# HW#1 Solution

*9/12/2019*

### HW 16.7 Productivity improvement.

```r
# HW 16.07
HW07 <- read.table(url("https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week2/CH16PR07.txt
# rename the variables
names(HW07)<- c("productivity", "expenditures", "firm")

HW07$expenditures<- as.factor(HW07$expenditures)
head(HW07)
```

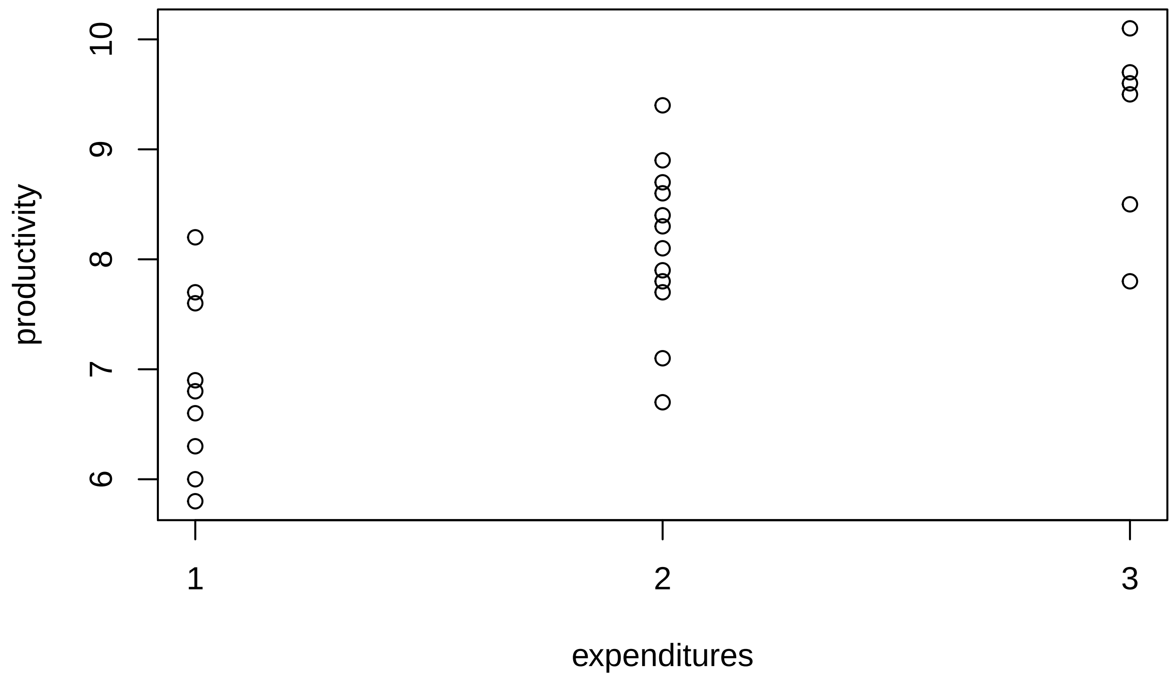
```
##   productivity expenditures firm
## 1          7.6            1    1
## 2          8.2            1    2
## 3          6.8            1    3
## 4          5.8            1    4
## 5          6.9            1    5
## 6          6.6            1    6
```

```r
str(HW07)
```

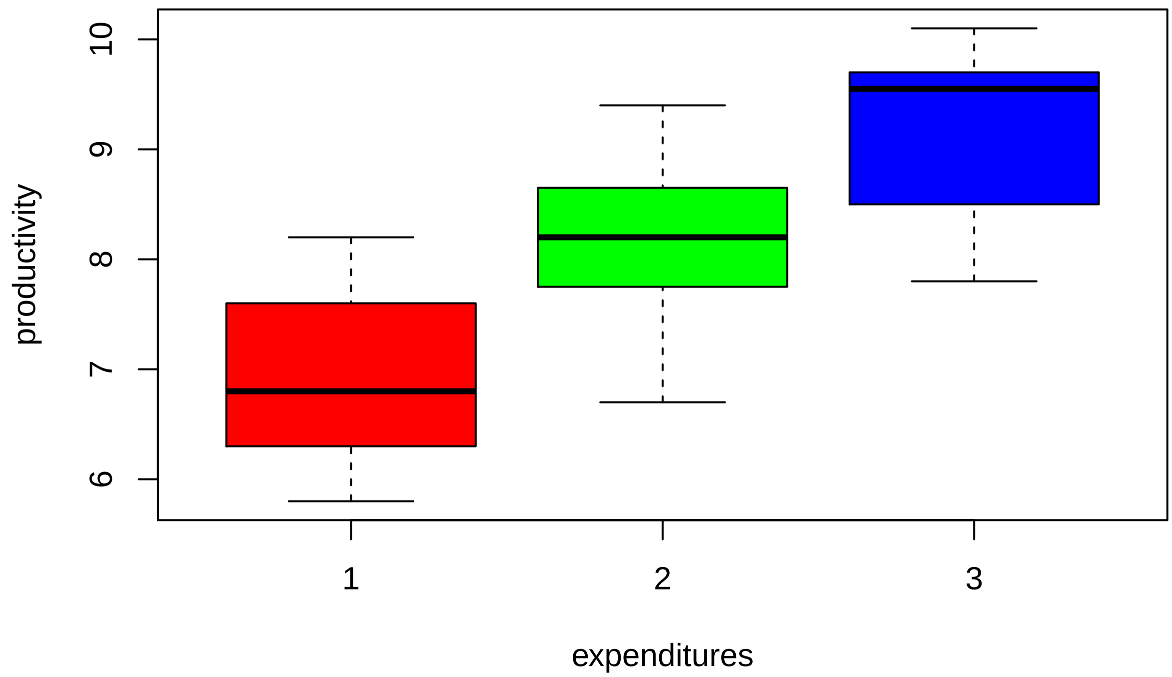
```
## 'data.frame':    27 obs. of  3 variables:
##  $ productivity: num  7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6 6.7 ...
##  $ expenditures: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 2 ...
##  $ firm        : int  1 2 3 4 5 6 7 8 9 1 ...
```

#### a. aligned dot plots:

```r
stripchart(productivity ~ expenditures, vertical = TRUE,  pch=1, data = HW07, xlab="expenditures")
```




```
boxplot(productivity ~ expenditures, data = HW07, col=rainbow(3))
```




**Results: the means appear to differ, and the varibilities appear to be similar.**

**b. Obtain the fitted values:**

```
fit <- aov(productivity ~ expenditures, data = HW07)
# factor level means= fitted values
predict(fit, newdata = data.frame(expenditures = factor(1:3)))
```

```
##        1        2        3
## 6.877778 8.133333 9.200000
```

**c. Obtain the residuals and sum:**

```
# residials
fit$residuals
```

```
##           1           2           3           4           5           6
##  0.72222222  1.32222222 -0.07777778 -1.07777778  0.02222222 -0.27777778
##           7           8           9          10          11          12
## -0.57777778  0.82222222 -0.87777778 -1.43333333 -0.03333333  1.26666667
##          13          14          15          16          17          18
```

```
##  0.46666667 -0.33333333 -0.43333333  0.76666667 -0.23333333  0.16666667
##          19          20          21          22          23          24
##  0.56666667 -1.03333333  0.26666667 -0.70000000  0.50000000  0.90000000
##          25          26          27
## -1.40000000  0.40000000  0.30000000
```

```
## sum of residuals
sum(fit$residuals)
```

```
## [1] 6.383782e-16
```

### d. ANOVA table

```
summary(fit)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## expenditures  2  20.12   10.06   15.72 4.33e-05 ***
## Residuals    24  15.36    0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### e-f.

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_a : \textit{not all} \quad \mu_i \quad \textit{are equal.}$$

**Results:** $F^*$**= 15.72 with a p-value <0.001, we reject** $H_0$ **and conclude** $H_a$**.**

## HW 16.10 Cash offers.

```
# HW 16.10
HW10 <- read.table(url("https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week2/CH16PR10.txt
dim(HW10)
```

```
## [1] 36  3
```

```
# rename the variables
names(HW10)<- c("offer", "age", "dealer")

HW10$age<- as.factor(HW10$age)

str(HW10)
```

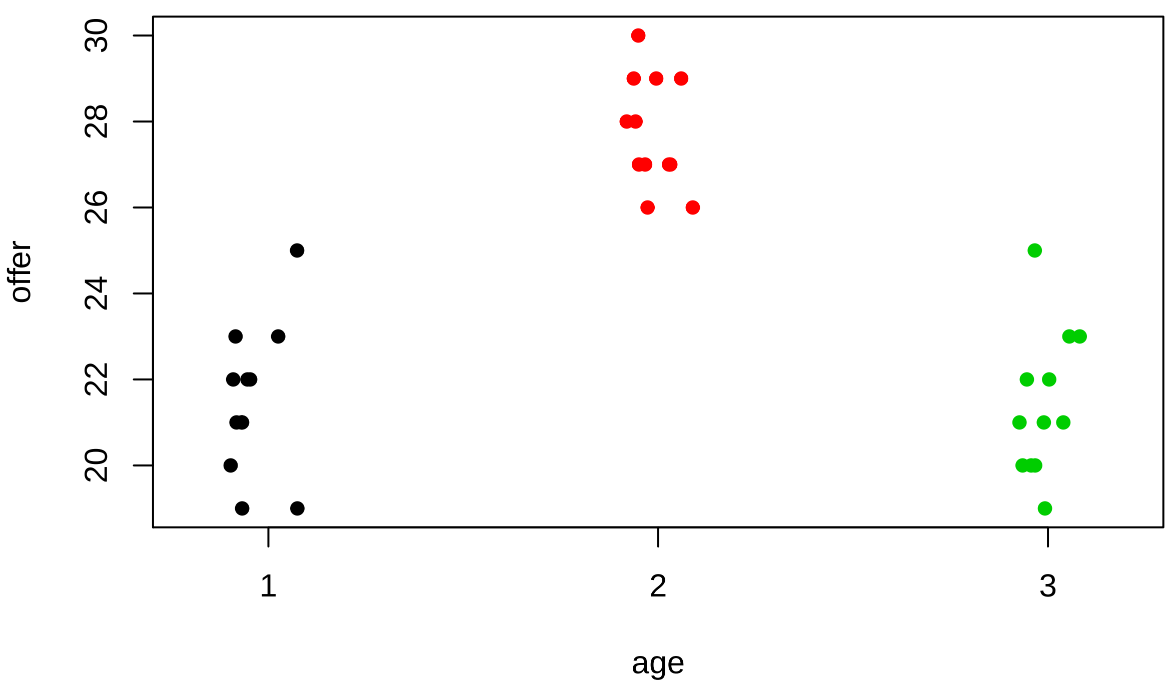
```
## 'data.frame':    36 obs. of  3 variables:
##  $ offer : num  23 25 21 22 21 22 20 23 19 22 ...
##  $ age   : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ dealer: int  1 2 3 4 5 6 7 8 9 10 ...
```

```r
head(HW10)
```

```
##   offer age dealer
## 1    23   1      1
## 2    25   1      2
## 3    21   1      3
## 4    22   1      4
## 5    21   1      5
## 6    22   1      6
```

**a. aligned dot plots:**

```r
# add jitter since some values are the same
stripchart(offer ~ age, vertical = TRUE,  data = HW10, xlab="age", method = "jitter", pch=16, col=1:3)
```




**Results: the means appear to differ, and the varibilities appear to be similar.**

**b. Obtain the fitted values:**

```r
fit <- aov(offer ~ age, data = HW10)

# factor level means= fitted values
predict(fit, newdata = data.frame(age = factor(1:3)))
```

```
##        1        2        3
## 21.50000 27.75000 21.41667
```

**c. Obtain the residuals :**

```r
# residials
fit$residuals
```

```
##          1          2          3          4          5          6
##  1.5000000  3.5000000 -0.5000000  0.5000000 -0.5000000  0.5000000
##          7          8          9         10         11         12
## -1.5000000  1.5000000 -2.5000000  0.5000000 -2.5000000 -0.5000000
##         13         14         15         16         17         18
##  0.2500000 -0.7500000 -0.7500000  1.2500000 -1.7500000  1.2500000
##         19         20         21         22         23         24
## -0.7500000  2.2500000  0.2500000 -0.7500000 -1.7500000  1.2500000
##         25         26         27         28         29         30
##  1.5833333 -1.4166667  3.5833333 -0.4166667  0.5833333  1.5833333
##         31         32         33         34         35         36
## -0.4166667 -1.4166667 -2.4166667 -1.4166667  0.5833333 -0.4166667
```

**d-f.**

```r
summary(fit)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## age          2  316.7  158.36    63.6 4.77e-12 ***
## Residuals   33   82.2    2.49                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_a : not\ all \quad \mu_i \quad are\ equal.$$

**Results: $F^*$= 63.6 with a p-value <0.001, we reject $H_0$ and conclude $H_a$. The mean cash offers are significantly different among the different age groups.**

## HW 16.25

```
mu <- c(7,8,9)
sigma2<- 0.9^2
power.anova.test(groups = length(mu), n = 9, between.var = var(mu), within.var = sigma2)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 3
##               n = 9
##     between.var = 1
##      within.var = 0.81
##       sig.level = 0.05
##           power = 0.9830563
##
## NOTE: n is number in each group
```

**Results: The power of the test is 0.983.**

## HW 16.27

```
mu <- c(22,28,22)
sigma2<- 1.6^2
power.anova.test(groups = length(mu), n = 12, between.var = var(mu), within.var = sigma2)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 3
##               n = 12
##     between.var = 12
##      within.var = 2.56
##       sig.level = 0.05
##           power = 1
##
## NOTE: n is number in each group
```

**Results: The power of the test is about 1.**

## HW 16.29

**We can define a power function that changes with Delta and Alpha**

```
Power_Fun <- function( Delta=10 , Alpha=0.05){
    sigma = 10
    r = 5
    power.anova.test(groups = r,  between.var =  Delta^2/(2*(r-1)),
                     within.var = sigma^2, power = 0.95,  sig.level = Alpha)
 }
```

**(a) Call the function for Delta= 10, 15,20,30 with with Alpha = 0.01**

```r
Power_Fun(Delta=10, Alpha=0.01)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 5
##               n = 49.99829
##     between.var = 12.5
##      within.var = 100
##       sig.level = 0.01
##           power = 0.95
##
## NOTE: n is number in each group
```

```r
Power_Fun(Delta=15, Alpha=0.01)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 5
##               n = 22.98059
##     between.var = 28.125
##      within.var = 100
##       sig.level = 0.01
##           power = 0.95
##
## NOTE: n is number in each group
```

```r
Power_Fun(Delta=20, Alpha=0.01)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 5
##               n = 13.53809
##     between.var = 50
##      within.var = 100
##       sig.level = 0.01
##           power = 0.95
##
## NOTE: n is number in each group
```

```r
Power_Fun(Delta=30, Alpha=0.01)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 5
```

```
##                n = 6.824308
##      between.var = 112.5
##       within.var = 100
##        sig.level = 0.01
##            power = 0.95
##
## NOTE: n is number in each group
```

**Results: For $\delta$ = 10, 15, 20, 30 with $\alpha$ = 0.01, the required sample sizes are 50, 23, 14, and 7, respectively, It suggested the sample size is decreasing with delta.**

**(b) Call the function for Delta= 10, 15,20,30 with Alpha = 0.05 (default in the function)**

```r
Power_Fun(Delta=10)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##           groups = 5
##                n = 38.10632
##      between.var = 12.5
##       within.var = 100
##        sig.level = 0.05
##            power = 0.95
##
## NOTE: n is number in each group
```

```r
Power_Fun(Delta=15)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##           groups = 5
##                n = 17.4883
##      between.var = 28.125
##       within.var = 100
##        sig.level = 0.05
##            power = 0.95
##
## NOTE: n is number in each group
```

```r
Power_Fun(Delta=20)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##           groups = 5
##                n = 10.28883
##      between.var = 50
##       within.var = 100
```

```
##       sig.level = 0.05
##           power = 0.95
##
## NOTE: n is number in each group
```

```r
Power_Fun(Delta=30)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 5
##               n = 5.187295
##     between.var = 112.5
##      within.var = 100
##       sig.level = 0.05
##           power = 0.95
##
## NOTE: n is number in each group
```

**Results: For $\delta$ = 10, 15,20,30 with $\alpha$ = 0.05, the required sample sizes are 39, 18, 11, and 6, respectively, It suggested that the sample size is increasing with smaller $\alpha$ (type I error) if all the other assumptions are the same.**