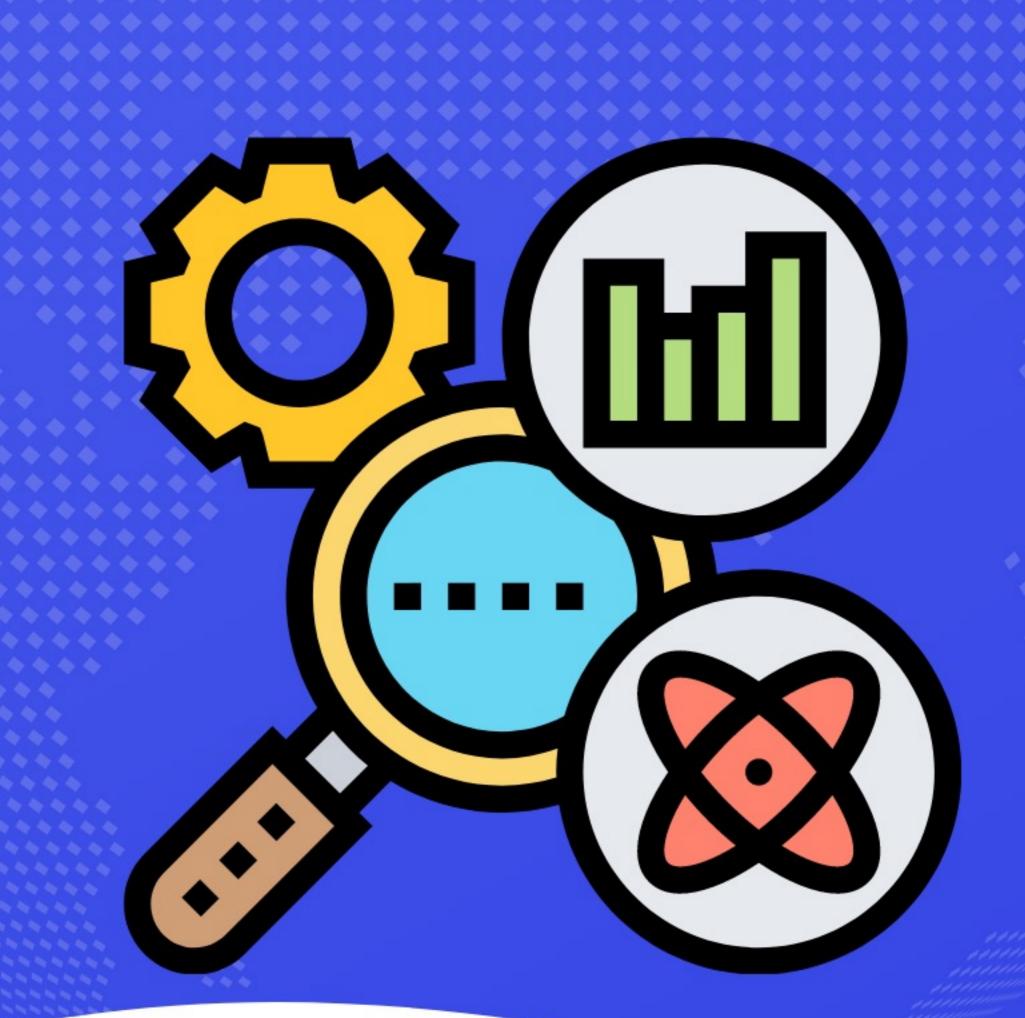
# P-value Ponderings: Beyond the Mythical Threshold

#### P-value Definition

The **p-value** is the probability of obtaining a result as or more extreme than the observed test statistic when the null hypothesis is true [1].





## What the p-value is and isn't

The p-value summarizes the **statistical compatibility** between the observed statistic and what we would expect if the null hypothesis is true [2]. The smaller the p-value, the more unusual the data would be. The p-value is **NOT**:

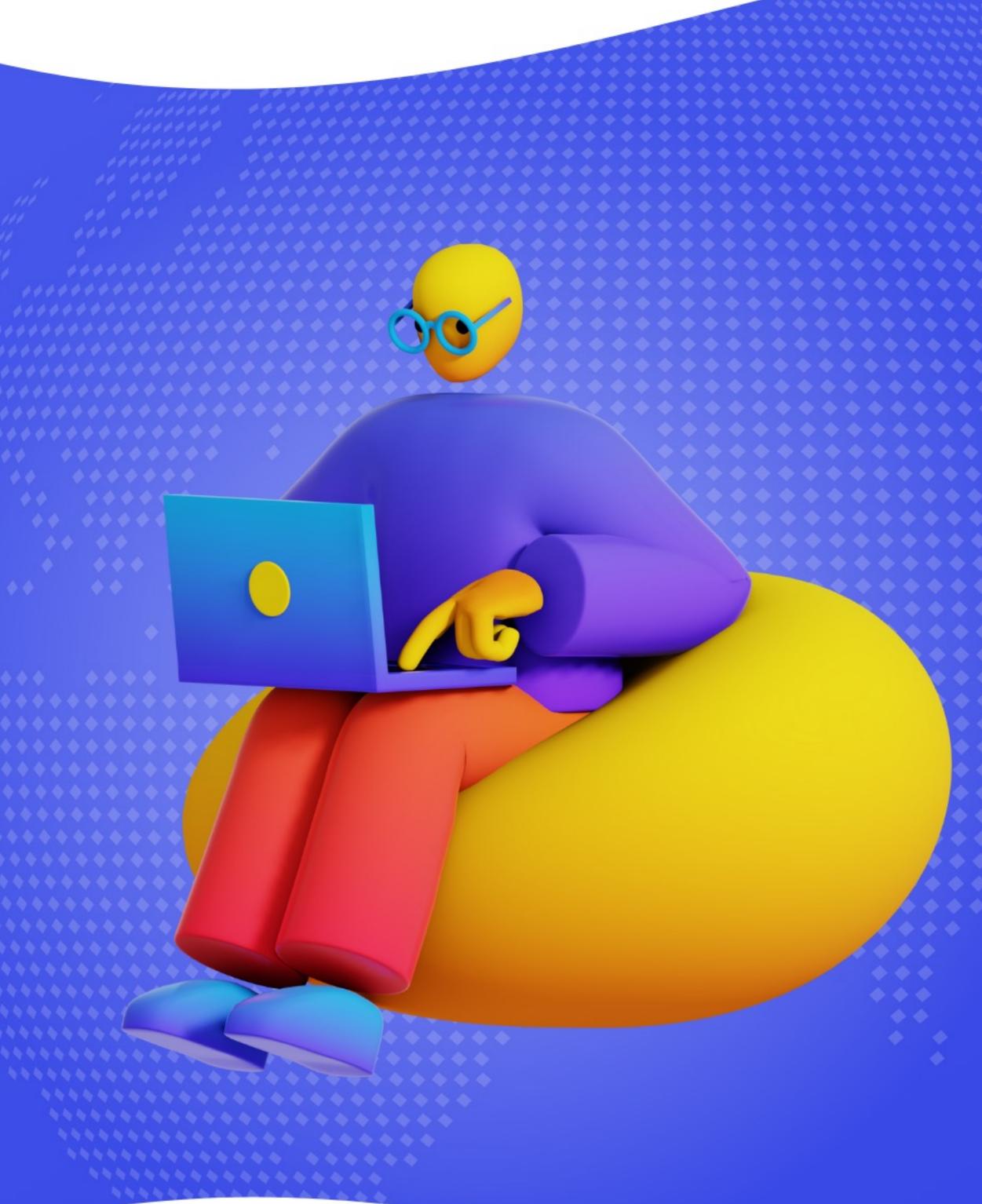
- The probability that the null hypothesis is true.
- The probability that the observed effect is produced by chance alone.
- The chance that the observed effect occurs if the null hypothesis is true.

## Criticisms of p-values

**Overstating the practical importance**: p-values alone should not dictate the *scientific* significance of an effect [3].

"All models are wrong": p-values originate from flawed models (including all assumptions and hypotheses). Small p-values signal imperfections in the model but do not tell us which assumptions are wrong [4].

**Robustness:** various decisions made during statistical analysis can dramatically influence p-values. Influential factors include sample size [5], correlated observations [6], choice of covariates and models [7], etc.

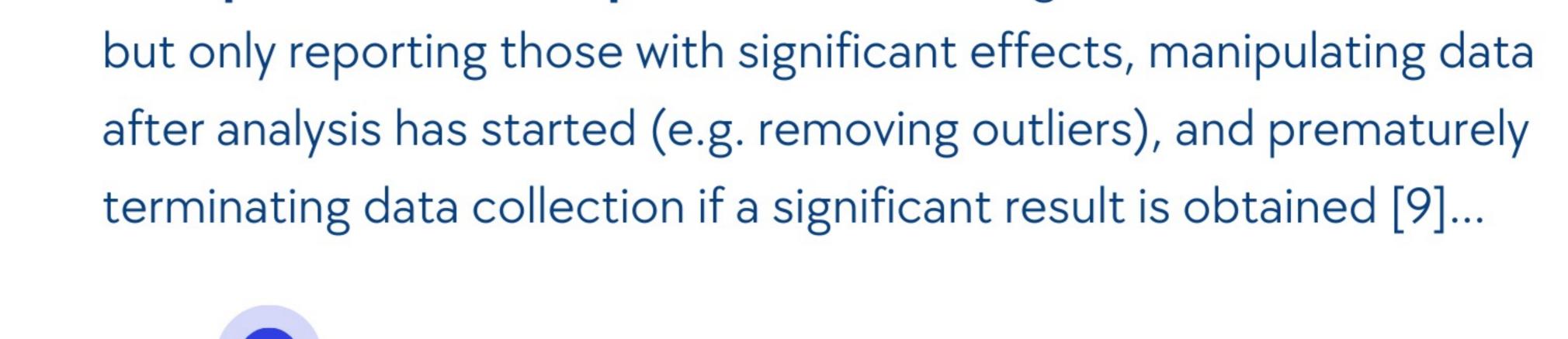


## Ethics of using p-values



researchers cherry-pick data or statistical analysis to make non-significant results significant [8].

**Definition:** P-hacking occurs when



Examples of unethical practices: collecting numerous variables



## Why is it important? Publication bias is based on the false

belief that achieving a statistical significance level shows a result is "real," hence some journals tend to publish mostly statistically significant results [3].



#### result in the publication of false positive findings,

- hinder the dissemination of accurate and honest results,
- adversely affect future research and systematic reviews
  - in the field and human knowledge as a whole [10].



• Researchers should report test statistics, degrees of freedom, effect size, and confidence intervals [6].

• Instead of binary inequalities (p < 0.05), report precise p-values [3].

- Bonferroni correction, Scheffé's technique, false discovery rate and erose validation can be used to check or provent p backing.
- cross-validation can be used to check or prevent p-hacking.

   P-values should not determine publication or interpretation [6].
- Scientific and statistical inferences should be made from results [6].

# 1. Greenland, S. (2019). Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their

- References
  - resolution with s-values. The American Statistician, 73(sup1), 106–114.

    2. Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European journal of epidemiology, 31(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3
  - 567(7748), 305–307. https://doi.org/10.1038/d41586-019-00857-9
    4. Trafimow, D., Haley, U., & Boje, D. (2022). Best way not to misuse P values is not to draw definitive conclusions

3. Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. Nature,

- about hypotheses. *BMJ Evidence-Based Medicine*. https://doi.org/10.1136/bmjebm-2022-111940 5. Sullivan, G. M., & Feinn, R. (2012). Using Effect Size-or Why the P Value Is Not Enough. *Journal of graduate medical education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1
- 6. Lazic, S. E. (2010). The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis?. *BMC* neuroscience, 11, 1-17.
- 7. Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632
- 8. Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of phacking in science. *PLoS biology*, *13*(3), e1002106.
- 9. Karpen, S. C. (2017). P value problems. *American journal of pharmaceutical education, 81*(9). 10. Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A. W., Cronin, E., ... & Williamson, P. R. (2008). Systematic

review of the empirical evidence of study publication bias and outcome reporting bias. PloS one, 3(8), e3081.

University of Toronto, 2023