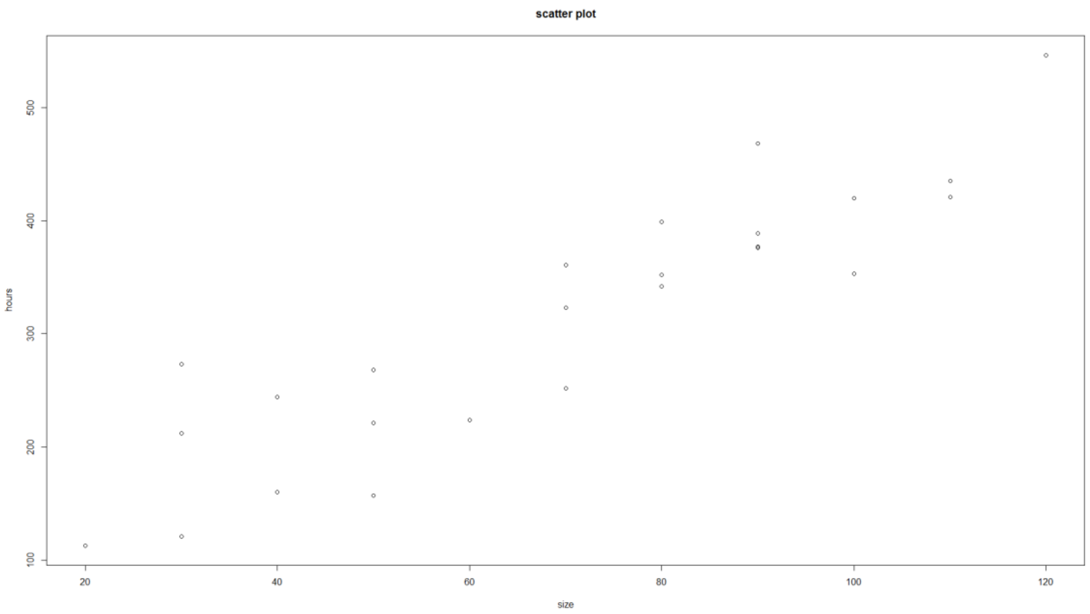
> ###testing constant variance#########
##Toluca Company example. To discover the relationship
##between lot size and labor hours required to produce the lot. Data
##on lot size and work hours for 25 recent production runs were utilized

```
tuluca<-read.table(file="C:/jenn/teaching/stat440540/data/CH1/CH01TA01.txt")
> tuluca
   V1  V2
1   80 399
2   30 121
3   50 221
4   90 376
> size<-tuluca$V1
```

- hours<-tuluca$V2

```
plot(size,hours, main="scatter plot")
```




```
> myfit<-lm(hours~size)
> b_0 <- myfit$coef[1]
> b_1 <- myfit$coef[2]
> plot(size,hours, main="Fitted Line Plot")
> abline(b_0, b_1)
> resid<-myfit$residuals
> residsquare<-resid^2
    myfit2<-lm(residsquare~size)
```

```
anova(myfit)
#myfit<-lm(hours~size)

Analysis of Variance Table

Response: hours
          Df  Sum Sq  Mean Sq  F value   Pr(>F)
size       1  252378  252378   105.88  4.449e-10 ***
Residuals 23   54825    2384

> anova(myfit2) # myfit2<-lm(residsquare~size)

Analysis of Variance Table

Response: residsquare
          Df     Sum Sq   Mean Sq  F value  Pr(>F)
size       1    7896142   7896142   1.0914   0.307
Residuals 23  166395896   7234604
```

BP statistic = (7896142/2)/(54825/25)^2 = 0.821, compared with chi^2 (0.95, 1) = 3.84, don't reject H_0, that the error variance is constant.

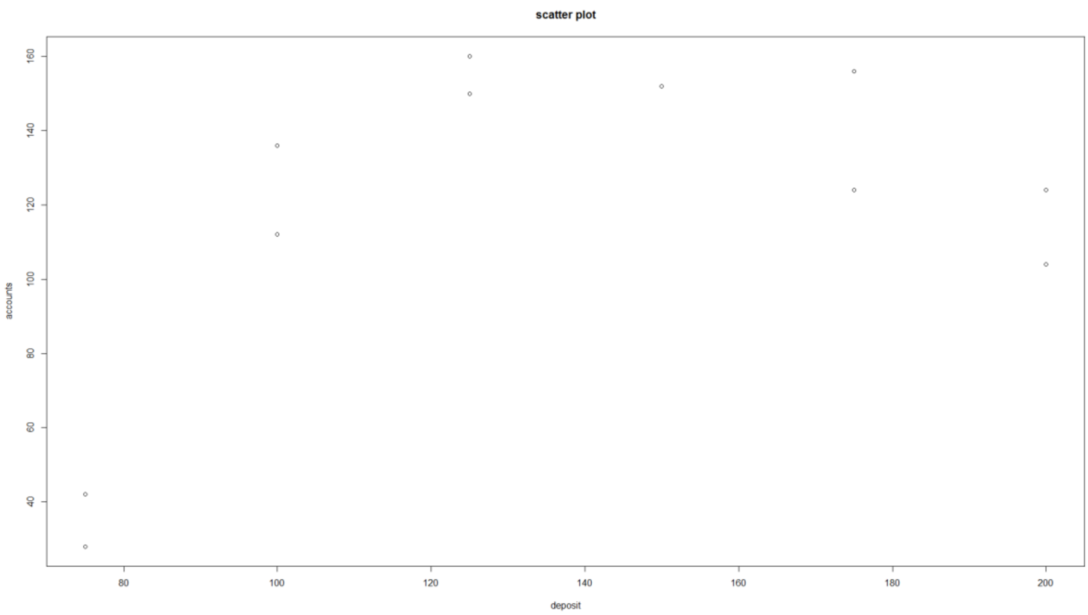
```
> bptest(hours~size,studentize=FALSE)

        Breusch-Pagan test

data:  hours ~ size
BP = 0.82092, df = 1, p-value = 0.3649

>
> shapiro.test(resid)

        Shapiro-Wilk normality test

data:  resid
W = 0.9789, p-value = 0.8626
```

p-value is 0.8626, do not reject null hypothesis of normality.

#######Fisher's lack of fit test###########

12 suburban branch offices of a commercial bank.
Holders of checking accounts were offered gifts for setting up money market accounts. Minimum deposits were required and the value of the gift was directly proportional to the minimum deposit. 6 levels were used. One bank dropped out. Linear fit is bad.

```
> bank<-read.table(file="C:/jenn/teaching/stat440540/data/CH3/CH03TA04.txt")
> bank
    V1  V2
1  125 160
2  100 112
3  200 124
4   75  28
5  150 152
6  175 156
7   75  42
8  175 124
9  125 150
10 200 104
11 100 136
> deposit<-bank$V1
> accounts<-bank$V2
plot(deposit,accounts, main="scatter plot")
```




```
> #full model
> myfit = aov(accounts~factor(deposit), data=bank)
> summary(myfit)
                 Df   Sum Sq  Mean Sq   F value   Pr(>F)
factor(deposit)   5   18735    3747      16.32    0.00409 **
Residuals         5   1148      230
```

```
> #reduced model
> myfit2<-lm(accounts~deposit)
> summary(myfit2)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  50.7225    39.3979   1.287     0.23
deposit       0.4867     0.2747   1.772     0.11
```

Residual standard error: 40.47 on 9 degrees of freedom
Multiple R-squared: 0.2586,   Adjusted R-squared: 0.1762
F-statistic: 3.139 on 1 and 9 DF, p-value: 0.1102

> summary(myfit)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(deposit)	5	18735	3747	16.32	0.00409 **
Residuals	5	1148	230		

> anova(myfit2)
Analysis of Variance Table

Response: accounts

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
deposit	1	5141.3	5141.3	3.1389	0.1102
Residuals	9	14741.6	1638.0		

>
F^* = MSLF/MSPE = [(14741.6 – 1148)/4]/(1148/5) = 3398.4/229.6 = 14.80, let \alpha = .01, then
F(.99; 4, 5) =11.4, F^* = 14.8>11.4, P value for the test is
0.006. we conclude $H_{\alpha}, that the regression function is not linear.

#generalized linear test
> anova(myfit, myfit2)
Analysis of Variance Table

Model 1: accounts ~ factor(deposit)
Model 2: accounts ~ deposit

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	5	1148				
2	9	14742	-4	-13594	14.801	0.005594 **

Reject the reduced model, conclude that the regression function is not linear.

```
##use boxcox to select transformation
library(MASS)
boxcox(plasma ~ age, lambda = seq(-2, 2, length = 10))

##find sses in TABLE 3.9

#W_I = K_1(Y_I^{\lambda} –1), if \lambda \neq 0
#W_I = K_2 (log_e Y_I), if \lambda =0
#Where
#K_2 = (\prod_{I=1}^n Y_I)^{1/n}
#K_1 = 1/\lambda K_2^{\lambda –1}

gmean <- exp(mean(log(plasma)))  #geometric mean K_2 in page 135
sse <- c()  #assign null vector to sse
lambda<-c()  #assign null vector to lambda
i <- 1  #assign 1 to i as initial value
##construct a loop to calculate sses for each lambda
for (lam in seq(-1,1,0.1)){
  if (lam != 0){
  tY <- (plasma^lam - 1) / (lam*gmean^(lam-1))
  }
else {
  tY <- log(plasma)*gmean
  }
  test <- anova(lm(tY~age))
  sse[i] <- test['Residuals','Sum Sq']
  lambda[i] <- lam
  i <- i+1
}
cbind(lambda,sse)

> cbind(lambda,sse)
      lambda      sse
 [1,]   -1.0 33.90887
 [2,]   -0.9 32.70442
 [3,]   -0.8 31.76453
 [4,]   -0.7 31.09066
 [5,]   -0.6 30.68680
 [6,]   -0.5 30.55961
 [7,]   -0.4 30.71859
 [8,]   -0.3 31.17631
 [9,]   -0.2 31.94867
[10,]   -0.1 33.05520
[11,]    0.0 34.51945
[12,]    0.1 36.36939
[13,]    0.2 38.63791
[14,]    0.3 41.36342
[15,]    0.4 44.59051
[16,]    0.5 48.37072
[17,]    0.6 52.76343
```

```
[18,]    0.7 57.83686
[19,]    0.8 63.66932
[20,]    0.9 70.35050
[21,]    1.0 77.98306
>
```