HW#5 Solution (week6 HW)

10/10/2019

HW 19.10 Cash offers

```
# HW 19.10
HW19  <- read.table(
   url("https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week6/CH19PR10.txt"))
# rename the variables
names(HW19)<- c("response", "age", "gender", "units")

HW19$age <- as.factor(HW19$age)
HW19$gender <- as.factor(HW19$gender)

# relabel it
levels(HW19$age) <- c("Young", "Middle", "Elderly")
levels(HW19$gender) <- c("Male", "Female")

head(HW19,8)
```

```
##   response   age gender units
## 1       21 Young   Male     1
## 2       23 Young   Male     2
## 3       19 Young   Male     3
## 4       22 Young   Male     4
## 5       22 Young   Male     5
## 6       23 Young   Male     6
## 7       21 Young Female     1
## 8       22 Young Female     2
```

```
str(HW19)
```

```
## 'data.frame':    36 obs. of  4 variables:
##  $ response: num  21 23 19 22 22 23 21 22 20 21 ...
##  $ age     : Factor w/ 3 levels "Young","Middle",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ gender  : Factor w/ 2 levels "Male","Female": 1 1 1 1 1 1 2 2 2 2 ...
##  $ units   : int  1 2 3 4 5 6 1 2 3 4 ...
```

a. fit two-way ANOVA analysis, get fitted value and residulas

```
fit  =  aov(response ~ age*gender, data = HW19)

# generate the combination of the factor levels= treatment (i,j)
Newdata = data.frame(age= rep(c("Young", "Middle", "Elderly"),c(2,2,2)),
                      gender= rep(c("Male", "Female"), 3))

(fitted = data.frame(Newdata, fitted.value=predict(fit,newdata= Newdata ) ))
```

```
##       age gender fitted.value
## 1   Young   Male     21.66667
## 2   Young Female     21.33333
## 3  Middle   Male     27.83333
## 4  Middle Female     27.66667
## 5 Elderly   Male     22.33333
## 6 Elderly Female     20.50000
```

b. Residuals and sum by treatments

```r
# There are six treatment, make residuals a 6*6 matrix
(ResbyTreatment <- matrix(fit$residuals, nrow=6, byrow=F ))
```

```
##            [,1]       [,2]       [,3]       [,4]       [,5] [,6]
## [1,] -0.6666667 -0.3333333  2.1666667 -1.6666667  2.6666667  2.5
## [2,]  1.3333333  0.6666667  1.1666667  1.3333333 -0.3333333 -1.5
## [3,] -2.6666667 -1.3333333 -1.8333333 -0.6666667  0.6666667 -0.5
## [4,]  0.3333333 -0.3333333  0.1666667  0.3333333 -1.3333333  0.5
## [5,]  0.3333333 -2.3333333 -0.8333333 -0.6666667 -0.3333333 -0.5
## [6,]  1.3333333  3.6666667 -0.8333333  1.3333333 -1.3333333 -0.5
```

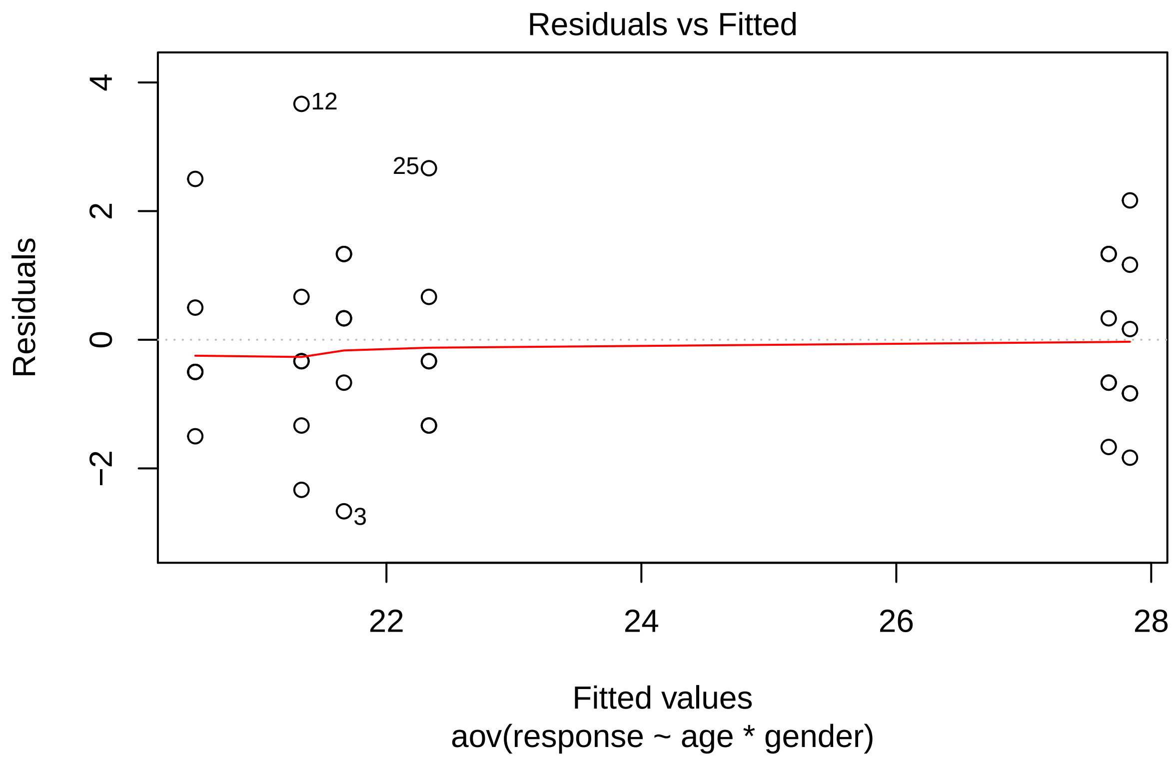
```r
# check if zero for each treatment,
apply(ResbyTreatment,  2, sum)
```

```
## [1] -2.553513e-15  3.996803e-15  2.914335e-15 -2.886580e-15 -4.440892e-16
## [6] -7.771561e-16
```

Results: The sum of residuals for each treatment are all zero.

c. Residual plot by treatment plot

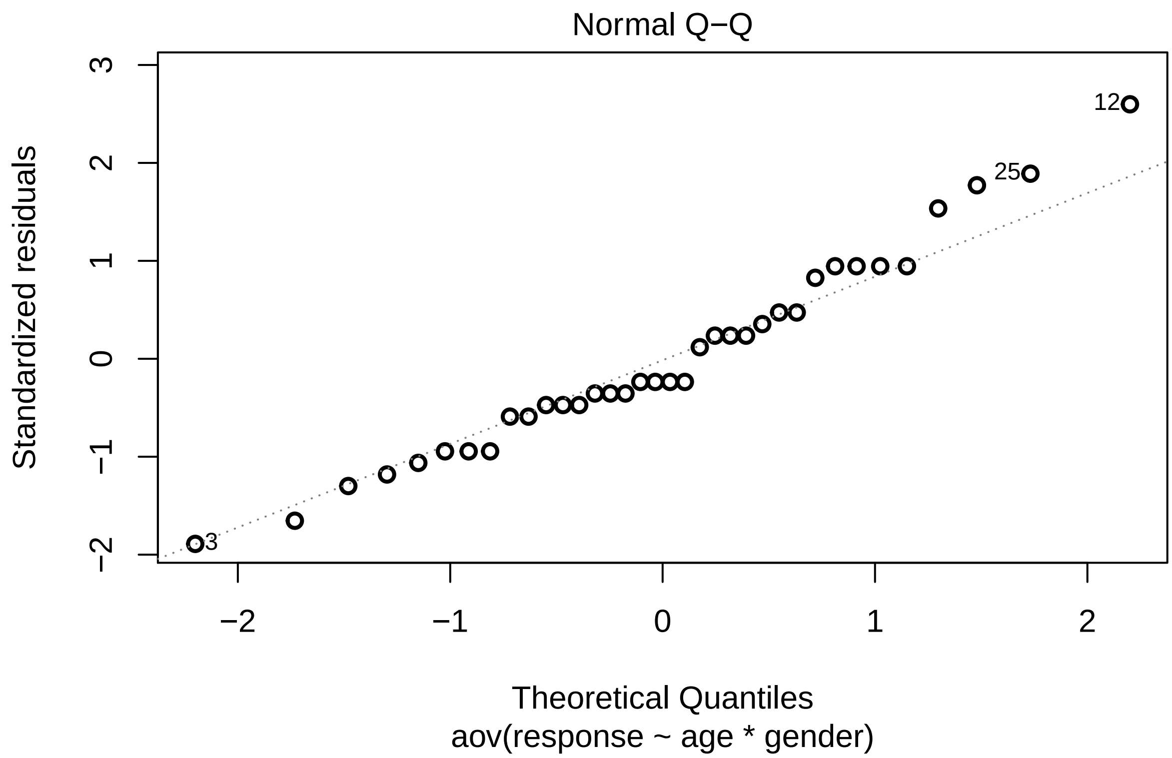
```r
plot(fit,1)
```



Results: The variances seem similar for different treatment levels.

d) Residual normal QQ plot

```
plot(fit,2, lwd=2)
```

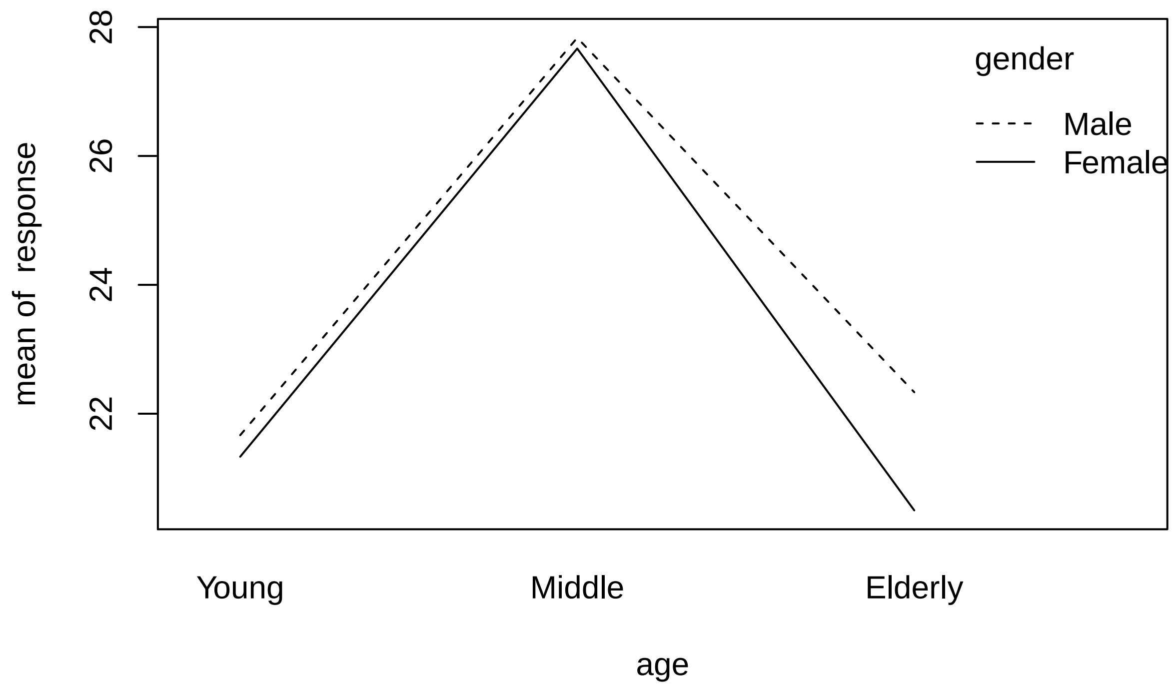


Results: The distribution of residuals does not appear to have a large deviation from the normal assumption.

HW 19.11 Cash offers

a) Estimated treatment means plot

```
with(HW19, interaction.plot(x.factor = age, trace.factor = gender, response = response))
```



Results: There appears to be an age effect with higher offers for middle age group; no gender effect was present.

b) ANOVA analysis

```
summary(fit)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## age          2  316.7  158.36  66.291 9.79e-12 ***
## gender       1    5.4    5.44   2.279    0.142
## age:gender   2    5.1    2.53   1.058    0.360
## Residuals   30   71.7    2.39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: The age effect account for most of total variability,

c-d, f)

- Testing interaction

H_0: all $(\alpha\beta)_{ij}$ equal zero

H_a: not all $(\alpha\beta)_{ij}$ equal zero.

$F = 1.058$ with a p-value =0.36, so we don't reject the null that there was no interaction.

- Testing for main effect
- Age factor:

H_0: all α_i equal zero

H_a: not all α_i equal zero.

$F = 66.291$ with a p-value <0.001, so we reject the null and conclude there is a significant age effect on cash offers.

- Gender factor:

H_0: all β_j equal zero

H_a: not all β_j equal zero.

$F = 2.279$ with a p-value= 0.142, so we don't reject the null that there was no gender effect.

Also, since the interaction is not important, it is meaningful to test for main effects. These results agree with the results from plot in (a)

(e) Upper bound

Kimball inequality yields the bound for the family level of significance:

$$\alpha \leq 1 - (1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_3) = 1(.95)(.95)(.95) = .143$$

(h)

```
# one way ANOVA
summary(aov(response ~ age, data=HW19))
```

```
##             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
```

```
summary(aov(response ~ age*gender, data=HW19))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## age          2  316.7  158.36  66.291 9.79e-12 ***
## gender       1    5.4    5.44   2.279    0.142
## age:gender   2    5.1    2.53   1.058    0.360
## Residuals   30   71.7    2.39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: Relationship:

SS(Age) in one-factor ANOVA= SS(Age) in two-factor ANOVA SSE in one-factor ANOVA= SSE+ SS(Gender) + SS(Age and Gender Interaction) in two-factor ANOVA. df for SSE in one-factor ANOVA = 33= 1+2+30 (the sum of three corresponding dfs) in two-factor ANOVA.

HW 19.30

a) CI for μ_{11} 95% CI

We can obtain this using "sample mean +/- t(.0975, df=30)*sqrt(MSE/6), or we can use emmeans function.

```
qt(.975, 30)
```

```
## [1] 2.042272
```

```
library(emmeans)
# get the estimate, SE, df and CI
(Est.mean = emmeans(fit, ~ age*gender))
```

```
##  age     gender emmean    SE df lower.CL upper.CL
##  Young   Male     21.7 0.631 30     20.4     23.0
##  Middle  Male     27.8 0.631 30     26.5     29.1
##  Elderly Male     22.3 0.631 30     21.0     23.6
##  Young   Female   21.3 0.631 30     20.0     22.6
##  Middle  Female   27.7 0.631 30     26.4     29.0
##  Elderly Female   20.5 0.631 30     19.2     21.8
##
## Confidence level used: 0.95
```

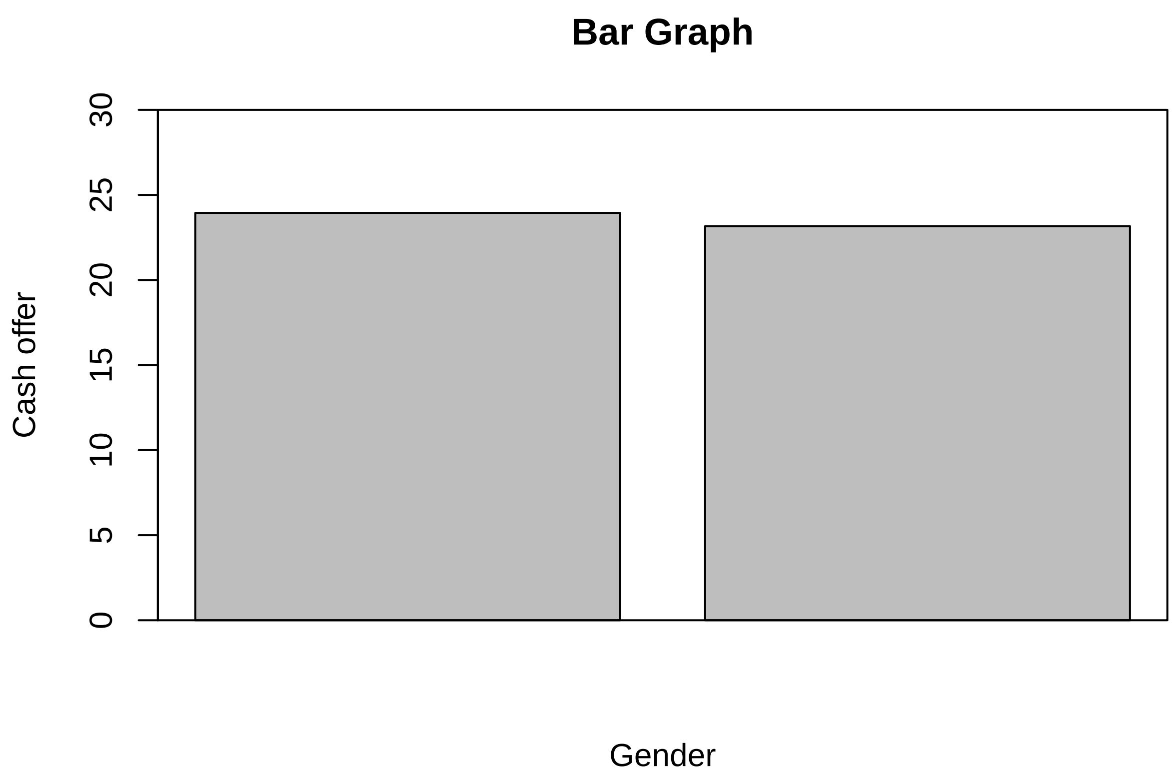
Results: The 95% CI for μ_{11} is [20.4 , 23.0] calculated from formula 21.7 +/- 2.042*0.631, or using the first row of the output for (Young , Male). If we repeat the experiment many times, we have 95% confidence that the mean cash offers for a young and male owner will fall in the interval [20.4 , 23.0].

b) bar graph of B-level means

```
FactorB.mean = with(HW19, by(response, gender, mean ))
(FactorB.mean = as.numeric(FactorB.mean))
```

```
## [1] 23.94444 23.16667
```

```
barplot(FactorB.mean, col='gray', xlab="Gender", ylab="Cash offer", main="Bar Graph", ylim=c(0,30))
box()
```



Results: The responses for male and female groups appear to be similar.

c). Get Estimate $\mu_{.1} - \mu_{.2}$ with 95% CI

method 1: use R functions

```
(Est.Bmean = emmeans(fit, ~ gender))
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
##  gender emmean    SE df lower.CL upper.CL
##  Male     23.9 0.364 30     23.2     24.7
##  Female   23.2 0.364 30     22.4     23.9
##
## Results are averaged over the levels of: age
## Confidence level used: 0.95
```

```
confint(pairs(Est.Bmean),  adjust = "none")
```

```
##  contrast      estimate    SE df lower.CL upper.CL
##  Male - Female    0.778 0.515 30   -0.274     1.83
##
## Results are averaged over the levels of: age
## Confidence level used: 0.95
```

method 2: use definitions

```r
(Mean.D = FactorB.mean[1]- FactorB.mean[2])
```

```
## [1] 0.7777778
```

```r
# MSE*2/(an)
(SE.D =  sqrt( 2.39*2/(3*6)))
```

```
## [1] 0.5153208
```

```r
Lower.limt =  Mean.D-  qt(.975, 30)* SE.D
Upper.limt =  Mean.D+ qt(.975, 30)* SE.D
paste("The 95% CI is [", round(Lower.limt, 3), "-", round(Upper.limt, 3),"].")
```

```
## [1] "The 95% CI is [ -0.275 - 1.83 ]."
```

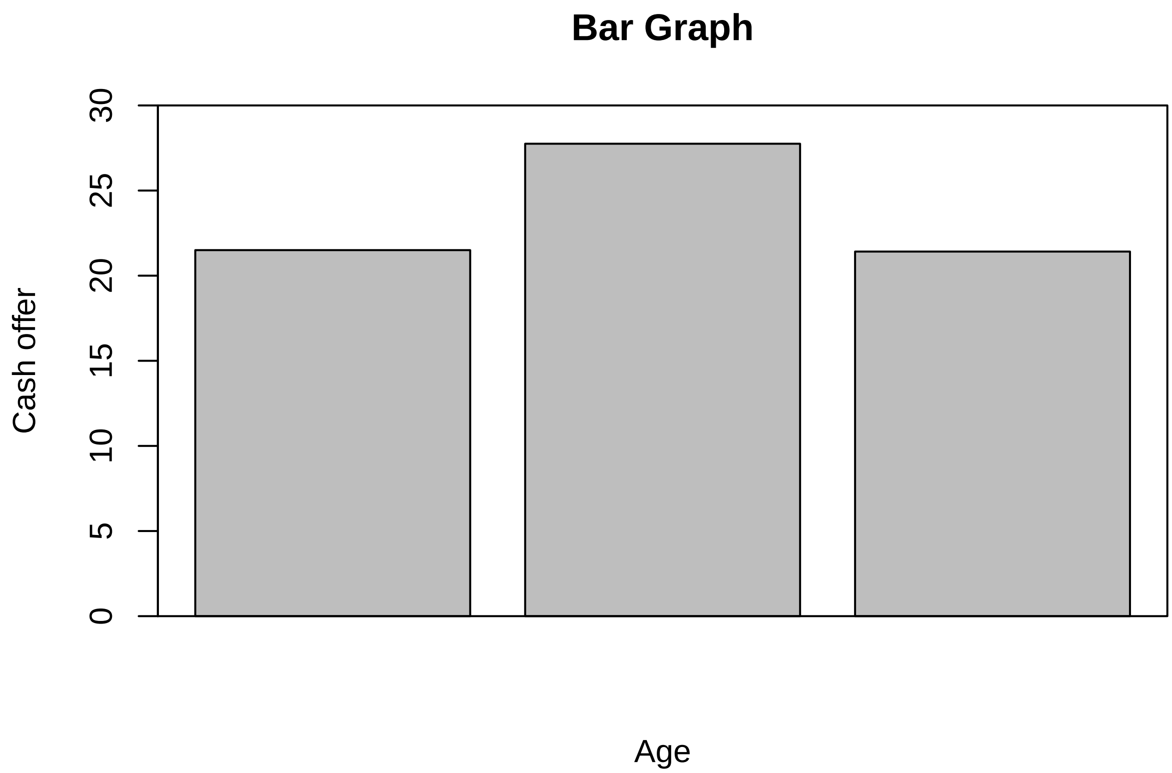
Results: The 95% CI is (-0.274, 1.83), suggesting no difference in the response between male and female. This is consistent with 19.11d and part (b).

d) bargraph for A

```r
FactorA.mean = with(HW19, by(response, age, mean ))
(FactorA.mean = as.numeric(FactorA.mean))
```

```
## [1] 21.50000 27.75000 21.41667
```

```r
barplot(FactorA.mean, col='gray', xlab="Age", ylab="Cash offer", main="Bar Graph", ylim=c(0,30))
box()
```



Results: The responses for 3 age groups appear to be different.

e) pairwise difference

```
Est.Amean = emmeans(fit, ~ age)
```

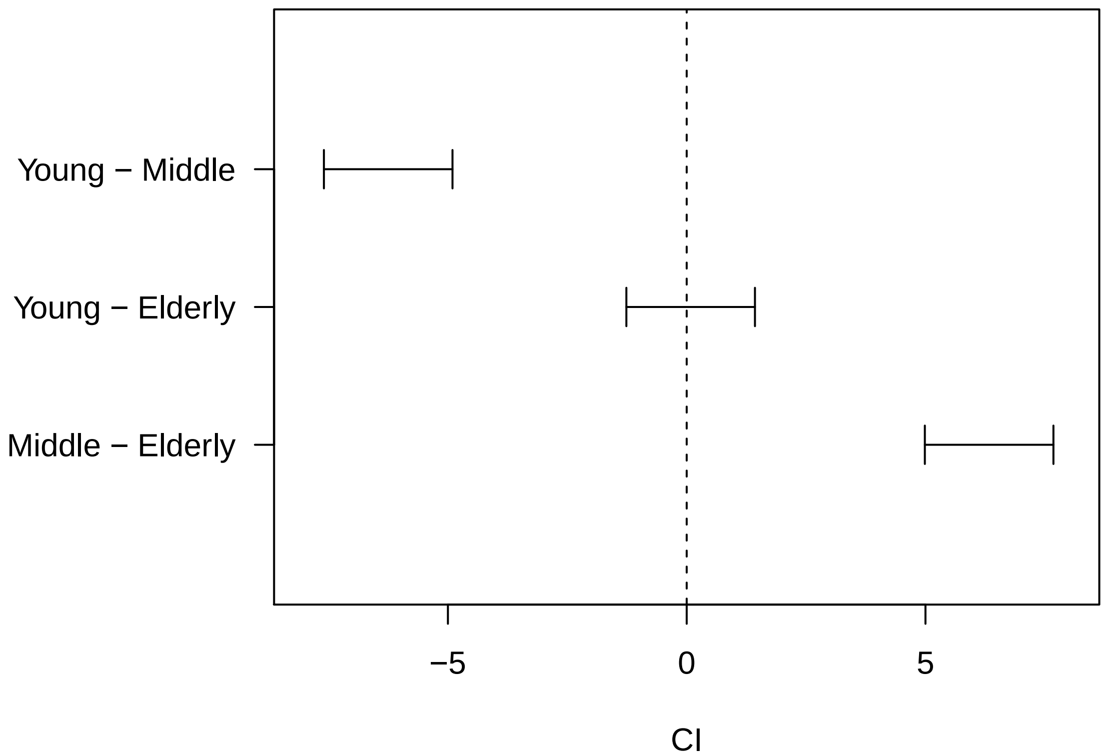
```
## NOTE: Results may be misleading due to involvement in interactions
```

```
(CIage <- confint(pairs(Est.Amean),  adjust = "Tukey", level=.90))
```

```
##  contrast         estimate    SE df lower.CL upper.CL
##  Young - Middle    -6.2500 0.631 30    -7.60    -4.90
##  Young - Elderly    0.0833 0.631 30    -1.26     1.43
##  Middle - Elderly   6.3333 0.631 30     4.99     7.68
##
## Results are averaged over the levels of: gender
## Confidence level used: 0.9
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

```
# making a plot for the CI
par(mar=c(4,8,3,3))
plot( x=c(-8,8), y= c(0,4), type='n', axes=F, xlab="CI", ylab="")
box()
```

```r
axis(1)
axis(2, at=1:3,  labels= rev(CIage[,1]),las=2)
# adding CI using arrows(x1,y1, x2, y2)
arrows(CIage[1,5],3 , CIage[1,6],3,angle=90, code=3, length = 0.1 )
arrows(CIage[2,5],2 , CIage[2,6],2,angle=90, code=3, length = 0.1 )
arrows(CIage[3,5],1 , CIage[3,6],1,angle=90, code=3, length = 0.1 )
abline(v=0, lty=2)
```



Results: The findings suggest the middle age group have bigger cash offers than the other two groups.

f) Is Tukey most efficient

```r
#df= (n-1)ab =30

df= (6-1)*2*3
alpha=0.1

#Bonferroni multiple
g=3 #
qt(1- alpha/(2*g), df)
```

```
## [1] 2.230625
```

```r
# Scheffe multiple r=a, df
r= 3
sqrt((r-1)*qf(1-alpha, (r-1) , df))
```

```
## [1] 2.231016
```

```r
#Tukey multiple, df=114
1/sqrt(2)* qtukey(1-alpha, nm=r, df)
```

```
## [1] 2.133506
```

Results: Tukey's procedure is most efficient since the Tukey's multiple is smallest and the CI by Tukey method will be narrowest compared to Bonferroni or Scheffe method.

g) Contrast and CI

```r
List = list(L1= c(1/2,  -1, 1/2))
L= contrast(Est.Amean, List, adjust="none")
confint(L)
```

```
##  contrast estimate    SE df lower.CL upper.CL
##  L1          -6.29 0.546 30    -7.41    -5.18
##
## Results are averaged over the levels of: gender
## Confidence level used: 0.95
```

Results: The estimate $\hat{L} = -6.29$ with 95% CI [-7.41 , -5.18].

f) linear combination $L = 0.3\mu_{12} + 0.6\mu_{22} + 0.1\mu_{32}$ (treatment mean)

```r
Est.mean
```

```
##  age     gender emmean    SE df lower.CL upper.CL
##  Young   Male     21.7 0.631 30     20.4     23.0
##  Middle  Male     27.8 0.631 30     26.5     29.1
##  Elderly Male     22.3 0.631 30     21.0     23.6
##  Young   Female   21.3 0.631 30     20.0     22.6
##  Middle  Female   27.7 0.631 30     26.4     29.0
##  Elderly Female   20.5 0.631 30     19.2     21.8
##
## Confidence level used: 0.95
```

```r
# We get the linear combination with c1-c6 match the order
# of the six treatments
List = list(L2= c(0, 0, 0, .3,  .6, .1))
L= contrast(Est.mean, List, adjust="none")
confint(L)
```

```
##  contrast estimate    SE df lower.CL upper.CL
##  L2           25.1 0.428 30     24.2     25.9
##
## Confidence level used: 0.95
```

Results: The $\hat{L} = 25.1$ with 95% CI [24.2, 25.9].

HW 19.38

```
Delta=3
Alpha=0.05
sigma = 1.5
r = 3
power.anova.test(groups = r, between.var = Delta^2/(2*(r-1)),
within.var = sigma^2, power = 0.90, sig.level = Alpha)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 3
##               n = 7.431865
##     between.var = 2.25
##      within.var = 2.25
##       sig.level = 0.05
##           power = 0.9
##
## NOTE: n is number in each group
```

Results: For each age group, we need 8 owners with different gender. Since the study is balanced, then each age and gender group, we need 4 owners.