STAT4355HW6

Code

This is an R script with the purpose of running multiple linear regression on football data

(a)

Hide

```
#load the data
ftbl <- read.csv(file = 'football.csv')
#build linear model
lm1 <- lm(y~x2+x7+x8, data = ftbl)
#model summary
summary(lm1)
```

```
Call:
lm(formula = y ~ x2 + x7 + x8, data = ftbl)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0370 -0.7129 -0.2043  1.1101  3.7049 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.808372   7.900859  -0.229 0.820899    
x2           0.003598   0.000695   5.177 2.66e-05 ***
x7           0.193960   0.088233   2.198 0.037815 *  
x8          -0.004816   0.001277  -3.771 0.000938 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.706 on 24 degrees of freedom
Multiple R-squared:  0.7863,	Adjusted R-squared:  0.7596 
F-statistic: 29.44 on 3 and 24 DF,  p-value: 3.273e-08
```

Hide

```
sigma(lm1)^2
```

```
[1] 2.91125
```

Fitted linear model:

y-hat = -1.808372 + 0.003598x2 + 0.193960x7 - 0.004816x8

i.

(σ^)2 = 2.91125

ii.

R^2 = 0.7863

iii.

adjusted R^2 = 0.7596

(b)

Hide

```
#test whether β2 = -β8
library(multcomp)
D1 <- matrix(c(0,1,0,1),1,4)
D1
```

```
     [,1] [,2] [,3] [,4]
[1,]    0    1    0    1
```

Hide

```
d1 <- 0
qf(0.05, 1, 24, lower.tail=FALSE)
```

```
[1] 4.259677
```

Hide

```
mytest1 <- glht(lm1, linfct=D1, rhs=d1)
summary(mytest1, test=Ftest())
```

```


     General Linear Hypotheses

Linear Hypotheses:
        Estimate
1 == 0 -0.001217

Global Test:
```

F <dbl>	DF1 <int>	DF2 <int>	Pr(>F) <dbl>
0.575409	1	24	0.4555029

1 row

i.

H0: β2 + β8 != 0

H1: β2 + β8 = 0

ii.

D = [0,1,0,1] d = 0

iii.

β-hat ~ N(β, σ2((X’X)^−1))

Dβ-hat ~ N(Dβ, σ2D((X'X)^−1)D')

iv.

F0 = ((Dβ-hat-d)'D((X'X)^-1)D' (Dβ-hat-d)/r)/(SSE/(n-p))

v.

F(1, 24)

vi.

F = 0.575409

p = 0.4555029

As p-value 0.456 is more than 0.05 = α, we accept H0 and conclude that passing yardage (β2) and the regression parameter for the opponents' yards rushing (β8) are not the same in magnitude, opposite in direction.

(c)

Hide

```
#test whether β2 = 0, β8 = 0, β7 = 0.2
library(multcomp)
D2 <- matrix(c(0,0,0,0,0,0,0,0,0,0,1,0),3,4)
D2
```

```
     [,1] [,2] [,3] [,4]
[1,]    0    0    0    0
[2,]    0    0    0    1
[3,]    0    0    0    0
```

Hide

```
d2 <- c(0,0,0.2)
qf(0.05, 1, 24, lower.tail=FALSE)
```

```
[1] 4.259677
```

Hide

```
mytest2 <- glht(lm1, linfct=D2, rhs=d2)
summary(mytest2, test=Ftest())
```

```
diag(.) had 0 or NA entries; non-finite result is doubtful
```

```

	 General Linear Hypotheses

Linear Hypotheses:
          Estimate
1 == 0    0.000000
2 == 0   -0.004815
3 == 0.2  0.000000

Global Test:
```

	F	DF1	DF2	Pr(>F)
	<dbl>	<int>	<int>	<dbl>
	14.22072	1	24	0.0009377699

1 row

i.

H0: $\beta2 = 0$, $\beta8 = 0$, $\beta7 = 0.2$

H1: $\beta2 != 0$, $\beta8 != 0$, $\beta7 != 0.2$

ii.

D = [(0,0,0,0),(0,0,0,0),(0,0,1,0)] (each parenthesis is a row)

d = (0, 0, 0.2)

iii.

β-hat ~ N(β, σ2((X'X)^−1))

Dβ-hat ~ N(Dβ, σ2D((X'X)^−1)D')

iv.

F0 = ((Dβ-hat-d)'D((X'X)^-1)D' (Dβ-hat-d)/r)/(SSE/(n-p))

v.

F = 14.22072

p = 0.0009377699

As p-value 0.001 is less than 0.05 = α, we reject H0 and conclude that passing yardage and the regression parameter for the opponents' yards rushing impact the number of games won and a unit increase in the team's rushing playes percent doesn't increase the number of games by 0.2.

(d)

Hide

```
#99 % confidence interval on the four individual coefficients
confint(lm1,level=0.99)
```

```
                    0.5 %       99.5 %
(Intercept) -23.906597837 20.289853719
x2            0.001654201  0.005541939
x7           -0.052823409  0.440743828
x8           -0.008387097 -0.001243891
```

(e)

Hide

```
#99 % confidence intervals on the mean number of games
#i.
newx1 <- data.frame(x2=2300,x7=56,x8=2100)
predict(lm1, newx1, interval='confidence', level=0.99)
```

```
       fit      lwr      upr
1 7.216424 6.159089 8.273758
```

Hide

```
#ii .
newx2 <- data.frame(x2=2900,x7=61,x8=1900)
predict(lm1, newx2, interval='confidence', level=0.99)
```

```
       fit      lwr      upr
1 11.30817 9.447357 13.16897
```

(f)

Hide

```
#99 % confidence intervals on a future observation
#i.
newx3 <- data.frame(x2=2300,x7=56,x8=2100)
predict(lm1, newx3, interval='prediction', level=0.99)
```

```
       fit     lwr     upr
1 7.216424 2.32845 12.1044
```

Hide

```
#ii .
newx4 <- data.frame(x2=2900,x7=61,x8=1900)
predict(lm1, newx4, interval='prediction', level=0.99)
```

```
       fit      lwr      upr
1 11.30817 6.185965 16.43037
```

(f)

The lengths of the PIs from (f) are are greater than the corresponding CIs from (e). Prediction intervals must account for both the uncertainty in estimating the population mean, plus the random variation of the individual values; as it takes into account the true error, the prediction interval is wider.

(h)

R code is within this notebook pdf.