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?

\[
\frac{\bar{X}_1 - \bar{X}_2}{\text{something}}
\]

(1)

The \textit{something} part is a bit unclear in their minds.
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
   female | 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

```
```

Phil Ender
Bayes for Undergrads
\[
\text{diff} = \text{mean(male)} - \text{mean(female)} \quad t = -3.7341
\]

Ho: \(\text{diff} = 0\) \hspace{1cm} \text{degrees of freedom} = 198

Ha: \(\text{diff} < 0\) \hspace{1cm} Ha: \(\text{diff} \neq 0\) \hspace{1cm} Ha: \(\text{diff} > 0\)

Pr(\(T < t\)) = 0.0001 \hspace{1cm} Pr(\(|T| > |t|\)) = 0.0002 \hspace{1cm} Pr(\(T > t\)) = 0.9999
. esize twosample write, by(female)

Effect size based on mean comparison

Obs per group:
  male = 91
  female = 109

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Estimate</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen’s d</td>
<td>-.5302296</td>
<td>-.8127436</td>
</tr>
<tr>
<td>Hedges’s g</td>
<td>-.5282182</td>
<td>-.8096604</td>
</tr>
</tbody>
</table>

Phil Ender  Bayes for Undergrads
The Goal

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.

\[ \text{Posterior} \propto \text{Likelihood} \times \text{Prior} \] (2)
In this presentation the t-distribution will be used in the likelihood model of \texttt{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-probably 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 be easier than `bayesmh`.

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

Yes, this is straightforward 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.
. fvset base none female

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

There is a lot of stuff here, so let’s deconstruct this command in chunks.
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((i.female, nocons)), 7))

The syntax for the t likelihood is t(sigma2, 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) mcmcsiz(50000) ///
hpd rseed(47)
```

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

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

mcmcsiz(50000) - Some researchers recommend 100,000 mcmc reps. Increasing the mcmcsiz 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.
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
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>StdDev</th>
<th>MCSE</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>write</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>50.34901</td>
<td>1.170282</td>
<td>.016223</td>
<td>48.16482</td>
</tr>
<tr>
<td>female</td>
<td>55.55363</td>
<td>.8070589</td>
<td>.010622</td>
<td>53.92307</td>
</tr>
<tr>
<td><strong>var</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>96.41478</td>
<td>16.442</td>
<td>.235399</td>
<td>66.63293</td>
</tr>
<tr>
<td>female</td>
<td>55.14833</td>
<td>8.864754</td>
<td>.118853</td>
<td>38.65227</td>
</tr>
</tbody>
</table>

**Note:** Output edited to fit space.
Let’s Inspect the Posterior Distribution

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

Because of duplicate rows there are 21,414 observations in the dataset.
. bayesgraph trace _all, byparm
. bayesgraph ac _all, by parm

Autocorrelations

write:0bn.female
write:1.female
var:0bn.female
var:1.female

Lag

Graphs by parameter
. bayesgraph histogram _all, normal byparm

Histograms

write:0bn.female
write:1.female
var:0bn.female
var:1.female

Density
Normal density
Graphs by parameter
Bayesstats Summary – Mean Difference

```stata
.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}

<table>
<thead>
<tr>
<th></th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean  StdDev</td>
</tr>
<tr>
<td>mean_dif</td>
<td>5.204619 1.420579 .018154 2.462988 8.021282</td>
</tr>
</tbody>
</table>
```

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: 0bn.female})
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)

<table>
<thead>
<tr>
<th></th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>StdDev</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>ES</td>
<td>0.8504517</td>
</tr>
</tbody>
</table>

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

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

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.
This example uses a subset of the hsbdemo dataset using `read` as the outcome of interest. The ROPE used was $\pm 3$.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>47</td>
<td>56.2766</td>
<td>92.46531</td>
</tr>
<tr>
<td>female</td>
<td>58</td>
<td>56.06897</td>
<td>93.11797</td>
</tr>
<tr>
<td>Total</td>
<td>105</td>
<td>56.1619</td>
<td>91.94469</td>
</tr>
</tbody>
</table>
bayesmh code for equivalence example

. 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) mcmcsizer(50000) hpd rseed(47)

The same as the first example.
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.
Bayesian t regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 55,000
Burn-in = 5,000
MCMC sample size = 50,000
Number of obs = 105
Acceptance rate = .1729
Efficiency: min = .07749
avg = .09009
max = .1024

Log marginal likelihood = -408.10901
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>read</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>56.17803</td>
<td>1.489381</td>
<td>.02082</td>
<td>53.29132 59.05976</td>
</tr>
<tr>
<td>female</td>
<td>55.91939</td>
<td>1.373013</td>
<td>.020109</td>
<td>53.27693 58.64388</td>
</tr>
<tr>
<td>sigma2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>81.09593</td>
<td>20.41242</td>
<td>.308998</td>
<td>46.10875 120.8572</td>
</tr>
<tr>
<td>female</td>
<td>85.04438</td>
<td>18.44299</td>
<td>.296289</td>
<td>52.59877 122.1864</td>
</tr>
</tbody>
</table>

MCMC standard errors can be reduced by increasing mcmcsizethe().
Bayesgraph Diagnostics

Bayesgraph diagnostics omitted due to time constraints but they looked pretty good.
Posterior summary statistics

MCMC sample size = 50,000

dif_mean : {read:1.female} - {read:0bn.female}

<table>
<thead>
<tr>
<th></th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>MCSE</td>
<td>[95% Cred. Interval]</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>dif_mean</td>
<td>-.2586332</td>
</tr>
<tr>
<td></td>
<td>2.014832</td>
</tr>
<tr>
<td></td>
<td>0.027907</td>
</tr>
<tr>
<td></td>
<td>-4.107825</td>
</tr>
<tr>
<td></td>
<td>3.813532</td>
</tr>
</tbody>
</table>

This time the 95 percent credible interval does include zero.
. 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.female} < 3

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>dif_mean</td>
<td>.86162</td>
<td>0.34530</td>
<td>.0046795</td>
</tr>
</tbody>
</table>

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:

I wish to thank Nikolay Balov of StataCorp for his assistance, particularly with the first version of \texttt{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.