

Bayesian Linear Regression

Michael Anderson, PhD

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

Sept 16, 2016

Outline

- 1 Null Hypothesis Significance Testing
 - One-Tailed Hypothesis Test
 - Two-Tailed Hypothesis Test
- 2 One vs. Two Binomial Proportions
- 3 Two Binomial Proportions
- 4 Example
- 5 Comparing Two Means

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.

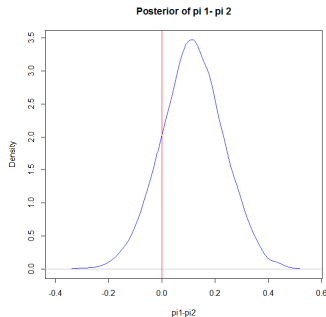
One-Tailed Hypothesis Test

- $H_0 : \pi_1 - \pi_2 \leq 0$ 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 refueling attempts are observed for 2 pilots (24 for each pilot).
- Pilot 1 has 7 disconnects and pilot 2 has 4 disconnects.
- Does the proportion of disconnects differ between the two pilots?

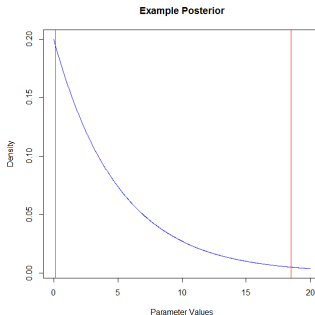
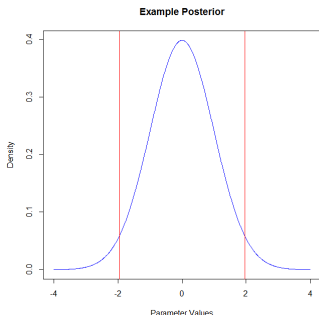
One-Tailed Hypothesis Test Example



- $P(\pi_1 - \pi_2 \leq 0 | \text{data}) = 0.1622$
- 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.

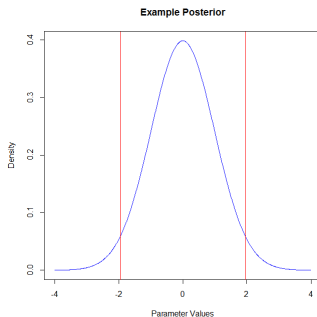


95% Highest Density Interval (HDI)

- 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.



One vs. Two Binomial Proportions

- Previously we have looked at estimating a single proportion.
- In many cases, however, we will be interested in comparing proportions from two or more groups.
- Estimation of the proportions will still be important but...
- Inference will typically focus on differences or ratios of these proportions.

Two Binomial Proportions

- Let θ_1 and θ_2 represent the proportions of two populations.
- We must first specify our prior beliefs about these proportions $p(\theta_1, \theta_2)$.

-

$$\begin{aligned}
 p(\theta_1, \theta_2 | x_1, \dots, x_n) &= p(x_1, \dots, x_n | \theta_1, \theta_2) p(\theta_1, \theta_2) / p(x_1, \dots, x_n) \\
 &= p(x_1, \dots, x_n | \theta_1, \theta_2) p(\theta_1, \theta_2) / \int \int p(x_1, \dots, x_n | \theta_1, \theta_2) p(\theta_1, \theta_2) d\theta_1 d\theta_2
 \end{aligned}$$

- If θ_1 and θ_2 are independent then $p(\theta_1, \theta_2) = p(\theta_1)p(\theta_2)$.
- So we can specify priors for θ_1 and θ_2 separately.
- Because the observations are independent, the likelihood for both groups can be specified separately.

Example

In a clinical trial 100 subjects were randomized to drug A (standard+placebo) and 100 to drug B (standard+treatment). For drug A, 60 subjects experienced an effect of interest whereas 75 subjects experienced this effect for drug B. How does the standard+treatment compare to the standard+placebo in regards to the effect of interest? Using non informative priors obtain 95% BCI for the ratio of these proportions.

- Obtain a Bayesian p-value for the test: $H_0 : \pi_B/\pi_A \leq 1$ vs. $H_1 : \pi_B/\pi_A > 1$.
- See files “drugtrial.bugs.txt” and “drug trial.R”

Comparing Two Means

Recall that the estimates for continuous data usually include a mean as well as a variance.

To compare two means from two populations, we will need to obtain

- likelihoods (two of them) for the two populations.
- priors (four of them) for the two means AND the two variances.

These additional items are easily incorporated in JAGS as illustrated in the following example.

Comparing Two Means

Suppose a physician prescribes an exercise regimen in addition to the blood pressure medication and records the systolic blood pressure (SBP) for 14 subjects after 3 months. Let's compare the mean SBP for medication only (TMT A) to the mean SBP for exercise + medication (TMT B).

```
list(N1=19,N2=14,a.tmt=c(121,94,119,122,142,168,116,172,155,
                        107,180,119,157,101,145,148,120,147,125),
     b.tmt=c(126,125,130,130,122,118,118,
             111,123,126,127,111,112,121))
```

As before, we could reasonably assume a Normal likelihood for SBP measures for both treatments.

Here we must estimate the means and variances for both groups separately. Thus we put diffuse priors on the means of both groups and diffuse priors on the precisions of both groups.

Specification of the Likelihood and Prior For Two Means

Putting this altogether we have:

- $p(a_1, \dots, a_{19} | \mu_1, \sigma_1^2) \sim \text{dnorm}(\mu_1, \tau_1)$
- $p(b_1, \dots, b_{14} | \mu_2, \sigma_2^2) \sim \text{dnorm}(\mu_2, \tau_2)$
- $p(\mu_1) \sim \text{dnorm}(0, 0.0001)$.
- $p(\mu_2) \sim \text{dnorm}(0, 0.0001)$.
- $p(\tau_1) \sim \text{dgamma}(.5, .01)$ and $\sigma_1^2 = 1/\tau_1$
- $p(\tau_2) \sim \text{dgamma}(.5, .01)$ and $\sigma_2^2 = 1/\tau_2$

Let's do this for the Systolic Blood Pressure Example

In Class Practice Problems

See files “sbp.2means.bugs.txt” and “SBP 2means.R”