

## References

- [1] Ashburner M, et al. (2000), “Gene ontology: tool for the unification of biology”, *The Gene Ontology Consortium. Nature Genetics*, **25(1)**, 25-29.
- [2] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- [3] Chen, H., Zeng, D.,& Yan, P. (2010). *Infectious disease informatics: Syndromic surveillance for public health and bio-defense*. New York: Springer.
- [4] Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS one*, 5(11), e14118.
- [5] Crowe, M., Rychtar, J., Rueppell, O., Chhetri, M., Remington, D., Gupta, S. (2013a). Proving the Proof- Interdisciplinary Undergraduate Research Positively Impacts Students. Topics from the 8th Annual UNCG Regional Mathematics, and Statistics Conference, *Springer Proceedings in Mathematics & Statistics*, Vol. 64, 25-29
- [6] Crowe, M., Raspet, E., Rychtar, J., Gupta, S. (2013b): Effect of Density and Extra Dung on Brood Parasitism in the Dung Beetle, *Onthophagus Taurus*. *Journal of Insect Behavior*, Vol. 26, No. 2, 253-259
- [7] Dubnicka, S., Blair, R. and Hettmansperger, T. (2002). Rank-based procedures for mixed paired and two-sample designs. *Journal of Modern Applied Statistical Methods*, 1(1), 32-41.
- [8] Eisnpor, R. and Habtzghi, D. (2013). Combing paired and two-sample data using a permutation test. *Journal of Data Science*, 11, 767-779.
- [9] Eysenback, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health infomatics methods to analyze search, communication and publication behavior on the Internet. *Journal of Medical Internet Research*, 11, 11-16. doi:10.2196/jmir.1157
- [10] Gao, X.L. (2015), “Penalized Weighted Low-rank Approximation for Robust Recovery of Recurrent Copy Number Variations”, *BMC Bioinformatics*, **16**:407.
- [11] Gao, X.L. (2016). A Flexible Shrinkage Operator for Fussy Grouped Variable Selection”, *Statistical Papers*. DOI:10.1007/s00362-016-0799-y, 1-24.
- [12] Gao, X.L. and Fang, Y. (2016). Penalized Weighted Least Squares for Outlier Detection and Robust Regression. <https://arxiv.org/abs/1603.07427>
- [13] Gao, X.L. and Feng, Y. (2018), “Penalized Weighted Least Absolute Deviation Regression”, *Statistics and Its Interface*, 1, 79 - 89.
- [14] Gould, L. H., Walsh, K. A., Vieira, A. R., Herman, K., Williams, I. T., Hall, A. J., & Cole, D. (2013). Surveillance for foodborne disease outbreaks—United States, 1998–2008. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 62(2), 1-34.
- [15] Gupta, S. N., Gupta, B. C. and Singh, S. (2002), “Estimation of Sensitivity Level of Personal Interview Survey Questions”, *Journal of Statistical Planning and Inference*, **100**, 239-247.
- [16] Gupta, S., Shabbir, J. and Sehra, S. (2010). Mean and Sensitivity Estimation in Optional Randomized Response Models. *Journal of Statistical Planning and Inference*, Vol. 140, 2870-2874
- [17] Gupta, S., Shabbir, J., Sousa, R., Corte-Real, P. (2012): Estimation of the Mean of a Sensitive Variable in the Presence of Auxiliary Information. *Communications in Statistics - Theory and Methods*, Vol. 41, No. 13-14, 2394-2404
- [18] Gupta, S., Tuck, A., Spears Gill, T., and Crowe, M. (2013): Optional Unrelated-Question Randomized Response Models. *Involve: A Journal of Mathematics*, Vol. 6 (4), 483-492
- [19] Gupta, S., Kalucha, G. and Shabbir, J. (2017): A Regression Estimator for Finite Population Mean of a Sensitive Variable Using an Optional Randomized Response Model. *Communications in Statistics - Simulation and Computation*, Vol. 46, No. 3, 2393 – 2405

- [20] Gupta, S., Erovenko, I., Oh, H., Rychtar, J., Taylor, D. (2017): Undergraduate Students Can do Meaningful Research. *Special Proceedings of 19th Annual Conference of the Society for Statistics and Computer Applications*, pp. 21-29
- [21] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986) “Robust Statistics. The Approach Based on Influence Functions.” New York: John Wiley and Sons.
- [22] Huang, Y., Vander Weele, T.J., and Lin, X. (2014). Joint analysis of snp and gene expression data in genetic association studies of complex diseases. *Ann Appl Stat.*, 8(1): 352–376.
- [23] Johnson, E. (2018). Permutation Tests for Mixed Paired and Two-Sample Designs (Unpublished master’s thesis). University of North Carolina at Greensboro, Greensboro, NC
- [24] Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. “Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks.” *BMC bioinformatics* 15.1 (2014): 276.
- [25] Kanehisa, M. and Goto, S. (2000), “KEGG: Kyoto Encyclopedia of Genes and Genomes”, *Nucleic Acids Research*. **28**, 27-30.
- [26] Lee, S., Wu, M.C., Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*, 13: 762–775.
- [27] Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzell, R. (2015). Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*.
- [28] Ma, Mahoney and Yu (2015). A Statistical Perspective on Algorithmic Leveraging, *Journal of Machine Learning Research*, 16 , 861-9
- [29] Magel, R. and Fu, R. (2014). Proposed nonparametric test for the mixed two-sample design. *Journal of Statistical Theory and Practice*, 8, 221-237.
- [30] Maity, A., Sullivan, P.F., Tzeng, J.Y. (2012). Multivariate phenotype association analysis by markerset kernel machine regression. *Genetic Epidemiology*, 36: 686–695.
- [31] Newkirk, R. W., Bender, J. B., & Hedberg, C. W. (2012). The potential capability of social media as a component of food safety and food terrorism surveillance systems. *Foodborne Pathogens and Disease*, 9, 120-124. doi:10.1089/fpd.2011.0990
- [32] Olanoff, D. (2015). Twitter Monthly Active Users Crawl To 316M, Dorsey: We are not satisfied.
- [33] She, Y. and Owen, A. (2011), “Outlier detection using nonconvex penalize regression”, *Journal of the American Statistical Association* **106(494)**, 626–639.
- [34] Spears-Gill, T., Tuck, A., Gupta, S., Crowe, M., Figueroa, J. (2013): A Field Test of Optional Unrelated Question Randomized Response Models: Estimates of Risky Sexual Behaviors. Topics from the 8th Annual UCG Regional Mathematics, and Statistics Conference, *Springer Proceedings in Mathematics and Statistics Series*, Vol. 64, 135-146
- [35] Suarez, D and Gupta, S. (2018): Variations of the Greenberg Unrelated Question Binary Model. *Involve: A Journal of Mathematics*, Vol. 11, No. 1, 119-126  
DOI: 10.2140/involve.2018.11.119
- [36] Suarez, D and Gupta, S., Johnson, E., and Mathena, P. (2018): Variations of the Greenberg Unrelated Question Model - a Field Test. Initial submission
- [37] Sun, J., Zheng, Y., Hsu, L. (2013) A unified mixed-effects model for rare-variant association in sequencing studies. *Genetic Epidemiology*, 37: 334–344.
- [38] Sun, J., Oualkacha, K., Forgetta, V., Zheng, H., Richards, B., Ciampi A., Greenwood, C., and the UK10K Consortium (2016). A method for analyzing multiple continuous phenotypes in rare variant association studies allowing for flexible correlations in variant effects. *European Journal of Human Genetics*, 24(9), p.1344-1351.
- [39] Sun J., Oualkacha K., Greenwood C., Lakhal-Chaieb L. (2018). Multivariate association test for rare variants controlling for cryptic and family relatedness. Accepted by *Canadian Journal of Statistics*.
- [40] Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso”, *J. Royal. Statist. Soc B.*, **58**, 267-288.

- [41] The UK10K Consortium (2015). The UK10K project identifies rare variants in health and disease. *Nature* 526, 82-90.
- [42] Wang, Yang and Stufken (2019). Information-Based Optimal Subdata Selection for Big Data Linear Regression, *Journal of the American Statistical Association*, Vol. 114, No. 525, Pages 393-405
- [43] Warner, S. L. (1965), “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias”, *Journal of the American Statistical Association*, **60**, 63-69.
- [44] Warner, S. L. (1971), “The Linear Randomized Response Model”, *Journal of the American Statistical Association*, 66, 884-888.
- [45] Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., Lin, X. (2011). Rare variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*, 89: 82–93.
- [46] Young, A., Gupta, S., Parks, R. (2019): A Binary Unrelated Question RRT Model Accounting for Untruthful Responding. To appear in *Involve – a Journal of Mathematics*
- [47] Yuan, M. and Lin, Y. (2006), “Model selection and estimation in regression with grouped variables”, *Journal of the Royal Statistical Society, Series B*, **68**, 49-67
- [48] Zhang, Q., Kazi, H. and Gupta, S. (2015): Modeling Risky Sexual Behaviors among College Students – Predictors of STD. *Collaborative Mathematics and Statistics Research. Springer International Publishing*, 2015, 75-79.