

Part I

Second-generation *p*-values: Introduction and Applications

Jeffrey D. Blume, and Megan H. Murray,

PhD

PhD

School of Data Science

Department of Biostatistics

University of Virginia

Vanderbilt University

About Us

- Jeffrey
 - Associate Dean and Professor in Data Science (UVA)
 - Areas of research include statistical inference, likelihood methods, second-generation *p*-values, prediction modeling, ROC curves, mediation modeling, missing data in prediction problems, and false discovery rates.
 - Website: www.statisticalevidence.com
 - UVA Profile: www.datascience.virginia.edu/people/jeffrey-blume
- Megan
 - PhD in Biostatistics at Vanderbilt: July 2022
 - Dissertation: ‘On second-generation *p*-values for equivalence testing and study planning, and flexible false discovery rate computation for classical *p*-values’
 - Fall 2022: Research Scientist at Eli Lilly Pharmaceuticals

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

ASA SGPV Short Course

Blume and Murray, 2022

Resources

- GitHub with Slides and Code
 - www.github.com/murraymegan/SGPV-ASA-Short-Course
- RStudio Desktop
 - www.rstudio.com/products/rstudio/download
- Interrupt or use Zoom chat for questions!
- For technical difficulties email Megan
 - megan.c.hollister@vanderbilt.edu

ASA SGPV Short Course

Blume and Murray, 2022

Synopsis

- Classical p -values are
 - Ubiquitous, Sacrosanct, Imperfect, Misused
 - Misunderstood (Significance vs. Hypothesis testing roles)
 - Openly debated in practice and theory
- Trend toward estimation in reporting of results
 - Report a estimation interval (e.g. confidence interval)
 - Does interval contain only clinically significant values?
- Second-generation p -values (SGPVs)
 - Embody and formalize this trend
 - Maintain and improve error rate control
 - Define clinically significant before looking at the data

ASA SGPV Short Course

Blume and Murray, 2022

Outline

- Evidential Metrics
- Second-generation p -value
 - Live coding using R
- Introductory examples
- High-dimensional examples
- Outrageous claim
- Statistical properties



ASA SGPV Short Course

Blume and Murray, 2022

Evidential metrics

Example:
Diagnostic Test

1. Measure of the strength evidence

- Axiomatic and intuitive justification
- Summary statistic, yardstick

Positive Test
Negative Test

2. Propensity to collect data that will yield a misleading #1

- Error rates
- Properties of the study design (!)

Sensitivity
Specificity

3. Probability that an observed #1 is misleading

- False Discovery rate, False Confirmation rate
- Chance that an observed result is mistaken
- Properties of the observed data (!)

PPV
NPV

ASA SGPV Short Course

Blume and Murray, 2022

Testing

Evidential Metric	What it measures	Hypothesis Testing	Significance Testing
1	strength of the evidence	Absent	Tail-area probability (<i>p</i> -value)
2	propensity for study to yield misleading evidence	Tail-area probability (error rates)	Absent
3	propensity for observed results to be misleading	misinterpret #2	misinterpret #1

- The tail-area probability is used to measure three distinct metrics

ASA SGPV Short Course

Blume and Murray, 2022

Second-generation p -value

- StatisticalEvidence.com
- Examine statistical properties later
- Retains strict error control

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

ASA SGPV Short Course

Blume and Murray, 2022

The p -value (what it is)

- Number between 0 and 1
- Smaller \Rightarrow support for an alternative hypothesis
- Larger \Rightarrow data are inconclusive
- Clinical significance is ignored
- Sample size confounds comparisons
- Interpretation
 - awkward
 - assumes null hypothesis true
 - rooted in inductive reasoning
- Not clear if/when ‘adjustments’ are necessary

ASA SGPV Short Course

Blume and Murray, 2022

The ^{2nd-generation} p -value (what we want it is)

Version 2.0

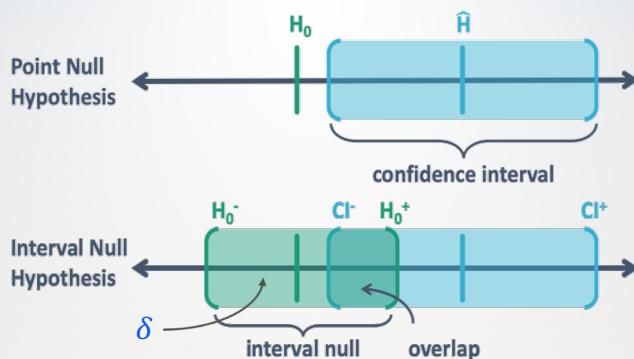
- ✓ Number between 0 and 1 →
 - near 0 supports alt
 - near 1 supports null
 - near $\frac{1}{2}$ inconclusive
- ✓ Smaller \Rightarrow support for an alternative hypothesis
 - Larger \Rightarrow data ~~are inconclusive~~ support null
 - Clinical significance is ~~ignored~~ incorporated
- ✗ Sample size confounds comparisons
 - Interpretation → Fraction of data-supported hypotheses that are null
 - ~~awkward~~ straightforward
 - assumes ~~null hypothesis true~~ conditions on observed data
 - ~~rooted in inductive reasoning~~ descriptive, summarizes
 - ~~Not~~ clear if/when 'adjustments' are necessary

Ideally, never

ASA SGPV Short Course

Blume and Murray, 2022

Illustration



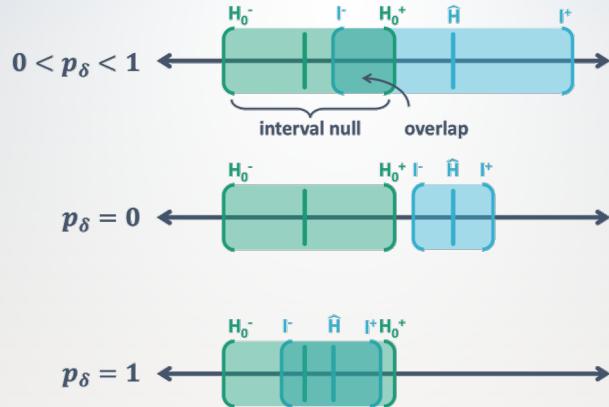
Point null hypothesis H_0 and interval null hypothesis $[H_0^-, H_0^+]$

Data-supported hypothesis \hat{H} and confidence interval $[CI^-, CI^+]$

ASA SGPV Short Course

Blume and Murray, 2022

Illustration

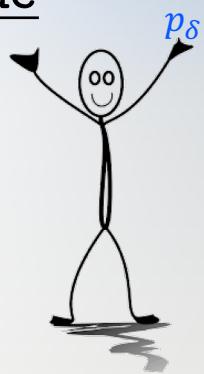


Works with confidence, credible, and support intervals

ASA SGPV Short Course

Blume and Murray, 2022

Second-generation p -value



- SGPV is in $[0,1]$ and denoted by p_δ
- δ for scientific significance
 1. $p_\delta = 0 \Rightarrow$ null **incompatible** with data
 2. $p_\delta = 1 \Rightarrow$ null **compatible** with data
 3. $0 < p_\delta < 1 \Rightarrow$ data are **inconclusive**
- Fraction of data-supported hypotheses that are null
- Retains strict error control, all rates $\rightarrow 0$

ASA SGPV Short Course

Blume and Murray, 2022

Definition

**Second-generation
p-value (SGPV)**

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

Proportion of data-supported hypotheses that are also null hypotheses

**Small-sample
correction factor**

shrinks proportion to $\frac{1}{2}$ when $|I|$ wide

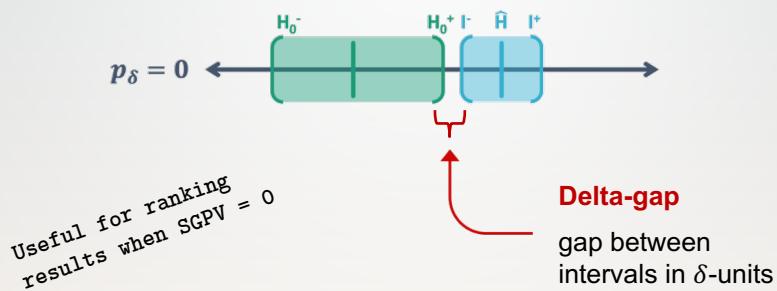
when $|I| > 2|H_0|$

ASA SGPV Short Course

Blume and Murray, 2022

The Delta-gap

When SGPV=0, there is a gap between the intervals. The length of that gap, in δ -units is the **delta-gap**.



ASA SGPV Short Course

Blume and Murray, 2022

Steps

- SGPV ~ the fraction of data-supported hypotheses that are null or practically null
1. Specify an interval null hypothesis or a point null with indifference zone
 2. Find confidence, support or credible interval
 3. Measure the fraction of interval (#2) that is in the null interval
 4. Apply small-sample correction factor, as necessary

ASA SGPV Short Course

Blume and Murray, 2022

COVID Clinical Trial



THE PREPRINT SERVER FOR HEALTH SCIENCES



HOME | SUBMIT | FAQ | BLOG | ALERTS / RSS | ABOUT

Advanced Search

Preprint

Ivermectin for Treatment of Mild-to-Moderate COVID-19 in the Outpatient Setting: A Decentralized, Placebo-controlled, Randomized, Platform Clinical Trial

Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV)-6 Study Group, Susanna Naglie
doi: <https://doi.org/10.1101/2022.06.10.22276252>

This article is a preprint and has not been certified by peer review [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.



Abstract Full Text Info/History Metrics

Preview PDF

Previous

Next

Posted June 12, 2022.

- Download PDF Email
 Print/Save Options Share
 Author Declarations Citation Tools
 Supplementary Material
 Data/Code

[Tweet](#) [Like 222](#)

COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv

Abstract

ASA SGPV Short Course

Blume and Murray, 2022

COVID Clinical Trial

- Randomized 1,591 patients to ivermectin treatment or placebo
- Mean time spent unwell was estimated using a longitudinal ordinal regression model; range was 0 to 14 days
- Patients reported each day their symptoms and severity, health care visits, and medications.

Results: “The difference in the amount of time spent feeling unwell with COVID was estimated to be 0.49 days in favor of ivermectin with a 95% credible interval of (0.15, 0.82).”

ASA SGPV Short Course

Blume and Murray, 2022

COVID Clinical Trial

Uncertainty Data Interval: (0.15, 0.82) days

Difference in mean time unwell between ivermectin treatment and placebo.

Hypothesis	Indifference or Null Zone	SGPV (p_δ)	Inference Outcome
3 hours difference	[−0.125, 0.125] days	$p_\delta = 0$	Consistent with alternative zone effects
12 hours difference	[−0.5, 0.5] days	$p_\delta = 0.522$	Inconclusive
18 hours difference	[−0.75, 0.75] days	$p_\delta = 0.896$	Inconclusive
1 day difference	[−1, 1] days	$p_\delta = 1$	Consistent with null zone effects
2 days difference	[−2, 2] days	$p_\delta = 1$	Consistent with null zone effects

ASA SGPV Short Course

Blume and Murray, 2022

Time for Code Part 1a!

ASA SGPV Short Course

Blume and Murray, 2022

10 Minute Break!

ASA SGPV Short Course

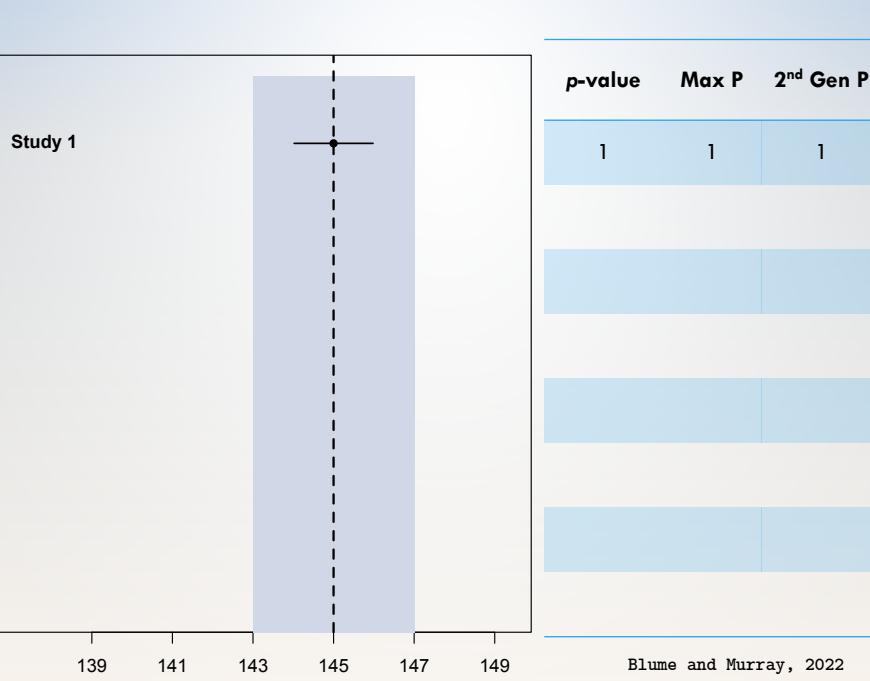
Blume and Murray, 2022

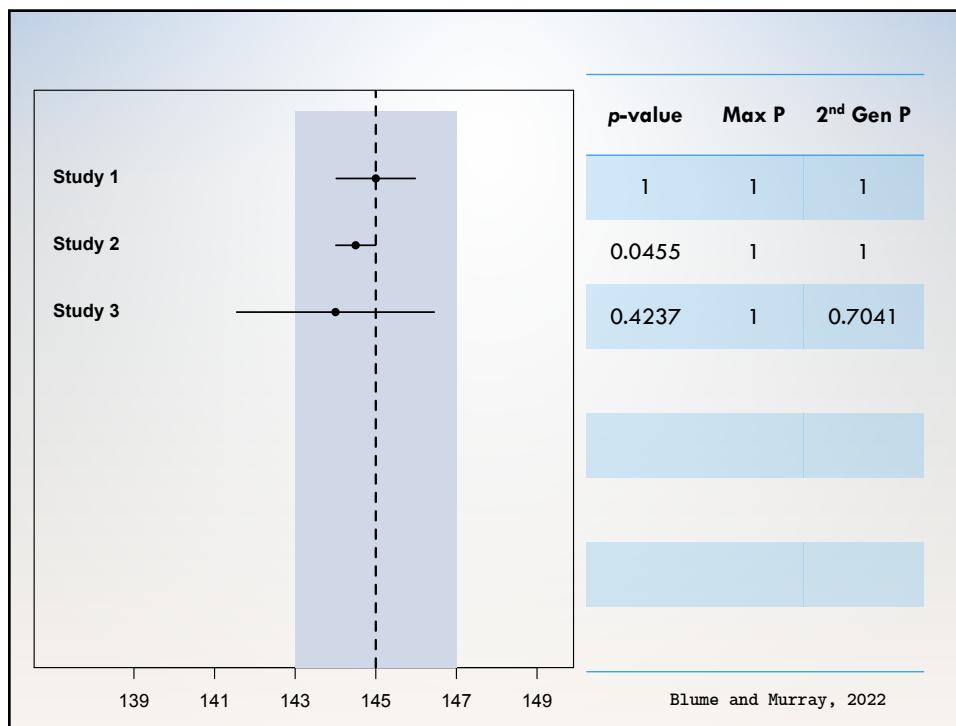
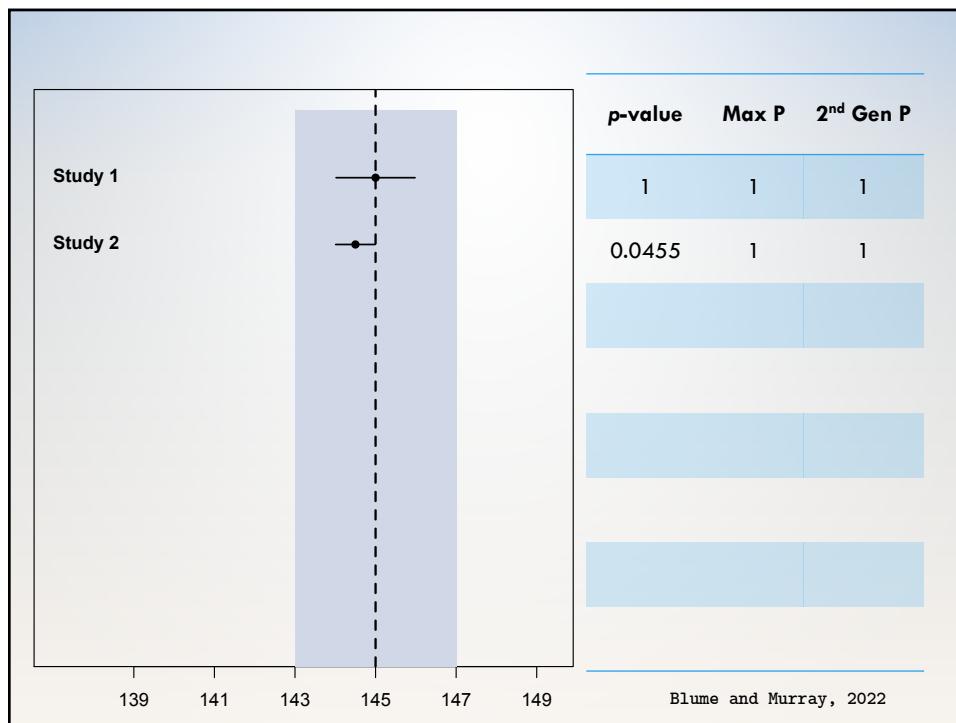
Systolic Blood Pressure

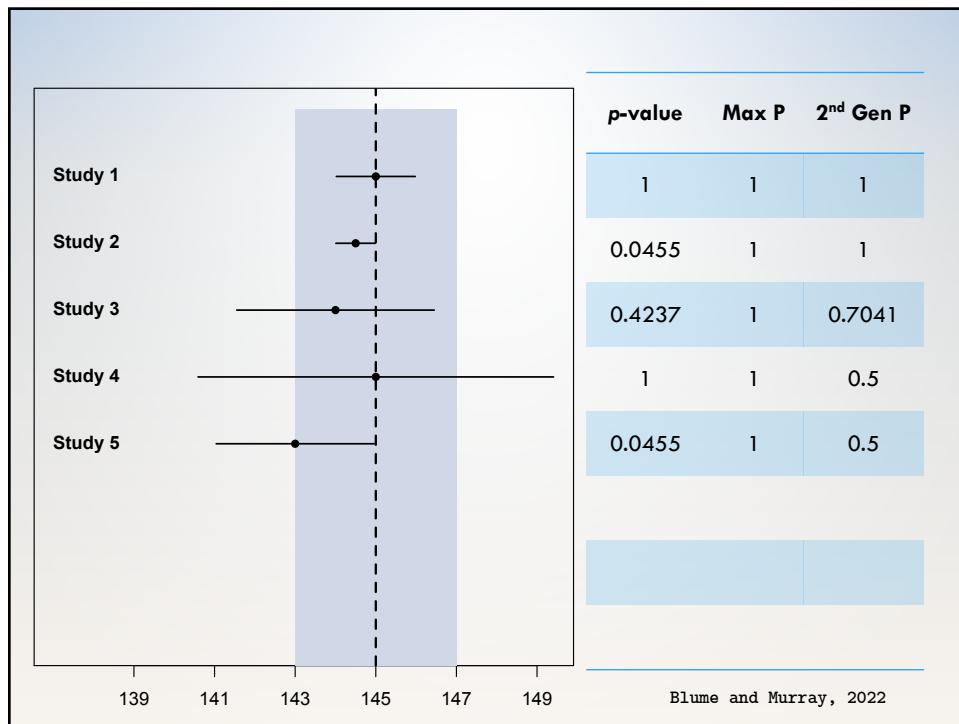
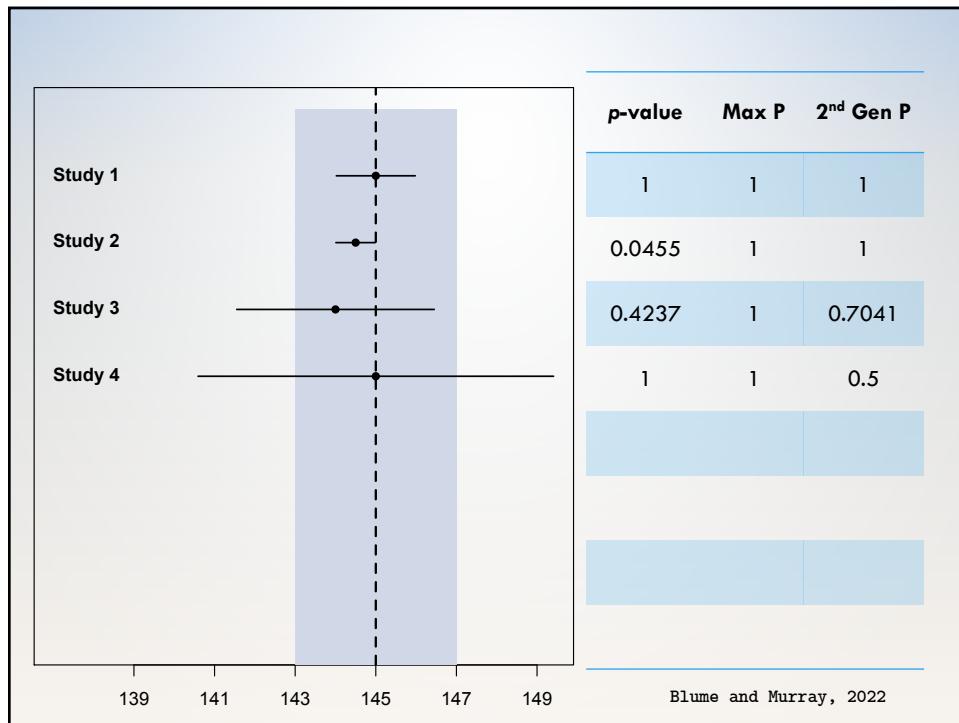
- SBP is reported to the nearest 2 mmHg
- Null Hypothesis: mean SPB is 145 mmHg
- Interval Null hypothesis: mean is 143 to 147 mmHg
- Results from 8 mock studies

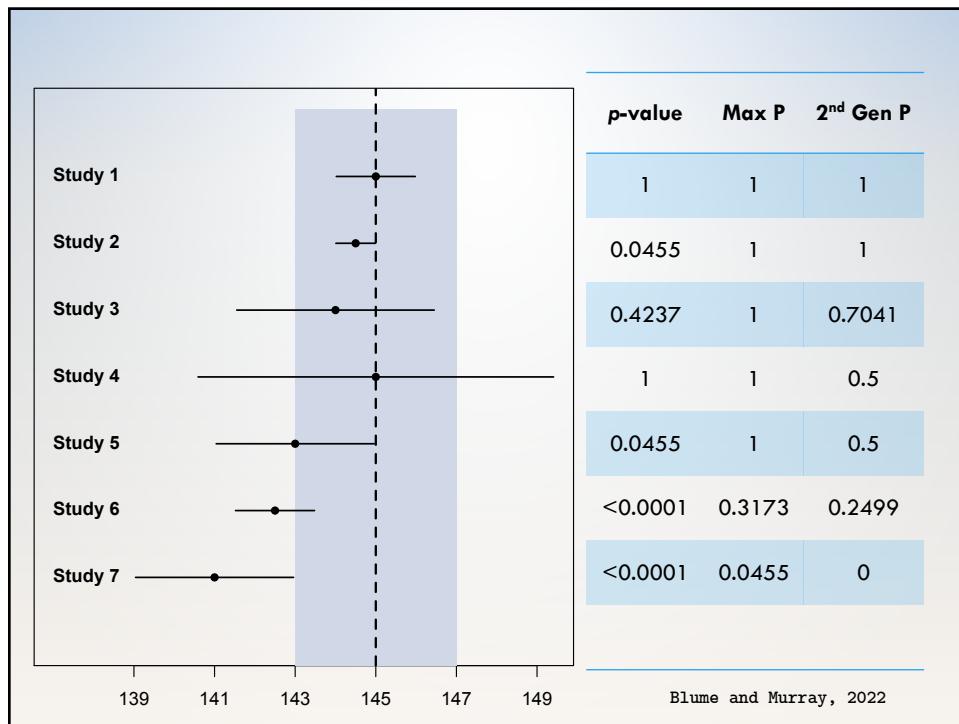
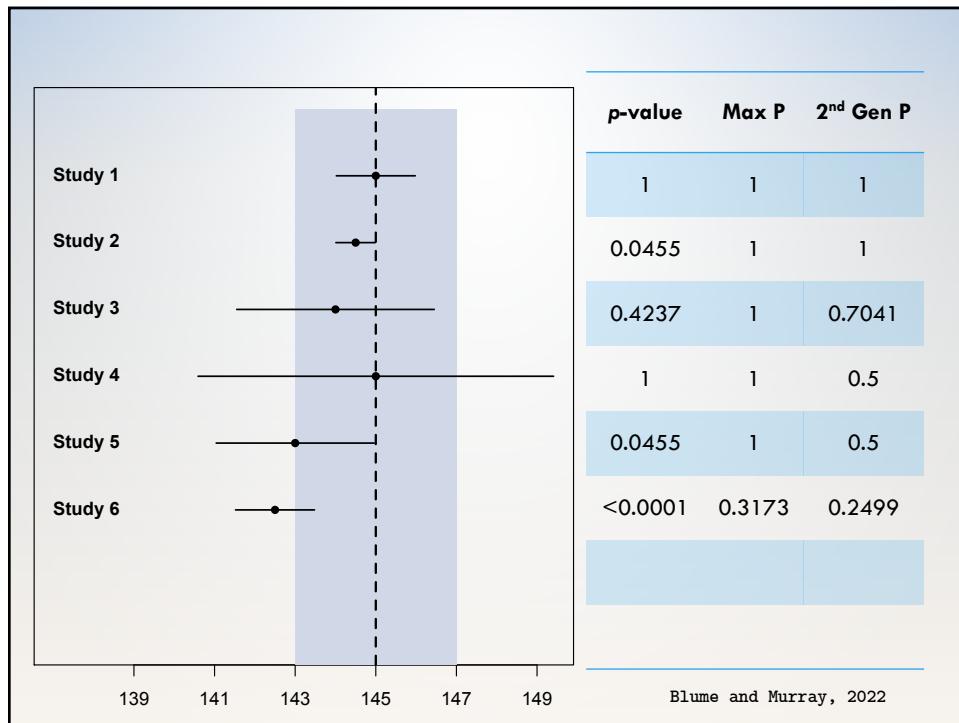
ASA SGPV Short Course

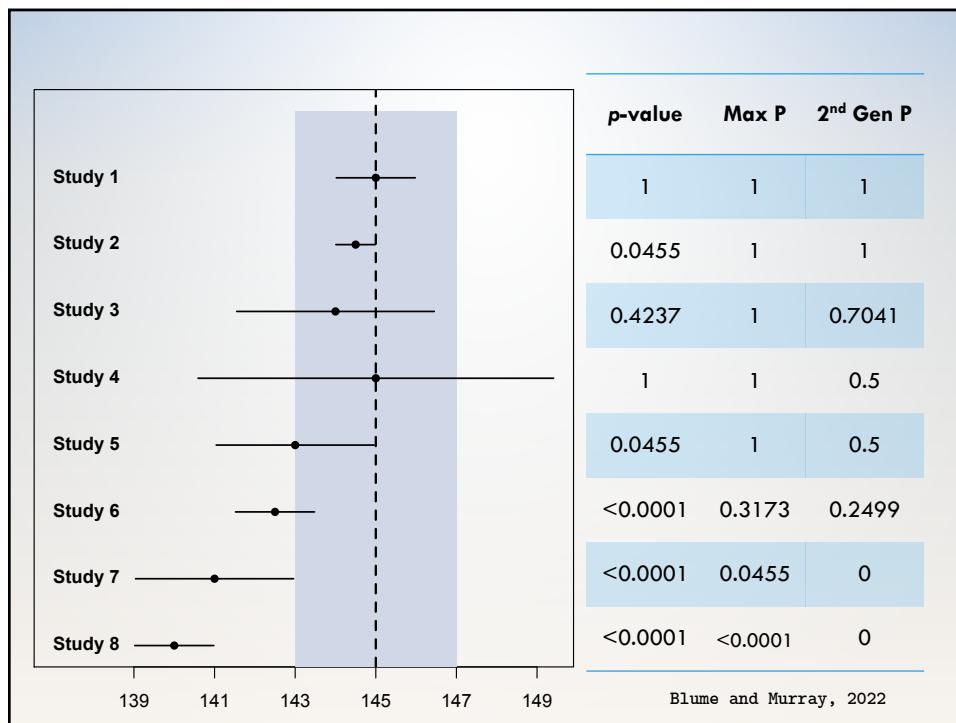
Blume and Murray, 2022











2x2 Tables & Odds Ratios

Exposure	Outcome	
	No	Yes
Exposed	35	65
Unexposed	50	50

- Scale can be important for inconclusive results
- A p_δ of 0 or 1 will always remain the same
- An inconclusive p_δ will vary slightly

OR = 1.86
95% CI: (1.05, 3.29)

Null: (0.9, 1.11)

$$p_\delta = \frac{(1.11 - 1.05)}{(3.29 - 1.05)}(1) = 0.024$$

log(or) = 0.62
95% CI: (0.05, 1.19)

Null: (-0.1, 0.1)

$$p_\delta = \frac{(0.1 - 0.05)}{(1.19 - 0.05)}(1) = 0.044$$

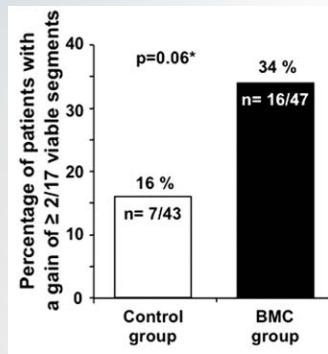
Bone Marrow in Acute Myocardial Infarction (BOMAMI)

- European Heart Journal (2011)
- Randomized multicenter study
- Intracoronary administration of autologous bone marrow cells (BMCs) can lead to a modest improvement in cardiac function
- Aim: Evaluate the effect of BMC therapy on myocardial viability in patients with decreased left ventricular ejection fraction (LVEF) after acute myocardial infarction (AMI)

ASA SGPV Short Course

Blume and Murray, 2022

BOMAMI Trial



Odds ratios	95% confidence interval	P-value
2.654	0.967 - 7.286	0.06

Null Interval: (0.9, 1.11)

$$p_\delta = \frac{(1.11 - 0.967)}{(7.286 - 0.967)} (15) = 0.34$$

ASA SGPV Short Course

Blume and Murray, 2022

Effect Measures for BOMAMI

	BMC	Control	Total	
Gain	16	7	23	
No Gain	31	36	67	
Total	47	43	90	
Risk	0.34	0.16		

Null Hypotheses

OR/RR: (0.9, 1.11)

RD: (-0.05, 0.05)

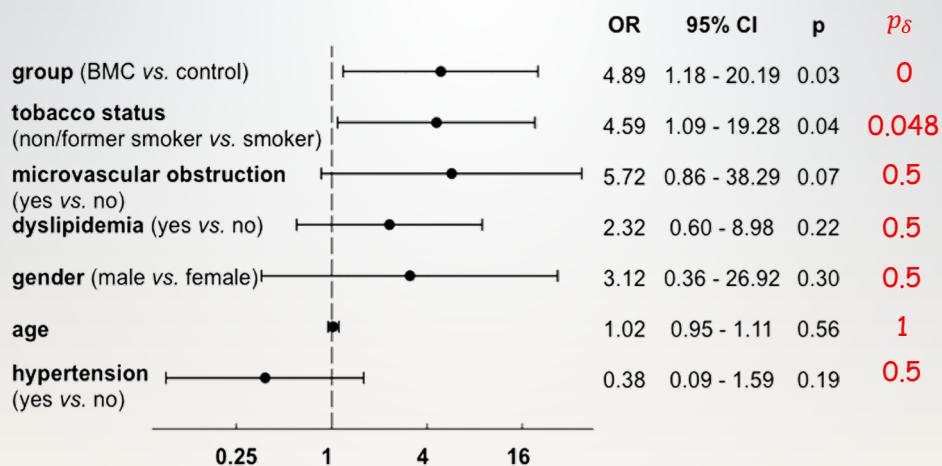
	Estimate	CI Lower	CI Upper	SGPV
Odds Ratio	2.65	0.967	7.286	0.34
Risk Ratio	2.09	0.953	4.589	0.37
Risk Difference	0.18	0.003	0.352	0.24

ASA SGPV Short Course

Blume and Murray, 2022

BOMAMI Trial

Logistic Regression with Null Zone: (0.9, 1.11)



ASA SGPV Short Course

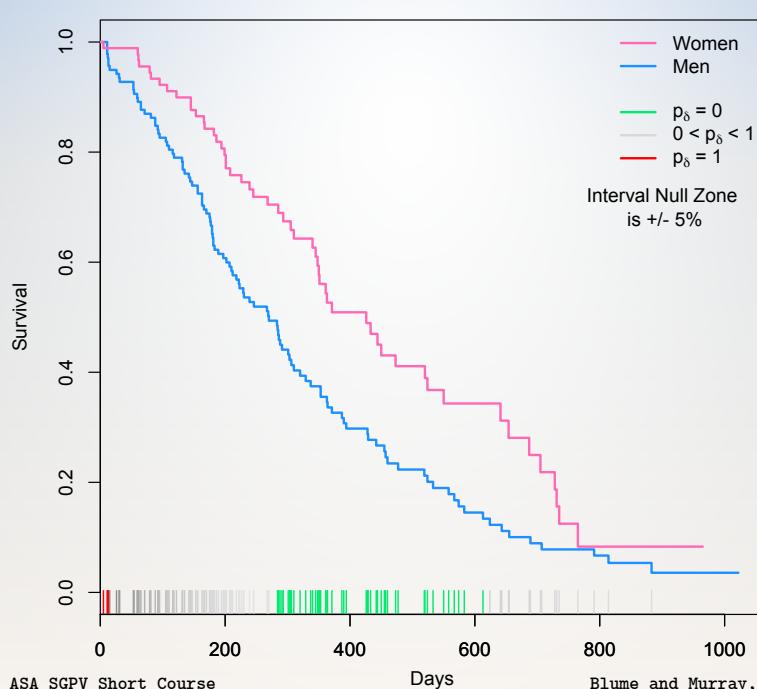
Blume and Murray, 2022

Lung Cancer Survival

- Survival time in patients with advanced lung cancer (days)
- Potential for gender dissimilarities
- Trial by North Central Cancer Treatment Group (1994)

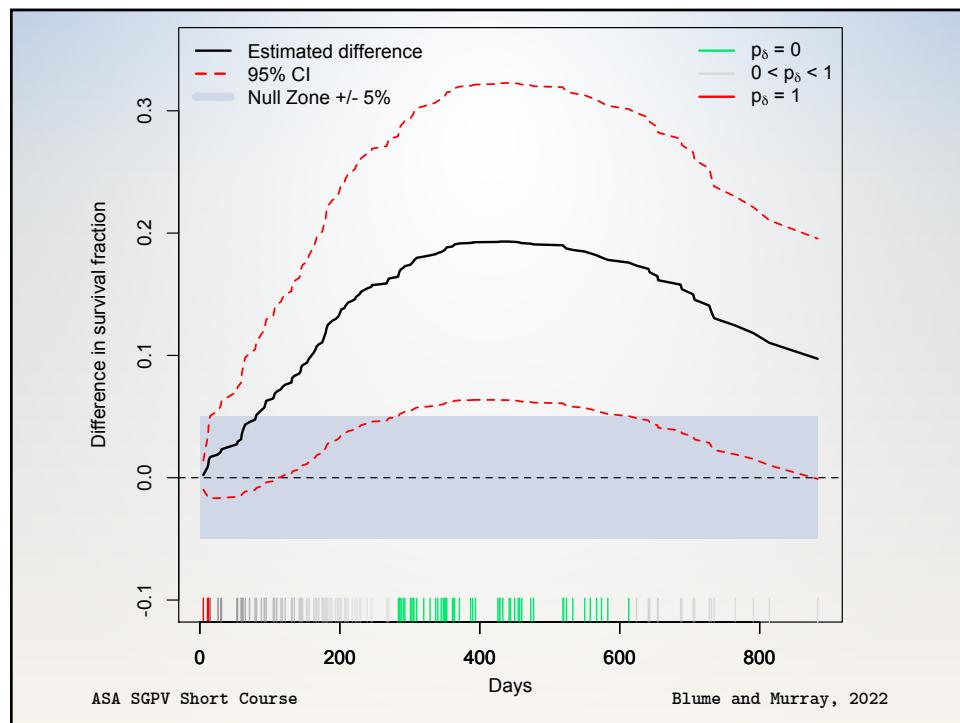
ASA SGPV Short Course

Blume and Murray, 2022



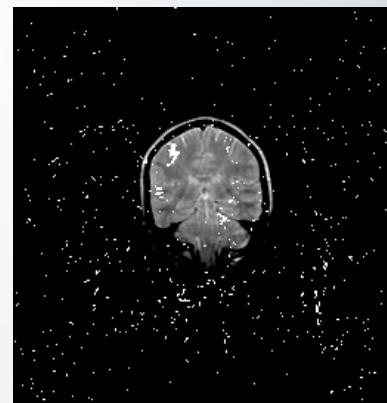
ASA SGPV Short Course

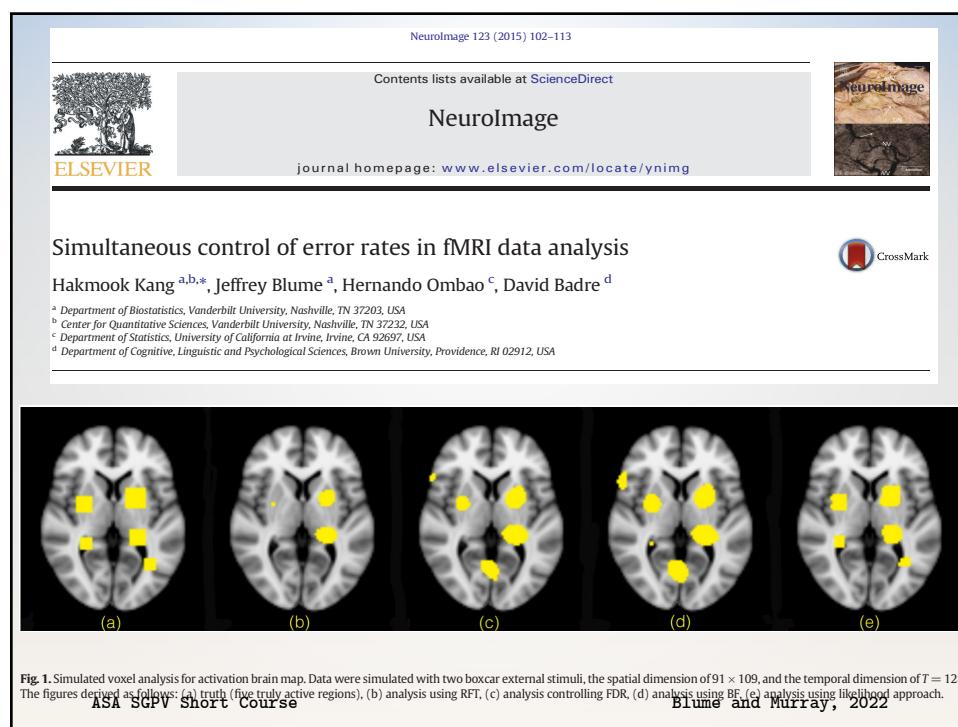
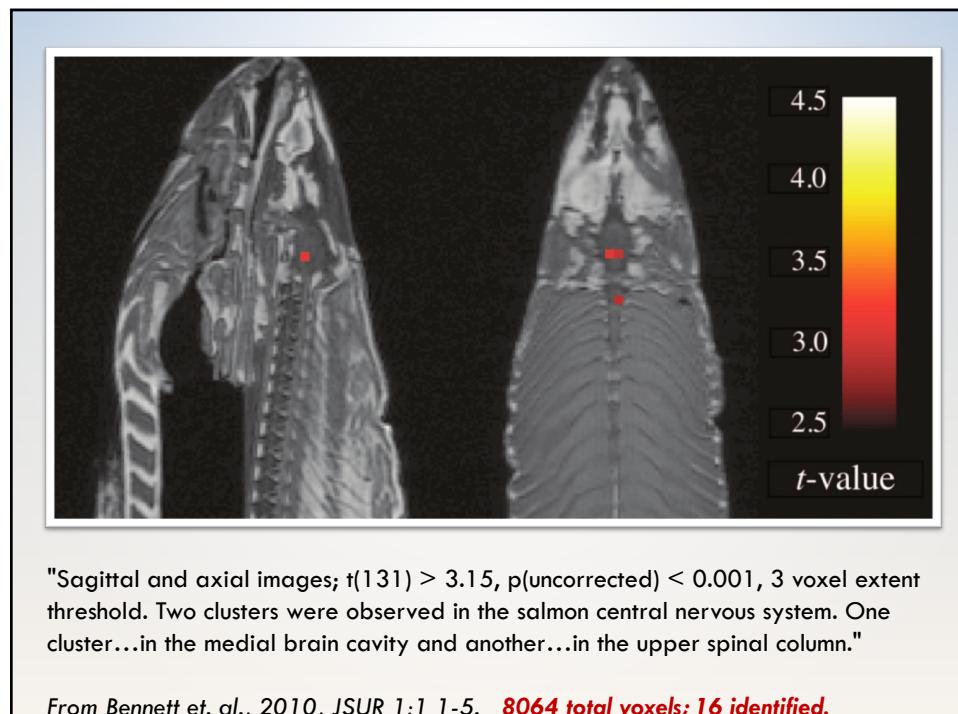
Blume and Murray, 2022



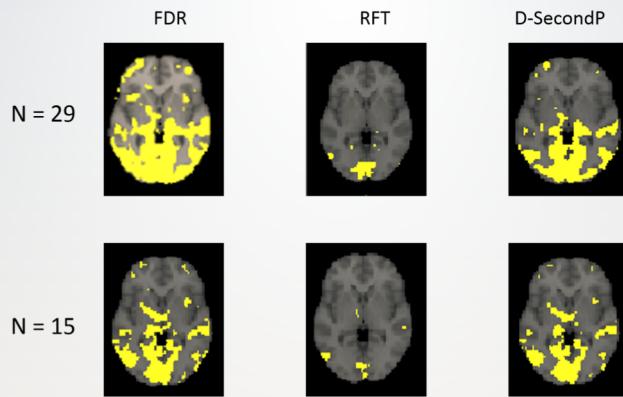
Setting interval null

- Before analyzing data (!)
- Measurement error
- Subject matter knowledge
- Impact of findings
- Community standard
- Get creative (fMRI example)
- Width not critical, buffer
- *The Atlantic salmon imaging saga*





SGPVs for Functional MRI Data



Lisa Lin Dissertation 2021: Data decimation results of activation maps for the 37th axial slice. The yellow blobs indicate activated areas resulted from FDR (first column), RFT (second column) and D-SecondP (third column). The first row corresponds to the activation maps with full sample size at 29. The second row corresponds to the activation maps with a randomly selected sample of size 15.

Blume and Murray, 2022

Time for Code Part 1b!

10 Minute Break!

ASA SGPV Short Course

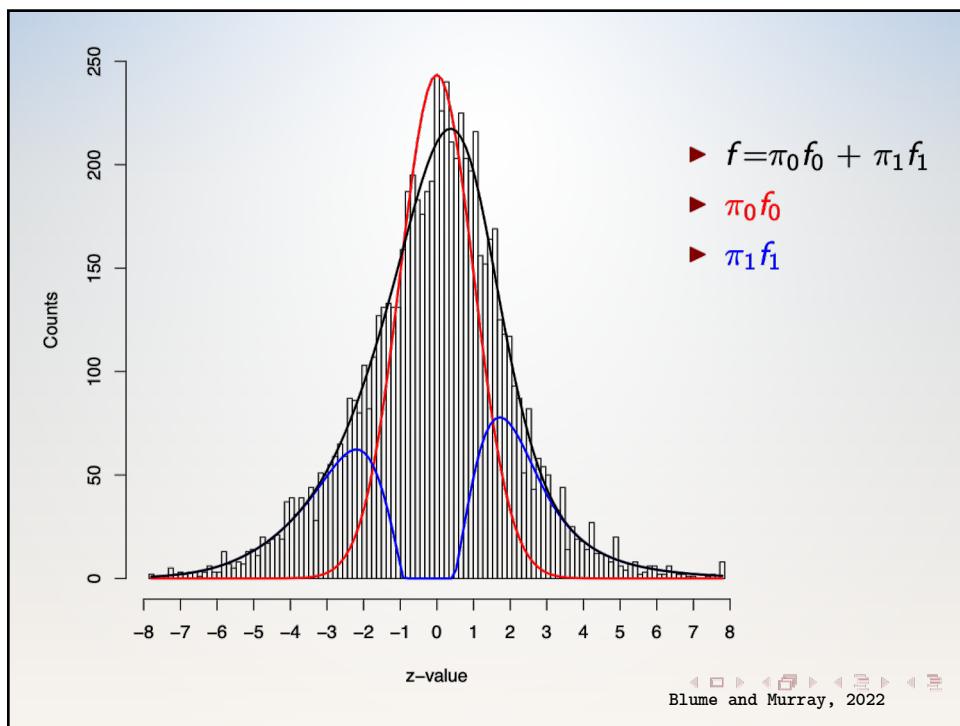
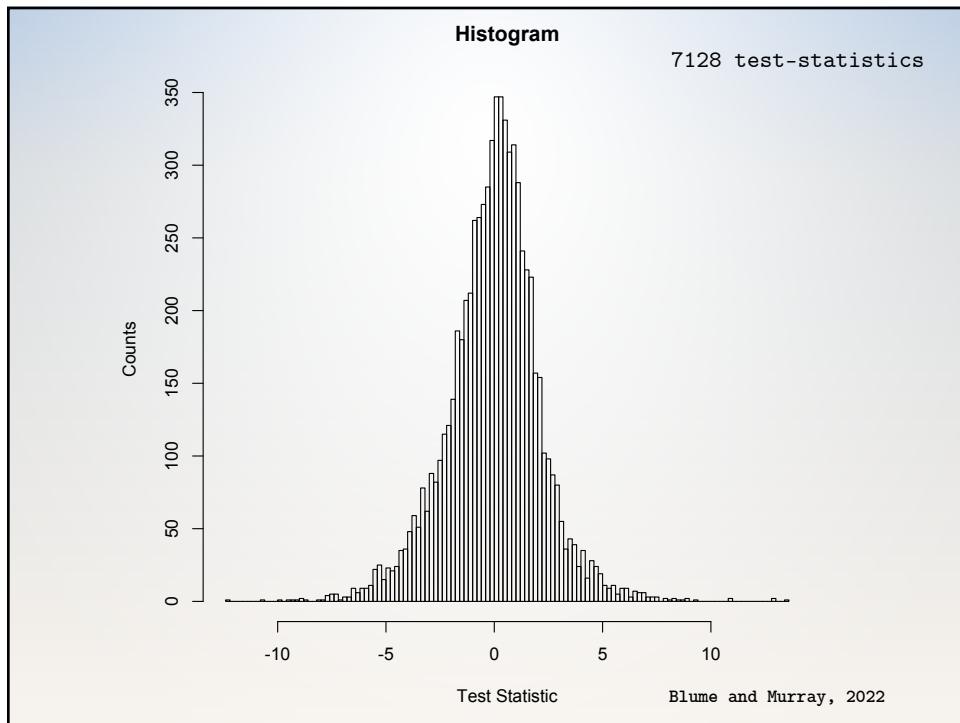
Blume and Murray, 2022

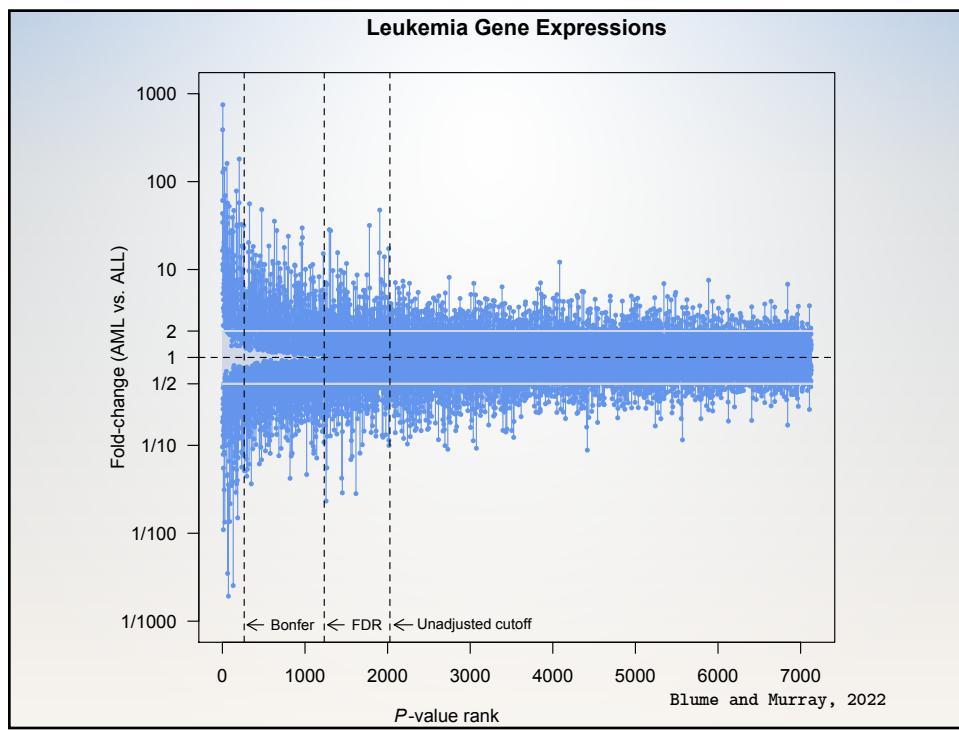
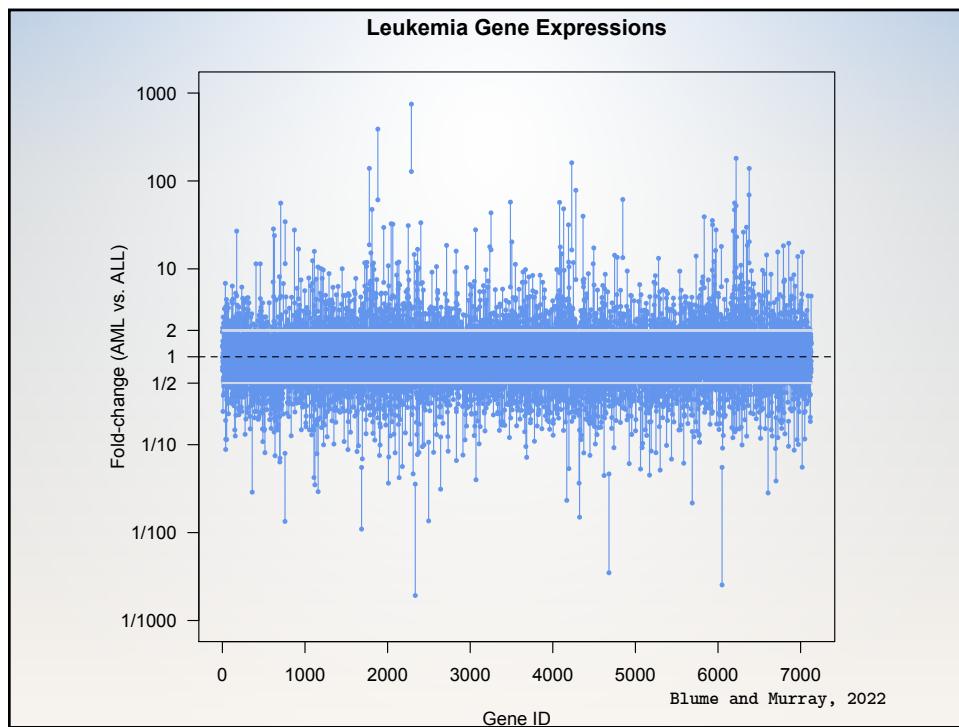
High-Dimensional Data: Leukemia gene expression

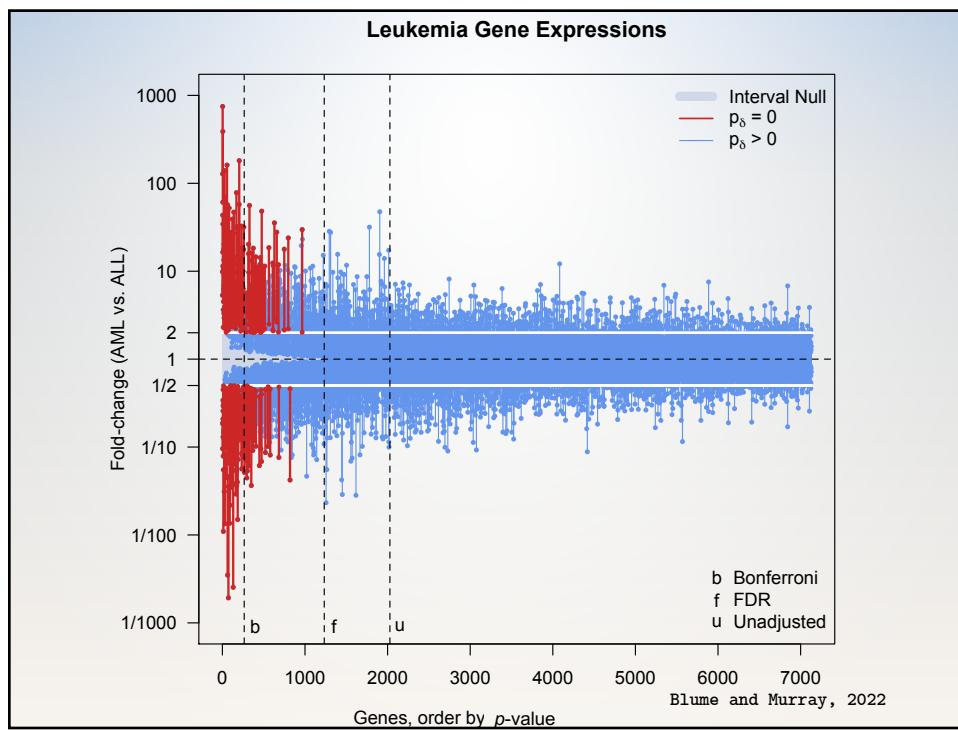
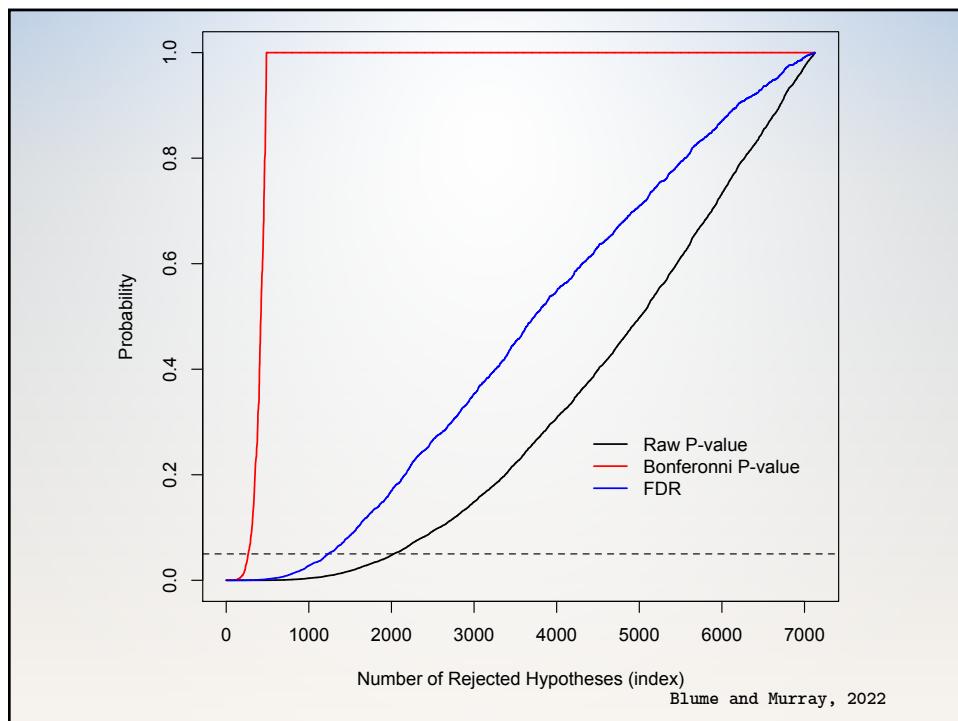
- Classifying acute leukemia by precursors
(Golub 1999, *Science*)
- 7128 genes ; 72 patients (47 ALL and 25 AML)
- Affymetrix chip collected expression levels
- Goal: Identify 'interesting' genes whose expression levels differ between All and AML subjects.
- Looking for fold changes of 2 or more

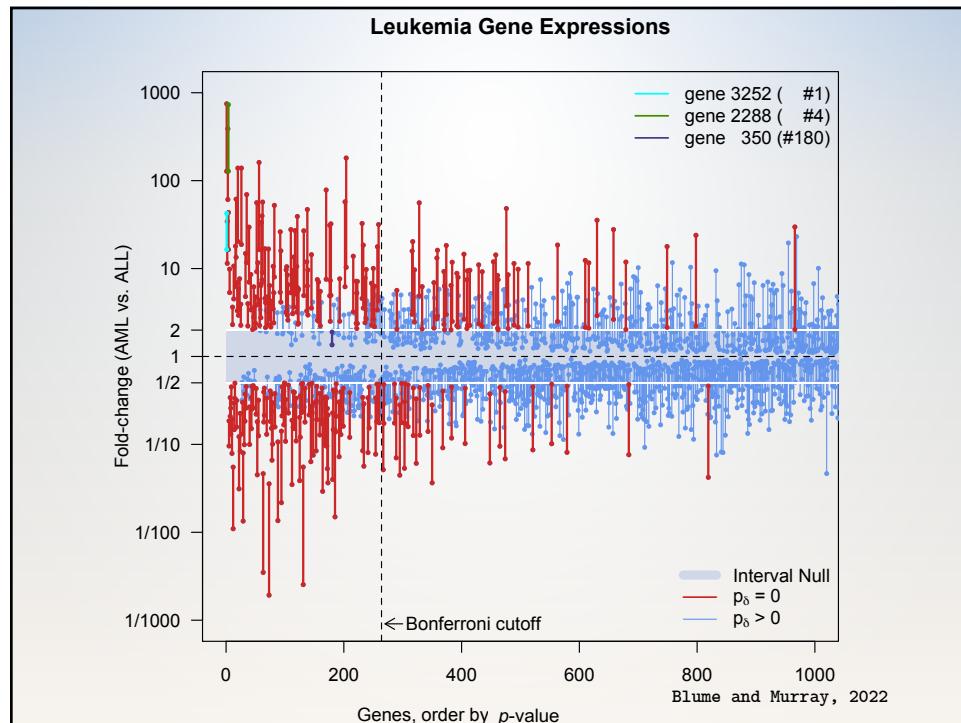
ASA SGPV Short Course

Blume and Murray, 2022









Cross-Tabulation of Leukemia Results

- Bonferroni vs ASA SGPV Short Course

	$1/2 < \text{Fold Change} < 2$ ($\delta = 0.3$)		$1/1.915 < \text{Fold Change} < 1.915$ ($\delta = 0.282$)	
	$p_\delta = 0$	$p_\delta > 0$	$p_\delta = 0$	$p_\delta > 0$
$p_{bon} < 0.05$	164	100	182	82
$p_{bon} > 0.05$	65	6799	82	6782
Total	229	6899	264	6864

Leukemia study findings

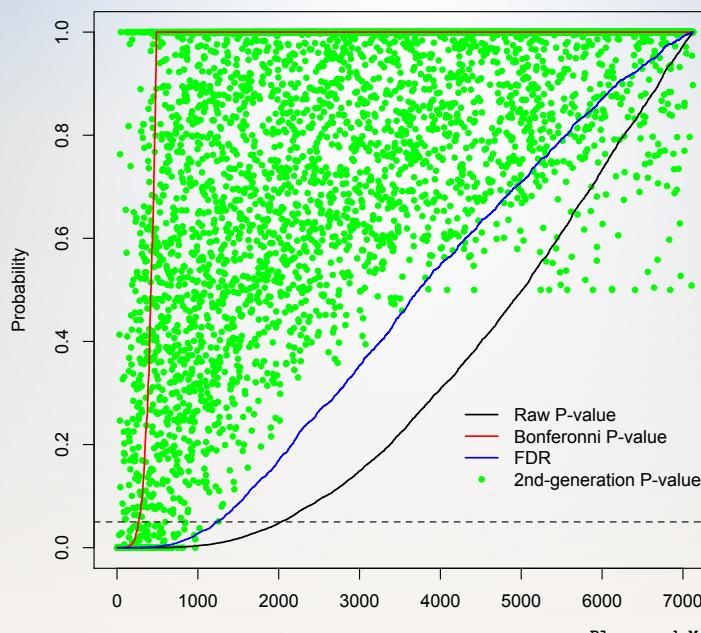
- Findings: Bonferroni 264, SGPV 229
 - Agree on 164 findings
 - Bonferroni +100, SGPV +65
- Effective Type I error rate: 0.037 vs. 0.032
- FDR of 2.45% captures all $p_\delta = 0$, 737 others
- Moving cutoff trades Type I for Type II errors
- SGPV changes the *ranking* of findings
 - Three categories now: null, alt, inconclusive
 - Null findings not illustrated here

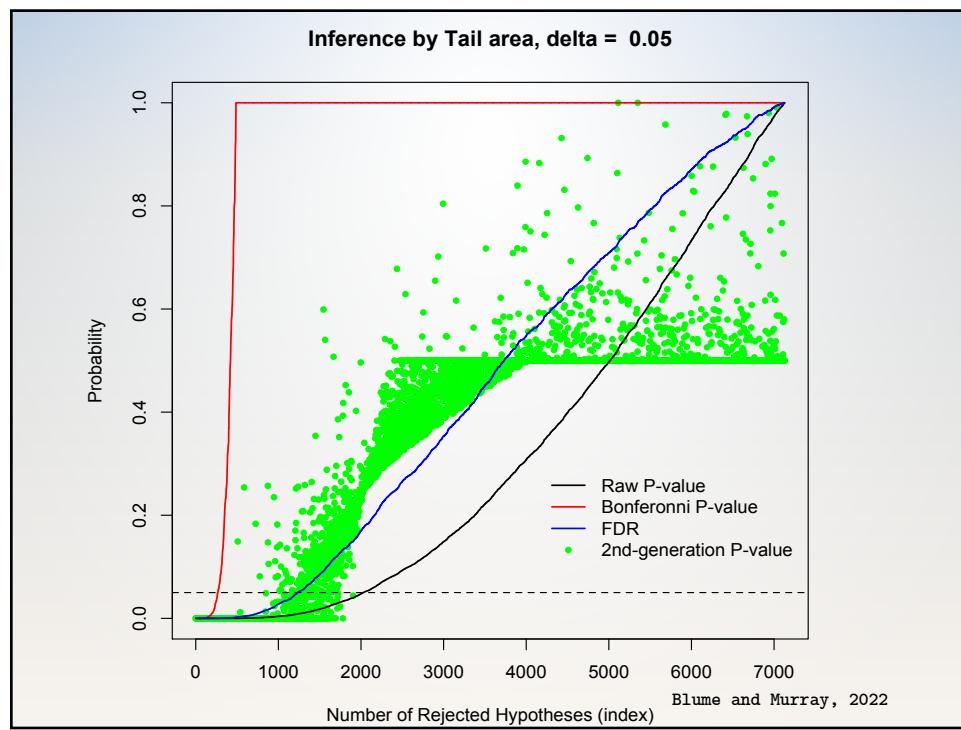
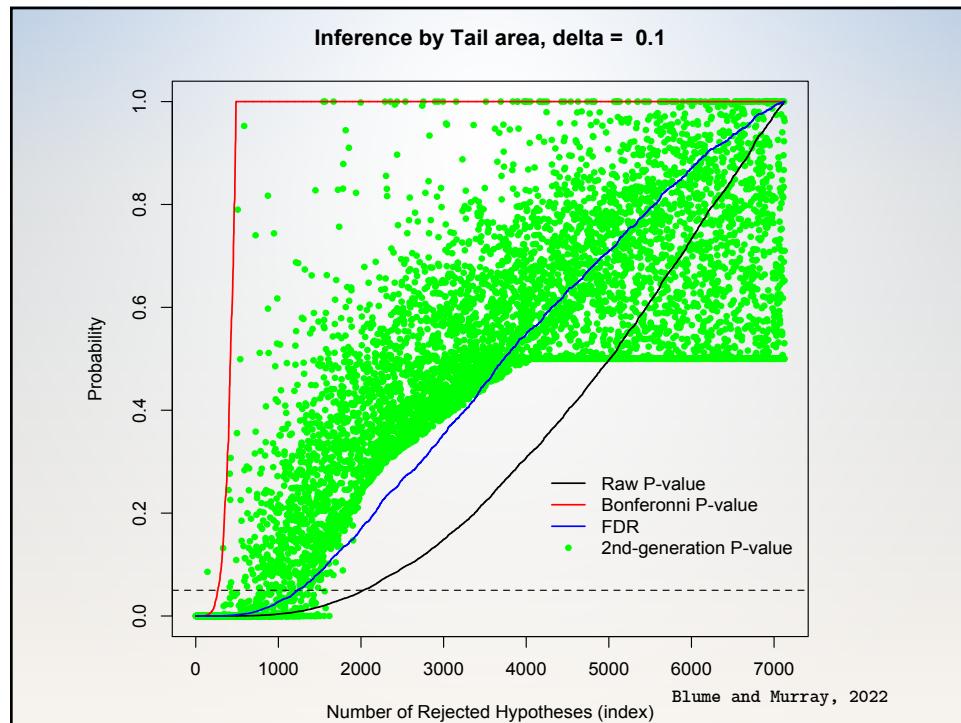
Some SGPV findings
have a priori
published validation

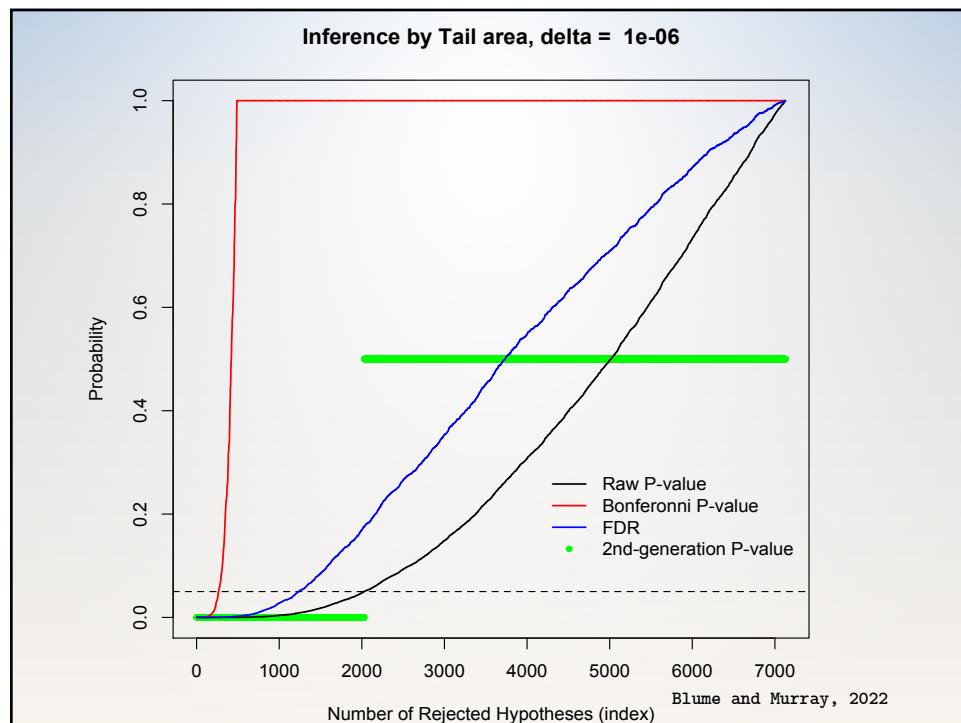
ASA SGPV Short Course

Blume and Murray, 2022

Inference by Tail area, delta = 0.3

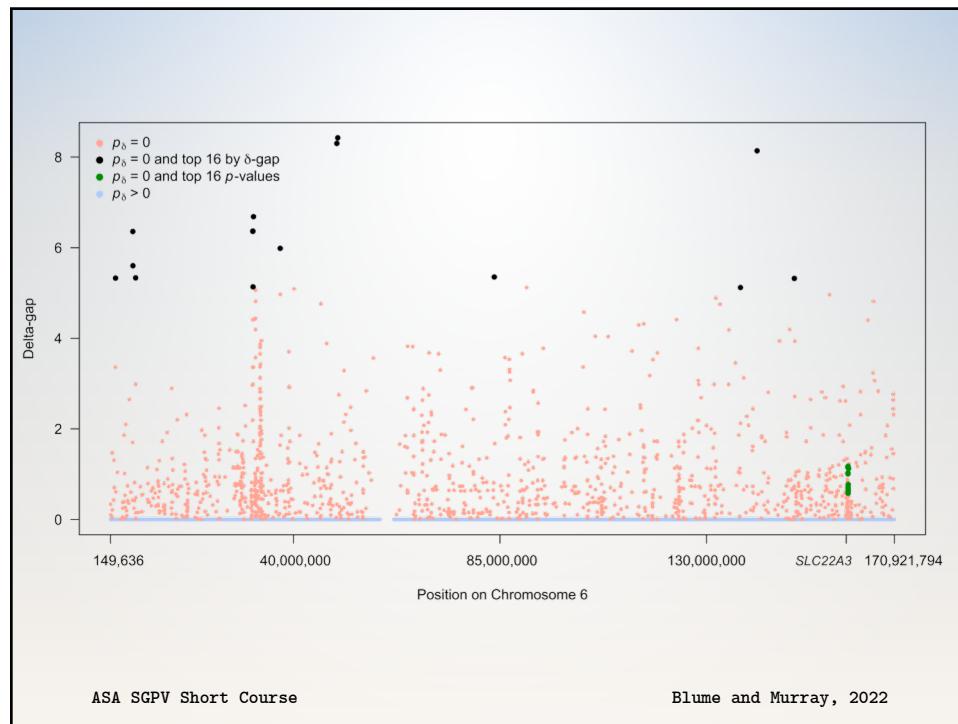
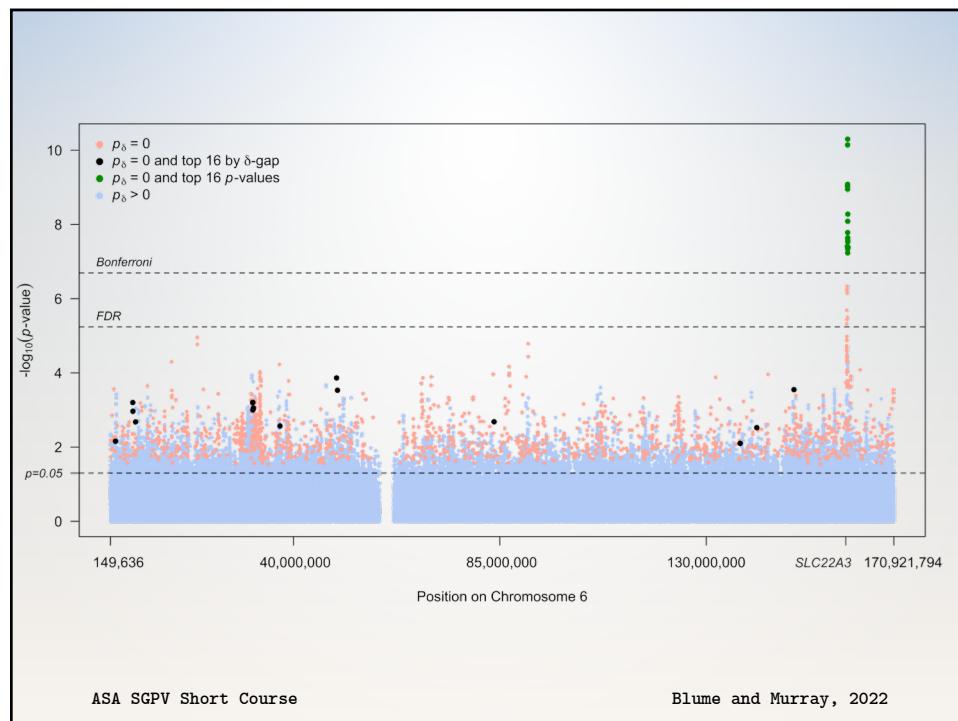


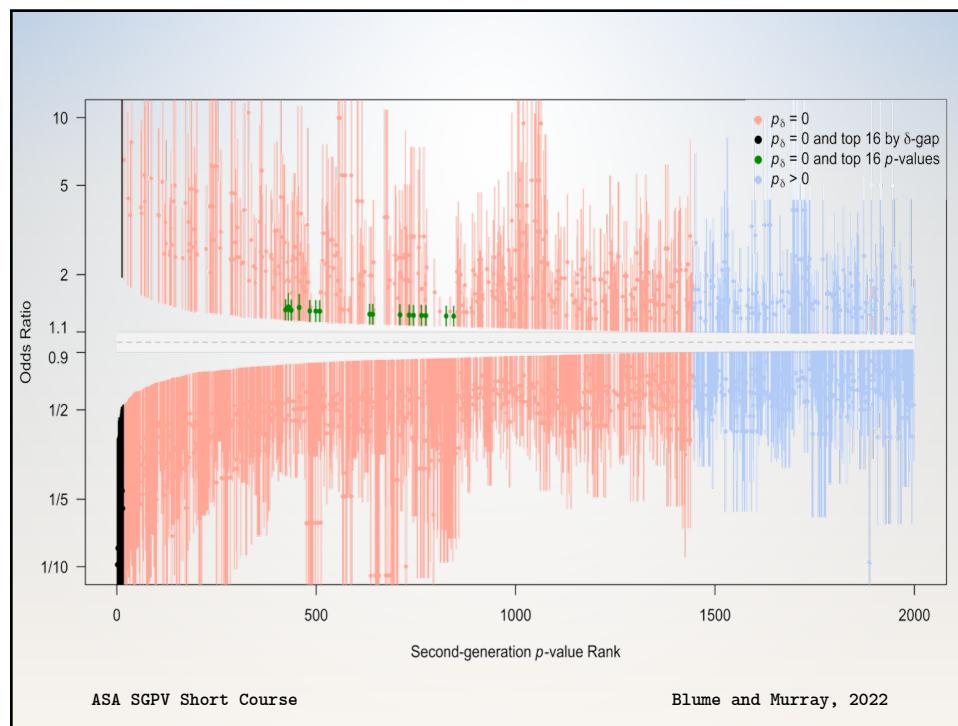
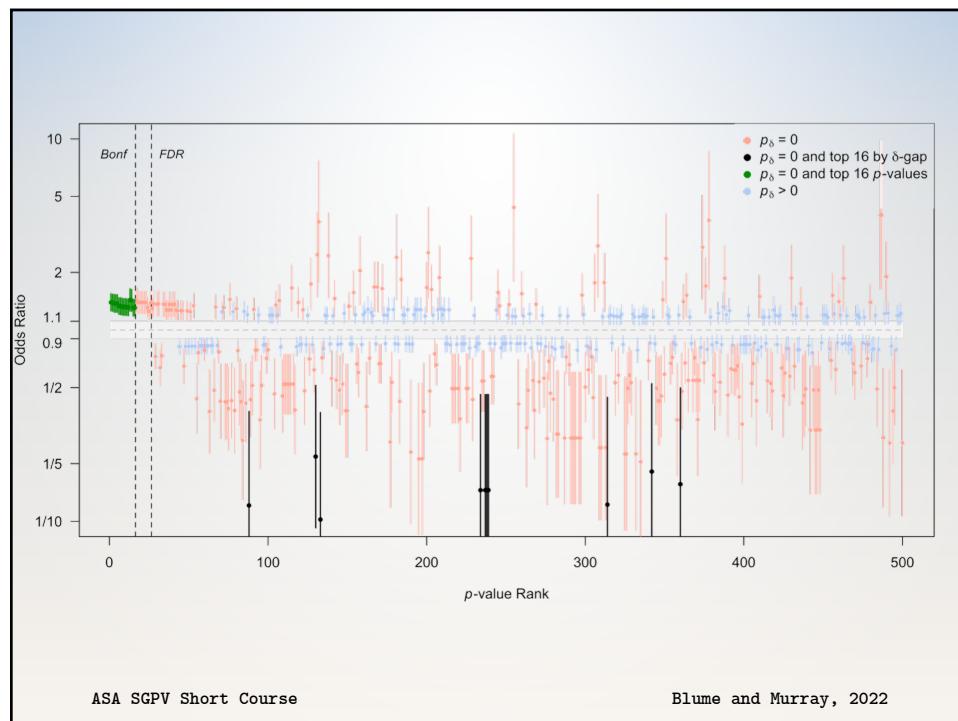


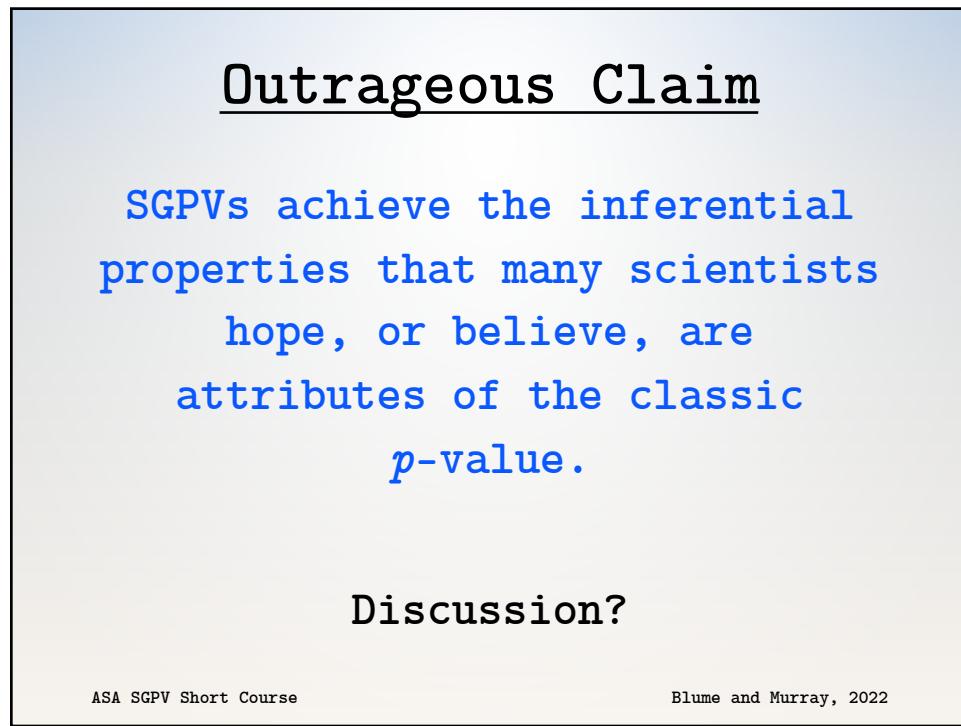
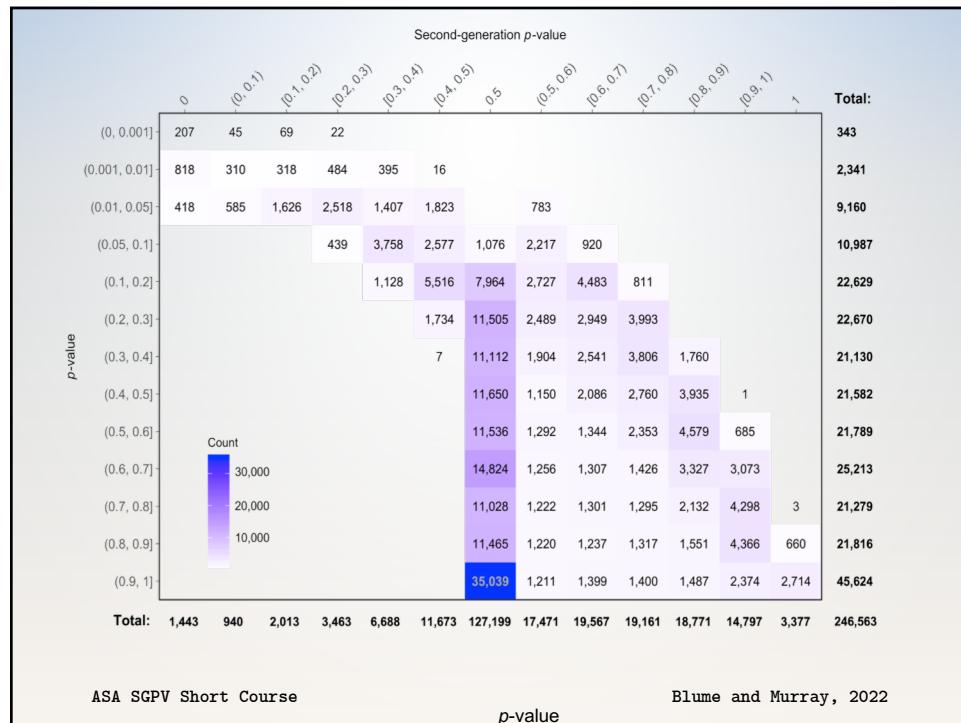


Prostate Cancer SNPs

- International Consortium for Prostate Cancer Genetics (Schaid and Chang 2055; ICPCG 2018)
- 3,894 subjects: 2,511 cases & 1,383 controls
- 247,000 single-nucleotide polymorphisms (SNPs) from Chromosome 6
- Goal: Identify 'interesting' SNPs potentially associated with prostate cancer
- Looking for odds ratios of <0.9 or >1.11







Time for Code Part 1c!

ASA SGPV Short Course

Blume and Murray, 2022

10 Minute Break!

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	$\text{SGPV } (p_\delta)$
2	Operating characteristics	$P(p_\delta = 0 H_0)$ $P(p_\delta = 1 H_1)$ $P(0 < p_\delta < 1 H)$
3	False discovery rates	$P(H_0 p_\delta = 0)$ $P(H_1 p_\delta = 1)$

ASA SGPV Short Course

Blume and Murray, 2022

Statistical Properties

Suppose interval I has coverage probability $1-\alpha$, then

Three ‘Error’ Rates

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

Will examine
these first

Two False Discovery Rates

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

Will graph to
illustrate

ASA SGPV Short Course

Blume and Murray, 2022

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

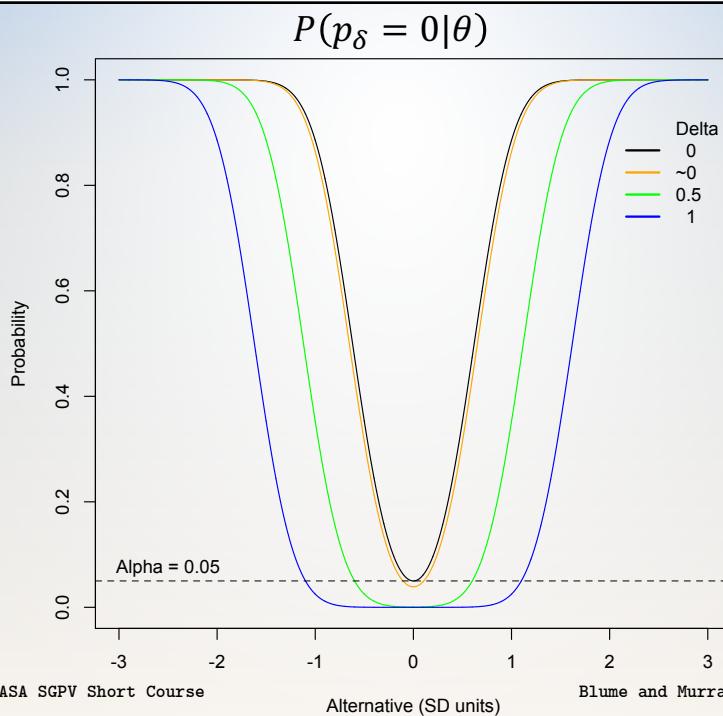
- θ_0 : point null, σ : standard deviation
- δ : half-width of indifference zone

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

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

ASA SGPV Short Course

Blume and Murray, 2022



ASA SGPV Short Course

Alternative (SD units)

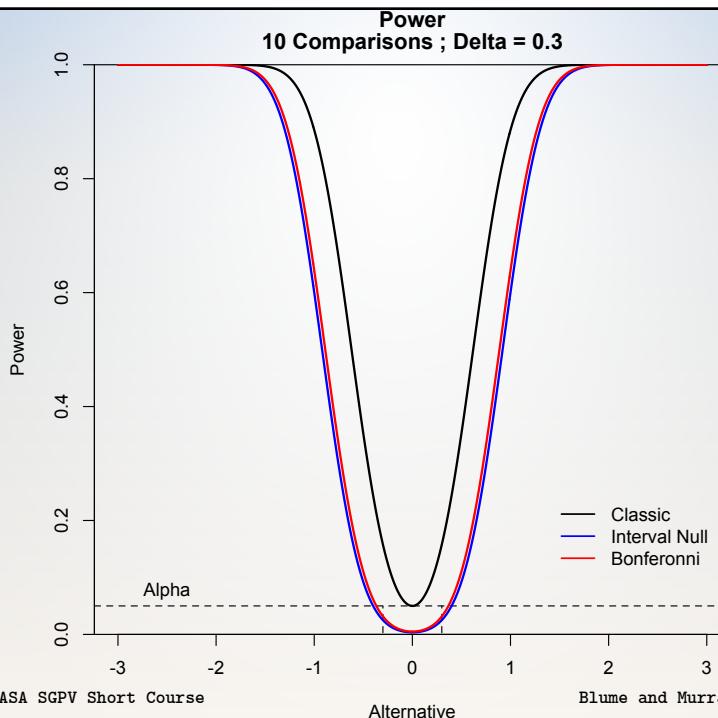
Blume and Murray, 2022

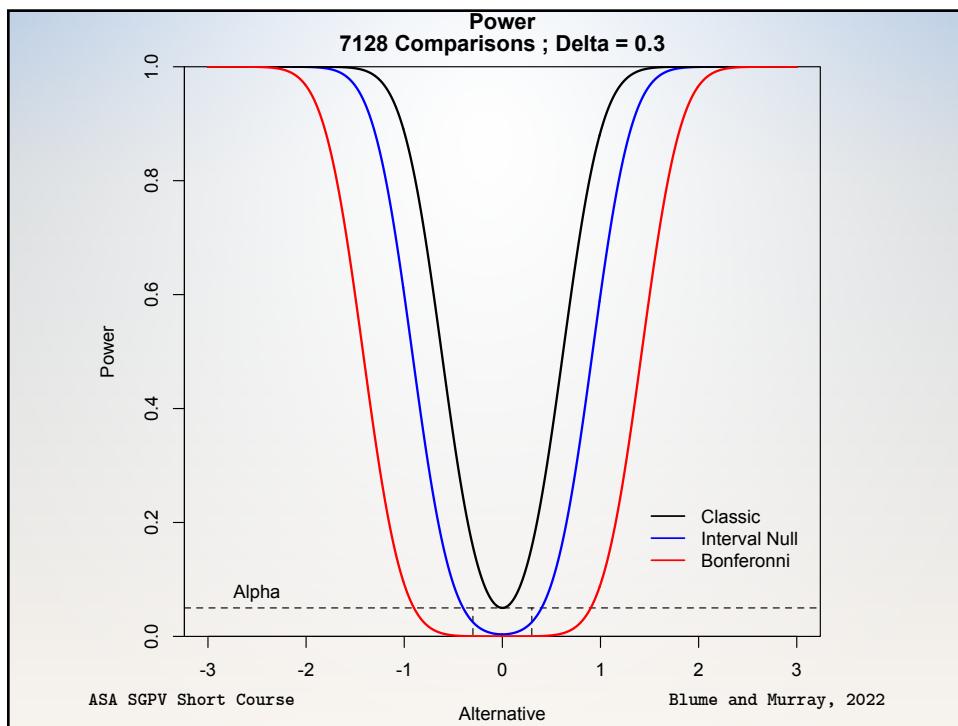
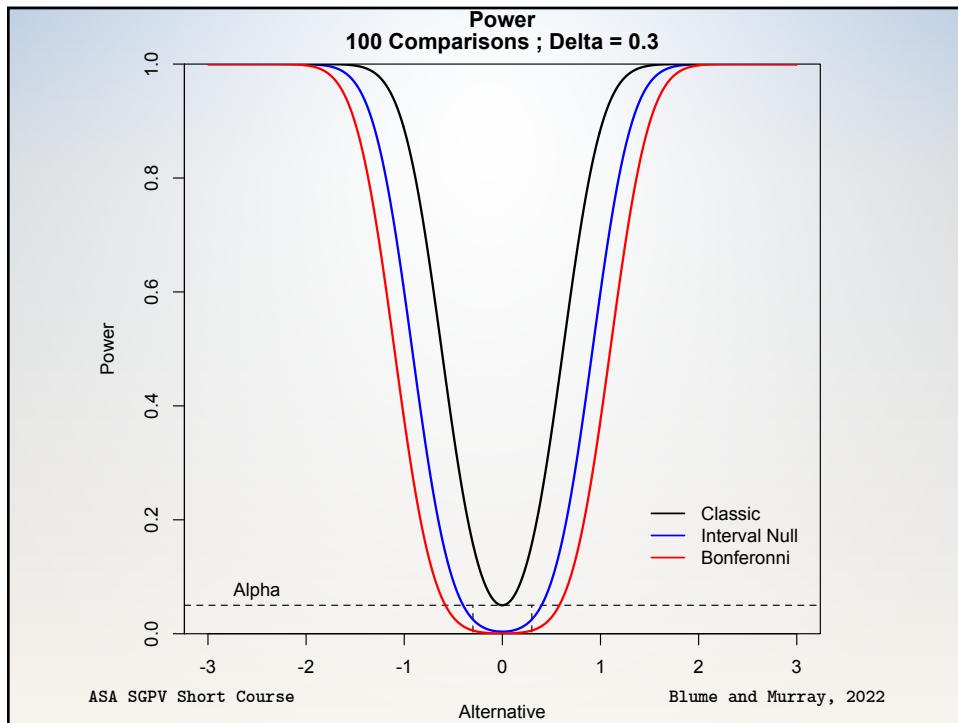
Compare with standard methods

- Second-generation p -value vs. Bonferroni correction
 - Adjusted for $\{10, 100, 7128\}$ comparisons
 - Leukemia data example
- Remember SGPV are not adjusted for comparisons
- Discuss?

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

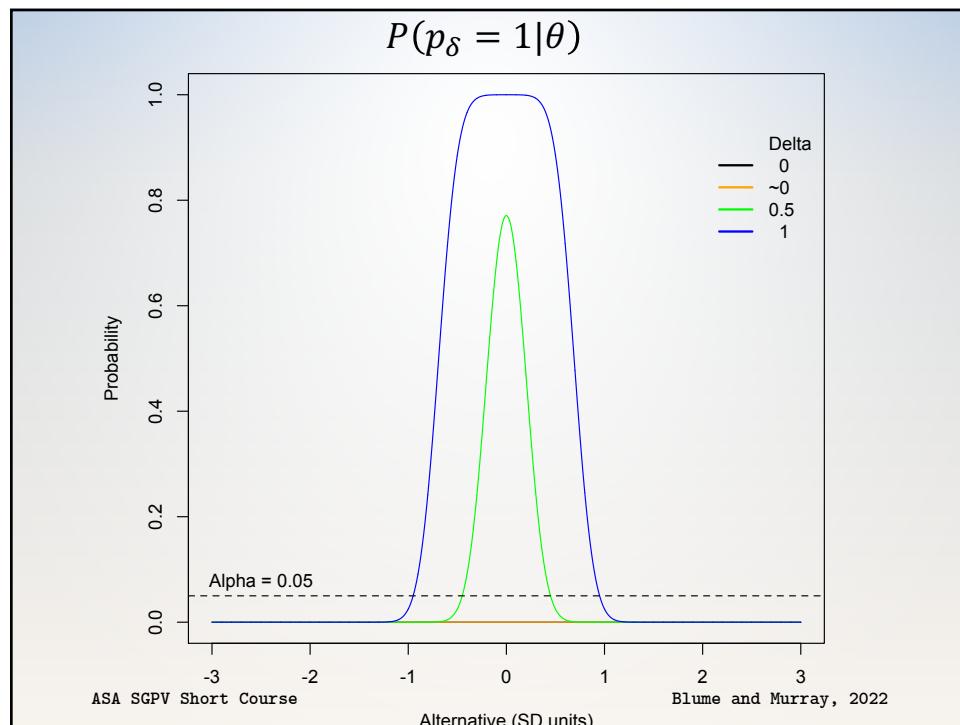
- How often are the data compatible with null?
- Sample size must be large enough to allow null interval to contain the interval estimate, so $(\delta > Z_{\alpha/2}/\sqrt{n})$ or $(\sqrt{n} > Z_{\alpha/2}/\delta)$
- This probability converges to 0 or 1 quickly

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

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

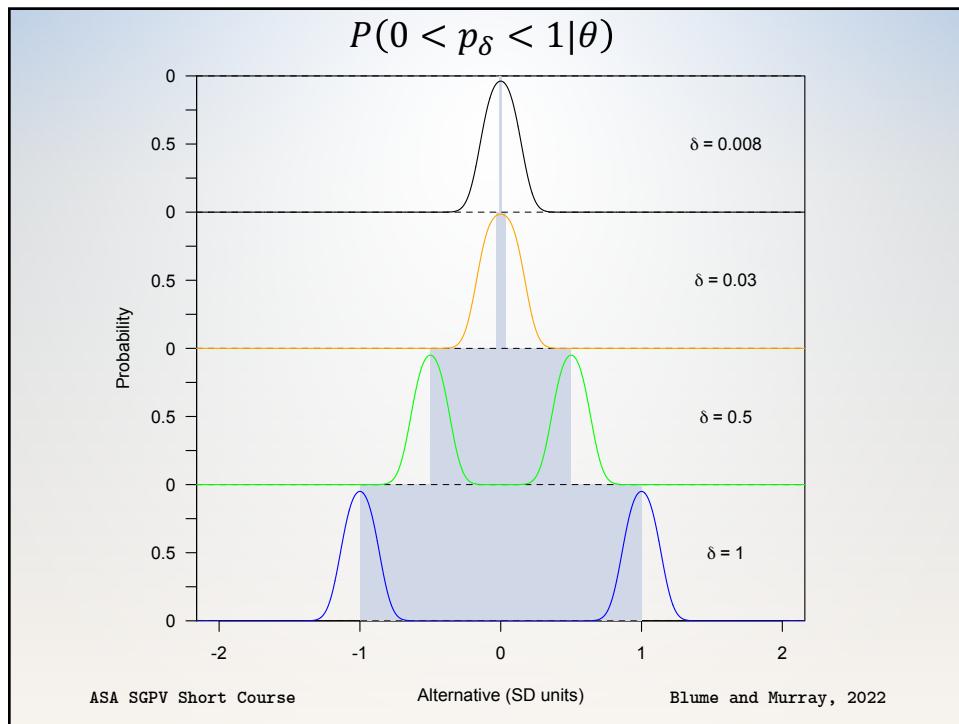
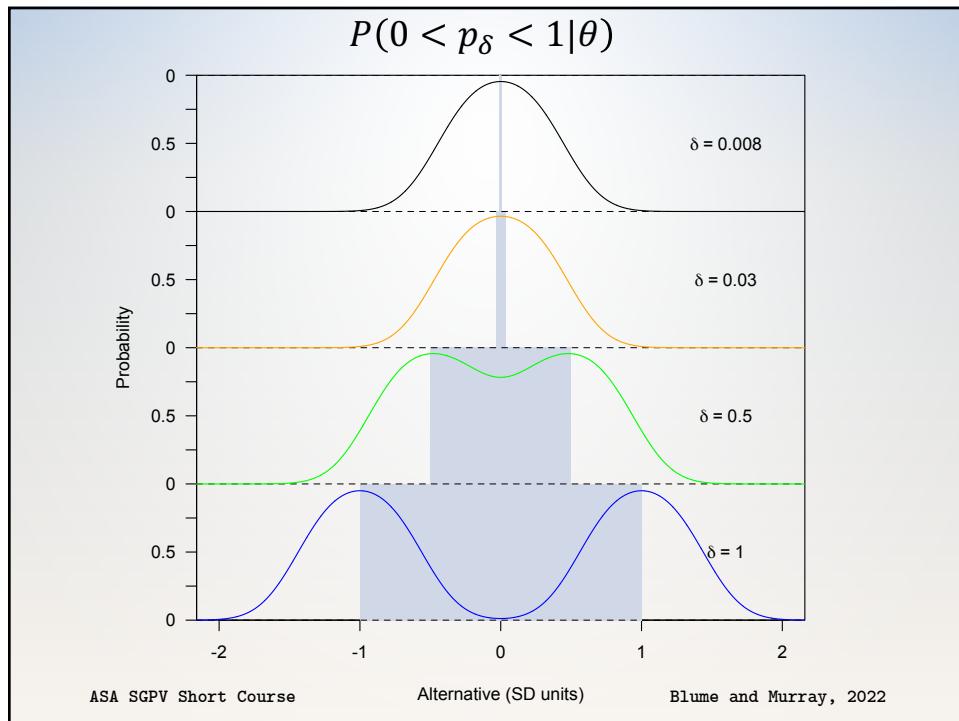
ASA SGPV Short Course

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
- $P(0 < p_\delta < 1|\theta) = 1 - P(p_\delta = 0|\theta) - P(p_\delta = 1|\theta)$



Statistical Properties

Suppose interval I has coverage probability $1-\alpha$, then

Three ‘Error’ Rates

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

Two False Discovery Rates (next section!)

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

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 (next section)
- Anchoring the scale of the effect size...
 - Eliminates most Type I Errors
 - Improves scientific translation of statistical model

Time for Code Part 1d!