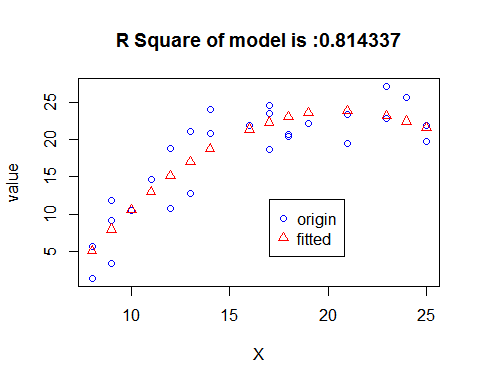
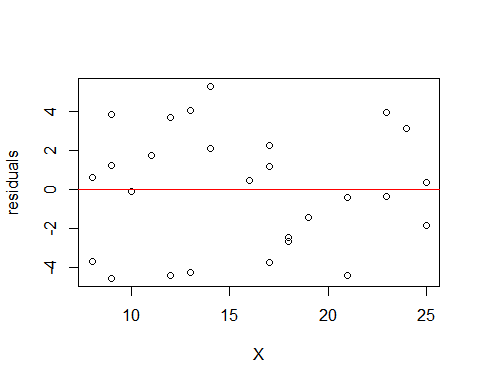
Data Analysis and Linear Models Homework 7

### 8.6

## Y X   
## Min. : 1.40 Min. : 8.00   
## 1st Qu.:12.30 1st Qu.:11.50   
## Median :20.40 Median :16.00   
## Mean :17.64 Mean :15.78   
## 3rd Qu.:22.45 3rd Qu.:20.00   
## Max. :27.10 Max. :25.00

#### a)Fit regression model(8.2)





## Analysis of Variance Table  
##   
## Model 1: Y ~ 1  
## Model 2: Y ~ X + I(X^2)  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 26 1284.81   
## 2 24 238.54 2 1046.3 52.633 1.678e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The plots *residuals* against linear predictor *X* show us that,The residual is evenly distributed around 0.And the F test of good of fitness in the variable table,the *p-value* is smaller than 0.05,which indicates that the full model is accepted,that is this model is work well;and R Square of the *model.1* is 0.8143372.Based on the two result,which indicats that this model fits well,that is,it is a good fit.

#### b)

To test wether or not there is a regression realation,that is to test the coefficient of quadratic interm is 0:

* H0: beta2 is 0
* H1: beta2 is not 0

## [1] "The computed F value is :25.453290"

## [1] "The computed p value is :0.000037"

According to the test above,we know that,the *p-value* is smaller than 0.01,so,we can conclude that,we have 99% confidence to reject the null hypothesis,that is ,the coefficient of quadratic item is not equal to 0.

#### c)

* Confidence intervals(CI)

Use the *Bonferroni Method*,We can get the confidence interval respetively when age=10,15,20,and the confidence intevel at 99 percenpt is (7.6742863,13.4661274),(18.1405546,22.1352916),(21.275656,26.295498)

#### d)

The prediction of model when age=15 is 20.1379231,and the 99% confidence of the prediction intervals is (11.1550164,29.1208298)

#### e)

To test whether the quadratic term can be dropped,that is ,the cofficient of the quadratic term is equal 0,that is

* H0: beta2 is 0
* H1: beta2 is not 0

And we have done this job in answer a,we know tha p-value of the test is 3.707829910^{-5},which is smaller than 0.01,so we conclude that,quadratic term can not be dropped in the model.

#### f)

##   
## Call:  
## lm(formula = Y ~ X + I(X^2), data = Steroid)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.5463 -2.5369 0.3868 2.1973 5.3020   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -26.32541 5.88154 -4.476 0.000157 \*\*\*  
## X 4.87357 0.77515 6.287 1.69e-06 \*\*\*  
## I(X^2) -0.11840 0.02347 -5.045 3.71e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.153 on 24 degrees of freedom  
## Multiple R-squared: 0.8143, Adjusted R-squared: 0.7989   
## F-statistic: 52.63 on 2 and 24 DF, p-value: 1.678e-09

According to the model,the cofficients of *Intercept* is -26.3254125,cofficients of *X* is 4.8735744,cofficients of *X^2* is -0.1184012,

so the fitted funtion is :Y=-26.3254125\*Intercept+4.8735744]\*X+-0.1184012\*X^2

### 8.21

According to what the topic mean,the linear regression equation is

* Y=b0+b1\*X1+b2\*X2+b3\*X3+bias

#### a)

Develop the response function for each type of protection category:

* When X2=1,X3=0,that is when type of protection is *Hard hat*, and Y=b0+b1\*X1+b2
* When X2=0,X3=1,taht is when type of protection is *Bump cap*,and Y=b0+b1\*X1+b3
* When X2=0,X3=0,that is when type of protection is *None*,and Y=b0+b1\*X1

#### b)

For each of following questions,the alternatives H0 and H1 for the appropriate is :

##### (1)

* H0: b2 is equal to 0
* H1: b2 is smaller than 0

##### (2)

* H0: b2 is equal to b3
* H1: b2 is not equal to b3

### 8.23

I do not agree with the interpretaion with result.Of course,accoding to the test in the result show for us,it just only tells *Winter* has no influence on sale except other interaction with other season.It is only true in this test.Hower,in other test,eg B2=0 or B3=0 or B0=0,they are ignored and not to be condidered in the topic.Besides,and interaction are also not be considered I think.So,I think the interpretion is biased.

### 8.34

#### a)

First-order linear regression model is :

* Y=b0+b1\*X1+b2\*X2+b3\*X3+bias

#### b)

The response functions for the three types of banks:

* When X2=1,X3=0,that is when type of bank is *Commercial*,Y=b0+b1\*X1+b2
* When X2=0,X3=1,that is when type of bank is *Mutual saving*,Y=b0+b1\*X1+b3
* When X2=-1,X3=-1,that is when type of bank is *Saving and loan*,Y=b0+b1\*X1-b2-b3

#### c)

##### (1)

* **b2** means if controlling for X1,when type of bank is *Commercial*,the constant term will increase by **b2** units

##### (2)

* **b3** means if controlling for X1,when type of bank is *Mutual saving*,the constant term will increase by **b3** units

##### (3)

* **-b2-b3** means if controlling for X1,when type of bank is *Saving and loan* ,the constant term will increase by **-b2-b3** units

### 8.39

## 'data.frame': 440 obs. of 17 variables:  
## $ Identification.number : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ County : Factor w/ 373 levels "Ada","Adams",..: 179 77 130 284 237 159 190 353 80 82 ...  
## $ State : Factor w/ 48 levels "AL","AR","AZ",..: 4 13 42 4 4 33 3 21 9 42 ...  
## $ Land.area : int 4060 946 1729 4205 790 71 9204 614 1945 880 ...  
## $ Total.population : int 8863164 5105067 2818199 2498016 2410556 2300664 2122101 2111687 1937094 1852810 ...  
## $ Percent.of.population18\_34: num 32.1 29.2 31.3 33.5 32.6 28.3 29.2 27.4 27.1 32.6 ...  
## $ Percent.of.population65 : num 9.7 12.4 7.1 10.9 9.2 12.4 12.5 12.5 13.9 8.2 ...  
## $ Num.active.physicians : int 23677 15153 7553 5905 6062 4861 4320 3823 6274 4718 ...  
## $ Num.hospital.beds : int 27700 21550 12449 6179 6369 8942 6104 9490 8840 6934 ...  
## $ Total.serious.crimes : int 688936 436936 253526 173821 144524 680966 177593 193978 244725 214258 ...  
## $ Percent.high : num 70 73.4 74.9 81.9 81.2 63.7 81.5 70 65 77.1 ...  
## $ Percent.bachelor : num 22.3 22.8 25.4 25.3 27.8 16.6 22.1 13.7 18.8 26.3 ...  
## $ Percent.below : num 11.6 11.1 12.5 8.1 5.2 19.5 8.8 16.9 14.2 10.4 ...  
## $ Percent.unemploy : num 8 7.2 5.7 6.1 4.8 9.5 4.9 10 8.7 6.1 ...  
## $ Per.capita.income : int 20786 21729 19517 19588 24400 16803 18042 17461 17823 21001 ...  
## $ Total.personal.income : int 184230 110928 55003 48931 58818 38658 38287 36872 34525 38911 ...  
## $ Geographic.region : int 4 2 3 4 4 1 4 2 3 3 ...

#### a) Fit the model

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X3 + X4 + X5, data = CDI.Sub)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1866.8 -207.7 -81.5 72.4 3721.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.075e+02 7.028e+01 -2.952 0.00332 \*\*   
## X1 5.515e-04 2.835e-04 1.945 0.05243 .   
## X2 1.070e-01 1.325e-02 8.073 6.8e-15 \*\*\*  
## X3 1.490e+02 8.683e+01 1.716 0.08685 .   
## X4 1.455e+02 8.515e+01 1.709 0.08817 .   
## X5 1.912e+02 8.003e+01 2.389 0.01731 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 566.1 on 434 degrees of freedom  
## Multiple R-squared: 0.9011, Adjusted R-squared: 0.8999   
## F-statistic: 790.7 on 5 and 434 DF, p-value: < 2.2e-16

#### b)

Accordint to the model above ,we can get the model Y=b0+b1\*X1 + b2\*X2 +b3\*X3+b4\*X4+b5\*X5, and the cofficients we can also get from the model fitted.

All the premise is if X1 if fixed:

* When X3=1,X4=0,X5=0,and Y=b0+b1\*X1+b2\*X2+b3
* When X3=0,X4=1,X5=0,and Y=b0+b1\*X1+b2\*X2+b4
* When X3=0,X4=0,X5=1,and Y=b0+b1\*X1+b2\*X2+b5

So,**b3-b4** is equal to the difference between north eastern region and north central region.

To examine whether the effect is different ,that is to

* H0: b3-b4 is equal to 0
* H1: b3-b4 is not equal 0

## [1] "The p-value of the test is : 0.482334"

## [1] "The 10 percent confidence interval is :(-126.415817,133.402068)"

Accordint to the result ,we know that the *p-value* is 0.4823336,which is smaller than 0.05,so we can not reject the null hypothesis,that is,*NC* and *NE* have no different effect on *actiave physicians*

#### c)

To test whether any geograph effects are present ,that is to test b3=b4=b5=0;we can get

* H0: b3,b4,b5 are all equal to 0
* H1: b3,b5,b5 are nor all equal to 0

## [1] "The p-value of test is 0.121032"

According to the result above,the *p-value* is 0.1210319,which is smaller than 0.10.So,at alpha=0.10 level,we can not reject the null hypothesis,that no geographic efffects are presnet.