IST772 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Wanyue Xiao
Date: September 15, 2020
Chapter Number: #5
Title of Chapter: Bayesian and Traditional Hypothesis Testing

The BEST package:
1. **The expression of 99.8% < 0 < 0.2%**
   * Shows *the proportion of mean differences in the MCMC run that were negative versus the proportion that were positive*.
   * **Interpretation**: 99.8% of the mean differences in the distribution were negative (meaning that manual transmissions were more efficient) and 0.2% were positive.
2. **95% Highest density interval (HDI):**
   * **Interpretation**: there is a 95% probability that the population mean difference between the two groups falls within this range.
   * HDI and confidence interval is not identical. The confidence interval is *calculated directly from the sample data* and conceptually *provides no direct evidence about the population mean difference*. Therefore, the **BEST HDI directly models the population parameters of interest** and shows us probability distributions for those parameters, whereas the **confidence interval uses sample data to compute one and only one example** of an upper and lower bound for the population mean.

t **distribution** is a family of symmetric distributions similar to the normal distribution *but with "thicker" tails that are a function of the greater uncertainty* involved in working with samples.

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

* To calculate t: -> equal to the difference in the sample means divided by the pooled standard error
* To calculate the df (degree of freedom): df = (n1-1)+(n2-1).

The Null Hypothesis Significance Test:
1. Begin by *asserting a null hypothesis* that there is no mean difference between the means of two groups.
2. *Choose an "alpha level" probability*, beyond which the null hypothesis will be rejected: a common alpha level is .05.
3. Collect data and conduct a statistical test *to calculate a significance value*, designated by the letter *p*.
4. If the calculated value of *p is less than the alpha level that was chosen above, then reject the null hypothesis.*
5. When the null hypothesis is rejected, this can be considered evidence in favor of some unspecified alternative hypothesis or t*he probability that any particular alternative hypothesis may be correct*.
6. If *p is greater than the alpha level that was chosen above*, then *fail to reject* the null hypothesis. Failing to reject the null hypothesis **does not mean that we accept the null hypothesis, rather that we have no good evidence either way.**
7. t.test() command

IST772 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Concerns of NHST:
- Foundational Assumption: One can use the ROPE to cope with the issue aroused by foundational assumption:
- The selection of alpha level: depends on the table of critical values.
    To interpret p-value 12.74/10000: If the null hypothesis were true, then only 14 out of 10,000 t-tests would yield a value of t larger in magnitude than 3.7671 (t value).

"Effect Size" refers to the strength or magnitude of the statistical finding.
- Cohen's *d* is conceptually simple: divide the mean difference by the pooled standard deviation of the two samples; **Interpretation**: Cohen's *d* showed a large effect size of *d* = –1.47, indicating that _manual transmissions were nearly one and a half standard deviations more fuel efficient than automatic transmissions._
- co.end.d() command in effsize package

Notation, Formula, and Notes on Bayes' Theorem & Markov-Chain Monte Carlo

$$p(H \mid D) = \frac{p(D \mid H) p(H)}{p(D)}$$

Bayes' theorem:
Here, *H* represents hypothesis and *D* represents data. To find out the result of *p(H|D)*, one shall know that the probability of this hypothesis being "true" given the statistical outcomes observed from these data. In Bayesian terms this is known as the **posterior probability**.
- **Likelihood** *p(D|H)*: the probability of observing these data when the hypothesis is "true";
- **Prior probability** *p(H)*: the baseline belief about the "truth" of the hypothesis;
- **Evidence** *p(D)*: the probability of observing these data under any and all conditions;

Markov-Chain Monte Carlo:
A Markov chain is a mathematical structure representing a matrix of state changes with probabilities. Markov chains model processes _where the possibilities for the next move only depend on the current position_.

Exercise Review

```
> t.test(PlantGrowth$weight[PlantGrowth$group=="ctrl"], PlantGrowth$weight[P
lantGrowth$group=="trt1"])

        Welch Two Sample t-test

data:  PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$weigh
t[PlantGrowth$group == "trt1"]
t = 1.1913, df = 16.524, p-value = 0.2504
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2875162  1.0295162
sample estimates:
mean of x mean of y
   5.032     4.661
```

6.

IST772 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

The value of t value, df, and p-value is 1.1913, 16.524, and 0.2504 respectively. Given that the p-value is 0.2504, which is far higher than 0.05, we fail to reject the hypothesis. The upper boundary is -0.2875 whereas the below boundary is 1.0295.

7.
Code:
plantBest <- BESTmcmc(PlantGrowth$weight[PlantGrowth$group=="ctrl"],
PlantGrowth$weight[PlantGrowth$group=="trt1"])
plot(plantBest)
Graph:

Difference of Means

mean = 0.389

14.2% < 0 < 85.8%

95% HDI

-0.368   1.13

$\mu_1 - \mu_2$
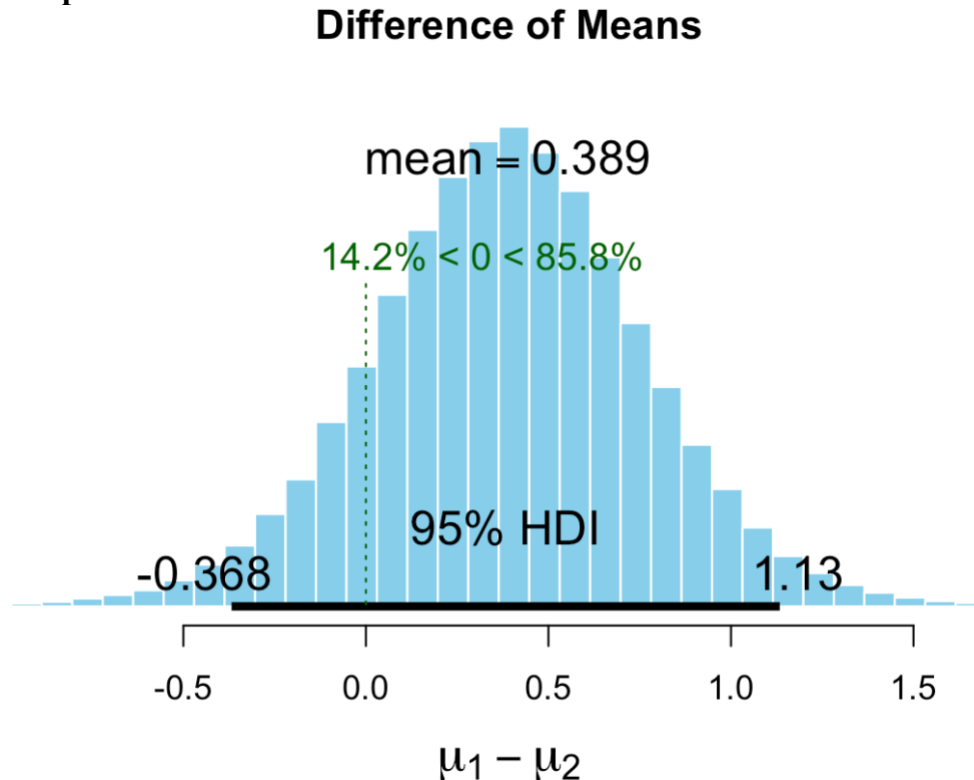
The HDI boundary values are -0.368 and 1.13. HDI is presented in the context of a probability distribution of estimated population mean differences. The interpretation is that 95% of the likely values of the population mean difference lie in the bell-shaped area between —0.368 and 1.13.

8.
A *t*-test tells you whether the difference between two sample means is "statistically significant" - not whether the two means are statistically different.
The interpretation of results of the null hypothesis: the result of null hypothesis can be used to reject or fail to reject the null hypothesis. A *t*-score with a *p*-value larger than 0.05 just states that the difference found is not "statistically significant". It does not say that there is no difference that can be relevant in statistics.
In this case, 2504 out of 10000 t-tests would yield a value of t smaller in magnitude than 1.1913, which is higher than the desired value 500 out of 10000. Therefore, we fail to reject the null hypothesis.

IST772 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

The confidence interval is a range of values that's likely to include a population value with a certain degree of confidence. In this case, there are 95 of those replication samples, proportionally, that the population difference in weight between group ctrl and group trt1 could be either a positive number or a negative number somewhere in the region of 0.371 plus or minus about 0.6585. Here the 0.371 is the central of the region which is also the best point estimate.
The HDI: Compared with confidence interval, HDI focus on difference of population mean instead of difference of sample mean. The interpretation is mentioned above in Question 7.

9.

```
> t.test(PlantGrowth$weight[PlantGrowth$group=="ctrl"], PlantGrowth$weigh
t[PlantGrowth$group=="trt2"])

        Welch Two Sample t-test

data:  PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$we
ight[PlantGrowth$group == "trt2"]
t = -2.134, df = 16.786, p-value = 0.0479
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.98287213 -0.00512787
sample estimates:
mean of x mean of y
    5.032     5.526
```

The interpretation of results of the null hypothesis: In this case, 475 out of 10000 t-tests would yield a value of t larger in magnitude than 1.1913, which is lower than the desired value 500 out of 10000. Therefore, we could reject the null hypothesis.
The confidence interval: In this case, there are 95 of those replication samples that the population difference in weight between group ctrl and group trt2 could be a negative number somewhere in the region of -0.4889 plus or minus about 0.494. Here the 0.4889 is the central of the region which is also the best point estimate.

Difference of Means

mean = -0.487

95.9% < 0 < 4.1%

95% HDI

-1.04   0.0813

$\mu_1 - \mu_2$
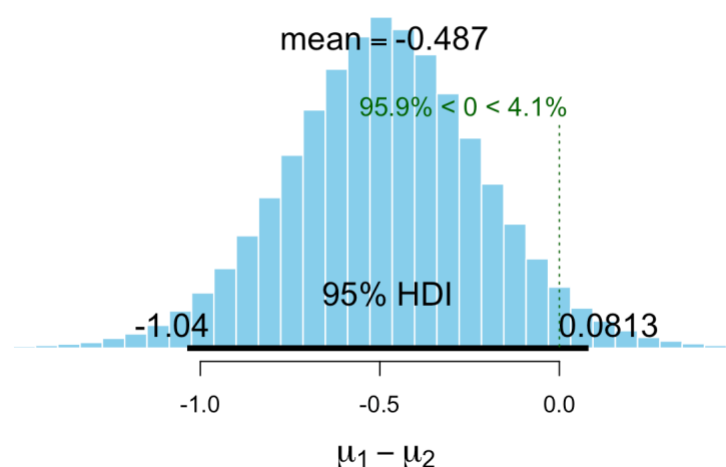
The HDI: The interpretation is that 95% of the likely values of the population mean difference lie in the bell-shaped area between –-1.04 and 0.0813.

10.

IST772 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

```
> t.test(rnorm(100000,mean=17.1,sd=3.8),rnorm(100000,mean=17.2,sd=3.8))

        Welch Two Sample t-test

data:  rnorm(1e+05, mean = 17.1, sd = 3.8) and rnorm(1e+05, mean = 17.2,
 sd = 3.8)
t = -6.6167, df = 199991, p-value = 3.681e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.14609766 -0.07932442
sample estimates:
mean of x mean of y
 17.09282  17.20553
```

The p-value is 3.681e-11, which is much smaller than the desired alpha level 0.01, let alone the commonly used on (which is 0.05). Therefore, one should reject the null hypothesis. However, due to the effect size, **with a large enough sample size any experiment can eventually reject the null hypothesis** and detect trivially small differences that turn out to be statistically significant. The large sample will reduce the standard error to close to zero. This in turn will artificially boost the t stat and commensurately lower the p value to close to 0%.

R Code Fragment and Explanation
BESTmcmc(y1, y2 = NULL, priors = NULL, …)
This function is the core of the BEST package. It calls JAGS and passes a description of the model, priors, and data, then retrieves and returns the MCMC samples for the parameters.

      y1   a numeric vector of data values.

      y2   a vector of values for a second group, or NULL if there is only
                  one group of observations.

      priors   an optional list of values controlling the priors.

Question for Class
1. If we fail to reject the null hypothesis, is that mean we only have one single answer, which is there is no difference between two groups? If not, what else answer we could have?
2. If we could reject the null hypothesis, is that indicate that there is a difference between those two groups?