Homework7

Samuel Mweni

10/27/2022

Table of Contents

load data

```
mmdata<-read.table("../homework7/CH01PR27.txt",header = F)
colnames(mmdata)<-c("Y","X")
head(mmdata,3)

##     Y  X
## 1 106 43
## 2 106 41
## 3  97 47

# tail check
tail(mmdata,2)

##     Y  X
## 59 70 72
## 60 74 76
```

a) Fit regression model (8.2). Plot the fitted regression function and the data. Does the quadratic

#regression function appear to be a good fit here? Find R2.

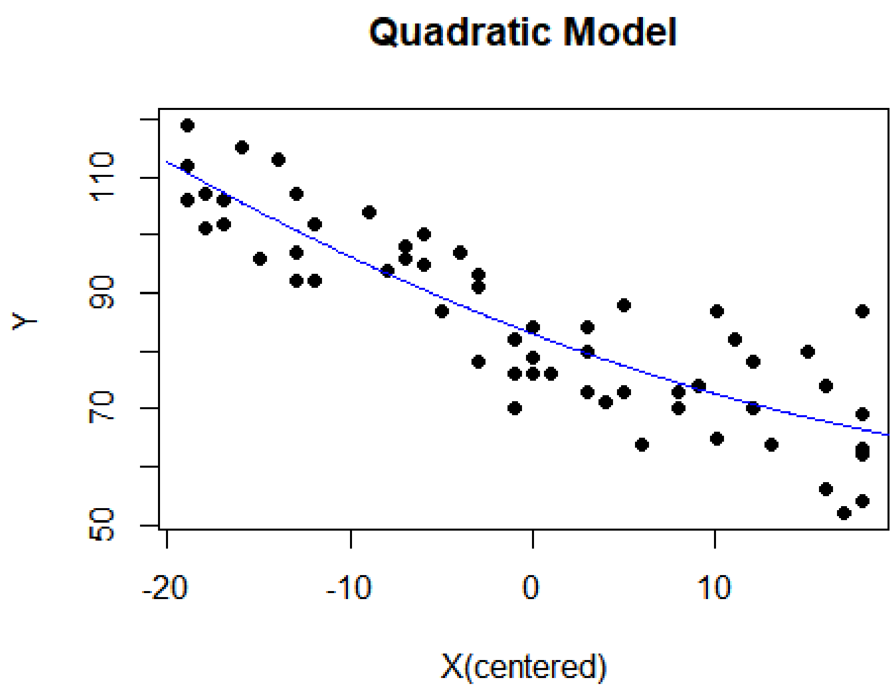
```
#center the variables
t<-(mmdata$X-mean(mmdata$X))
t_power<- t^2
```

```r
mmdata<-cbind(mmdata,t,t_power)

mmdata.reg=lm(Y~t + t_power,data = mmdata)
mmdata.reg

##
## Call:
## lm(formula = Y ~ t + t_power, data = mmdata)
##
## Coefficients:
## (Intercept)            t      t_power
##    82.93575     -1.18396      0.01484
```

Fit the regression model 8.2 Y_hat = 82.93575-1.18396x +0.01484

plot the fitted regression function and data

```r
plot(mmdata$t,mmdata$Y,main="Quadratic
Model",xlab="X(centered)",ylab="Y",pch=19)
x=seq(-20,20,by =.1)
y=82.93575-1.18396*x+0.01484*x^2
lines(x,y,col="blue")
```



Does the quadratic appears to be a good fit?

Yes it is a good fit

Find R^2 R^2= 0.7632

summary

```
summary(mmdata.reg)

##
## Call:
## lm(formula = Y ~ t + t_power, data = mmdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.086  -6.154  -1.088   6.220  20.578
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 82.935749   1.543146  53.745   <2e-16 ***
## t           -1.183958   0.088633 -13.358   <2e-16 ***
## t_power      0.014840   0.008357   1.776   0.0811 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.026 on 57 degrees of freedom
## Multiple R-squared:  0.7632, Adjusted R-squared:  0.7549
## F-statistic: 91.84 on 2 and 57 DF,  p-value: < 2.2e-16
```

Test whether or not there is a regression relation; use (){ = .05. State the alternatives, decisionrule, and conclusion.

Hypothesis H0 : β1 = β11,= 0,Ha :not both β1 and β11 = 0. Decision rule, and conclusion: - we reject Ho If p_value is < alpha

conclusion , we reject the Ho in this case because p_value (2.2e-16) is less than 0.05.there is sufficient evidence to indicate that a regression relation exist between muscle mass and the centered age variable and its squared value .

#-Estimate the mean muscle mass for women aged 48 years; use a 95 percent confidenceinterval. Interpret your interval.

```
mmdata.reg2<- lm(Y~X,data = mmdata)
predict(mmdata.reg2,list(X=48),se.fit =T,interval ="confidence",level=0.95)

## $fit
##        fit      lwr      upr
## 1 99.22678 96.20318 102.2504
##
## $se.fit
## [1] 1.510501
##
## $df
## [1] 58
```

```
##
## $residual.scale
## [1] 8.173177
```

From the out above , the std.err. for Y_hat(mean) is 1.510501.The 95% CI for Ymean is :Yhat+-t(df=45-2,1-alpha/2) x se(yhat_mean).Putting the values to the formula the 90% CI is:

```
tcrit<-qt(0.975,df=57)
c(99.22678-tcrit*1.510501,99.22678+tcrit*1.510501)

## [1]  96.20205 102.25151
```

At confidence level of 95%, the estimate mean musle mass for women aged 48 years is

96.20205 ≤ E(Yh)≤ 102.25151

#d Predict the muscle mass for a woman whose age is 48 years; use a 95 percent prediction interval. Interpret your interval.

```
mmdata.reg2<- lm(Y~X,data = mmdata)
predict(mmdata.reg2,list(X=48),se.fit =T,interval ="prediction",level=0.95)

## $fit
##        fit      lwr      upr
## 1 99.22678 82.58934 115.8642
##
## $se.fit
## [1] 1.510501
##
## $df
## [1] 58
##
## $residual.scale
## [1] 8.173177
```

From the output above Yhat_mean=99.22678 and the std.err for yhat_mean is 1.510501. to compute the standard error of the individual prediction

se(yhat_mean)=√(se(yhat_mean))^2 +MSE)

```
se_EY_new <-(sqrt((1.510501)^2+(8.173177)^2))
se_EY_new

## [1] 8.311584
```

so, by inserting the values to the formula at 95% PI is;

```
tcrit<-qt(0.975,df=57)
c(99.22678-tcrit*8.311584,99.22678+tcrit*8.311584)

## [1]  82.58312 115.87044
```

At confidence level of 95%, the predic mass for a woman aged 48 years is

82.58312 ≤ E(Yh)≤ 115.87044

e) Test whether the quadratic term can be dropped from the regression model; use (){ = .05.State the alternatives, decision rule, and conclusion.

Full regression model: Y_hat = 82.93575-1.18396x +0.01484x^2

Hypothesis

H0 : β1 =0 Ha :β11≠ 0

Reduced model

yhat =82.93575-1.18396x +0*x^2

yhat =82.93575-1.18396x

Partial F-test

```
mmdata.full.reg=lm(Y~t+t_power,data =mmdata)
mmdata.reduced.reg<- lm(Y~t,data = mmdata)
anova(mmdata.reduced.reg,mmdata.full.reg)

## Analysis of Variance Table
##
## Model 1: Y ~ t
## Model 2: Y ~ t + t_power
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     58 3874.4
## 2     57 3671.3  1    203.13 3.1538 0.08109 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We obtain our SSE from the anova table output

THE F-PARTIAL =(SSE(R)-SSE(F)/(dfe(R)-dfe(f)))/(SSE(F)/dfe(F))

```
      =(203.13)/(58-57)/3671.3/57

      =3.153763

F_partial<-((203.13)/(1)/(3671.3/57))
F_partial

## [1] 3.153763
```

The P-value calculation

```
p_value<-1-pf(3.153763,1,57)
p_value

## [1] 0.08108997
```

p-value =P(F(df1=dfe(R)-dfe(F),df2=dfe(F))) >Fpartial)
=P(F(df1=1,df2=57)>3.153763)=0.08108997

Decision rule

Reject Ho if p-value < alpha(0.025)

Statistical conclusion

since p-value (0.08108997) is greater than alpha (0.025),we do not reject Ho . Therefore ,we do not have signficant evidence to support that x^2 is needed in the model ,so x^2 can be dropped from the model when x is in the model.