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, PhD
Department of Biostatistics
Vanderbilt University

Course Layout

- Slides Part I: Introduction, applications, and statistical properties
  - Coding Part I
- Lunch (12:00-1:00pm)
- Slides Part II: Equivalence tests and false discovery rates
  - Coding Part II
- Slides Part III: SGPV Variable Selection
  - Coding Part III
- Questions and Discussion

Outline

- Equivalence Tests
  - Two One-Sided Tests (TOST)
  - Comparison to SGPVs
- False Discovery Rates
  - R Packages
    - sgpv::fdrisk()
    - FDRestimation::p.fdr()

Equivalence Tests

- Establish bioequivalence between data and an established equivalence range or interval null
- 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
  - $H_0 = [\theta^-, \theta^+]$

TOST Test

- Most popular frequentist test is the Two One-Sided t-tests (TOST) (Schuirmann 1987)
  - Flips the null and alternative (be careful)
  - Uses the $(1-2\alpha)\%$ confidence interval (be careful)

- Tests are ordinary, one-sided, α-level t-tests

- If *both* one-sided tests reject then conclude the evidence is contained in the equivalence range

TOST Test

- **Fail to reject the (original) null:** The confidence interval is outside of the indifference zone

$$H_0\colon (\theta < \theta^- \ or \ \theta > \theta^+)$$

- **Reject the (original) null:** The confidence interval is contained within the indifference zone

$$H_1\colon (\theta \geq \theta^- \ and \ \theta \leq \theta^+)$$

- Reported p-value is the p-value of largest magnitude from the two one-sided tests

$$p_T = max\{p_{T_1}, p_{T_2}\}$$

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

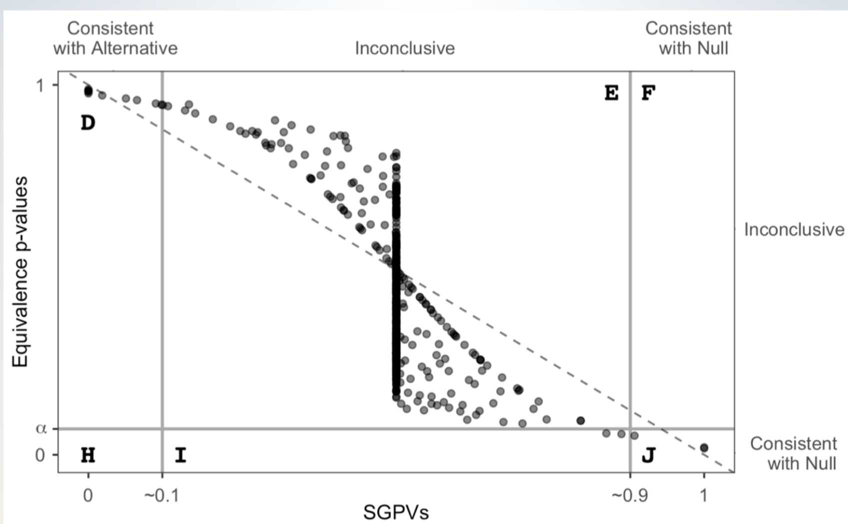
		SGPV Outcomes		
		Consistent with the alternative (SGPV near 0)	**Inconclusive** (SGPV near ½)	**Consistent with the null** (SGPV near 1)
Equivalence Tests Outcomes	**Consistent with the alternative** (*p*-value is unable to indicate this)	Not applicable **A**	Not applicable **B**	Not applicable **C**
	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 v SGPV Simulation

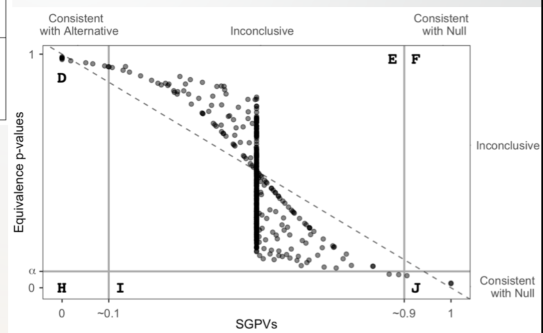
- Simulate TOST and SGPV reported p-values
  - Data generated under the null, $N(0,1)$
  - Sample size of n=6
  - Yields 70% power for $\Delta = 1$ with 5% type 1 error
  - Indifference zone is $[\theta^-, \theta^+] = [-0.375, 0.375]$
  - Uncertainty interval is 95% confidence interval
  - 500 iterations (for illustration)

TOST vs. SGPV comparison

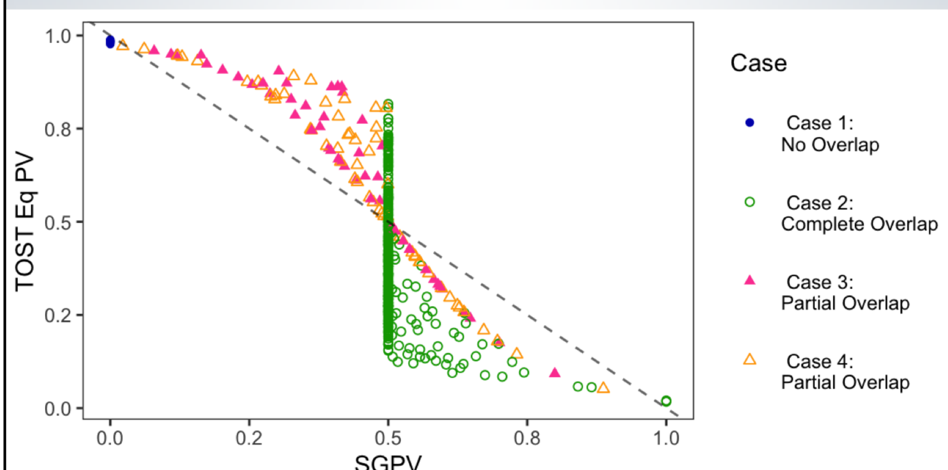


TOST vs. SGPV comparison

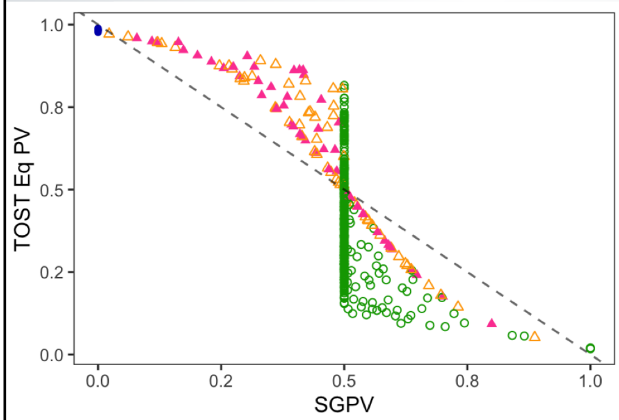
		SGPV Outcomes		
		Consistent with the alternative (SGPV near 0)	**Inconclusive** (SGPV near ½)	**Consistent with the null** (SGPV near 1)
Equivalence Tests Outcomes	**Consistent with the alternative** (*p*-value is unable to indicate this)	Not applicable **A**	Not applicable **B**	Not applicable **C**
	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**

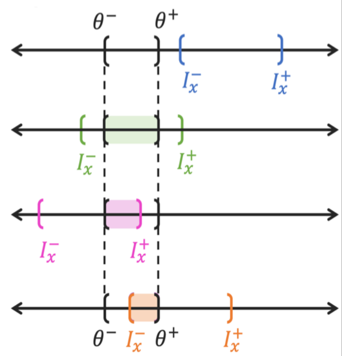


TOST vs. SGPV comparison



TOST vs. SGPV comparison





ASA SGPV Short Course Blume and Murray, 2022

Relationship

- Derived the mathematical relationship in all cases of overlap
- Case 3 and 4 of partial overlap:

$$p_\delta = \left[\left(\frac{1}{2c_\alpha}\right)\mathrm{F}_{\mathrm{n}}^{-1}(1-p_T) + \frac{1}{2}\right] \times max\left\{\frac{\frac{c_\alpha S}{\sqrt{n}}}{(\theta^+ - \theta^-)}, 1\right\}$$

- Limiting behavior under the null

$$P(Case\ 2) \to 1$$

$p_\delta \to 1$ (right)

$p_T \to 0$ (bottom)

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	• Indicates when data agree with null or alternative without additional testing
• No measure of overlap included in computation	• Includes overlap in reported p-value

Comparison

- TOST and SGPV are not one-to-one unless the variance is known

- TOST has significant limitations
- *SGPV is more flexible and easier to interpret*

Time for Code Part 2a!

10 Minute Break!

Statistical Properties

Suppose interval I has coverage probability 1-α, then

Three 'Error' Rates

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

Two False Discovery Rates

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

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%

False discovery rates

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

- False Discovery Rate (FDR)

Error rates

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

- False Confirmation Rate (FCR)

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

Error rates

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

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)

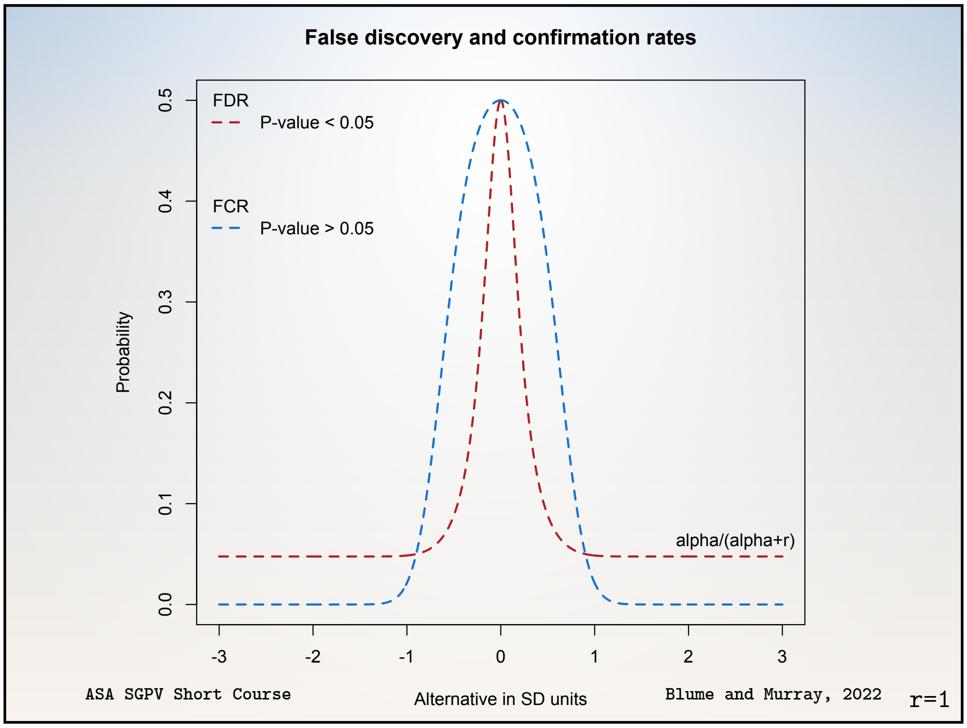
$$P(H_1|p_\delta = 1) = \left[1 + \frac{P(p_\delta = 1|H_0)}{P(p_\delta = 1|H_1)} \frac{1}{r}\right]^{-1}$$

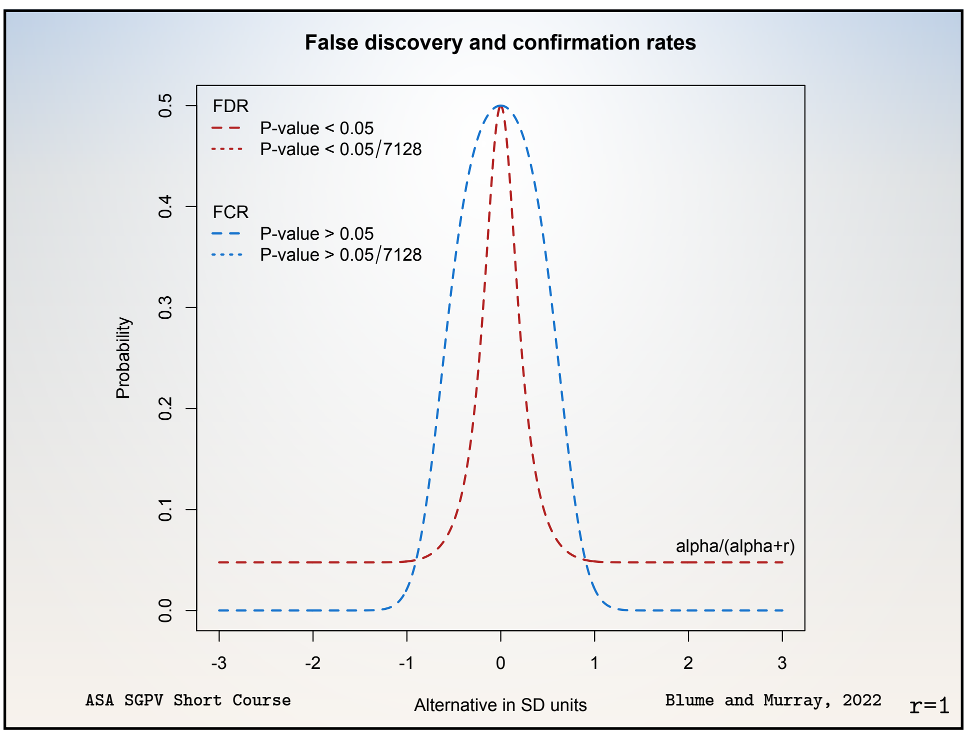
Error Rates

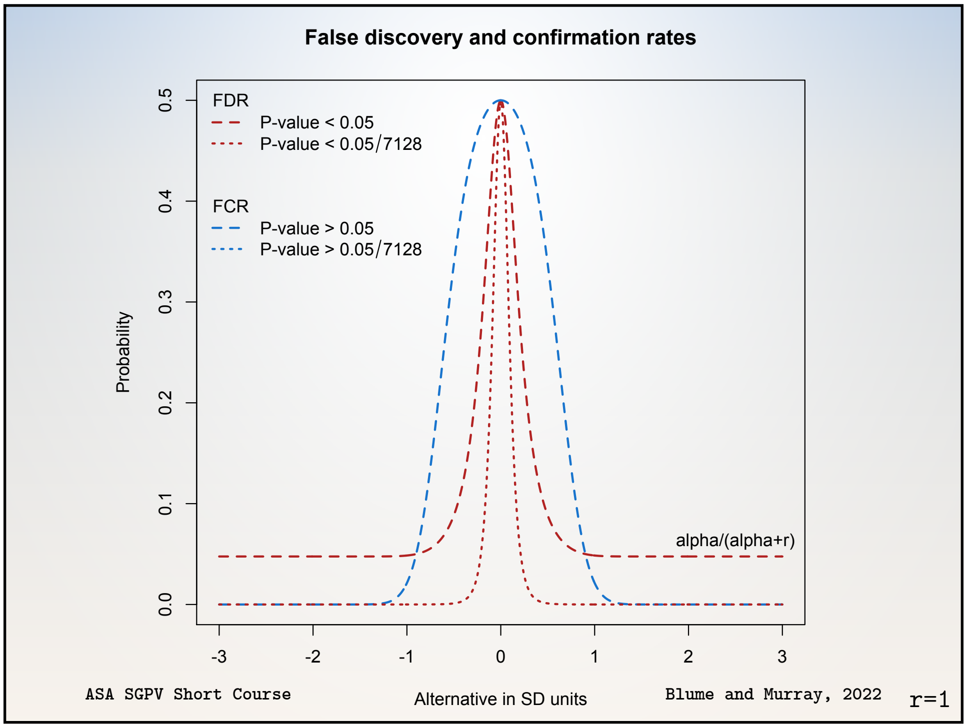
$r = P(H_1)/P(H_0)$

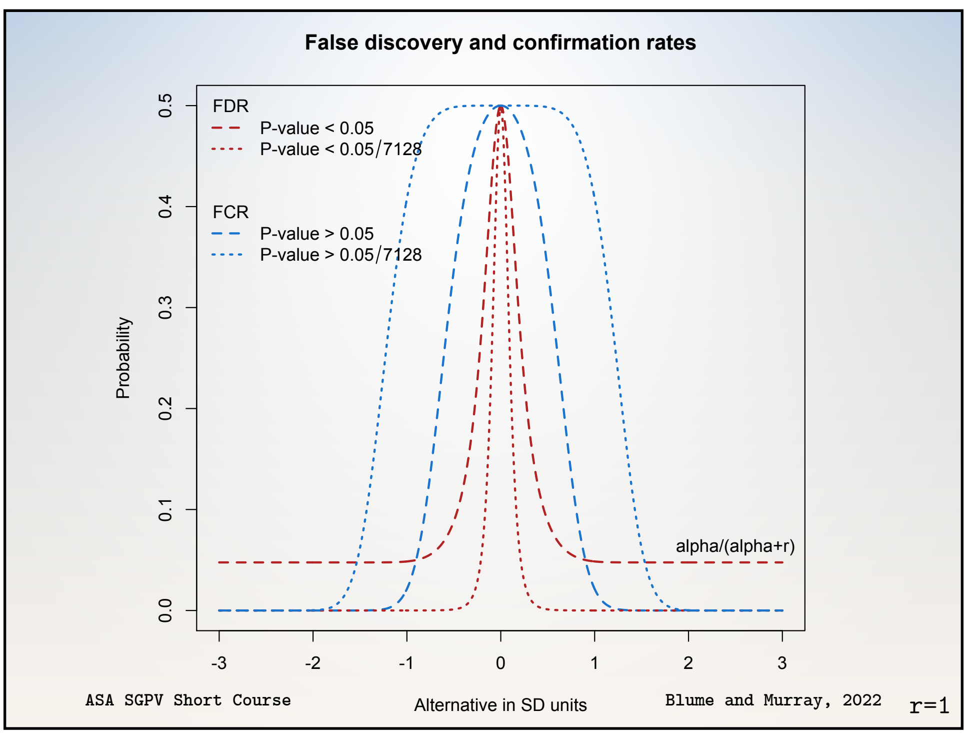
ASA SGPV Short Course

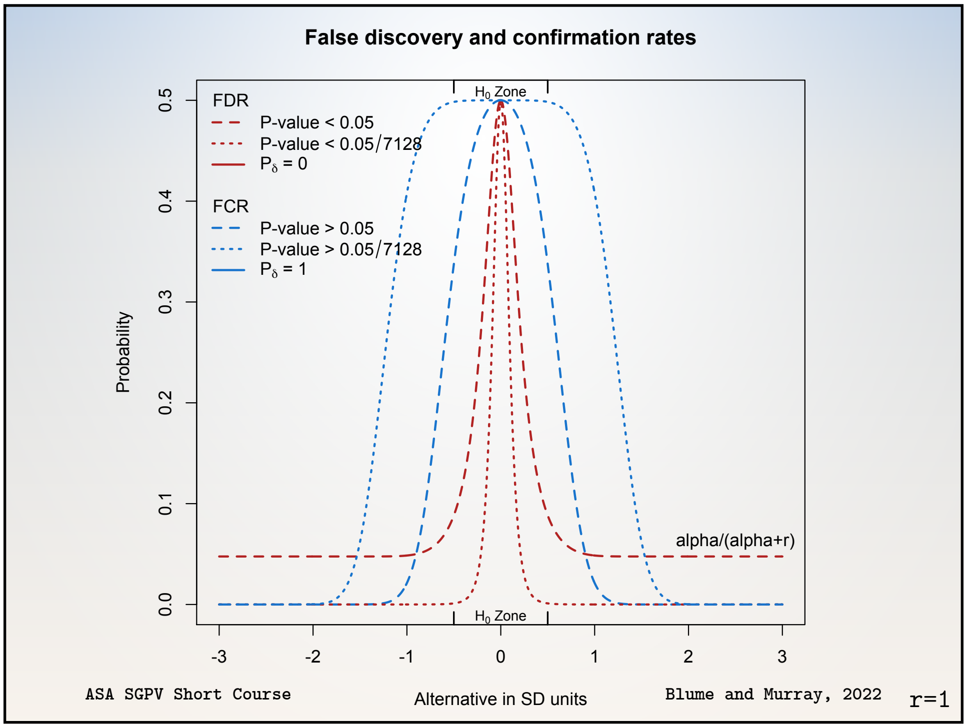
Blume and Murray, 2022

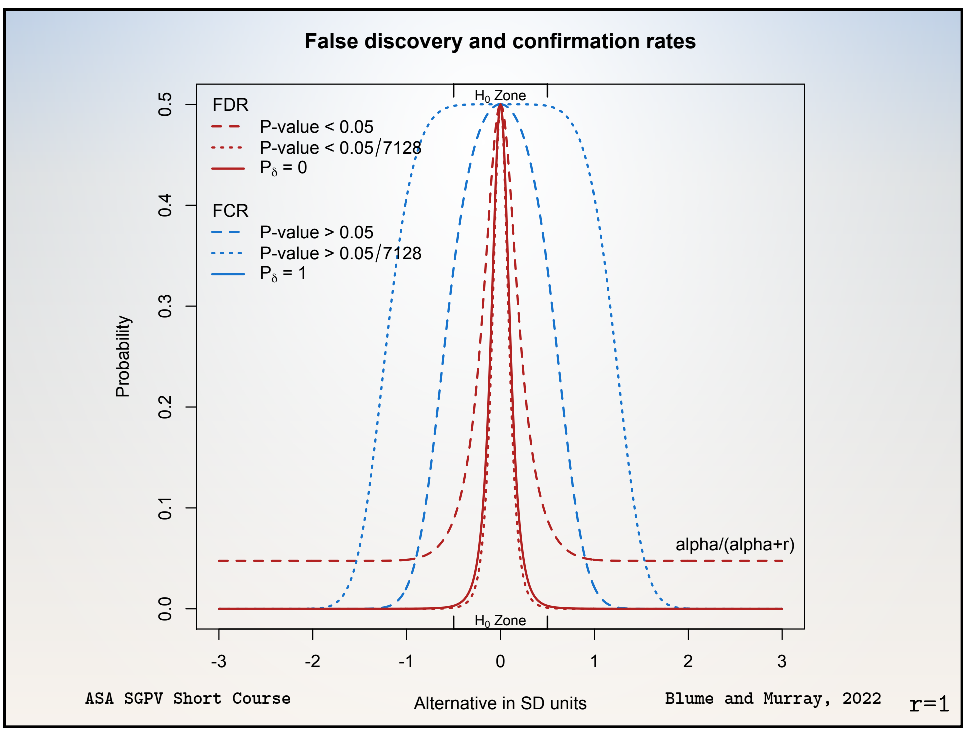

False discovery and confirmation rates
FDR
P-value < 0.05
FCR
P-value > 0.05
alpha/(alpha+r)
Probability
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
Alternative in SD units
ASA SGPV Short Course
Blume and Murray, 2022
r=1

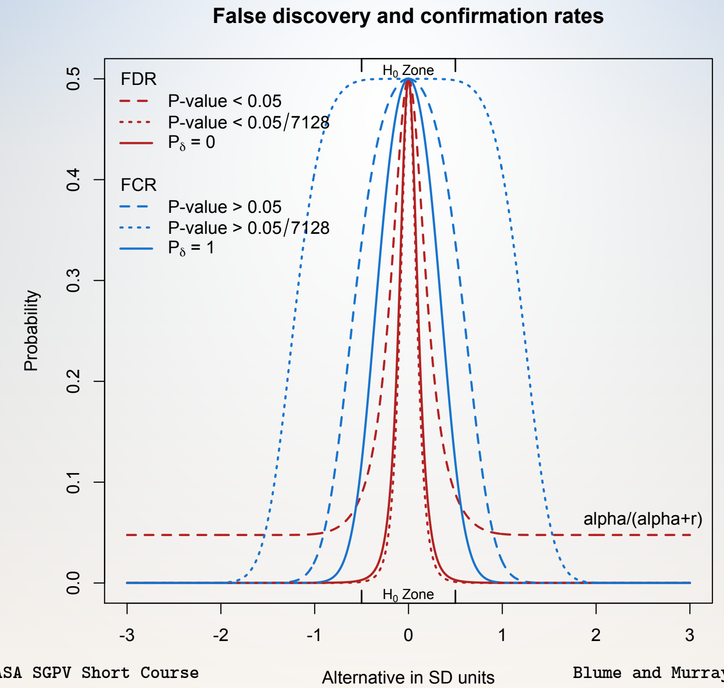

False discovery and confirmation rates
FDR
P-value < 0.05
P-value < 0.05/7128
FCR
P-value > 0.05
P-value > 0.05/7128
alpha/(alpha+r)
Probability
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
Alternative in SD units
ASA SGPV Short Course
Blume and Murray, 2022
r=1


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
Probability
0.0
0.1
0.2
0.3
0.4
0.5
alpha/(alpha+r)
-3
-2
-1
0
1
2
3
ASA SGPV Short Course
Alternative in SD units
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
Probability
0.0
0.1
0.2
0.3
0.4
0.5
alpha/(alpha+r)
-3
-2
-1
0
1
2
3
ASA SGPV Short Course
Alternative in SD units
Blume and Murray, 2022
r=1


False discovery and confirmation rates
FDR
P-value < 0.05
P-value < 0.05/7128
$P_\delta = 0$
FCR
P-value > 0.05
P-value > 0.05/7128
$P_\delta = 1$
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
$P_\delta = 0$
FCR
P-value > 0.05
P-value > 0.05/7128
$P_\delta = 1$
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



ASA SGPV Short Course Blume and Murray, 2022 r=1

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.

ASA SGPV Short Course Blume and Murray, 2022

Time for Code Part 2b!

10 Minute Break!