Bayesian Statistics
Bayesian Hypothesis Testing

Michael Anderson, PhD
Hélène Carabin, DVM, PhD

Department of Biostatistics and Epidemiology
The University of Oklahoma Health Sciences Center

May 20, 2016

Outline

Null Hypothesis Significance Testing (NHST)

In frequentist analyses, hypothesis testing involves Null Hypothesis Significance Testing (NHST).
How it works:

- Establish Null and Alternative Hypotheses.
- Observe data collected to test these hypotheses.
- Determine if the Null hypotheses is rejected or not.
- Interpret this in terms of the parameter.

This has at least the following shortcomings.

- What do we conclude if the Null is NOT rejected?
- Reject/Fail to reject decision gives no indication of uncertainty.
- Ignore prior knowledge of the subject matter.
- P-values/CI depend on researcher's intention.
- Multiple comparisons require corrections that depend on researcher's intention.

Null Hypothesis Significance Testing (NHST)

P-value

The probability of observing the data $\mathbf{x}$, or anything more extreme, given the Null Hypothesis, H_0, is true (ie $P(\mathbf{x}|H_0)$).

Bayesian P-value

The probability the Null Hypothesis, H_0, is true given the data, $\mathbf{x}$ (ie $P(H_0|\mathbf{x})$).

Bayesian Approach to Hypothesis Testing

Hypotheses can be examined in view of the posterior distribution.
Advantages to NHST:

- Direct interpretation of the believability of θ.
- Does not depend on researcher's intention.
- Is responsive to the analyst's prior beliefs.
- Multiple comparisons don't present a problem.

Approaches for Bayesian Hypothesis Testing.

- One-Tailed Hypothesis.
- Two-Tailed Hypothesis.
  - 95% Highest Density Interval.
  - Region of Practical Equivalence (ROPE).

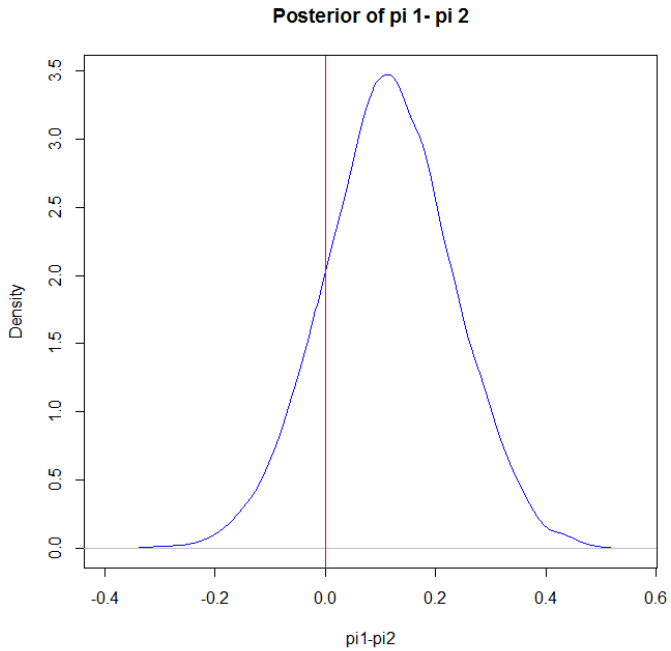
One-Tailed Hypothesis Test

- $H_0 : \pi_1 - \pi_2 \leq 0$ vs. $H_1 : \pi_1 - \pi_2 > 0$
- Obtain the posterior of $\pi_1 - \pi_2$.
- Compute $P(H_0|data)$ from the posterior.
- (i.e. Compute $P(\pi_1 - \pi_2 \leq 0|data)$)

Example

- 48 subjects with cancer are observed at 2 locations (24 from each location).
- Location 1 has 7 females and location 2 has 4 females.
- Does the proportion of females differ between the two locations?

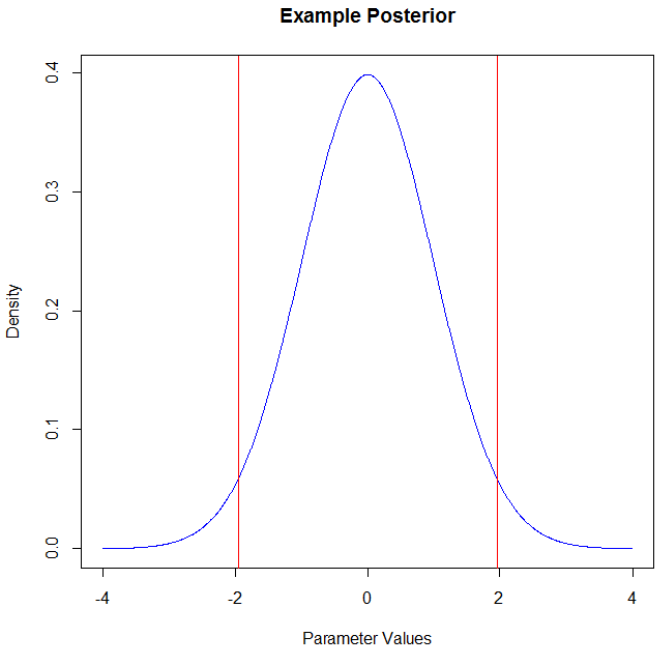
One-Tailed Hypothesis Test Example

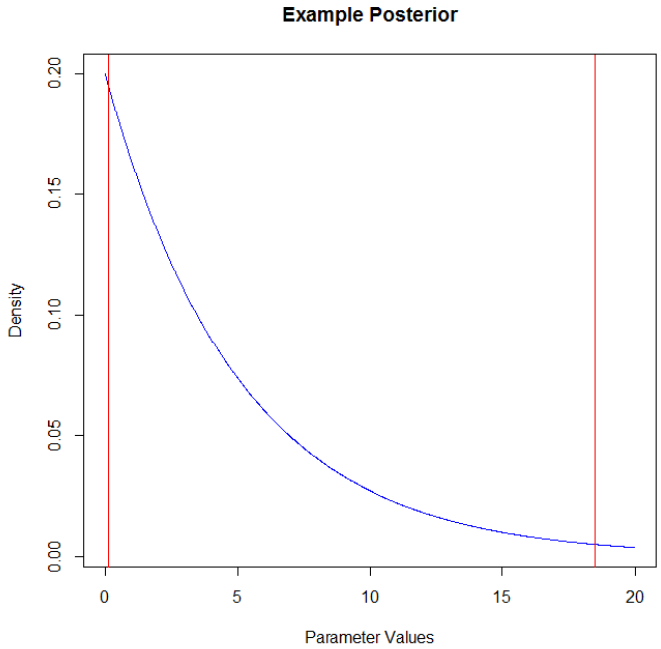


- $P(\pi_1 - \pi_2 \leq 0 | data) = 0.1622$
- Compare to a NHST p-value of 0.3133.

Two-Tailed Hypothesis Test

- $H_0 : \pi_1 - \pi_2 = 0$ vs. $H_1 : \pi_1 - \pi_2 \neq 0$
- One may obtain a 95% Bayesian Credible Interval (BCI) from the posterior.
- This is simply the 2.5% and 97.5%-iles from the posterior.
- This may not always be the best interval to use.



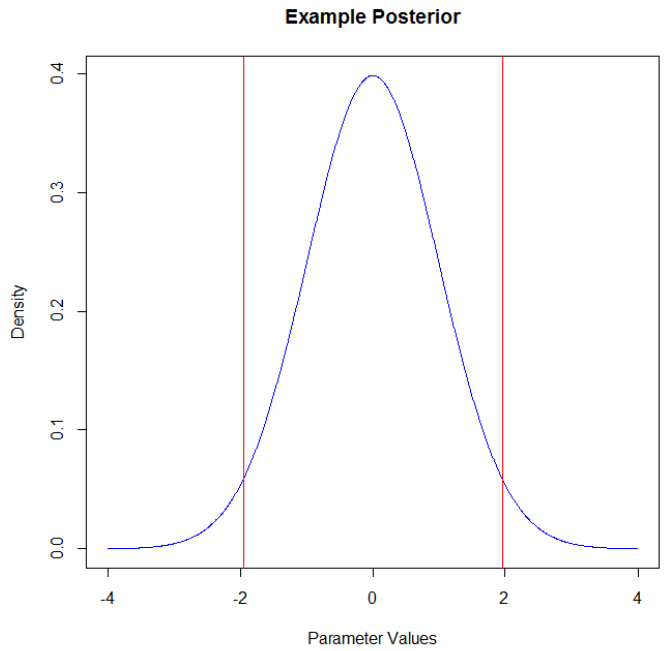


95% Highest Density Interval (HDI)

- Instead, consider a 95% Highest Density Interval (HDI).
- Identifies 95% of the area with the highest posterior probability.
- We can use the HDI to form conclusions about hypotheses.

95% Highest Density Interval (HDI)

- $H_0 : \pi_1 - \pi_2 = 0$ vs. $H_1 : \pi_1 - \pi_2 \neq 0$
- Compute a 95% HDI .
- This gives you the plausible values of the parameter.
- Can decide to Reject or Fail to reject given the interval.



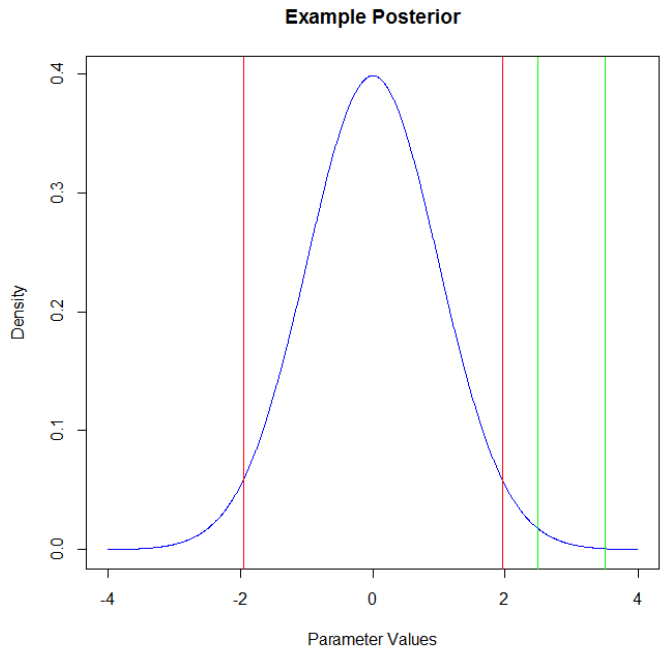
Region of Practical Equivalence (ROPE)

What if we want to obtain $P(H_0|data)$?

- Identify a Region of Practical Equivalence (ROPE).
- This is a region of values that are for practical purposes, the same
- Identify where the ROPE lies relative to the 95% HDI.
- Conclusions about the Null and Alternative depending on the overlap.
- Can be used to compute a Bayesian p-value about the test.
  - Bayesian p-value = area of HDI in ROPE / area of the HDI

Region of Practical Equivalence (ROPE)

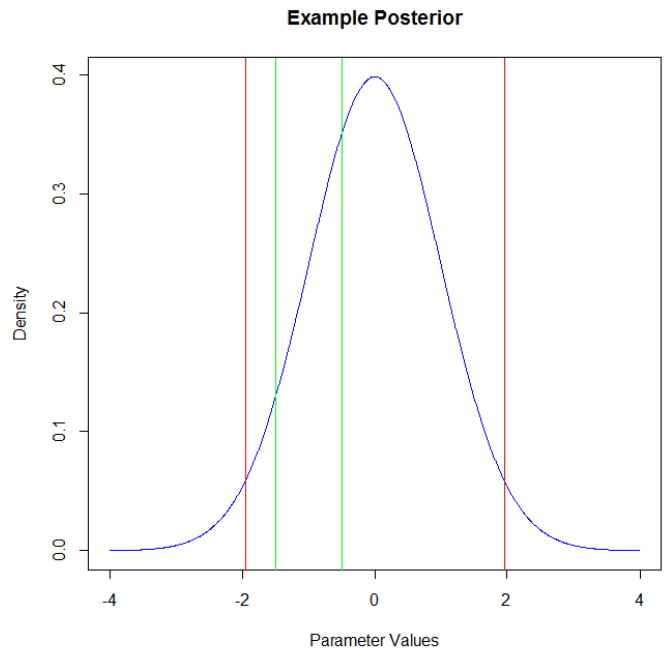
- Rejection-A parameter value is NOT credible, or rejected, if its entire ROPE lies outside of the 95% HDI.



Note: The Bayesian p-value is zero in this case.

Region of Practical Equivalence (ROPE)

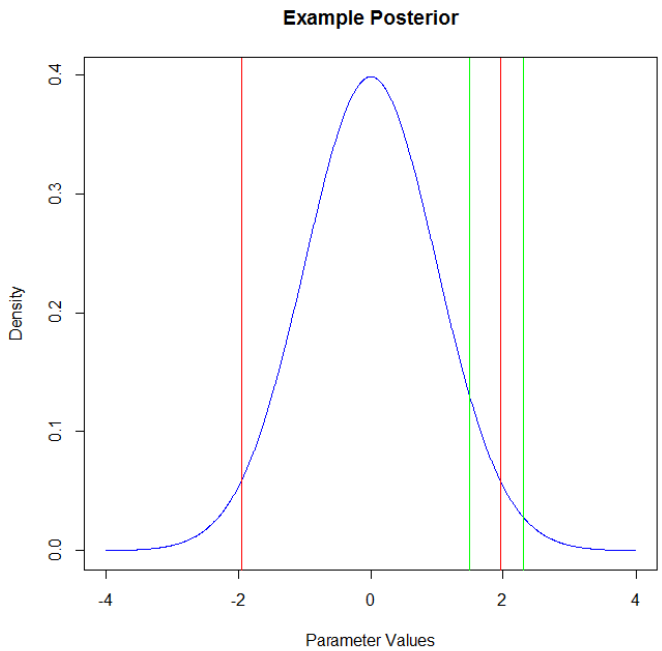
- Acceptance-A parameter value is declared to be accepted if its entire ROPE is completely contained in the 95% HDI.



Note: The Bayesian p-value is one in this case.

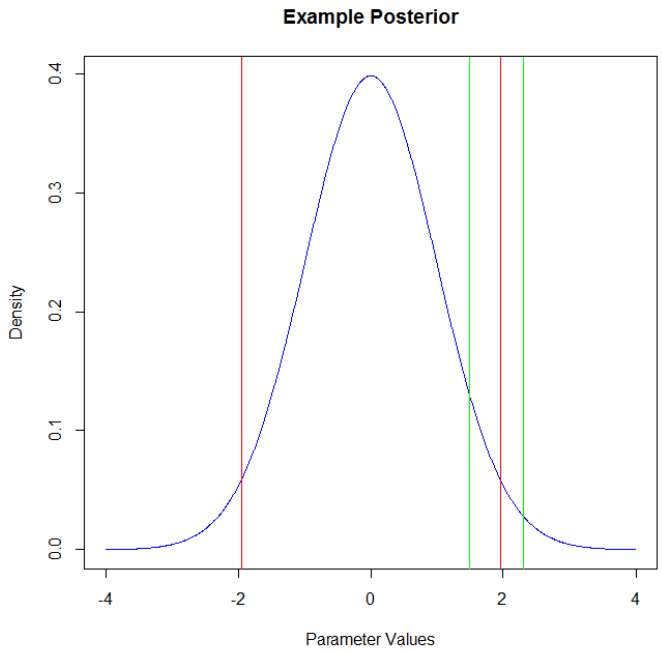
Region of Practical Equivalence (ROPE)

What about a ROPE that is partially contained within the 95% HDI?



Region of Practical Equivalence (ROPE)

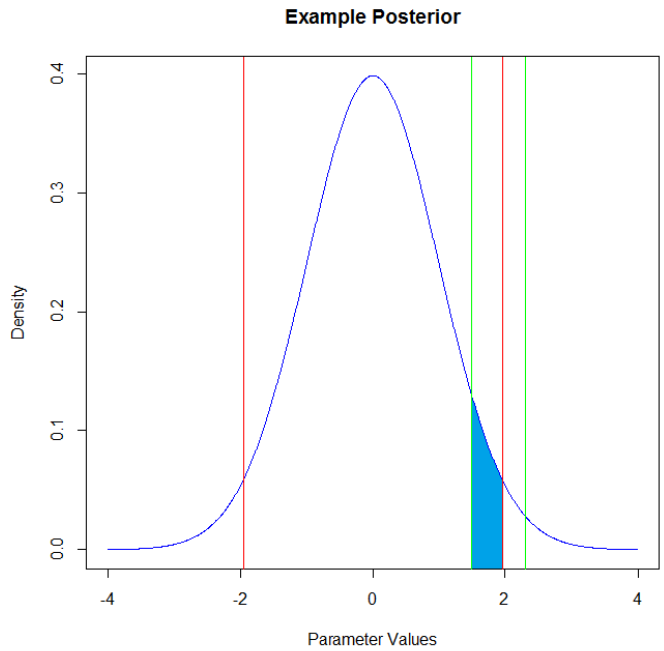
What about a ROPE that is partially contained within the 95% HDI?



- Compute the 95% HDI.

Region of Practical Equivalence (ROPE)

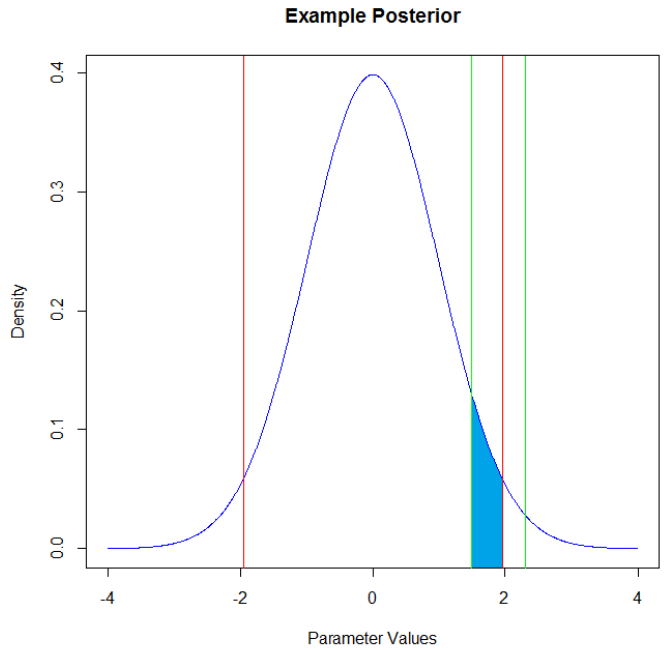
What about a ROPE that is partially contained within the 95% HDI?



- Compute the 95% HDI.
- Compute the area of the overlap of the ROPE with HDI.

Region of Practical Equivalence (ROPE)

What about a ROPE that is partially contained within the 95% HDI?



- Compute the 95% HDI.
- Compute the area of the overlap of the ROPE with HDI.
- This ratio is the probability the parameter is practically equivalent to the Null.

Region of Practical Equivalence (ROPE)

For a two-tailed test, the Bayesian p-value is VERY dependent on the chosen ROPE.

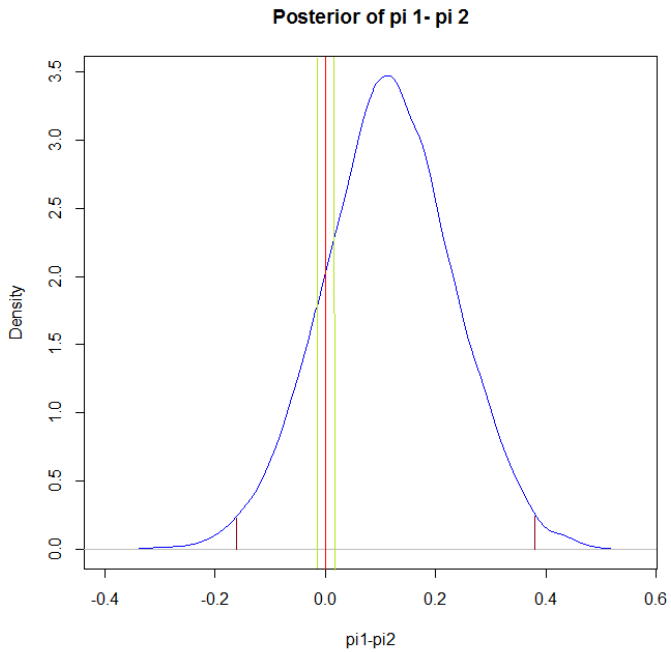
- wide ROPE implies lower prob of falsely rejecting the Null.
- wide ROPE implies lower prob of rejecting a false Null.
- ROPE can be thought of as Equivalence Limits from Equivanlence Testing.
- Reporting the 95% HDI conveys more information than "Accept/Reject".
- Reporting the 95% HDI allows reader to use their own ROPE for comparison.

Example

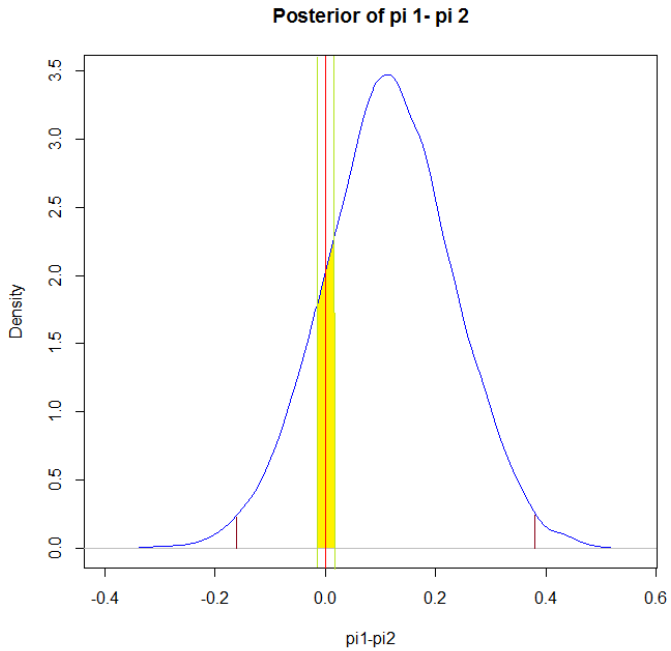
Example

- 48 subjects with cancer are observed at 2 locations (24 from each location).
- Location 1 has 7 females and location 2 has 4 females.
- Does the proportion of females differ between the two locations?
- $H_0 : \pi_1 - \pi_2 = 0$ vs. $H_1 : \pi_1 - \pi_2 \neq 0$.
- In consulation with experts, a ROPE of -0.02 to 0.02 is obtained.

Example


Posterior of pi 1- pi 2
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-0.4
-0.2
0.0
0.2
0.4
0.6
pi1-pi2

Example



Because the ROPE is completely contained within the 95% HDI we fail to reject H_0 and two-tailed Bayesian p-value of 1.0000 is obtained.