1/28/2010 36-402/608 ADA-II H. Seltman

Breakout #6 Results

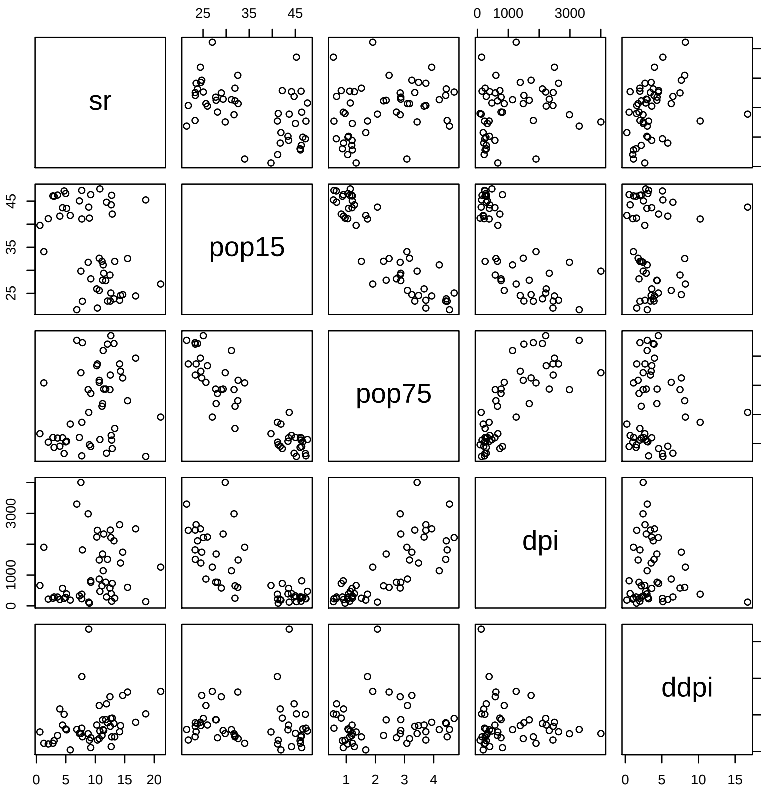
Under the life-cycle savings hypothesis as developed by Franco Modigliani, the savings ratio (aggregate personal saving divided by disposable income) is explained by per-capita disposable income, the percentage rate of change in per-capita disposable income, and two demographic variables: the percentage of population less than 15 years old and the percentage of the population over 75 years old. The data are averaged over the decade 1960-1970 to remove the business cycle or other short-term fluctuations.

```
data(LifeCycleSavings)
lcs = LifeCycleSavings
dim(lcs) # [1] 50  5
sapply(lcs, function(v) mean(is.na(v)))
#   sr pop15 pop75   dpi  ddpi
#    0     0     0     0     0

?LifeCycleSavings
# sr     numeric  aggregate personal savings
# pop15  numeric  % of population under 15
# pop75  numeric  % of population over 75
# dpi    numeric  real per-capita disposable income
# ddpi   numeric  % growth rate of dpi

rownames(lcs)[1:5]
#[1] "Australia" "Austria"   "Belgium"   "Bolivia"   "Brazil"

pairs(lcs)
dev.copy(pdf, "BO6pairs.pdf"); dev.off()
```



1: Which predictors look most useful? What transformation might be worth checking? What do you see in terms of outliers (in the X and Y directions)?

```
m0 = lm(sr~., lcs)
summary(m0)
#Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
#(Intercept) 28.5660865  7.3545161   3.884 0.000334 ***
#pop15       -0.4611931  0.1446422  -3.189 0.002603 **
#pop75       -1.6914977  1.0835989  -1.561 0.125530
#dpi         -0.0003369  0.0009311  -0.362 0.719173
#ddpi         0.4096949  0.1961971   2.088 0.042471 *
#Residual standard error: 3.803 on 45 degrees of freedom
#Multiple R-squared: 0.3385,     Adjusted R-squared: 0.2797
```

2: Summarize what you learn from the regression results.

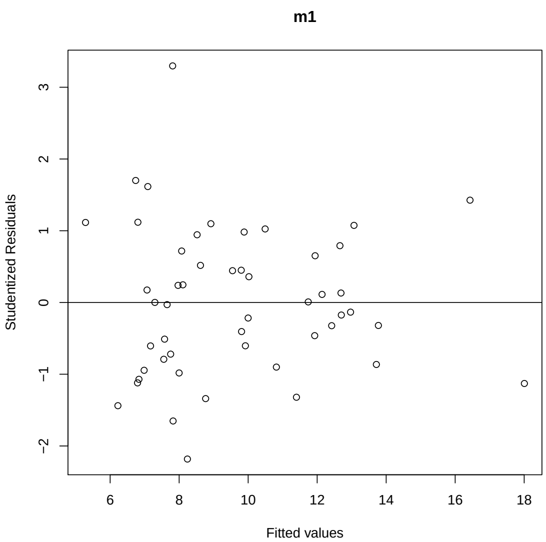
```
library(MASS) # for stepAIC()
m1 = stepAIC(lm(sr~.^2,lcs))
summary(m1)
#Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
#(Intercept) 16.5287997  4.3729241   3.780 0.000459 ***
#pop15       -0.2023669  0.0981090  -2.063 0.044943 *
#dpi         -0.0027411  0.0011774  -2.328 0.024457 *
#ddpi         0.0462479  0.2439993   0.190 0.850521
#dpi:ddpi     0.0008171  0.0003593   2.274 0.027802 *
#Residual standard error: 3.698 on 45 degrees of freedom
#Multiple R-squared: 0.3745,     Adjusted R-squared: 0.3189
```

3: Summarize what you learn from the new regression results.

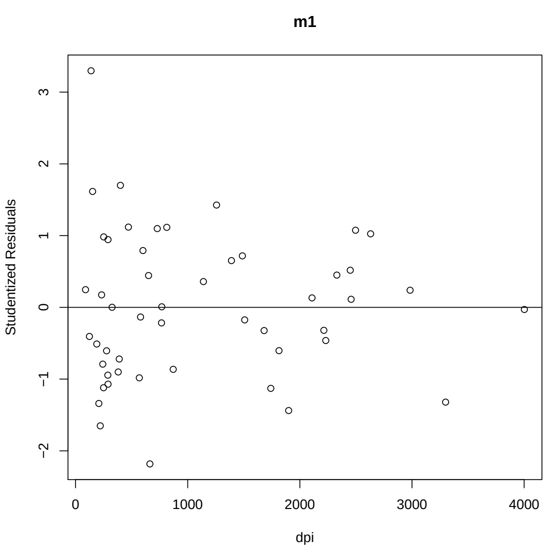
```
# A residual plotting function
rp = function(mdl, xname=NULL, fname=NULL) {
  res = rstudent(mdl)
  if (is.null(xname)) {
    x = fitted(mdl)
    xname = "Fitted values"
  } else {
    x = mdl$model[,xname]
  }
  plot(x, res, xlab=xname, ylab="Studentized Residuals",
       main = deparse(substitute(mdl)))
  abline(h=0)
  if (!is.null(fname)) {
    dev.copy(pdf, paste(fname,".pdf",sep=""))
    dev.off()
  }
  invisible(NULL)
}
```

4: Comment on the R code.

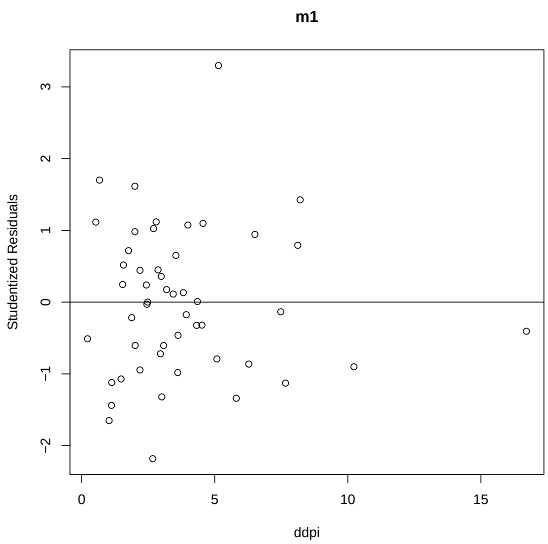
```
rp(m1, fname="BO6RF")
```




```
rp(m1, "dpi", "BO6Rdpi")
```



```
rp(m1, "ddpi", "BO6Rddpi")
```

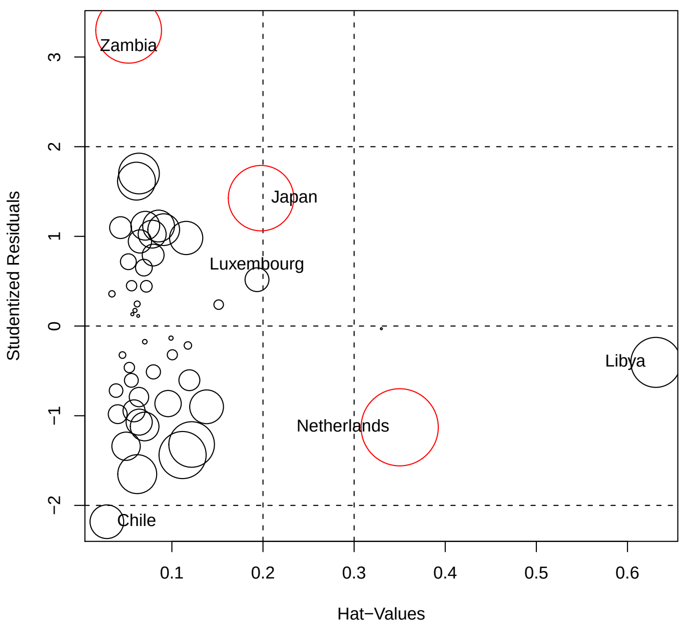


5: Comment on the residual plots

```
summary(influence.measures(m1), digits=2)
#Potentially influential observations of
#         lm(formula = sr ~ ., data = lcs) :
#              dfb.1_ dfb.pp15 dfb.dpi dfb.ddpi dfb.dp:d
#Luxembourg      0.12    -0.14    0.15     0.07    -0.20
#Netherlands     0.12    -0.15    0.51     0.26    -0.73
#United States   0.01    -0.01   -0.02     0.00     0.00
#Zambia         -0.17     0.25    0.06     0.27    -0.13
#Libya           0.09    -0.01   -0.24    -0.51     0.26
#
#             dffit cov.r cook.d  hat
#Luxembourg     0.25  1.35   0.01 0.19
#Netherlands   -0.83  1.49   0.14 0.35
#United States -0.02  1.67   0.00 0.33
#Zambia         0.78  0.39   0.10 0.05
#Libya         -0.53  2.98   0.06 0.63
```

6: Comment on the flagged influence measures.

```
library(car) # for influencePlot()
out = influencePlot(m1) # Right click to stop identifying points
dev.copy(pdf, "BO6IP.pdf"); dev.off()
```




```
influence.measures(m1)$infmat[out,] # (repeats manually deleted)
#                 dfb.1_    dfb.pp15      dfb.dpi     dfb.ddpi     dfb.dp:d
#Chile       -0.03674514 -0.04284283  0.03470584  0.08878191  0.01305477
#Japan       -0.01171377  0.02913492 -0.44015044 -0.09193093  0.53233681
#
#                 dffit      cov.r      cook.d         hat
#Chile       -0.3747104 0.6891229 0.02591532 0.02864327
#Japan        0.7082681 1.1127700 0.09807518 0.19783607
```

7: Comment on the additional, manually flagged influence measures

```
round(rbind(lcs[out,], allmean=apply(lcs,2,mean), allsd=apply(lcs,2,sd)),2)
#               sr pop15 pop75     dpi  ddpi
#Chile        0.60 39.74  1.34  662.86  2.67
#Japan       21.10 27.01  1.91 1257.28  8.21
#Luxembourg  10.35 21.80  3.73 2449.39  1.57
#Netherlands 14.65 24.71  3.25 1740.70  7.66
#Zambia      18.56 45.25  0.56  138.33  5.14
#Libya        8.89 43.69  2.07  123.58 16.71
#allmean      9.67 35.09  2.29 1106.76  3.76
#allsd        4.48  9.15  1.29  990.87  2.87
```

8: Comment on possible actions to be taken.