414HW2

Ningyuan Wang

9/18/2020

1.

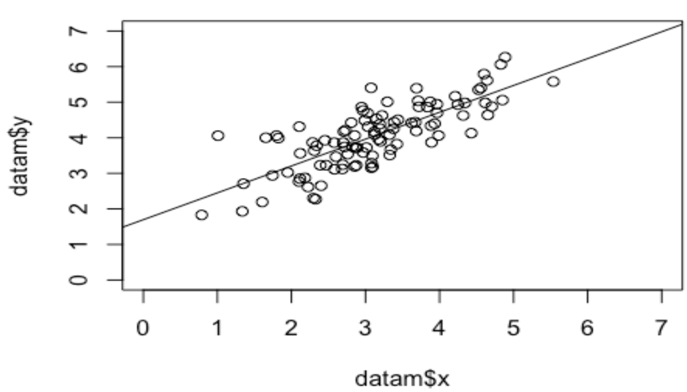
```
# generate observations
library(MASS)
mu <- c(3,4)
sigma <- matrix(c(1.0,0.8,0.8,1.0),nrow=2)
set.seed(123)
datam <- data.frame(mvrnorm(100,mu,sigma))
colnames(datam) <- c("x","y")
```

a. The scatter plot is in part b. To save the space, we only show the code here.

```
plot(datam$x, datam$y, xlim = c(0,7), ylim = c(0,7))
```

b. The values for alpha_hat is 1.6969772, beta_hat is 0.7547914, sigma_hat is 0.5701, and r_squared is 0.6063.

```
lm = lm(y~x, data = datam)
lm$coefficients

## (Intercept)           x
##   1.6969772   0.7547914

summary(lm)$sigma

## [1] 0.5701041

summary(lm)$r.squared

## [1] 0.606287

plot(datam$x, datam$y, xlim = c(0,7), ylim = c(0,7))+ abline(lm)
```



c. The value of F-statistic is 150.91, and p-value is < 2.2e-16. Because p-value < alpha(=0.05), reject H_0.

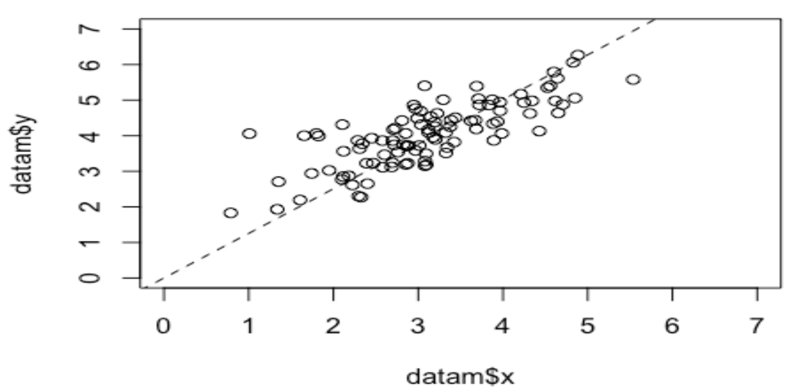
```
# reduced model
lm0 = lm(y~1, datam)
# compare two models
anova(lm0, lm)

## Analysis of Variance Table
##
## Model 1: y ~ 1
## Model 2: y ~ x
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     99 80.901
## 2     98 31.852  1    49.049 150.91 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d. The value of F-statistic is 5.5854, and p-value is 0.02008. Because p-value < alpha(=0.05), reject H_0.

```
lm1 = lm(y~offset(0.9*x), datam)
anova(lm1, lm)

## Analysis of Variance Table
##
## Model 1: y ~ offset(0.9 * x)
## Model 2: y ~ x
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     99 33.667
## 2     98 31.852  1    1.8154 5.5854 0.02008 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.

```
m0 <- lm(y ~ x -1 , data=datam)
summary(m0)

plot(datam$x, datam$y, xlim = c(0,7), ylim = c(0,7))+ abline(m0, lty=2)
```



b. No. The sum of residuals for the model without an intercept is 13.79104.

```r
sum(residuals(m0))

## [1] 13.79104
```

c. Hand-writttten part was attached below.

```r
sum(datam$x * residuals(m0))

## [1] -4.010681e-15

sum(fitted.values(m0) * residuals(m0))

## [1] -4.458239e-15

# cor(datam) # correlation coeeficeint r
# anova(lm) # anova table for the model
```

2c. To find out $\sum_{i=1}^{n} x_i e_i$ and $\sum_{i=1}^{n} \hat{y}_i e_i$,

we start from the least square criterion.

Model: $y_i = \beta X_i + \varepsilon_i$

$$\min_{\beta} \epsilon'\epsilon = \sum_{i=1}^{n} (y_i - \beta X_i)^2$$

$\Downarrow$ normal equation

$$\frac{\partial \epsilon'\epsilon}{\partial \beta} = -2\sum_{i=1}^{n} (y_i - \beta x_i) x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \beta X_i) X_i = 0$$

where $\sum_{i=1}^{n} (y_i - \beta X_i) X_i = \sum_{i=1}^{n} x_i e_i$

so, $\boxed{\sum_{i=1}^{n} x_i e_i = 0}$

Then, $\boxed{\sum_{i=1}^{n} \hat{y}_i e_i = \sum_{i=1}^{n} \hat{\beta} x_i e_i = 0}$ since $\sum_{i=1}^{n} x_i e_i = 0$

3a. we already have: $\hat{\beta} = \dfrac{\sum (X_i - \bar{X}) y_i}{\sum (X_i - \bar{X})^2} = \dfrac{S_{xy}}{S_{xx}}$

so, $r = \dfrac{S_{xy}}{(S_{xx} S_{yy})^{\frac{1}{2}}} = \dfrac{\hat{\beta} (S_{xx}^{\frac{1}{2}})^2}{(\cancel{S_{xx}} S_{yy})^{\frac{1}{2}}} = \boxed{\hat{\beta} \sqrt{\dfrac{S_{xx}}{S_{yy}}}}$

r and $\hat{\beta}$ have the same sign.

3b. we already know: $\hat{y}_i = \bar{y} + \beta (X_i - \bar{X})$

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{S_{yy}} = \frac{\sum (\hat{\beta} (x_i - \bar{x}))^2}{S_{yy}} = \boxed{\frac{\hat{\beta}^2 S_{xx}}{S_{yy}}}$$

In 3a, $r^2 = \hat{\beta}^2 \dfrac{S_{xx}}{S_{yy}} = R^2$