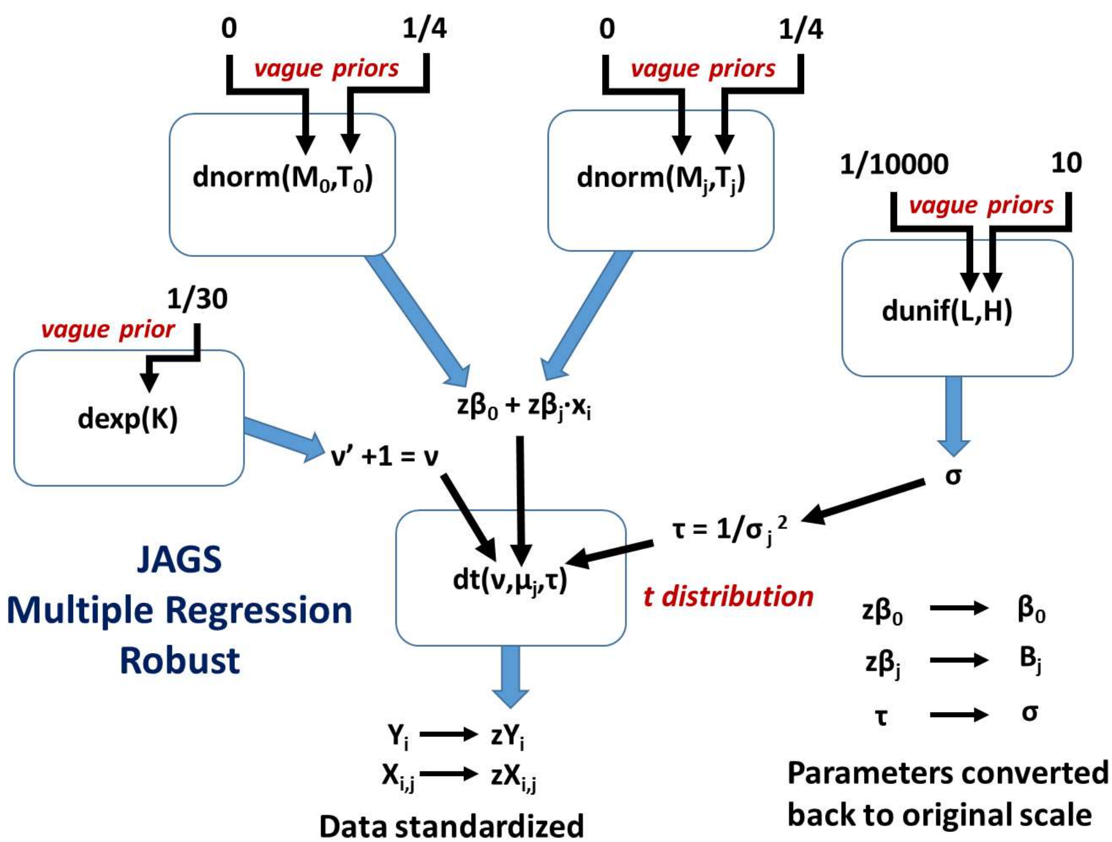
ORIGIN ≡ 0

Robust MCMC - Multiple Regression

Multiple Regression is a straightforward extension of Simple Regression where the dependent variable y, with values indexed by *i*, are interpreted to be related to multiple independent variables x_j. The linear function then becomes $y_i = \beta_0 + \beta_j x_{i,j} + \varepsilon_i$. The error term ε is commonly assumed belong to a Normal distribution. However in MCMC Robust regression using JAGS, the three parameter t-distribution is commonly utilized instead. JAGS Scaffolds are derived from the the text by J.K. Kruschke (K): *Doing Bayesian Data Analysis - A Tutorial with R, JAGS, and Stan*, available at https://sites.google.com/site/doingbayesiandataanalysis/.



The model description is nearly identical to that seen in Simple Regression (040 MCMC) except for the index *j* for each new independent variable on regression slope parameters β_j. As in other forms of Multiple regression, it is best for the independent variables to be uncorrelated. Otherwise in MCMC, marginal posterior distributions of parameters are wide (mostly reflecting vague priors). Correlations between independent variables can be observed in posterior pairwise plots, although even these will fail to display three-way and higher order dependencies. One approach to resolving this problem is to use Principal Components as summary independent variables, since these are guaranteed not to be correlated.

```
# THE DATA.
 y = data[,yName]
 x =
as.matrix(data[,xName],ncol-
=length(xName))
 cat("\nCORRELATION
MATRIX OF PREDICTORS:\n ")
 show( round(cor(x),3) )
 cat("\n")
 flush.console()
```

R script and report from: `Jags-YmetMulti-Mrobust.R`

```
CORRELATION MATRIX OF PREDICTORS:
            Spend  PrcntTake
Spend       1.000      0.593
PrcntTake   0.593      1.000
```

from: `Jags-YmetMulti-Mrobust.R`

```
# Specify the data in a list, for later shipment to JAGS:
dataList = list(
  x = x ,
  y = y ,
  Nx = dim(x)[2] ,
  Ntotal = dim(x)[1]
)
# THE MODEL.
modelString = "
# Standardize the data:
data {
  ym <- mean(y)
  ysd <- sd(y)
  for ( i in 1:Ntotal ) {
    zy[i] <- ( y[i] - ym ) / ysd
  }
  for ( j in 1:Nx ) {
    xm[j]  <- mean(x[,j])
    xsd[j] <-   sd(x[,j])
    for ( i in 1:Ntotal ) {
      zx[i,j] <- ( x[i,j] - xm[j] ) / xsd[j]
    }
  }
}
# Specify the model for standardized data:
 model {
  for ( i in 1:Ntotal ) {
    zy[i] ~ dt( zbeta0 + sum( zbeta[1:Nx] * zx[i,1:Nx] ) ,
1/zsigma^2 , nu )
  }
  # Priors vague on standardized scale:
  zbeta0 ~ dnorm( 0 , 1/2^2 )
  for ( j in 1:Nx ) {
    zbeta[j] ~ dnorm( 0 , 1/2^2 )
  }
  zsigma ~ dunif( 1.0E-5 , 1.0E+1 )
  nu ~ dexp(1/30.0)
  # Transform to original scale:
  beta[1:Nx] <- ( zbeta[1:Nx] / xsd[1:Nx] )*ysd
  beta0 <- zbeta0*ysd  + ym - sum( zbeta[1:Nx] *
xm[1:Nx] / xsd[1:Nx] )*ysd
  sigma <- zsigma*ysd
 }
 " # close quote for modelString
 # Write out modelString to a text file
```

I notice that K changes initial ranges for the priors here compared with Simple Regression. The new values, still quite vague, are reported in the model graph above. Potential difference in behavior or MCMC results in JAGS due to these changes has not been investigated.

Results from Standard Multiple Regression:

```
#Run standard linear regression:
LM=lm(SATT~Spend+PrcntTake,
        data=myData)
summary(LM)
confint(LM,level=0.95)
>sd(myData$SATT)
```

```
> summary(LM)
Call:
lm(formula = SATT ~ Spend + PrcntTake, data = myData)

Residuals:
    Min      1Q  Median      3Q     Max 
-88.400 -22.884   1.968  19.142  68.755 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 993.8317    21.8332  45.519  < 2e-16 ***
Spend        12.2865     4.2243   2.909  0.00553 ** 
PrcntTake    -2.8509     0.2151 -13.253  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 32.46 on 47 degrees of freedom
Multiple R-squared:  0.8195,	Adjusted R-squared:  0.8118 
F-statistic: 106.7 on 2 and 47 DF,  p-value: < 2.2e-16
```

```
> sd(myData$SATT)
[1] 74.82056
```

```
> confint(LM,level=0.95)
                 2.5 %      97.5 %
(Intercept) 949.908859 1037.754459
Spend         3.788291   20.784746
PrcntTake    -3.283679   -2.418179
```

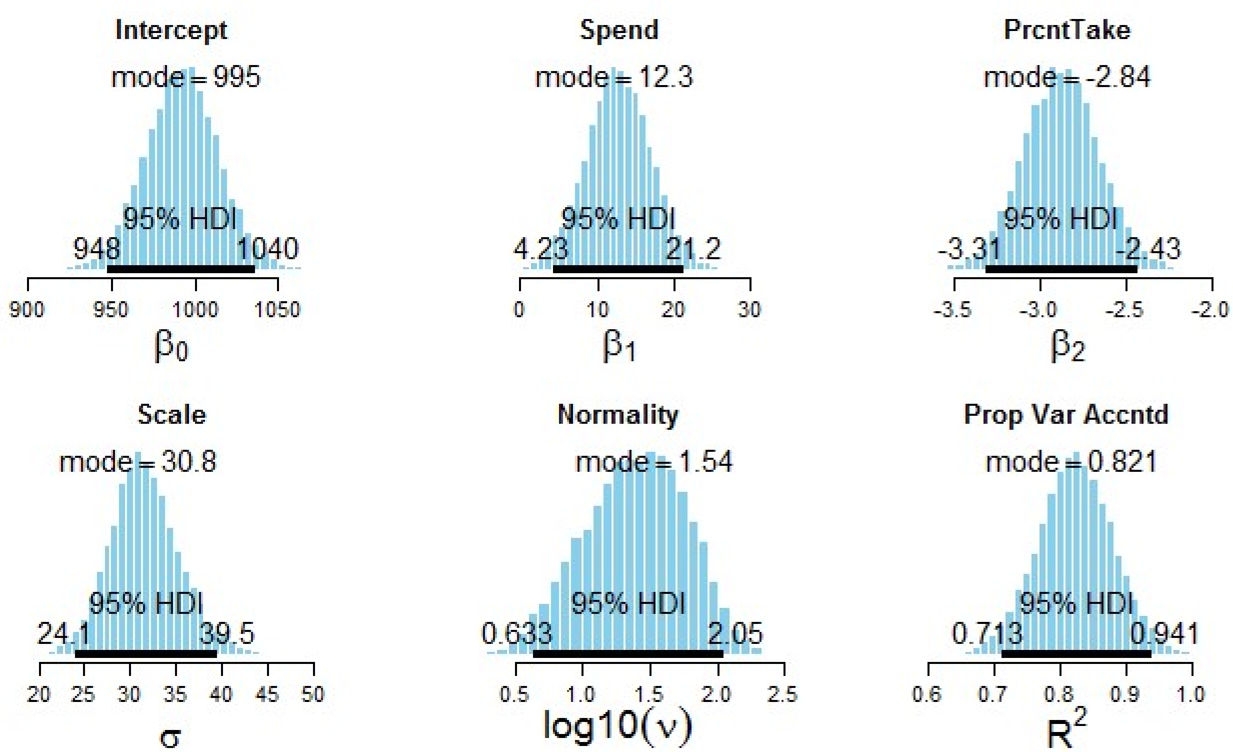
MCMC Results:

> show(summaryInfo)

	Mean	Median	Mode	ESS	HDImass
CHAIN	2.000000e+00	2.000000e+00	1.997412541	1.5	0.95
beta0	9.919843e+02	9.920485e+02	988.274883014	13590.9	0.95
beta[1]	1.274904e+01	1.272370e+01	12.602759142	12837.4	0.95
beta[2]	-2.876866e+00	-2.877040e+00	-2.849980901	13397.0	0.95
sigma	3.155556e+01	3.131390e+01	30.294068984	13866.9	0.95
zbeta0	-4.025332e-04	-5.621735e-04	-0.001656074	15000.0	0.95
zbeta[1]	2.322152e-01	2.317535e-01	0.229550801	12837.4	0.95
zbeta[2]	-1.029020e+00	-1.029085e+00	-1.019405887	13397.0	0.95
zsigma	4.217499e-01	4.185200e-01	0.404890042	13866.9	0.95
nu	3.468730e+01	2.540330e+01	11.874055562	10747.7	0.95
log10(nu)	1.388136e+00	1.404890e+00	1.525492918	11981.5	0.95

	HDIlow	HDIhigh	CompVal	PcntGtCompVal	ROPElow	ROPEhigh
CHAIN	1.0000000	3.000000	NA	NA	NA	NA
beta0	948.7530000	1036.160000	NA	NA	NA	NA
beta[1]	4.0370600	21.063000	NA	NA	NA	NA
beta[2]	-3.3308400	-2.449830	NA	NA	NA	NA
sigma	24.2185000	39.451800	NA	NA	NA	NA
zbeta0	-0.1262110	0.119524	NA	NA	NA	NA
zbeta[1]	0.0735324	0.383649	NA	NA	NA	NA
zbeta[2]	-1.1914000	-0.876276	NA	NA	NA	NA
zsigma	0.3236880	0.527285	NA	NA	NA	NA
nu	1.8808400	95.717300	NA	NA	NA	NA
log10(nu)	0.6489033	2.079966	NA	NA	NA	NA

	PcntLtROPE	PcntInROPE	PcntGtROPE
CHAIN	NA	NA	NA
beta0	NA	NA	NA
beta[1]	NA	NA	NA
beta[2]	NA	NA	NA
sigma	NA	NA	NA
zbeta0	NA	NA	NA
zbeta[1]	NA	NA	NA
zbeta[2]	NA	NA	NA
zsigma	NA	NA	NA
nu	NA	NA	NA
log10(nu)	NA	NA	NA



Summary of Findings:

Parameter:	Simple Regression: point estimate	95% CI	MCMC using JAGS: mode	95% HDI
β_0	993.8317	[949.91 - 1037.75]	995	[948 - 1040]
β_1	12.2865	[3.788 - 20.785]	12.3	[4.23 - 21.2]
β_2	-2.8509	[-3.284 - -2.418]	-2.84	[-3.31 - -2.43]
σ	74.82056		30.8	[24.1 - 39.5]

Normality parameter ν = 1.54 indicates that y_i are essentially Normally Distributed according to the criterion ν > log10(30) = 1.4771.