Problem 2

```
> mydata = read.csv("schooltest.csv")
> n1 = 13
> n2 = 12
> q = 6
> Y1 = mydata[1:13,3:8]
> Y2 = mydata[14:25,3:8]
> sigma = (t(Y1)%*%scale(Y1,scale=F) +
t(Y2)%*%scale(Y2,scale=F))/(n1+n2-2)
> Y1_bar = as.matrix(1/n1 * t(rep(1, n1))) %*% as.matrix(Y1)
> Y2_bar = as.matrix(1/n2 * t(rep(1, n2))) %*% as.matrix(Y2)
> Z = solve(sqrt(1/n1+1/n2)) %*% (Y1_bar-Y2_bar)
> Z
     SATERW25 SATERW75 SATMath25 SATMath75    ACT25 ACT75
[1,] 40.83268 21.05685  93.19459  64.69177 3.506807 1.249
> W = (n1 + n2-2) * sigma
> W
           SATERW25  SATERW75 SATMath25 SATMath75     ACT25  ACT75
SATERW25  60132.692 47720.192 66934.615 56938.462 3874.8077 2825.0
SATERW75  47720.192 40980.609 52409.615 48121.795 3018.5577 2317.5
SATMath25 66934.615 52409.615 82130.769 68523.077 4180.3846 3080.0
SATMath75 56938.462 48121.795 68523.077 63758.974 3591.5385 2770.0
ACT25      3874.808  3018.558  4180.385  3591.538  267.9423  190.5
ACT75      2825.000  2317.500  3080.000  2770.000  190.5000  145.0
> T2= (n1 + n2-2)*Z%*%solve(W)%*%t(Z)
> T2
         [,1]
[1,] 17.91281
> F = (n1+n2-1-q)/(q*(n1+n2-2))*T2
> F
         [,1]
[1,] 2.336453
> p = 1 - pf(F,q,n1+n2-1-q)
> p
           [,1]
[1,] 0.07608356
```

Since p-value=0.076 > 0.05, we can not reject the null hypothesis and conclude that there is no mean differences between the Big Ten and the Pac-12 under 0.05 significance level.

Problem 3.(a)

```
> library(msos)
> coastal = read.csv("coastalcities.csv",header = T)
> Y = as.matrix(coastal[,-c(1:3)])
> tmp = cbind(coastal$Latitude^0, coastal$Latitude^1,
coastal$Latitude^2,coastal$Latitude^3)
> x = qr.Q(qr(tmp))
> z = cbind(1,cos(1:12*2*pi/12),sin(1:12*2*pi/12))
> bsm = bothsidesmodel(x,Y,z)
> bsmh1 = bothsidesmodel.hotelling(x,Y,z,rows=4,cols=1:3)
> bsm$Beta[4,]
[1] 6.602545 2.061428 4.433489
> bsmh1$Hotelling$T2
[1] 20.83118
> bsmh1$Hotelling$F
[1] 6.429378
> bsmh1$Hotelling$pvalue
[1] 0.002230121
```

Since p-value = 0.00223 < 0.05, we reject the null hypothesis and conclude that cubic latitude terms are needed under significance level 0.05.

Problem 3.(b)

```
bsmh2 = bothsidesmodel.hotelling(x,Y,z,rows=3:4,cols=2:3)
> bsm$Beta[3:4,2:3]
           [,1]       [,2]
[1,] 0.04341989 0.03296079
[2,] 2.06142759 4.43348899
> bsmh2$Hotelling$T2
[1] 15.69799
> bsmh2$Hotelling$F
[1] 3.779146
> bsmh2$Hotelling$pvalue
[1] 0.01493896
```

Since p-value = 0.0149 < 0.05, we reject the null hypothesis and conclude that quadratic and cubic latitude terms can affect cyclic (cosine/sine) aspects under significance level 0.05.

Problem 4

```
> data("caffeine")
> Y = as.matrix(caffeine[,-1])
> x = cbind(caffeine[,1]^0,caffeine[,1]^1,caffeine[,1]^2)
> z = cbind(1,c(-1,1))
>
> cps = NULL
> for(l in (1:2)) for(p in (1:3)) {
+   pattern = matrix(0,ncol=2,nrow=3)
+   pattern[1:p,1:l] = 1
+   bsm = bothsidesmodel(x,Y,z,pattern)
+   cps = rbind(cps,c(l,p,bsm$ResidSS,bsm$Dim,bsm$Cp))
+ }
> df = data.frame(cps)
> colnames(df)[1:5]=c("l*","p*","ResidSS*","d*","Cp*")
> df
  l* p* ResidSS* d*      Cp*
1  1  1 58.64024  4 66.64024
2  1  2 57.36501  5 67.36501
3  1  3 57.29646  6 69.29646
4  2  1 51.62460  5 61.62460
5  2  2 44.36132  7 58.36132
6  2  3 44.00000  9 62.00000
```

From the result, when l*=2, p*=2, the model has lowest Cp*. Hence, keeping constant and linear terms gives lowest Cp*.