Module 4: Inference on Means and Covariances

Weixing Song

August 15, 2018

First, we use the following R-codes to install all necessary packages.

```
list.of.packages=c("pracma","car","biotools")
if(length(which(!list.of.packages %in% installed.packages()))){
  install.packages(list.of.packages[!list.of.packages %in% installed.packages()])
}
```

1. Inference on A Normal Population Mean

1.1 Hypothesis Testing

Example 4.1 Perspiration from 20 healthy females was analyzed. Three components, $X_1 =$sweat rate, $X_2 =$sodium content, and $X_3 =$potassium content, were measured. Assume that $(X_1, X_2, X_3)'$ follows a multivariate normal distribution with mean vector $\mu = (\mu_1, \mu_2, \mu_3)'$. Test the null hypothesis

$$H_0 : \mu' = [4, 50, 10] \quad \text{versus} \quad H_1 : \mu' \neq [4, 50, 10].$$

R-Codes

```
HT2.test=function(data,mu0,siglevel)
 {
  n=nrow(data);
  p=ncol(data);
  xbar=colMeans(data);
  S=var(data);
  Tsq=round(n*t(xbar-mu0)%*%solve(S)%*%(xbar-mu0),2)
  Fv=round((n-1)*p*qf(1-siglevel,p,n-p)/(n-p),2)
  if(Tsq>Fv)
    {
     noquote(cbind(Tsq, Fv, "sig."))
    }
  else
    {
     noquote(cbind(Tsq, Fv, "non-sig."))
    }
}

sweat=read.table("S:/Workshop/Data/T5-1.DAT")
HT2.test(sweat,c(4,50,10),0.1)

##           Fv
## [1,] 9.74 8.17 sig.
```

1.2 Confidence Region for Population mean vector only.

Example Data for radiation from microwave ovens were collected. Two variables are considered

$$x_1 = \sqrt[4]{\text{measured ratdiation with door closed}}, \quad x_2 = \sqrt[4]{\text{measured ratdiation with door open}}.$$

Construct a 95% confidence interval of $\boldsymbol{\mu} = (\mu_1, \mu_2)$.

```r
HT2.cfl=function(data,siglevel)
 {
  n=nrow(data);
  p=ncol(data);
  xbar=colMeans(data);
  S=var(data);
  lam=eigen(S)$values;
  ev=eigen(S)$vectors;
  Fv=(n-1)*p*qf(1-siglevel,p,n-p)/((n-p)*n)
  len=sqrt(Fv*lam);
  cat("The confidence region of the population mean is an ellipsoid.\n\n")
  cat("The center of the ellipsoid:\n")
  cat(xbar,"\n\n")
  cat("The directions of the ellipsoid:\n")
  print(ev);
  cat("\n","The length of axis in each direction:\n")
  cat(len,"\n")

  rad=sqrt(Fv)
  conf.elip = ellipse(center=xbar, shape=S, radius=rad, col="blue",draw=F,fill=T,grid=T)
  plot(conf.elip, type='l')
  points(xbar[1],xbar[2],pch=19,cex=1.5)
}

data1=read.table("S:/Workshop/Data/T4-1.DAT")
data2=read.table("S:/Workshop/Data/T4-5.DAT")
library(pracma);
library(car);
```
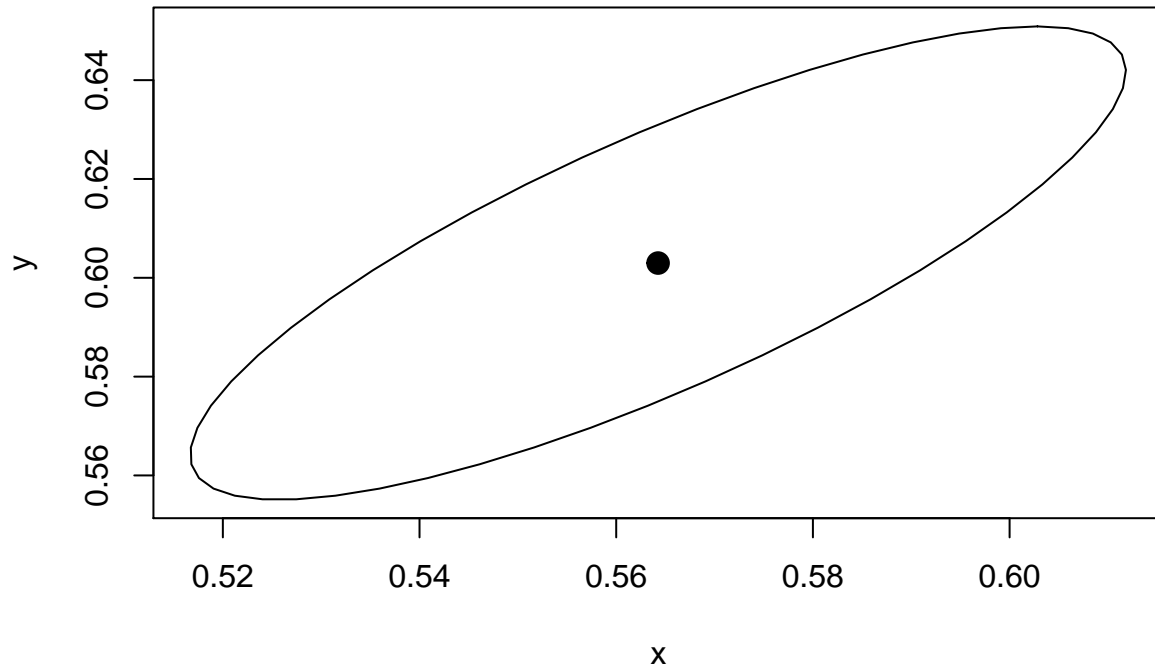
```
##
## Attaching package: 'car'

## The following object is masked from 'package:pracma':
##
##      logit
```

```r
data=nthroot(as.matrix(cbind(data1,data2)),4)
HT2.cfl(data,0.05)
```

```
## The confidence region of the population mean is an ellipsoid.
##
## The center of the ellipsoid:
## 0.5642575 0.6029812
##
## The directions of the ellipsoid:
##             [,1]        [,2]
## [1,] 0.7041574 -0.7100439
## [2,] 0.7100439  0.7041574
##
```

```
##  The length of axis in each direction:
## 0.06424195 0.02075877
```


1.3 Simultaneous Confidence Intervals

Example 4.2 (Constructing simultaneous confidence intervals and ellipses). The scores obtained by $n = 87$ college students on the College Level Examination Program (CLEP) subtest X_1, and the College Qualification Test (CQT) subtests X_2 and X_3 are given in Table 5.2 on page 228 for X_1 =social science and history, X_2 =verbal, and X_3=science. Construct simultaneous confidence intervals for μ_1, μ_2 and μ_3.

```
sci.conf=function(data,coeA,clevel,type="B")
 {
  xbar=colMeans(data);
  S=var(data);
  n=nrow(data);
  p=ncol(data);
  Fv=(n-1)*p*qf(clevel,p,n-p)/((n-p)*n)
  tv=qt(1-(1-clevel)/(2*nrow(coeA)),n-1)
  tvn=qt(1-(1-clevel)/2,n-1)
  chiv=qchisq(clevel,p);
  if(type=="S")
  {
  lend=coeA%*%xbar-sqrt(Fv*diag(coeA%*%S%*%t(coeA)))
  rend=coeA%*%xbar+sqrt(Fv*diag(coeA%*%S%*%t(coeA)))
  cat("\n The simultaneous confidence intervals (Scheffe) are:\n\n")
  }
```

```r
else if (type=="B")
{
lend=coeA%*%xbar-tv*sqrt(diag(coeA%*%S%*%t(coeA))/n)
rend=coeA%*%xbar+tv*sqrt(diag(coeA%*%S%*%t(coeA))/n)
cat("\n The simultaneous confidence intervals (Bonferroni) are:\n\n")
}
else if (type=="L")
{
lend=coeA%*%xbar-sqrt(chiv)*sqrt(diag(coeA%*%S%*%t(coeA))/n)
rend=coeA%*%xbar+sqrt(chiv)*sqrt(diag(coeA%*%S%*%t(coeA))/n)
cat("\n The simultaneous confidence intervals (Large Sample) are:\n\n")
}
else if (type=="NS")
{
lend=coeA%*%xbar-tvn*sqrt(diag(coeA%*%S%*%t(coeA))/n)
rend=coeA%*%xbar+tvn*sqrt(diag(coeA%*%S%*%t(coeA))/n)
cat("\n The simultaneous confidence intervals (NS) are:\n\n")
}

result=(cbind(lend,rend));
dimnames(result)=list(seq(length(lend)),c("Left Limit","Right Limit"))
print(result)
}
data=read.table("S:/Workshop/Data/T5-2.DAT")
A=rbind(c(1,0,0),c(0,1,0),c(0,0,1),c(0,1,-1))
sci.conf(data,A,0.95,"S")
```

```
##
##  The simultaneous confidence intervals (Scheffe) are:
##
##    Left Limit Right Limit
## 1  502.99940   550.17302
## 2   51.21484    58.16447
## 3   23.63855    26.61432
## 4   26.43153    32.69491
```

```r
sci.conf(data,A,0.95,"B")
```

```
##
##  The simultaneous confidence intervals (Bonferroni) are:
##
##    Left Limit Right Limit
## 1  505.74026   547.43216
## 2   51.61863    57.76069
## 3   23.81145    26.44143
## 4   26.79544    32.33100
```

```r
sci.conf(data,A,0.95,"L")
```

```
##
##  The simultaneous confidence intervals (Large Sample) are:
##
##    Left Limit Right Limit
## 1  503.74533   549.42709
## 2   51.32473    58.05458
```

```
## 3    23.68560    26.56727
## 4    26.53057    32.59587
```

```
 sci.conf(data,A,0.95,"NS")
```

```
##
##   The simultaneous confidence intervals (NS) are:
##
##    Left Limit Right Limit
## 1  510.34352   542.82889
## 2   52.29678    57.08253
## 3   24.10183    26.15105
## 4   27.40663    31.71981
```

2. Paired Comparison

2.1 Hypothesis Testing

Example 4.3 (Checking for a mean difference with paired observations) Municipal wastewater treatment plants are required by law to monitor their discharges into rivers and streams on a regular basis. Concern about the reliability of data from one of these self-monitoring programs led to a study in which samples of effluent were divided and sent to two laboratories for testing. One-half of each sample was sent to the Wisconsin State Laboratory of Hygiene, and one-half was sent to a private commercial laboratory routinely used in the monitoring program. Measurements of biochemical oxygen demand (BOD) and suspended solids (SS) were obtained, for $n = 11$ sample splits, from the two laboratories.

Hypothesis Testing: $H_0 : \delta = 0$ **versus** $H_1 : H_0$ **is not true.**

```
HT2.test=function(data,mu0,siglevel)
 {
  n=nrow(data);
  p=ncol(data);
  xbar=colMeans(data);
  S=var(data);
  Tsq=round(n*t(xbar-mu0)%*%solve(S)%*%(xbar-mu0),2)
  Fv=round((n-1)*p*qf(1-siglevel,p,n-p)/(n-p),2)
  if(Tsq>Fv)
    {
     noquote(cbind(Tsq, Fv, "sig."))
    }
  else
    {
     noquote(cbind(Tsq, Fv, "non-sig."))
    }
}

waste=read.table("S:/Workshop/Data/T6-1.DAT")
d1=waste$V1-waste$V3
d2=waste$V2-waste$V4
mydata=cbind(d1,d2)
HT2.test(mydata,c(0,0),0.05)
```

```
##             Fv
## [1,] 13.64 9.46 sig.
```

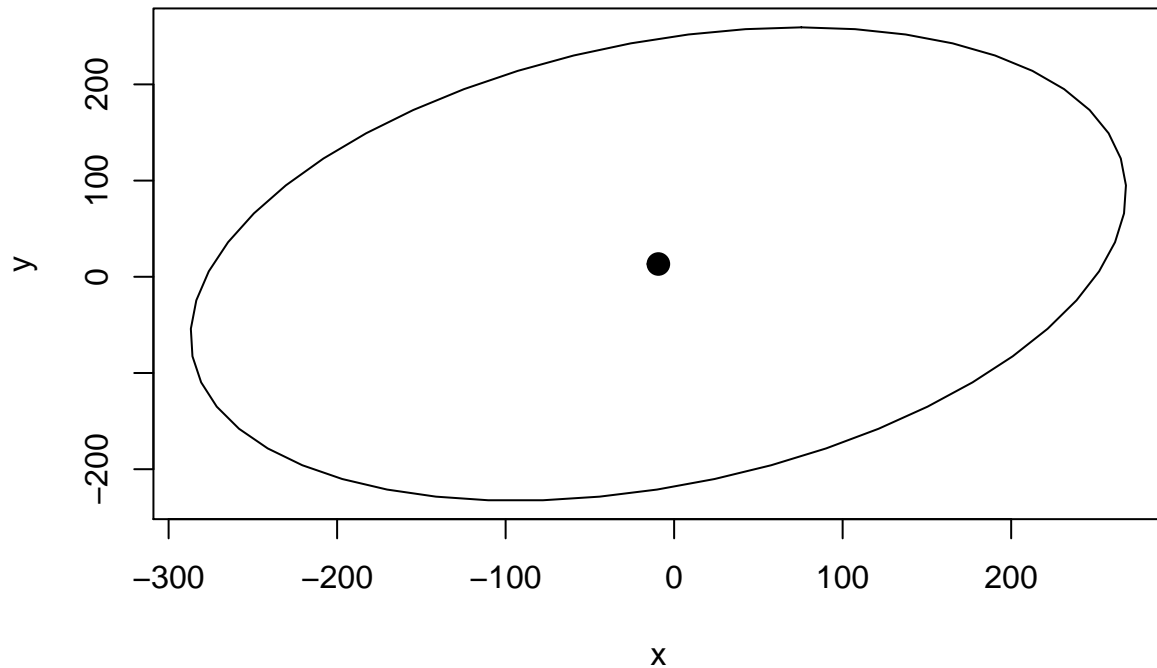## 2.2. Confidence Region

Confidence Region for *delta*

Construct a 95% confidence interval of δ.

```r
HT2.cfl=function(data,siglevel)
 {
  n=nrow(data);
  p=ncol(data);
  xbar=colMeans(data);
  S=var(data);
  lam=eigen(S)$values;
  ev=eigen(S)$vectors;
  Fv=(n-1)*p*qf(1-siglevel,p,n-p)/((n-p)*n)
  len=sqrt(Fv*lam);
  cat("The confidence region of the population mean is an ellipsoid.\n\n")
  cat("The center of the ellipsoid:\n")
  cat(xbar,"\n\n")
  cat("The directions of the ellipsoid:\n")
  print(ev);
  cat("\n","The length of axis in each direction:\n")
  cat(len,"\n")

  conf.elip = ellipse(center=xbar, shape=S, radius=len, col="blue",draw=F,fill=T,grid=T)
  plot(conf.elip, type='l')
  points(xbar[1],xbar[2],pch=19,cex=1.5)
}


library(pracma);
library(car);
HT2.cfl(mydata,0.05)
```

```
## The confidence region of the population mean is an ellipsoid.
##
## The center of the ellipsoid:
## -9.363636 13.27273
##
## The directions of the ellipsoid:
##            [,1]       [,2]
## [1,] 0.3324812 -0.9431099
## [2,] 0.9431099  0.3324812
##
##  The length of axis in each direction:
## 19.66569 12.02364
```

Simultaneous Confidence Interval

Construct simultaneous confidence intervals for δ_1, δ_2.

```r
sci.conf=function(data,coeA,clevel,type="B")
 {
  xbar=colMeans(data);
  S=var(data);
  n=nrow(data);
  p=ncol(data);
  Fv=(n-1)*p*qf(clevel,p,n-p)/((n-p)*n)
  tv=qt(1-(1-clevel)/(2*nrow(coeA)),n-1)
  chiv=qchisq(clevel,p);
  if(type=="S")
  {
  lend=coeA%*%xbar-sqrt(Fv*diag(coeA%*%S%*%t(coeA)))
  rend=coeA%*%xbar+sqrt(Fv*diag(coeA%*%S%*%t(coeA)))
  cat("\n The simultaneous confidence intervals (Scheffe) are:\n\n")
  }
  else if (type=="B")
  {
  lend=coeA%*%xbar-tv*sqrt(diag(coeA%*%S%*%t(coeA))/n)
  rend=coeA%*%xbar+tv*sqrt(diag(coeA%*%S%*%t(coeA))/n)
  cat("\n The simultaneous confidence intervals (Bonferroni) are:\n\n")
  }
```

```
else if (type=="L")
{
lend=coeA%*%xbar-chiv*sqrt(diag(coeA%*%S%*%t(coeA))/n)
rend=coeA%*%xbar+chiv*sqrt(diag(coeA%*%S%*%t(coeA))/n)
cat("\n The simultaneous confidence intervals (Large Sample) are:\n\n")
}

result=(cbind(lend,rend));
dimnames(result)=list(seq(length(lend)),c("Left Limit","Right Limit"))
print(result)
}

A=rbind(c(1,0),c(0,1))
sci.conf(mydata,A,0.95,"S")
```

```
##
##  The simultaneous confidence intervals (Scheffe) are:
##
##   Left Limit Right Limit
## 1 -22.453272     3.72600
## 2  -5.700119    32.24557
```

3. Repeated Measures Design

3.1 Hypothesis Testing

Example 4.4 (Anesthetics effects on animals)
```
rep.test=function(data,conA,siglevel)
{
 n=nrow(data);
 q=ncol(data);
 xbar=colMeans(data);
 S=var(data);
 Tsq=round(n*t(conA%*%xbar)%*%solve(conA%*%S%*%t(conA))%*%(conA%*%xbar),2)
 Fv=round((n-1)*(q-1)*qf(1-siglevel,q-1,n-q+1)/(n-q+1),2)
 if(Tsq>Fv)
   {
    noquote(cbind(Tsq, Fv, "sig."))
   }
 else
   {
    noquote(cbind(Tsq, Fv, "non-sig."))
   }
}
C=rbind(c(-1,-1,1,1),c(1,-1,1,-1),c(1,-1,-1,1));
mydata=read.table("S:/Workshop/Data/T6-2.DAT")
rep.test(mydata,C,0.05)
```

```
##              Fv
## [1,] 116.02 10.93 sig.
```

3.2. Simultaneous Confidence Interval

```
sci.rep=function(data,coeA,clevel)
 {
  xbar=colMeans(data);
  S=var(data);
  n=nrow(data);
  q=ncol(data);
  Fv=(n-1)*(q-1)*qf(clevel,q-1,n-q+1)/((n-q+1)*n)
  lend=coeA%*%xbar-sqrt(Fv*diag(coeA%*%S%*%t(coeA)))
  rend=coeA%*%xbar+sqrt(Fv*diag(coeA%*%S%*%t(coeA)))
  cat("\n The simultaneous confidence intervals are:\n\n")
  result=(cbind(lend,rend));
  dimnames(result)=list(seq(length(lend)),c("Left Limit","Right Limit"))
  print(result)
 }
 C=rbind(c(-1,-1,1,1),c(1,-1,1,-1),c(1,-1,-1,1));
 mydata=read.table("S:/Workshop/Data/T6-2.DAT")
 sci.rep(mydata,C,0.95)
```

```
##
##  The simultaneous confidence intervals are:
##
##    Left Limit Right Limit
## 1  135.65030   282.98128
## 2 -114.72708    -5.37818
## 3  -78.72858    53.14964
```

4. Comparing Several Multivariate Population Means

4.1 When summary statistics are available

Example 4.5 A Multivariate analysis of Wisconsin nursing home data.

```
  n1=271;
  n2=138;
  n3=107;
  g=3;

  xbar1=c(2.066,.480,.082,.360)
  xbar2=c(2.167,.596,.124,.418)
  xbar3=c(2.273,.521,.125,.383)
  S1=matrix(c(.291,-.001,.002,.010,-.001,.011,.000,.003,.002,.000,.001,.000,
              .010,.003,.000,.010),4,4)
  S2=matrix(c(.561,.011,.001,.037,.011,.025,.004,.007,.001,.004,.005,.002,
              .037,.007,.002,.019),4,4)
  S3=matrix(c(.261,.030,.003,.018,.030,.017,-.000,.006,.003,-.000,.004,.001,
              .018,.006,.001,.013),4,4)
  p=length(xbar1);

  #calculate statistics
```

```
  W=(n1-1)*S1+(n2-1)*S2+(n3-1)*S3 # W different from text so different values for test statistic
  xbar=(n1*xbar1+n2*xbar2+n3*xbar3)/(n1+n2+n3)
  B=n1*(xbar1-xbar)%*%t(xbar1-xbar)+n2*(xbar2-xbar)%*%t(xbar2-xbar)
    +n3*(xbar3-xbar)%*%t(xbar3-xbar)
```

```
##              [,1]         [,2]         [,3]         [,4]
## [1,] 2.01015843 0.0216292762 0.335125882 0.0398763135
## [2,] 0.02162928 0.0002327307 0.003605950 0.0004290686
## [3,] 0.33512588 0.0036059498 0.055870898 0.0066480256
## [4,] 0.03987631 0.0004290686 0.006648026 0.0007910423
```

```
  Lamb=det(W)/det(B+W)


 #trasformed statistic to F


  n=n1+n2+n3
  T=((n-p-2)/p)*(1-sqrt(Lamb))/sqrt(Lamb)
  crit=qf(.99,2*p,2*(n-p-2))
  c(T,crit)
```

```
## [1] 15.101769  2.528682
```

```
#the approximation for log Lamb
  Tlog=-(n-1-(p+g)/2)*log(Lamb)
  critlog=qchisq(.99,p*(g-1))
  c(Tlog,critlog)
```

```
## [1] 114.51416  20.09024
```

4.2 When raw data are available

Example 4.6 The relationship of size and shape for painted turtles is studied. The following table contains their measurements on the carapaces of 24 female and 24 male turtles. Test for equality of the two population mean vectors using $\alpha = 0.05$.

```
mydata=read.table("S:/Workshop/Data/T6-9.DAT")
X=mydata[1:24,-4]
Y=mydata[25:48,-4]
S1=cov(X)
S2=cov(Y)
T2=cov(rbind(X,Y))
W=23*(S1+S2);
BW=47*T2;
Lamb=det(W)/det(BW)
result=manova(cbind(V1,V2,V3)~V4, data=mydata)

summary(result)
```

```
##           Df  Pillai approx F num Df den Df    Pr(>F)
## V4         1 0.61143   23.078      3     44 3.967e-09 ***
## Residuals 46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(result,test="Roy")
```

```
##             Df     Roy approx F num Df den Df     Pr(>F)
## V4          1 1.5735    23.078        3      44 3.967e-09 ***
## Residuals 46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary.aov(result)
```

```
##  Response V1 :
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## V4          1  6165.3  6165.3  20.889 3.656e-05 ***
## Residuals  46 13576.6   295.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response V2 :
##             Df Sum Sq Mean Sq F value    Pr(>F)
## V4          1 2451.0 2451.02  22.104 2.376e-05 ***
## Residuals  46 5100.8  110.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response V3 :
##             Df Sum Sq Mean Sq F value    Pr(>F)
## V4          1 1541.3  1541.3  40.563 8.087e-08 ***
## Residuals  46 1747.9    38.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.3 Simultaneous Confidence Intervals for Treatment Means

Example 4.7 A Multivariate analysis of Wisconsin nursing home data.

```r
Multi.Sci=function(p,g,n1,n2,n3,xbar1,xbar2,xbar3,S1,S2,S3,clevel)
{
  n=n1+n2+n3;
  W=(n1-1)*S1+(n2-1)*S2+(n3-1)*S3 # W different from text so different values for test statistic
  tv=qt(1-(1-clevel)/(p*g*(g-1)),n-g)

  L12=xbar1-xbar2-tv*sqrt(diag(W)*(1/n1+1/n2)/(n-g))
  R12=xbar1-xbar2+tv*sqrt(diag(W)*(1/n1+1/n2)/(n-g))

  L13=xbar1-xbar3-tv*sqrt(diag(W)*(1/n1+1/n3)/(n-g))
  R13=xbar1-xbar3+tv*sqrt(diag(W)*(1/n1+1/n3)/(n-g))

  L23=xbar2-xbar3-tv*sqrt(diag(W)*(1/n2+1/n3)/(n-g))
  R23=xbar2-xbar3+tv*sqrt(diag(W)*(1/n2+1/n3)/(n-g))

  cat("SCI: Trt1-Trt2\n")
  result=cbind(L12,R12)
  print(result)
  cat("\n\n")

  cat("SCI: Trt1-Trt3\n")
```

```
    result=cbind(L13,R13)
    print(result)
    cat("\n\n")

    cat("SCI: Trt2-Trt3\n")
    result=cbind(L23,R23)
    print(result)
 }

  n1=271;
  n2=138;
  n3=107;
  g=3;

  xbar1=c(2.066,.480,.082,.360)
  xbar2=c(2.167,.596,.124,.418)
  xbar3=c(2.273,.521,.125,.383)
  S1=matrix(c(.291,-.001,.002,.010,-.001,.011,.000,.003,.002,.000,.001,.000,.010,.003,.000,.010),4,4)
  S2=matrix(c(.561,.011,.001,.037,.011,.025,.004,.007,.001,.004,.005,.002,.037,.007,.002,.019),4,4)
  S3=matrix(c(.261,.030,.003,.018,.030,.017,-.000,.006,.003,-.000,.004,.001,.018,.006,.001,.013),4,4)
  p=length(xbar1);

  Multi.Sci(p,g,n1,n2,n3,xbar1,xbar2,xbar3,S1,S2,S3,0.95)
```

```
## SCI: Trt1-Trt2
##              L12          R12
## [1,] -0.28081737  0.07881737
## [2,] -0.15404725 -0.07795275
## [3,] -0.05760551 -0.02639449
## [4,] -0.09234919 -0.02365081
##
##
## SCI: Trt1-Trt3
##              L13          R13
## [1,] -0.40331955 -0.0106804531
## [2,] -0.08253891  0.0005389142
## [3,] -0.06003765 -0.0259623456
## [4,] -0.06050148  0.0145014801
##
##
## SCI: Trt2-Trt3
##              L23          R23
## [1,] -0.327485749 0.11548575
## [2,]  0.028136213 0.12186379
## [3,] -0.020221711 0.01822171
## [4,] -0.007308795 0.07730879
```

3. Testing for Equality of Covariance Matrices

Example 4.8 Researchers interested in assessing pulmonary function in non pathological populations asked subjects to run on a treadmill until exhaustion. Samples of air were collected at definite intervals and the gas contents analyzed. The results on 4 measures of oxygen consumption for 25 males and 25 females are given

in the following data set. The variables were

$$
\begin{aligned}
X1 &= \text{resting volume } O_2 \text{ (L/min)} \\
X2 &= \text{resting volume } O_2 \text{ (mL/kg/min)} \\
X3 &= \text{maximum volume } O_2 \text{ (L/min)} \\
X4 &= \text{maximum volume } O_2 \text{ (mL/kg/min)}
\end{aligned}
$$

Assume (X_1, X_2, X_3, X_4) is normal. Test the hypothesis $H_0 : \Sigma_1 = \Sigma_2 = \Sigma$.

```
require(biotools)
```

```
## Loading required package: biotools

## Loading required package: rpanel

## Loading required package: tcltk

## Package `rpanel', version 1.1-4: type help(rpanel) for summary information

## Loading required package: tkrplot

## Loading required package: MASS

## Loading required package: lattice

## Loading required package: SpatialEpi

## Loading required package: sp

## ---
## biotools version 3.1

##
```

```
mydata=read.table("S:/Workshop/Data/T6-12.DAT")
boxM(mydata[,1:4], mydata$V5)
```

```
##
##  Box's M-test for Homogeneity of Covariance Matrices
##
## data:  mydata[, 1:4]
## Chi-Sq (approx.) = 58.002, df = 10, p-value = 8.637e-09
```