HW#4, Nan Deng

(1)

```r
college <- read.csv("/Users/CandiceDeng\ 1/Desktop/STATS500/HW#3/college.csv")
line_overall <- lm(college$gradrat ~ college$csat+college$private+college$stufac+college$
rmbrd+college$act, data=college)
library(car)
```

(a)

```r
matrix_a <- cbind(rbind(0,0,0),rbind(0,0,0),diag(3),rbind(0,0,0))
lht(line_overall,matrix_a)
```

```
## Linear hypothesis test
##
## Hypothesis:
## college$private = 0
## college$stufac = 0
## college$rmbrd = 0
##
## Model 1: restricted model
## Model 2: college$gradrat ~ college$csat + college$private + college$stufac +
##     college$rmbrd + college$act
##
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    117 11217
## 2    114 10895  3    322.56 1.1251  0.342
```

The p-value is 0.342 which is greater than 0.05, the hypothesis should not be rejected.

(b)

```r
matrix_b <- cbind(0,0,0,0,1,-1)
lht(line_overall,matrix_b)
```

```
## Linear hypothesis test
##
## Hypothesis:
## college$rmbrd - college$act = 0
##
## Model 1: restricted model
## Model 2: college$gradrat ~ college$csat + college$private + college$stufac +
##     college$rmbrd + college$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
```

The p-value is around 0.006061 which is less than 0.05, the hypothesis should be rejected.

(2)

(a)

```r
library(faraway)
```

```
##
## Attaching package: 'faraway'

## The following objects are masked from 'package:car':
##
##     logit, vif

data(prostate)
attach(prostate)
line_2a <- lm(lpsa~lcavol+lweight+age+lbph+svi+lcp+gleason+pgg45)
summary(line_2a)

##
## Call:
## lm(formula = lpsa ~ lcavol + lweight + age + lbph + svi + lcp +
##     gleason + pgg45)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7331 -0.3713 -0.0170  0.4141  1.6381
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.669337   1.296387   0.516  0.60693
## lcavol       0.587022   0.087920   6.677 2.11e-09 ***
## lweight      0.454467   0.170012   2.673  0.00896 **
## age         -0.019637   0.011173  -1.758  0.08229 .
## lbph         0.107054   0.058449   1.832  0.07040 .
## svi          0.766157   0.244309   3.136  0.00233 **
## lcp         -0.105474   0.091013  -1.159  0.24964
## gleason      0.045142   0.157465   0.287  0.77503
## pgg45        0.004525   0.004421   1.024  0.30886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7084 on 88 degrees of freedom
## Multiple R-squared:  0.6548, Adjusted R-squared:  0.6234
## F-statistic: 20.86 on 8 and 88 DF,  p-value: < 2.2e-16

qt(0.975,88)

## [1] 1.98729

qt(0.95,88)

## [1] 1.662354

confint(line_2a,"age",level=0.95)

##           2.5 %      97.5 %
## age -0.04184062 0.002566267

confint(line_2a,"age",level=0.90)

##             5 %          95 %
## age -0.0382102 -0.001064151
```

95% CI for βage = −0.019637 ± 1.98729 × 0.011173 = (−0.041841, 0.002567) 90% CI for βage = −0.019637 ± 1.662354 × 0.011173 = (−0.038210, −0.001064) For Hypothesis Analysis, 95% CI for βage does pass the 0 point while 90% CI for βage does not, H0 should not be rejected when α=0.05 and should be rejected when α=0.10. If using p-value which is equal to 0.08229, if 0.05<p-value<0.10, the result will be the same.

(b)

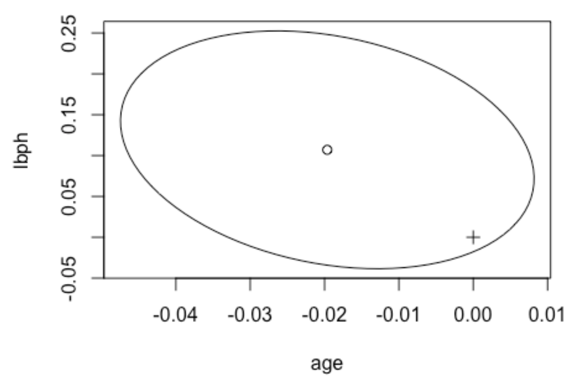
```
library(ellipse)

##
## Attaching package: 'ellipse'

## The following object is masked from 'package:car':
##
##     ellipse

## The following object is masked from 'package:graphics':
##
##     pairs

plot(ellipse(line_2a, c(4,5),level=0.95),type="l")
points(line_2a$coef[4],line_2a$coef[5])
points(0, 0, pch=3)
```



Considering that dot 0 is inside the confidence region, the null hypothesis failed to be rejected.

(c)

```
x <- data.frame(lcavol=1.44692,lweight=3.62301,age=65,lbph=0.3001,svi=0,lcp=-0.79851,
 gleason=7,pgg45=15)
predict(line_2a,x)

##        1
## 2.389053

predict(line_2a,x,level=0.95,interval="confidence")

##        fit      lwr      upr
## 1 2.389053 2.172437 2.605669

predict(line_2a,x,level=0.95,interval="predict")
```

```
##          fit       lwr      upr
## 1 2.389053 0.9646584 3.813447
```

(3)

(a)

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{TSS-RSS}{TSS}$$

$$= 1 - \left(1 + F \cdot \frac{p-1}{n-p}\right)^{-1}$$

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

When R2→0, F is equal to 0; when R2→1, F is infinity.

(b)

$\hat{y} = Py$, $e = (I-P)y$

$$Cov(\hat{y}, e) = cov(Py, (I-P)y)$$

$$= P \cdot cov(y, y)(I-P)^T$$

$$= P \cdot cov(X\beta+\varepsilon, X\beta+\varepsilon)(I-P)^T$$

$$= P \cdot cov(\varepsilon, \varepsilon)(I-P)^T$$

$$= P \cdot Var(\varepsilon)(I-P)^T$$

$$= P \cdot \sigma^2 I (I-P)^T$$

$$= \sigma^2 P(I-P) = \sigma^2 (P-P^2)$$

Since $P = P^2$, $Cov(\hat{y}, e) = \sigma^2 \cdot 0 = 0$