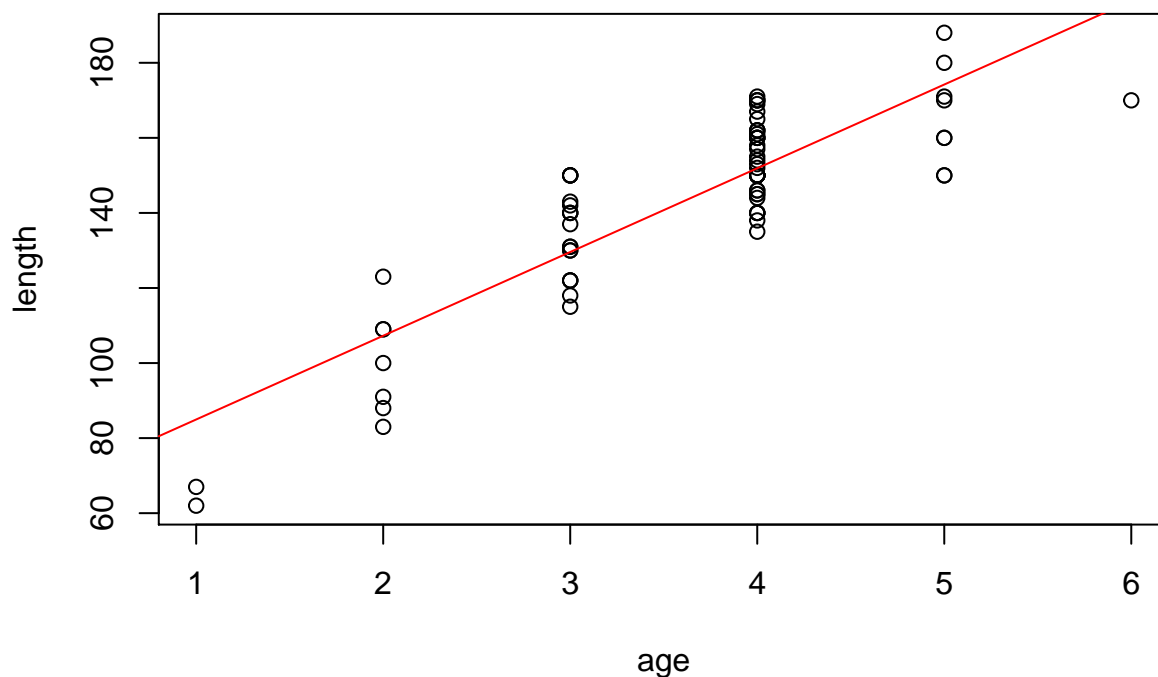


# assignment1

460386740

4/18/2018

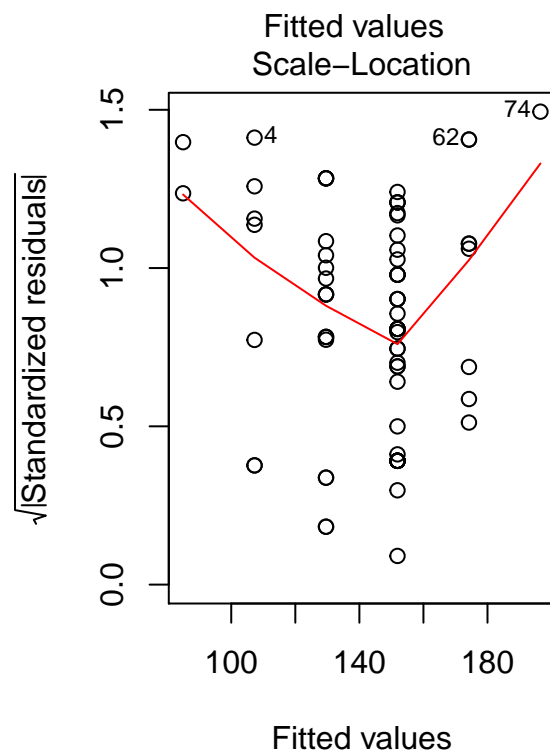
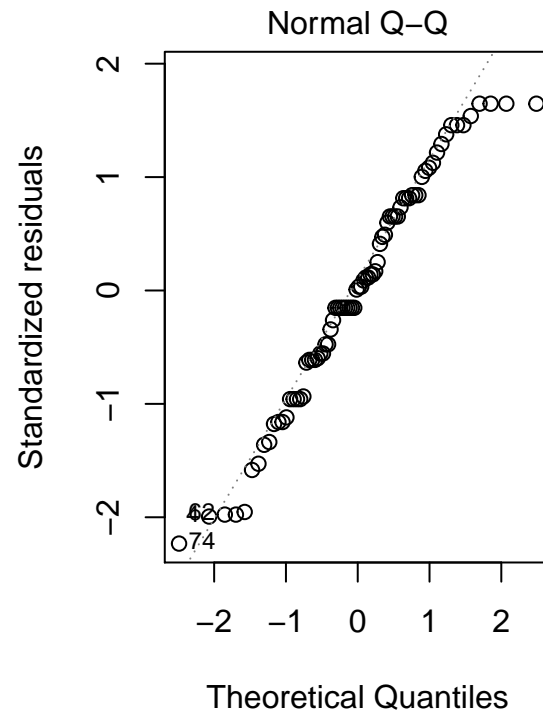
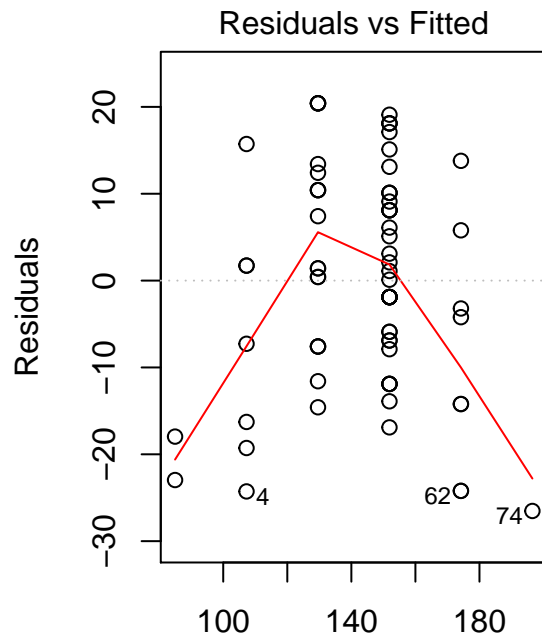
```
blue=read.table("/Users/shirleyma/Desktop/bluegills.txt",header = T)
attach(blue)
#(a)
#the linear model would be  $y = 62.64898 + 22.31229x$ 
M=lm(length~age,data=blue)
plot(length~age,data=blue)
abline(coef(M),col="red")
```



```
coef(M)
```

```
## (Intercept)      age
##    62.64898    22.31229
```

```
#(b)
#for testing the linear relationship, we focus on the first two graphs, the first plot shows the
#residual has a non-linear pattern which is quadratic, this indicates a bad linear model fit.
#the second graph shows most residuals are normally distributed well except the end part.
#so the model assumption may not hold.
#the third graph shows the residuals are not spread equally along the range of predictors.
#so this data source may not be a good fit of linear model.
par(mfrow=c(1,2))
plot(M,which = c(1,2,3))
```



```
#(c)
M2=lm(length~as.factor(age),blue)
anova(M2)
```

```
## Analysis of Variance Table
##
## Response: length
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## as.factor(age) 5 36046 7209.2 58.9 < 2.2e-16 ***
## Residuals 72 8813 122.4
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##(d)
#the fitted model is  $y = 97.77 - 112.31x + 109.89x^2 - 35.73x^3 + 5.16x^4 - 0.28x^5$ 
age.2 = age ^ 2
age.3 = age ^ 3
age.4 = age ^ 4
age.5 = age ^ 5

M.poly5 = lm(length ~ age+age.2+age.3+age.4+age.5, data = blue)
summary(M.poly5)
```

```
##
## Call:
## lm(formula = length ~ age + age.2 + age.3 + age.4 + age.5, data = blue)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-19.8947	-7.8293	-0.6289	7.8716	22.5714

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	97.7771	100.2951	0.975	0.333
age	-112.3131	194.4594	-0.578	0.565
age.2	109.8934	135.0990	0.813	0.419
age.3	-35.7385	43.1896	-0.827	0.411
age.4	5.1614	6.4697	0.798	0.428
age.5	-0.2803	0.3676	-0.763	0.448

```
##
## Residual standard error: 11.06 on 72 degrees of freedom
## Multiple R-squared: 0.8035, Adjusted R-squared: 0.7899
## F-statistic: 58.9 on 5 and 72 DF, p-value: < 2.2e-16
```

```
##(e)
M.poly4=update(M.poly5, ~.-age.5)
summary(M.poly4)
```

```
##
## Call:
## lm(formula = length ~ age + age.2 + age.3 + age.4, data = blue)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-19.345	-8.345	-1.691	7.060	23.197

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	28.7011	42.9179	0.669	0.506
age	27.9439	62.9092	0.444	0.658
age.2	9.7392	31.5255	0.309	0.758

```
## age.3      -3.1834      6.5027  -0.490    0.626
## age.4       0.2413      0.4717   0.512    0.610
##
## Residual standard error: 11.03 on 73 degrees of freedom
## Multiple R-squared:  0.802, Adjusted R-squared:  0.7911
## F-statistic: 73.9 on 4 and 73 DF,  p-value: < 2.2e-16
```

```
#coefficient of age.4 is not significant and p-value is large,so remove it
M.poly3=update(M.poly4,~.-age.4)
summary(M.poly4)
```

```
##
## Call:
## lm(formula = length ~ age + age.2 + age.3 + age.4, data = blue)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.345  -8.345  -1.691   7.060  23.197
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  28.7011     42.9179   0.669   0.506
## age          27.9439     62.9092   0.444   0.658
## age.2         9.7392     31.5255   0.309   0.758
## age.3        -3.1834      6.5027  -0.490   0.626
## age.4         0.2413      0.4717   0.512   0.610
##
## Residual standard error: 11.03 on 73 degrees of freedom
## Multiple R-squared:  0.802, Adjusted R-squared:  0.7911
## F-statistic: 73.9 on 4 and 73 DF,  p-value: < 2.2e-16
```

```
M.poly2=update(M.poly3,~.-age.3)
summary(M.poly2)
```

```
##
## Call:
## lm(formula = length ~ age + age.2, data = blue)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.846  -8.321  -1.137   6.698  22.098
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.622     11.016   1.237   0.22
## age          54.049      6.489   8.330 2.81e-12 ***
## age.2        -4.719      0.944  -4.999 3.67e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.91 on 75 degrees of freedom
## Multiple R-squared:  0.8011, Adjusted R-squared:  0.7958
## F-statistic: 151.1 on 2 and 75 DF,  p-value: < 2.2e-16
```

*#now the model seems good which matches with the quadratic guess at the beginning.*

```
 #(e)  
M.new = lm(length ~ age+age.2, data = blue)  
s1=deviance(M.new)  
s0=deviance(M.poly5)  
n=length(age)  
fobs=((s1-s0)/(6-3))/(s0/(n-3))  
1-pf(fobs,6-3,n-6)
```

```
## [1] 0.8206683
```

```
1-pf(fobs,6-3,n-3)
```

```
## [1] 0.8206726
```