

# STT864 LAB1

*Nan Cao*

*January 31, 2016*

```
library(car) #####Q1 # import of dataset setwd("C:/Users/nan66/documents/msu/stt864/LAB1")
MSID<-read.table(file="MicroarraysampleIDs.txt") SNPID<-read.table(file="SNPsampleIDs.txt") CXM<-
read.table(file="ChromoXmicroarray.txt",head=T) CXSNP<-read.table(file="ChromoXsnp.txt",head=T)
```

```
# Q2
# $CSF2RA
```

```
CSF2RA<-CXM[which(CXM$IDENTIFIER=="CSF2RA")[1,],] CSF2RAdata<-CSF2RA[,c(3:49)]
ordered.CSF2RAdata<-CSF2RAdata[,order(colnames(CSF2RAdata))] colnames(ordered.CSF2RAdata)<-
MSID[,2]
```

```
#Q3 Q4 Q5
# the nearest Q
```

```
nearestSNPs<-CXSNP[c(1:2),] which(CXM$IDENTIFIER=="CSF2RA") SNPsX<-CXSNP[1,c(1:78)]
colnames(SNPsX)<-SNPID[,2] nomissingSNPs<-SNPsX[,SNPsX!="4"] finalSNPs<-nomissingSNPs[,colnames(nomissingSNPs)]
finalMicroarray<-ordered.CSF2RAdata[,colnames(ordered.CSF2RAdata)%in%colnames(nomissingSNPs)]
finalSNPs0<-finalSNPs[,order(colnames(finalSNPs))] finalMicroarray0<-finalMicroarray[,order(colnames(finalMicroarray))]
finaldataX<-rbind(finalMicroarray0,finalSNPs0)
```

```
#the nearest two SNPs
```

```
nearest2SNPs<-CXSNP[c(1:2),] SNPs2X<-CXSNP[c(1:2),c(1:78)] colnames(SNPs2X)<-SNPID[,2]
nomissingSNPs2X<-SNPs2X[, (SNPs2X[,1]!="4") & (SNPs2X[,2]!="4")] finalSNPs2X<-nomissingSNPs2X[,colnames(nomissingSNPs2X)]
finalMicroarray2X<-ordered.CSF2RAdata[,colnames(ordered.CSF2RAdata)%in%colnames(nomissingSNPs2X)]
finalSNPs02X<-finalSNPs2X[,order(colnames(finalSNPs2X))] finalMicroarray02X<-finalMicroarray2X[,order(colnames(finalMicroarray02X))]
finaldata2X<-rbind(finalMicroarray02X,finalSNPs02X)
```

```
# summary statistics????
```

```
x0<-as.numeric(finalSNPs0)-1 y0<-as.numeric(finalMicroarray0) lm0<-lm(y0~x0)
```

```
x00<-as.numeric(finalSNPs0)-1 y00<-as.numeric(finalMicroarray0) lm00<-lm(y00~x00)
```

```
x1<-as.numeric(finalSNPs0!=1) y1<-as.numeric(finalMicroarray0) lm1<-lm(y1~x1)
```

```
x2<-as.numeric(finalSNPs0!=3) y2<-as.numeric(finalMicroarray0) lm2<-lm(y2~x2)
```

```
xvec4AA<-c(1,1,0) xvec4aa<-c(1,0,-1) xvec4Aa<-c(1,1,-1) Xmat<-matrix(0,length(finalSNPs0),3) for (i in
1:length(finalSNPs0)) { Xmat[i,]<-xvec4AA (finalSNPs0[i]=="1")+xvec4Aa(finalSNPs0[i]=="2")+xvec4aa*(finalSNPs0[i]=="3")
}
```

```
betahat<-solve(t(Xmat)%Xmat)%t(Xmat)%y0 lm3<-lm(y0~Xmat[,2]+Xmat[,3])
```

```
# Confidence interval for b1-b2 or hypothesis testing for
# H0: b1=b2 vs H1: b1≠ b2
```

```

n<-length(y0) rankx<-qr(Xmat)$rank SSE<-sum(epsilonhat^2) MSE<-SSE/(n-rankx) sigmahat<-
sqrt(MSE) cvec<-c(0,1,-1) cbeta<-cvec%%solve(t(Xmat)%%Xmat)%%t(Xmat)%%y0 varcbeta<-
t(cvec)%%solve(t(Xmat)%%Xmat)%%cvec low95cbeta<-cbeta-qt(0.975,n-rankx)sigmahatsqrt(varcbeta)
upp95cbeta<-cbeta+qt(0.975,n-rankx)sigmahatsqrt(varcbeta) CI95cbeta<-c(low95cbeta,upp95cbeta) Tn<-
cbeta/(sigmahatsqrt(varcbeta)) pvalue<-2*(1-pt(abs(Tn),n-rankx))

```

```
#Q7
```

```
hist(as.matrix(CSF2RA[,c(4:49)]),main="histgram of CSF2RA")
```

```
#Q8
```

```

MA<-as.numeric(finalMicroarray02X[1,]) snp1<-as.numeric(finalSNPs02X[1,]) snp2<-as.numeric(finalSNPs02X[2,])
scatterplot(snp1|MA) scatterplot(snp2|MA) boxplot(MA_snp1) boxplot(MA_snp2)

```

```
# Q11
```

```
# Two-way ANOVA models
```

```

snp1.factor<-as.factor(snp1) snp2.factor<-as.factor(snp2) y1<-MA lm2anova<-lm(y1~snp1.factor*snp2.factor)
summary(lm2anova) anova(lm2anova)

```

```
# Q12
```

```
## Check residuals
```

```
predy<-Xmat%%solve(t(Xmat)%%Xmat)%%t(Xmat)%%y0 epsilonhat<-y0-predy plot(predy, epsilonhat)
```

```
# Q13
```

```
## Check normal distribution#Q13
```

```
qqnorm(y0)
```

```
#QQnorm & interaction plot
```

```
qqnorm(lm2anova$res) interaction.plot(snp1.factor,snp2.factor,y1)
```