References and further reading (Tutorial on Statistically Sound Pattern Discovery)

1. A. Agresti: Categorical Data Analysis. 3rd edit. Wiley Series in Probability and Statistics. Wiley, 2013.

2. Y. Benjamini and Y. Hochberg: Controlling the false discovery rate: A practical and powerful approach to multiple testing". Journal of the Royal Statistical Society, Series B 57(1):289-300, 1995.

3. Y. Benjamini and M. Leshno: Statistical Methods for Data Mining. In O. Maimon and L. Rokach (eds.) The Data Mining and Knowledge Discovery Handbook, pages 565-87. Springer, 2005.

4. E. Edgington and P. Onghena: Randomization Tests. 4th edit. Statistics: A Series of Textbooks and Monographs. Chapman & Hall/CRC, 2007.

5. A. Gionis, H. Mannila, T. Mielikäinen and P. Tsaparas: Assessing data mining results via swap randomization. ACM Transactions on Knowledge Discovery from Data 1(3):14:1–14:32, 2007.

6. S. Hanhijärvi: Multiple Hypothesis Testing in Pattern Discovery. In Discovery Science, Proceedings of the 14th International Conference, Lecture Notes in Computer Science vol. 6926, pages 122-134. Springer, 2011.

7. S. Hanhijärvi, M. Ojala, N. Vuokko, K. Puolamäki, N. Tatti and H. Mannila: Tell me something I dont know: Randomization strategies for iterative data mining. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 379-388. ACM, 2009.

8. R. Hubbard and M.J. Bayarri: P Values are not Error Probabilities. Discussion papers series 2003-26, Department of Statistical Science, University Duke University, 2003.

9. W. Hämäläinen: Efficient search for statistically significant dependency rules in binary data. (PhD thesis) Department of Computer Science, University of Helsinki, Finland, 2010. `http://www.cs.uef.fi/~whamalai/research.html`

10. W. Hämäläinen: Kingfisher: an efficient algorithm for searching for both positive and negative dependency rules with statistical significance measures. Knowledge and Information Systems: An International Journal (KAIS) 32(2):383-414, 2012.

11. S. Lallich, O. Teytaud and E. Prudhomme: Association Rule Interestingness: Measure and Statistical Validation. In F. Guillet and H.J. Hamilton (eds.) Quality Measures in Data Mining. Studies in Computational Intelligence volume 43. Pages 251-275. Springer, 2007.

12. S. Lallich, B. Vaillant and P. Lenca: Parametrised measures for the evaluation of association rule interestingness. In Proceedings of the 11th Symposium on Applied Stochastic Models and Data Analysis (AS-MDA'05), pages 220-229. 2005.

13. E.L. Lehmann: The Fisher, Neyman–Pearson Theories of Testing Hypotheses: One Theory or Two? Journal of the American Statistical Association 88(424): 1242-1249, 1993.

14. E.L. Lehmann and J.P. Romano: Testing Statistical Hypothesis. 3rd edit. Springer, 2008.

15. J. Lijffijt, P. Papapetrou and K. Puolamäki: A statistical significance testing approach to mining the most informative set of patterns. Data Mining and Knowledge Discovery. In press.

16. R.J. Little: Calibrated Bayes: A Bayes/Frequentist Roadmap. The American Statistician 60:213-223, 2006.

17. N. Megiddo and R. Srikant: Discovering predictive association rules. In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD), pages 274-278. AAAI Press, 1998.

18. P. Novak, N. Lavrac and G.I. Webb: Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. Journal of Machine Learning Research 10: 377-403, 2009.

19. G.I. Webb: Discovering Significant Patterns. Machine Learning 68(1):1-33, 2007.

20. G.I. Webb: Layered critical values: a powerful direct-adjustment approach to discovering significant patterns. Machine Learning 71(2-3):307-323, 2008.

21. G.I. Webb: Self-Sufficient Itemsets: An Approach to Screening Potentially Interesting Associations Between Items. Transactions on Knowledge Discovery from Data 4: 3:1-3:20, 2010.

22. G.I. Webb: Magnum Opus version 4.6.3 (Computer Software). `http://www.giwebb.com`, 2010.

23. G.I. Webb: Filtered-top-k Association Discovery. WIREs Data Mining and Knowledge Discovery 1(3):183-192. Wiley, 2011.

24. G.I. Webb and J. Vreeken: Efficient Discovery of the Most Interesting Associations. Transactions on Knowledge Discovery from Data 8(3): 15:1-15:31, 2014.

25. G.I. Webb and S. Zhang: K-Optimal Rule Discovery. Data Mining and Knowledge Discovery 10(1):39–79, 2005.

26. P.H. Westfall and S.S. Young: Resampling-based multiple testing: Examples and methods for p-value adjustment, John Wiley & Sons, 1993.