HW5

Ran Zhang

2/28/2019

Problem 1

Ho: mu(X1) = mu(X2) = mu(X3) Contrast matrix

```
C1 <- matrix(c(1,-1,0,1,0,-1),nrow=2,ncol=3,byrow=T)
```

Calculate the test Statistic

```
xbar1 <- c(46.1,57.3,50.4)
mu0_1 <- c(0,0,0)
q1 <- nrow(C1)
n1 <- 40
S1 <- matrix(c(101.3, 63.0, 71.0, 63.0, 80.2, 55.6, 71.0, 55.6, 97.4), nrow=3, ncol=3, b
yrow=TRUE)
T1 <- (n1*(n1-q1))/(q1*(n1-1))*t(C1%*%xbar1) %*% solve(C1%*%S1%*%(t(C1))) %*% (C1%*%xbar
1)
```

Critical value

```
alpha1 <- 0.05
qf(alpha1, q1, (n1-q1), lower.tail = F)
```

```
## [1] 3.244818
```

Comment: As T1 value is larger than the critical value, we reject H0, so the means of three indices are not equal.

Problem 2a

Build a different contrast matrix

```
C3 <- matrix(c(0,0,1,-1,0,1,0,-1,1,0,0,-1),nrow=3,ncol=4,byrow=T)
C3
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    0    0    1   -1
## [2,]    0    1    0   -1
## [3,]    1    0    0   -1
```

Problem 2b

Load the data and as data2b

```
data2b <- as.matrix(read.table("T6-2.dat"))
colnames(data2b) <- c("trt1","trt2", "trt3","trt4")
head(data2b)
```

```
##      trt1 trt2 trt3 trt4
## [1,]  426  609  556  600
## [2,]  253  236  392  395
## [3,]  359  433  349  357
## [4,]  432  431  522  600
## [5,]  405  426  513  513
## [6,]  324  438  507  539
```

Build a function as contrast

```
T2.contrast <- function(data.matrix, contrast.matrix, alpha=0.05){
  dat <- data.matrix
  C <- contrast.matrix
  # Sample mean and variance
  xbar <- colMeans(dat)
  S <- cov(dat)
  # sample size and observations
  n <- nrow(dat)
  q <- nrow(C)
  # Intermediate quantities
  invCSC <- solve(C%*%S %*% (t(C)))
  Cxbar <- C %*% xbar
  # Test statistic for this data
  T2 <- n*(n-q)/((n-1)*q)*(t(Cxbar)) %*% invCSC %*% (Cxbar)
  # Critical value
  critical_value_F <- qf(p=0.05, df1=q, df2=n-q, lower.tail = F)
  # p-value
  pv <- pf(T2, df1=q, df2 = n-q, lower.tail = F)
  # Display the results
  results <- data.frame(T2 = T2, critical = critical_value_F, df1 = q, df2 = n-q, pvalue
=pv)
  return(results)
}
```

Test the data with new contrast input from 2a

```
T2.contrast(data2b,C3)
```

```
##         T2 critical df1 df2       pvalue
## 1 34.37521 3.238872   3  16 3.317767e-07
```

Comment: The results are same as we got from the lectures.

Problem 2c

Test interactions Vectorize the data for the interaction plot

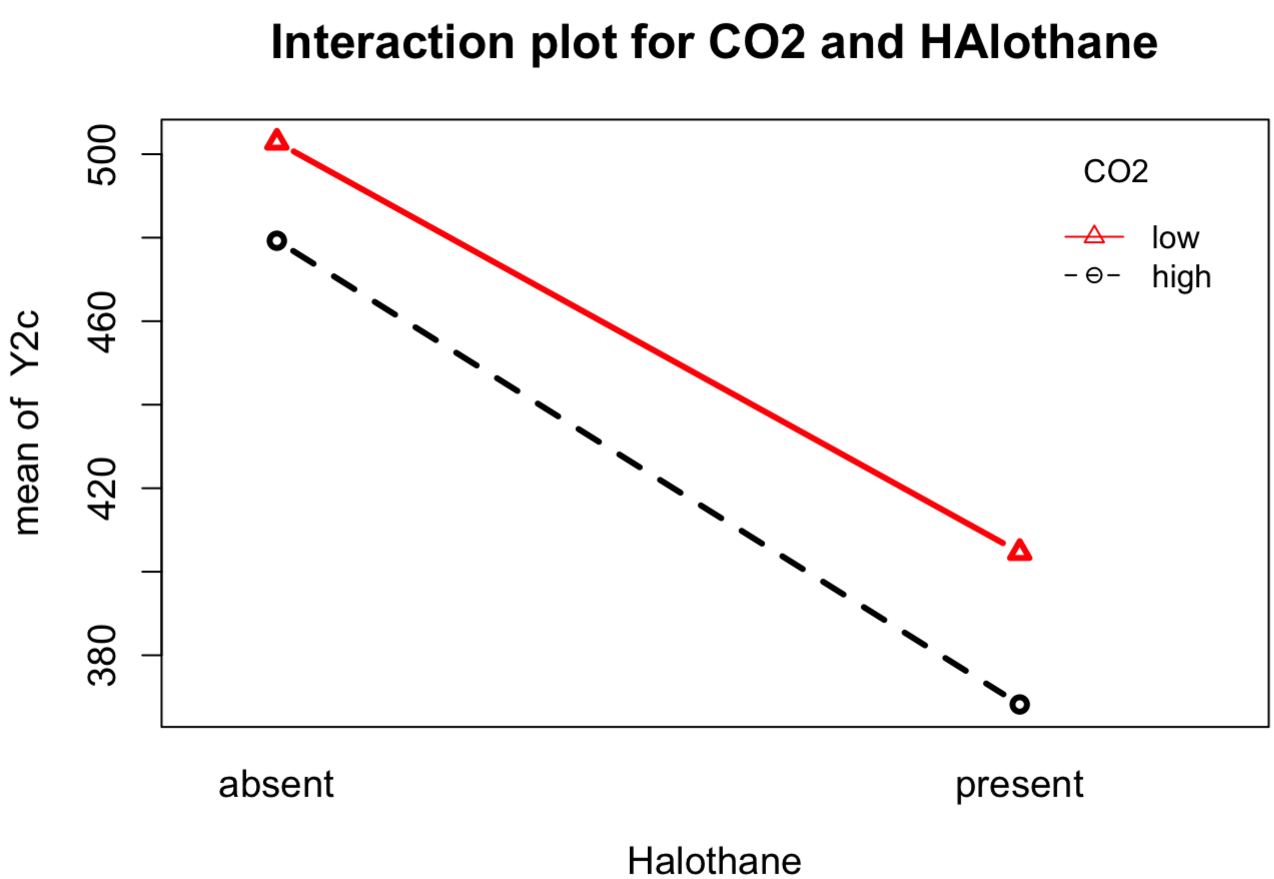
```
Y2c <- c(data2b)
```

Define the treatment combinations for different variables

```
CO2 <- c(rep("high",19), rep("low",19),rep("high",19),rep("low",19))
Halothane <- c(rep("present",19), rep("present",19), rep("absent",19), rep("absent",19))
```

Interaction plot by interaction.plot function

```
interaction.plot(Halothane,CO2,Y2c,lwd=3,col=c(1,2),cex.axis=1.2,cex.lab=1.2,main="Inter
action plot for CO2 and HAlothane", cex.main=1.5, type="b", pch=1:2)
```



Test main effects of halothane

```
C2c1 <- matrix(c(1,1,-1,-1),nrow=1,ncol=4,byrow=T)
T2.contrast(data2b,C2c1)
```

```
##         T2 critical df1 df2       pvalue
## 1 88.25581 4.413873   1  18 2.314867e-08
```

Test main effects of CO2

```
C2c2 <- matrix(c(1,-1,1,-1),nrow=1,ncol=4,byrow=T)
T2.contrast(data2b,C2c2)
```

```
##          T2 critical df1 df2      pvalue
## 1 13.18751 4.413873   1  18 0.001908795
```

Comment: There is no interaction effect, and the main effects of CO2 and Halothane are statistically significant as the p-value of them both smaller than 0.05.

Problem 2d

Ho: (miu1 - miu2) = 2*(miu3-miu4) same as u1-u2-2u3+2u4 = 0 The contrast matrix

```
C2d <- matrix(c(1,-1,-2,2),nrow=1,ncol=4,byrow=T)
T2.contrast(data2b,C2d)
```

```
##           T2 critical df1 df2    pvalue
## 1 0.1319143 4.413873   1  18 0.7206858
```

Comment: As the p-value is bigger than 0.05, we fail to reject the hypothesis test, so there is no significant evidence that the CO2 effect is twice when Halothane is present than Halothane is absent.

Problem 3a

Input the data

```
data3 <- read.table("T6-9.dat", header = F)
colnames(data3) <- c("Length","Width","Height","Gender")
str(data3)
```

```
## 'data.frame':    48 obs. of  4 variables:
##  $ Length: int  98 103 103 105 109 123 123 133 133 133 ...
##  $ Width : int  81 84 86 86 88 92 95 99 102 102 ...
##  $ Height: int  38 38 42 42 44 50 46 51 51 51 ...
##  $ Gender: Factor w/ 2 levels "female","male": 1 1 1 1 1 1 1 1 1 1 ...
```

Test for the equality of genders' means Display the sample means of two gender

```
data3.M <- data3[data3$Gender=="male", ]
data3.F <- data3[data3$Gender=="female", ]
Xbar.M <- colMeans(data3.M[,1:3])
Xbar.F <- colMeans(data3.F[,1:3])
```

Difference between the means

```
Xbar.diff <- Xbar.M - Xbar.F
rbind(Xbar.M,Xbar.F, Xbar.diff)
```

```
##               Length     Width    Height
## Xbar.M    113.37500  88.29167  40.70833
## Xbar.F    136.04167 102.58333  52.04167
## Xbar.diff -22.66667 -14.29167 -11.33333
```

Test for differences in mean for Length

```
test_length <- t.test(data3.M[,1], data3.F[,1],var.equal = T)
test_length
```

```
## 
##  Two Sample t-test
## 
## data:  data3.M[, 1] and data3.F[, 1]
## t = -4.5705, df = 46, p-value = 3.656e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -32.64934 -12.68399
## sample estimates:
## mean of x mean of y 
##  113.3750  136.0417
```

Test for differences in mean for Width

```
test_width <- t.test(data3.M[,2], data3.F[,2],var.equal = T)
test_width
```

```
## 
##  Two Sample t-test
## 
## data:  data3.M[, 2] and data3.F[, 2]
## t = -4.7015, df = 46, p-value = 2.376e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -20.410526  -8.172807
## sample estimates:
## mean of x mean of y 
##  88.29167 102.58333
```

Test for differences in mean for Height

```
test_height <- t.test(data3.M[,3], data3.F[,3],var.equal = T)
test_height
```

```
##
##  Two Sample t-test
##
## data:  data3.M[, 3] and data3.F[, 3]
## t = -6.3689, df = 46, p-value = 8.087e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.915222  -7.751445
## sample estimates:
## mean of x mean of y
##  40.70833  52.04167
```

Problem 3b Define significant level and number of variables

```
alpha3b <- 0.05
p3b <- 3
```

Covariance of each group

```
S.M <- cov(data3.M[,1:3])
S.F <- cov(data3.F[,1:3])
```

Sample size of each group

```
n.M <- nrow(data3.M)
n.F <- nrow(data3.F)
```

Pooled covariance

```
S.pool3b <- ((n.M-1)*S.M+(n.F-1)*S.F)/(n.M+n.F-2)
```

Bonferroni intervals

```
lower_3b <- (Xbar.M-Xbar.F)-qt(alpha3b/(2*p3b), df=n.M+n.F-2, lower.tail = F)*sqrt(diag
(S.pool3b)*(1/n.M+1/n.F))
upper_3b <- (Xbar.M-Xbar.F)+qt(alpha3b/(2*p3b), df=n.M+n.F-2, lower.tail = F)*sqrt(diag
(S.pool3b)*(1/n.M+1/n.F))
int.bonf3b <- cbind(lower_3b,upper_3b)
int.bonf3b
```

```
##         lower_3b   upper_3b
## Length -34.98915 -10.344185
## Width  -21.84471  -6.738628
## Height -15.75477  -6.911898
```

Comment: As p-value is much smaller than 0.05, there is a significant evidence that there is a difference between male and female.

Problem 4a

Load the data and screenshot the data

```
library(car)
```

```
## Loading required package: carData
```

```
library(carData)
data4 <- as.matrix(Pottery[,-1])
site <- Pottery[,-1]
str(Pottery)
```

```
## 'data.frame':    26 obs. of  6 variables:
##  $ Site: Factor w/ 4 levels "AshleyRails",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ Al  : num  14.4 13.8 14.6 11.5 13.8 10.9 10.1 11.6 11.1 13.4 ...
##  $ Fe  : num  7 7.08 7.09 6.37 7.06 6.26 4.26 5.78 5.49 6.92 ...
##  $ Mg  : num  4.3 3.43 3.88 5.64 5.34 3.47 4.26 5.91 4.52 7.23 ...
##  $ Ca  : num  0.15 0.12 0.13 0.16 0.2 0.17 0.2 0.18 0.29 0.28 ...
##  $ Na  : num  0.51 0.17 0.2 0.14 0.2 0.22 0.18 0.16 0.3 0.2 ...
```

```
head(Pottery)
```

```
##        Site   Al   Fe   Mg   Ca   Na
## 1 Llanedyrn 14.4 7.00 4.30 0.15 0.51
## 2 Llanedyrn 13.8 7.08 3.43 0.12 0.17
## 3 Llanedyrn 14.6 7.09 3.88 0.13 0.20
## 4 Llanedyrn 11.5 6.37 5.64 0.16 0.14
## 5 Llanedyrn 13.8 7.06 5.34 0.20 0.20
## 6 Llanedyrn 10.9 6.26 3.47 0.17 0.22
```

```
levels(Pottery$Site)
```

```
## [1] "AshleyRails" "Caldicot"    "IsleThorns"  "Llanedyrn"
```

Extract each groups

```
Ash <- data4[Pottery$Site=="AshleyRails",]
Cal <- data4[Pottery$Site=="Caldicot",]
Isl <- data4[Pottery$Site=="IsleThorns",]
Lla <- data4[Pottery$Site=="Llanedyrn",]
```

Mean of each gourp & overall mean

```
Xbar_Ash <- colMeans(Ash[,1:5])
Xbar_Ash
```

```
##     Al     Fe     Mg     Ca     Na
## 17.320  1.512  0.606  0.052  0.048
```

```
Xbar_Cal <- colMeans(Cal[,1:5])
Xbar_Cal
```

```
##     Al     Fe     Mg     Ca     Na
## 11.700  5.415  3.855  0.295  0.050
```

```
Xbar_Isl <- colMeans(Isl[,1:5])
Xbar_Isl
```

```
##     Al     Fe     Mg     Ca     Na
## 18.180  1.712  0.674  0.026  0.054
```

```
Xbar_Lla <- colMeans(Lla[,1:5])
Xbar_Lla
```

```
##         Al         Fe         Mg         Ca         Na
## 12.5642857  6.3721429  4.8264286  0.2021429  0.2507143
```

```
Xbar_all <- colMeans(data4[,1:5])
Xbar_all
```

```
##         Al         Fe         Mg         Ca         Na
## 14.4923077  4.4676923  3.1415385  0.1465385  0.1584615
```

Problem 4b

Grouping factors

```
Site4b <- Pottery$Site
```

MANOVA test

```
lmres <- lm(data4~Site4b)
summary(Manova(lmres))
```

```
##
## Type II MANOVA Tests:
##
## Sum of squares and products for error:
##            Al          Fe          Mg          Ca         Na
## Al 48.2881429  7.08007143  0.60801429  0.10647143 0.58895714
## Fe  7.0800714 10.95084571  0.52705714 -0.15519429 0.06675857
## Mg  0.6080143  0.52705714 15.42961143  0.43537714 0.02761571
## Ca  0.1064714 -0.15519429  0.43537714  0.05148571 0.01007857
## Na  0.5889571  0.06675857  0.02761571  0.01007857 0.19929286
##
## ------------------------------------------
##
## Term: Site4b
##
## Sum of squares and products for the hypothesis:
##             Al          Fe          Mg         Ca         Na
## Al  175.610319 -149.295533 -130.809707 -5.8891637 -5.3722648
## Fe -149.295533  134.221616  117.745035  4.8217866  5.3259491
## Mg -130.809707  117.745035  103.350527  4.2091613  4.7105458
## Ca   -5.889164    4.821787    4.209161  0.2047027  0.1547830
## Na   -5.372265    5.325949    4.710546  0.1547830  0.2582456
##
## Multivariate Tests: Site4b
##                  Df test stat  approx F num Df   den Df     Pr(>F)
## Pillai            3   1.55394   4.29839     15 60.00000 2.4129e-05 ***
## Wilks             3   0.01230  13.08854     15 50.09147 1.8404e-12 ***
## Hotelling-Lawley  3  35.43875  39.37639     15 50.00000 < 2.22e-16 ***
## Roy               3  34.16111 136.64446      5 20.00000 9.4435e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Extract the univariate ANOVA results

```
summary(Anova(lmres),univariate = T, multivariate = F)
```

```
##
##  Type II Sums of Squares
##           df      Al      Fe     Mg       Ca      Na
## Site4b     3 175.610 134.222 103.35 0.204703 0.25825
## residuals 22  48.288  10.951  15.43 0.051486 0.19929
##
##  F-tests
##           Al    Fe    Mg    Ca  Na
## Site4b 26.67 89.88 49.12 29.16 9.5
##
##  p-values
##        Al          Fe          Mg          Ca          Na
## Site4b 1.6269e-07 1.6794e-12 6.4522e-10 7.5455e-08 0.00032093
```

Comment: As the p-value is smaller than 0.05, so we reject the hypothesis test. The group means are not equal among them. The top block gives W and the middle block gives B, the bottom block shows the method testing resluts.

Problem 4c

Load the library

```
library(emmeans)
```

Create a list to store the results

```
p4c <- 5
pair.lst <- vector("list",p4c)
```

Name the list

```
names(pair.lst) <- colnames(data4)
```

Run the emmeans for each variable to estimate the group means etc

```
for (j in 1:p4c){
  wts <- rep(0,p4c)
  wts[j] <- 1
  pair.lst[[j]] <- emmeans(lmres,"Site4b",weights = wts)
}

pair.lst
```

```
## $Al
##  Site4b      emmean    SE df lower.CL upper.CL
##  AshleyRails   17.3 0.663 22    15.95     18.7
##  Caldicot      11.7 1.048 22     9.53     13.9
##  IsleThorns    18.2 0.663 22    16.81     19.6
##  Llanedyrn     12.6 0.396 22    11.74     13.4
##
## Results are averaged over the levels of: rep.meas
## Confidence level used: 0.95
##
## $Fe
##  Site4b      emmean    SE df lower.CL upper.CL
##  AshleyRails   1.51 0.316 22    0.858     2.17
##  Caldicot      5.42 0.499 22    4.380     6.45
##  IsleThorns    1.71 0.316 22    1.058     2.37
##  Llanedyrn     6.37 0.189 22    5.981     6.76
##
## Results are averaged over the levels of: rep.meas
## Confidence level used: 0.95
##
## $Mg
##  Site4b      emmean    SE df lower.CL upper.CL
##  AshleyRails  0.606 0.375 22   -0.171     1.38
##  Caldicot     3.855 0.592 22    2.627     5.08
##  IsleThorns   0.674 0.375 22   -0.103     1.45
##  Llanedyrn    4.826 0.224 22    4.362     5.29
##
## Results are averaged over the levels of: rep.meas
## Confidence level used: 0.95
##
## $Ca
##  Site4b      emmean     SE df lower.CL upper.CL
##  AshleyRails  0.052 0.0216 22  0.00713   0.0969
##  Caldicot     0.295 0.0342 22  0.22406   0.3659
##  IsleThorns   0.026 0.0216 22 -0.01887   0.0709
##  Llanedyrn    0.202 0.0129 22  0.17533   0.2290
##
## Results are averaged over the levels of: rep.meas
## Confidence level used: 0.95
##
## $Na
##  Site4b      emmean     SE df lower.CL upper.CL
##  AshleyRails  0.048 0.0426 22  -0.0403    0.136
##  Caldicot     0.050 0.0673 22  -0.0896    0.190
##  IsleThorns   0.054 0.0426 22  -0.0343    0.142
##  Llanedyrn    0.251 0.0254 22   0.1980    0.303
##
## Results are averaged over the levels of: rep.meas
## Confidence level used: 0.95
```

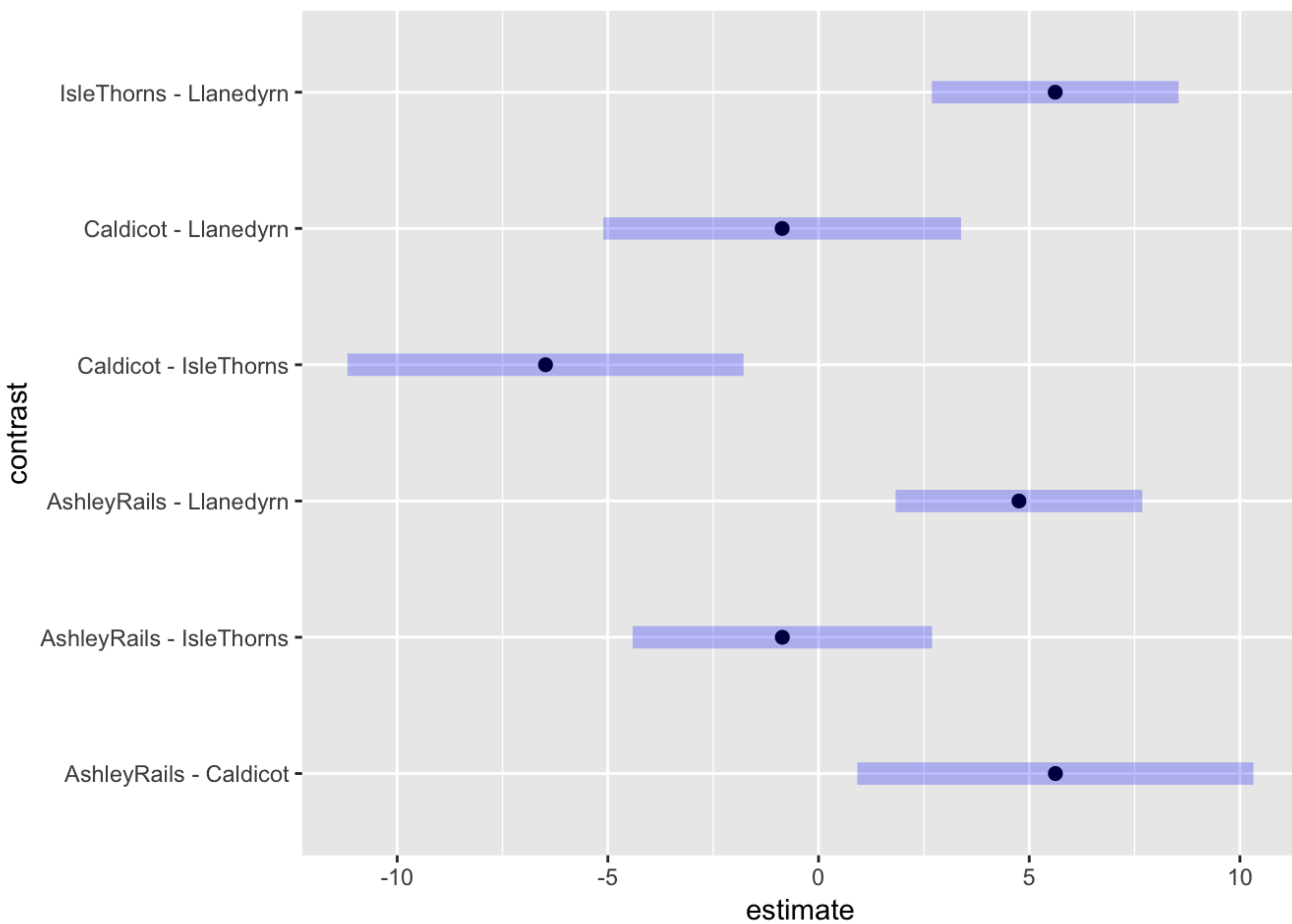
Adjust alpha due to the multiple comparisons

```
g4 <- 5
alpha_old <- 0.05
nc <- p4c * g4 *(g4-1)/2
alphanew <- 0.05/nc
alphanew
```

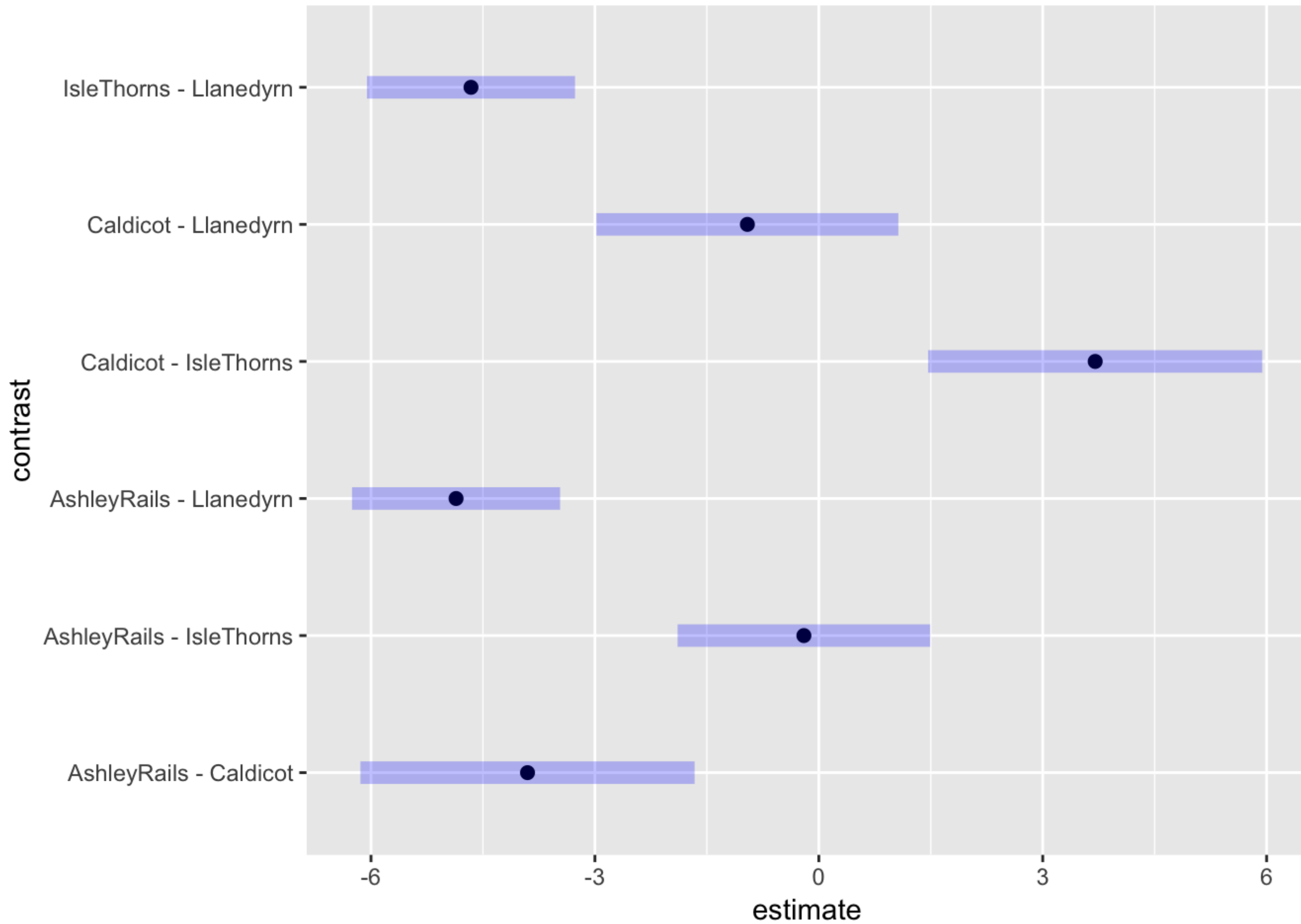
```
## [1] 0.001
```

Pair-wise differences for all 5 variables

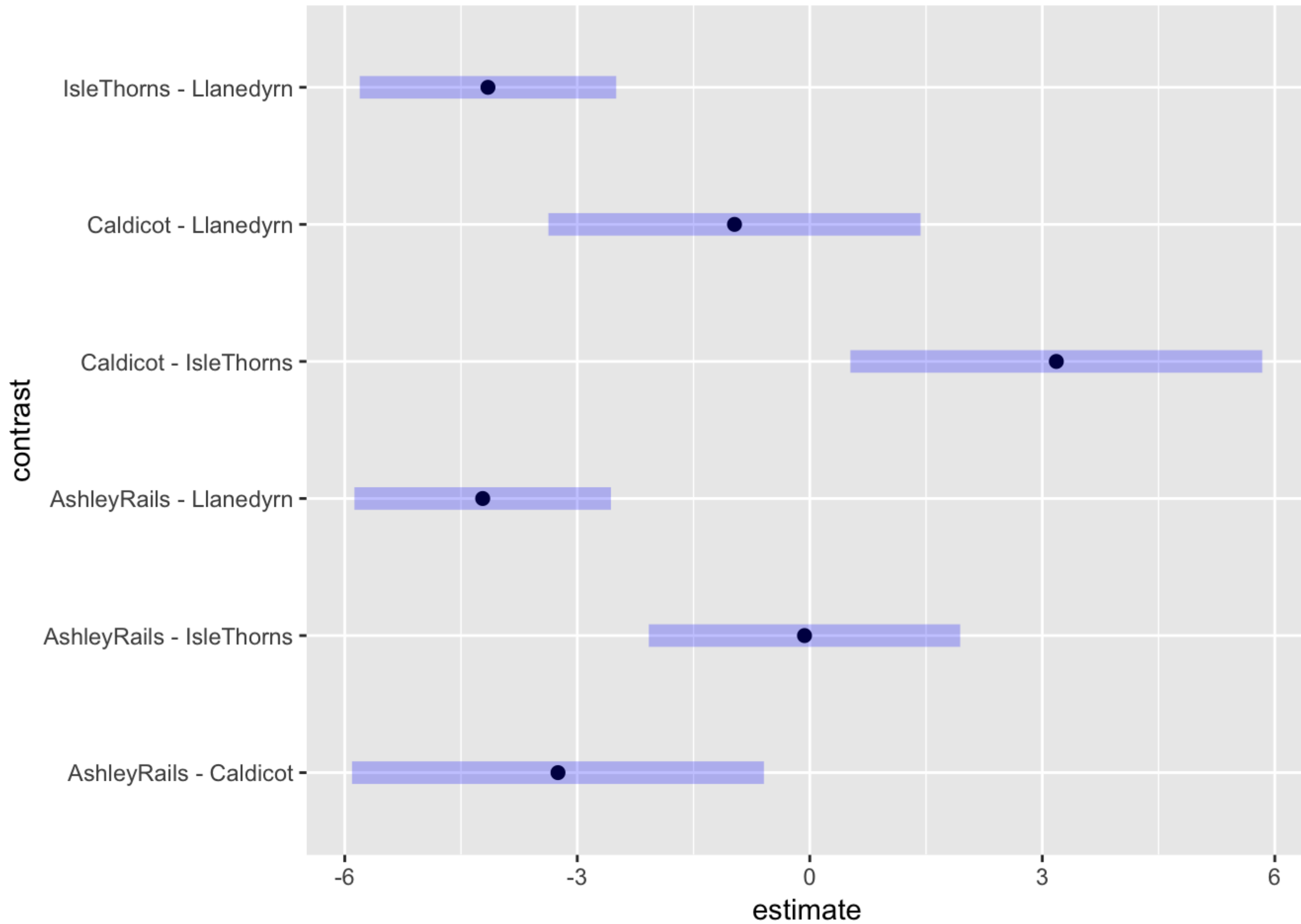
```
par(mfrow=c(1,5))
cont1 <- contrast(pair.lst$Al, "pairwise")
bb1 <- confint(cont1, level = 1-alphanew, adj="none")
plot(bb1, main="Al variable contrast")
```



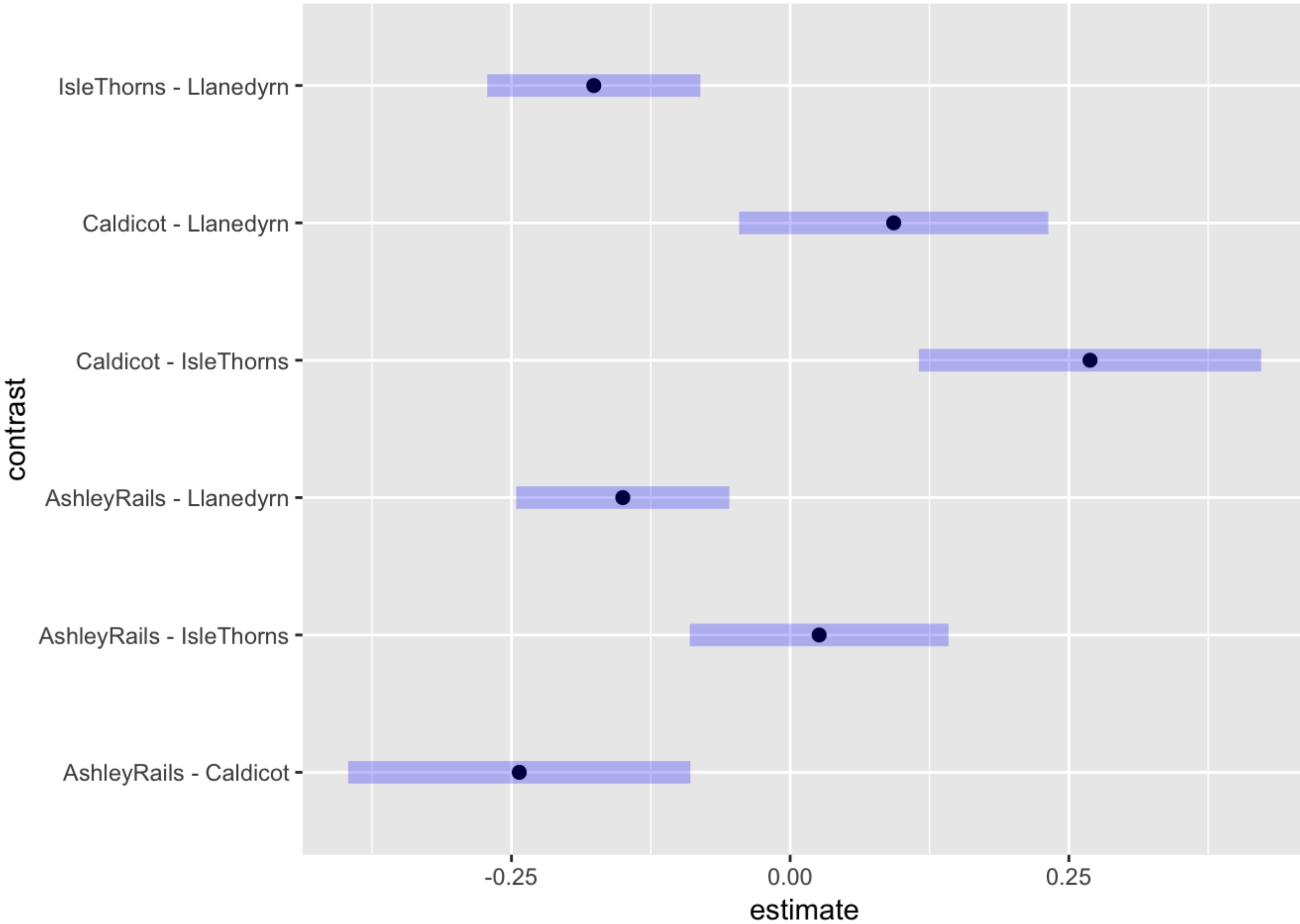

```
cont2 <- contrast(pair.lst$Fe, "pairwise")
bb2 <- confint(cont2, level = 1-alphanew, adj="none")
plot(bb2, main="Fe variable contrast")
```



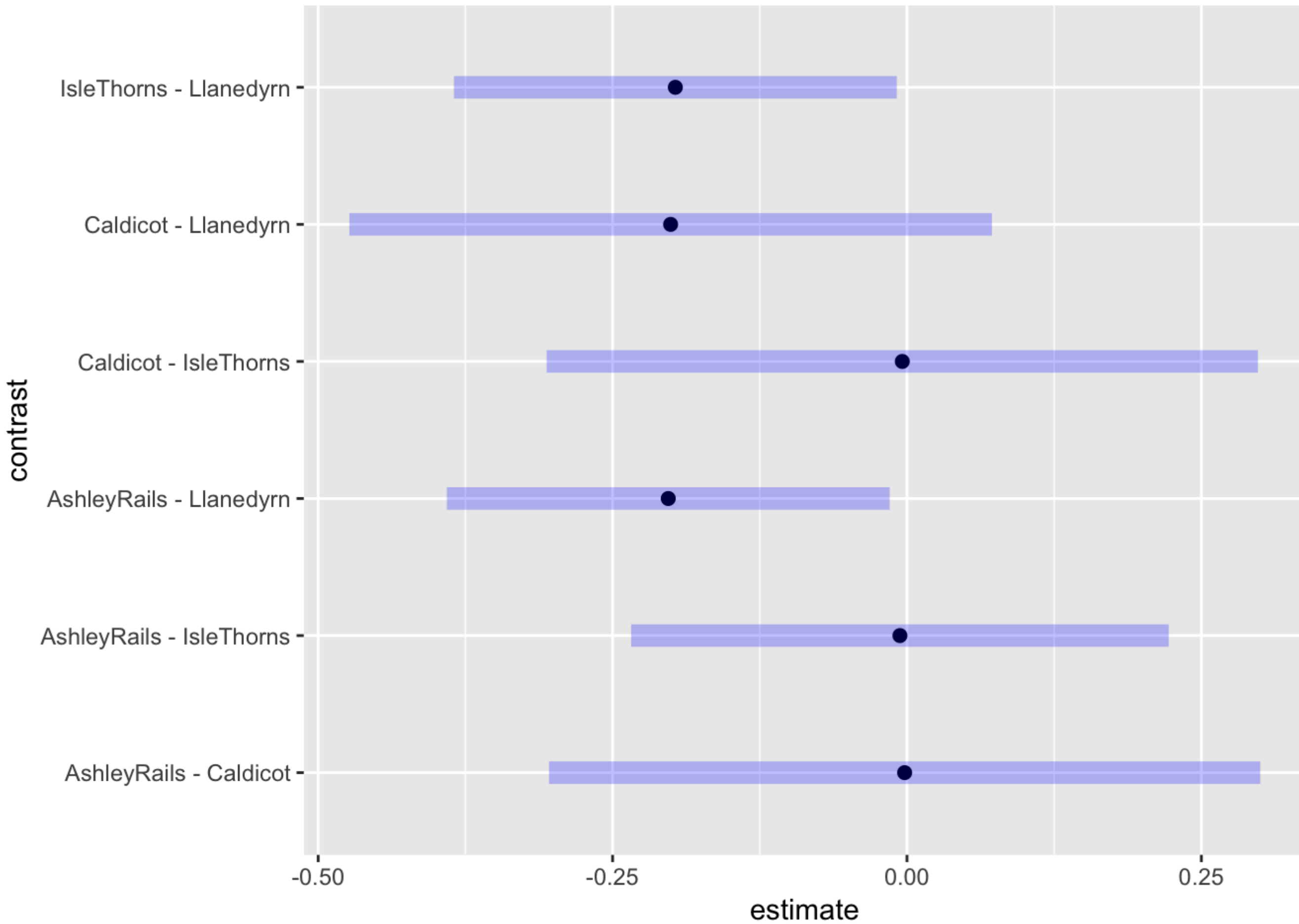

```
cont3 <- contrast(pair.lst$Mg, "pairwise")
bb3 <- confint(cont3, level = 1-alphanew, adj="none")
plot(bb3, main="Mg variable contrast")
```




```
cont4 <- contrast(pair.lst$Ca, "pairwise")
bb4 <- confint(cont4, level = 1-alphanew, adj="none")
plot(bb4, main="Ca variable contrast")
```




```
cont5 <- contrast(pair.lst$Na, "pairwise")
bb5 <- confint(cont5, level = 1-alphanew, adj="none")
plot(bb5, main="Na variable contrast")
```




```
par(mfrow=c(1,1))
```

Problem 4d

Assumptions: All distributions have different means with same variance, and independent from each other.

Problem 5a

Load the data

```
data5 <- read.table("T6-17.dat")
colnames(data5) <- c("Location", "Variety", "Yield", "SdMatKer", "Size")
head(data5)
```

```
##   Location Variety Yield SdMatKer Size
## 1        1       5 195.3    153.1 51.4
## 2        1       5 194.3    167.7 53.7
## 3        2       5 189.7    139.5 55.5
## 4        2       5 180.4    121.1 44.4
## 5        1       6 203.0    156.8 49.8
## 6        1       6 195.9    166.0 45.8
```

Perfom MANOVA

```
x5 <- as.matrix(data5[,3:5])
Location <- as.factor(data5[,1])
Variety <- as.factor(data5[,2])
res5a <- lm(x5~Location*Variety)
fit5a <- manova(res5a)
summary(fit5a, test="Wilks")
```

```
##                  Df    Wilks approx F num Df den Df   Pr(>F)
## Location          1 0.106516  11.1843      3      4 0.020502 *
## Variety           2 0.012444  10.6191      6      8 0.001928 **
## Location:Variety  2 0.074300   3.5582      6      8 0.050794 .
## Residuals         6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 5b

Comment: As p-value of the interaction is just a bit of higher than 0.05, so we cannot say there is no interaction between the variety and location, so the effects of location and variety are not additive.

Problem 5c

Three univariate ANOVA

```
summary.aov(res5a)
```

```
##  Response Yield :
##                  Df  Sum Sq Mean Sq F value  Pr(>F)
## Location          1   0.701   0.701  0.0404 0.84743
## Variety           2 196.115  98.057  5.6460 0.04177 *
## Location:Variety  2 205.102 102.551  5.9048 0.03824 *
## Residuals         6 104.205  17.367
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response SdMatKer :
##                  Df  Sum Sq Mean Sq F value  Pr(>F)
## Location          1  162.07  162.07  2.7617 0.14761
## Variety           2 1089.01  544.51  9.2786 0.01459 *
## Location:Variety  2  780.70  390.35  6.6517 0.03003 *
## Residuals         6  352.10   58.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response Size :
##                  Df  Sum Sq Mean Sq F value  Pr(>F)
## Location          1  72.521  72.521  4.5882 0.07594 .
## Variety           2 284.102 142.051  8.9872 0.01567 *
## Location:Variety  2  85.952  42.976  2.7190 0.14435
## Residuals         6  94.835  15.806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 5d Comment: The interaction effect means that as the changes of the location results the changes the relationship between the variety and the responses.