Part II

Second-generation p-values: equivalence tests, statistical properties, and false discovery rates

Jeffrey D. Blume, PhD
School of Data Science
University of Virginia

and

Megan H. Murray, almost PhD (July 6th defense)
Department of Biostatistics
Vanderbilt University

Course Layout

- Slides Part I
  - Introduction and applications
- Coding Part I
- Slides Part II
  - Equivalence tests, statistical properties, and false discovery rates
- Coding Part II
- Questions and Discussion

Outline

- → Equivalence Tests
  - → Two One-Sided Tests (TOST)
- → Statistical Properties
  - → $P(p_\delta = 0|H)$, $P(p_\delta = 1|H)$, and $P(0 < p_\delta < 1|H)$
- → False Discovery Rates
  - → R Packages
- → Code Part 2
- → Review of Topics Learned

ASA SGPV Short Course Blume and Murray, 2022

Equivalence Tests

- Establish bioequivalence between data and an established range
- Example: A pharmaceutical company tests for drug approval by comparing new drug's performance to an approved drug's performance
- Uses an interval null or equivalence range
  - $H_0 = [\theta^-, \theta^+]$
- Most popular: TOST, Bayesian ROPE, test of Anderson and Hauck, and SGPV

ASA SGPV Short Course Blume and Murray, 2022

TOST Test

- Two One-Sided Tests (Schuirmann 1987)
- Tests are ordinary, one-sided, α-level t-tests
- Flips the null and alternative hypotheses, and tests if the data $I_x = (I_x^-, I_x^+)$ are outside the equivalence range $[\theta^-, \theta^+]$
- $(H_{0_1}: \theta < \theta^-)$ and $(H_{0_2}: \theta > \theta^+)$
- If *both* one-sided tests reject then conclude the evidence is contained in the equivalence range
- $p_T = max\{p_{T_1}, p_{T_2}\}$

ASA SGPV Short Course | Blume and Murray, 2022

SGPV Definition

Second-generation *p*-value (SGPV)

$$p_\delta = \frac{|I \cap H_0|}{|I|} \times \max\left\{\frac{|I|}{2|H_0|}, 1\right\}$$

Proportion of data-supported hypotheses that are also null hypotheses

Small-sample correction factor

shrinks proportion to ½ when $|I|$ wide

when $|I| > 2|H_0|$

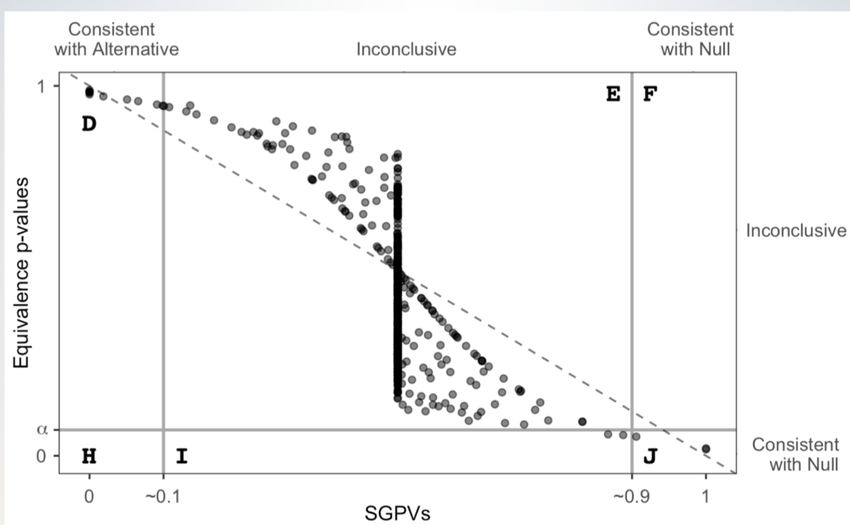
ASA SGPV Short Course | Blume and Murray, 2022

TOST vs. SGPV comparison

		SGPV Outcomes		
		Consistent with the alternative (SGPV near 0)	**Inconclusive** (SGPV near ½)	**Consistent with the null** (SGPV near 1)
Equivalence Tests Outcomes	**Consistent with the alternative** (*p*-value is unable to indicate this)	Not applicable **A**	Not applicable **B**	Not applicable **C**
	Inconclusive (*p*-value is non-significant)	Can occur **D**	Can occur **E**	Never occurs **F**
	Consistent with the null (*p*-value is significant)	Never occurs **H**	Can occur in small samples **I**	Can occur **J**

ASA SGPV Short Course Blume and Murray, 2022

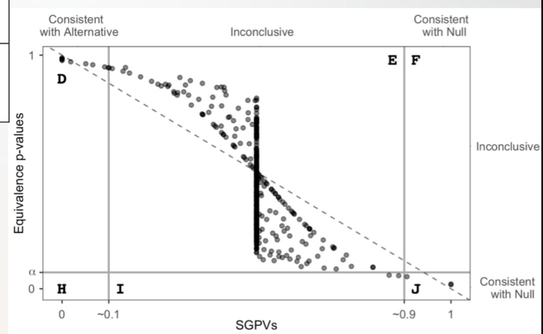
TOST vs. SGPV comparison



ASA SGPV Short Course Blume and Murray, 2022

TOST vs. SGPV comparison

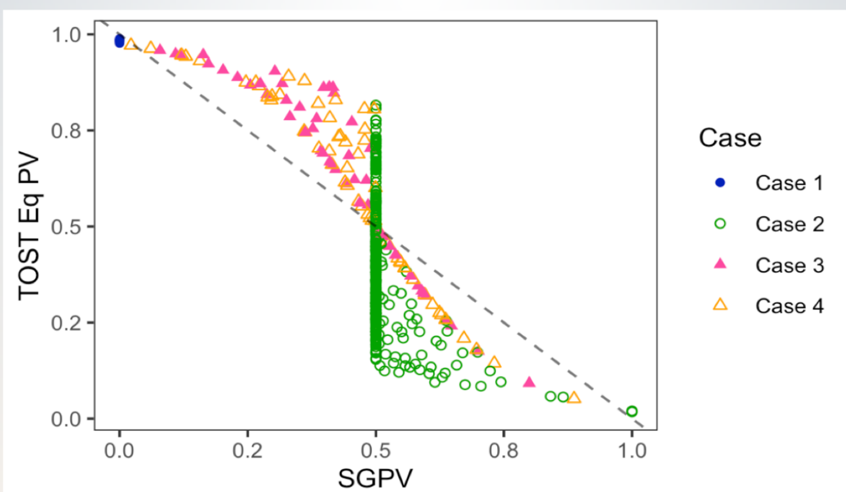
		SGPV Outcomes		
		Consistent with the alternative (SGPV near 0)	**Inconclusive** (SGPV near ½)	**Consistent with the null** (SGPV near 1)
Equivalence Tests Outcomes	**Consistent with the alternative** (*p*-value is unable to indicate this)	Not applicable A	Not applicable B	Not applicable C
	Inconclusive (*p*-value is non-significant)	Can occur D	Can occur E	Never occurs F
	Consistent with the null (*p*-value is significant)	Never occurs H	Can occur in small samples I	Can occur J



ASA SGPV Short Course

Blume and Murray, 2022

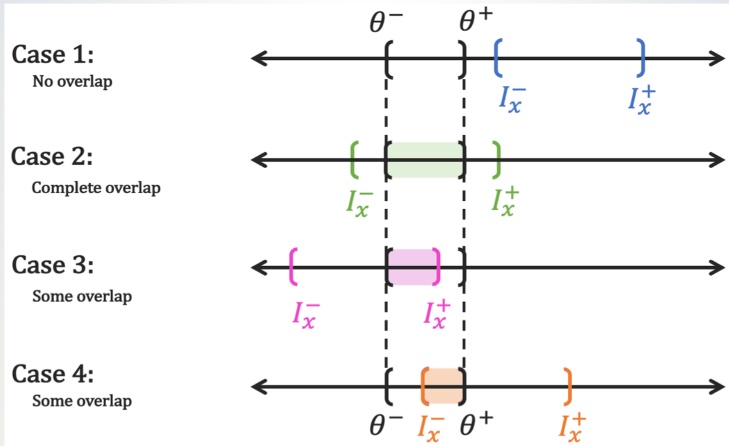
TOST vs. SGPV comparison



ASA SGPV Short Course

Blume and Murray, 2022

Equivalence Tests



ASA SGPV Short Course

Blume and Murray, 2022

TOST vs. SGPV comparison

TOST	SGPV
2 inference outcomes	3 inference outcomes
Conclusions only about $(1-2\alpha)\%$ confidence interval	Any uncertainty data interval can be used
Type I Error is ultra-conservative (distribution of p_T is non-uniform)	Type I error is accurately assessed (limited by width of data interval)
Not uniformly most powerful	Has additional statistical properties explained in next section
No measure of overlap included in computation	Indicates when data agree with null or alternative without additional testing

ASA SGPV Short Course

Blume and Murray, 2022

Second-generation p-value

- Statistical properties in TAS & PLOS One
- Retains strict error control

Evidential Metric	What it measures	SPGV
1	Summary measure	SGPV (p_δ)
2	Operating characteristics	$P(p_\delta = 0 \mid H_0)$ $P(p_\delta = 1 \mid H_1)$ $P(0 < p_\delta < 1 \mid H)$
3	False discovery rates	$P(H_0 \mid p_\delta = 0)$ $P(H_1 \mid p_\delta = 1)$

Statistical Properties

Suppose interval I has coverage probability 1-α, then

Three 'Error' Rates

1. $P(p_\delta = 0|H_0) \leq \alpha$ and $\to 0$ as $n \to \infty$
2. $P(p_\delta = 1|H_1) \leq \alpha$ and $\to 0$ as $n \to \infty$
3. $P(0 < p_\delta < 1|H)$ controlled through sample size

Will examine these first

Two False Discovery Rates

1. $P(H_0 \mid p_\delta = 0)$
2. $P(H_1 \mid p_\delta = 1)$

Will graph to illustrate

Statistical Properties

- Three Inferential Categories
  1. $p_\delta = 0 \Rightarrow$ data incompatible with null
  2. $p_\delta = 1 \Rightarrow$ data compatible with null
  3. $0 < p_\delta < 1 \Rightarrow$ data are inconclusive

- Three 'error' rates
  1. $P(p_\delta = 0|H)$ when H is null
  2. $P(p_\delta = 1|H)$ when H is not null
  3. $P(0 < p_\delta < 1|H)$ when H is either

- Assume H makes statements about a parameter θ
- Large sample setting

ASA SGPV Short Course Blume and Murray, 2022

Statistical Properties

- **How often are the data incompatible with null?**

- Examine $P(p_\delta = 0|\theta)$ as θ varies
  - Power function

- This probability
  - converges to one for alternatives not near the edge of interval null
  - converges to zero for null hypotheses not near the edge of the null set
  - converges to alpha for hypotheses approaching or on the edge of the null set

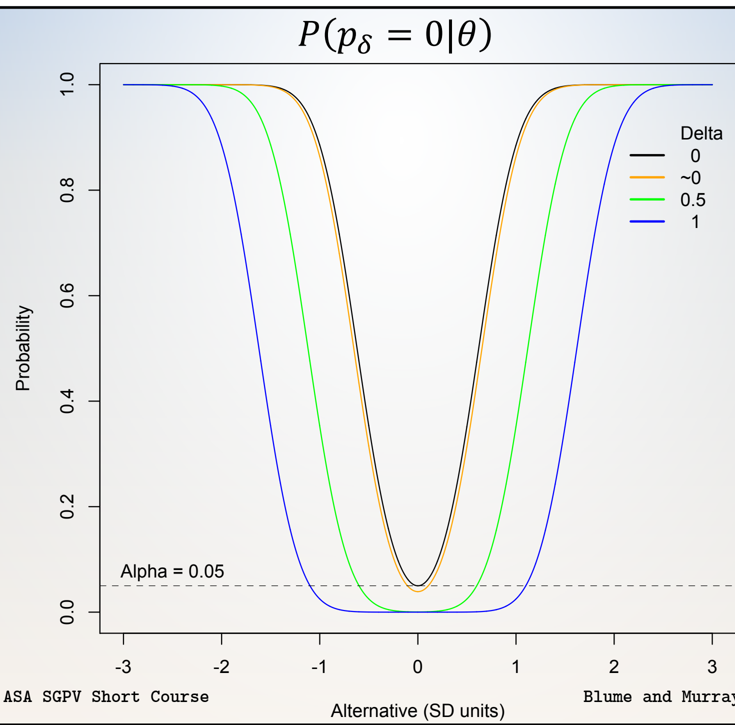
ASA SGPV Short Course Blume and Murray, 2022

'Power' Function

- θ_0: point null, σ: standard deviaiton
- δ: half-width of indifference zone

$$P(p_\delta = 0|\theta) = \Phi\left[\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - \frac{\sqrt{n}\delta}{\sigma} - Z_{\alpha/2}\right] + \Phi\left[-\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - \frac{\sqrt{n}\delta}{\sigma} - Z_{\alpha/2}\right]$$

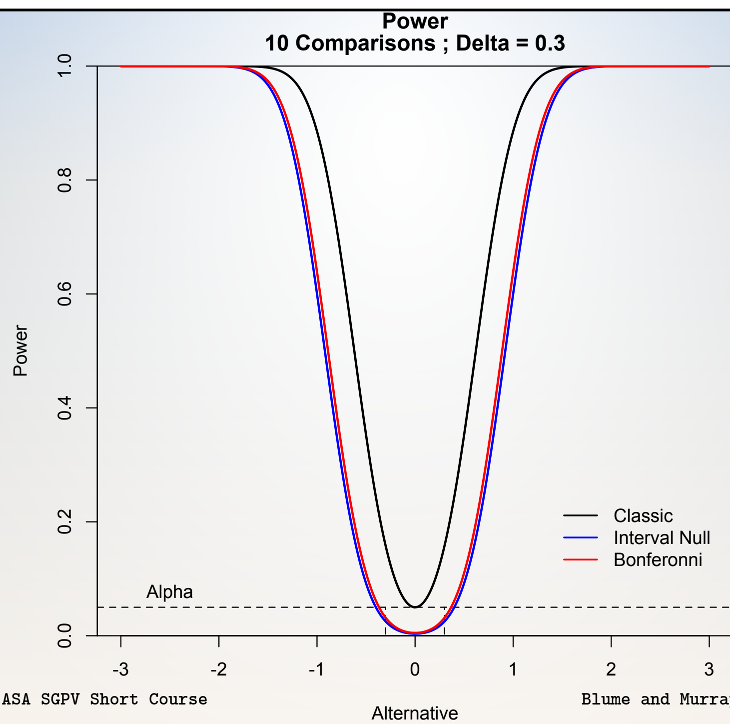
$$P_{\theta_0}(p_\delta = 0|\theta_0) = 2\Phi\left[-\frac{\sqrt{n}\delta}{\sigma} - Z_{\alpha/2}\right]$$

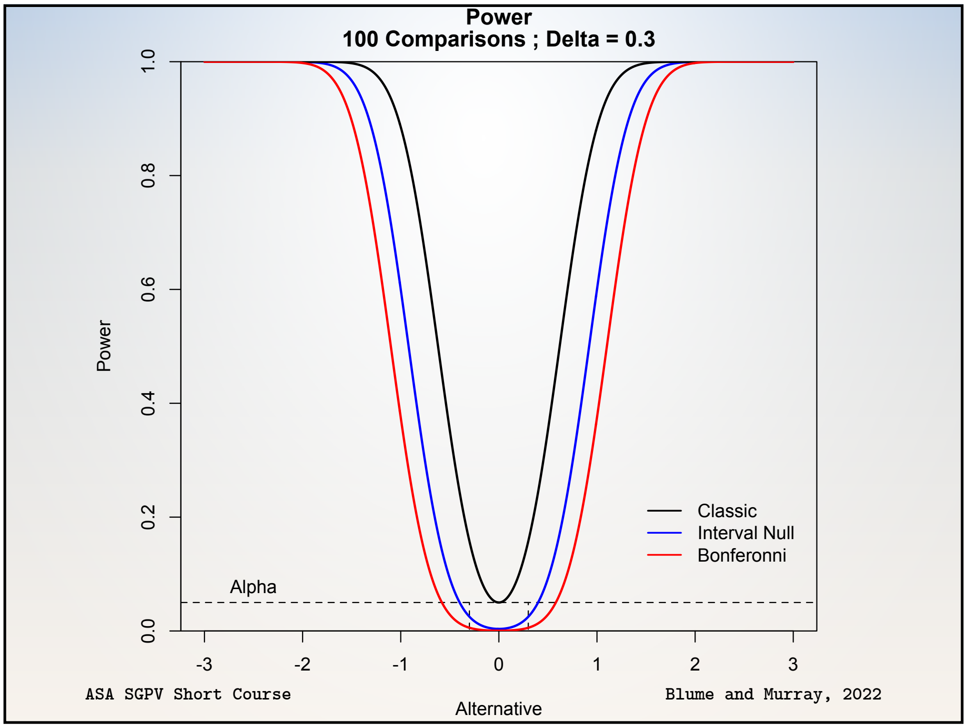


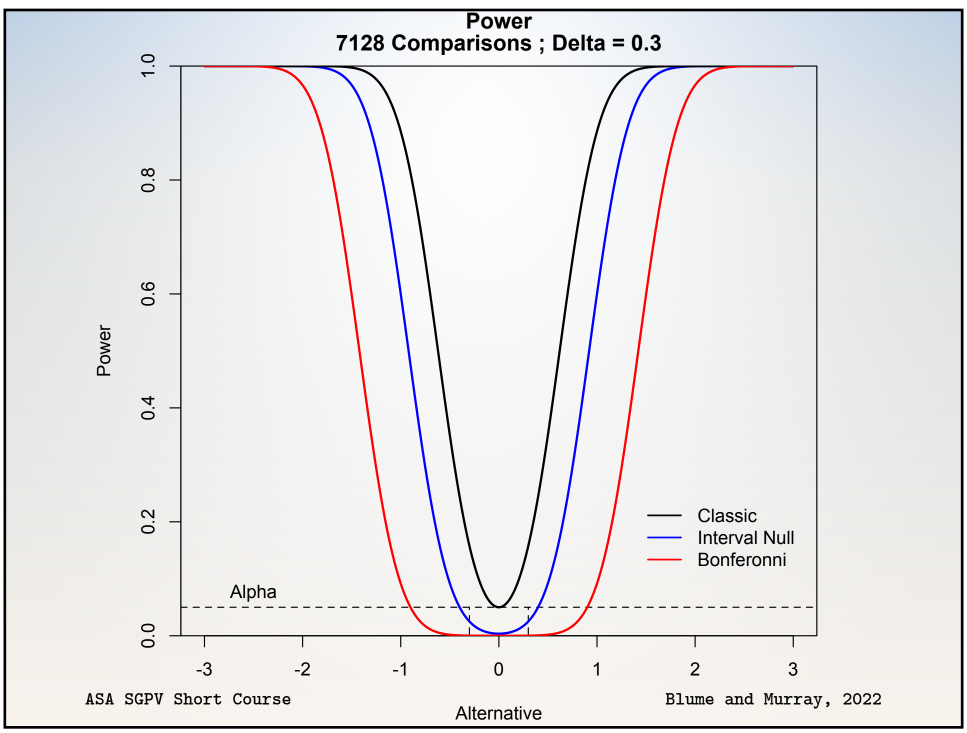
Compare with standard methods

- Second-generation p-value vs. Bonferroni correction
  - Adjusted for {10, 100, 7128} comparisons
  - Leukemia data example

- Remember SGPV are not adjusted for comparisons
- Discuss?




Power
100 Comparisons ; Delta = 0.3
Power
0.0
0.2
0.4
0.6
0.8
1.0
Alpha
Classic
Interval Null
Bonferonni
-3
-2
-1
0
1
2
3
Alternative
ASA SGPV Short Course
Blume and Murray, 2022


Power
7128 Comparisons ; Delta = 0.3
Power
0.0
0.2
0.4
0.6
0.8
1.0
Alpha
Classic
Interval Null
Bonferonni
-3
-2
-1
0
1
2
3
Alternative
ASA SGPV Short Course
Blume and Murray, 2022

Compatible with Null

- How often are the data compatible with null?
- Examine $P(p_\delta = 1|\theta)$ when θ is null or practically null
  - Essentially opposite of power function
- Sample size must be large enough to allow the null interval to contain the interval estimate
- This probability
  - converges to zero or one quickly for alternatives not near the edge of interval null

ASA SGPV Short Course — Blume and Murray, 2022

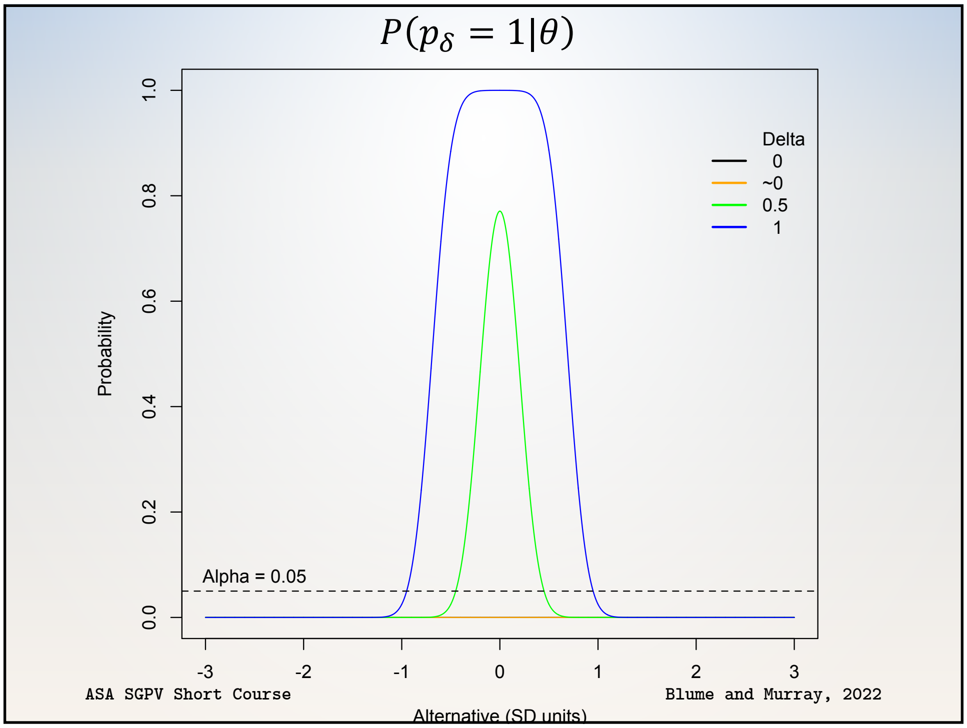
'Null Power' Function

- **How often are the data compatible with null?**
- Sample size must be large enough to allow null interval to contain the interval estimate, so $(\delta > Z_{\alpha/2}/\sqrt{n})$ or $(\sqrt{n} > Z_{\alpha/2}/\delta)$
- This probability converges to 0 or 1 quickly

$$P(p_\delta = 1|\theta) = \Phi\left[\frac{\sqrt{n}(\theta_0 + \delta)}{\sigma} - \frac{\sqrt{n}\theta}{\sigma} - Z_{\alpha/2}\right] - \Phi\left[\frac{\sqrt{n}(\theta_0 - \delta)}{\sigma} - \frac{\sqrt{n}\theta}{\sigma} + Z_{\alpha/2}\right]$$

when $\delta > Z_{\alpha/2}/\sqrt{n}$

ASA SGPV Short Course — Blume and Murray, 2022

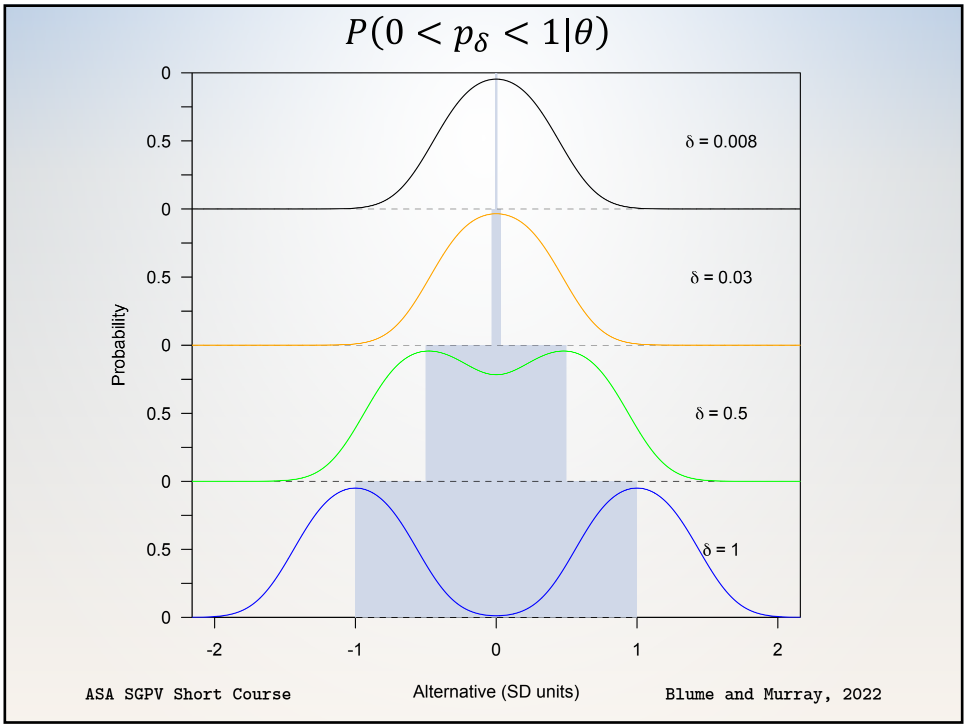


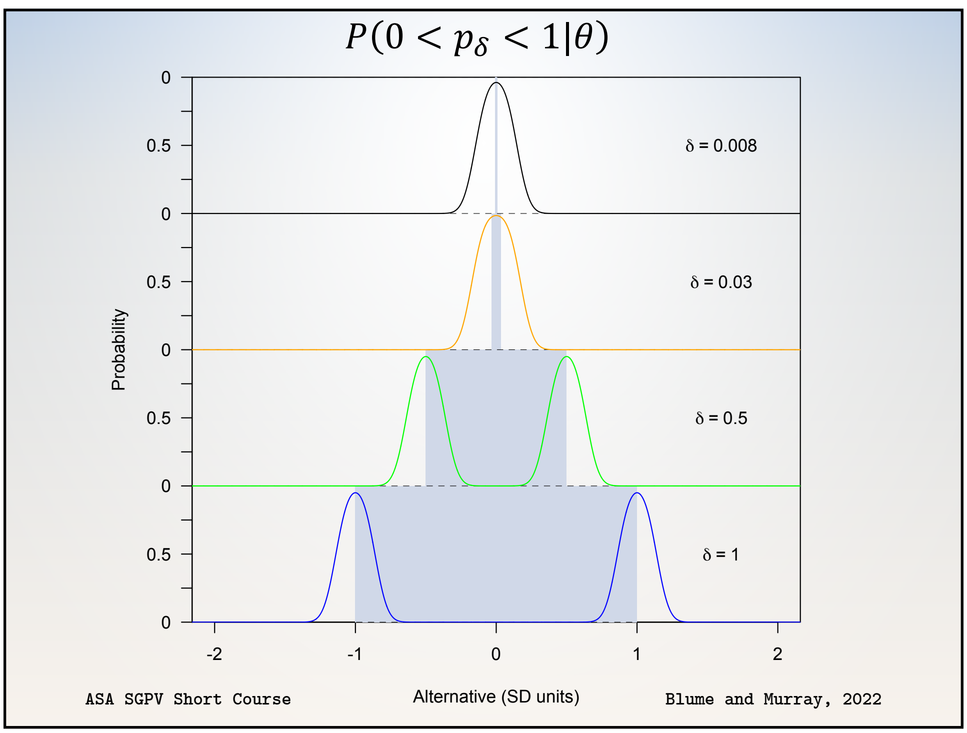
Probability of Inconclusive Data

- How often are the data inconclusive?
- Examine $P(0 < p_\delta < 1|\theta)$ for various θ
- This probability
  - drives sample size projections
  - is maximized when H is near the interval null edge
  - decreases quickly as H moves away from edge of null
- $P(0 < p_\delta < 1|\theta) = 1 - P(p_\delta = 0|\theta) - P(p_\delta = 1|\theta)$

ASA SGPV Short Course

Blume and Murray, 2022


$P(0 < p_\delta < 1|\theta)$
Probability
δ = 0.008
δ = 0.03
δ = 0.5
δ = 1
0
0.5
-2
-1
0
1
2
Alternative (SD units)
ASA SGPV Short Course
Blume and Murray, 2022


$P(0 < p_\delta < 1|\theta)$
Probability
δ = 0.008
δ = 0.03
δ = 0.5
δ = 1
0
0.5
-2
-1
0
1
2
Alternative (SD units)
ASA SGPV Short Course
Blume and Murray, 2022

Statistical Properties

Suppose interval I has coverage probability 1-α, then

Three 'Error' Rates

1. $P(p_\delta = 0|H_0) \leq \alpha$ and $\to 0$ as $n \to \infty$
2. $P(p_\delta = 1|H_1) \leq \alpha$ and $\to 0$ as $n \to \infty$
3. $P(0 < p_\delta < 1|H)$ controlled through sample size

Two False Discovery Rates

1. $P(H_0 \mid p_\delta = 0)$
2. $P(H_1 \mid p_\delta = 1)$

ASA SGPV Short Course Blume and Murray, 2022

False Discovery Rates

- FDR for 5 SGPV=0 findings; computed under various null and alternative configurations (w/ flat prior).

SNP ID	SGPV rank	p-value rank	OR	1/8 SI lower limit	1/8 SI upper limit	FDR_1	FDR_2	FDR_3
kgp4568244_C	1	133	0.10	0.03	0.37	2.9%	17.1%	3.3%
kgp8051290_G	13	2002	15.58	1.95	124.68	4.3%	30.3%	4.9%
kgp4497498_A	28	255	4.37	1.80	10.64	2.5%	8.6%	3.1%
rs3123636_G	423	1	1.39	1.26	1.55	0.01%	0.1%	0.4%
kgp7460928_G	1443	3310	1.78	1.11	2.87	2.4%	2.0%	3.0%

ASA SGPV Short Course Blume and Murray, 2022

False discovery rates

- Impact of α=0.05 vs α=0.05/7128 (7128 comparisons)

Error rates

- False Discovery Rate (FDR)

$$P(H_0|p<\alpha)=\left[1+\frac{(1-\beta)}{\alpha}r\right]^{-1}$$

- False Confirmation Rate (FCR)

$$P(H_1|p>\alpha)=\left[1+\frac{(1-\alpha)}{\beta}\frac{1}{r}\right]^{-1}$$

$r=P\,(H_1)/P\,(H_0)$

Error rates

ASA SGPV Short Course

Blume and Murray, 2022

False discovery rates

- **Second-generation *p*-values**
- False Discovery Rate (FDR)

Error Rates

$$P(H_0|p_\delta=0)=\left[1+\frac{P(p_\delta=0|H_1)}{P(p_\delta=0|H_0)}r\right]^{-1}$$

- False Confirmation Rate (FCR)

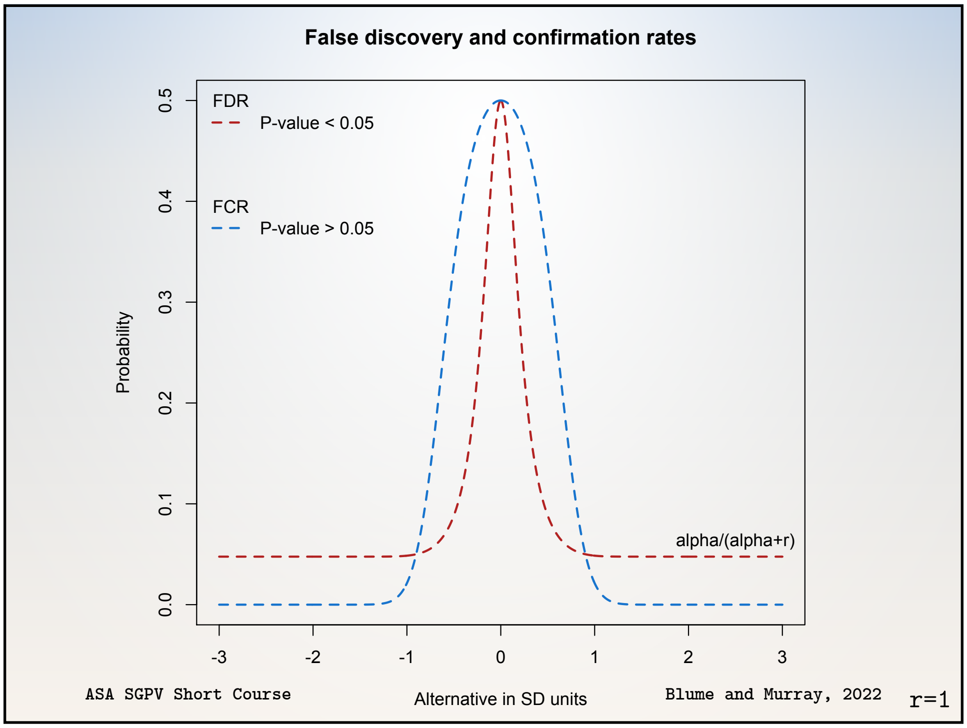
$$P(H_1|p_\delta=1)=\left[1+\frac{P(p_\delta=1|H_0)}{P(p_\delta=1|H_1)}\frac{1}{r}\right]^{-1}$$

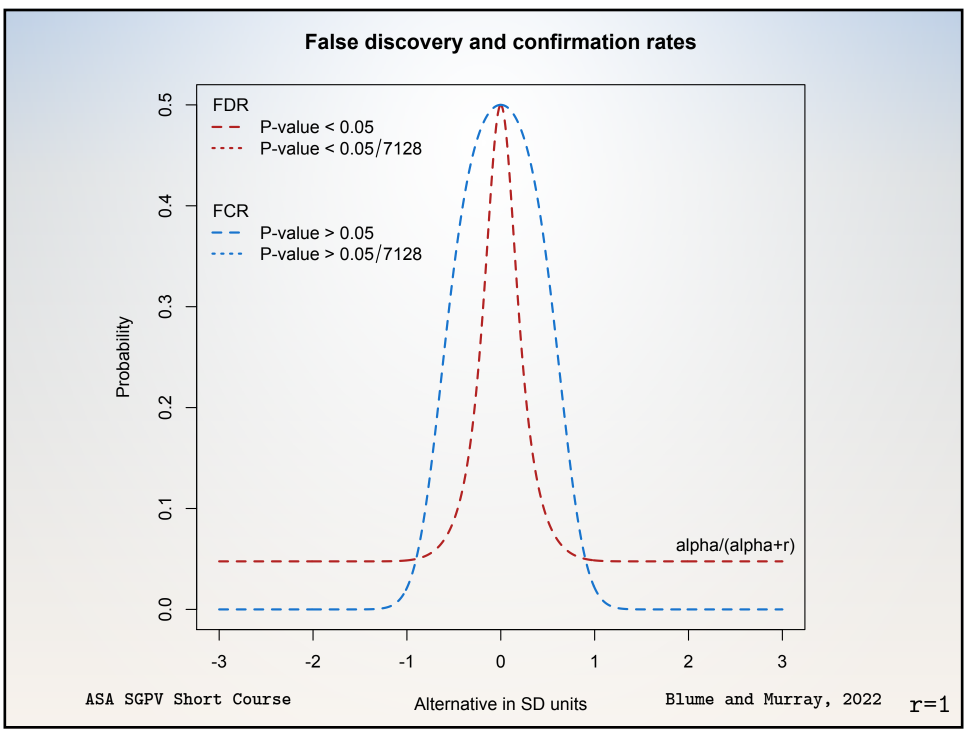
$r=P\,(H_1)/P\,(H_0)$

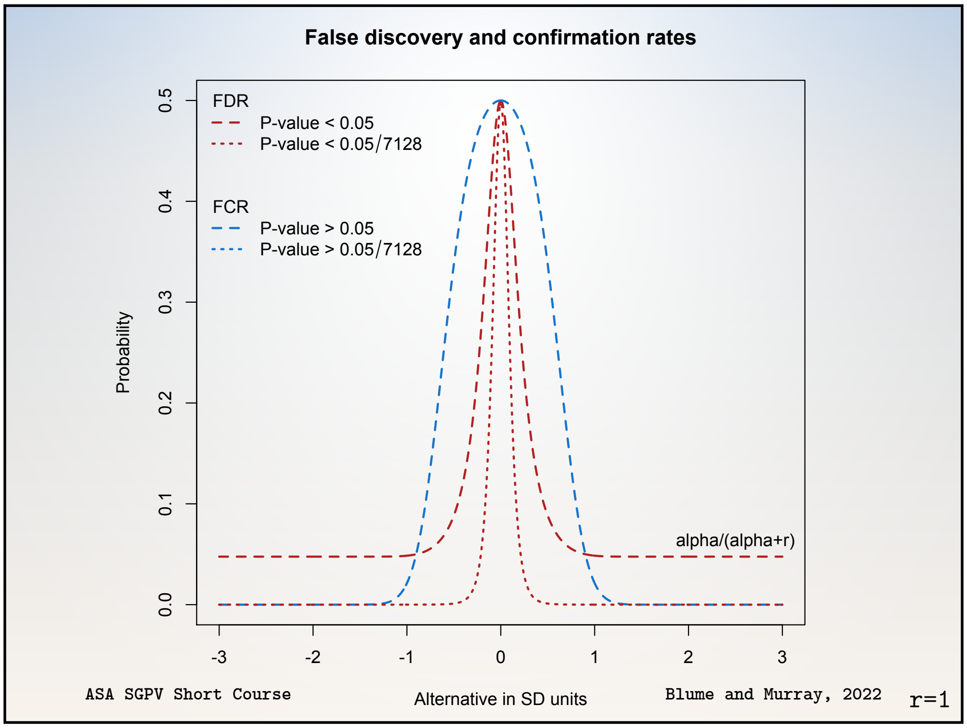
Error Rates

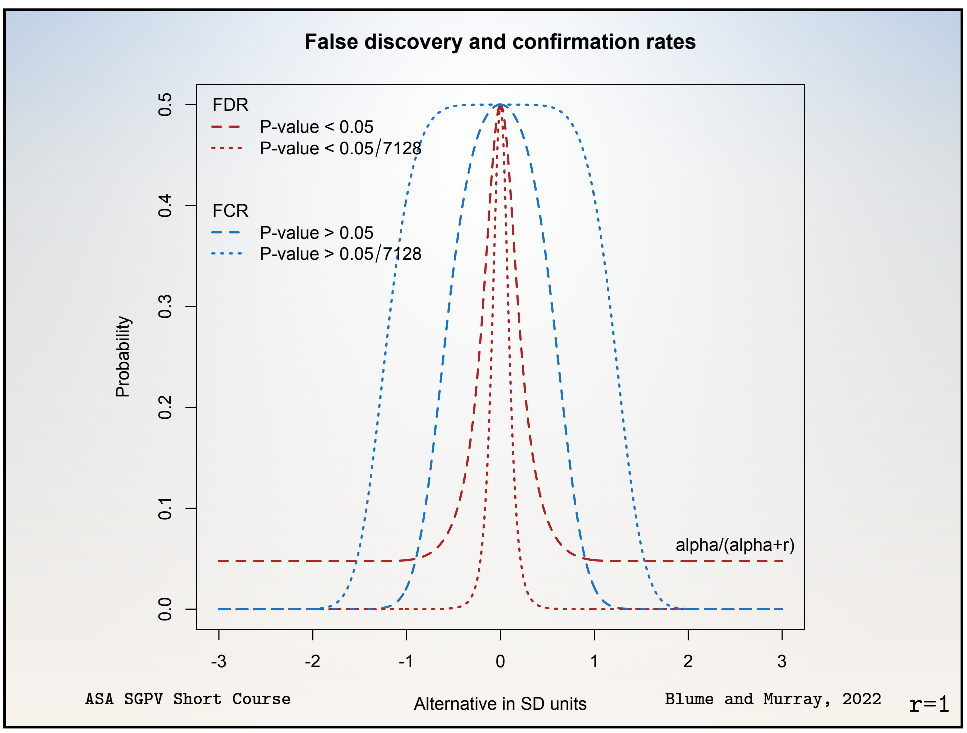
ASA SGPV Short Course

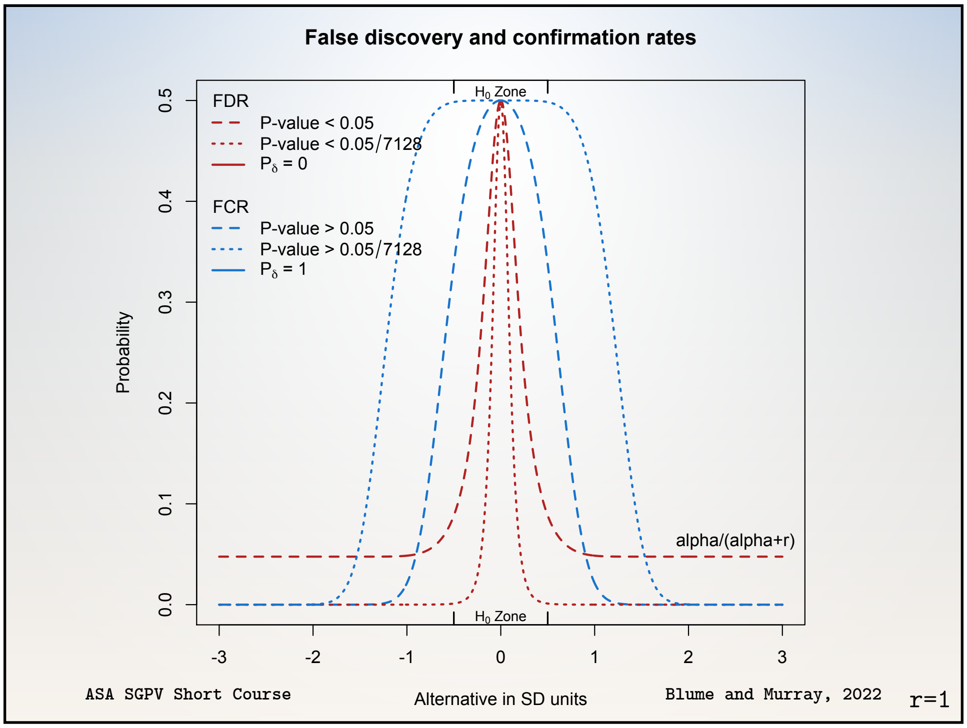
Blume and Murray, 2022

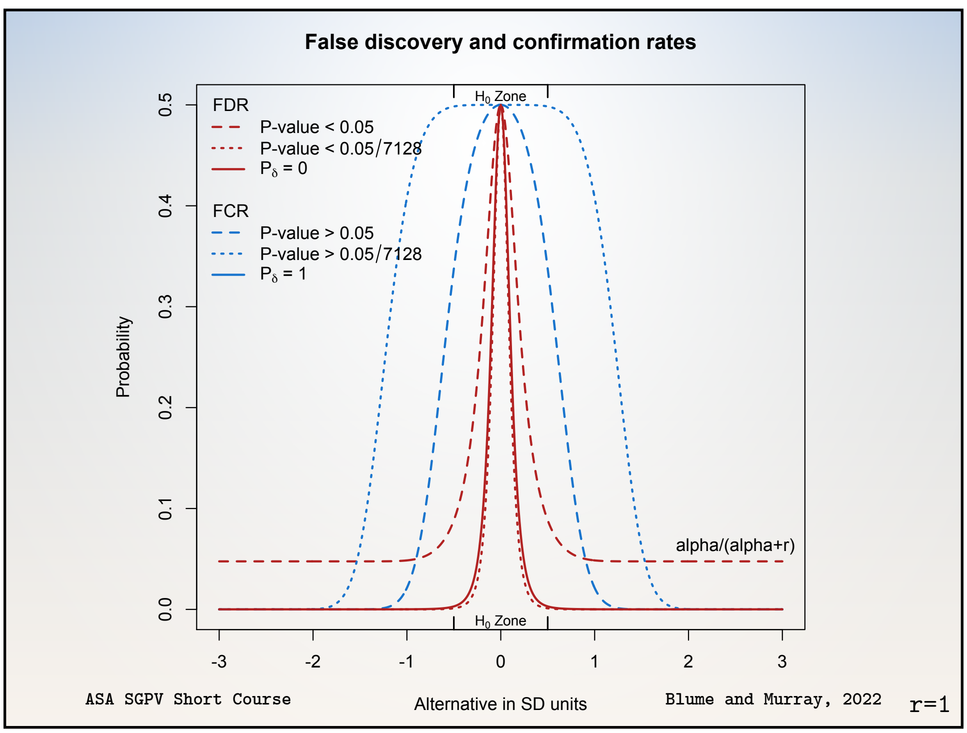

False discovery and confirmation rates
FDR
P-value < 0.05
FCR
P-value > 0.05
alpha/(alpha+r)
Probability
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
Alternative in SD units
ASA SGPV Short Course
Blume and Murray, 2022
r=1


False discovery and confirmation rates
FDR
P-value < 0.05
P-value < 0.05/7128
FCR
P-value > 0.05
P-value > 0.05/7128
alpha/(alpha+r)
Probability
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
Alternative in SD units
ASA SGPV Short Course
Blume and Murray, 2022
r=1

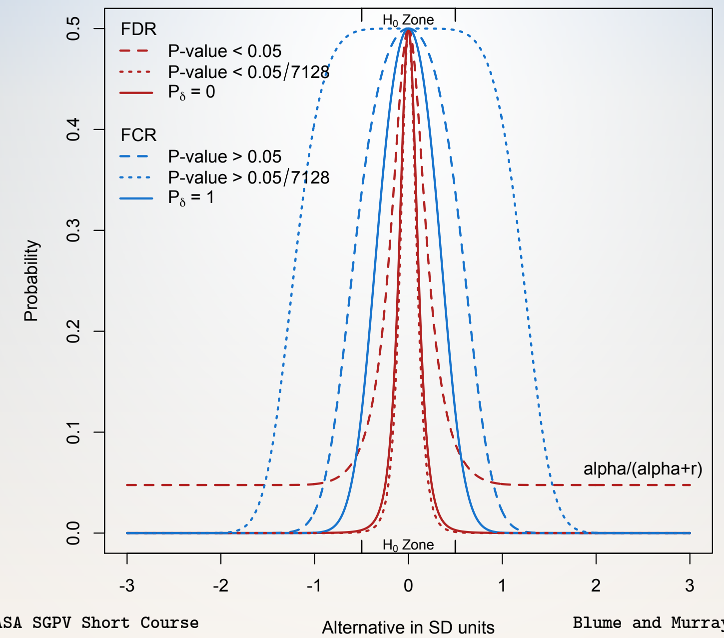

False discovery and confirmation rates
FDR
P-value < 0.05
P-value < 0.05/7128
FCR
P-value > 0.05
P-value > 0.05/7128
alpha/(alpha+r)
Probability
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
Alternative in SD units
ASA SGPV Short Course
Blume and Murray, 2022
r=1


False discovery and confirmation rates
FDR
P-value < 0.05
P-value < 0.05/7128
FCR
P-value > 0.05
P-value > 0.05/7128
alpha/(alpha+r)
Probability
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
Alternative in SD units
ASA SGPV Short Course
Blume and Murray, 2022
r=1


False discovery and confirmation rates
FDR
P-value < 0.05
P-value < 0.05/7128
P_δ = 0
FCR
P-value > 0.05
P-value > 0.05/7128
P_δ = 1
H_0 Zone
H_0 Zone
alpha/(alpha+r)
Probability
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
Alternative in SD units
ASA SGPV Short Course
Blume and Murray, 2022
r=1


False discovery and confirmation rates
FDR
P-value < 0.05
P-value < 0.05/7128
P_δ = 0
FCR
P-value > 0.05
P-value > 0.05/7128
P_δ = 1
H_0 Zone
H_0 Zone
alpha/(alpha+r)
Probability
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
Alternative in SD units
ASA SGPV Short Course
Blume and Murray, 2022
r=1

False discovery and confirmation rates



ASA SGPV Short Course

Blume and Murray, 2022

r=1

Remarks

- Second-generation *p*-values…
  - Has three "Error" rates
  - Allows Type I and II rate to converge to zero
  - Control changes of inconclusive results
  - Controls error rate using *science*
  - Reduces the false discovery rate
- Anchoring the scale of the effect size…
  - Eliminates most Type I Errors
  - Improves scientific translation of statistical model

ASA SGPV Short Course

Blume and Murray, 2022

FDR R Packages

- SGPVs
  - Valerie Welty
  - sgpv::fdrisk()
  - This function computes the false discovery risk (sometimes called the "empirical bayes FDR") for a second-generation *p*-value of 0, or the false confirmation risk for a second-generation *p*-value of 1.
- Raw p-values
  - FDRestimation::p.fdr()
  - This function computes FDRs and Method Adjusted p-values.
  - Methods include: Benjamini-Hochberg, Benjamini-Yeukateli, Bonferroni, Holm, Hochberg, and Sidak.

Time for Code Part 2!

Review of Topics Learned

→ Traditional p-values are flawed

→ Second-generation p-value framework and definition

→ Outrageous claim: The SGPV achieves the inferential properties that many scientists hope, or believe, are attributes of the classic p-value.

→ Comparison to Equivalence Tests: Two One-Sided Tests (TOST)

→ Statistical Properties of SGPVs

→ $P(p_\delta = 0|H)$, $P(p_\delta = 1|H)$, and $P(0 < p_\delta < 1|H)$

→ False Discovery Rates

→ Code

→ Compute and apply SGPVs

→ Reproduce shown examples

→ Compute FDRs

Things we couldn't get to:

- SGPV variable selection (regression modeling) by Yi Zuo
  - R Package: ProSGPV
    - Paper: f1000research.com/articles/11-58
    - Vignettes:
      - https://github.com/zuoyi93/ProSGPV/tree/master/vignettes
      - cran.r-project.org/web/packages/ProSGPV/vignettes/linear-vignette.html
- How to adjust for collaborator uncertainty when choosing the null zone
- Comparing SGPVs to machine learning methods
- Sequential monitoring by Jonathan Chipman
  - https://arxiv.org/abs/2204.10678

Acknowledgements

- Collaborators
  - William D. Dupont
  - Robert A. Greevy
  - Lucy D'Agostino McGowan
  - Jonathan Chipman
  - Valerie Welty
  - Lisa Lin
  - Jeffrey R. Smith
  - Yi Zuo
  - Thomas G. Stewart
  - Vanderbilt SEDS Lab

- Website
  - www.statisticalevidence.com

ASA SGPV Short Course   Blume and Murray, 2022

Questions?

ASA SGPV Short Course   Blume and Murray, 2022