5.(a)

```
> library(msos)
> data("caffeine")
> y = as.matrix(caffeine[,-1])
> x =
cbind(1,c(rep(-1,9),rep(0,10),rep(1,9)),c(rep(1,9),rep(-9/5,10),
rep(1,9)))
> z = matrix(c(1,-1,1,1),nrow = 2,byrow = T)
> results = NULL
> models = NULL
> for(p in (1:3)) for(l in (1:2)) {
+   pattern = matrix(0,ncol = 2, nrow = 3)
+   pattern[1:p,1:l] = 1
+   bothsidesmodel = bothsidesmodel.mle(x,y,z,pattern)
+   results =
c(p,l,bothsidesmodel$Dev,bothsidesmodel$Dim,bothsidesmodel$BIC)
+   models = rbind(models,results)
+ }
> bic = models[,5]
> p = exp(-(bic-max(bic))/2)
> p = 100*p/sum(p)
> final = cbind(models,p)
> colnames(final) = c("p*","l*","Deviance","Dimension"," BIC
","probability")
> final
        p* l* Deviance Dimension     BIC  probability
results  1  1 121.0264         4 134.3552    7.316568
results  1  2 114.8222         5 131.4832   30.757055
results  2  1 119.8117         5 136.4727    2.538011
results  2  2 106.9465         7 130.2720   56.359200
results  3  1 119.7257         6 139.7189    0.500718
results  3  2 106.4904         9 136.4803    2.528448
```

5.(b)

From the result, when l*=2, p*=2, the model has the highest probability(56.36%). Hence, keeping constant and linear terms gives highest probability.

5.(c)

Difference effect in the model is captured when l* = 2, 30.757055+56.359200+2.528448 = 89.64. Hence, the chance that the difference effect is in the model is 89.64%.

5.(d)

```
> best_pattern = cbind(c(1,1,0),c(1,1,0))
> best_bothsidesmodel = bothsidesmodel.mle(x,y,z,best_pattern)
> best_bothsidesmodel$Beta
          [,1]       [,2]
[1,] 7.1607143 -0.4821429
[2,] 0.4444444 -0.5555556
[3,] 0.0000000  0.0000000
```