> crime<- read.table("http://www.statsci.org/data/general/uscrime.txt",header=TRUE)

# remove the response variable (it's in the 16th column)

> vars<-crime[-16]

> pca<-prcomp(vars, scale = TRUE)

> summary(pca)

Importance of components:

PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8

Standard deviation 2.4534 1.6739 1.4160 1.07806 0.97893 0.74377 0.56729 0.55444

Proportion of Variance 0.4013 0.1868 0.1337 0.07748 0.06389 0.03688 0.02145 0.02049

Cumulative Proportion 0.4013 0.5880 0.7217 0.79920 0.86308 0.89996 0.92142 0.94191

PC9 PC10 PC11 PC12 PC13 PC14 PC15

Standard deviation 0.48493 0.44708 0.41915 0.35804 0.26333 0.2418 0.06793

Proportion of Variance 0.01568 0.01333 0.01171 0.00855 0.00462 0.0039 0.00031

Cumulative Proportion 0.95759 0.97091 0.98263 0.99117 0.99579 0.9997 1.00000

#get the eigenvector of the matrix

> eigen<-pca$rotation

> eigen

PC1 PC2 PC3 PC4 PC5 PC6

M -0.30371194 0.06280357 0.1724199946 -0.02035537 -0.35832737 -0.449132706

So -0.33088129 -0.15837219 0.0155433104 0.29247181 -0.12061130 -0.100500743

Ed 0.33962148 0.21461152 0.0677396249 0.07974375 -0.02442839 -0.008571367

Po1 0.30863412 -0.26981761 0.0506458161 0.33325059 -0.23527680 -0.095776709

Po2 0.31099285 -0.26396300 0.0530651173 0.35192809 -0.20473383 -0.119524780

LF 0.17617757 0.31943042 0.2715301768 -0.14326529 -0.39407588 0.504234275

M.F 0.11638221 0.39434428 -0.2031621598 0.01048029 -0.57877443 -0.074501901

Pop 0.11307836 -0.46723456 0.0770210971 -0.03210513 -0.08317034 0.547098563

NW -0.29358647 -0.22801119 0.0788156621 0.23925971 -0.36079387 0.051219538

U1 0.04050137 0.00807439 -0.6590290980 -0.18279096 -0.13136873 0.017385981

U2 0.01812228 -0.27971336 -0.5785006293 -0.06889312 -0.13499487 0.048155286

Wealth 0.37970331 -0.07718862 0.0100647664 0.11781752 0.01167683 -0.154683104

Ineq -0.36579778 -0.02752240 -0.0002944563 -0.08066612 -0.21672823 0.272027031

Prob -0.25888661 0.15831708 -0.1176726436 0.49303389 0.16562829 0.283535996

Time -0.02062867 -0.38014836 0.2235664632 -0.54059002 -0.14764767 -0.148203050

PC7 PC8 PC9 PC10 PC11 PC12

M -0.15707378 -0.55367691 0.15474793 -0.01443093 0.39446657 0.16580189

So 0.19649727 0.22734157 -0.65599872 0.06141452 0.23397868 -0.05753357

Ed -0.23943629 -0.14644678 -0.44326978 0.51887452 -0.11821954 0.47786536

Po1 0.08011735 0.04613156 0.19425472 -0.14320978 -0.13042001 0.22611207

Po2 0.09518288 0.03168720 0.19512072 -0.05929780 -0.13885912 0.19088461

LF -0.15931612 0.25513777 0.14393498 0.03077073 0.38532827 0.02705134

M.F 0.15548197 -0.05507254 -0.24378252 -0.35323357 -0.28029732 -0.23925913

Pop 0.09046187 -0.59078221 -0.20244830 -0.03970718 0.05849643 -0.18350385

NW -0.31154195 0.20432828 0.18984178 0.49201966 -0.20695666 -0.36671707

U1 -0.17354115 -0.20206312 0.02069349 0.22765278 -0.17857891 -0.09314897

U2 -0.07526787 0.24369650 0.05576010 -0.04750100 0.47021842 0.28440496

Wealth -0.14859424 0.08630649 -0.23196695 -0.11219383 0.31955631 -0.32172821

Ineq 0.37483032 0.07184018 -0.02494384 -0.01390576 -0.18278697 0.43762828

Prob -0.56159383 -0.08598908 -0.05306898 -0.42530006 -0.08978385 0.15567100

Time -0.44199877 0.19507812 -0.23551363 -0.29264326 -0.26363121 0.13536989

PC13 PC14 PC15

M -0.05142365 0.04901705 0.0051398012

So -0.29368483 -0.29364512 0.0084369230

Ed 0.19441949 0.03964277 -0.0280052040

Po1 -0.18592255 -0.09490151 -0.6894155129

Po2 -0.13454940 -0.08259642 0.7200270100

LF -0.27742957 -0.15385625 0.0336823193

M.F 0.31624667 -0.04125321 0.0097922075

Pop 0.12651689 -0.05326383 0.0001496323

NW 0.22901695 0.13227774 -0.0370783671

U1 -0.59039450 -0.02335942 0.0111359325

U2 0.43292853 -0.03985736 0.0073618948

Wealth -0.14077972 0.70031840 -0.0025685109

Ineq -0.12181090 0.59279037 0.0177570357

Prob -0.03547596 0.04761011 0.0293376260

Time -0.05738113 -0.04488401 0.0376754405

# get the first 4 pc

> pc<-pca$x[,1:4]

#fit a linear regression model with the these 4 pc

> crimepc<-as.data.frame(cbind(pc,crime$Crime))

> modelpca<-lm(V5~.,crimepc)

> summary(modelpca)

Call:

lm(formula = V5 ~ ., data = crimepc)

Residuals:

Min 1Q Median 3Q Max

-557.76 -210.91 -29.08 197.26 810.35

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 905.09 49.07 18.443 < 2e-16 \*\*\*

PC1 65.22 20.22 3.225 0.00244 \*\*

PC2 -70.08 29.63 -2.365 0.02273 \*

PC3 25.19 35.03 0.719 0.47602

PC4 69.45 46.01 1.509 0.13872

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 336.4 on 42 degrees of freedom

Multiple R-squared: 0.3091, Adjusted R-squared: 0.2433

F-statistic: 4.698 on 4 and 42 DF, p-value: 0.003178

#Get the parameters for the original model (scaled)

> beta0<-modelpca$coefficients[1]

> beta0

(Intercept)

905.0851

> betas<-modelpca$coefficients[2:5]

> betas

PC1 PC2 PC3 PC4

65.21593 -70.08312 25.19408 69.44603

#coefficents equals beta times eigenvector matrix

> Vjk<-eigen[,1:4] #k=1,2,3,4 here

> Vjk

PC1 PC2 PC3 PC4

M -0.30371194 0.06280357 0.1724199946 -0.02035537

So -0.33088129 -0.15837219 0.0155433104 0.29247181

Ed 0.33962148 0.21461152 0.0677396249 0.07974375

Po1 0.30863412 -0.26981761 0.0506458161 0.33325059

Po2 0.31099285 -0.26396300 0.0530651173 0.35192809

LF 0.17617757 0.31943042 0.2715301768 -0.14326529

M.F 0.11638221 0.39434428 -0.2031621598 0.01048029

Pop 0.11307836 -0.46723456 0.0770210971 -0.03210513

NW -0.29358647 -0.22801119 0.0788156621 0.23925971

U1 0.04050137 0.00807439 -0.6590290980 -0.18279096

U2 0.01812228 -0.27971336 -0.5785006293 -0.06889312

Wealth 0.37970331 -0.07718862 0.0100647664 0.11781752

Ineq -0.36579778 -0.02752240 -0.0002944563 -0.08066612

Prob -0.25888661 0.15831708 -0.1176726436 0.49303389

Time -0.02062867 -0.38014836 0.2235664632 -0.54059002

> alpha<-Vjk %\*% betas #aj=sum(beta\_k\*Vjk),

e.g. for predictor M, beta\_k= 65.21593 -70.08312 25.19408 69.44603

Vjk= -0.30371194 0.06280357 0.1724199946 -0.02035537

a\_(M)= 65.21593\*(-0.30371194)-70.08312\*0.06280357+25.19408\*0.1724199946+69.44603\* (-0.02035537)= -21.277963

> alpha

[,1]

M -21.277963

So 10.223091

Ed 14.352610

Po1 63.456426

Po2 64.557974

LF -14.005349

M.F -24.437572

Pop 39.830667

NW 15.434545

U1 -27.222281

U2 1.425902

Wealth 38.607855

Ineq -27.536348

Prob 3.295707

Time -6.612616

Beta0 and alpha are the intercept and coefficients for the original model(scaled).

For the un-scaled model:

#get the un-scaled intercept

> beta\_org <-beta0-sum(alpha\* mean/sd)

> beta\_org

(Intercept)

1666.485

>

#get the un-scaled coefficents for each input

> alpha\_org<- alpha/sd

> alpha\_org

[,1]

M -16.9307630

So 21.3436771

Ed 12.8297238

Po1 21.3521593

Po2 23.0883154

LF -346.5657125

M.F -8.2930969

Pop 1.0462155

NW 1.5009941

U1 -1509.9345216

U2 1.6883674

Wealth 0.0400119

Ineq -6.9020218

Prob 144.9492678

Time -0.9330765