**AAA weARE 14: March 18, 3:30-4:30 p.m.**

**How Useful are *p*-values for Inference?**

**PANELISTS (alphabetical by last name):**

[Sudipta Basu](https://www.fox.temple.edu/about-fox/directory/dr-sudipta-basu/), Temple U. (moderator)

[Bill Cready](https://jindal.utdallas.edu/faculty/william-cready), U. Texas at Dallas

[Sanjay Kallapur](https://www.isb.edu/en/research-thought-leadership/faculty/faculty-directory/sanjay-kallapur.html), Indian School of Business

[Murray Lindsay](https://www.cpacanada.ca/en/career-and-professional-development/event-biographies/r-murray-lindsay), U. Lethbridge

[Ron Wasserstein](https://www.amstat.org/ASA/About/Ronald-L-Wasserstein.aspx), American Statistical Association

**WEBINAR PRE-READINGS**

Wasserstein, R. L., Lazar, N. A., 2016, ASA statement on statistical significance and *p*-values, *The American Statistician* 70 (2), [131-133](https://doi.org/10.1080/00031305.2016.1154108).

Wasserstein, R. L., Schirm, A. L., Lazar, N. A., 2019, Moving to a world beyond *p* < 0.05, *The American Statistician* 73 (Supplement 1), [1-19](https://doi.org/10.1080/00031305.2019.1583913).

**REFERENCES BY PANELISTS AND AUDIENCE DURING SESSION (and similar resources):**

Bernard, V. L., Thomas, J. K., 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13 (4), [305-340](https://doi.org/10.1016/0165-4101(90)90008-R).

Carver, R. P., 1978. The case against statistical significance testing. *Harvard Educational Review* 48 (3), [378-399](https://doi.org/10.17763/haer.48.3.t490261645281841).

Cohen, J., 1994. The earth is round (*p* < .05). *The American Psychologist* 49 (12), [997-1003](https://doi.org/10.1037/0003-066X.49.12.997).

Cox, D. R., 1986. Some general aspects of the theory of statistics*. International Statistical Review* 54 (2), [117-126](https://doi.org/10.2307/1403138).

Cready, W. M., 2019. Complacency at the gates. *Significance* 16 (4), [18-19](https://doi.org/10.1111/j.1740-9713.2019.01297.x).

Cready, W. M., 2020. The “Big N” audit quality kerfuffle. [Working paper](http://dx.doi.org/10.2139/ssrn.3511585).

Cready, W. M., He, J., Lin, W., Shao, C., Wang, D., Zhang, Y., 2021. Is there a confidence interval for that? A critical examination of null outcome reporting in accounting research. [Working paper](http://dx.doi.org/10.2139/ssrn.3131251).

Gigerenzer, G., Krauss, S., Vitouch, 0., 2004. The null ritual: What you always wanted to know about significance testing but were afraid to ask. [Pp. 391-408](https://tuprd-my.sharepoint.com/personal/sudiptab_temple_edu/Documents/Sudipta%20Basu/AAA%20Committees/2021%20P-Values%20Webinar/%5bSearch%20domain%20library.mpib-berlin.mpg.de/ft/gg/GG_Null_2004.pdf%5d%20library.mpib-berlin.mpg.de/ft/gg/GG_Null_2004.pdf) in Kaplan, D. (Ed.) *The Sage Handbook of Quantitative Methodology for the Social Sciences.* Sage, Thousand Oaks, CA.

Hill, A. B., 1956. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58 (5), [295-300](https://doi.org/10.1177/003591576505800503).

Hubbard, R., Lindsay, R. M., 2013. From significant difference to significant sameness: A proposal for a paradigm shift in business research. *Journal of Business Research* 66 (9), [1377-1388](https://doi.org/10.1016/j.jbusres.2012.05.002).

Kallapur, S., 2021. Scientific inference in accounting research: Beyond *p* < 0.05, *Studies in Accounting Research* (Work-in-progress), American Accounting Association, Sarasota FL.

Leek, J. T., Jager, L. R., 2014. Is most published research really false? *Annual Review of Statistics and its Application* 4, [109-122](https://doi.org/10.1146/annurev-statistics-060116-054104).

Lindley, D.V., 1957. A statistical paradox. *Biometrika* 44 (1-2), [187-192](https://doi.org/10.1093/biomet/44.1-2.187).

Lindsay, R. M., Ehrenberg, A. S. C., 1993. The design of replicated studies. *The American Statistician* 47 (3), [217-228](https://www.jstor.org/stable/2684982).

Meehl, P. E., 1967. Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science* 34 (2), [103-115](https://www.jstor.org/stable/186099).

Rozeboom, W. W., 1960. The fallacy of the null hypothesis significance test. *Psychological Bulletin* 57 (5), [416-428](https://doi.org/10.1037/h0042040).

Shafer, G., 2020. On the nineteenth century origins of significance testing and p-hacking. *The Game-theoretic Probability and Finance Project*. [Working paper 55](http://www.probabilityandfinance.com/articles/55.pdf).

Tong, C., 2019. Statistical inference enables bad science; statistical thinking enables good science, *The American Statistician* 73 (Supplement 1), [246-261](https://doi.org/10.1080/00031305.2018.1518264).

Wilkinson, L., and the Task Force on Statistical Inference, 1999. Statistical methods in psychological journals: Guidelines and explanations. *American Psychologist* 54 (8), [594-604](https://doi.org/10.1037/0003-066X.54.8.594).

Ziliak, S. T., McCloskey, D. N., 2004. Size matters: the standard error of regressions in The American Economic Review. *Journal of Socio-Economics* 33 (5), [527-546](https://doi.org/10.1016/j.socec.2004.09.024).

Ziliak, S. T., McCloskey, D. N., 2008. *The Cult of Statistical Significance: How the Standard Error costs us jobs, justice, and lives*. [University of Michigan Press](https://www.press.umich.edu/186351), Ann Arbor, MI.

**Polling Questions for 3/18 webinar**

Poll questions adapted from [Gigerenzer et al. (2004, pp. 2-3)](https://tuprd-my.sharepoint.com/personal/sudiptab_temple_edu/Documents/Sudipta%20Basu/AAA%20Committees/2021%20P-Values%20Webinar/%5bSearch%20domain%20library.mpib-berlin.mpg.de/ft/gg/GG_Null_2004.pdf%5d%20library.mpib-berlin.mpg.de/ft/gg/GG_Null_2004.pdf)

**Polling question 1**

Suppose you have a treatment that you suspect may alter performance on a certain task. You compare the means of your control and experimental groups (say, 20 subjects each), using a simple independent means t-test and your result is significant (t = 2.7, df = 18, p = .01). Please mark each statement below as “true” or “false.” False means that the statement does not follow logically from the above premises.

(1) You have absolutely disproved the null hypothesis (i.e., there is no difference between the population means). ® True False ®

(2) You have found the probability of the null hypothesis being true. ® True False ®

(3) You have absolutely proved your experimental hypothesis (that there is a difference between the population means). ® True False ®

**Polling question 2**

Suppose you have a treatment that you suspect may alter performance on a certain task. You compare the means of your control and experimental groups (say, 20 subjects each), using a simple independent means t-test and your result is significant (t = 2.7, df = 18, p = .01). Please mark each statement below as “true” or “false.” False means that the statement does not follow logically from the above premises.

(4) You can deduce the probability of the experimental hypothesis being true. ® True False ®

(5) You know, if you decide to reject the null hypothesis, the probability that you are making the wrong decision. ® True False ®

(6) You have a reliable experimental finding in the sense that if, hypothetically, the experiment were repeated a great number of times, you would obtain a significant result on 99% of occasions. ® True False ®