HW#2 Solution

9/19/2019

HW 17.8 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)
```

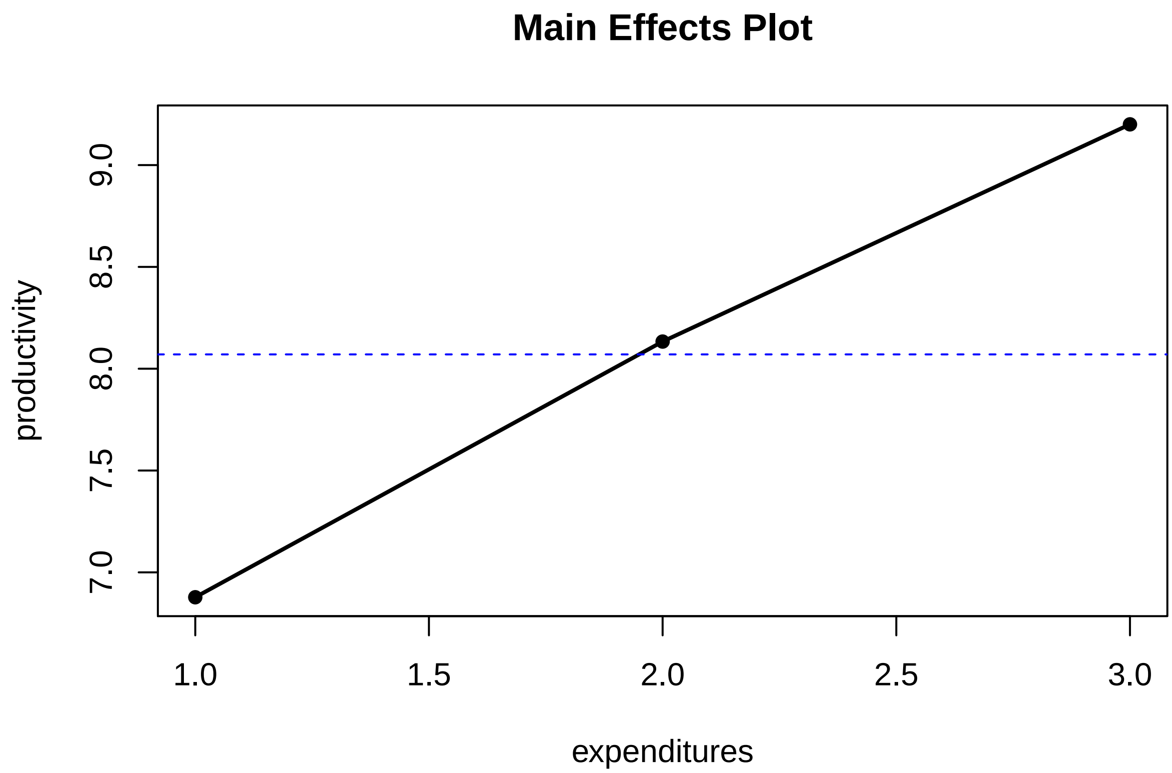
```
## 'data.frame':    27 obs. of  3 variables:
##  $ productivity: num  7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6 6.7 ...
##  $ expenditures: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 2 ...
##  $ firm        : int  1 2 3 4 5 6 7 8 9 1 ...
```

a. Line plot

```r
fit <- aov(productivity ~ expenditures, data = HW07)
(Factor_means = predict(fit, newdata = data.frame(expenditures = factor(1:3))))
```

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

```r
Overall_means = mean(Factor_means)
plot(1:3, Factor_means, type='o', pch=16, lwd=2,
     xlab="expenditures", ylab="productivity", main="Main Effects Plot")
abline(h= Overall_means, lty=2, col='blue')
```



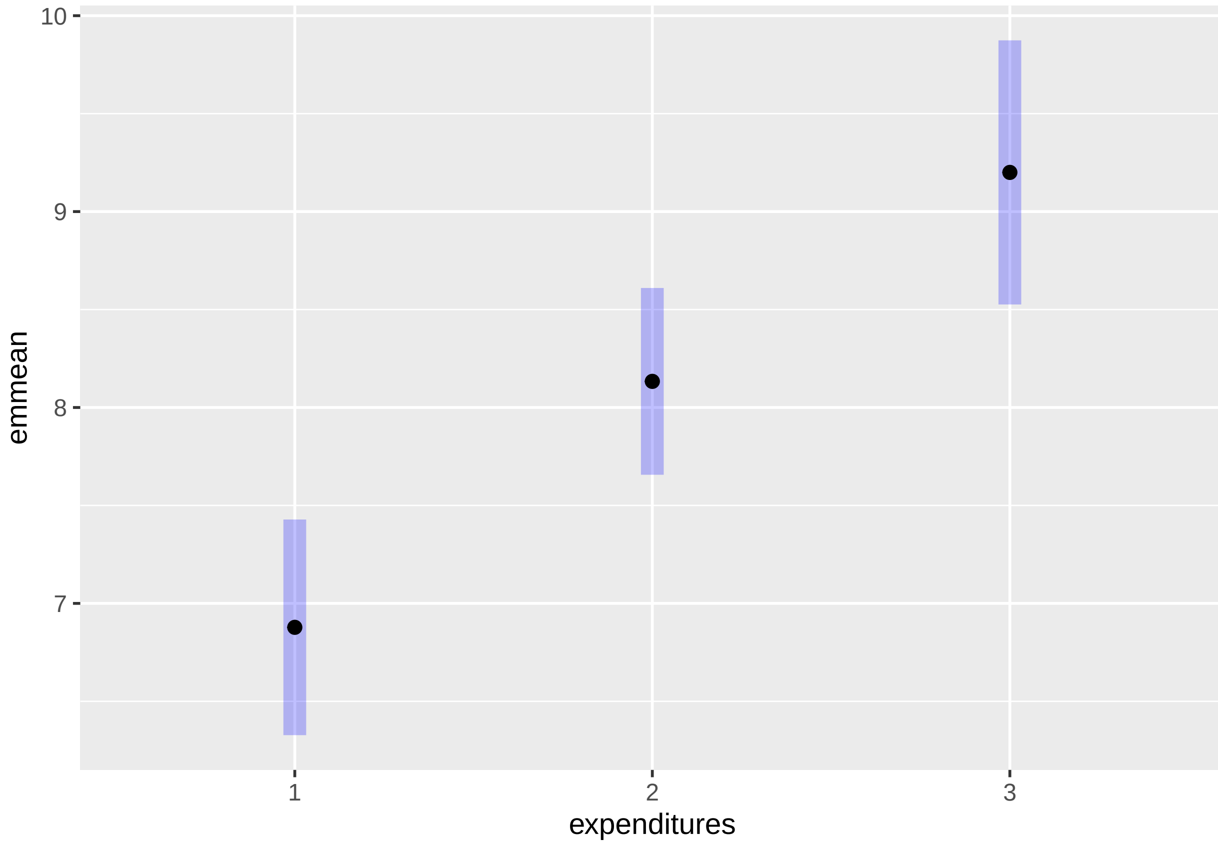
Results: It suggests productivity is increasing with expenditures.

b. Estimate mean and 95% CI for μ_3

```r
library(emmeans)
(Est.mean<- emmeans(fit, ~ expenditures))
```

```
##  expenditures emmean    SE df lower.CL upper.CL
##  1              6.88 0.267 24     6.33     7.43
##  2              8.13 0.231 24     7.66     8.61
##  3              9.20 0.327 24     8.53     9.87
##
## Confidence level used: 0.95
```

```r
plot(Est.mean, horizontal=F)
```



Results: The mean is 9.2 with 95% CI (8.53-9.87).

c. Obtain the 95% CI for mu2- mu1

```r
pairs(Est.mean, adjust = "none" )
```

```
##  contrast estimate    SE df t.ratio p.value
##  1 - 2       -1.26 0.353 24 -3.559  0.0016
##  1 - 3       -2.32 0.422 24 -5.507  <.0001
##  2 - 3       -1.07 0.400 24 -2.666  0.0135
```

Results: We can first obtain estimate is -(-1.26)=1.26, with SE =0.53, df=24. Then we get CI as follows:

```r
LCI=  1.26 - qt(.975, 24)*0.353
UCI=  1.26 + qt(.975, 24)*0.353
paste("95% CI is (", round(LCI,2), round(UCI,2), ").")
```

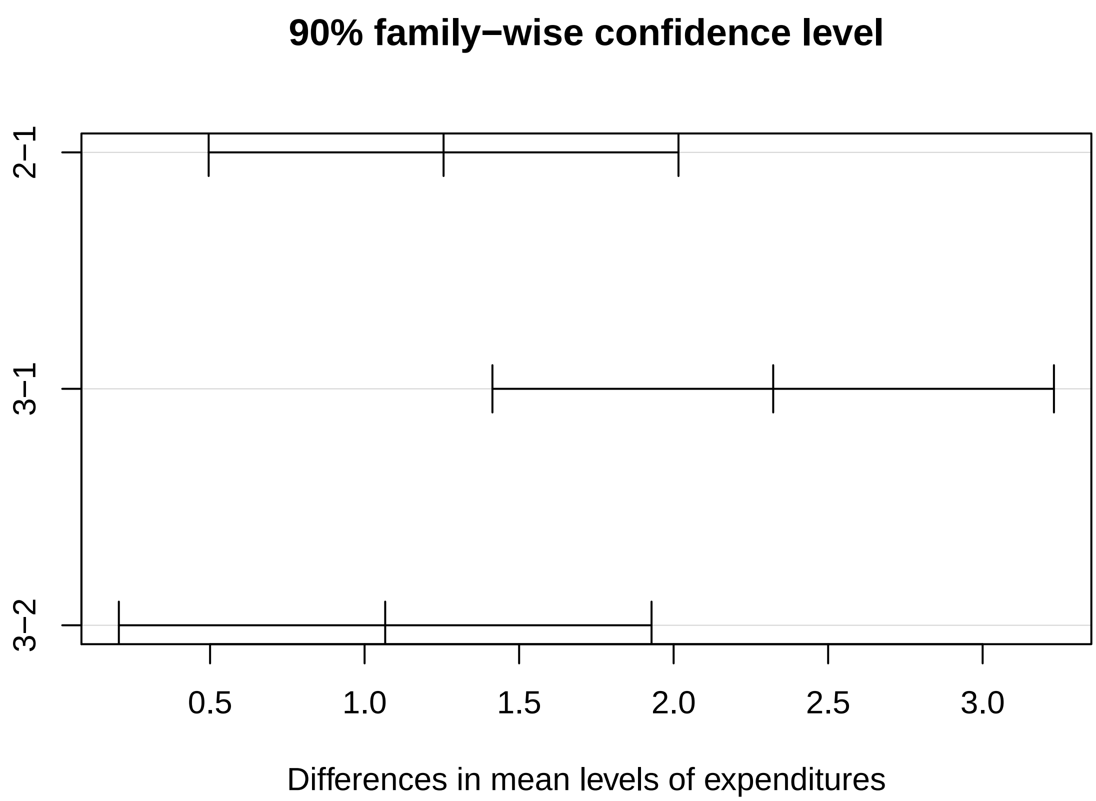
```
## [1] "95% CI is ( 0.53 1.99 )."
```

d. Tukey

```r
TukeyHSD(fit, conf.level = 0.90)
```

```
##   Tukey multiple comparisons of means
##     90% family-wise confidence level
##
## Fit: aov(formula = productivity ~ expenditures, data = HW07)
##
## $expenditures
##         diff       lwr      upr     p adj
## 2-1 1.255556 0.4954165 2.015695 0.0043755
## 3-1 2.322222 1.4136823 3.230762 0.0000335
## 3-2 1.066667 0.2047500 1.928583 0.0347870
```

```r
plot(TukeyHSD(fit, conf.level = 0.90))
```



e. compute the three mutiples

```r
#Tukey multiple
1/sqrt(2)* qtukey(.90, nm=3, df=24)
```

```
## [1] 2.154636
```

```r
#Bonferroni multiple,
g=3 #
qt(1- 0.1/(2*g), 24)
```

```
## [1] 2.25775
```

```r
# Scheffe multiple r-1=2, n_T-r=24
sqrt(2*qf(.90, 2, 24))
```

```
## [1] 2.253145
```

Results: Yes. Tukey multiple =2.15, which is less than Bonferroni multiple=2.257 and Scheffe multiple =2.253.

HW 17.11 Cash offers data

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

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

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

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

str(HW10)
```

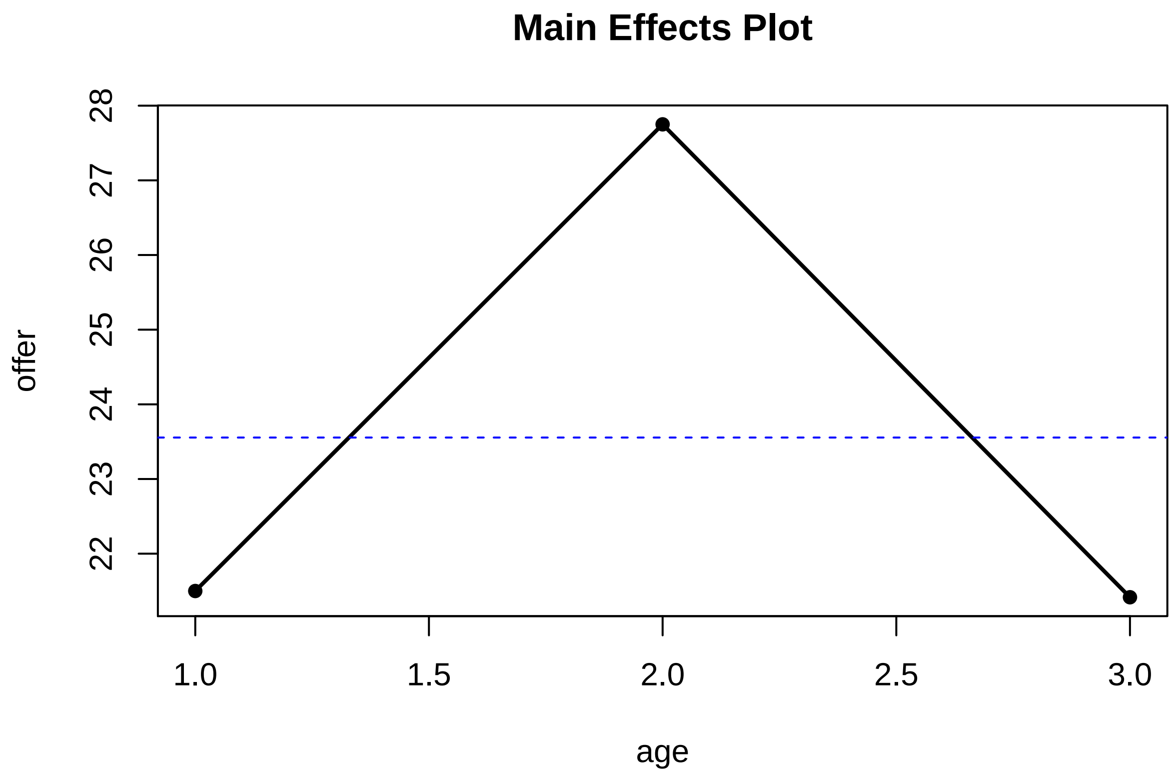
```
## 'data.frame':    36 obs. of  3 variables:
##  $ offer : num  23 25 21 22 21 22 20 23 19 22 ...
##  $ age   : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ dealer: int  1 2 3 4 5 6 7 8 9 10 ...
```

a. main effect plots

```r
fit <- aov(offer ~ age, data = HW10)
(Factor_means = predict(fit, newdata = data.frame(age = factor(1:3))))
```

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

```r
Overall_means = mean(Factor_means)
plot(1:3, Factor_means, type='o', pch=16, lwd=2,
     xlab="age", ylab="offer", main="Main Effects Plot")
abline(h= Overall_means, lty=2, col='blue')
```



Results: It suggests a nonlinear relationship. The middle age group seems to get a larger case offer..

b. Obtain 99% CI for younger age (mu1)

```
Est.mean<- emmeans(fit, ~ age)
confint(Est.mean, adjust="none", level=.99)
```

```
##  age emmean    SE df lower.CL upper.CL
##  1     21.5 0.456 33     20.3     22.7
##  2     27.8 0.456 33     26.5     29.0
##  3     21.4 0.456 33     20.2     22.7
##
## Confidence level used: 0.99
```

Results: For younger age , estimate=21.5, 99% CI is (20.3. 22.7)

c. Get 99% CI for mu3-mu1

```r
pairs(Est.mean, adjust = "none" )
```

```
##  contrast estimate    SE df t.ratio p.value
##  1 - 2     -6.2500 0.644 33 -9.702  <.0001
##  1 - 3      0.0833 0.644 33  0.129  0.8979
##  2 - 3      6.3333 0.644 33  9.831  <.0001
```

Results: We can first obtain estimate is D1=mu3-mu1= -0.0833, SE=0.644; df=33

```r
# For D1
D1 = -0.0833
SE = 0.644
LCI=  D1 - qt(1-0.01/2, 33)* 0.644
UCI=  D1 + qt(1-0.01/2, 33)* 0.644
paste("99% CI is D1 ,  (", round(LCI,2), round(UCI,2), ").")
```

```
## [1] "99% CI is D1 ,  ( -1.84 1.68 )."
```

d. test if a contrast =0,

$$H_0: \quad L = \mu_1 - 2\mu_2 + \mu_3 = 0$$

$$H_0: \quad L = \mu_1 - 2\mu_2 + \mu_3 \neq 0$$

Rejection rule, $t^* = \hat{L}/s(\hat{L})$, and we reject null if $|t^*| > t(1-\alpha/2, n_T - r)$.

```r
L = list(L1= c(1,-2,1))
contrast(Est.mean, L, adjust="none")
```

```
##  contrast estimate   SE df t.ratio p.value
##  L1          -12.6 1.12 33 -11.278 <.0001
```

Results: we reject the null hypothesis that this contrast is zero, or equivalently,

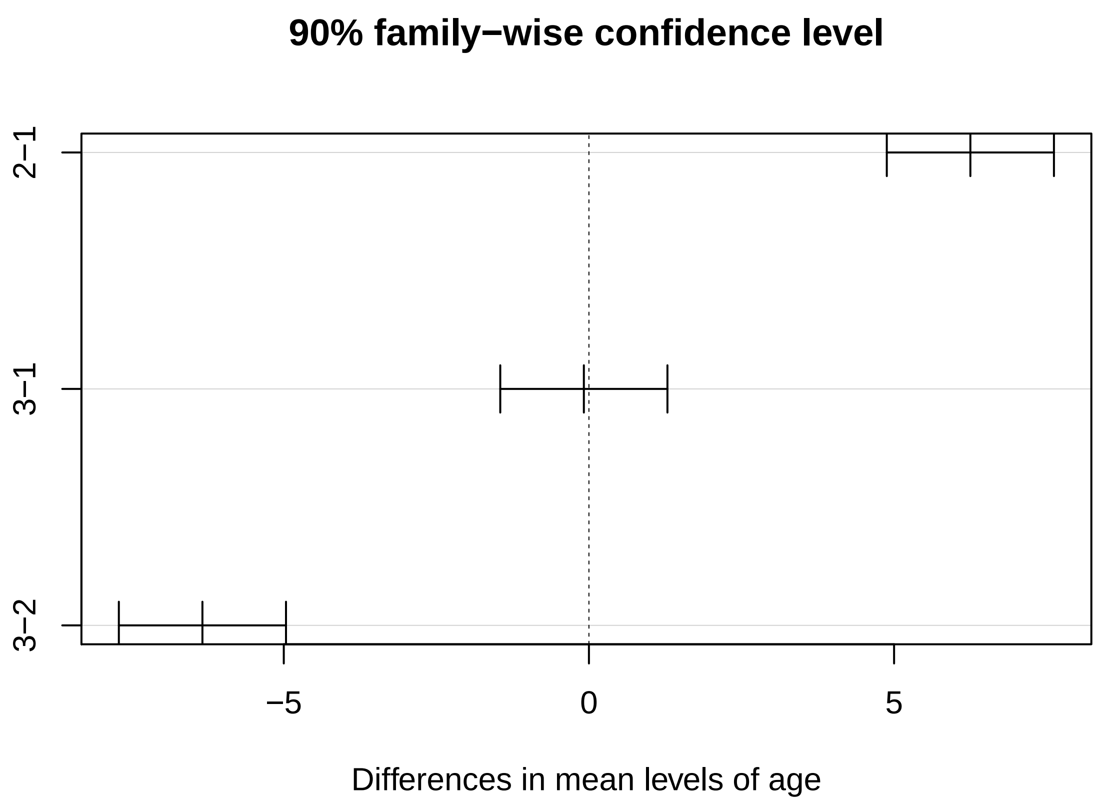
$\mu_2 - \mu_1 = \mu_3 - \mu_2$.

e. Tukey

```r
TukeyHSD(fit, conf.level = 0.90)
```

```
##   Tukey multiple comparisons of means
##     90% family-wise confidence level
##
## Fit: aov(formula = offer ~ age, data = HW10)
##
## $age
##            diff       lwr       upr     p adj
## 2-1  6.25000000  4.880508  7.619492 0.0000000
## 3-1 -0.08333333 -1.452825  1.286158 0.9908192
## 3-2 -6.33333333 -7.702825 -4.963842 0.0000000
```

```r
plot(TukeyHSD(fit, conf.level = 0.90))
```



f. compute the 2 mutiples

```r
#Tukey multiple
1/sqrt(2)* qtukey(.90, nm=3, df=33)
```

```
## [1] 2.125907
```

```r
#Bonferroni multiple,
g=3 #
qt(1- 0.1/(2*g), 33)
```

```
## [1] 2.220913
```

Results: No. Tukey multiple =2.13, which is less than Bonferroni multiple=2.22.