you only have time to read one article I would suggest you to read the article item [146] in this list; If you have time to read two articles I would suggest you to read items [146] and [141]; If you have time to read one book, it is item [2] (**The Lady Tasting Tea**) that I would recommend; If you have 30 minutes to watch a video, it is the very last item [152] I would recommend:

<https://www.youtube.com/watch?v=iJ4kqk3V8jQ> online video presented by Professor Geoff Cumming. I truly believe, either as a statistician or as a researcher/scientist who need to use statistical analysis seriously, it is worth spending your precious time to read at least some of the references listed here.

The message is clear: Null Hypothesis Significance Testing (NHST) should hardly have a place in statistical inference or scientific reasoning; any attempt to claim a 'statistical significance' by dichotomizing (or categorizing) a continuous testing measure (e.g., a p -value, a confidence interval, or Bayes factors) is not logically defensible in theory, flawed technically, and damaging in practice; with data sets obtained from non-repetitive studies (whether or not an experimental design) the best statistical data analysis that one can do is what-if analysis (namely, no conclusive/confirmatory statement can be made about the population true value based on sampling distribution ground).

List of a selection of literature related to concerns/criticisms on NHST (in chronological order):

Part I: Books

1. Edited by Denton E. Morrison and Ramon E. Henkel (1970). ***The Significance Test Controversy***. Routledge, Taylor & Francis Group.
2. David Salsburg (2001). ***The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century***. Henry Holt and Company.
3. Burham, K. and Anderson, D. (2002). ***Model Selection and Multimodel Inference: a practical information-theoretic approach***. Springer.
4. E.T. Jaynes (edited by G. Larry Bretthorst) (2003). ***Probability Theory: the logic of science***. Cambridge University Press.
5. Richard A. Berk (2004). ***Regression Analysis: A Constructive Critique***. SAGE.
6. Stephen T. Ziliak and Deirdre N. McCloskey (2008). ***The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives***. The University of Michigan Press.
7. Raymond Hubbard (2015). ***Corrupt Research: The case for reconceptualizing empirical management and social science***. SAGE Publications, Inc.
8. Richard McElreath (2016). ***Statistical Rethinking: A Bayesian Course with Examples in R and Stan***. CRC Press, Taylor & Francis Group.
9. Weichung Joe Shih and Joseph Aisner (2016). ***Statistical Design and Analysis of Clinical Trials: Principles and Methods***. CRC Press, Taylor & Francis Group.
10. Geoff Cumming and Robert Calin-Jageman (2017). ***Introduction to The New Statistics: Estimation, Open Science, & Beyond***. Routledge.
11. Hadley Wickham & Garrett Golemund (2017). ***R for Data Science: Import, Tidy, Transform, Visualize, and Model Data***. O'Reilly.
12. Richard F. Harris (2017). ***Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions***. BASIC BOOKS.
13. Edited by Vladik Kreinovich, Nguyen Ngoc Thach, Nguyen Duc Trung, and Dang Van Thanh (2019). ***Beyond Traditional Probabilistic Methods in Economics***. Springer
14. David Spiegelhalter (2019). ***The Art of Statistics: How to learn from data***. BASIC BOOKS, New York.

Part II: Articles (including book chapters if any)

15. F. Yates (1951). **The Influence of Statistical Methods for Research Workers on the Development of the Science of Statistics.** Journal of the American Statistical Association, Vol. 46, No. 253, pp. 19-34
16. William W. Rozeboom (1960). **THE FALLACY OF THE NULL-HYPOTHESIS SIGNIFICANCE TEST.** Psychological Bulletin, Vol. 57, No. 5, 416-428
17. David Bakan (1966). **The Test of Significance in Psychological Research.** Psychological Bulletin, Vol. 66, No. 6, 423-437.
18. Ronald N. Giere (1972). **The Significance Test Controversy.** The British Journal for the Philosophy of Science, Vol. 23, No. 2, pp. 170-181.
19. W. Edwards Deming (1975). **On Probability As a Basis For Action.** The American Statistician, Vol. 29, No. 4, pp. 146-152.
20. George E. P. Box (1976). **Science and Statistics.** Journal of the American Statistical Association, Vol. 71, No. 356, pp. 791-799.
21. Leonard J. Savage (1976). **On Rereading R. A. Fisher.** The annals of Statistics, Vol. 4, No. 3, 441-500.
22. Ronald P. Carver (1978). **The Case Against Statistical Significance Testing.** Harvard Educational Review, Vol. 48, Issue 3, pages 378-399.
23. Mario Bunge (1981). **Four concepts of probability.** Appl. Math. Modelling, Vol. 5, pp. 306-312.
24. C. Chatfield (1985). **The Initial Examination of Data.** Journal of the Royal Statistical Society. Series A (General), Vol. 148, No. 3, pp. 214-253.
25. Terry Speed (1986). **Questions, Answers and Statistics.** ICOTS 2, 18-28.
26. Martin J. Gardner and Douglas G. Altman (1986). **Confidence intervals rather than P values: estimation rather than hypothesis testing.** Statistics in Medicine, British Medical Journal, Vol. 292, pp. 746-750.
27. James O. Berger and Thomas Sellke (1987). **Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence.** Journal of the American Statistical Association, Vol. 82, No. 397, pp.112-122.
28. Steven N. Goodman & Richard Royall (1988). **Evidence and Scientific Research.** (Commentary) American Journal of Public Health, Vol. 78, No. 12, pp. 1568-1574.
29. Nigel G. Yoccoz (1991). **Use, Overuse, and Misuse of Significance Tests in Evolutionary Biology and Ecology.** Bulletin of the Ecological Society of America, Vol. 72, No. 2, pp. 106-111.
30. E.L. Lehmann (1993). **The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?** Journal of the American Statistical Association, Vol. 88, No. 424, 201-208.
31. Gerald J. Hahn and William Q. Meeker (1993). **Assumptions for Statistical Inference.** The American Statistician, Vol. 47, No. 1, pp. 1-11
32. Ronald P. Carver (1993). **The Case Against Statistical Significance Testing, Revisited.** Journal of Experimental Education, 61(A), 287-292.
33. Rama Menon (1993). **Statistical Significance Testing Should be Discontinued in Mathematics Education Research.** Mathematics Education Research Journal, Vol. 5, No. 1, 4-18.
34. Steven N. Goodman (1993). **p Values, Hypothesis Tests, and Likelihood: Implications for Epidemiology of a Neglected Historical Debate.** American Journal of Epidemiology, Vol. 137, No.5, pp.485-496.

35. Jacob Cohen (1994). **The Earth Is Round ($p < .05$)**. American Psychologist, Vol.49, No. 12, 997-1003.
36. Bruce Thompson (1994). **The Concept of Statistical Significance Testing**. Practical Assessment, Research & Evaluation, Vol. 4, No. 5, Available online: <http://PAREonline.net/getvn.asp?v=4&n=5>.
37. Bruce Thompson (1994). **The Pivotal Role of Replication in Psychological Research: Empirically Evaluating the Replicability of Sample Results**. Journal of Personality 62:2, 157-176.
38. Ruma Falk & Charles W. Greenbaum (1995). **Significance Test Die Hard: The Amazing Persistence of a Probabilistic Misconception**. Theory & Psychology 5(1), 75-98.
39. R.E. Kirk (1996). **Practical significance: A concept whose time has come**. Educational and Psychological Measurement, 56, 746-759.
40. Marks R. Nester (1996). **An Applied Statistician's Creed**. Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 45, No. 4, 401-410.
41. Frank L. Schmidt (1996). **Statistical Significance Testing and Cumulative Knowledge in Psychology: Implications for Training of Researchers**. Psychological Methods, Vol. 1, No. 2, 115-129.
42. Bruce Thompson (1996). **AERA Editorial Policies Regarding Statistical Significance Testing: Three Suggested Reforms**. Educational Researcher, Vol. 25, No. 2, pp. 26-30.
43. Robert P. Abelson (1997). **On the Surprising Longevity of Flogged Horses: Why There Is a Case for the Significance Test**. Psychological Science, Volume 8 issue 1, pp. 12-15.
44. Patrick E. Shrout (1997). **Should Significance Tests Be Banned? Introduction to a Special Section Exploring the Pros and Cons**. Psychological Sciences, Vol. 8, No. 1, 1-2.
45. Frank L. Schmidt (1997). **Eight common but false objections to the discontinuation of significance testing in the analysis of research data, in book: What if there were no significance tests?** Editors: Lisa L. Harlow, Stanley A. Mulaik, James H. Steiger, Publisher: Lawrence Erlbaum Associates.
46. Janet M. Lang, Kenneth J. Rothman, and Cristina I. Cann (1998). **That Confounded P-Value**. Epidemiology, Volume 9, Number 1, 7-8.
47. James E. McLean and James M. Ernest (1998). **The Role of Statistical Significance Testing In Educational Research**. Research in the Schools, Vol. 5, No. 2, 15-22.
48. James Currall (1998). **Review on the book 'Statistical Significance: Rationale, Validity and Utility' (Siu L. CHOW, 1996)**. Journal of the Royal Statistical Society. Series D (The Statistician), Vol. 47, No. 2, pp. 394-395
49. Robert W. Frick; Gerd Gigerenzer (1998). **Two individual reviews on the book 'Statistical Significance: Rationale, Validity and Utility' (Siu L. CHOW, 1996)**. BEHAVIORAL AND BRAIN SCIENCES (1998) 21:2, 199-200.
50. Tapan K. Nayak (1998). **Review on the book 'Statistical Significance: Rationale, Validity and Utility' (Siu L. CHOW, 1996)**. TECHNOMETRICS, MAY 1998, VOL. 40, NO. 2
51. David H. Krantz (1999). **The Null Hypothesis Testing Controversy in Psychology**. Journal of the American Statistical Association, Vol. 44, No. 448, pp. 1372-1381.
52. Douglas H. Johnson (1999). **The Insignificance of Statistical Significance Testing**. Journal of Wildlife Management, 63(3): 763-772.
53. Howard Wainer (1999). **One Cheer for Null Hypothesis Significance Testing**. Psychological Methods, Vol. 4, No. 2, 212-213.
54. Anderson, D. R., Burnham, K. P., and Thompson, W. L. (2000). **Null Hypothesis Testing: Problems, Prevalence, and an Alternative**. Journal of Wildlife Management, 64, 912-923.

55. John I. Marden (2000). **Hypothesis Testing: From p Values to Bayes Factors.** Journal of the American Statistical Association, Vol. 95, No. 452, 1316-1320.
56. Raymond S. Nickerson (2000). **Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy.** Psychological Methods, Vol. 5, No. 2, 241-301.
57. Charles Poole (2001). **Low P -values or Narrow Confidence Intervals: Which Are More Durable?** Epidemiology, Vol. 12, No. 3, 291-294.
58. Joachim Krueger (2001). **Null Hypothesis Significance Testing: On the Survival of a Flawed Method.** American Psychologist, Vol. 56, No. 1, 16-26. DOI: 10.1037//0003-066X.56.1.16.
59. Jonathan A. C. Sterne and George Davey Smith (2001). **Sifting the evidence-what's wrong with significance tests?** BMJ, 322:226-31.
60. Gerd Gigerenzer (2002). **The Superego, the Ego, and the Id in Statistical Reasoning.** Print publication date: 2002; Published to Oxford Scholarship Online: October 2011; DOI: 10.1093/acprof:oso/9780195153729.001.0001.
61. Jeffrey A. Gliner, Nancy L. Leech, and George A. Morgan (2002). **Problems With Null Hypothesis Significance Testing (NHST): What Do the Textbooks Say?** The Journal of Experimental Education, 71(1), 83-92.
62. Shlomo S. Sawilowsky (2003). **Deconstructing arguments from the case against hypothesis testing.** Journal of Modern Applied Statistical Methods, 2(2), 467-474. Available at: http://digitalcommons.wayne.edu/coe_tbf/17
63. Michael D. Jennions and Anders Pape Moller (2003). **A survey of the statistical power of research in behavioral ecology and animal behaviour.** Behavioral Ecology Vol. 14 No. 3: 438-445.
64. Raymond Hubbard, M. J. Bayarri, Kenneth N. Berk and Matthew A. CarltonSource (2003). **Confusion over Measures of Evidence (p 's) versus Errors (α 's) in Classical Statistical Testing.** The American Statistician, Vol. 57, No. 3, pp. 171-182.
65. Shinichi Nakagawa (2004). **A farewell to Bonferroni: the problems of low statistical power and publication bias.** Behavioural Ecology, Vol. 15, No. 6: 1044-1045, doi:10.1093/beheco/arh107.
66. Gerd Gigerenzer (2004). **Mindless statistics.** The Journal of Socio-Economics 33, 587-606.
67. Ioannidis JPA (2005). **Why most published research findings are false.** PLoS Med 2: e124. doi:10.1371/journal.pmed.0020124
68. Nekane Balluerka, Juana Gomez, and Dolores Hidalgo (2005). **The Controversy over Null Hypothesis Significance Testing Revisited.** Methodology European Journal of Research Methods for the Behavioral and Social Sciences, Vol. 1(2):55-70, DOI 10.1027/1614-1881.1.2.55
69. Editorial (2006). **Some experimental design and statistical criteria for analysis of studies in manuscripts submitted for consideration for publication.** Animal Feed Science and Technology 129, 1-11.
70. Andrew Gelman and Hal Stern (2006). **The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant.** The American Statistician, November 2006, Vol. 60, No. 4, 328-331.
71. Stephen Gorard (2006). **Towards a judgement-based statistical analysis.** British Journal of Sociology of Education, 27:1, 67-80, DOI: 10.1080/01425690500376663
72. Goodman S, Greenland S (2007). **Why most published research findings are false: Problems in the analysis.** PLoS Med 4(4): e168. doi:10.1371/journal.pmed.0040168
73. Raymond Hubbard and J. Scott Armstrong (2006). **Why We Don't Really Know What Statistical Significance Means: A Major Educational Failure.** Journal of Marketing Education, Vol. 28, pp. 114-120.

74. James M. Gibbons, Neil M.J. Crout and John R. Healey (2007). **What role should null-hypothesis significance tests have in statistical education and hypothesis falsification?** (Letter to editor) *TRENDS in Ecology and Evolution* Vol.22 No.9, 445-446.
75. Shinichi Nakagawa and Innes C. Cuthill (2007). **Effect size, confidence interval and statistical significance: a practical guide for biologists.** *Biol. Rev.*, 82, pp. 591-605, doi:10.1111/j.1469-185X.2007.00027.x.
76. Zab Mosenifar (2007). **Population Issues in Clinical Trials.** *Proc Am Thorac Soc* Vol 4. pp 185–188, DOI: 10.1513/pats.200701-009GC.
77. Timothy R. Levine, et al. (2008). **A Critical Assessment of Null Hypothesis Significance Testing in Quantitative Communication Research.** *Human Communication Research* 34, 171–187.
78. Aris Spanos (2008). **Review of Stephen T. Ziliak and Deirdre N. McCloskey's *The cult of statistical significance: how the standard error costs us jobs, justice, and lives.* Ann Arbor (MI): The University of Michigan Press, 2008, xxiii+322 pp.** *Erasmus Journal for Philosophy and Economics*, Volume 1, Issue 1, pp. 154-164.
79. Stephen T. Ziliak and Deirdre N. McCloskey (2008). **Science is judgment, not only calculation: a reply to Aris Spanos's review of *The cult of statistical significance.*** *Erasmus Journal for Philosophy and Economics*, Volume 1, Issue 1, pp. 165-170.
80. Timothy R. Levine, Rene Weber, Craig Hullett, Hee Sun Park, & Lisa L. Massi Lindsey (2008). **A Critical Assessment of Null Hypothesis Significance Testing in Quantitative Communication Research.** *Human Communication Research* 34, pp. 171–187. doi:10.1111/j.1468-2958.2008.00317.x.
81. Stuart H. Hurlbert and Celia M. Lombardi (2009). **Final Collapse of the Neyman-Pearson decision theoretic framework and rise of the neoFisherian.** *Ann. Zool. Fennici* 46: 311-349.
82. Stephen R. Cole and Elizabeth A. Stuart (2010). **Generalizing Evidence From Randomized Clinical Trials to Target Populations The ACTG 320 Trial.** *American Journal of Epidemiology*, 172:107–115.
83. Joseph Lee Rodgers (2010). **The Epistemology of Mathematical and Statistical Modeling: A Quiet Methodological Revolution.** *American Psychologist*, Vol. 65, No. 1, 1–12. DOI: 10.1037/a0018326.
84. Stephen Gorard (2010). **All evidence is equal: the flaw in statistical reasoning.** *Oxford Review of Education*, Vol. 36, No. 1, pp. 63-77.
85. Andreas Stang, Charles Poole, and Oliver Kuss (2010). **The ongoing tyranny of statistical significance testing in biomedical research.** *Eur J Epidemiol* 25:225-230. DOI 10.1007/s10654-010-9440-x
86. Daniel Greco (2011). **Significance Testing in Theory and Practice.** *Brit. J. Phil. Sci.* 62, 607–637. doi:10.1093/bjps/axq023.
87. Douglas G. Altman (2011). **How to obtain the P value from a confidence interval.** *BMJ*, 343:d2304, doi: <https://doi.org/10.1136/bmj.d2304>
88. James Tabery (2011). **Commentary: Hogben vs the Tyranny of Averages.** *International Journal of Epidemiology*, 40:1458–1460, doi:10.1093/ije/dyr031
89. John P. A. Ioannidis (2012). **Why Science Is Not Necessarily Self-Correcting.** *Perspectives on Psychological Science* 7(6) 645-654. DOI: 10.1177/1745691612464056.
90. Andrew Gelman (2013). **P Values and Statistical Practice.** *Epidemiology*, Volume 24, Number 1, 69-72.
91. Jesper W. Schneider (2013). **Caveats for using statistical significance tests in research assessments.** *Journal of Informetrics* 7, 50– 62.

92. Andreas Stang and Charles Poole (2013). **The researcher and the consultant: a dialogue on null hypothesis significance testing.** *Eur J Epidemiol* (2013) 28:939–944, DOI 10.1007/s10654-013-9861-4
93. Dalson Britto Figueiredo Filho, et al. (2013). **When is statistical significance not significant?** *Brazilian political science review*, 7(1), pages 31-55.
94. Andrew Gelman and Eric Loken (2014). **The Statistical Crisis in Science: Data-dependent analysis – a “garden of forking paths” – explains why many statistically significant comparisons don’t hold up.** *American Scientist*, Volume 102, pp. 460-465.
95. Andrew Gelman and John Carlin (2014). **Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors.** *Perspectives on Psychological Science*, Vol. 9(6) 641-651.
96. Regina Nuzzo (2014). **Statistical Errors: *p* values, the ‘gold standard’ of statistical validity, are not as reliable as many scientists assume.** *Nature*, Vol. 506, 150-152.
97. Geoff Cumming (2014). **The New Statistics: Why and How.** *Psychological Science*, Vol. 25(1), 7-29, DOI: 10.1177/0956797613504966
98. Gerd Gigerenzer & Julian N. Marewski (2014). **Surrogate Science: The Idol of a Universal Method for Scientific Inference.** *Journal of Management*, Vol. 41, No. 2, pp. 421-440. DOI: 10.1177/0149206314547522.
99. Paul A. Murtaugh (2014). **In defense of *P* values.** *Ecology*, 95(3), 2014, pp. 611–617.
100. S. Gorard (2014). **The widespread abuse of statistics by researchers: what is the problem and what is the ethical way forward?** *Psychology of education review*, 38 (1). pp. 3-10.
101. P. White (2014). **A Response to Gorard: The widespread abuse of statistics by researchers: What is the problem and what is the ethical way forward?** *The Psychology of Education Review*, 38(1), pp. 24-28.
102. Editorial (2014). **Business Not as Usual.** *Psychological Science*, Vol. 25(1) 3-6. DOI: 10.1177/0956797613512465.
103. Dave Neale (2015). **Defending the logic of significance testing: a response to Gorard.** *Oxford Review of Education*, 41:3, 334-345, DOI: 10.1080/03054985.2015.1028526
104. Jesper W. Schneider (2015). **Null hypothesis significance tests. A mix-up of two different theories: the basis for widespread confusion and numerous misinterpretations.** *Scientometrics*, 102: 411-432, DOI 10.1007/s11192-014-1251-5.
105. Gerd Gigerenzer & Julian N. Marewski (2015). **Surrogate Science: The Idol of a Universal Method for Scientific Inference.** *Journal of Management*, Vol. 41 No. 2, February 2015 421–440, DOI: 10.1177/0149206314547522.
106. Jose D. Perezgonzalez (2015). **Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing.** *Frontiers in Psychology*, Volume 6, Article 223.
107. Roger Peng (2015) **The reproducibility crisis in science: A statistical counterattack.** *Science: significance*, pp.30-32. The Royal Statistical Society.
108. Ronald L. Wasserstein & Nicole A. Lazar (2016). **The ASA's Statement on *p*-Values: Context, Process, and Purpose.** *The American Statistician*, 70:2, 129-133, DOI:10.1080/00031305.2016.1154108.
109. John Concato & John A. Hartigan (2016). ***P* values: from suggestion to superstition.** *J Investig Med* 2016;64:1166–1171. doi:10.1136/jim-2016-000206
110. Blakeley B. McShane and David Gal (2016). **Blinding Us to the Obvious? The Effect of Statistical Training on the Evaluation of Evidence.** *Management Science* 62(6):1707-1718. <http://dx.doi.org/10.1287/mnsc.2015.2212>

111. Kenneth J. Rothman (2016). **Disengaging from statistical significance.** Eur J Epidemiol (2016) 31:443–444. DOI 10.1007/s10654-016-0158-2
112. Sander Greenland, et al. (2016). **Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations.** Eur J Epidemiol (2016) 31:337–350, DOI 10.1007/s10654-016-0149-3
113. Andrew Gelman (2016). **The Problems With P-Values are not Just With P-Values.** Online discussion of the ASA Statement on Statistical Significance and P-Values, The American Statistician, 70.
114. Jeehyoung Kim and Heejung Bang (2016). **Three common misuses of P values.** Dent Hypotheses, 7(3): 73–80. doi:10.4103/2155-8213.190481
115. Kenneth J. Rothman (2016). **Disengaging from statistical significance.** Eur J Epidemiol 31:443–444 DOI 10.1007/s10654-016-0158-2
116. Robert E. Kass, Brian S. Caffo, Marie Davidian, Xiao-Li Meng, Bin Yu and Nancy Reid (2016). **Ten Simple Rules for Effective Statistical Practice.** (Editorial) PLOS Computational Biology | DOI:10.1371/journal.pcbi.1004961.
117. Steven N. Goodman, Daniele Fanelli, John P. A. Ioannidis (2016). **What does research reproducibility mean?** Sci Transl Med 8, 341ps12341ps12, DOI: 10.1126/scitranslmed.aaf5027
118. Amrhein et al. (2017). **The earth is flat ($p > 0.05$): significance thresholds and the crisis of unreplicable research.** PeerJ 5:e3544; DOI 10.7717/peerj.3544
119. Donald Berry (2017). **A p-Value to Die For.** Journal of the American Statistical Association, 112:519, 895-897, DOI: 10.1080/01621459.2017.1316279
120. Robert Matthews (2017). **The ASA's p-value statement, one year on.** The Royal Statistical Society, In Practice, 38-41.
121. Joseph Kang, Jaeyoung Hong, Precious Esie, Kyle T. Bernstein, and Sevgi Aral (2017). **An Illustration of Errors in Using the P Value to Indicate Clinical Significance or Epidemiological Importance of a Study Finding.** Sex Transm Dis., 44(8): 495–497. doi:10.1097/OLQ.0000000000000635
122. Brian D. Haig (2017). **Tests of Statistical Significance Made Sound.** Educational and Psychological Measurement, Vol. 77(3) 489–506
123. Denes Szucs and John P.A. Ioannidis (2017). **When Null Hypothesis Significance Testing Is Unsuitable for Research: A Reassessment.** Frontiers in Human Neuroscience, Volume 11, Article 390.
124. Timothy L. Lash (2017). **The Harm Done to Reproducibility by the Culture of Null Hypothesis Significance Testing.** American Journal of Epidemiology, Vol. 186, No. 6, DOI: 10.1093/aje/kwx261
125. Sander Greenland (2017). Invited Commentary: **The Need for Cognitive Science in Methodology.** American Journal of Epidemiology, Vol. 186, No. 6 DOI: 10.1093/aje/kwx259
126. Andrew Gelman (2018). **The Failure of Null Hypothesis Significance Testing When Studying Incremental Changes, and What to Do About It.** Personality and Social Psychology Bulletin, Vol. 44(1) 16-23.
127. Benjamin et al. (2018). **Redefine Statistical Significance.** Nature Human Behaviour, 2, 6–10.
128. Jeffrey R. Spence and David J. Stanley (2018). **Concise, Simple, and Not Wrong: In Search of a Short-Hand Interpretation of Statistical Significance.** Frontiers in Psychology, Volume 9, Article 2185.

129. Harry Crane (2018). **The Impact of *P*-hacking on “Redefine Statistical Significance”**. *Basic and Applied Social Psychology*, 40:4, 219-235, DOI: 10.1080/01973533.2018.1474111.
130. Gerd Gigerenzer (2018). **Statistical Rituals: The Replication Delusion and How We Got There**. *Advances in Methods and Practices in Psychological Science*, Vol. 1(2) 198–218.
131. Van Calster B, Steyerberg, EW, Collins GS, and Smits T. (2018). **Consequences of relying on statistical significance: Some illustrations**. *Eur J Clin Invest*. 48:e12912. <https://doi.org/10.1111/eci.12912> .
132. Valentin Amrhein, Sander Greenland, Blake McShane (2019). **Retire statistical significance**. *Nature*, Vol. 567, 305: Comment.
133. Ronald D. Fricker Jr., Katherine Burke, Xiaoyan Han & William H. Woodall (2019). **Assessing the Statistical Analyses Used in *Basic and Applied Social Psychology* After Their *p*-Value Ban**. *The American Statistician*, 73:sup1, 374-384, DOI: 10.1080/00031305.2018.1537892
134. Blakeley B. McShane, et al. (2019). **Abandon Statistical Significance**. *The American Statistician*, Vol. 73, No. S1, 235-245: Statistical Inference in the 21st Century.
135. Christopher Tong (2019). **Statistical Inference Enables Bad Science; Statistical Thinking Enables Good Science**. *The American Statistician*, Vol. 73, No. S1, 246-261: Statistical Inference in the 21st Century.
136. Dana P. Turner, Hao Deng and Timothy T. Houle (Guest Editorial, 2019). **Statistical Hypothesis Testing: Overview and Application**. *Headache*, pages 302-307. doi: 10.1111/head.13706.
137. Deborah G. Mayo (2019). ***P* - value thresholds: Forfeit at your peril**. *Eur J Clin Invest.*, 49:e13170. <https://doi.org/10.1111/eci.13170>
138. Andrew Gelman (2019). **When we make recommendations for scientific practice, we are (at best) acting as social scientists**. *Eur J Clin Invest.*, 49:e13165. DOI: 10.1111/eci.13165
139. Tom E. Hardwicke & John P.A. Ioannidis (2019). **Petitions in scientific argumentation: Dissecting the request to retire statistical significance**. *Eur J Clin Invest.*, 49:e13162. <https://doi.org/10.1111/eci.13162>
140. Horbert Hirschauer, Sven Gruner, oliver Mußhoff and Claudia Becker (2019). **Twenty Steps Towards an Adequate Inferential Interpretation of *p*-Values in Econometrics**. *Journal of Economics and Statistics*, 239(4):703–721
141. Raymond Hubbard, Brian D. Haig & Rahul A. Parsa (2019). **The Limited Role of Formal Statistical Inference in Scientific Inference**. *The American Statistician*, 73:sup1, 91-98, DOI: 10.1080/00031305.2018.1464947
142. Raymond Hubbard (2019). **Will the ASA's Efforts to Improve Statistical Practice be Successful? Some Evidence to the Contrary**. *The American Statistician*, 73:sup1, 31-35, DOI: 10.1080/00031305.2018.1497540
143. Rob Herbert (2019). **Research Note: Significance testing and hypothesis testing: meaningless, misleading and mostly unnecessary**. *Journal of Physiotherapy*, 65, 178-181.
144. Valentin Amrhein, David Trafimow & Sander Greenland (2019). **Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don't Expect Replication**. *The American Statistician*, Vol. 73, No. S1, 262-270: Statistical Inference in the 21st Century.
145. Vincent S. Staggs (2019). **Why statisticians are abandoning statistical significance**. Guest Editorial, *Res Nurs Health*, 42:159–160, DOI: 10.1002/nur.21947.
146. Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar (2019). **Moving to a World Beyond “*p*<0.05”**. *The American Statistician*, Vol. 73, No. S1, 1-19: Editorial.

147. Caiyun Liao, Andrew L. Speirs, Sierra Goldsmith, & Sherman J. Silber (2020). **When “facts” are not facts: what does p value really mean, and how does it deceive us?** Journal of Assisted Reproduction and Genetics, 37:1303-1310, <https://doi.org/10.1007/s10815-020-01751-4>.

Part III: Online materials

148. <http://www.stats.org.uk/statistical-inference/> the link for Statistical Inference (and What is Wrong With Classical Statistics) – a long list of references.
149. <https://fionaresearch.files.wordpress.com/2013/06/fidler-phd-2006.pdf> Fiona Fidler’s PhD thesis **“FROM STATISTICAL SIGNIFICANCE TO EFFECT ESTIMATION: STATISTICAL REFORM IN PSYCHOLOGY, MEDICINE AND ECOLOGY.”**
150. <https://learningstatisticswithr.com/book/> Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.6.1). 2019-01-11, Danielle Navarro (UNSW, Australia)
151. <https://www.fharrell.com/post/introduction/> Frank Harrell, author of an influential book on regression modeling and currently both a biostatistics professor and a statistician at the Food and Drug Administration sums up “some of his personal philosophy of statistics” here.
152. <https://www.youtube.com/watch?v=iJ4kqk3V8jQ> online video presented by Professor Geoff Cumming, La Trobe University, Australia