Homework 4

Emily Lu

5/7/2019

1. This problem uses the data set UN11 from the alr4 package.

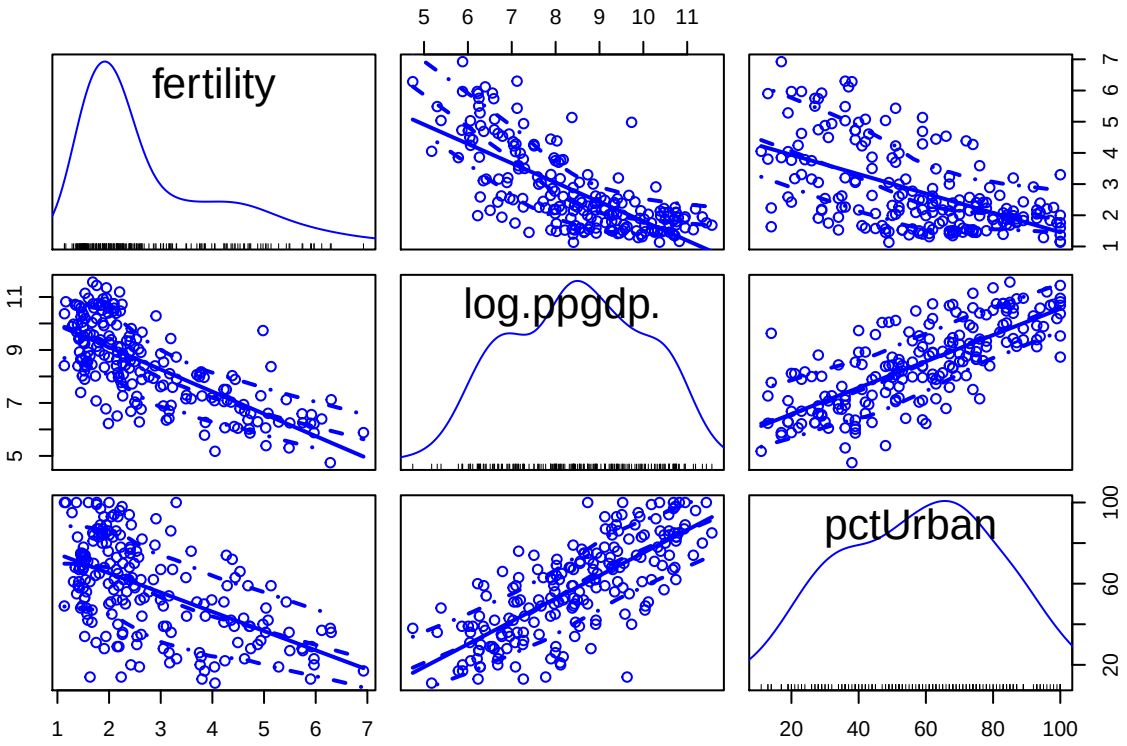
```
library(alr4)
```

```
## Loading required package: car
## Loading required package: carData
## Loading required package: effects
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
attach(UN11)
```

a) Examine the scatterplot matrix for (fertility, log(ppgdp), pctUrban), and comment on the marginal relationships.

```
scatterplotMatrix(~fertility + log(ppgdp) + pctUrban)
```



From the scatterplot matrix above, we can see the following: a negative correlation between fertility and pctUrban, negative correlation between fertility and log(ppgdp), and positive correlation between log(ppgdp) and pctUrban.

b) Fit the two simple regressions fertility $\sim$ log(ppgdp) and fertility $\sim$ pctUrban, and verify that the slopes are significantly different from zero at any conventional level of significance.

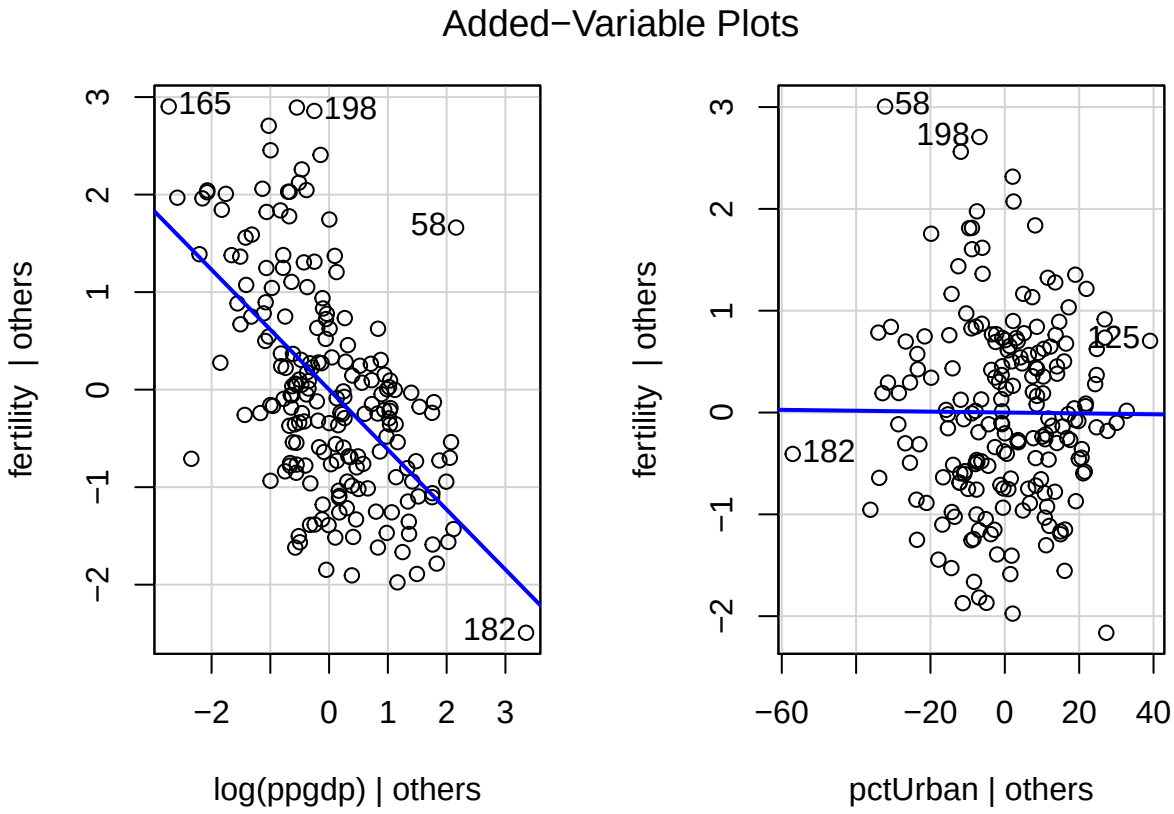
```r
fit1 <- lm(fertility ~ log(ppgdp) + pctUrban)
summary(fit1)
```

```
##
## Call:
## lm(formula = fertility ~ log(ppgdp) + pctUrban)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.15114 -0.64929 -0.06604  0.63253  2.99102
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.9932699  0.3993367  20.016   <2e-16 ***
## log(ppgdp)  -0.6151425  0.0641565  -9.588   <2e-16 ***
## pctUrban    -0.0004393  0.0042656  -0.103    0.918
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9328 on 196 degrees of freedom
## Multiple R-squared:   0.52,  Adjusted R-squared:  0.5151
## F-statistic: 106.2 on 2 and 196 DF,  p-value: < 2.2e-16
```

From the summary above, we can see that the slopes are significantly different from zero at any conventional level of significance.

c) Obtain the added-variable plots for both predictors. Based on the added-variable plots, does log(ppgdp) seem to be useful after adjusting for pcturban, and similarly, does pcturban seem to be useful after adjusting for log(ppgdp)?

```r
avPlots(fit1)
```



The AV plot for log(ppgdp) after pctUrban shows that log(ppgdp) is still correlated with fertility even after accounting for the effects of pcturban. However, that of pctUrban after log(ppgdp) shows that it is not useful when log(ppgdp) is already in the model.

2. Consider a multiple linear regression model with two continuous predictors:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,\ \epsilon \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

a) Suppose that x_{i1} and x_{i2} are exactly related in that $x_{i1} = 2.2x_{i2}$ for all i. For example, x_{i2} could be weight in kilograms and x_{i1} weight in pounds for the i-th individual. Describe the appearance of the added-variable plot for x_{i2} after adjusting for x_{i1}.

The added-variable plot for x_{i2} after adjusting for x_{i1} would have a vertical slope.

b) Suppose instead that x_{i1} and x_{i2} are not perfectly correlated, but that $Y_i = 3x_{i1}$, i.e. Y_i is perfectly correlated with x_{i1}. Describe the appearance of the added-variable plot for x_{i2}.

Given that x_{i1} and x_{i2} are not perfectly correlated, the added-variable plot for x_{i2} would not have a significant slope.