hw chap2

2.1

```
Y = c(1907, 3000, 1287, 1947, 2700, 2273, 2373, 3113, 3260, 2493)
X = c(5.1, 7.8, 3.5, 4.5, 7.1, 5.6, 6.2, 8.0, 8.8, 6.4)
lm.sol<-lm(Y~1+X)
CC=summary(lm.sol, digits = digits, maxsum = maxsum)
print(CC)
```

```
##
## Call:
## lm(formula = Y ~ 1 + X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -128.591  -70.978   -3.727   49.263  167.228
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   140.95     125.11   1.127    0.293
## X             364.18      19.26  18.908 6.33e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 96.42 on 8 degrees of freedom
## Multiple R-squared:  0.9781, Adjusted R-squared:  0.9754
## F-statistic: 357.5 on 1 and 8 DF,  p-value: 6.33e-08
```

```
new <- data.frame(X=7)
a =predict(lm.sol,new,interval="prediction")
print(a)
```

```
##        fit      lwr      upr
## 1 2690.227 2454.971 2925.484
```

```
a =predict(lm.sol,new,interval="confidence")
print(a)
```

```
##        fit     lwr      upr
## 1 2690.227 2613.35 2767.105
```

```
new <- data.frame(X=7)
a =predict(lm.sol,new,interval="prediction")
print(a)
```

```
##        fit      lwr      upr
## 1 2690.227 2454.971 2925.484
```

```
a =predict(lm.sol,new,interval="confidence")
print(a)
```

```
##        fit     lwr      upr
## 1 2690.227 2613.35 2767.105
```

回归方程

$$Y = 140.95 + 364.18X$$

6.33e-08<0.05 R^2=0.9754

系数，方程，相关系数都通过检验

2.

预测值2690.227 置信区间[2613.35, 2767.105] 预测区间[2454.971 ,2925.484]

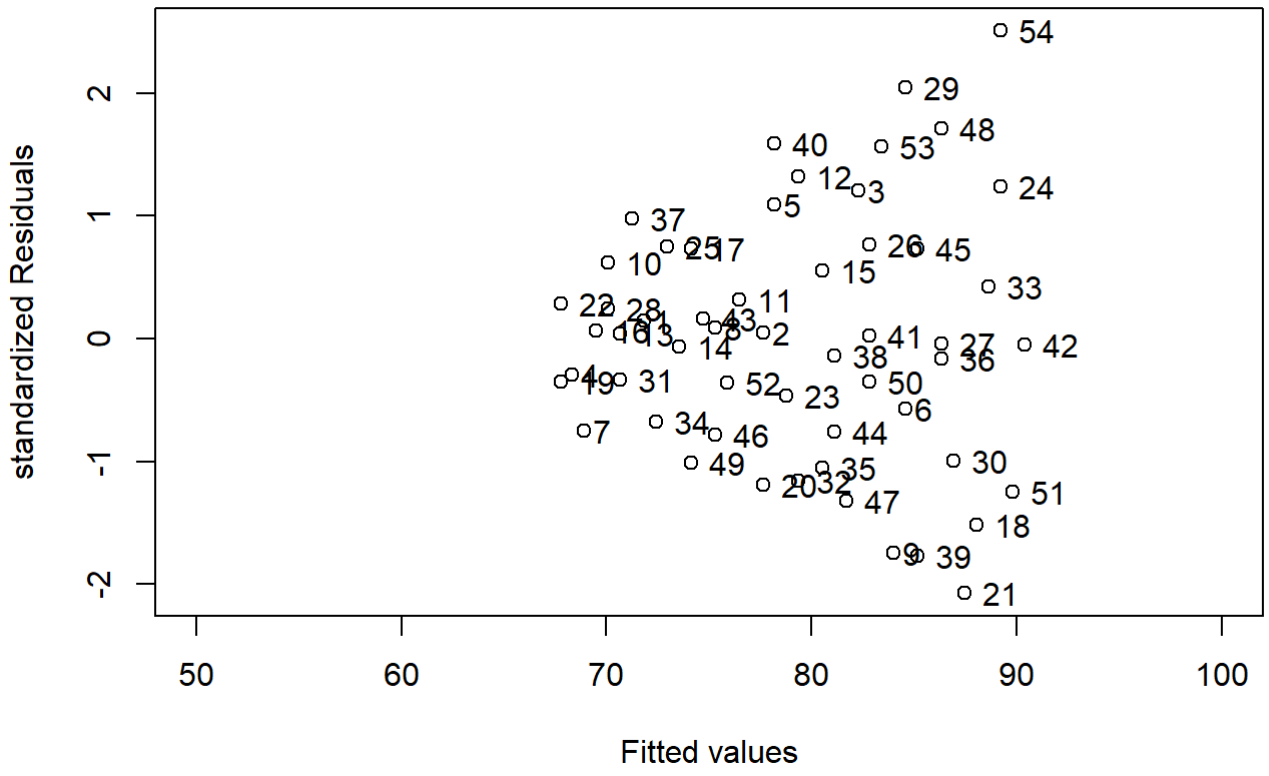
2.2

2

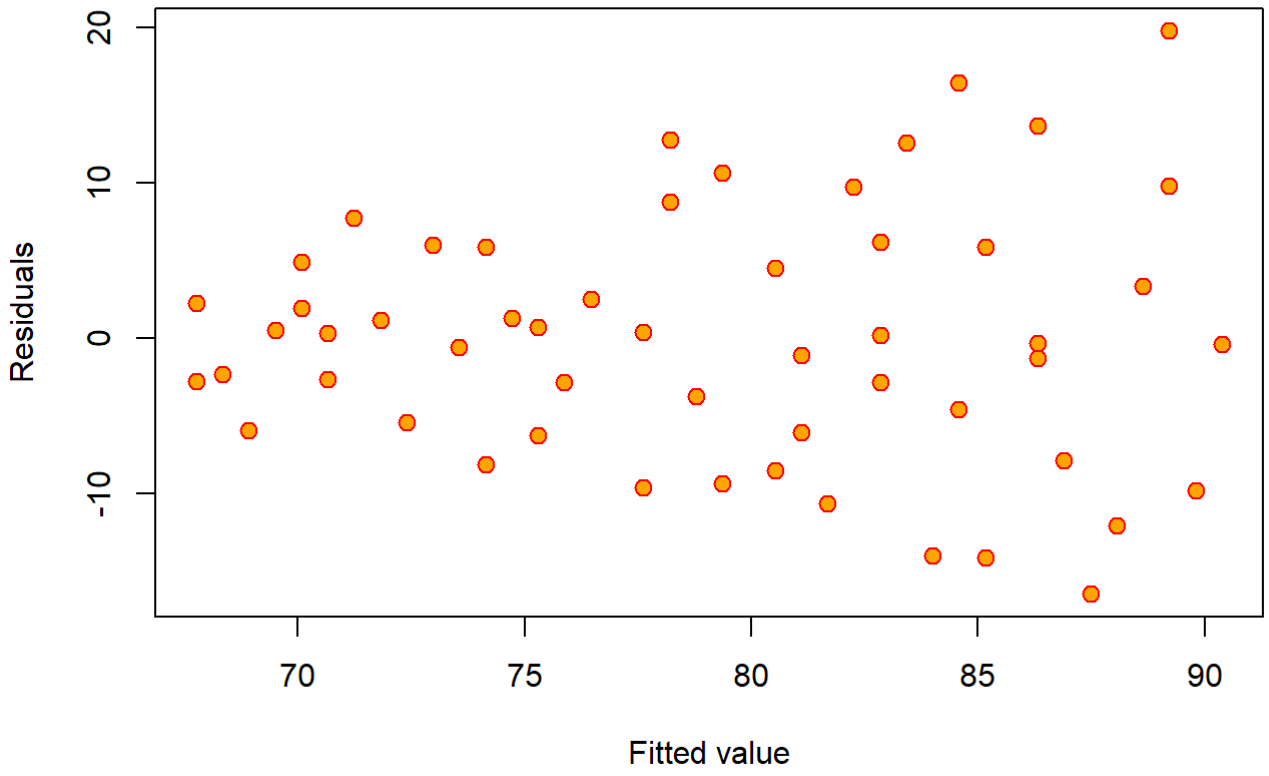
```
Y=c(73, 78, 92, 66, 87, 80, 63, 76, 70, 75, 79, 90, 71, 73, 85, 70, 80, 76, 65, 68, 71, 70, 75,
99, 79, 89, 86, 72, 101, 79, 68, 70, 92, 67, 72, 85, 79, 80, 71, 91, 83, 90, 76, 75, 91, 69, 71
, 100, 66, 80, 80, 73, 96, 109)
X = c(27, 37, 45, 21, 38, 49, 22, 33, 48, 24, 35, 40, 25, 30, 42, 23, 31, 55, 20, 37, 54, 20, 3
9, 57, 29, 46, 52, 24, 49, 53, 25, 40, 56, 28, 42, 52, 26, 43, 50, 38, 46, 59, 32, 43, 50, 33,
44, 52, 31, 46, 58, 34, 47, 57)

lm.sol<-lm(Y~1+X)
CC=summary(lm.sol, digits = digits, maxsum = maxsum)
print(CC)
```

```
##
## Call:
## lm(formula = Y ~ 1 + X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.4786  -5.7877  -0.0784   5.6117  19.7813
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 56.15693    3.99367  14.061  < 2e-16 ***
## X            0.58003    0.09695   5.983 2.05e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.146 on 52 degrees of freedom
## Multiple R-squared:  0.4077, Adjusted R-squared:  0.3963
## F-statistic: 35.79 on 1 and 52 DF,  p-value: 2.05e-07
```

```
pre<- fitted.values(lm.sol)
rst = rstandard(lm.sol)
plot (pre, rst, xlab = "Fitted values" ,ylab = "standardized Residuals", xlim = c(50,100))
text (pre, rst, adj = c (- 0.5,0.5))
```



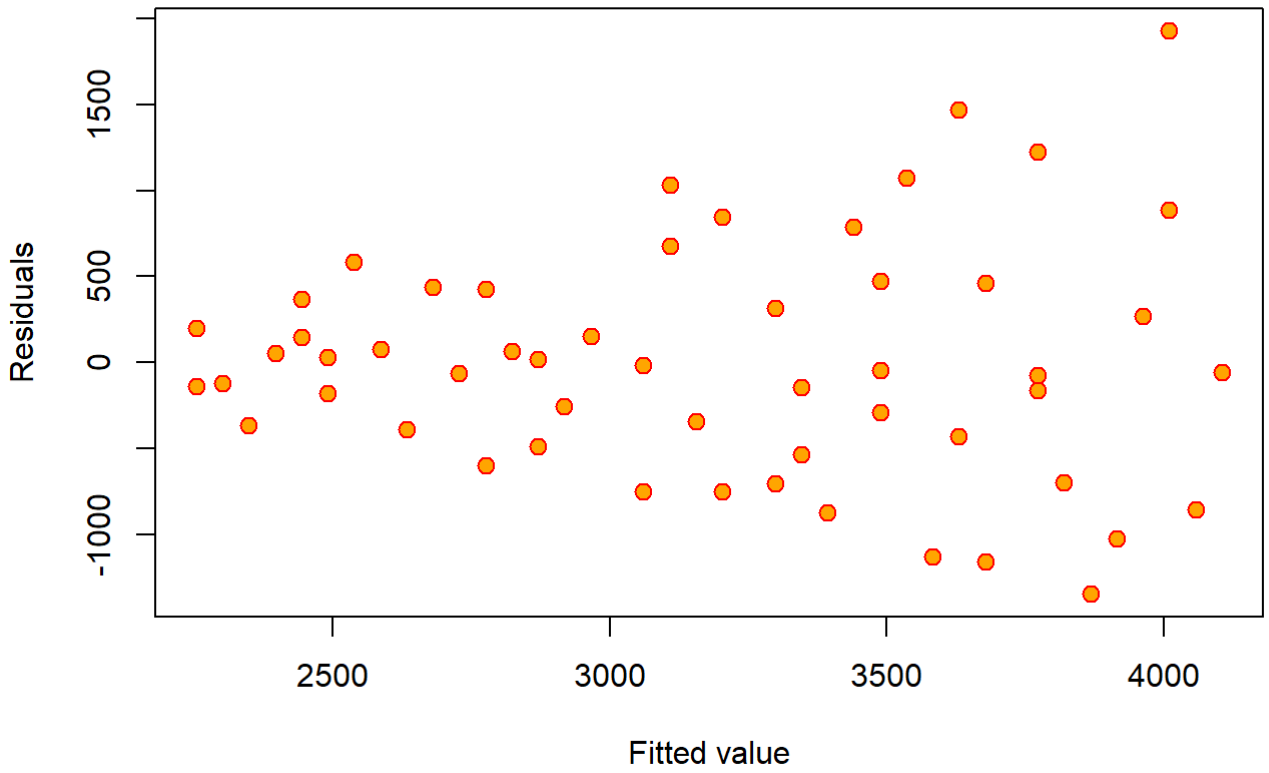

```
plot (fitted (lm.sol) , resid (lm.sol) ,cex = 1.2, pch = 21,col = "red" , bg = "orange",xlab =
"Fitted value", ylab = "Residuals")
```




```
# boxcox ( lm.sol, lambda = seq(0,1, by = 0.1) )##告号Box- cox变换后，作回归分析
lambda<- 2; Ylam <-(Y^lambda-1) / lambda
lm.lam <- lm ( Ylam ~ X) ; summary ( lm.lam)
```

```
##
## Call:
## lm(formula = Ylam ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1347.87  -420.43   -49.88   410.23  1929.65
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1303.192    337.285   3.864 0.000311 ***
## X             47.494      8.188   5.800 3.97e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 687.9 on 52 degrees of freedom
## Multiple R-squared:  0.3928, Adjusted R-squared:  0.3812
## F-statistic: 33.65 on 1 and 52 DF,  p-value: 3.971e-07
```

```
plot (fitted ( lm.lam) , resid (lm.lam),
cex = 1.2, pch= 21,col = "red", bg = "orange" ,xlab ="Fitted value" , ylab = "Residuals")
```



回归方程

$$Y = 56.15693 + 0.58003X$$

系数，方程，相关系数都通过检验，R较小 由图可知为异方差，

用λ=2,作boxcox变换后基本满足要求

2-3

```
Y=c(64, 51, 60, 76, 71, 96, 61, 77, 54, 93, 77, 95, 81, 54, 93, 168, 93, 99)
X2 = c(52, 58, 23, 37, 19, 46, 34, 50, 24, 44, 65, 56, 44, 36, 31, 58, 29, 51)
X1 = c(0.4, 12.6, 0.4, 10.9, 3.1, 23.1, 0.6, 23.1, 4.7, 21.6, 1.7, 23.1, 9.4, 1.9, 10.1, 26.8,
11.6, 29.9)
X3 = c(158, 112, 163, 111, 37, 114, 157, 134, 59, 73, 123, 168, 46, 143, 117, 202, 173, 124)
lm.sol<-lm(Y~1+X1+X2+X3)
CC=summary(lm.sol, digits = digits, maxsum = maxsum)
print(CC)
```

```
##
## Call:
## lm(formula = Y ~ 1 + X1 + X2 + X3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.349 -11.383  -2.659  12.095  48.807
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 43.65007   18.05442   2.418  0.02984 *
## X1           1.78534    0.53977   3.308  0.00518 **
## X2          -0.08329    0.42037  -0.198  0.84579
## X3           0.16102    0.11158   1.443  0.17098
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.97 on 14 degrees of freedom
## Multiple R-squared:  0.5493, Adjusted R-squared:  0.4527
## F-statistic: 5.688 on 3 and 14 DF,  p-value: 0.009227
```

```
lm.step<- step(lm.sol)
```

```
## Start:  AIC=111.27
## Y ~ 1 + X1 + X2 + X3
##
##        Df Sum of Sq    RSS    AIC
## - X2    1      15.7 5599.4 109.32
## <none>              5583.7 111.27
## - X3    1     830.6 6414.4 111.77
## - X1    1    4363.4 9947.2 119.66
##
## Step:  AIC=109.32
## Y ~ X1 + X3
##
##        Df Sum of Sq     RSS    AIC
## <none>               5599.4 109.32
## - X3    1     833.2  6432.6 109.82
## - X1    1    5169.5 10768.9 119.09
```

X2,X3的系数不能通过t检验（p值分别为0.84579，0.17098）

逐步回归舍去X2,再跑：

```
Y=c(64, 51, 60, 76, 71, 96, 61, 77, 54, 93, 77, 95, 81, 54, 93, 168, 93, 99)
# X2 = c(52, 58, 23, 37, 19, 46, 34, 50, 24, 44, 65, 56, 44, 36, 31, 58, 29, 51)
X1 = c(0.4, 12.6, 0.4, 10.9, 3.1, 23.1, 0.6, 23.1, 4.7, 21.6, 1.7, 23.1, 9.4, 1.9, 10.1, 26.8,
11.6, 29.9)
X3 = c(158, 112, 163, 111, 37, 114, 157, 134, 59, 73, 123, 168, 46, 143, 117, 202, 173, 124)
lm.sol<-lm(Y~1+X1+X3)
CC=summary(lm.sol, digits = digits, maxsum = maxsum)
print(CC)
```

```
##
## Call:
## lm(formula = Y ~ 1 + X1 + X3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.713 -11.324  -2.953  11.286  48.679
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  41.4794    13.8834   2.988  0.00920 **
## X1            1.7374     0.4669   3.721  0.00205 **
## X3            0.1548     0.1036   1.494  0.15592
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.32 on 15 degrees of freedom
## Multiple R-squared:  0.5481, Adjusted R-squared:  0.4878
## F-statistic: 9.095 on 2 and 15 DF,  p-value: 0.002589
```

```
# lm.step<- step(lm.sol)
```

X3系数P值仍大于0.05，可以删去X3.

2-4

假设:

$$H_0：无差别$$
$$H_1：有差别$$

```
lamp <- data.frame (X= c (148,76,393,520,236,134,55,166,415,153,513,264,433,94,535,327,214,135,
280,304,335,643,216,536,128,723,258,380,594,465),A= factor (rep (1:3,c (10,10,10) )))
lamp.aov <- aov (X ~ A,data = lamp)
summary ( lamp.aov)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## A            2 198772   99386   3.482 0.0452 *
## Residuals   27 770671   28543
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value = 0.0452<0.05

拒绝假设，有差别

```
rm(list = ls())
lamp <- data.frame (X= c (148,76,393,520,236,134,55,166,415,153,513,264,433,94,535,327,214,135,
280,304,335,643,216,536,128,723,258,380,594,465),A= factor (rep (1:3,c (10,10,10) )))
# lamp <- data.frame (X= c ( 2,4,3,2,4,7,7,2,2,5,4,5,6,8,5,10,7,12,12,6,6,7,11,6,6,7,9,5,5,10,
6,3,10),A= factor (rep (1:3,c (11,10,12) )))
lamp.aov <- lm(X ~ A,data = lamp)
anova( lamp.aov)
```

```
## Analysis of Variance Table
##
## Response: X
##           Df Sum Sq Mean Sq F value  Pr(>F)
## A          2 198772   99386  3.4819 0.04515 *
## Residuals 27 770671   28543
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# print((lamp))
# print(X)
attach(lamp)
tapply(X,A,mean)
```

```
##     1     2     3
## 229.6 309.9 427.8
```

```
g = pairwise.t.test(X,A)
print(g)
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  X and A
##
##   1     2
## 2 0.297 -
## 3 0.042 0.261
##
## P value adjustment method: holm
```

1-3的p-value = 0.042<0.05

故1-3有差别，1-2,2-3无差别

2-5

oneway.text:

```
rm(list = ls())
lamp <- data.frame (X= c (148,76,393,520,236,134,55,166,415,153,513,264,433,94,535,327,214,135,
280,304,335,643,216,536,128,723,258,380,594,465),A= factor (rep (1:3,c (10,10,10) )))

# lamp <- data.frame (X= c ( 2,4,3,2,4,7,7,2,2,5,4,5,6,8,5,10,7,12,12,6,6,7,11,6,6,7,9,5,5,10,
6,3,10),A= factor (rep (1:3,c (11,10,12) )))
# lamp.aov <- lm(X ~ A,data = lamp)
# anova( lamp.aov)
# # print((lamp))
# print(X)
# attach(lamp)
# tapply(X,A,mean)
# g = pairwise.t.test(X,A)
# print(g)
oo = oneway.test(X~A,data = lamp)
print(oo)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  X and A
## F = 2.9699, num df = 2.000, denom df = 17.775, p-value = 0.07713
```

p-value = 0.07713

kruskal.test:

```
rm(list = ls())
lamp <- data.frame (X= c (148,76,393,520,236,134,55,166,415,153,513,264,433,94,535,327,214,135,
280,304,335,643,216,536,128,723,258,380,594,465),A= factor (rep (1:3,c (10,10,10) )))

# lamp <- data.frame (X= c ( 2,4,3,2,4,7,7,2,2,5,4,5,6,8,5,10,7,12,12,6,6,7,11,6,6,7,9,5,5,10,
6,3,10),A= factor (rep (1:3,c (11,10,12) )))
# lamp.aov <- lm(X ~ A,data = lamp)
# anova( lamp.aov)
# print((lamp))
# print(X)
# attach(lamp)
# tapply(X,A,mean)
# g = pairwise.t.test(X,A)
# print(g)
oo = kruskal.test(X~A,data = lamp)
print(oo)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  X by A
## Kruskal-Wallis chi-squared = 5.4684, df = 2, p-value = 0.06495
```

p-value = 0.06495

正态性检验:

```
rm(list = ls())
lamp <- data.frame (X= c (148,76,393,520,236,134,55,166,415,153,513,264,433,94,535,327,214,135,
280,304,335,643,216,536,128,723,258,380,594,465),A= factor (rep (1:3,c (10,10,10) )))

# lamp <- data.frame (X= c ( 2,4,3,2,4,7,7,2,2,5,4,5,6,8,5,10,7,12,12,6,6,7,11,6,6,7,9,5,5,10,
6,3,10),A= factor (rep (1:3,c (11,10,12) )))
# lamp.aov <- lm(X ~ A,data = lamp)
# anova( lamp.aov)
cc = with ( lamp, tapply (X,A, shapiro.test))
print(cc)
```

```
## $`1`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.87975, p-value = 0.1296
##
##
## $`2`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.95268, p-value = 0.7003
##
##
## $`3`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.97137, p-value = 0.9033
```

p-value = [0.1296,0.7003,0.9033] 通过检验

方差齐性检验:

```
rm(list = ls())
lamp <- data.frame (X= c (148,76,393,520,236,134,55,166,415,153,513,264,433,94,535,327,214,135,
280,304,335,643,216,536,128,723,258,380,594,465),A= factor (rep (1:3,c (10,10,10) )))

# lamp <- data.frame (X= c ( 2,4,3,2,4,7,7,2,2,5,4,5,6,8,5,10,7,12,12,6,6,7,11,6,6,7,9,5,5,10,
6,3,10),A= factor (rep (1:3,c (11,10,12) )))
# lamp.aov <- lm(X ~ A,data = lamp)
# anova( lamp.aov)
# cc = with ( lamp, tapply (X,A, shapiro.test))
# print(cc)
# print((lamp))
# print(X)
# attach(lamp)
# tapply(X,A,mean)
# g = pairwise.t.test(X,A)
# # print(g)
oo = bartlett.test(X~A,data = lamp)
print(oo)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  X by A
## Bartlett's K-squared = 0.79406, df = 2, p-value = 0.6723
```

p-value = 0.6723，通过检验

综上，正态性检验更准确