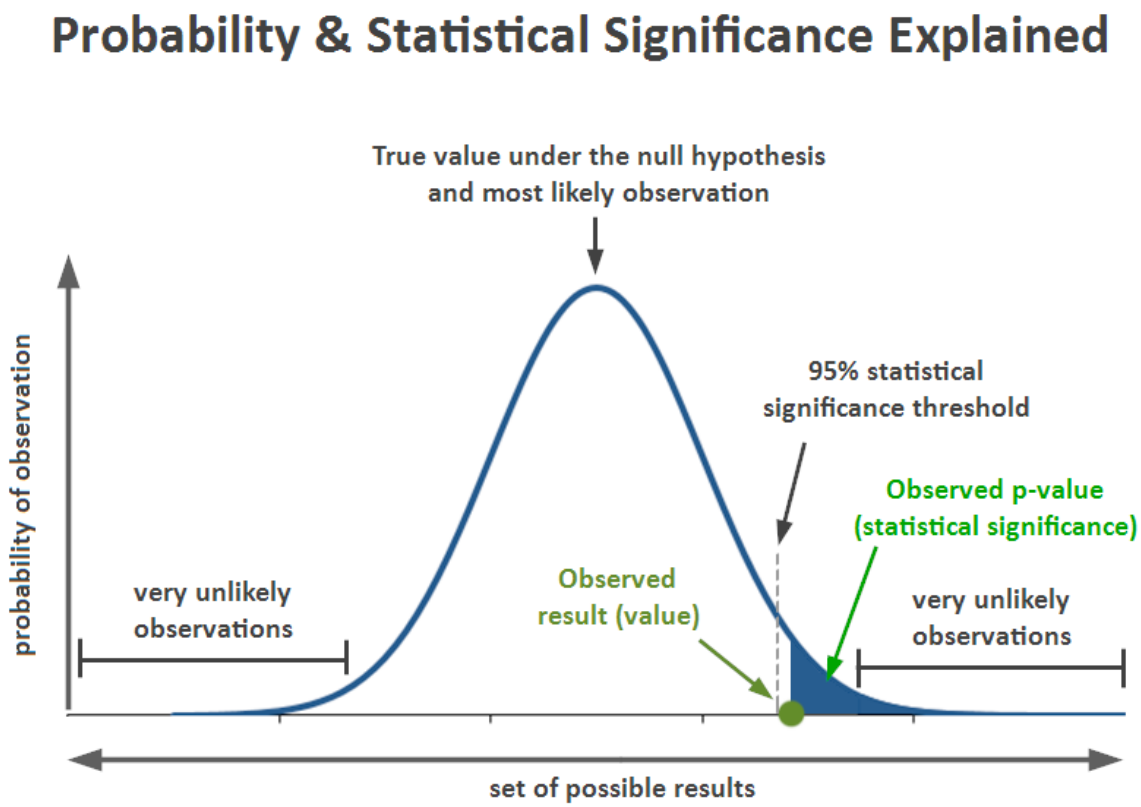
IST772 Bayesian Inference (Week 5)

Pre-class activity:

The following problem is modified from Haller and Krauss (2002): "Suppose you have a drug that you suspect may alter performance on a certain task. You compare the means of control and experimental groups (n=20 in each group). Further, suppose you use the independent means t-test and your result is (t = 2.7, df=18, p=0.0112). Consider these interpretations of your results:

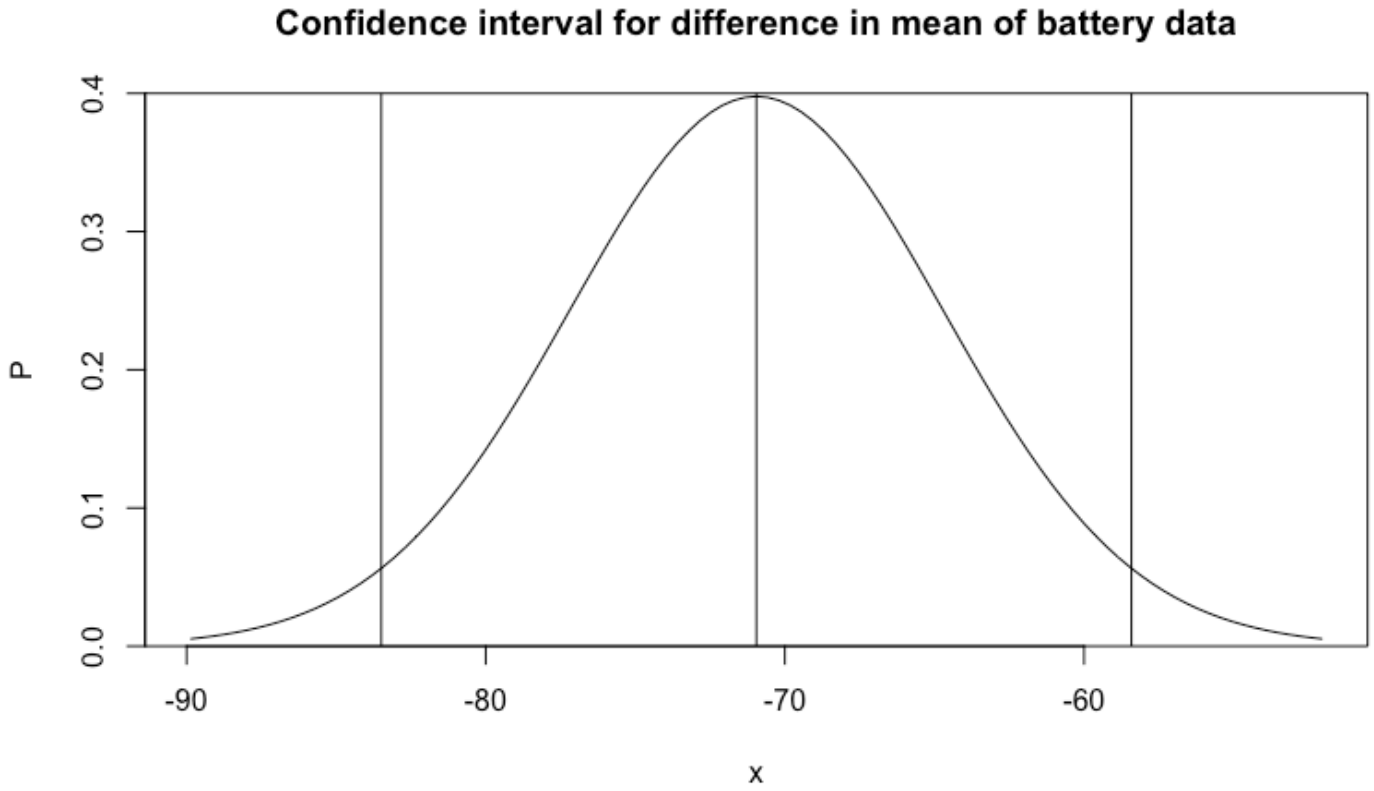
1. You have disproved the null hypothesis (that is, the hypothesis of no difference between the population means).
2. You have calculated the probability of the null hypothesis being true.
3. You have proved your experimental hypothesis (there is a difference between the population means).
4. You can deduce/calculate the probability of the experimental hypothesis being true from the provided results.
5. Assuming you decide to reject the null hypothesis, you know the probability that you are making the wrong decision.
6. You have a reliable experimental finding in the sense that if, hypothetically, you repeated the experiment many times, you would obtain a significant result on 99% of the replications.

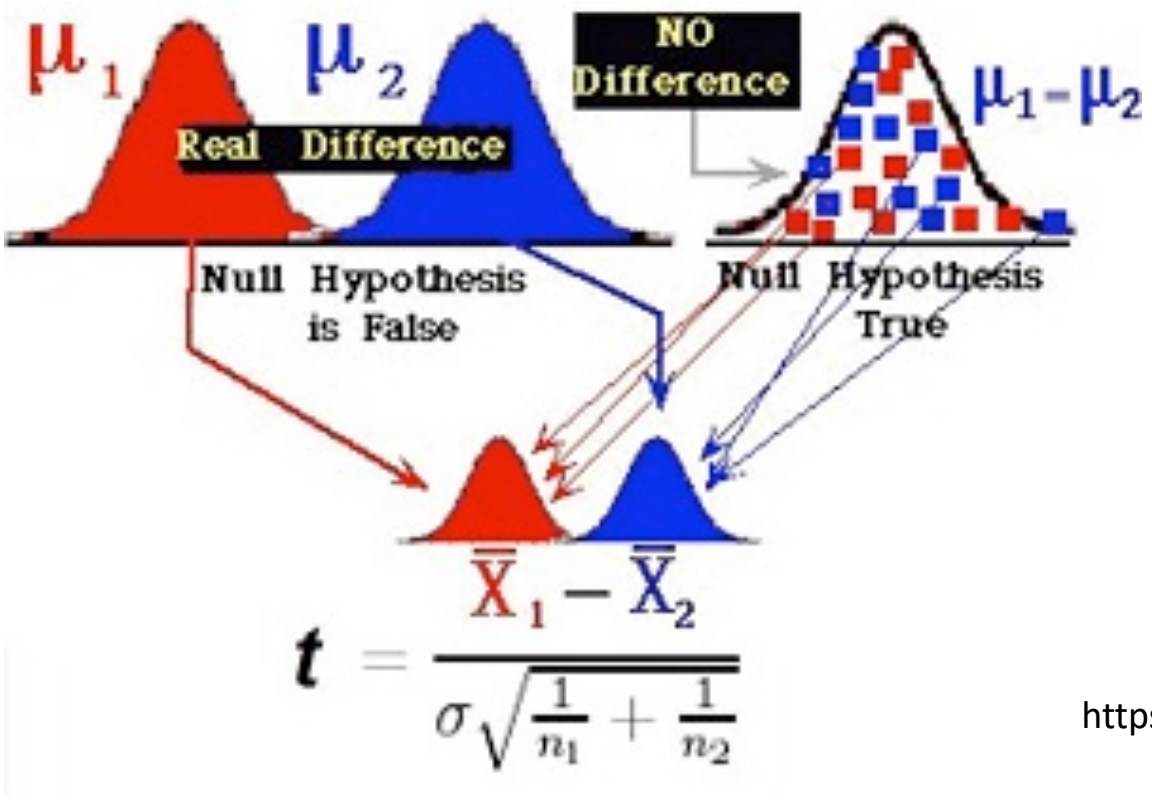
Send me a private chat message indicating which, if any, of these statements is false. It is possible that more than one statement is false.

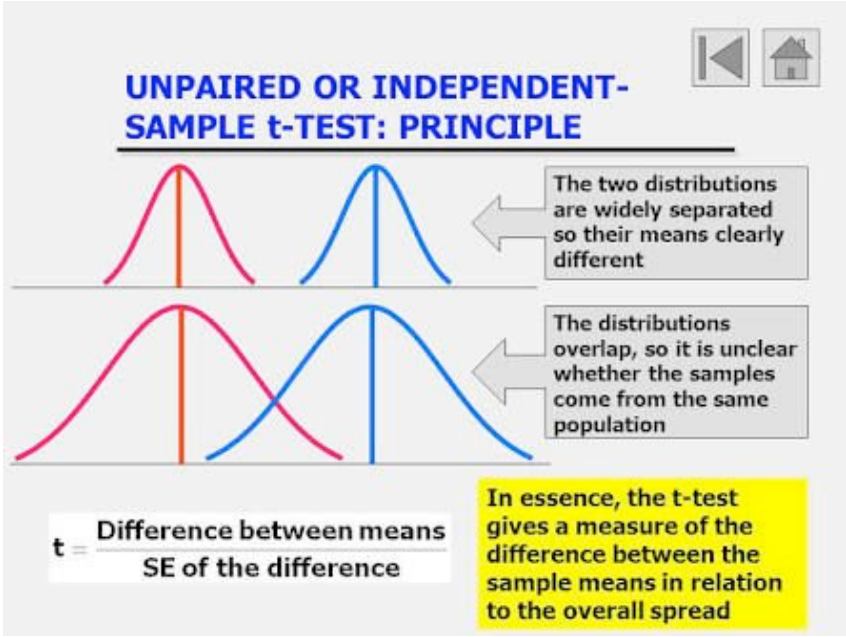


http://blog.analytics-toolkit.com/wp-content/uploads/2017/09/2017-09-11-Statistical-Significance-P-Value-1.png

Confidence interval is same logic, in reverse







https://www.pinterest.com/pin/516084438535895017/

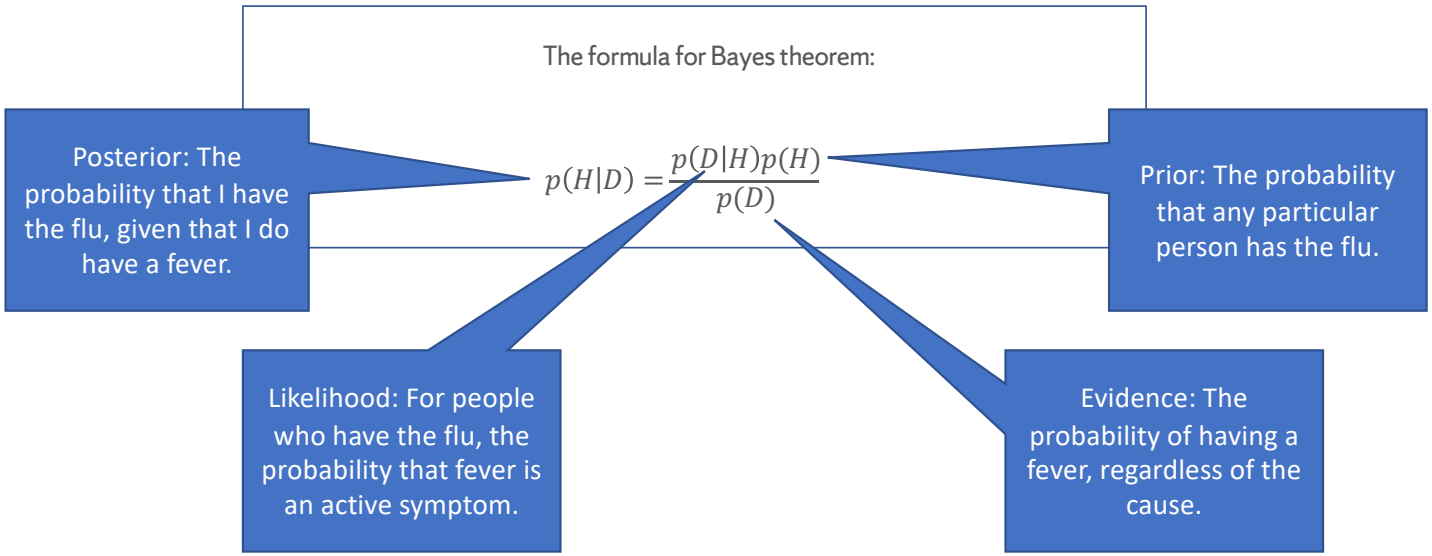
From: http://currentnursing.com/biostatistics/inferential_statistics_t-test.html

David Blackwell

(picture and dates from
https://en.wikipedia.org/wiki/David_Blackwell)

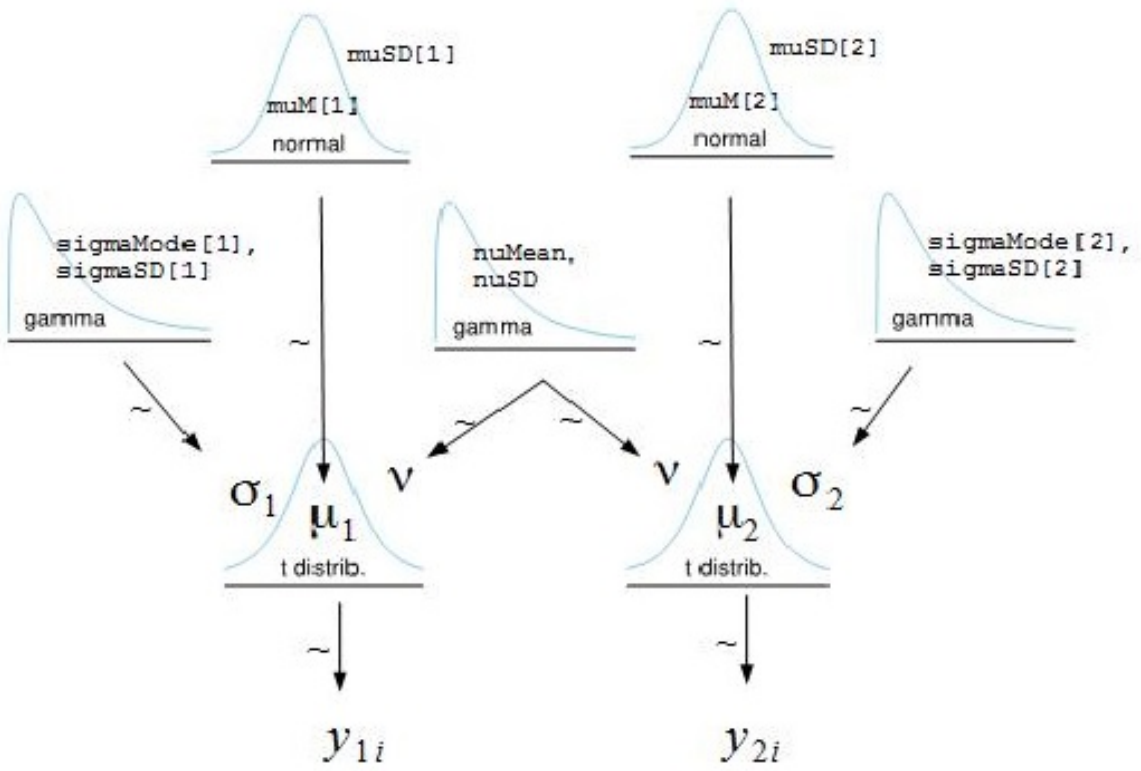
- American statistician and mathematician who made significant contributions to game theory, probability theory, information theory, and Bayesian statistics
- Seventh African American to receive a Ph.D. in Mathematics, in 1941 at age 22, from University of Illinois
- First Black tenured faculty member at UC Berkeley in 1955, after 12 years at different HBCUs
- Wrote one of the first textbooks on Bayesian statistics in 1969
- First African American inducted into the National Academy of Sciences, in 1965; also elected a member of the AAAS and many other honours

Bayes Theorem: I have a fever. . . Do I have the flu?



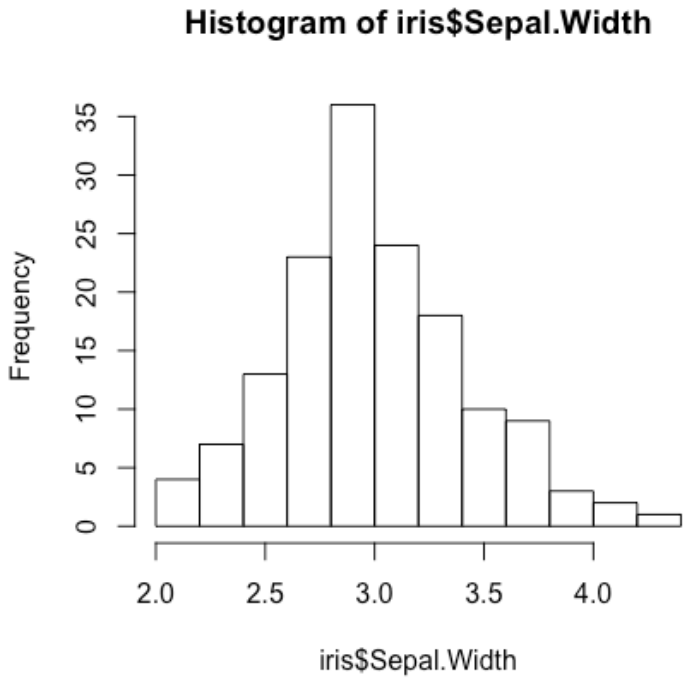
In practice, using multivariate/metric data instead of simple event counts, the denominator cannot be directly calculated, as its definition contains an integral with no closed form expression. We therefore must use an iterative computational technique to estimate p(D). This is the purpose of MCMC estimation.

BESTmcmc model



MCMC Estimation – Metropolis-Hastings Demo

- A Markov Chain is a series of linked values where the value of t+1 depends only on the value of t
- A Markov "kernel" defines the transition between value t and value t+1
- The Metropolis-Hastings algorithm is one such kernel: We can model an arbitrary distribution using a simple rule. . .



Mean: 3.057333; SD: 0.4358663

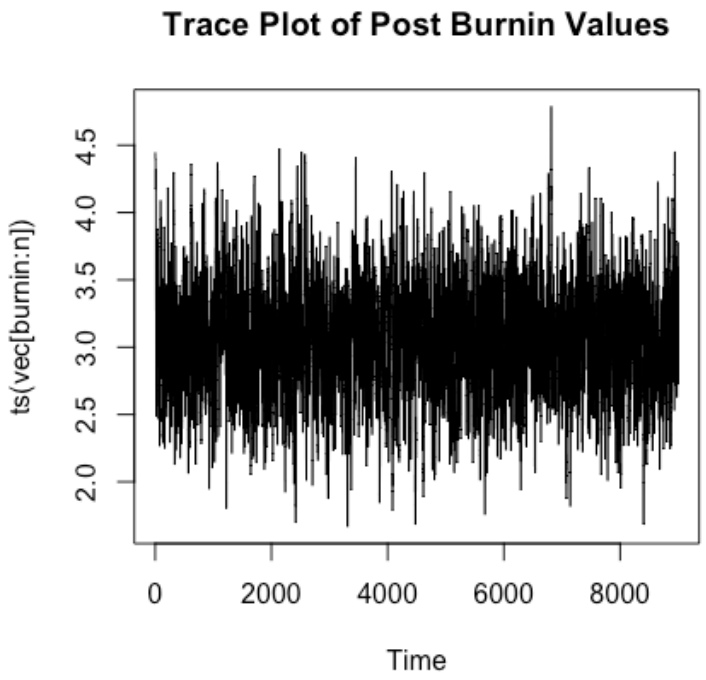
Accept or Reject a Candidate Value Based on its Likelihood in the Proposal Distribution

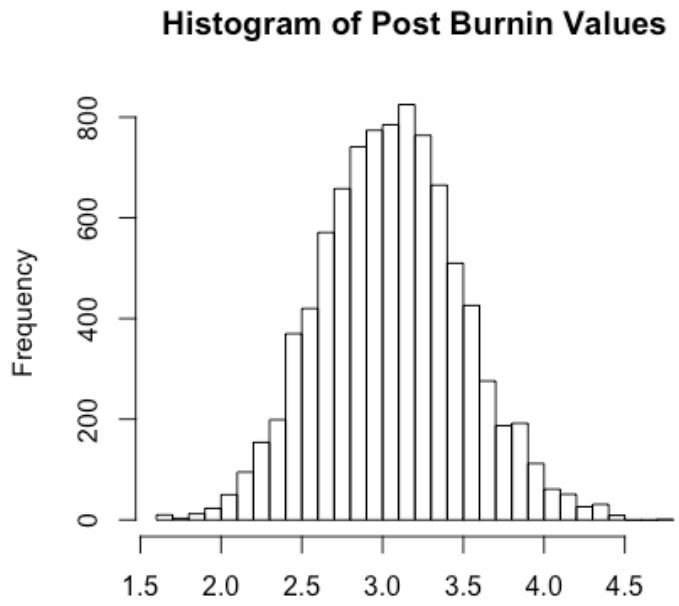
```
target <- rnorm(1, 0, sd=sigma) # Draw one value from a proposal normal distribution
can <- x + target  # Candidate value links to previous value: Makes it a Markov Chain

# Acceptance probability as the min of 1 or the ratio of density values of candidate and x.
aprob <- min(1, dnorm(can, mean=mu, sd=sigma)/dnorm(x, mean=mu, sd=sigma))

# Accept/reject the candidate value by
# comparing to a random draw from uniform(0, 1)
if (aprob >= runif(1)) { x <- can }

vec[i] <- x # Either keep the old x or use the new value depending upon the previous test
```

The relative probability in favor of the candidate instead of the previous value (x).

See: https://phylogeny.uconn.edu/mcmc-robot/

Results of 10,000 Candidates (less 1000 burnin)





median	mean	var	std.dev
3.057871002	3.057546233	0.192490420	0.438737302

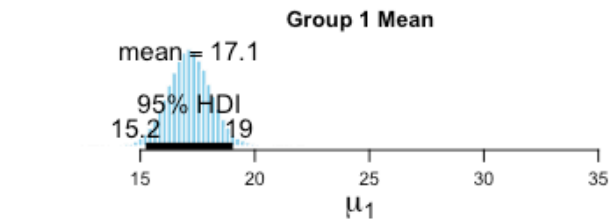
Original data: Mean: 3.057333; SD: 0.4358663

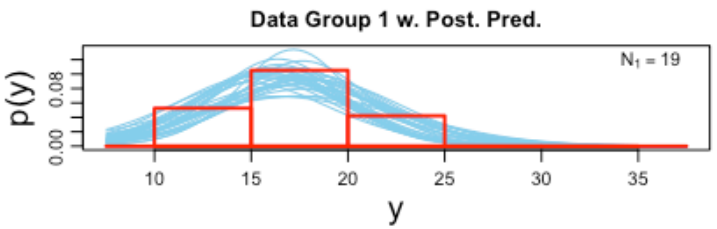
Breakout –

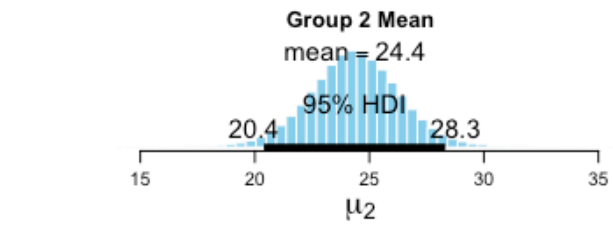
- Open notebook `1. Week5Breakout1.Rmd`
- Compare battery life for NiCad and Li-Ion groups using the BEST procedure:
  - If you have installed JAGS, the rjags package, and the BEST package, you will be able to do this in your version of R
  - If you have not installed these, you can use a website: http://www.sumsar.net/best_online/
- Produce some diagnostics and interpret your results
- Share your code on https://codeshare.io/aJDyRX

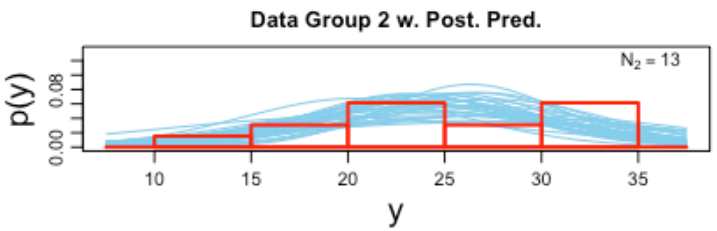
BEST: Some Explanations

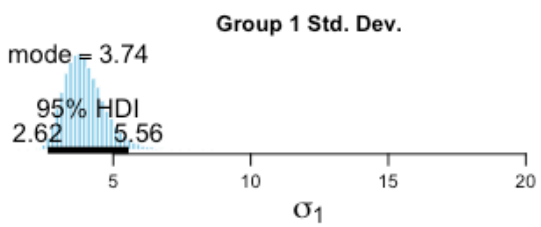
- The plotAll() command draws diagnostic plots for a variety of aspects of the Bayesian estimation
- The following slides explain some of that output
- The whole display as produced by R appears in the next slide, followed by several slides that focus on particular parameters

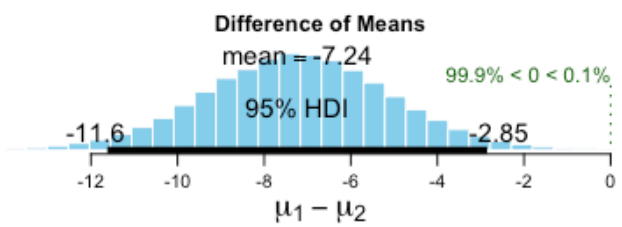

Group 1 Mean
mean = 17.1
95% HDI
15.2
19
15
20
25
30
35
μ1

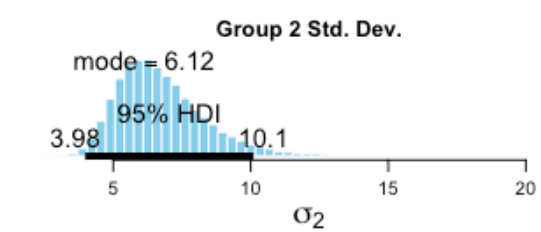

Data Group 1 w. Post. Pred.
N1 = 19
p(y)
0.00
0.08
10
15
20
25
30
35
y

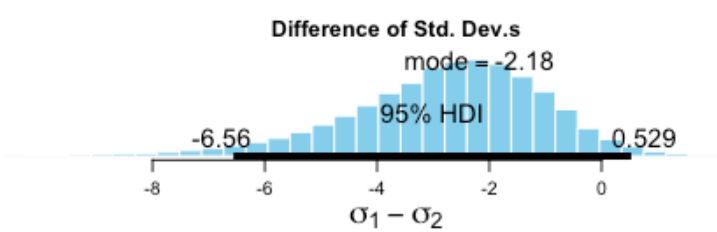

Group 2 Mean
mean = 24.4
95% HDI
20.4
28.3
15
20
25
30
35
μ2

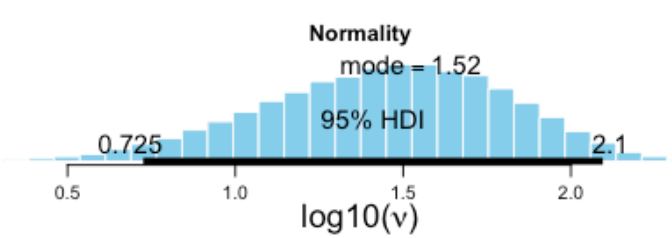

Data Group 2 w. Post. Pred.
N2 = 13
p(y)
0.00
0.08
10
15
20
25
30
35
y

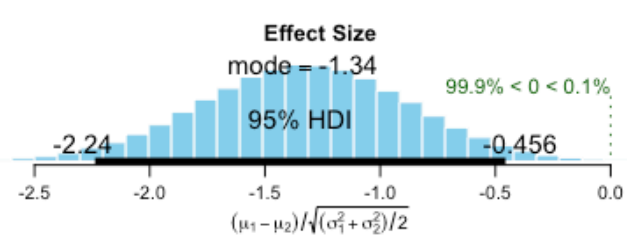

Group 1 Std. Dev.
mode = 3.74
95% HDI
2.62
5.56
5
10
15
20
σ1

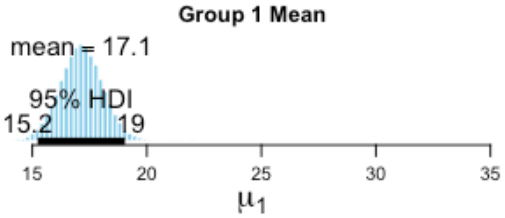

Difference of Means
mean = -7.24
99.9% < 0 < 0.1%
95% HDI
-11.6
-2.85
-12
-10
-8
-6
-4
-2
0
μ1 − μ2

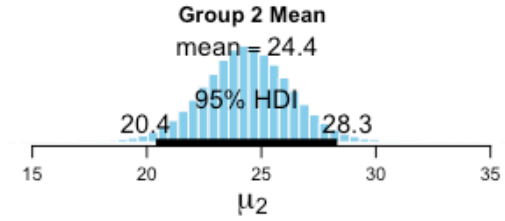

Group 2 Std. Dev.
mode = 6.12
95% HDI
3.98
10.1
5
10
15
20
σ2


Difference of Std. Dev.s
mode = -2.18
95% HDI
-6.56
0.529
-8
-6
-4
-2
0
σ1 − σ2


Normality
mode = 1.52
95% HDI
0.725
2.1
0.5
1.0
1.5
2.0
log10(ν)

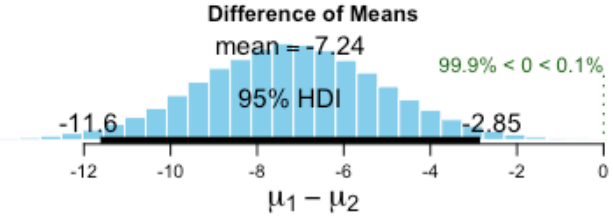

Effect Size
mode = -1.34
99.9% < 0 < 0.1%
95% HDI
-2.24
-0.456
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
(μ1 − μ2)/√((σ1² + σ2²)/2)





The histograms to the left show the means from each of the 100,000 MCMC simulations for each of the two groups. The X-axis is calibrated identically between the two histograms to simplify the visual comparison.

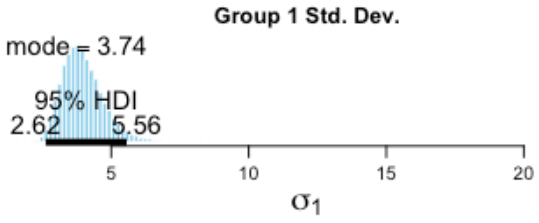
For all of these histograms, the dark line on the X-axis shows the extent of the “highest density interval.” Each HDI represents the 95% of the possibilities for the posterior distribution of the given parameter, in this case the means.

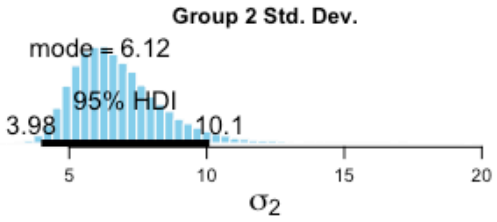


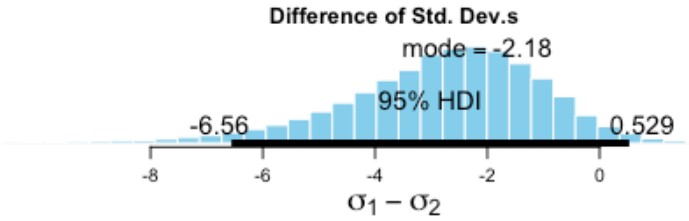
The histogram above shows the difference between the group means for each of the 100,000 MCMC simulations.

The BEST procedure also models the standard deviations for each of the 100,000 MCMC simulation runs. As a result, we can also compare SDs between the two groups. The difference in SDs is modeled in the histogram below.

Does the HDI for the difference in SDs overlap with zero? Is there a credible difference between the SDs of the two groups?

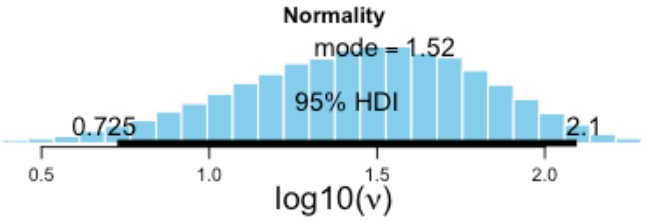






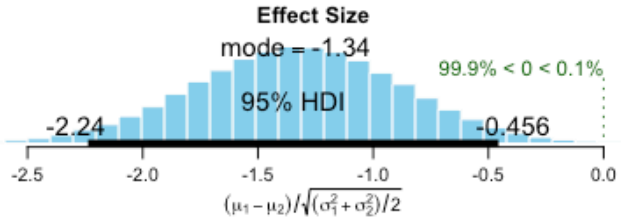
For each of the 100,000 MCMC simulation runs, the BEST procedure calculates an index of normality, notated below as the log (base 10) of "nu."

What we are looking for is values of log10(nu) that are in the region of 1.5. To the extent that the HDI of this display does NOT overlap with 1.5, there may be severe non-normality. The BEST procedure is robust against non-normality whereas the classical t-test is not.



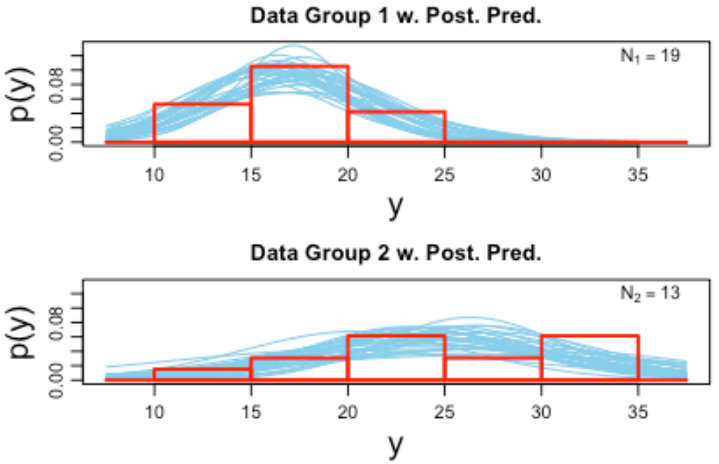
The histogram below models the "effect size" of each of the simulation runs. Effect size is calculated as Cohen's D, which is simply the difference in means divided by the pooled standard deviation (a kind of average SD across the two groups).

If the effect size HDI does not overlap with zero, this is suggestive of a credible difference between means. Remembering that Cohen's D is calibrated in SDs, anything with an absolute value > 0.8 is considered a large effect.

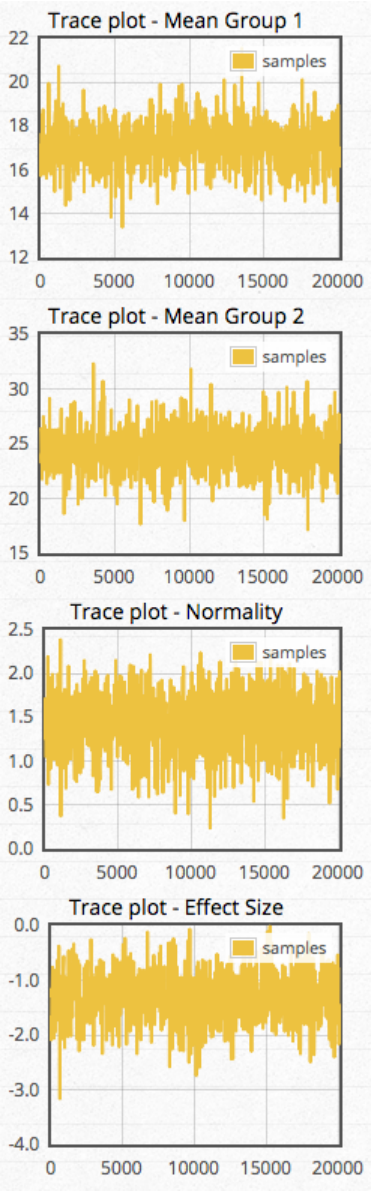


These displays provide a diagnostic showing whether it is appropriate to be using the BEST procedure. The blue curves show a random sample of plots of the t distribution, using the mean, SD, and normality coefficient generated by a particular run of the MCMC process. The red histogram is the actual sample data.

To the extent that data are highly multi-modal or weirdly distributed, the t-distribution may be a bad fit.



For example, the rightmost category (30-35) in the Data Group 2 histogram seems like a second mode for these data. With only n=13 observations, however, there is not enough data to really say for sure. In the absence of other anomalies, this is not a severe enough problem to raise a concern.



These trace plots do not appear in the plotAll() output from BESTmcmc(), but they do appear on Rasmus Baath's webpage. We will also see examples of these trace plots later in the course.

Each trace plot represents the same data that you saw in the HDI histograms on previous pages, but rather than displaying them as a histogram, they show as a time series.

RB's webpage defaults to 20,000 MCMC simulations. For each simulation run, the various parameters generated for that step in the robot's path are displayed as a new dot just to the right of the previous dot. You can watch these unfold in real time as the Javascript processes your data.

RB says that these should look like a hairy caterpillar. We want to see roughly equal variation around a reasonably stable mean (for the topmost plot about 17 mpg) throughout the whole run, with just a few small spikes here and there.

Week 5 practice exam

- Open the file week5practiceexam.docx in Blackboard and answer the questions in the document
- You can upload the exam on the LMS if you want feedback on your answers; otherwise, just post saying you do not need feedback
- Points are given for taking the exam

Homework and Practice Exam

- Make sure you are using the updated syllabus that I distributed at the beginning of the semester (on the wall and in the handouts folder).
- Feel free to submit your exam file to the LMS if you are unsure about any of your answers.
- The homework for week five is based on exercises 6 through 10 on pages 86 and 87 but with changes in the notebook.