Name: Myisha Hassan
Date: 9/25/25
Paper Title: Of beauty, sex, and power: Statistical challenges in estimating small effects
Author Names: Andrew Gelman and David Weakliem
Year Published: 2008

Open questions for discussion in class:

- How should we interpret statistically insignificant but theoretically motivated results?
- Are small-sample studies in social science just simply unreliable?
- Can Bayesian priors help address the problem of small effect sizes in practice?
- How can researchers practically apply Bayesian inference with informative priors when there isn't a strong body of previous literature to determine a plausible effect size?

The topic areas covered by the paper are:

- Concepts of Type M (magnitude) and Type S (sign) statistical errors
- The challenges in estimating small effect in studies and interpreting results in the "gray area of significance"
- Statistical interpretation of small effects and its challenges
- Sex ratio research and evolutionary psychology: critique of how statistically weak findings particularly in evolutionary psychology are published and amplified by popular media.
- Limitations of classical inference and power analysis
- Using Bayesian inference for weak effects

The previous approaches to this problem were:

- Using thresholds like a 5% level of statistical significance to determine whether a result is worthy of being published.
- Focusing mainly on Type 1 (false positive) and Type 2 (false negative) errors since these don't fully address the issue of overestimated effect size
- Interpreting statistically significant results from small samples as strong evidence of a real and important effect without considering that the estimate's magnitude might be inflated.

Outline the basic new approach or approaches to this problem:

- Reliance on p-values and statistical significance in small-sample studies
- Classical inference without incorporating prior expectations
- Media and peer review emphasis on "significant" results, even if unreliable

Critical assumptions made include:

- Based on a large body of previous research, any true effect of a factor like parental beauty on the human sex ratio is expected to be very small (on the order of 1% point)
- The current system of scientific publication and media reporting is biased toward overstating the magnitude and certainty of findings from small studies,

particularly if the results are startling or novel.

The performance of the techniques discussed in the paper was measured in what manner:

- comparison of confidence intervals vs. expected effect sizes
- Simulation of error rates (Type 1, Type 2, Type M, Type S) under various scenarios to show even a significant result would be unreliable .
- Bayesian posterior probabilities with different prior distributions
- Reanalyzing real-world datasets with a standard regression to show that the initial claim was not statistically significant

What background techniques are used in the paper that you are not familiar with:

- Diffuse prior distribution
- Bayesian prior specification with Cauchy distributions
- Type M (magnitude) and Type S (sign) errors
- Power analysis / power calculations
- Trivers-Willard hypothesis
- Odds ratio

The following terms were defined:

- Generalized Trivers-Willard hypothesis: theory that parents' ability to vary the sex ratio of their offspring evolved to increase the expected number of descendants.
- Type 1 error: statistically significant result when there is no real effect.
- Type 2 error: Failing to get a statistically significant result when there is a real effect.
- Type M (magnitude) error: An error where a study produces a statistically significant estimate of an effect that is much larger in magnitude than the true effect.
- Type S (sign) error: An error where a study produces a statistically significant estimate that has the opposite sign (direction) of the true effect

I rate and justify the value of this paper as:

9/10. The paper shows a flaw in how science has been practiced and reported so far and uses interesting real world examples to show how studies with small sample sizes can produce results that are both statistically significant and wildly misleading. The authors also offer solution frameworks by bringing up Type M and Type S errors to show how a finding can be exaggerated or even in the wrong direction. They also connect with Bayesian priors which anchor estimates to realistic expectations and provide more credible interpretations. This makes the paper especially valuable in this class to me because it shows early on why priors matter and how they prevent overinterpretation of noisy data.