Copier_Maintenance_Model

Shah Shoib

April 8, 2016

Linkedin Profile

Basic Introduction to Linear Regression

Linear analysis is a statistical methodology tgat utilizes the relationship between two or more quantitative varaibles so that a response or outcome varriable cab be predicted from the other or others.

Four conditions that comprises the simple linear regression model are:

- The mean of the response E(Yi) at each value of the predictor Xi is the linear function of the Xi
- The errors are independent
- The errors at each value of the predictor Xi are normally distributed
- The errors at each value of the predictor Xi have equal variance (Homoscedasticity)

Here is a problem

Copier maintenance. The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data below have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of copiers serviced and Y is the total number of minutes spent by the service person. Assume that first-order regression model (1.1) is appropriate.

```
#Reading the table from the URL
copier_data <- read.table("http://www.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/Kutn
attach(copier_data)
 #Checking the top 3 data rows and number of data in the given dataset
head(copier_data,3);dim(copier_data)
```

```
##   Time No_Copiers
## 1   20          2
## 2   60          4
## 3   46          3
```

```
## [1] 45  2
```

```
summary(copier_data)
```

```
##       Time          No_Copiers
##  Min.   :  3.00   Min.   : 1.000
##  1st Qu.: 36.00   1st Qu.: 2.000
##  Median : 74.00   Median : 5.000
```

```
##  Mean   : 76.27   Mean   : 5.111
##  3rd Qu.:111.00   3rd Qu.: 7.000
##  Max.   :156.00   Max.   :10.000
```

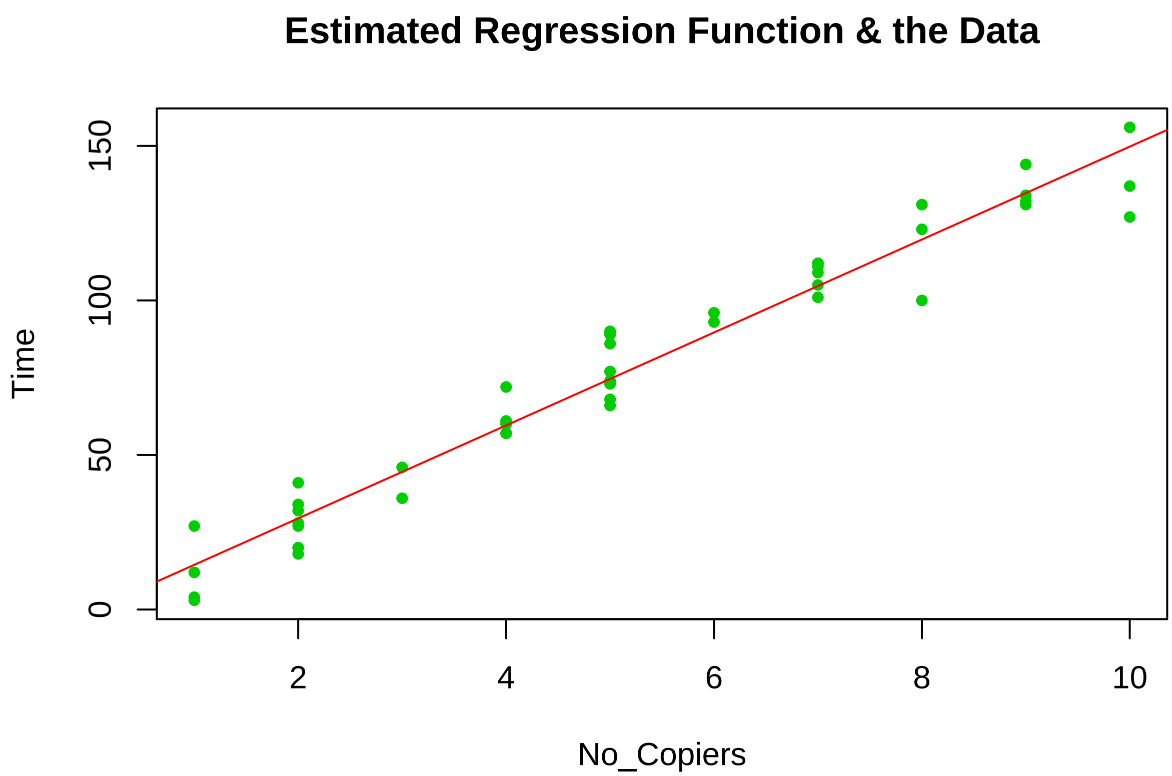
```r
# Estimated regression function.
copier.lm <- lm(Time~No_Copiers,data=copier_data)
#Summary of Regression Model
summary(copier.lm)
```

```
##
## Call:
## lm(formula = Time ~ No_Copiers, data = copier_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7723  -3.7371   0.3334   6.3334  15.4039
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5802     2.8039  -0.207    0.837
## No_Copiers   15.0352     0.4831  31.123   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.914 on 43 degrees of freedom
## Multiple R-squared:  0.9575, Adjusted R-squared:  0.9565
## F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16
```

```r
#Extracting fitted values
yhat <- -0.5802 + 15.0352*No_Copiers

#Plot the estimated regression function and the data.
plot(No_Copiers,Time,main="Estimated Regression Function & the Data",pch=20,col=491)
abline(copier.lm,coef=coef(copier.lm),lty=1,col="red")
```

```
## Warning in abline(copier.lm, coef = coef(copier.lm), lty = 1, col =
## "red"): 'a' and 'b' are overridden by 'coef'
```




```
anova(copier.lm)
```

```
## Analysis of Variance Table
##
## Response: Time
##             Df Sum Sq Mean Sq F value    Pr(>F)
## No_Copiers   1  76960   76960  968.66 < 2.2e-16 ***
## Residuals   43   3416      79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Obtain a point estimate of the mean service time when X = 5 copiers are
#serviced.
yhat <- -0.5802 + 15.0352*5


#Obtaining the residuals
e <- copier.lm$residuals

#Obtaining the Error sum of squares (SSE) and Error mean square (MSE)
SSE <- anova(copier.lm)[2,2]
#MIN Q = sum(copier.lm$residuals)^2
MSE <- anova(copier.lm)[2,3]
sqrt(MSE)
```

```
## [1] 8.913508
```

```
b1 <-15.0352

#Obtain point estimates of variance and sd?
xbar = mean(No_Copiers)
var.b1 <- MSE/sum((No_Copiers-xbar)^2)
s.b1 <- sqrt(var.b1)

#Estimate the change in the mean service time when the number of copiers
#serviced increases by one. Use a 90 percent confidence interval. #Interpret your confidence interval.

t<-qt(1-.05,45-2)
#The 90% confidence interval is

rightb1<-b1-t*s.b1
leftb1<-b1+t*s.b1
c(rightb1,leftb1)
```

```
## [1] 14.2231 15.8473
```

```
# H0:b1 equal = 0 & Ha:b1 not equal = 0

t.star = (b1-14)/s.b1
#if |t.star| less than equal to t(1-alpha/2;n-2),concule H0
#if |t.star| greater than t(1-alpha/2;n-2),concule Ha
n=45
alpha=.05
qt(1-alpha/2,n-2)
```

```
## [1] 2.016692
```

```
#Since 31.12316 > 2.016692, we conclude Ha that b1 is not equal to 0 or
#that  there is a linear association between X and Y. The value of t
#-statistic is 31.12316

#Conduct a t test to determine whether or not there is a linear associatio
#n between X and Y here; control the a risk at .10. State the alternatives
#,decision rule, and conclusion. What is the P-value of your test?


pt( 2.1428,45-2)
```

```
## [1] 0.9810845
```

```
# H0:b1 equal = 14 & Ha:b1 not equal = 14

var.b0 <- MSE*(1/n+(xbar^2/sum(No_Copiers-xbar)^2))
sb.b0 <- sqrt(var.b0)
b0<--0.5802

t.star = (b0)/sb.b0
```

```r
t<-qt(1-.05,45-2)
p.value = pt(-abs(t.star), df=length(Time)-1)

#Set up the basic ANOVA table in the format ofTable 2.2. Which elements
#of your table are additive? Also set up the ANOVA table in the format
#of Table 2.3. How do the two tables differ?

(anova(copier.lm))
```

```
## Analysis of Variance Table
##
## Response: Time
##            Df Sum Sq Mean Sq F value    Pr(>F)
## No_Copiers  1  76960   76960  968.66 < 2.2e-16 ***
## Residuals  43   3416      79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
SSR <- anova(copier.lm)[1,2]
SSE <- anova(copier.lm)[2,2]
MSE <- anova(copier.lm)[2,3]
SSTO<- sum(anova(copier.lm)[,2])

pvalue <- anova(copier.lm)[1,5]
MSR <- SSR/1
#Conduct an F test to determine whether or not there is a linear associati
#on between time spent and number of copiers serviced; use a = .10. State
#the alternatives, decision rule, and conclusion.

(Fstar <- MSR/MSE)
```

```
## [1] 968.6572
```

```r
#F-statistics
qf(1-.10,1,43)
```

```
## [1] 2.825999
```

```r
#H0: b1=0
#Ha:b1 != 0

#Decision rule when the risk  of a Type I error is to be controlled at
#alpha

# if Fstar <= F(1-apha,1,n-2), conclude H0
# if Fstar > F(1-apha,1,n-2), conclude Ha

#Since Fstar = 968.65 we conclude Ha

#By how much, relatively, is the total variation in number of minutes
#spent on a call- reduced when the number of copiers serviced is introduce
```

```r
#d into the analysis? Is this a relatively small or large reduction? What
#is the name of this measure?

Rsquare <- SSR/SSTO #Coefficient of determination

#Calculate r and attach the appropriate sign.
sqrt(Rsquare)
```

```
## [1] 0.978517
```

```r
#Which measure, r or R2, has the more clear-cut operational interpretation?
Rsquare
```

```
## [1] 0.9574955
```

```r
#problem 4.3 page 172

#Extracting the regression coefficients
b0 <- summary(copier.lm)$coefficients[1, 1] #Intercept
b1 <- summary(copier.lm)$coefficients[2, 1] #Slope

#Extracting standard errors for the regression coefficients
sd.b0 <- summary(copier.lm)$coefficients[1, 2]
sd.b1 <- summary(copier.lm)$coefficients[2, 2]

# Calulating Benferroni Interval
alpha <- .05
g<-2
n<-45
bfi <- qt(1-alpha/(2*g),n-2)
bfi
```

```
## [1] 2.322618
```

```r
# Benferroni Joint Confidence Interval
# For intercept b0
bfi_b0_right <- b0 - bfi*sd.b0
bfi_b0_left <- b0 + bfi*sd.b0
c(bfi_b0_right,bfi_b0_left)
```

```
## [1] -7.092642  5.932329
```

```r
# For intercept b1
bfi_b1_right <- b1 - bfi*sd.b1
bfi_b1_left <- b1 + bfi*sd.b1
c(bfi_b1_right,bfi_b1_left)
```

```
## [1] 13.91322 16.15728
```

```r
# Benferron Joint Confidence Interval for beta0 and beta1 are
paste(bfi_b0_right,"<=beta0<=",bfi_b0_left)
```

```
## [1] "-7.09264201944444 <=beta0<= 5.93232870351749"
```

```r
paste(bfi_b1_right,"<=beta1<=",bfi_b1_left)
```

```
## [1] "13.9132207611322 <=beta1<= 16.1572753224187"
```

```r
# As per the recommendation suggested by the consultant for beta0 and
#beta1 lies in the confidence interval calulated using benferroni
#confidence interval

detach(copier_data)
```