Centered Model for the SAT Data

```
> sat = read.table("http://www.utstat.toronto.edu/~brunner/data/legal/openSAT.data.txt")
> summary(lm(GPA ~ VERBAL + MATH, data=sat))

Call:
lm(formula = GPA ~ VERBAL + MATH, data = sat)

Residuals:
     Min       1Q   Median       3Q      Max
-2.24875 -0.35113  0.04659  0.38745  1.03527

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.6062975  0.4414062   1.374    0.171
VERBAL      0.0023072  0.0005522   4.178 4.42e-05 ***
MATH        0.0009999  0.0006093   1.641    0.102
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5484 on 197 degrees of freedom
Multiple R-squared: 0.1161, Adjusted R-squared: 0.1071
F-statistic: 12.93 on 2 and 197 DF,  p-value: 5.284e-06

> # Center the variables
> attach(sat)
> x1 = VERBAL-mean(VERBAL); x2 = MATH-mean(MATH) # Could do y = GPA-mean(GPA)
> cmodel = lm(GPA ~ x1 + x2) # Or lm(Y ~ 0 + x1 + x2)
> summary(cmodel)

Call:
lm(formula = GPA ~ x1 + x2)

Residuals:
     Min       1Q   Median       3Q      Max
-2.24875 -0.35113  0.04659  0.38745  1.03527

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.6300000  0.0387764  67.825  < 2e-16 ***
x1          0.0023072  0.0005522   4.178 4.42e-05 ***
x2          0.0009999  0.0006093   1.641    0.102
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5484 on 197 degrees of freedom
Multiple R-squared: 0.1161,    Adjusted R-squared: 0.1071
F-statistic: 12.93 on 2 and 197 DF,  p-value: 5.284e-06
```

Centered Model for the trees Data

```
> # Use of centering to reduce multicolinearity in polynomial regression
>
> with(trees,    # Do the following using trees
+ {Gsq = Girth^2
+ cGirth = Girth-mean(Girth)
+ cGsq = cGirth^2
+ cor( cbind(Girth,Gsq,cGirth,cGsq) )
+ } # End expression to be evaluated
+ ) # End within trees

           Girth       Gsq    cGirth      cGsq
Girth  1.0000000 0.9930404 1.0000000 0.4379578
Gsq    0.9930404 1.0000000 0.9930404 0.5407885
cGirth 1.0000000 0.9930404 1.0000000 0.4379578
cGsq   0.4379578 0.5407885 0.4379578 1.0000000

>
> forest = within(trees,    # Do the following to trees
+ {cGirth = Girth-mean(Girth)
+ cGsq = cGirth^2
+ cHeight = Height-mean(Height)
+ } # End expression to be evaluated
+ ) # End within trees

> model = lm(Volume ~ cHeight+cGirth+cGsq, data=forest); summary(model)

Call:
lm(formula = Volume ~ cHeight + cGirth + cGsq, data = forest)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2928 -1.6693 -0.1018  1.7851  4.3489

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 27.61093    0.64314  42.931  < 2e-16 ***
cHeight      0.37639    0.08823   4.266 0.000218 ***
cGirth       4.23255    0.19630  21.561  < 2e-16 ***
cGsq         0.26862    0.04590   5.852 3.13e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.625 on 27 degrees of freedom
Multiple R-squared:  0.9771,   Adjusted R-squared:  0.9745
F-statistic: 383.2 on 3 and 27 DF,  p-value: < 2.2e-16
```