414HW3

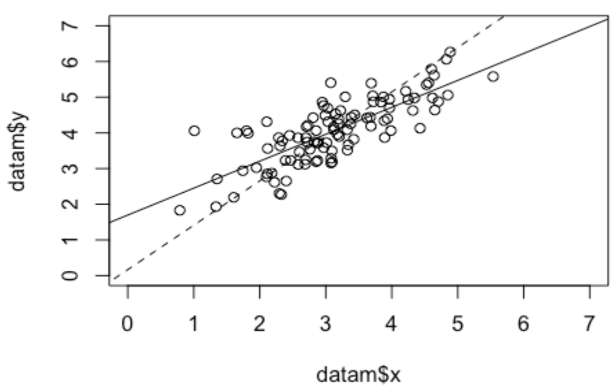
Ningyuan Wang

9/25/2020

1.

a. In the following plot, the solid line represents the linear model with y as the response and x as the predictor. The dashed line represents the linear model reversing the roles of x and y.

```
lm = lm(y~x, data = datam)
lm_rev = lm(x~y, data = datam)
# summary(lm_rev)
newy = seq(-1, 8, length = 50)
newx = coef(lm_rev)[1] + coef(lm_rev)[2]*newy
plot(datam$x, datam$y, xlim = c(0,7), ylim = c(0,7))+ abline(lm)

lines(newx, newy, lty = 2)
```



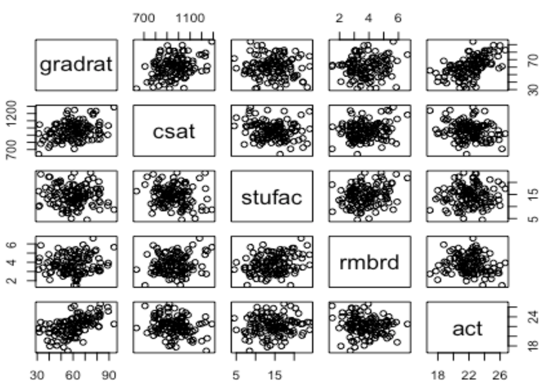
b. The R_squared, T-stats associated with the predicotors and their respective p-values, F-stats and p-values from the two models are the same.

```
# r_squared
summary(lm)$r.squared

## [1] 0.606287

summary(lm_rev)$r.squared

## [1] 0.606287

# t-stats
summary(lm)$coefficients

##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 1.6969772 0.19998330  8.485595 2.309177e-13
## x           0.7547914 0.06144191 12.284635 1.495934e-21

summary(lm_rev)$coefficients
```

```
##                 Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -0.1348023 0.27138014 -0.4967285 6.204930e-01
## y            0.8032510 0.06538664 12.2846353 1.495934e-21

# F-stats
summary(lm)$fstatistic

##    value    numdf    dendf
## 150.9123   1.0000  98.0000

summary(lm_rev)$fstatistic

##    value    numdf    dendf
## 150.9123   1.0000  98.0000

# summary(lm)
# summary(lm_rev)
```

c. Please check the hand-written part below.

2.




```
lm_dat = lm(gradrat ~ csat + private + stufac + rmbrd + act, data = dat)
# summary(lm_dat)
```

b. F-stat is 1.1251 and p-value is 0.342. Since p-value >0.05, do not reject H0.

```
lm_t1 = lm(gradrat ~ csat + act, data = dat)
anova(lm_t1, lm_dat)

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

c. F-stat is 0.6348 and p-value is 0.4272 Since p-value >0.05, do not reject H0.

```
lm_t2 = lm(gradrat ~ offset(0.05*csat) + private + stufac  + rmbrd + act, data = dat)
anova(lm_t2, lm_dat)
```

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

d. F-stat is 44.565 and p-value is close to 0. Since p-value <0.05, reject H0.

```
lm_t3 = lm(gradrat ~ csat + private + stufac, data = dat)
anova(lm_t3, lm_dat)

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

e. F-stat is 7.8216 and p-value is 0.006. Since p-value <0.05, reject H0.

```
lm_t4 = lm(gradrat ~ csat + private + stufac  + I(rmbrd + act), data = dat)
anova(lm_t4, lm_dat)

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

3.

beta0_hat is 1, beta1_hat is -0.5, and beta2_hat is 0.5. The estimates for beta0_hat and beta1_hat remain unchanged if beta2 is 0.

```
dat3 = matrix(c(1,0,-1,1,0,2), ncol = 2)
dat3 = as.data.frame(dat3)
colnames(dat3) = c("x", "y")

lm3 = lm(y~x + I(3*x^2 -2),  data = dat3)
lm3$coefficients

##    (Intercept)              x I(3 * x^2 - 2)
##            1.0           -0.5            0.5

lm4 = lm(y~x, data = dat3)
lm4$coefficients

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

1c. under $H_0: \beta = 0$

$T_{y|x} = \dfrac{\hat{\beta}}{se(\hat{\beta})}$, we already know $\hat{\beta} = \dfrac{S_{xy}}{S_{xx}}$

& $\hat{Var}(\hat{\beta}) = \dfrac{\hat{\sigma}^2}{S_{xx}}$

since $\hat{\sigma}^2 = \dfrac{1}{n-2} SSE = \dfrac{SST - SSR}{n-2} = \dfrac{S_{yy} - \hat{\beta}^2 S_{xx}}{n-2}$

$= \dfrac{S_{xx}S_{yy} - S_{xy}^2}{(n-2) S_{xx}}$

so, $\hat{Var}(\hat{\beta}) = \dfrac{S_{xx}S_{yy} - S_{xy}^2}{(n-2) S_{xx}^2}$

$T_{y|x} = \dfrac{S_{xy}}{\cancel{S_{xx}}} \dfrac{\sqrt{(n-2)}\ \cancel{S_{xx}}}{\sqrt{S_{xx}S_{yy} - S_{xy}^2}} = \dfrac{S_{xy}\sqrt{n-2}}{\sqrt{S_{xx}S_{yy} - S_{xy}^2}}$

If we reverse the role between x and y, then

$S_{xx} \to S_{yy}$ and $S_{yy} \to S_{xx}$; $S_{xy} = S_{yx}$ unchanged.

$T_{x|y} = \dfrac{S_{xy}\sqrt{n-2}}{\sqrt{S_{yy}S_{xx} - S_{xy}^2}} = T_{y|x}$

4a. we already know: $R^2 = \dfrac{SSR}{SST} \Rightarrow SSR = R^2 \cdot SST$

$SSE = SST - SSR$

$= (1-R^2) SST$

$F = \dfrac{(SST - SSE)/(p-1)}{SSE/(n-p)} = \dfrac{R^2 \cdot \cancel{SST}}{(p-1)} \cdot \dfrac{(n-p)}{(1-R^2)\cancel{SST}}$

$= \dfrac{n-p}{p-1} \cdot \dfrac{R^2}{1-R^2}$

note: $\dfrac{n-p}{p-1}$ is always >1

As $R^2 \to 0$, $F \to 0^+$

As $R^2 \to 1$, $F \to \infty$

4b. $cov(\hat{y}, e) = cov(Py, (I_n - P)y)$

$= P\,cov(y, y)(I_n - P)$

$= \sigma^2 (I_n - P)P$ since P is idempotent

$= 0$