Hotelling's T^2 and MANOVA

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.

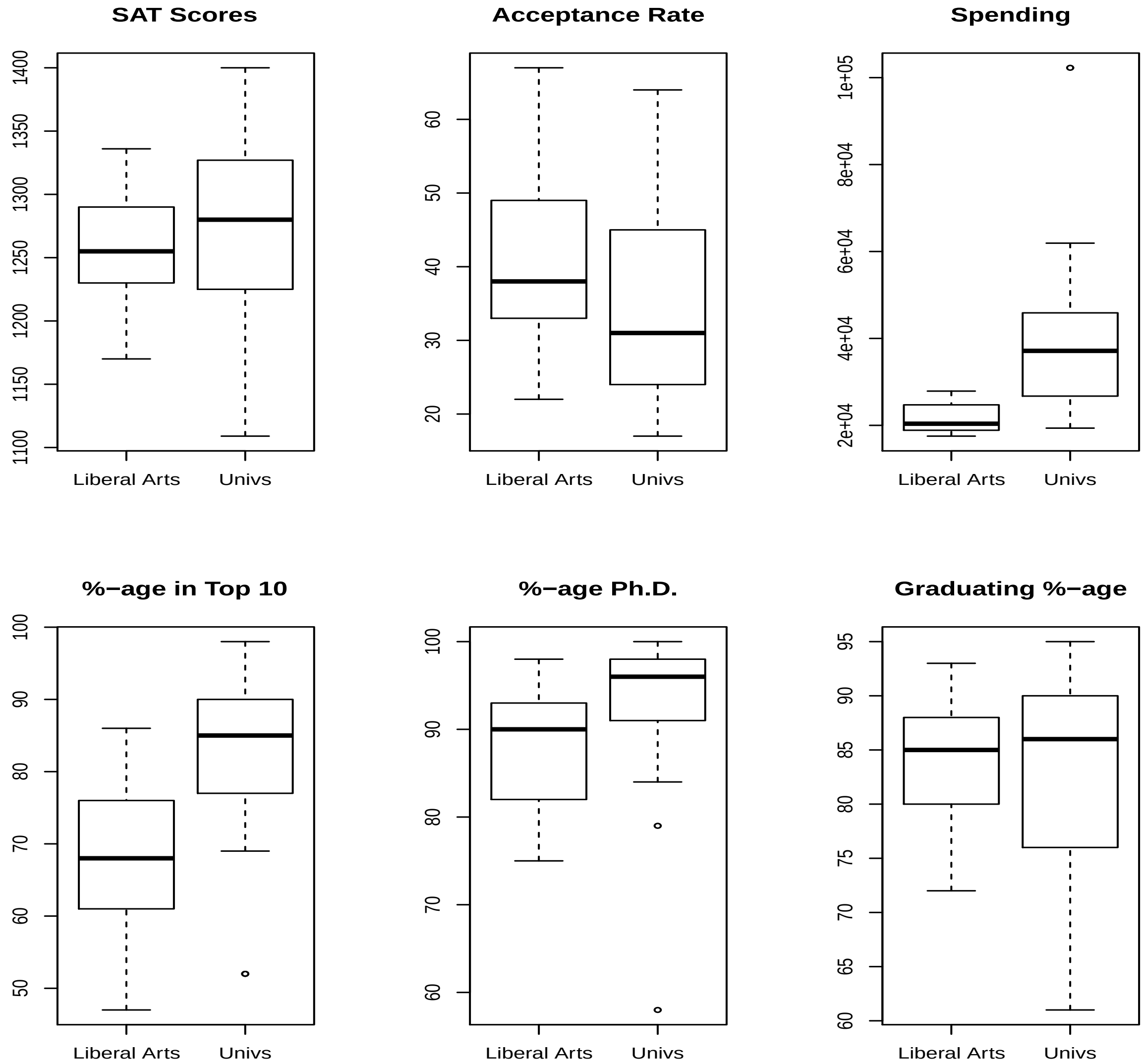
```
> d<-read.table("colleges.data",header=T)
> d
              name       it  sat acc    dps t10 phd gp
1          Amherst Lib_Arts 1315  22  26636  85  81 93
2       Swarthmore Lib_Arts 1310  24  27487  78  93 88
.......................................................
.......................................................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

> dat1<-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")))

> dat1la<-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")))

> dat1u<-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(dat1la[,1],dat1u[,1],names=c("Liberal Arts","Univs"),main="SAT Scores")
> boxplot(dat1la[,2],dat1u[,2],names=c("Liberal Arts","Univs"),main="Acceptance Rate")
> boxplot(dat1la[,3],dat1u[,3],names=c("Liberal Arts","Univs"),main="Spending")
> boxplot(dat1la[,4],dat1u[,4],names=c("Liberal Arts","Univs"),main="%-age in Top 10")
> boxplot(dat1la[,5],dat1u[,5],names=c("Liberal Arts","Univs"),main="%-age Ph.D.")
> boxplot(dat1la[,6],dat1u[,6],names=c("Liberal Arts","Univs"),main="Graduating %-age")
```




```
> x1b<-matrix(apply(dat1la,2,mean),ncol=1)
> x2b<-matrix(apply(dat1u,2,mean),ncol=1)

> t(x1b)
        [,1]  [,2]     [,3]  [,4]  [,5]  [,6]
[1,] 1256.64 40.56 21755.56 67.24 88.24 84.12
> t(x2b)
        [,1]  [,2]     [,3]  [,4]  [,5]  [,6]
[1,] 1271.28 35.12 38738.84 81.64 92.88 82.84

> Sp<-(cov(dat1la)+cov(dat1u))/2

> (43/(48*6))*(25/2)*(t(x1b-x2b)%*%solve(Sp)%*%(x1b-x2b))
         [,1]
[1,] 8.440519

> 1-pf(8.440519,6,43)
[1] 4.506612e-06
```

```
> m1<-manova(dat1~d$it)
> summary.aov(m1)

 Response sat :
            Df Sum Sq Mean Sq F value Pr(>F)
d$it         1   2679    2679  0.6852 0.4119
Residuals   48 187685    3910

 Response acc :
            Df Sum Sq Mean Sq F value Pr(>F)
d$it         1  369.9   369.9  2.1187 0.1520
Residuals   48 8380.8   174.6

 Response dps :
            Df     Sum Sq    Mean Sq F value    Pr(>F)
d$it         1 3605397494 3605397494  22.146 2.177e-05 ***
Residuals   48 7814347896  162798914

 Response t10 :
            Df Sum Sq Mean Sq F value    Pr(>F)
d$it         1 2592.0  2592.0  19.567 5.559e-05 ***
Residuals   48 6358.3   132.5

 Response phd :
            Df  Sum Sq Mean Sq F value  Pr(>F)
d$it         1  269.12  269.12  4.2034 0.04583 *
Residuals   48 3073.20   64.03

 Response gp :
            Df  Sum Sq Mean Sq F value Pr(>F)
d$it         1   20.48   20.48  0.3539 0.5547
Residuals   48 2778.00   57.88


> summary(m1)
          Df Pillai approx F num Df den Df    Pr(>F)
d$it       1 0.5408   8.4405      6     43 4.507e-06 ***
Residuals 48

> summary(m1,test="Wilks")
          Df  Wilks approx F num Df den Df    Pr(>F)
d$it       1 0.4592   8.4405      6     43 4.507e-06 ***
Residuals 48

> summary(m1,test="Hotelling")
          Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
```

```
d$it       1            1.1777   8.4405      6     43 4.507e-06 ***
Residuals 48

> summary(m1,test="Roy")
          Df    Roy approx F num Df den Df    Pr(>F)
d$it       1 1.1777   8.4405      6     43 4.507e-06 ***
Residuals 48


> t.test(d$sat[d$it=="Lib_Arts"],d$sat[d$it=="Univ"],var.equal=T)$p.value
[1] 0.4119055

> t.test(d$acc[d$it=="Lib_Arts"],d$acc[d$it=="Univ"],var.equal=T)$p.value
[1] 0.1520223

> t.test(d$dps[d$it=="Lib_Arts"],d$dps[d$it=="Univ"],var.equal=T)$p.value
[1] 2.177199e-05

> t.test(d$t10[d$it=="Lib_Arts"],d$t10[d$it=="Univ"],var.equal=T)$p.value
[1] 5.559154e-05

> t.test(d$phd[d$it=="Lib_Arts"],d$phd[d$it=="Univ"],var.equal=T)$p.value
[1] 0.04582705

> t.test(d$gp[d$it=="Lib_Arts"],d$gp[d$it=="Univ"],var.equal=T)$p.value
[1] 0.5547268


> p.adjust(c(0.4119055,0.1520223,2.177199e-05,5.559154e-05,0.04582705,0.5547268))
[1] 0.8238110000 0.4560669000 0.0001306319 0.0002779577 0.1833082000
[6] 0.8238110000
```