Illustration of 2 Indpendent Samples Hotelling's T^2

Observations pertaining to six educational variables, which are thought to reflect the quality of education provided by an acedemic institution, are collected for 25 Liberal Arts colleges and 25 Research Universities for their under-graduate programmes, in the US. The six educational variables of interest are as folows: `sat`, giving the median combined Math and Verbal SAT score of the students; `acc`, indicating the percentage of applicants that get accepted; `dps`, providing the money spent per student in dollars; `t10`, yielding the percentage of students in the institution, who were in the top 10% of their graduating higher secondary class; `phd`, showing the percentage of faculty at the institution that have PhD degrees; and `gp`, giving the graduating percentage of students from the institution.

The data file `colleges.data` contains observations on these six variables in its last six of the eight columns, with the `headers` of the first two columns being `name`, giving the name of the institution and `it`, indicating the institution type - assuming two possible levels of `Lib_Arts` (Liberal Arts college) or `Univ` (Research University). It has 51 rows including the first "`header`" row. The objective is first of all to see if the two types institutions differ with respect to these six educational variables at all or not, and if so which ones.

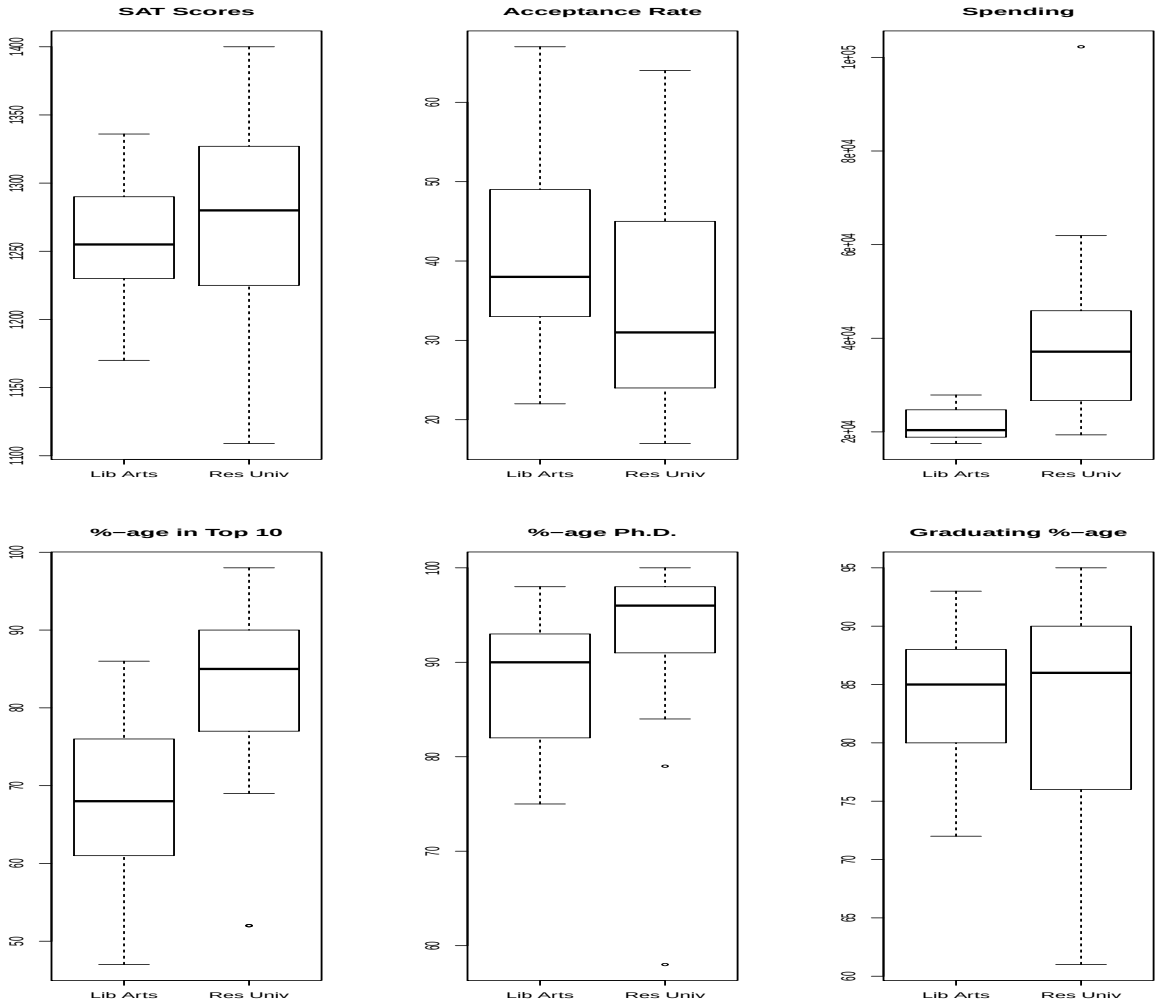
```
> d<-read.table("colleges.data",header=T)
> d
              name       it  sat acc    dps t10 phd gp
1          Amherst Lib_Arts 1315  22  26636  85  81 93
2       Swarthmore Lib_Arts 1310  24  27487  78  93 88
.....................................................
.....................................................
49      Washington     Univ 1225  54  39883  71  98 76
50  U_of_Rochester     Univ 1155  56  38597  52  96 73

> data<-matrix(c(d$sat,d$acc,d$dps,d$t10,d$phd,d$gp),ncol=6,
+ dimnames=list(d$name,c("sat","acc","dps","t10","phd", "gp")))

> data.la<-matrix(c(d$sat[d$it=="Lib_Arts"],d$acc[d$it=="Lib_Arts"],
+ d$dps[d$it=="Lib_Arts"],d$t10[d$it=="Lib_Arts"],d$phd[d$it=="Lib_Arts"],
+ d$gp[d$it=="Lib_Arts"]),ncol=6,dimnames=list(d$name[d$it=="Lib_Arts"],
+ c("sat","acc","dps","t10","phd", "gp")))

> data.ru<-matrix(c(d$sat[d$it=="Univ"],d$acc[d$it=="Univ"],d$dps[d$it=="Univ"],
+ d$t10[d$it=="Univ"],d$phd[d$it=="Univ"],d$gp[d$it=="Univ"]),ncol=6,
+ dimnames=list(d$name[d$it=="Univ"],c("sat","acc","dps","t10","phd", "gp")))

> boxplot(data.la[,1],data.ru[,1],names=c("Lib Arts","Res Univ"),main="SAT Scores")
> boxplot(data.la[,2],data.ru[,2],names=c("Lib Arts","Res Univ"),main="Acceptance Rate")
> boxplot(data.la[,3],data.ru[,3],names=c("Lib Arts","Res Univ"),main="Spending")
> boxplot(data.la[,4],data.ru[,4],names=c("Lib Arts","Res Univs"),main="%-age in Top 10")
> boxplot(data.la[,5],data.ru[,5],names=c("Lib Arts","Res Univs"),main="%-age Ph.D.")
> boxplot(data.la[,6],data.ru[,6],names=c("Lib Arts","Res Univs"),main="Graduating %-age")
```




```
>hot2is <- function(x1,x2)
+ {
+  p1<-ncol(x1)
+  p2<-ncol(x2)
+  if(p1 != p2)
+   print("Dimensions do not match.")
+  else
+  {
+   p<-p1
+   n1<-nrow(x1)
+   n2<-nrow(x2)
+   x1b<-matrix(apply(x1,2,mean),ncol=1)
+   x2b<-matrix(apply(x2,2,mean),ncol=1)
+   Sp<-((n1-1)*cov(x1)+(n2-1)*cov(x2))/(n1+n2-2)
+   T2<-((n1*n2)/(n1+n2))*(t(x1b-x2b)%*%solve(Sp)%*%(x1b-x2b))
+   F.obs<-((n1+n2-p-1)/(p*(n1+n2-2)))*T2
+   p.value<-1-pf(F.obs,p,n1+n2-2)
+   return(list(T2=T2,F.obs=F.obs,p.value=p.value))
+  }
+ }
```

```
> hot2is(data.la,data.ru)
$T2
         [,1]
[1,] 56.53185
$F.obs
         [,1]
[1,] 8.440519
$p.value
             [,1]
[1,] 2.869683e-06

> data.frame(Var= c("sat","acc","dps","t10","phd", "gp"),
+            LCL=(x1b-x2b)-sqrt(((48*6)/43)*qf(0.95,6,43)*(2/25)*diag(Sp)),
+            UCL=(x1b-x2b)+sqrt(((48*6)/43)*qf(0.95,6,43)*(2/25)*diag(Sp)))

  Var           LCL          UCL
1 sat    -84.335431    55.055431
2 acc     -9.287617    20.167617
3 dps -31204.476185 -2762.083815
4 t10    -27.228059    -1.571941
5 phd    -13.558363     4.278363
6  gp     -7.199219     9.759219

> tsat<-t.test(d$sat~d$it,var.equal=T)
> tacc<-t.test(d$acc~d$it,var.equal=T)
> tdps<-t.test(d$dps~d$it,var.equal=T)
> tt10<-t.test(d$t10~d$it,var.equal=T)
> tphd<-t.test(d$phd~d$it,var.equal=T)
> tgp<-t.test(d$gp~d$it,var.equal=T)

> data.frame( Var=c("sat","acc","dps","t10","phd", "gp"),
+             LCL=c(tsat$conf[1],tacc$conf[1],tdps$conf[1],
+                   tt10$conf[1],tphd$conf[1],tgp$conf[1]),
+             UCL=c(tsat$conf[2],tacc$conf[2],tdps$conf[2],
+                   tt10$conf[2],tphd$conf[2],tgp$conf[2]),
+             p.value=c(tsat$p.value,tacc$p.value,tdps$p.value,
                   tt10$p.value,tphd$p.value,tgp$p.value))

  Var           LCL           UCL      p.value
1 sat    -50.200848  2.092085e+01 4.119055e-01
2 acc     -2.074504  1.295450e+01 1.520223e-01
3 dps -24239.391235 -9.727169e+03 2.177199e-05
4 t10    -20.945288 -7.854712e+00 5.559154e-05
5 phd     -9.190436 -8.956439e-02 4.582705e-02
6  gp     -3.046370  5.606370e+00 5.547268e-01
```

```
> raw_p<-c(0.4119055,0.1520223,2.177199e-05,5.559154e-05,0.04582705,0.5547268)

>  p.adjust(raw_p,method="bonf")
[1] 1.0000000000 0.9121338000 0.0001306319 0.0003335492 0.2749623000
[6] 1.0000000000

>  p.adjust(raw_p,method="holm")
[1] 0.8238110000 0.4560669000 0.0001306319 0.0002779577 0.1833082000
[6] 0.8238110000

>  p.adjust(raw_p,method="hoch")
[1] 0.5547268000 0.4560669000 0.0001306319 0.0002779577 0.1833082000
[6] 0.5547268000

>  p.adjust(raw_p,method="hommel")
[1] 0.5547268000 0.4560669000 0.0001306319 0.0002779577 0.1833082000
[6] 0.5547268000

>  p.adjust(raw_p,method="fdr")
[1] 0.4942866000 0.2280334500 0.0001306319 0.0001667746 0.0916541000
[6] 0.5547268000

>  p.adjust(raw_p,method="BY")
[1] 1.0000000000 0.5586819525 0.0003200483 0.0004085978 0.2245525450
[6] 1.0000000000
```