p–values Had a Good Run: A Primer on the 'New Statistics'

Rob Cribbie
Quantitative Methods Program
Department of Psychology
York University

What’s the plan …

- We are going to scratch the surface in terms of modern issues related to data analysis in the behavioral sciences
- There will less of a focus on the actual methods used to conduct the analyses (i.e., the “how”, and more of a focus on the “why”, although we will also dabble in some applied work

Day One

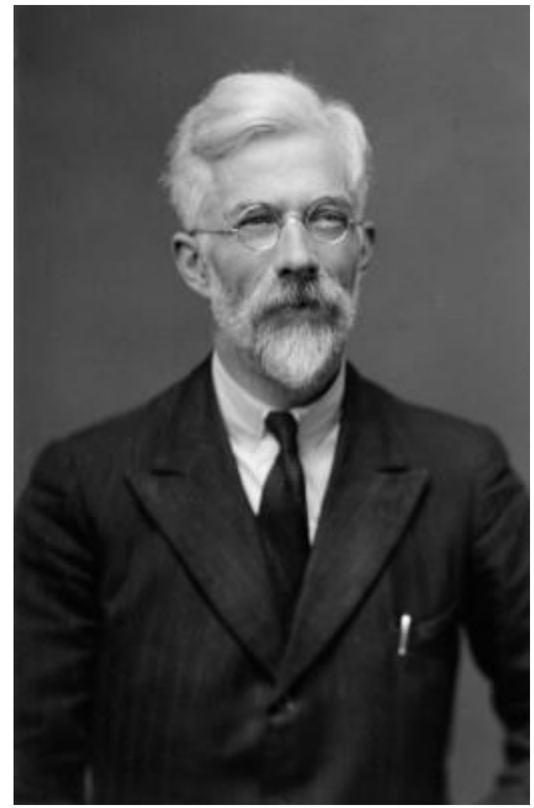
- Part 1: History of Null Hypothesis Significance Testing
- Part 2: Problems with Null Hypothesis Significance Testing
- Part 3: Multiplicity Issue and Null Hypothesis Significance Testing

Part 1: History of NHST

- Modern Null Hypothesis Significance Testing can be traced back to:
  - Fisher's Significance Testing
  - Neyman–Pearson Hypothesis Testing

Fisher's Significance Testing

- Ronald Aylmer Fisher was a biologist and statistician
- He was the main force behind tests of significance and can be considered the most influential figure in modern data analytic techniques



Main Goal of Fisherian Testing

- The primary motivation behind Fisher's approach to significance testing was to find the probability of the data, given the null hypothesis

- Highlights
  - There is no alternative hypothesis
  - Power is of no interest
  - There is no alpha (α) level (a priori Type I error rate)

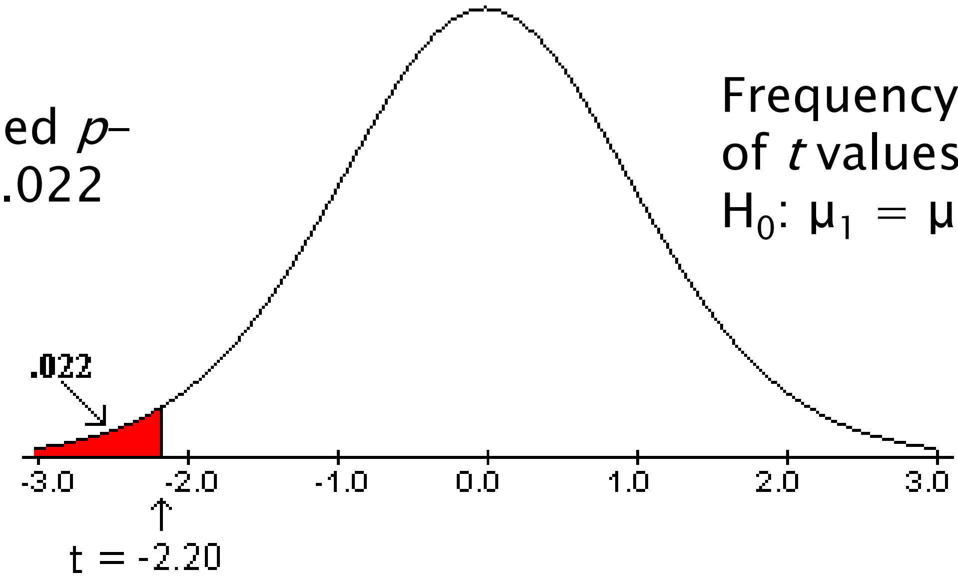
Steps Involved in Fisher's Approach to Significance Testing

- Let's use as an example comparing two independent populations
  - Step 1: Select an appropriate test
    - Independent Samples *t*–test
  - Step 2: State H_0
    - H_0: $\mu_1 = \mu_2$
      - Could also be a directional hypothesis
        - E.g., H_0: $\mu_1 \geq \mu_2$
      - Could also test differences other than 0 (nil hypothesis)
        - E.g., H_0: $\mu_1 - \mu_2 = 5$ or H_0: $\mu_1 - \mu_2 \leq 5$

Steps Involved in Fisher's Approach to Significance Testing

- Step 3: Calculate the *p*–value, assuming H_0 is true
  - *p*–value: probability of finding a test statistic more extreme than that found, assuming H_0 is true

One–tailed *p*–value = .022



Frequency distribution of *t* values under H_0: $\mu_1 = \mu_2$ for df = 15

Steps Involved in Fisher's Approach to Significance Testing

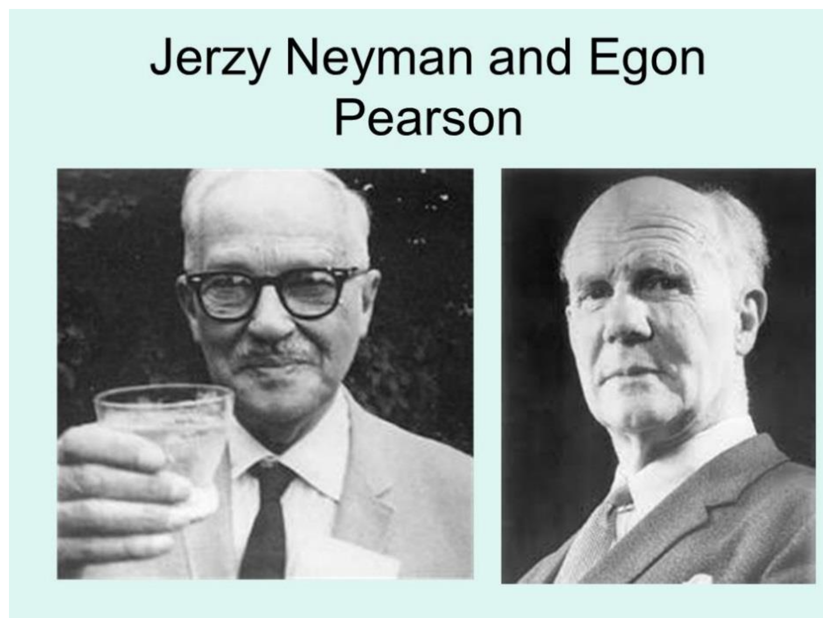
- Step 4: Statistical Decision
  - Is the *p*–value small enough to conclude that the results were highly unlikely if H_0 is true?
    - Typically made relative to some cutoff (e.g., .01, .05), however cutoffs need not be specified
      - What's important is that *p*–values of .049 and .051 are very similar probabilistically
    - Exact *p*–values are important since the magnitude of the probability is of utmost importance
    - The *p*–value provides information regarding the plausibility of H_0
      - Smaller *p*–values provide greater evidence against H_0

Steps Involved in Fisher's Approach to Significance Testing

- Step 5: Interpret the Findings
  - If a result is deemed statistically significant, one of two statements is true
    - A rare mistake has occurred
    - H_0 does not accurately represent the true state of affairs
  - Non-significant results provide useful information, such as whether results were in the expected direction and the magnitude of the effects
    - Non-significant results can even provide information that can be used to strengthen support for H_0

Neyman and Pearson's Hypothesis Testing Approach

- Jerzy Neyman and Egon Pearson sought to improve Fisher's approach to statistical significance testing
- Their approach greatly expanded on the principles and procedures outlined by Fisher



Steps Involved in Neyman and Pearson's Hypothesis Testing

- Step 1: State the Research Hypothesis
  - State what result is expected, including the smallest meaningful effect size (MES)
    - This is used to establish appropriate hypotheses or conduct power analyses
- Step 2: Select an Appropriate Test Statistic
  - Note that since "power" is a concept in Neyman and Pearson, tests can be based on differences in power (e.g., parametric vs nonparametric)

Steps Involved in Neyman and Pearson's Hypothesis Testing

- Step 3: State the Null Hypothesis
  - Similar in nature to Fisher's H_0 (e.g., H_0: $\mu_1 = \mu_2$)
  - Power analyses based on MES should be conducted, such that the null includes inconsequential effects
    - In other words, important effects should be found with a high probability
  - A new concept is the idea of an α level
    - Under Neyman-Pearson only a single α level is chosen, where Fisher was more flexible (concern was the magnitude of *p*)
  - Also central to the Neyman-Pearson approach is the minimization of the risk of Type I errors (rejecting H_0 when it is true)

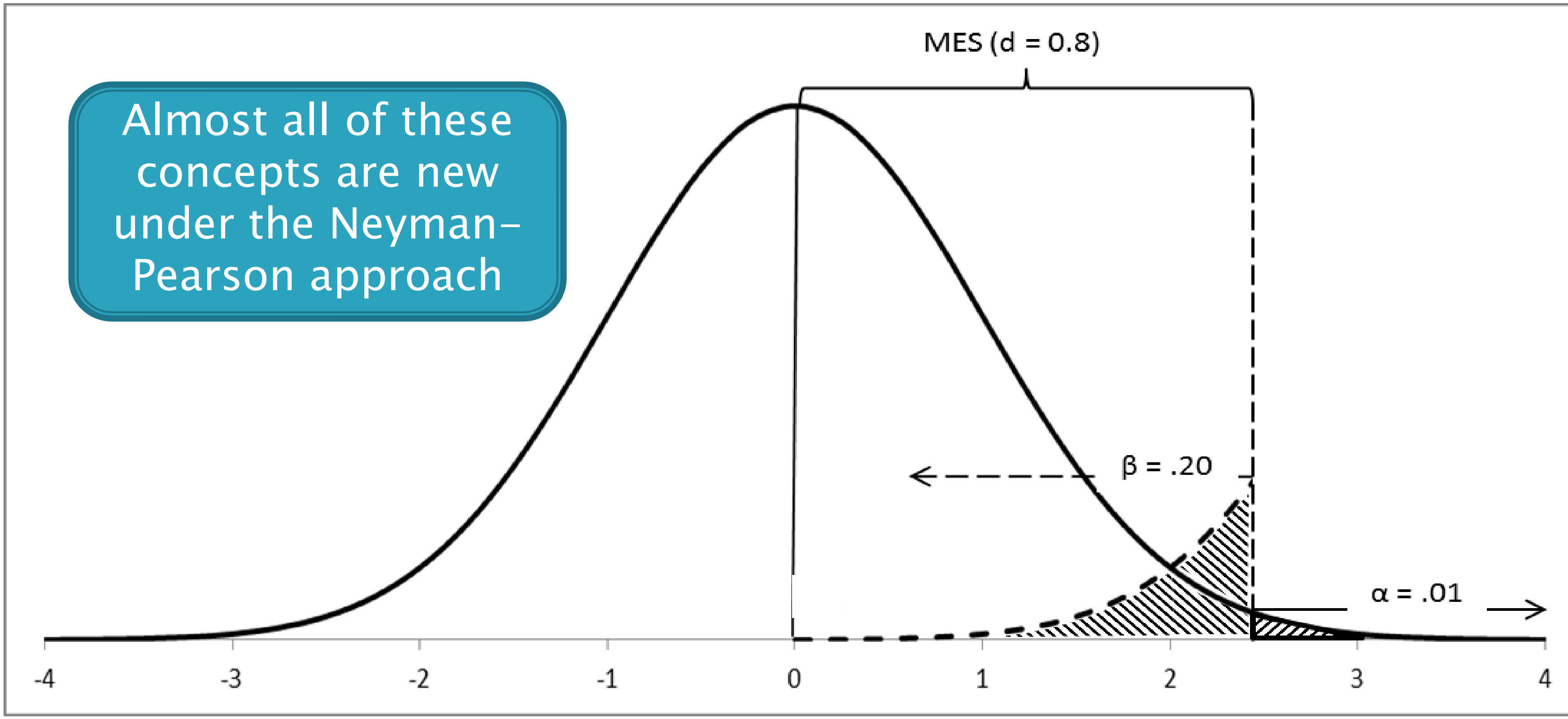
Steps Involved in Neyman and Pearson's Hypothesis Testing

- Step 4: State the Alternate Hypothesis (H_A)
  - The concept of H_A is novel under the Neyman–Pearson approach
    - H_a: $\mu_1 \neq \mu_2$
    - The presence of H_A permits power analyses and introduces the concept of a Type II error (β, not rejecting H_0 when it is false)
      - Neyman and Pearson proposed 20% ($\beta = .20$) as an upper ceiling for β, and the value of alpha ($\beta = \alpha$) as its lower floor

Neyman–Pearson Hypothesis Testing

		THE TRUTH	
		The null hypothesis (H_0) is true (H_a is false)	**The null hypothesis (H_0) is not true** (H_a is true)
THE DECISION THE ANALYST MAKES	**Reject H_0** (support H_a)	**TYPE I (α) error/ Alpha Risk/ p – value** **Overreacting** **$(1 - \alpha)$ = the Confidence level of the test**	**Correct Decision $(1 - \beta)$** **Power of the test**
	Fail to Reject H_0 (do not support Ha)	**Correct Decision**	**TYPE II (β) error/ Beta Risk** **Underreacting**

Smallest Meaningful Effect Size, Power and the Null Hypothesis



Ho

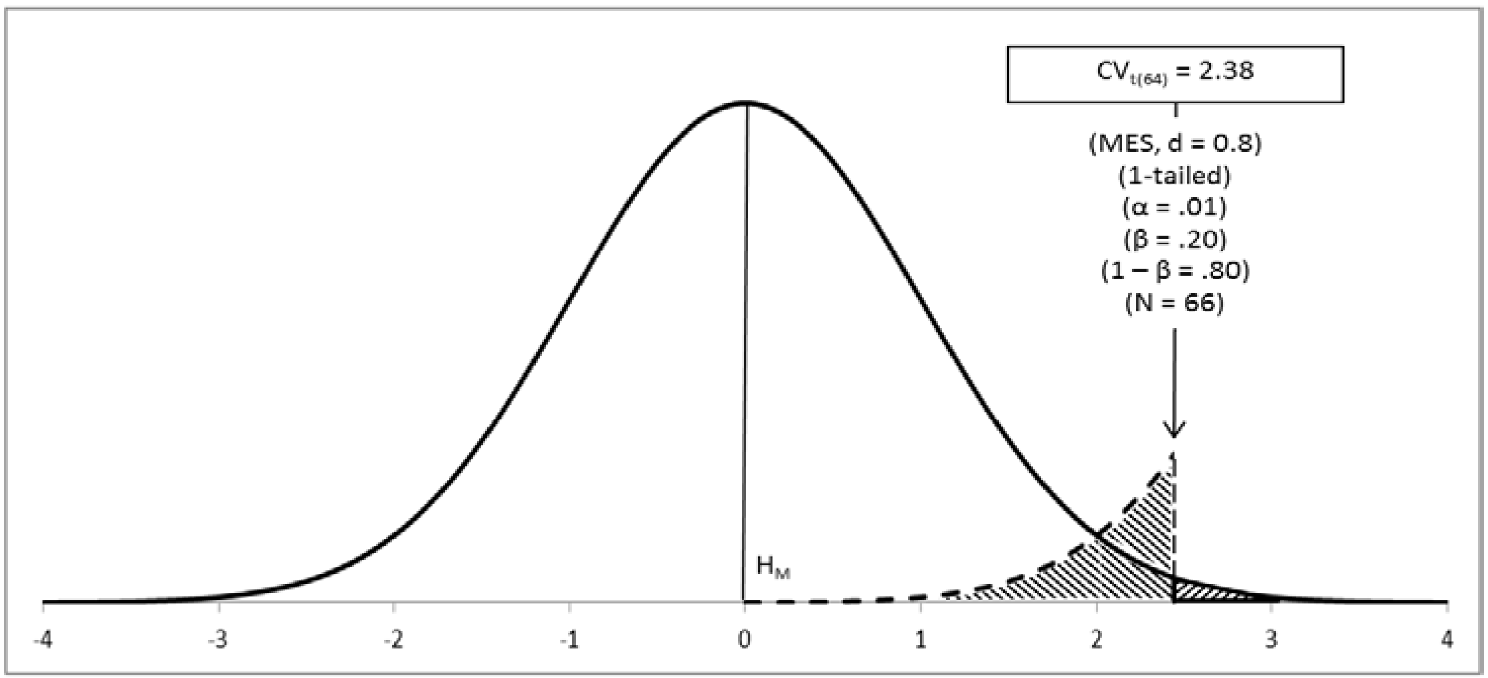
Steps Involved in Neyman and Pearson's Hypothesis Testing

- Step 5: Conduct a Power Analysis
  - What sample size is required to ensure that $\beta < .20$ $(1 - \beta = .80)$?
    - There is no reason to conduct a low-power study (i.e., $1 - \beta < .80$)
    - β should fall between α and .20
      - *If it is desired to have β less than α, than the hypotheses should be reversed* (N&P)

    - Controlling for errors in the long run is very important!

Steps Involved in Neyman and Pearson's Hypothesis Testing

- Step 6: Determine the Critical Value for the Test Statistic



Steps Involved in Neyman and Pearson's Hypothesis Testing

- Step 7: Compare the test statistic to the critical value or the *p*–value to α
- Step 8: Make a decision regarding H_0/H_a
  - Reject or retain H_0
  - Unlike Fisher, the hypothesis decision is most important, not the magnitude of the *p*–value
- To summarize, the Neyman–Pearson approach emphasizes a priori decisions, including MES, error rates, power/sample size, etc., and focuses more on decisions regarding hypotheses than the magnitude of *p*–values

Modern NHST

- Modern null hypothesis significance testing borrows from both Fisher and Neyman–Pearson
  - Procedurally, most researchers follow Neyman–Pearson
  - Philosophically, however, many researchers are more in favour of Fisher's approach in terms of evaluating evidence against H_0 through quantifying the magnitude of the *p*–value

Discussion Point

- If you had to choose one of the methods as the primary method for your field, which would it be?