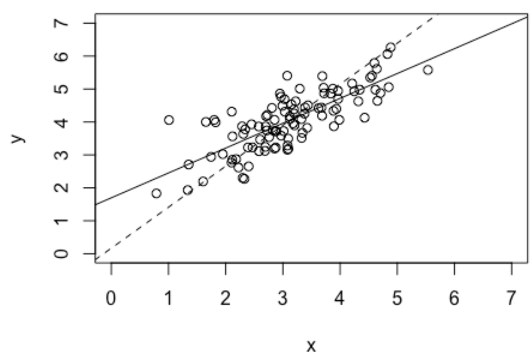
HW#3, Nan Deng

(1)

```r
library(MASS)
mu <- c(3,4)
sigma <- matrix(c(1.0,0.8,0.8,1.0),nrow=2)
set.seed(123)
datam <- data.frame(mvrnorm(100,mu,sigma))
colnames(datam) <- c("x","y")
```

(a)

```r
library(faraway)
fit_line1 <- lm(y ~ x,data=datam)
plot(datam$x,datam$y,xlab="x",ylab="y",xlim=c(0,7),ylim=c(0,7))
fit_line2 <- lm(x ~ y,data=datam)
plot(datam$x,datam$y,xlab="x",ylab="y",xlim=c(0,7),ylim=c(0,7))
abline(fit_line1)
newy <- c(-10,10)
newx <- fit_line2$coefficients[1]+fit_line2$coefficients[2]*newy
lines(newx,newy,lty=2)
```



(b)

```r
summary(fit_line1)

##
## Call:
## lm(formula = y ~ x, data = datam)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.18120 -0.44633 -0.01159  0.36843  1.60020
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.69698    0.19998   8.486 2.31e-13 ***
## x            0.75479    0.06144  12.285  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5701 on 98 degrees of freedom
```

```
## Multiple R-squared:  0.6063, Adjusted R-squared:  0.6023
## F-statistic: 150.9 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
summary(fit_line2)
```

```
##
## Call:
## lm(formula = x ~ y, data = datam)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.11649 -0.35759  0.01514  0.40905  1.24599
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.13480    0.27138  -0.497     0.62
## y            0.80325    0.06539  12.285   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5881 on 98 degrees of freedom
## Multiple R-squared:  0.6063, Adjusted R-squared:  0.6023
## F-statistic: 150.9 on 1 and 98 DF,  p-value: < 2.2e-16
```

The R^2, p-value and F-statistics remain the same in these two models.

(c)

(d)

(e)

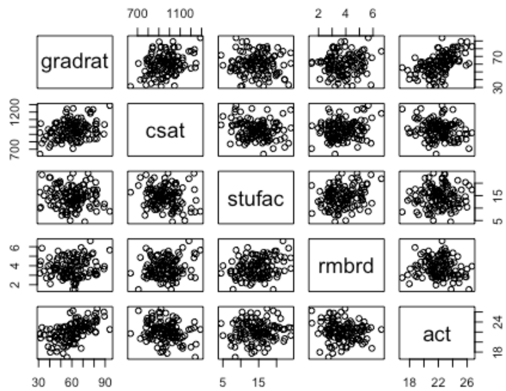
```r
for (i in 1:5) {
    yi_hat <- fit_line1$fitted.values[i]
    xi <- datam[i,1]
    left <- (yi_hat-mean(datam[,2]))/sd(datam[,2])
    right <- cor(x=datam)[1,2]*(xi-mean(datam[,1]))/sd(datam[,1])
    print(cat(paste("i=",i),paste("Left=",left),paste("Right=",right)))
}
```

```
## i= 1 Left= -0.356395169413281 Right= -0.356395169413283NULL
## i= 2 Left= -0.350162752477095 Right= -0.350162752477095NULL
## i= 3 Left= 1.19980471582391 Right= 1.19980471582391NULL
## i= 4 Left= 0.0476069981594529 Right= 0.0476069981594529NULL
## i= 5 Left= 0.253666333569284 Right= 0.253666333569285NULL
```

(2)

(a)

```
college <- read.csv("/Users/CandiceDeng\ 1/Desktop/STATS500/HW#3/college.csv")
plot(cbind(college[1:2],college[4:6]))
```



(b)

```
line_overall <- lm(college$gradrat ~ college$csat+college$private+college$stufac+college$
rmbrd+college$act, data=college)
line_b <- lm(college$gradrat ~ college$csat+college$act, data=college)
anova(line_overall,line_b)

## Analysis of Variance Table
##
## Model 1: college$gradrat ~ college$csat + college$private + college$stufac +
##     college$rmbrd + college$act
## Model 2: college$gradrat ~ college$csat + college$act
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    114 10895
## 2    117 11217 -3   -322.56 1.1251  0.342
```

The p-value 0.342 is greater than 0.05, so the hypothesis cannot be rejected.

(c)

```
line_c <- lm(college$gradrat ~ offset(0.05*college$csat)+college$private+college$stufac+c
ollege$rmbrd+college$act, data=college)
anova(line_overall,line_c)

## Analysis of Variance Table
##
## Model 1: college$gradrat ~ college$csat + college$private + college$stufac +
##     college$rmbrd + college$act
## Model 2: college$gradrat ~ offset(0.05 * college$csat) + college$private +
##     college$stufac + college$rmbrd + college$act
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    114 10895
## 2    115 10955 -1   -60.668 0.6348 0.4272
```

The p-value 0.427 is greater than 0.05, so the hypothesis cannot be rejected.

(d)

```
line_d <- lm(college$gradrat ~ college$csat+college$private+college$stufac, data=college)
anova(line_overall, line_d)

## Analysis of Variance Table
##
## Model 1: college$gradrat ~ college$csat + college$private + college$stufac +
##     college$rmbrd + college$act
## Model 2: college$gradrat ~ college$csat + college$private + college$stufac
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1    114 10895
## 2    116 19412 -2   -8517.9 44.565 5.018e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value 5.018e-15 is less than 0.05, so the hypothesis should be rejected.

(e)

```
line_e <- lm(college$gradrat ~ I(college$act+college$rmbrd)+college$private+college$stufa
c+college$csat, data=college)
anova(line_overall, line_e)

## Analysis of Variance Table
##
## Model 1: college$gradrat ~ college$csat + college$private + college$stufac +
##     college$rmbrd + college$act
## Model 2: college$gradrat ~ I(college$act + college$rmbrd) + college$private +
##     college$stufac + college$csat
##   Res.Df   RSS Df Sum of Sq      F   Pr(>F)
## 1    114 10895
## 2    115 11642 -1   -747.49 7.8216 0.006061 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value around 0.006 is less than 0.05, so the hypothesis should be rejected.

(3)

```
x_linear <- c(1,0,-1)
y_linear <- c(1,0,2)
fit <- lm(y_linear ~ x_linear+I(3*(x_linear*x_linear)-2))
fit$coefficients

##                        (Intercept)                          x_linear
##                                1.0                              -0.5
## I(3 * (x_linear * x_linear) - 2)
##                                0.5

fit2 <- lm(y_linear ~ x_linear)
fit2$coefficients

## (Intercept)    x_linear
##         1.0        -0.5
```

If $\beta_2=0$， β_0 is 1.0 and β_1 is -0.5, which remains unchanged.