

Second-Generation p -values: Statistical Properties

Jeffrey D. Blume, PhD

Departments of Biostatistics,
Biomedical Informatics

VANDERBILT  UNIVERSITY

Outline

- Review
 - Second-generation p -value definition
 - Evidential Metrics
 - Frequentist Properties
 - False Discovery Rates
 - Prostate Cancer SNP data (~247,000)
 - FDR example: False Discovery / Confirmation Rates
 - Leukemia Data
 - $\alpha=0.05$ vs $\alpha=0.05/7128$ vs SG p -value



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 $|I| > 2|H_0|$

Second-generation p-values

Blume, Nov 2019

Evidential metrics

1. **Measure of the strength evidence**
 - Axiomatic and intuitive justification
 - Summary statistic, yardstick
2. **Propensity to collect data that will yield a misleading #1**
 - Error rates
 - Properties of the study design (!)
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 (!)

Second-generation p-values

Blume, Nov 2019

Likelihood

Law of Likelihood: The hypothesis that does a better job at predicting the observed events is better supported by the data.

| Evidential Metric | What it measures | Likelihood |
|-------------------|---|--|
| 1 | strength of the evidence | Likelihood Ratio (LR) |
| 2 | propensity for study to yield misleading evidence | $P(LR > k H_0)$ $P(LR < 1/k H_1)$ |
| 3 | propensity for observed results to be misleading | $P(H_0 LR = k)$ $P(H_1 LR = 1/k)$ |

Second-generation p-values

Blume, Nov 2019

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

Second-generation p-values

Blume, Nov 2019

Second-generation p -value

- StatisticalEvidence.com
- 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 | $P(H_0 p_\delta = 0)$ $P(H_1 p_\delta = 1)$ |

Second-generation p -values

Blume, Nov 2019

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

Second-generation p -values

Blume, Nov 2019

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

Second-generation p-values

Blume, Nov 2019

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

Second-generation p-values

Blume, Nov 2019

'Power' Function

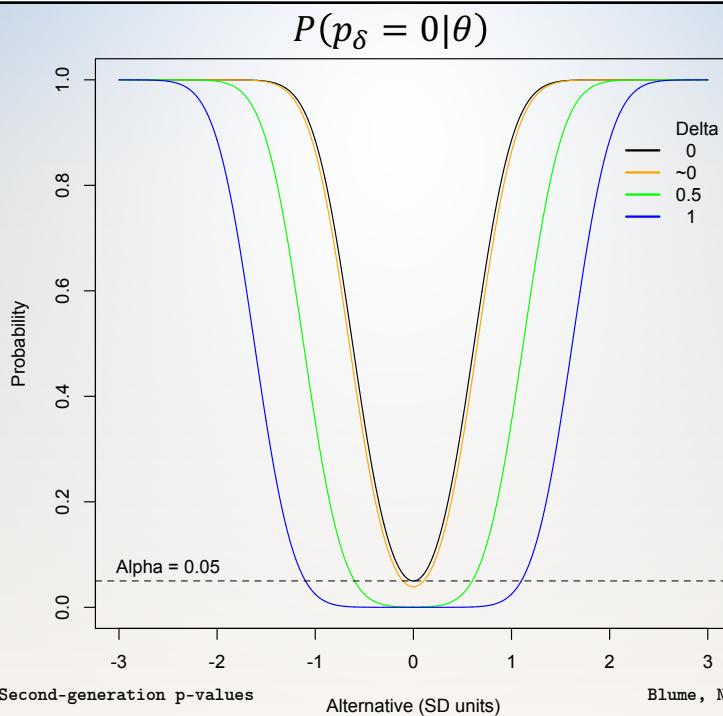
- θ_0 : point null, σ : standard deviation
- δ : half-width of indifference zone

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

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

Second-generation p-values

Blume, Nov 2019

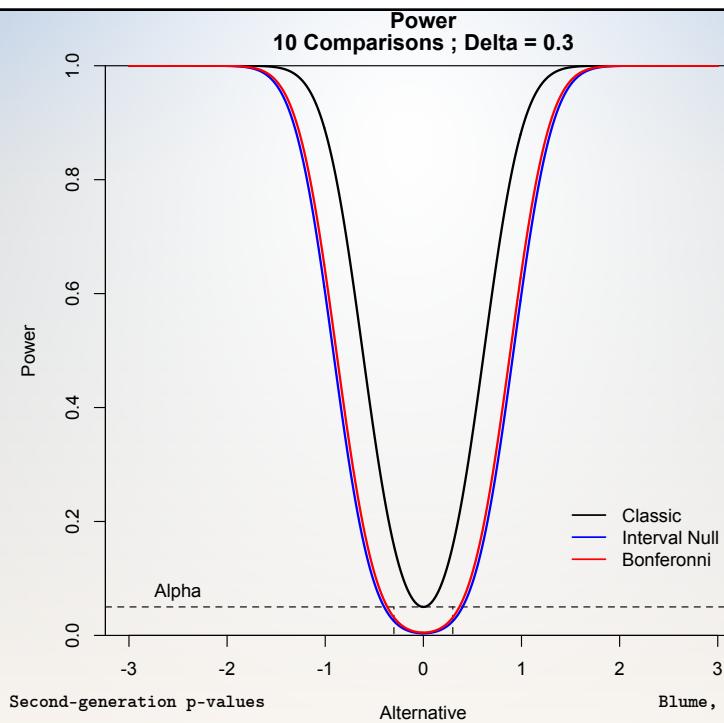


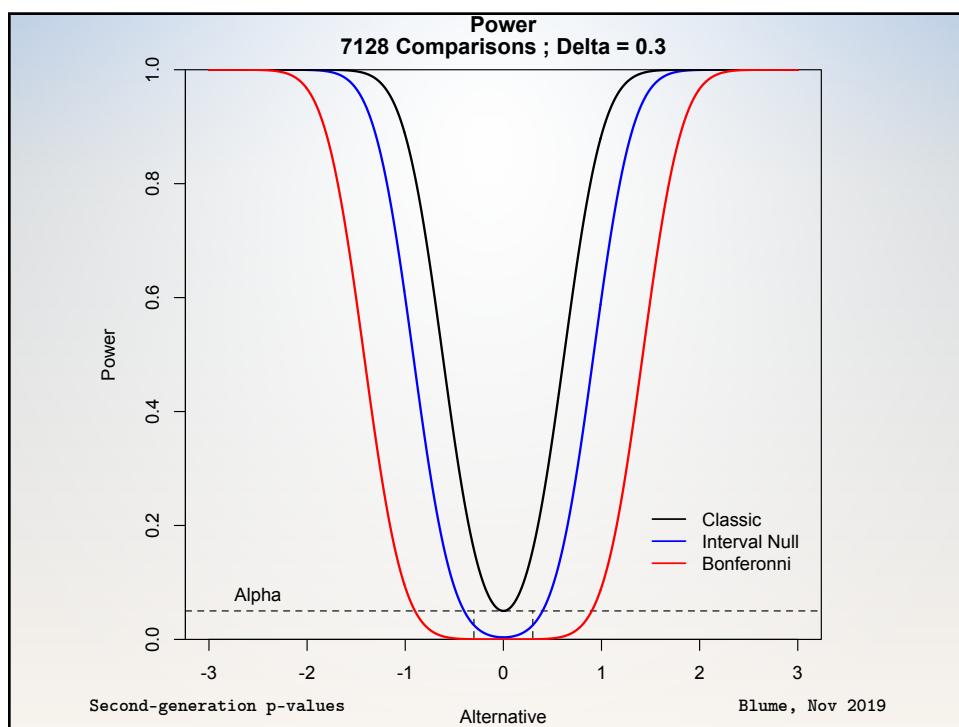
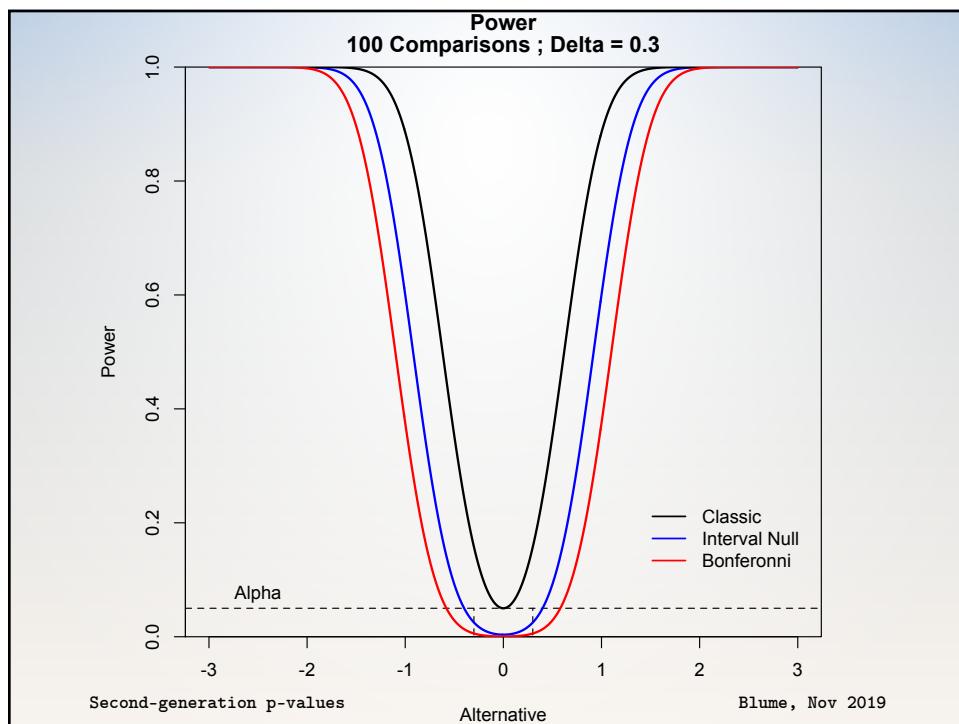
Compare with standard methods

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

Second-generation p-values

Blume, Nov 2019





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

Second-generation p-values

Blume, Nov 2019

‘Null Power’ Function

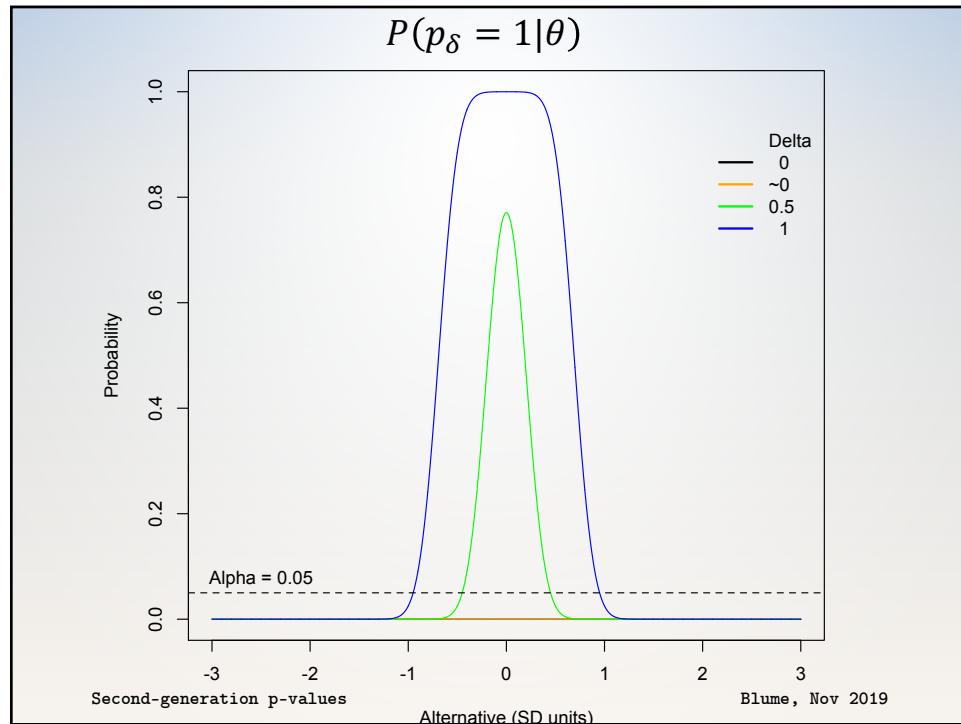
- How often are the data compatible with null?
- Sample size must be large enough to allow null interval to contain the interval estimate, so $(\delta > Z_{\alpha/2}/\sqrt{n})$ or $(\sqrt{n} > Z_{\alpha/2}/\delta)$
- This probability converges to 0 or 1 quickly

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

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

Second-generation p-values

Blume, Nov 2019

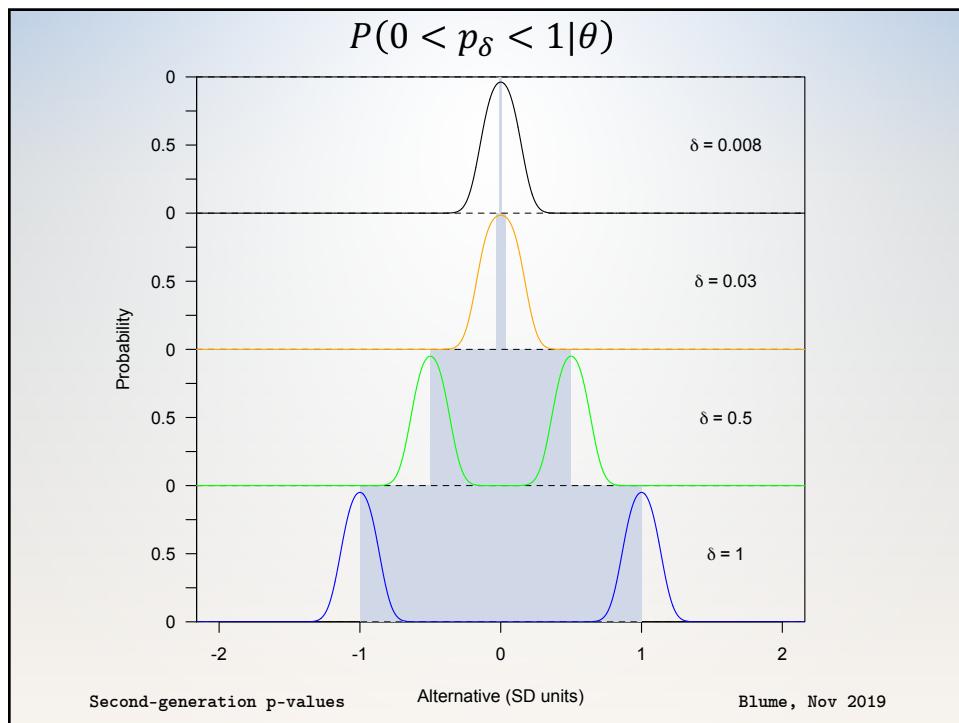
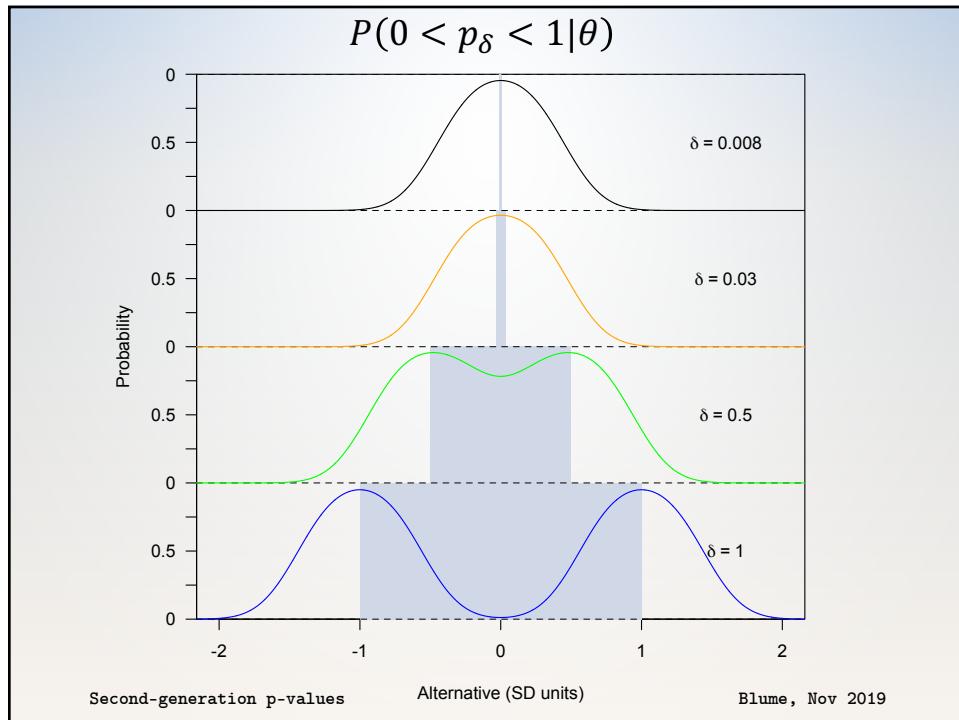


Probability of Inconclusive Data

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

Second-generation p-values

Blume, Nov 2019



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

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

Second-generation p-values

Blume, Nov 2019

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% |

Second-generation p-values

Blume, Nov 2019

False discovery rates

- Impact of $\alpha=0.05$ vs $\alpha=0.05/7128$ (7128 comparisons)

- False Discovery Rate (**FDR**)

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

- False Confirmation Rate (**FCR**)

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

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

Second-generation p-values

Blume, Nov 2019

False discovery rates

- Second-generation *p*-values

- False Discovery Rate (**FDR**)

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

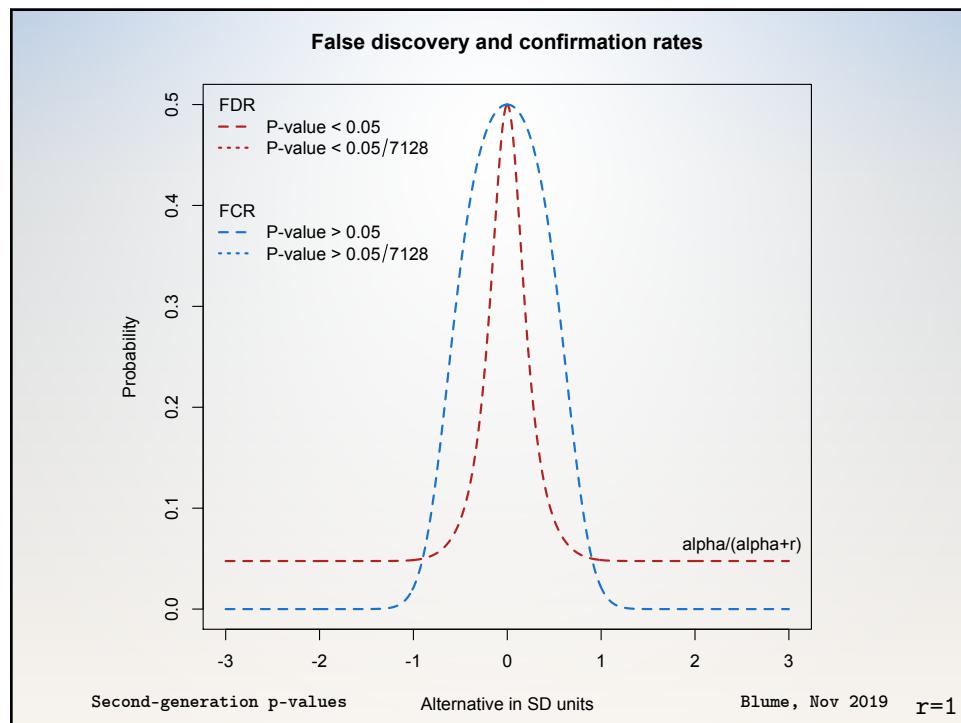
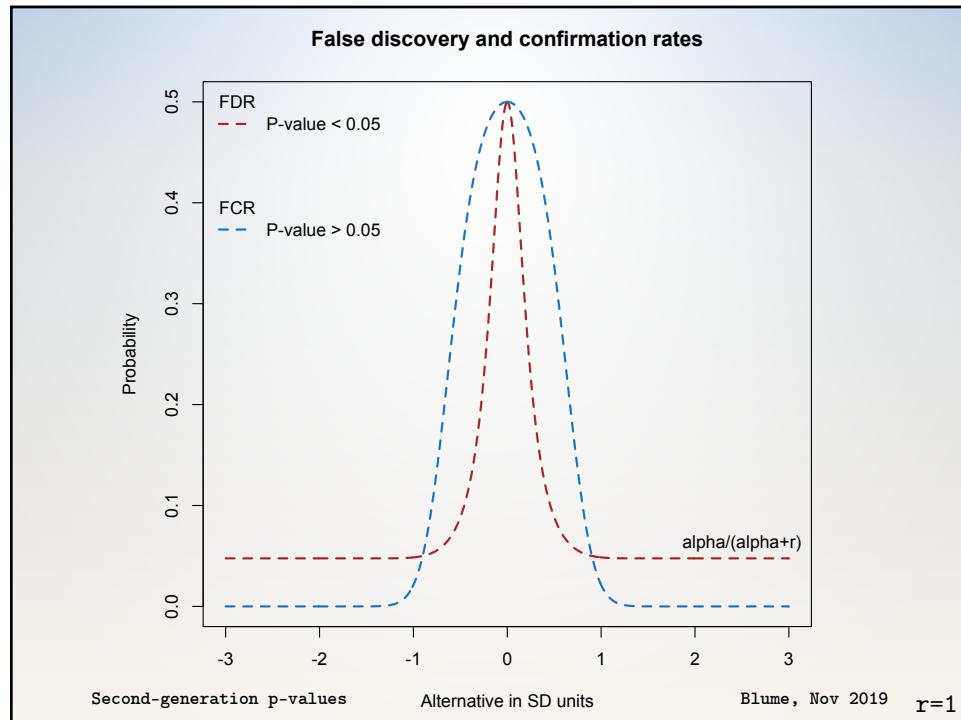
- False Confirmation Rate (**FCR**)

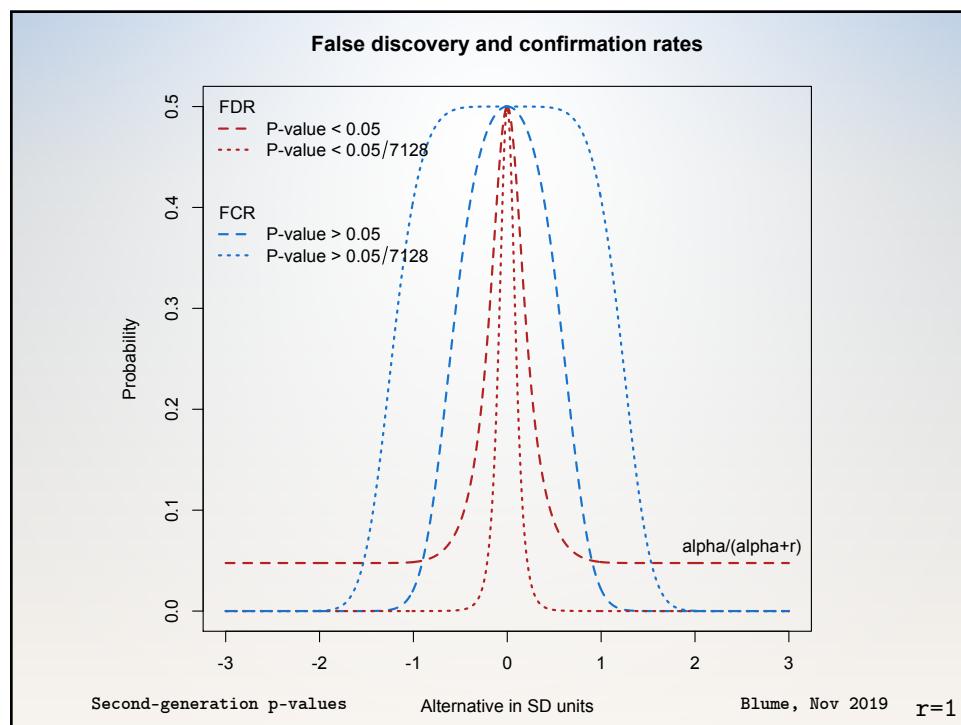
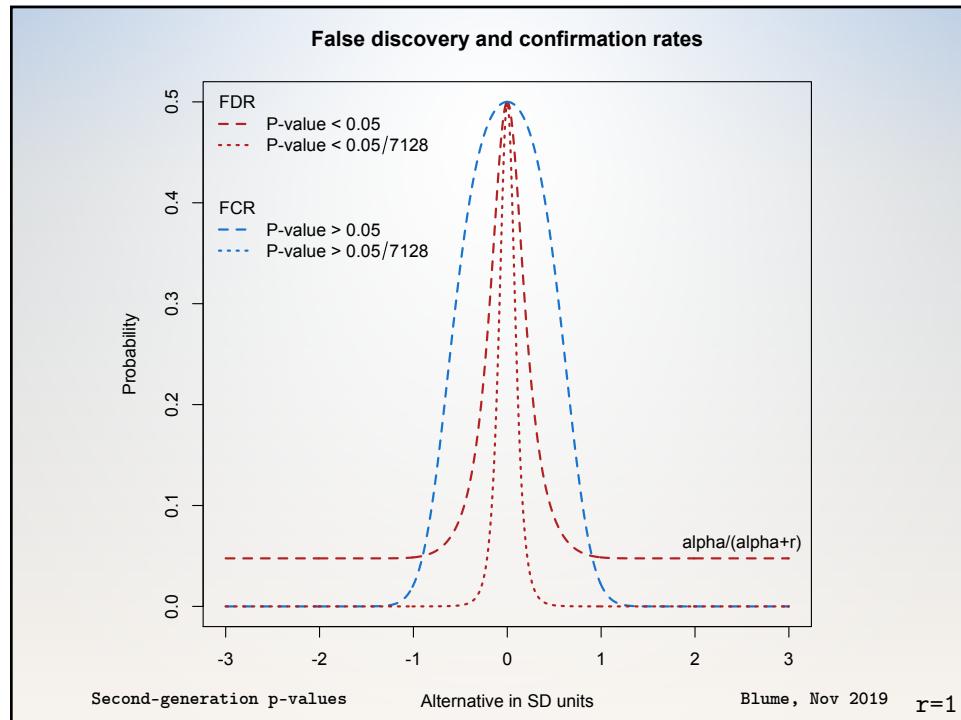
$$P(H_1|p_\delta = 1) = \left[1 + \frac{P(p_\delta = 1|H_0)}{P(p_\delta = 1|H_1)} \frac{1}{r} \right]^{-1}$$

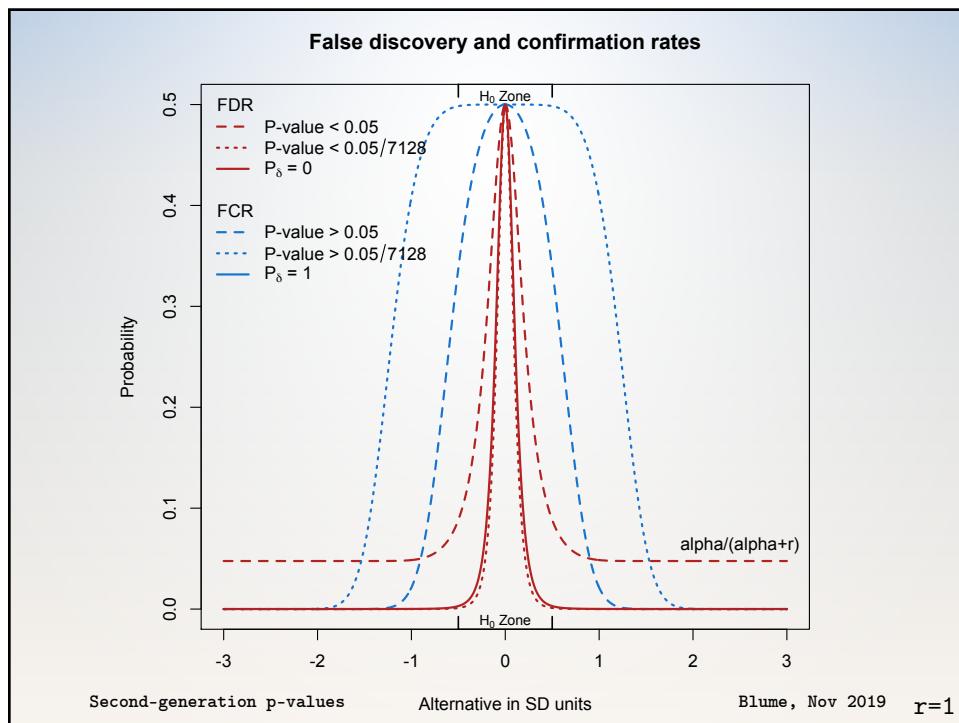
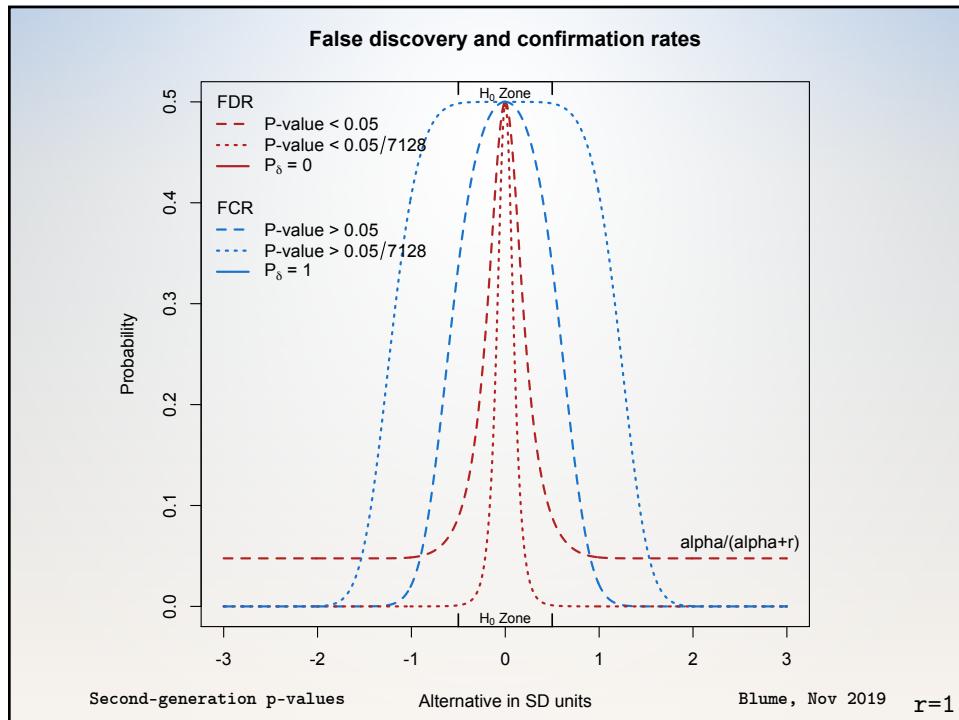
$$r = P(H_1)/P(H_0)$$

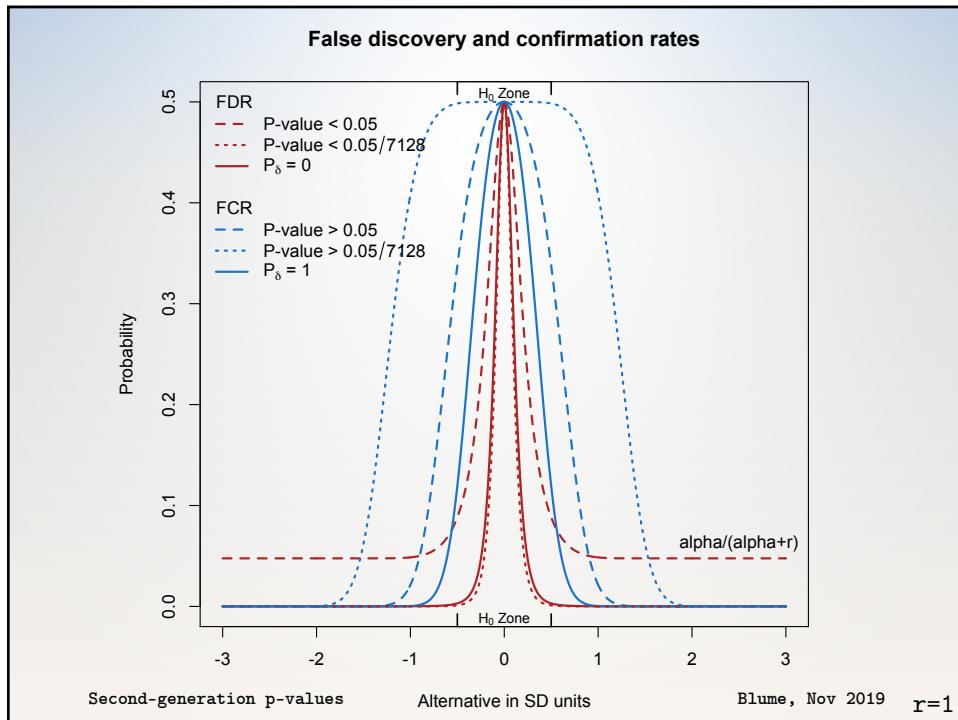
Second-generation p-values

Blume, Nov 2019









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

Second-generation p-values

Blume, Nov 2019

Acknowledgements

- Collaborators

- William D. Dupont
- Robert A. Greevy
- Lucy D'Agostino McGowan
- Valerie Welty
- Jeffrey R. Smith

- Website / Papers / Code

- statisticalevidence.com
- PLOS One ; TAS (In Press)
- Google “Second-Generation *p*-value”
- devtools::install_github("weltybiostat/sgpv")

Outrageous Claim (!?)

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

Second-generation *p*-values

Blume, Nov 2019