HW6

DANDONG TU

2017/10/2

1

a)

```
library(alr4)
```

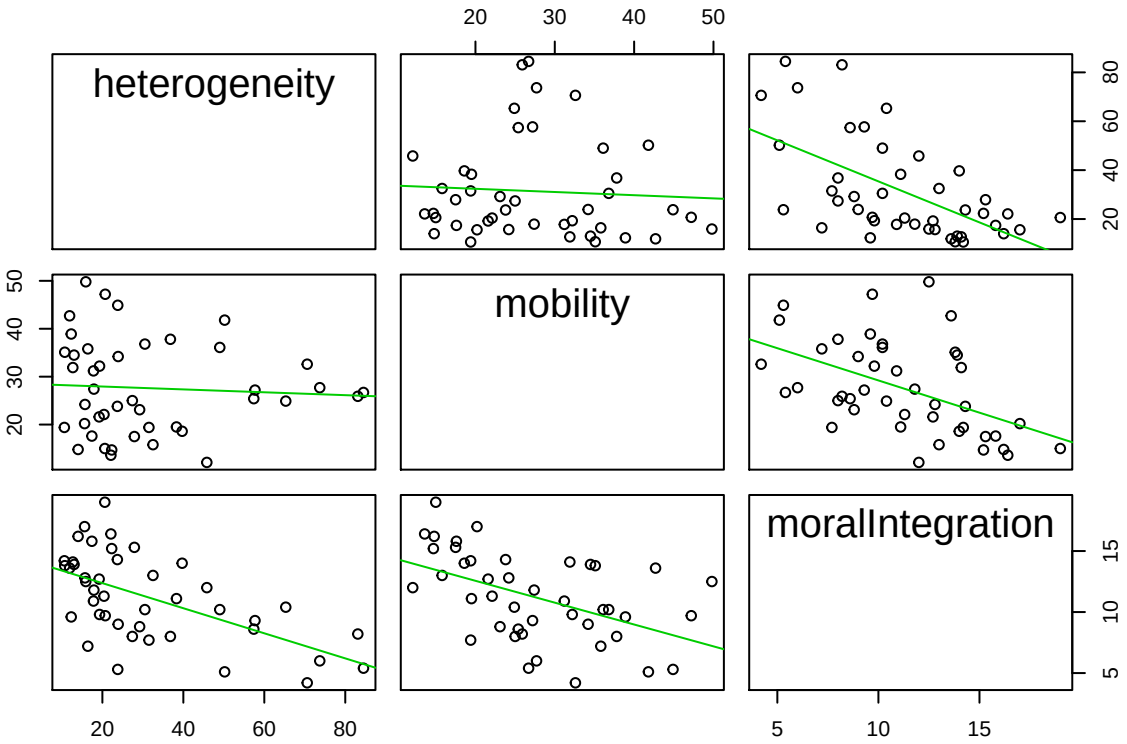
```
## Loading required package: car
```

```
## Loading required package: effects
```

```
##
## Attaching package: 'effects'
```

```
## The following object is masked from 'package:car':
##
##     Prestige
```

```
data1=read.table("/Users/dandongtu/Downloads/Angell.txt",header= TRUE)
scatterplotMatrix(~ heterogeneity+mobility+moralIntegration,
                  data=data1,
                  diagonal="none",
                  smoother=FALSE)
```



From the scatter polt matrix we observed relationship between the response and predictors. The least-squares lines indicate the slightly negative relationships with response for both predictors and these are reasonalble

summaries.

b)

```
lm.h=lm(data1$moralIntegration~data1$heterogeneity)
summary(lm.h)
```

```
##
## Call:
## lm(formula = data1$moralIntegration ~ data1$heterogeneity)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6780 -2.6099  0.2493  2.2971  6.6931
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          14.42355    0.82507  17.482  < 2e-16 ***
## data1$heterogeneity -0.10275    0.02212  -4.645 3.49e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.926 on 41 degrees of freedom
## Multiple R-squared:  0.3448, Adjusted R-squared:  0.3288
## F-statistic: 21.58 on 1 and 41 DF,  p-value: 3.486e-05
```

From the summary we observed the linear regress of $y = -0.1028x + 14.4236$. The $\hat{\beta}_1 = -0.1028$ and $\hat{\beta}_0 = 14.4236$. Meanwhile the coefficient of determination which is R^2 is equal to 0.329 which indicates that model explains 32.9% of the predictor of the response data.

c)

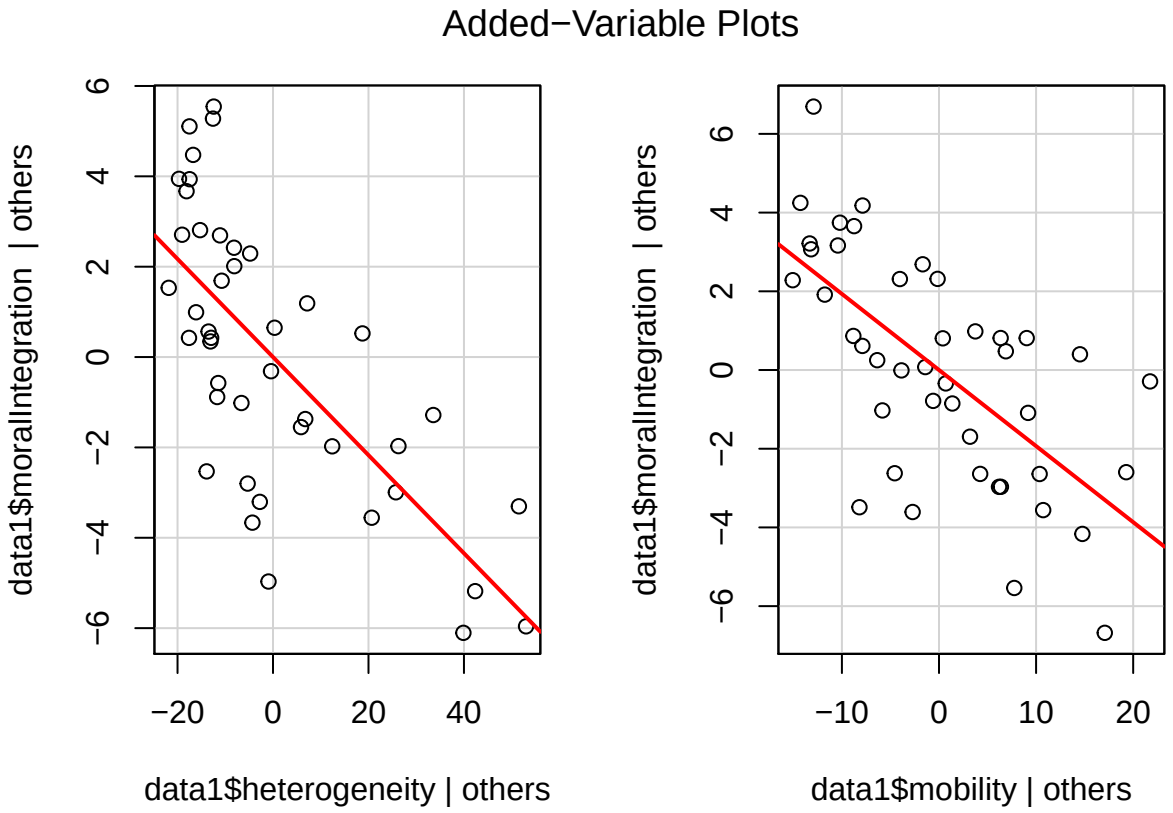
```
lm.both=lm(data1$moralIntegration~data1$heterogeneity+data1$mobility)
summary(lm.both)
```

```
##
## Call:
## lm(formula = data1$moralIntegration ~ data1$heterogeneity + data1$mobility)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.071 -1.194 -0.206  1.738  4.195
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          19.94076    1.19265  16.720  < 2e-16 ***
## data1$heterogeneity -0.10856    0.01699  -6.389 1.34e-07 ***
## data1$mobility      -0.19331    0.03543  -5.456 2.74e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.243 on 40 degrees of freedom
```

```
## Multiple R-squared:  0.6244, Adjusted R-squared:  0.6056
## F-statistic: 33.25 on 2 and 40 DF,  p-value: 3.126e-09
```

In summary we have $\hat{\beta}_1 = -0.1086$ and $\hat{\beta}_2 = -0.1933$, also $\hat{\beta}_0 = 19.9408$ which shows the regression $y = -0.1086x_1 - 0.1933x_2 + 19.9408$ Also we observed coefficient of determination with value of 0.606 which indicates that model explains 60.6% of the predictors of the response data, that is far away larger than the 32.9% that we observed in part(b)

```
avPlots(lm.both)
```



From the add-variable plots, we observed the influence from the original model(which contains only one predictors) to new model(contains two predictors).Specifically, its the additional contribution of x1 on y once the contribution of x2 has been already accounted for.

d)

```
m1=lm(data1$moralIntegration~data1$heterogeneity)
summary(m1)
```

```
##
## Call:
## lm(formula = data1$moralIntegration ~ data1$heterogeneity)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6780 -2.6099  0.2493  2.2971  6.6931
##
## Coefficients:
```

```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         14.42355    0.82507  17.482  < 2e-16 ***
## data1$heterogeneity -0.10275    0.02212  -4.645 3.49e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.926 on 41 degrees of freedom
## Multiple R-squared:  0.3448, Adjusted R-squared:  0.3288
## F-statistic: 21.58 on 1 and 41 DF,  p-value: 3.486e-05
```

By observing the summary of heterogeneity to moralIntegration, we conclude that for one additional in heterogeneity , moralIntegration reduce by 0.1028 units. $\hat{\beta}_1 = -0.1028$ $se(\hat{\beta}_1|x) = 0.0221$ $t = \frac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_1|x)} = -2.8$ p-value=3.5e-05

```
confint(m1,level = 0.97)
```

```
##                           1.5 %      98.5 %
## (Intercept)         12.5685788 16.27852516
## data1$heterogeneity -0.1524859 -0.05301849
```

Hypothesis test: H0: $\beta_1 = 0$ H1:$\beta_1 \neq 0$ The 97% confidence interval for the intercept is [12.569,16.279] which does not include 0 and so the intercept is significantly different from 0. The 97% confidence interval for the slope is[-0.152,-0.053] which also does not include 0 and so the slope is significantly different from 0, so that we reject Ho.

e)

```
Angell=data1
set.seed(100)
n = dim(Angell)[1]
Angell$social = with(Angell, heterogeneity+mobility+rnorm(n,0,.1))
mod1 = lm(moralIntegration ~ heterogeneity + mobility + social, data = Angell)
summary(mod1)
```

```
##
## Call:
## lm(formula = moralIntegration ~ heterogeneity + mobility + social,
##     data = Angell)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2357 -1.1764 -0.2883  1.7623  4.3731
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     20.030      1.200  16.685   <2e-16 ***
## heterogeneity   -4.077      4.527  -0.900    0.373
## mobility        -4.165      4.531  -0.919    0.364
## social           3.968      4.527   0.876    0.386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.25 on 39 degrees of freedom
## Multiple R-squared:  0.6316, Adjusted R-squared:  0.6033
```

```
## F-statistic: 22.29 on 3 and 39 DF,  p-value: 1.427e-08
```

```
confint(mod1,level = 0.97)
```

```
##                    1.5 %    98.5 %
## (Intercept)    17.326232 22.734091
## heterogeneity -14.273969  6.120805
## mobility      -14.371261  6.041701
## social         -6.228576 14.163707
```

Hypothesis test: H0: $\beta_1 = 0$ H1:$\beta_1 \neq 0$ From the result, we observe that the 97% confidence interval for the intercept is [17.15,22.72] which does not include 0 and so the intercept is significantly different from 0 Also,the 97% confidence interval for the slope is [-7.11, 6.76] which includs 0 so we cannot conclude that it is different from 0. Compare the results with part d, we see a big difference.

```
Angell1=Angell[,c(2,3,5)]
crossprod(scale(Angell1,center = TRUE, scale = FALSE))
```

```
##               heterogeneity  mobility    social
## heterogeneity    17498.4065 -525.6928 16974.211
## mobility          -525.6928 4024.4898  3502.694
## social           16974.2106 3502.6938 20482.549
```

From the covariance matrix, we observed super large value within heterogeneity~social (16977) come with high value of 17498(heterogeneity~heterogeneity) and 20452(social~social). This large value means the two predictors (heterogeneity and social) are highly correlated. Thus, after adding the predictor "social" to the new model(with social), it makes a huge influence. That is the reason why we observed such contradictory results.