Homework 3

Jing Leng

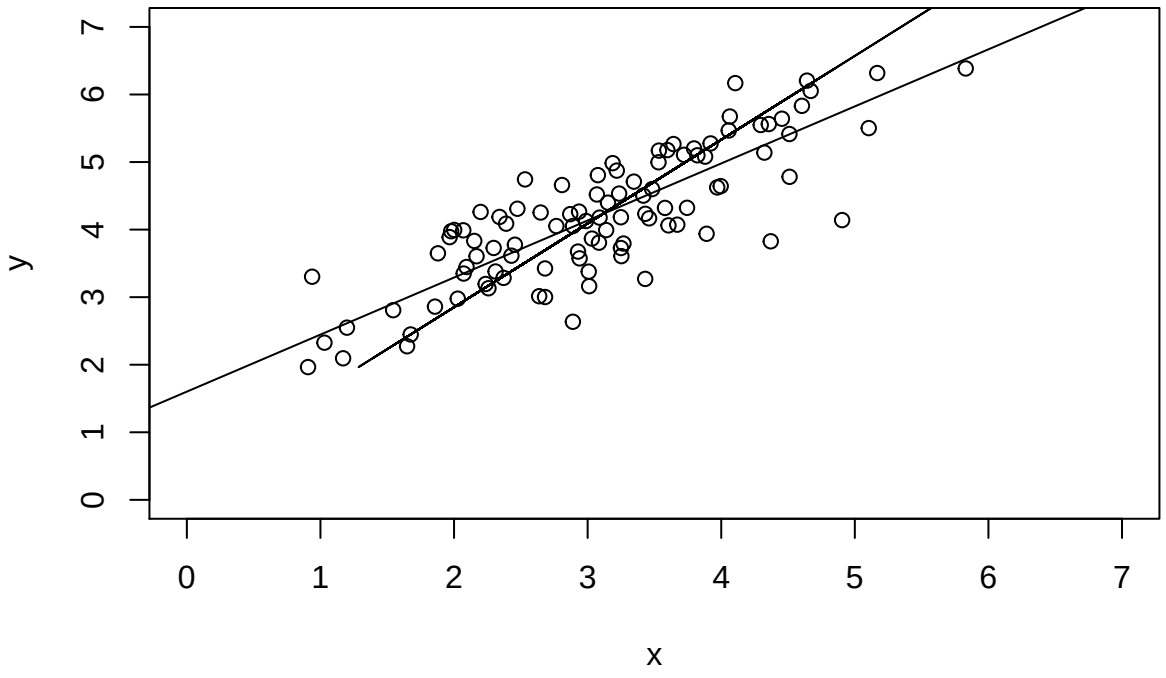
September 25, 2014

1

a)

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

plot(y~x, datam, xlim=c(0,7),ylim=c(0,7))
lm <- lm(y~x, datam)
abline(lm)
lm2 <- lm(x~y, datam)
lines(lm2$fitted.values, datam$y)
```



b)

```
summary(lm)
```

```
##
## Call:
```

```
## lm(formula = y ~ x, data = datam)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6028 -0.3819  0.0135  0.4142  1.7251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.6018     0.1909    8.39  3.7e-13 ***
## x             0.8444     0.0584   14.46  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.586 on 98 degrees of freedom
## Multiple R-squared:  0.681,  Adjusted R-squared:  0.678
## F-statistic:  209 on 1 and 98 DF,  p-value: <2e-16
```

```
summary(lm2)
```

```
##
## Call:
## lm(formula = x ~ y, data = datam)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4265 -0.3802 -0.0474  0.3604  1.8648
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2994     0.2427   -1.23     0.22
## y             0.8065     0.0558   14.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.573 on 98 degrees of freedom
## Multiple R-squared:  0.681,  Adjusted R-squared:  0.678
## F-statistic:  209 on 1 and 98 DF,  p-value: <2e-16
```

The R^2, t-statistcs, F-statistcs and their respective p-values are the same.

c)

$$T_{y|x} = \frac{\sqrt{S_{xx}}(\hat{\beta} - \beta)}{\hat{\sigma}} = \frac{\sqrt{S_{xx}}(\frac{S_{xy}}{S_{xx}} - \beta)}{\sqrt{S_{yy} - \frac{S_{xy}^2}{X_{xx}}}} = \frac{S_{xy} - S_{xx}\beta}{\sqrt{S_{yy}S_{xx} - S_{xy}^2}} = \frac{S_{xy}}{\sqrt{S_{yy}S_{xx} - S_{xy}^2}}$$

For $T_{y|x}$, x and y are interchangeable, thus $T_{x|y} = T_{y|x}$.

d).

e)

```
left <- (lm$fitted.values - mean(datam$y))/(sd(datam$y))
r = sqrt(1-(sum(lm$residuals^2))/(sum((datam$y - mean(datam$y))^2)))
right <- r*(datam$x - mean(datam$x))/(sd(datam$x))
sum(abs(left - right))
```

```
## [1] 5.196e-14
```

First 5 elements for left: -0.0283, 1.6291, 1.2738, 0.4341, 0.7019.
First 5 elements for right: -0.0283, 1.6291, 1.2738, 0.4341, 0.7019.

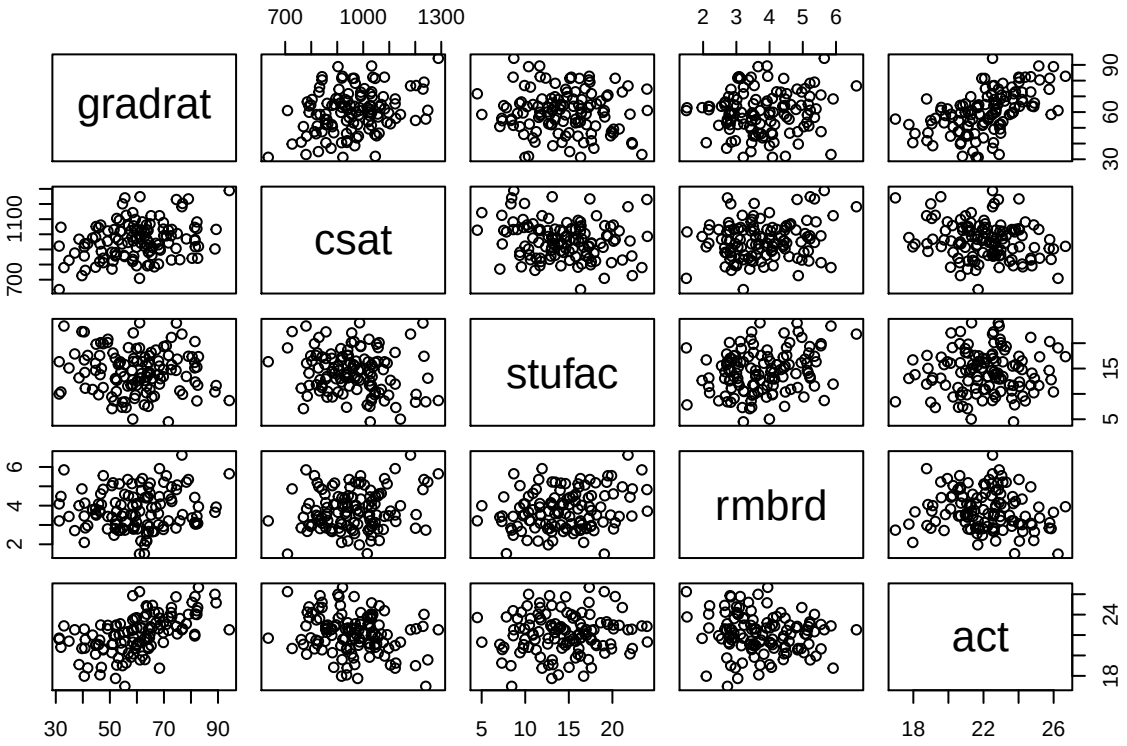
2

a)

```
college <- read.csv('college.csv')
head(college)
```

```
##   gradrat   csat private stufac rmbrd   act lenroll
## 1   59.16 1031.4       1  9.102 2.934 23.61   4.615
## 2   88.89  903.0       1 10.384 3.655 25.99   7.753
## 3   70.88  989.7       0 19.487 5.545 23.04   6.156
## 4   58.67  955.7       0 22.028 4.438 20.15   6.334
## 5   68.32  904.2       1 15.484 4.467 23.66   5.704
## 6   54.67 1201.0       1  8.357 2.832 19.00   5.870
```

```
pairs(college[c(1,2,4,5,6)])
```



b)

```
lm2 <- lm(gradrat ~ . - lenroll - private - stufac - rmbrd, data = college)
anova(lm, lm2)
```

```
## Warning: models with response '"gradrat"' removed because response differs
## from model 1

## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value Pr(>F)
## x          1   71.9    71.9     209 <2e-16 ***
## Residuals 98   33.7     0.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F-statistics is 1.1251, p-value is 0.342. We fail to reject the null hypothesis. ###c)

```
summary(lm)$coefficients
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   1.6018    0.19086   8.392 3.664e-13
## x             0.8444    0.05838  14.464 4.699e-26
```

```r
coe <- summary(lm)$coefficients
tstats <- (coe[2, 1] - 0.05)/coe[2,2]
tstats
```

```
## [1] 13.61
```

```r
pv <- 2*pt((coe[2, 1] - 0.05)/coe[2,2], 120 - 6)
pv
```

```
## [1] 2
```

t-statistics is 13.6078, p-value is 2. We fail to reject the null hypothesis.

d)

```r
lm3 <- lm(gradrat ~ . - lenroll - private - act, data = college)
anova(lm, lm3)
```

```
## Warning: models with response '"gradrat"' removed because response differs
## from model 1
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value Pr(>F)
## x          1   71.9    71.9     209 <2e-16 ***
## Residuals 98   33.7     0.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-statistics is 45.289, p-value is close to 0. We reject the null hypothesis.

e)

```r
lm4 <- lm(gradrat ~ I(rmbrd + act) + csat + private + stufac, college)
anova(lm, lm4)
```

```
## Warning: models with response '"gradrat"' removed because response differs
## from model 1
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value Pr(>F)
## x          1   71.9    71.9     209 <2e-16 ***
## Residuals 98   33.7     0.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-statistics is 7.8216, p-value is 0.0061. We reject the null hypothesis.

3

```
x <- c(1, 0, -1)
y <- c(1, 0, 2)
lm <- lm(y ~ x + I(3*x^2 -2))
lm2 <- lm(y ~ x)
summary(lm)
```

```
##
## Call:
## lm(formula = y ~ x + I(3 * x^2 - 2))
##
## Residuals:
## ALL 3 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.0         NA      NA       NA
## x                  -0.5         NA      NA       NA
## I(3 * x^2 - 2)      0.5         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:     1,   Adjusted R-squared:   NaN
## F-statistic:  NaN on 2 and 0 DF,  p-value: NA
```

```
summary(lm2)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##    1    2    3
##  0.5 -1.0  0.5
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.000      0.707    1.41     0.39
## x             -0.500      0.866   -0.58     0.67
##
## Residual standard error: 1.22 on 1 degrees of freedom
## Multiple R-squared:  0.25,   Adjusted R-squared:  -0.5
## F-statistic: 0.333 on 1 and 1 DF,  p-value: 0.667
```

β_0 and β_1 remain unchanged no matter whether $\beta_2 = 0$.