**Tor Thogersen**

**DA 410 -- Multivariate Analysis -- Winter 2018**

**Project 3**

**Part 1**: The data shown in the next page were collected by Spicer et al. (1987) in an investigation of sudden infant death syndrome (SIDS). The two groups here consist of 16 SIDS victims and 49 controls. The Factor68 variable arises from spectral analysis of 24 hour recordings of electrocardiograms and respiratory movements made on each child. All the infants have a gestational age of 37 weeks or more and were regarded as full term. Enter the data in a Note Pad and save it as **SIDS.txt**

****

**Part 2:** Use R or SAS to construct the discriminant function based on all four variables and use it to classify a baby with HR = 100, BW = 3000, Factor68 = 0.3, Gesage = 40.

> ### using R built in function for Linear discriminant Analyts for SIDS data

> SIDS.dis <- lda(group ~ HR + BW + Factor68 + Gesage, data = SIDS)

> SIDS.dis

Call:

lda(group ~ HR + BW + Factor68 + Gesage, data = SIDS)

Prior probabilities of groups:

1 2

0.7538462 0.2461538

Group means:

HR BW Factor68 Gesage

1 129.2408 3437.857 0.3108163 39.38776

2 132.9500 2964.688 0.4018125 39.25000

Coefficients of linear discriminants:

LD1

HR 0.0007620693

BW -0.0012289400

Factor68 10.6336856317

Gesage -0.0333426816

> #### Use R construct the discriminant function based on all four variables and use it to classify a baby with HR = 100, BW = 3000, Factor68 = 0.3, Gesage = 40.

>

> newobs <- rbind(c(100, 3000, .3, 40))

> newobs

[,1] [,2] [,3] [,4]

[1,] 100 3000 0.3 40

> SIDSobs <- rbind(c(100, 3000, .3, 40))

> SIDSobs

[,1] [,2] [,3] [,4]

[1,] 100 3000 0.3 40

> dimnames(SIDSobs) <- list(NULL,c('HR','BW', 'Factor68', 'Gesage'))

> SIDSobs <- data.frame(SIDSobs)

> predict(SIDS.dis, newdata = SIDSobs)$class

[1] 1

Levels: 1 2

> # Posterior probabilities of this applicant being in each group:

> predict(SIDS.dis, newdata = SIDSobs)$posterior

1 2

1 0.8505037 0.1494963

> predict(SIDS.dis, newdata = SIDSobs

+ )

$class

[1] 1

Levels: 1 2

$posterior

1 2

1 0.8505037 0.1494963

$x

LD1

1 -0.002763367

> ##### Misclassification rate of LDA rule:

> SIDSgroup <-predict(SIDS.dis, SIDS, method='plug-in')$class

> table(SIDSgroup,SIDS$group)

SIDSgroup 1 2

1 46 9

2 3 7

> # cross-validation rate of LDA rule:

> correct<-rep(0,times=nrow(SIDS))

> for (j in 1:nrow(SIDS)){

+ mysidsdis <- lda(grouping=SIDS$group[-j], x=SIDS[-j, 2:5])

+ mysidspred <- predict(mysidsdis,newdata=SIDS[j,2:5])$class

+ correct[j] <- (mysidspred==SIDS$group[j])

+ }

> cv.sids.misclass <- 1-mean(correct)

> cv.sids.misclass

[1] 0.2