

Bayes for Undergrads

Phil Ender

UCLA Statistical Consulting Group (Ret)

Stata Conference Columbus - July 19, 2018

The UCLA Department of Statistics teaches Stat 10: Introduction to Statistical Reasoning for undergraduates.

It is service course for a number social science and biological science departments. The course is ten weeks long and covers topics from simple probability up to simple linear regression including the two-group Student's t-test.

How much do students retain after 10 weeks of Intro to Statistical Reasoning?

Sadly, not much.

They remember the mean and something about the normal distribution. And, they almost all remember the two-group t-test.

There's something almost magical about the attraction of the t-test to students.

What do students remember about the t-test?

$$\frac{\bar{X}_1 - \bar{X}_2}{\textit{something}} \quad (1)$$

The **something** part is a bit unclear in their minds.

t-Test Example

Tradition Null Hypothesis Significance Testing

```
. use hsbdemo, clear
```

```
. ttest write, by(female)
```

Two-sample t test with equal variances

```
-----+-----  
Group | Obs   Mean   StdErr   StdDev   [95% Conf. Int.]  
-----+-----  
male  |  91  50.1209  1.08027  10.3052  47.97473  52.26703  
femal|109  54.9908  .779069  8.13372  53.44658  56.53507  
-----+-----  
combin|200   52.775  .670237  9.47859  51.45332  54.09668  
-----+-----  
diff  |      -4.86995  1.30419                -7.44184 -2.298059  
-----+-----
```

t-Test Example Example Continued

```
diff = mean(male) - mean(female)      t = -3.7341
Ho: diff = 0                          degrees of freedom = 198

Ha: diff < 0      Ha: diff != 0      Ha: diff > 0
Pr(T < t) = 0.0001 Pr(|T| > |t|) = 0.0002 Pr(T > t) = 0.9999
```

t-Test Example – Effect Size

```
. esize twosample write, by(female)
```

Effect size based on mean comparison

Obs per group:

male = 91

female = 109

Effect Size	Estimate	[95% Conf. Interval]	
Cohen's d	-.5302296	-.8127436	-.2464207
Hedges's g	-.5282182	-.8096604	-.2454859

Teach students the principles and practice of the Markov chain Monte Carlo Bayesian analysis using something that the students can relate to. Namely, the t-test.

Unfortunately, there is no Bayes prefix for the t-test command.

Instead, we will use the **bayesmh** command to create something the students can relate to.

Use **bayesmh** to generate posterior distributions of the means and variances for each of the two groups.

And, from the posterior distributions of the means we can construct an analysis that is equivalent to the two-group t-test.

The following relationship sets the stage for the several parts of the **bayesmh** command.

$$Posterior \propto Likelihood \times Prior \quad (2)$$

In this presentation the t-distribution will be used in the likelihood model of **bayesmh** to describe the data.

I want to emphasize the point that the t-distribution is not being used as a probability distribution for hypothesis testing. It is only being used to describe the distribution of the data.

We don't need no stinkin' assumptions!

This may not be completely true. However, we don't need assumptions about normality and homogeneity of variance that are required when using the t probability distribution to test hypotheses.

Remember we are using the t -distribution likelihood as a description of our data not as a probability distribution used for statistical hypothesis testing.

Using Bayes prefix would easier than **bayesmh**

```
. bayes, hpd: regress write i.female
```

Yes, this is straight forward but it does not correspond to the students' mental image of the t-test with the differences between two means.

Using **bayesmh** we can construct an analysis that parallels their mental framework.

The Bayesmh Comand

```
. fvset base none female

. bayesmh write i.female, noconstant           ///
    likelihood(t({var:i.female, nocons}), 7))  ///
prior({write:}, normal(0, 10000))             ///
prior({var:}, igamma(.01, .01))              ///
init({var:} 1) block({var:})                  ///
burnin(5000) mcmcsize(50000)                  ///
hpd rseed(47)
```

There is a lot of stuff here, so let's deconstruct this command in chunks.

Bayesmh Deconstruction - The Model

```
. fvset base none female  
. bayesmh write i.female, noconstant
```

To get separate estimates for both males and females we need to set the base level for female to none along with using no constant for the model.

```
likelihood(t(({var:i.female, nocons}), 7))
```

The syntax for the t likelihood is $t(\textit{sigma}2, \textit{df})$.
Again make use of the nocons option to get separate variances for each group.

Use a smallish degrees of freedom for fatter tails than the normal distribution. This could help with outliers.


```
prior({write:}, normal(0, 10000)) ///  
prior({var:}, igamma(.01, .01)) ///
```

Somewhat noninformative priors for means and variances.

We could have used t-distribution prior for the means. Andrew Gelman might consider that to be a weakly informative prior.

Bayesmh Deconstruction - Options

```
init({var:} 1) block({var:})          ///  
burnin(5000) mcmcsize(50000)        ///  
hpd rseed(47)
```

`init({var:} 1)` - Better starting value for variance than the default `init` of zero.

`block({var:})` - Helps with mixing and improves the efficiency of the Metropolis–Hastings algorithm.

`mcmcsize(50000)` - Some researchers recommend 100,000 mcmc reps. Increasing the `mcmcsize` would help in reducing the MCSE.

`hpd` - Highest posterior density credible intervals alternative to equal-tailed credible intervals.

Model summary

Likelihood:

```
write ~ t(xb_write,{var:i.female,nocons},7)
```

Priors:

```
{write:i.female} ~ normal(0,10000)
```

```
{var:i.female} ~ igamma(.01,.01)
```

(1) Parameters are elements of the linear form `xb_write`.

Bayesmh Output – Header

Bayesian t regression

Random-walk Metropolis-

Hastings sampling

MCMC iterations = 55,000

Burn-in = 5,000

MCMC sample size = 50,000

Number of obs = 200

Acceptance rate = .244

Efficiency: min = .09757

avg = .1071

Log marginal likelihood = -750.11755

max = .1155

Bayesmh Output – Estimates Table

	Mean	StdDev	MCSE	HPD [95% Cred. Interval]	
write					
male	50.34901	1.170282	.016223	48.16482	52.73893
female	55.55363	.8070589	.010622	53.92307	57.07884
var					
male	96.41478	16.442	.235399	66.63293	129.1073
female	55.14833	8.864754	.118853	38.65227	72.65642

Note: Output edited to fit space.

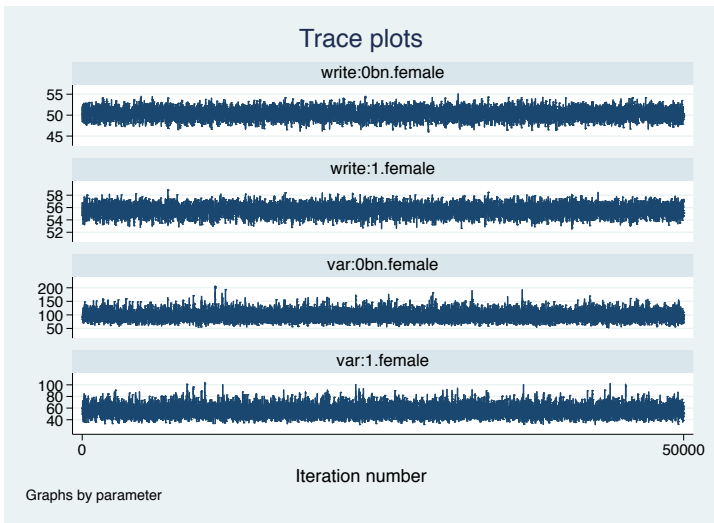
Let's Inspect the Posetrior Distribution

_index	eq1_p1	eq1_p2	eq2_p1	eq2_p2	_freq
1	52.1539	55.3361	92.666557	59.85294	1
2	51.269785	54.716995	92.666557	59.85294	2
4	50.002058	55.864413	92.666557	59.85294	2
6	48.446471	56.748254	92.666557	59.85294	3
9	49.404953	56.641649	92.666557	59.85294	1
...					
49987	50.353773	55.533455	86.964364	45.956056	2
49989	49.253494	55.048986	99.922864	50.792015	1
49990	49.825816	55.10641	99.922864	50.792015	6
49996	49.825816	55.10641	70.6489	63.027343	3
49999	49.825816	55.10641	92.526761	60.513473	2

Because of duplicate rows there are 21,414 observations in the dataset.

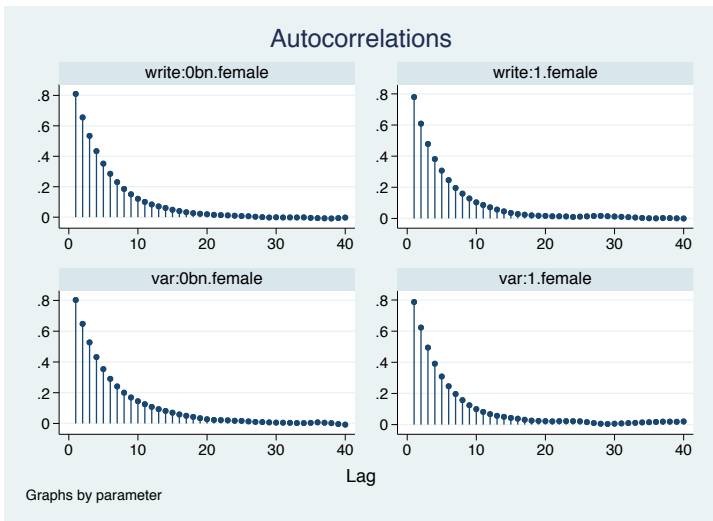
Bayesgraph Trace

```
. bayesgraph trace _all, byparm
```



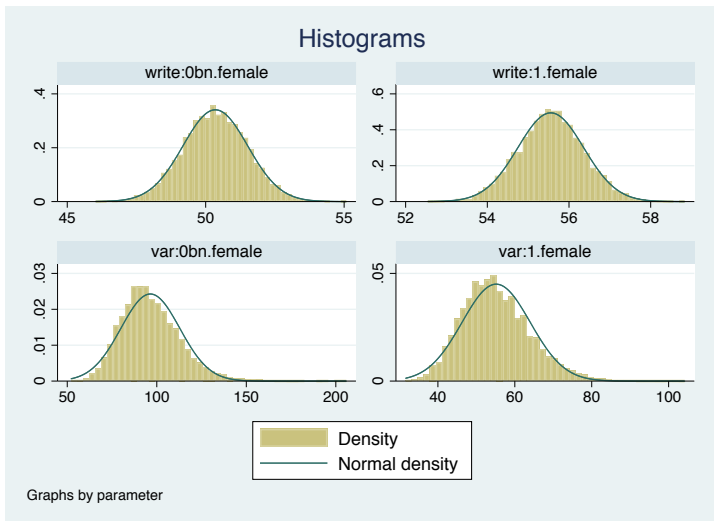
Bayesgraph Autocorrelation

```
. bayesgraph ac _all, byparm
```



Bayesgraph Histogram

```
. bayesgraph histogram _all, normal byparm
```



Bayesstats Summary – Mean Difference

```
. bayesstats summary (mea_dif:{write:1.female}- \\
>                               {write:0bn.female}), hpd
```

Posterior summary statistics MCMC sample size = 50,000

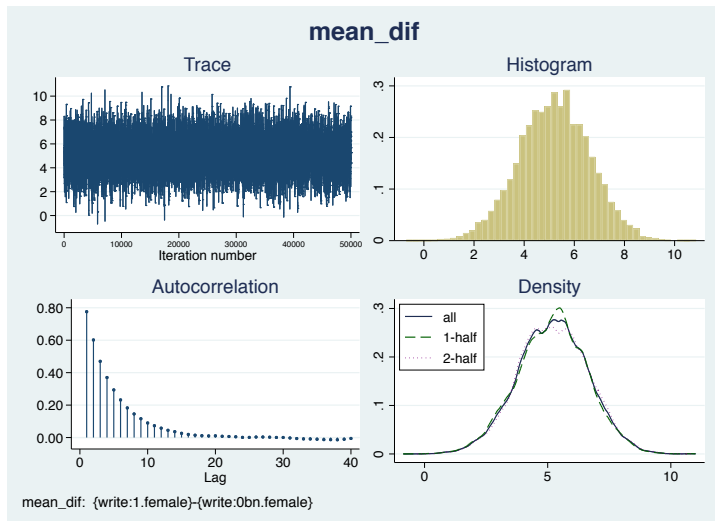
```
mean_dif : {write:1.female}-{write:0bn.female}
```

```
-----
          |                                     HPD
          |      Mean  StdDev  MCSE  [95% Cred. Interval]
-----+-----
mean_dif |  5.204619  1.420579  .018154  2.462988  8.021282
```

95 percent of the posterior differences in means falls in the HPD credible interval. It does not include zero.

Bayesgraph Diagnostics

bayesgraph diagnostics (mean_dif:{write:1.female}-{write:0b



Bayestest Interval

```
. bayestest interval (mean_dif:{write:1.female}- \\
> {write:0bn.female}), lower(0)
```

```
Interval tests      MCMC sample size =    50,000
mean_dif : {write:1.female}-{write:0bn.female} > 0
```

```
-----
          |      Mean      Std. Dev.      MCSE
-----+-----
mean_dif |      .9999      0.01000      .0000447
-----
```

99.99 percent of the posterior differences in means were greater than zero, i.e., the female mean is greater than the male mean with a probability of .99+.

Bayesstats Summary – Effect Size

```
. bayesstats summary (ES:({write:1.female}- ///  
>           {write:0bn.female})/ ///  
>(sqrt({var:1.female}+{var:0bn.female})/2)), hpd
```

Posterior summary statistics MCMC sample size = 50,000

```
ES : ({write:1.female}-{write:0bn.female})/  
      (sqrt({var:1.female}+{var:0bn.female})/2)
```

				HPD	
	Mean	StdDev	MCSE	[95% Cred. Interval]	
ES	.8504517	.2380119	.003057	.3994647	1.332236

A few Advantages of the Bayesian Approach

Credible intervals mean what students think confidence intervals mean. Avoids misunderstandings of p-values.

Does not depend on large sample theory.

Does not depend on normal distribution or homogeneity of variance to estimate the probability that group means differ.

It is possible to test differences in variances in the same way as testing differences in means.

Allows researchers to test the NULL.

Testing the NULL

In traditional statistical hypothesis testing, failure to reject the null hypothesis tells you nothing about the probability of the the NULL being true.

However, using the Bayesian approach it is possible to get an estimate of the probability for the NULL.

The Bayesian approach to testing the NULL involves defining a Region Of Practical Equivalence (ROPE). The ROPE is an interval which the researcher believes that values to be clinically or meaningfully equal.

If the preponderance of credible intervals fall within the ROPE then the researcher may conclude that the null hypothesis is likely to be true. You can obtain the probability of the difference in means falling within the ROPE using the **bayestest interval** post-estimation command.

The data for equivalence example

This example uses a subset of the hsbdemo dataset using **read** as the outcome of interest. The ROPE used was ± 3 .

female	N	mean	variance
male	47	56.2766	92.46531
female	58	56.06897	93.11797
Total	105	56.1619	91.94469

```
. fvset base none female

. bayesmh read i.female, noconstant          ///
  likelihood(t(({sigma2: i.female, nocons}), 7))  ///
prior({read:}, normal(50, 10000))           ///
prior({sigma2:}, igamma(.01, .01))         ///
init({sigma2:} 1) block({sigma2:})         ///
burnin(5000) mcmcsize(50000) hpd rseed(47)
```

The same as the first example.

Equivalence Output – Model Summary

Model summary

Likelihood:

```
read ~ t(xb_read, {sigma2:i.female,nocons},7)
```

Priors:

```
{read:i.female} ~ normal(50,10000)  
{sigma2:i.female} ~ igamma(.01,.01)
```

(1) Parameters are elements of the linear form `xb_read`.

Equivalence Output – Header

```
Bayesian t regression           MCMC iterations = 55,000
Random-walk Metropolis-        Burn-in = 5,000
Hastings sampling              MCMC sample size = 50,000
                                Number of obs = 105
                                Acceptance rate = .1729
                                Efficiency:  min = .07749
                                                avg = .09009
                                max = .1024
Log marginal likelihood = -408.10901
```

Equivalence Output – Estimates Table

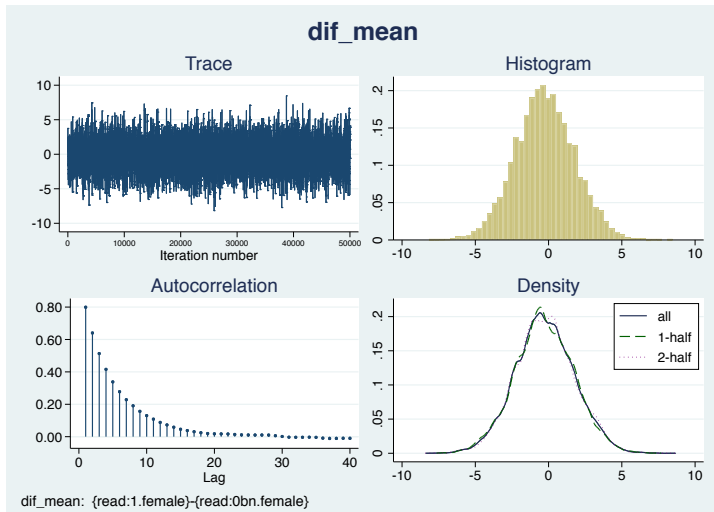
	Mean	Std. Dev.	MCSE	HPD [95% Cred. Interval]	
read					
male	56.17803	1.489381	.02082	53.29132	59.05976
female	55.91939	1.373013	.020109	53.27693	58.64388
sigma2					
male	81.09593	20.41242	.308998	46.10875	120.8572
female	85.04438	18.44299	.296289	52.59877	122.1864

MCMC standard errors can be reduced by increasing `mcmcsize()`.

bayesgraph diagnostics omitted due to time constraints but they looked pretty good.

Bayesgraph Diagnostics

```
. bayesgraph diagnostics (dif_mean:{read:1.female}- \\  
> {read:0bn.female})
```



Bayestest Interval

```
. bayestest interval (dif_mean:{read:1.female}- ///  
> {read:0bn.female}), lower(-3) upper(3)
```

Interval tests MCMC sample size = 50,000

```
    dif_mean : -3 < {read:1.female}-{read:0bn.f  
              emale} < 3
```

```
-----  
                  |           Mean       Std. Dev.        MCSE  
-----+-----  
    dif_mean |       .86162        0.34530       .0046795  
-----
```

86 percent of the posterior differences in means fell within the ROPE. This is pretty good evidence for the equivalence of means for **read** in the two groups.

The idea for the approach in this presentation came from the following article:

Kruschke, J.K. (2012). Bayesian Estimation Supersedes the t Test. *Journal of Experimental Psychology*, 142(2), 573–603.

I wish to thank Nikolay Balov of StataCorp for his assistance, particularly with the first version of **bayesmh** which did not have the t-distribution likelihood built-in.

The two-group Student's t-test provides an excellent framework for introducing undergraduate stat students to the Markov Chain Monte Carlo method of Bayesian analysis.

This concludes my presentation.