

ORIGIN ≡ 0

Robust Two Sample Analysis

The JAGS analog to the two-sample t-test follows directly from the format described in 010 & 020 MCMC Worksheets. Scaffolds are derived from J.K. Kruschke (K): *Doing Bayesian Data Analysis - A Tutorial with R, JAGS, and Stan*, available at <https://sites.google.com/site/doingbayesiandataanalysis/>. The only difference in One-sample versus Two-sample analysis is specifying a second variable in JAGS, derived from a column in the original data, that identifies group membership in the standard long-form data format. In K's scaffold, information about group column name is passed from the Driver level to the function `genMCMC()` at the 2nd level which in turn codes a numeric variable for JAGS.

```
from; Jags-Ymet-Xnom2grp-RrobustHetDRIVER
```

```
yName="Score"
```

```
xName="Group"
```

```
# Load the relevant model into R's working memory:
```

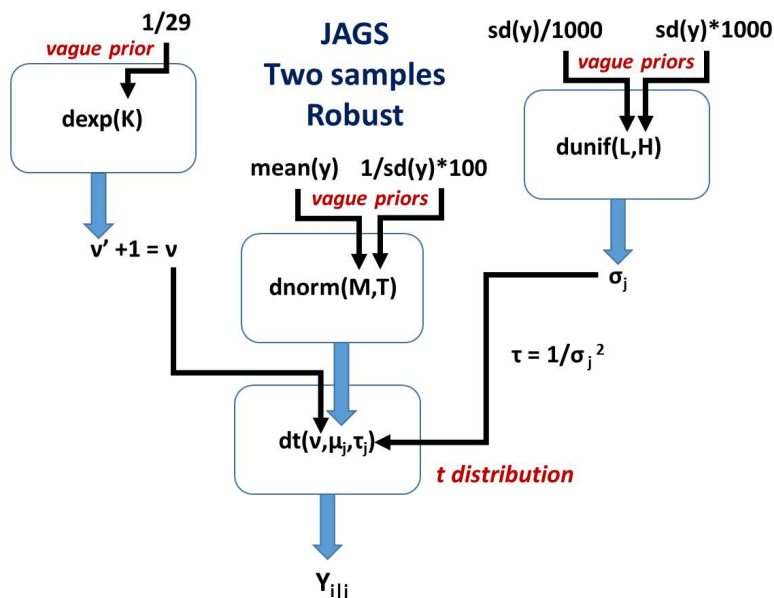
```
source("Jags-Ymet-Xnom2grp-MrobustHet.R")
```

```
#-----
```

```
# Generate the MCMC chain:
```

```
mcmcCoda = genMCMC( datFrm=myDataFrame , yName=yName , xName=xName ,
```

```
  numSavedSteps=50000 , saveName=fileNameRoot )
```



```
from: Jags-Ymet-X2grp-MrobustHet.R
```

```
# THE DATA.
```

```
y = as.numeric(datFrm[,yName])
```

```
x = as.numeric(as.factor(datFrm[,xName]))
```

```
xLevels = levels(as.factor(datFrm[,xName]))
```

```
Ntotal = length(y)
```

```
#-----
```

```
# THE MODEL.
```

```
modelString = "
```

```
model {
```

```
  for ( i in 1:Ntotal ) {
```

```
    y[i] ~ dt( mu[x[i]] , 1/sigma[x[i]]^2 , nu )
```

```
  }
```

```
  for ( j in 1:2 ) { # 2 groups
```

```
    mu[j] ~ dnorm( meanY , 1/(100*sdY)^2 )
```

```
    sigma[j] ~ dunif( sdY/1000 , sdY*1000 )
```

```
  }
```

```
  nu ~ dexp(1/30.0)
```

```
}
```

```
" # close quote for modelString
```

The Two-sample Robust model structure looks identical in structure to the One-sample model. The difference lies in addition of index j , with values of 1 or 2, specifying group membership. This means that for the distribution of y_i , there are two mean parameters μ_1 and μ_2 and two precision parameters τ_1 and τ_2 for each group. Priors for each of the four distribution parameters for y_i are carried out independently. There is, however, only one normality parameter ν , so only a single prior is determined for this exactly as in the One-sample model.

Two Sample t-tests:

For comparison with MCMC results, the following are results from running standard t-tests with and without assuming equality of variance between the two group:

```
> t.test(Score~Group,data=myDataFrame, var.equal=FALSE) # unequal variances
```

```
Welch Two Sample t-test
```

```
data: Score by Group
```

```
t = -1.958, df = 111.44, p-value = 0.05273
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-15.70602585  0.09366161
```

```
sample estimates:
```

```
mean in group Placebo mean in group Smart Drug
```

```
100.0351 107.8413
```

```
> t.test(Score~Group,data=myDataFrame, var.equal=TRUE) #equal variances
```

```
Two Sample t-test
```

```
data: Score by Group
```

```
t = -1.9249, df = 118, p-value = 0.05665
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-15.8369851  0.2246208
```

```
sample estimates:
```

```
mean in group Placebo mean in group Smart Drug
```

```
100.0351 107.8413
```

```
> var.test(Score~Group,data=myDataFrame)
```

```
F test to compare two variances
```

```
data: Score by Group
```

```
F = 0.49457, num df = 56, denom df = 62, p-value = 0.00823
```

```
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
```

```
0.2962264 0.8313852
```

```
sample estimates:
```

```
ratio of variances
```

```
0.4945694
```

```
> sd(myDataFrame$Score[myDataFrame$Group=="Placebo"])
```

```
[1] 17.8945
```

```
> sd(myDataFrame$Score[myDataFrame$Group=="Smart Drug"])
```

```
[1] 25.4452
```

The variance ratio F-test (see Biostatistics 170) indicates that the Welch Test above is more appropriate.

Two Sample MCMC Results:

> show(summaryInfo)

	Mean	Median	Mode	ESS	HDI _{mass}
mu[1]	99.2601668	99.2724000	99.2606903	30169.3	0.95
mu[2]	107.1396293	107.1440000	107.0112339	29957.3	0.95
muDiff	7.8794625	7.8838000	7.9411225	29253.3	0.95
sigma[1]	11.3367363	11.2154000	11.2390852	11999.9	0.95
sigma[2]	17.9232785	17.7795000	17.4576198	11537.8	0.95
sigmaDiff	6.5865422	6.5073000	6.4381478	22301.2	0.95
nu	3.8650320	3.5061300	2.9775310	5588.0	0.95
log10(nu)	0.5570688	0.5448280	0.5417733	7816.5	0.95
effSz	0.5319058	0.5291614	0.5007781	26530.7	0.95

	HDI _{low}	HDI _{high}	CompVal	PcntGtCompVal
mu[1]	95.79930000	102.9110000	NA	NA
mu[2]	101.78200000	112.3780000	NA	NA
muDiff	1.69020000	14.3730000	0	99.20802
sigma[1]	8.08601000	14.7977000	NA	NA
sigma[2]	12.84890000	23.3849000	NA	NA
sigmaDiff	1.16020000	12.0011100	0	99.31401
nu	1.61057000	6.8828200	NA	NA
log10(nu)	0.26902467	0.8689435	NA	NA
effSz	0.08978173	0.9767674	0	99.20802

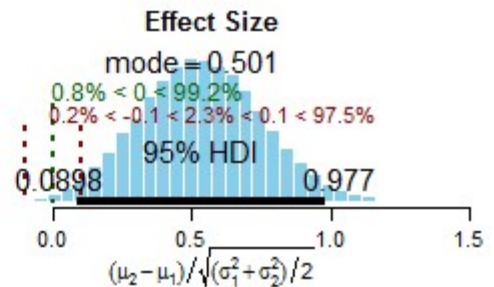
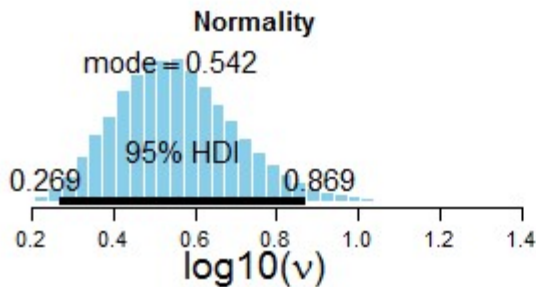
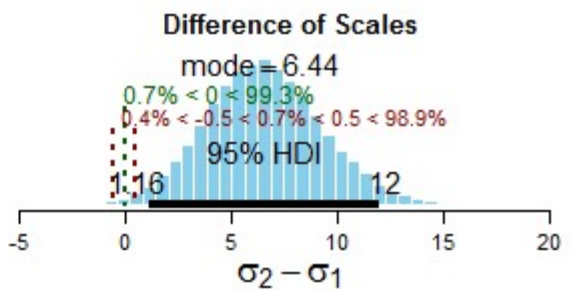
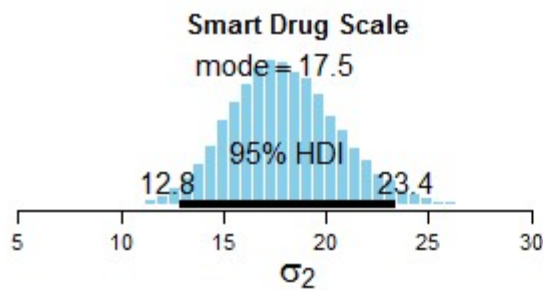
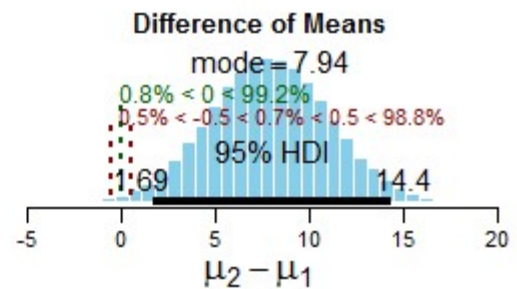
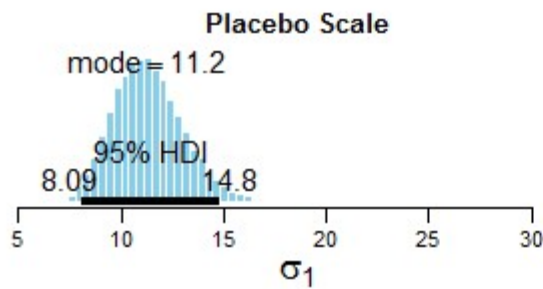
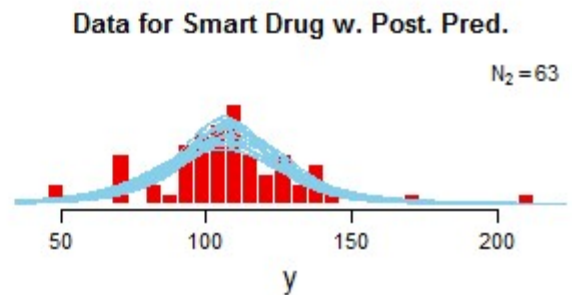
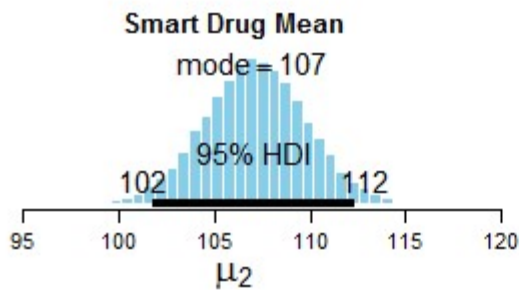
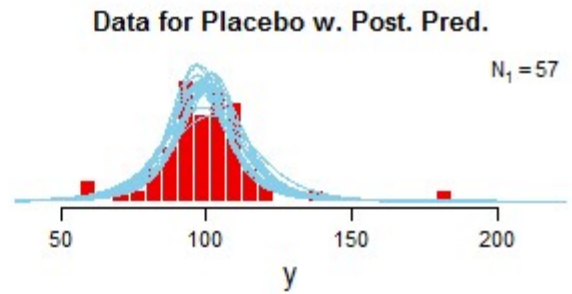
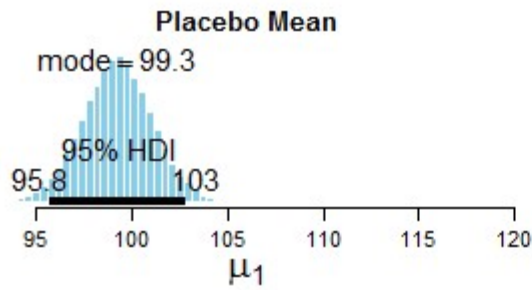
	ROPE _{low}	ROPE _{high}	PcntLtROPE	PcntInROPE	PcntGtROPE
mu[1]	NA	NA	NA	NA	NA
mu[2]	NA	NA	NA	NA	NA
muDiff	-0.5	0.5	0.5079898	0.6839863	98.80802
sigma[1]	NA	NA	NA	NA	NA
sigma[2]	NA	NA	NA	NA	NA
sigmaDiff	-0.5	0.5	0.4139917	0.6779864	98.90802
nu	NA	NA	NA	NA	NA
log10(nu)	NA	NA	NA	NA	NA
effSz	-0.1	0.1	0.2019960	2.3179536	97.48005

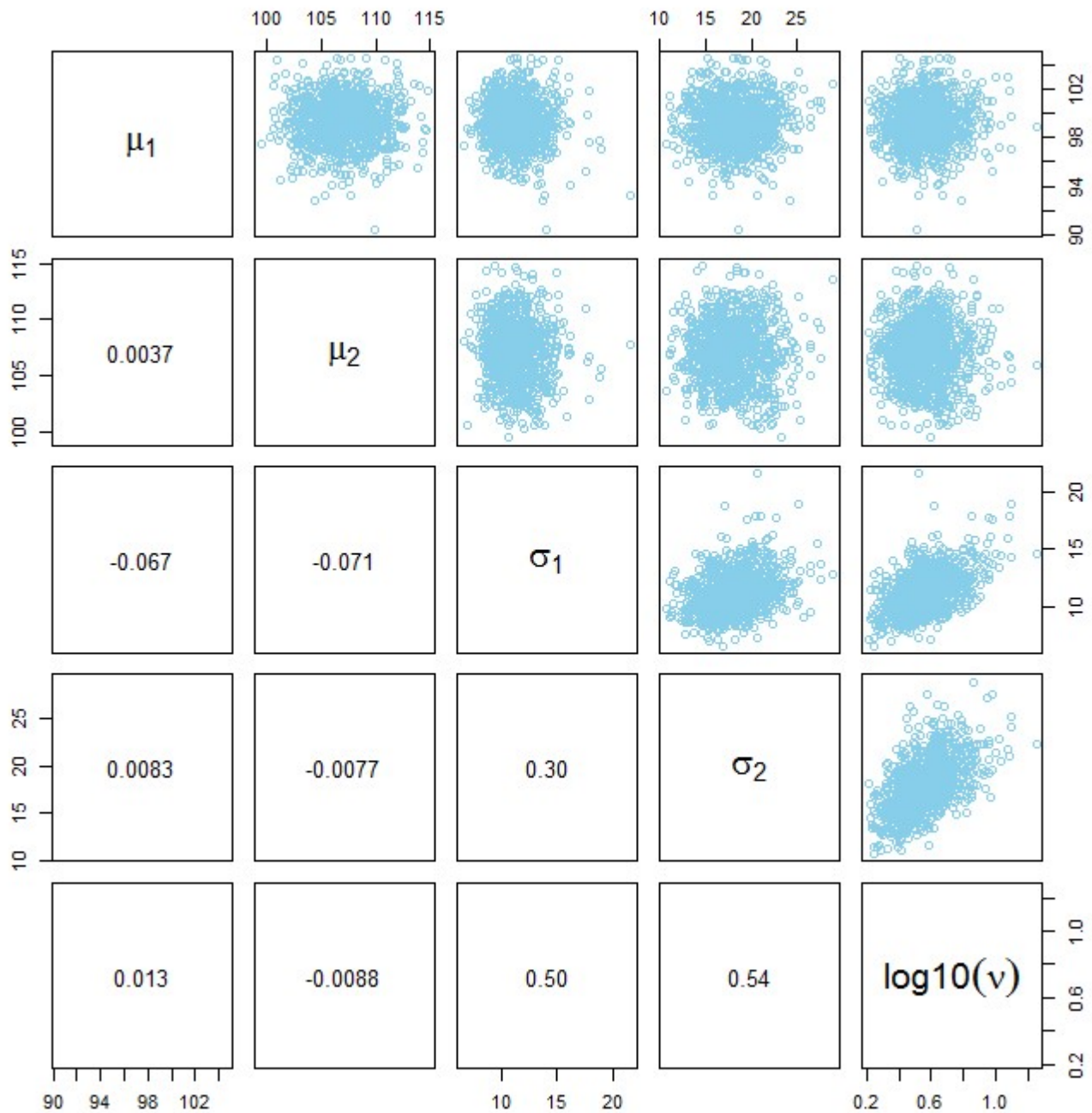
Summary of Findings:

The Variance Ratio test shows strong preference for the two groups having different variances (p-value = 0.8%). Similar results are seen in the MCMC plot for Differences in Scales (0.7 % for zero difference in the mode and outside the HDI). The MCMC Post Prediction Plot shows a difference quite clearly both in the distribution of the original data (red histogram) and the different shapes of the probable t-distributions for each group. The fit looks good. The Normality plot with mode = 0.542, quite a bit less than $\log_{10}(30) = 1.4771$, indicates preference for the t-distribution versus Normal distribution as a better fit for the data, given the evident data outliers. As a result of using the t-distribution, lower variance is seen in the fit of MCMC compared with measurement of standard deviation from the samples. MCMC difference in modes of μ includes only 0.8% of the distribution and outside the 95% HDI - impressive) whereas Welsh p-value for difference in means is 0.0523 (5.23% - not significant).

How do the Welsh/Variance Ratio tests and MCMC Results compare?

Parameter:	Welsh Test:	95% CI	MCMC using JAGS:	95% HDI
μ	sample means: Placebo = 100.0351 Smart Drug = 107.8413	for the difference: [0.0934 - 15.706]	modes: Placebo = 99.3 Smart Drug = 107	for the difference: [1.69 - 14.4]
σ	sample sd: Placebo = 17.8945 Smart Drug = 25.4452 Ratio = 0.4946	for the ratio: [0.2962 - 0.8314]	modes: Placebo = 11.2 Smart Drug = 17.5	for the difference: [1.16 - 12]





Correlation of v with both shape parameters σ_1 and σ_2 is to be expected from the definition of the t -distribution. Shape parameters σ_1 and σ_2 are show some correlation with each other. Otherwise correlations are low.