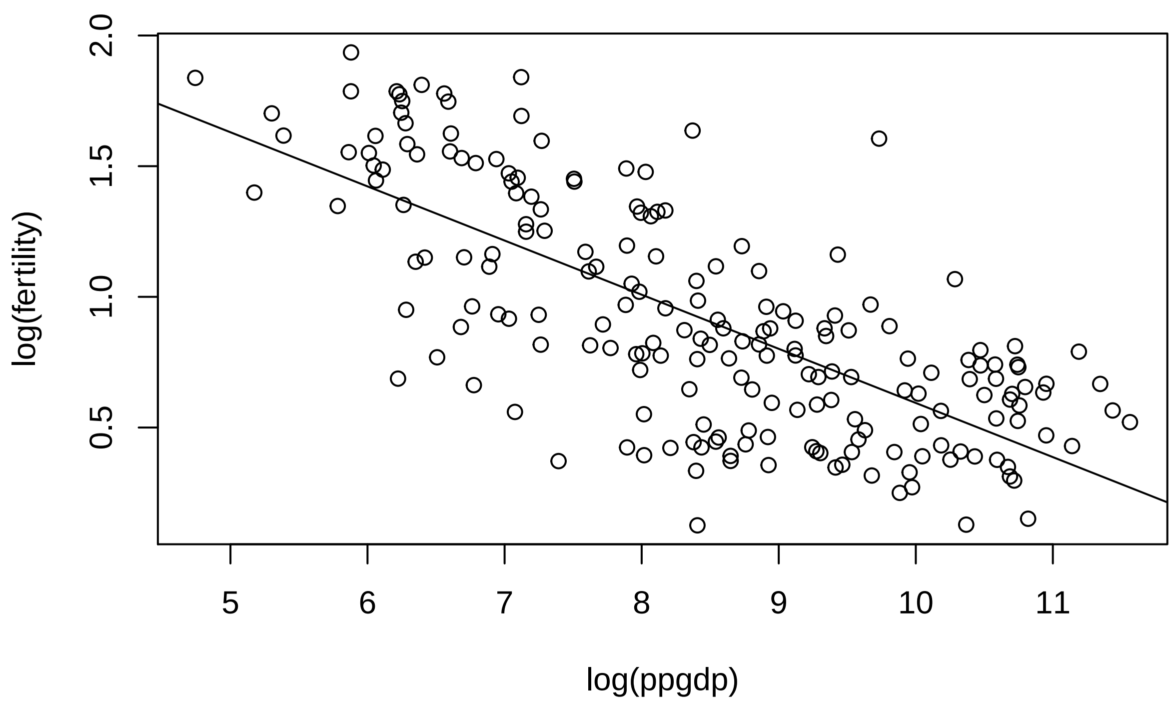
PSTAT 126 Project 3

```r
library(alr4)
```

1a) Drawing scatter plot and line of reggression

```r
attach(UN11)
x = log(UN11$ppgdp)
y = log(UN11$fertility)
fit = lm(log(fertility) ~ log(ppgdp))
plot(log(fertility) ~ log(ppgdp))
abline(fit$coef[1],fit$coef[2],color = 'red')
```



##1b) Coefficient

```r
summary(fit)
```

```
##
## Call:
## lm(formula = log(fertility) ~ log(ppgdp))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.79828 -0.21639  0.02669  0.23424  0.95596
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.66551    0.12057   22.11   <2e-16 ***
## log(ppgdp)  -0.20715    0.01401  -14.79   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3071 on 197 degrees of freedom
## Multiple R-squared:  0.526,  Adjusted R-squared:  0.5236
## F-statistic: 218.6 on 1 and 197 DF,  p-value: < 2.2e-16
```

##1c) Predicting the Log(x = 43140.9)

```
fit = lm(y ~ x)
data1 = data.frame(x = log(43140.9))
predict(fit,newdata = data1 ,interval = "predict", level = .95)
```

```
##         fit        lwr      upr
## 1 0.4547578 -0.1554686 1.064984
```

##1d) Finding the Confidence Interval

```
lower_bound = exp(-0.1554686)
lower_bound
```

```
## [1] 0.856014
```

```
upper_bound = exp(1.064984)
upper_bound
```

```
## [1] 2.900793
```

##2a) Calculating the Anova

```
library(faraway)
attach(prostate)
x = prostate$lcavol
y = prostate$lpsa
fit1 = lm(y ~ x)
anov = anova(fit1)
anov
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x          1 69.003  69.003  111.27 < 2.2e-16 ***
## Residuals 95 58.915   0.620
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##2b) Explanation of the Anova

In the ANOVA test the S_{xx} is the quantity that is being represented. The S_{xx} in this specific problem is the value that determines the variability of lpsa. Therefore, the value of the sse is not being represented.

2c) F-testing

$H_0: B_1 = 0 $
$H_1 : B_1 \neq 0$

```r
f_test = var.test(x,y)
f_test
```

```
##
##  F test to compare two variances
##
## data:  x and y
## F = 1.0425, num df = 96, denom df = 96, p-value = 0.8387
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6971165 1.5591181
## sample estimates:
## ratio of variances
##           1.042539
```

Our p-value is .8387 which is greater than our $\alpha = .05$. Therefore we have enough significant evidence to conclude that B1=0. Therefore we fail to reject the null hypothesis.

2d) T-testing

```r
t.test(x,y)
```

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = -6.7364, df = 191.92, p-value = 1.84e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.4587654 -0.7979895
## sample estimates:
## mean of x mean of y
##  1.350010  2.478387
```

3a) using the lm() function

```r
library(faraway)
attach(fat)
x = fat$brozek
y = fat$age
fit2 = lm( y ~ x)
fit2
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##     35.9807       0.4702
```

```r
summary(fit2)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.547  -9.154  -1.204   8.701  35.099
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 35.98072    2.01376  17.867  < 2e-16 ***
## x            0.47016    0.09844   4.776 3.04e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.09 on 250 degrees of freedom
## Multiple R-squared:  0.08362,    Adjusted R-squared:  0.07996
## F-statistic: 22.81 on 1 and 250 DF,  p-value: 3.045e-06
```

3b Stating the variation

```r
summary(fit2)
```

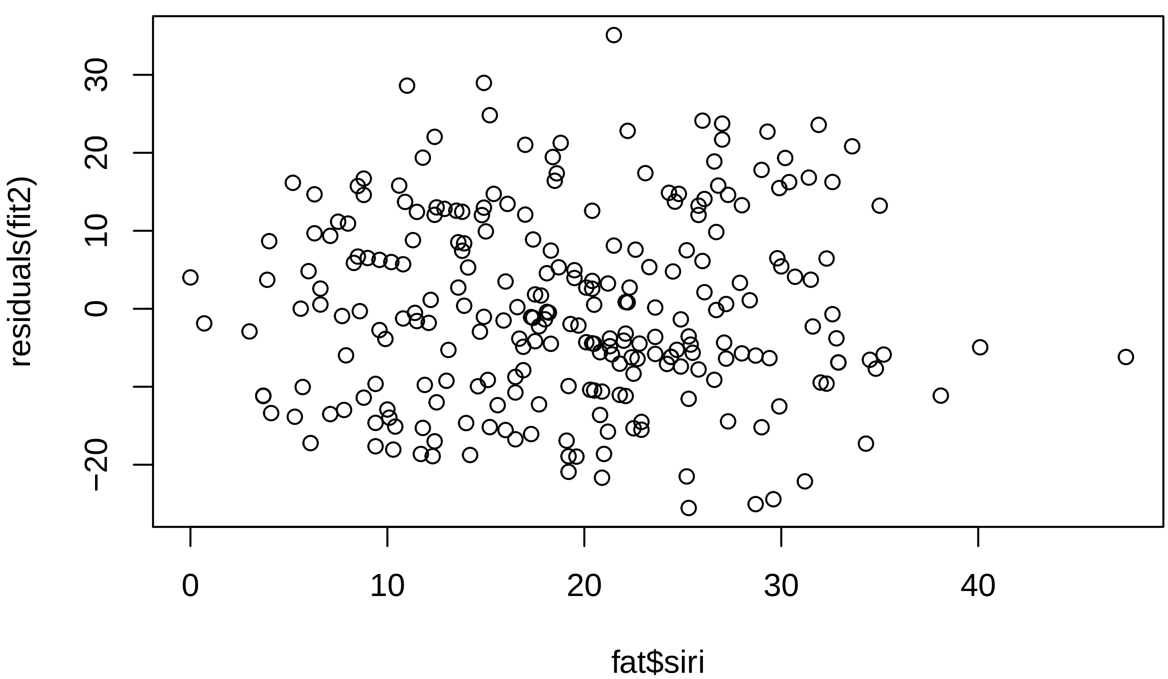
```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.547  -9.154  -1.204   8.701  35.099
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 35.98072    2.01376  17.867  < 2e-16 ***
## x            0.47016    0.09844   4.776 3.04e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.09 on 250 degrees of freedom
## Multiple R-squared:  0.08362,    Adjusted R-squared:  0.07996
## F-statistic: 22.81 on 1 and 250 DF,  p-value: 3.045e-06
```

The multiple R is equal to .08362 which is the variance of our model.

3c) ploting the residuals

```r
plot(fat$siri, residuals(fit2))
```



Siri is contributing to our model because its explaining more variation than our original model.

3d) using the lm() function with 3 datasets

```r
lm(fat$brozek~fat$age+fat$siri)
```

```
##
## Call:
## lm(formula = fat$brozek ~ fat$age + fat$siri)
##
## Coefficients:
## (Intercept)      fat$age     fat$siri
##    1.260382    -0.001486     0.926583
```