Homework7

Samuel Mweni

10/27/2022

Table of Contents

No table of contents entries found.

#Question 4
a a. Fit regression model (8.58) using the number of cases shipped (X d and the binary variable(X3) as predictors.

load data

```
rdata<-read.table("../homework7/hello.txt",header = F)
colnames(rdata)<-c("Y","X1","X2","X3")

tail(rdata,3)
```

```
##       Y     X1   X2 X3
## 50 4499 290455 7.99  0
## 51 4186 411750 7.83  0
## 52 4342 292087 7.77  0
```

Fit a regression model

```
rdata.reg=lm(Y~X1+I(X1^2)+X3+X1:X3,data = rdata)
rdata.reg
```

```
##
## Call:
## lm(formula = Y ~ X1 + I(X1^2) + X3 + X1:X3, data = rdata)
##
## Coefficients:
## (Intercept)           X1      I(X1^2)           X3        X1:X3
##   3.925e+03    1.570e-03   -1.154e-09    6.505e+02   -8.870e-05
```

```
summary(rdata.reg)
```

```
##
## Call:
## lm(formula = Y ~ X1 + I(X1^2) + X3 + X1:X3, data = rdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -288.253 -102.112   -7.251   72.363  294.646
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.925e+03  6.262e+02   6.268 1.06e-07 ***
## X1           1.570e-03  3.755e-03   0.418   0.6778
## I(X1^2)     -1.154e-09  5.481e-09  -0.211   0.8341
## X3           6.505e+02  2.801e+02   2.322   0.0246 *
## X1:X3       -8.870e-05  8.760e-04  -0.101   0.9198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 145.2 on 47 degrees of freedom
## Multiple R-squared:  0.6865, Adjusted R-squared:  0.6599
## F-statistic: 25.74 on 4 and 47 DF,  p-value: 2.476e-11
```

our regression model is yhat=3925 -0.00152X1-1.154e-09 I(X1^2)+(650.2X3)-(8.870e-05 X1:X3)

#b Test whether or not the interaction terms and the quadratic term can be dropped from the model; use a = .05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Hypothesis Ho:B2 =B4 =0 Ha : Atleast one inequality

The full Regression model Yi =Bo +B1(X1)-B2($x^2$1) +B3(X3)-B4(X1X3)+e

The reduced model Yi =Bo +B1(X1)+B3(X3)+e

```
rdata.reg<- lm(Y~X1+X3,data =rdata)
rdata.reg2<-lm(Y~X1+I(X1^2)+X3+X1:X3,data = rdata)
anova(rdata.reg,rdata.reg2)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X3
## Model 2: Y ~ X1 + I(X1^2) + X3 + X1:X3
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     49 992204
## 2     47 991173  2    1031.2 0.0244 0.9759
```

use formula based calculation

SSE(F)=SSE(X1,I(X1^2),X3,X1:X3)=991173 SSE(R)=SSE(X1,X2)=992204 The F - partial =(SSE(R)-SSE(F)/(dfe(R)-dfe(f)))/(SSE(F)/dfe(F)) = ((1031.2)/(49-47))/(991173/47) =(1031.2)/(2)/(991173/7) =0.0244901 p-value calculation

```
p_value<-1-pf(0.0244901,2,47)
p_value
```

```
## [1] 0.9758198
```

p-value =P(F(df1=dfe(R)-dfe(F),df2=dfe(F))) >Fpartial) =P(F(df1=1,df2=57)>0.02444901)=0.9758598.

Decision rule Reject Ho if p-value is <alpha

Statistical conclusion since p-value (0.9758598)is greater than alpha(0.025),we do not reject Ho.Therefore,we do not have signficance evidence to support that $x1^2$,X1X3 needed in the model so $x1^2$,X1X3 can be droped from the model when X1 and X2 are in the model.

#C Why would we wish to include number of cases (X1) in the regression when ourinterest is in estimating the effects of holiday on labor hours? This is because at this point we do not have evidence that the number of cases is insignificant in this model such that it can be removed and not affecting the model.Hence to ensure better results we include the number of cases.