

Project 3

Code ▼

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

Hide

```
library(MASS)
file <- "C:\\Users\\Taterthot\\Desktop\\da 410\\a4\\babydeath.txt"
sids <- read.table(file, header = TRUE, sep = "", dec = ".")
sids
```

| Group <int> | HR <dbl> | BW <int> | Factor68 <dbl> | Gesage <int> |
|----------------|-------------|-------------|-------------------|-----------------|
| 1 | 115.6 | 3060 | 0.291 | 39 |
| 1 | 108.2 | 3570 | 0.277 | 40 |
| 1 | 114.2 | 3950 | 0.390 | 41 |
| 1 | 118.8 | 3480 | 0.339 | 40 |
| 1 | 76.9 | 3370 | 0.248 | 39 |
| 1 | 132.6 | 3260 | 0.342 | 40 |
| 1 | 107.7 | 4420 | 0.310 | 42 |
| 1 | 118.2 | 3560 | 0.220 | 40 |
| 1 | 126.6 | 3290 | 0.233 | 38 |
| 1 | 138.0 | 3010 | 0.309 | 40 |

1-10 of 65 rows

Previous 1 2 3 4 5 6 7 Next

Hide

NA

PART 2

** Use R 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.

Make sure you include clear command lines and relevant output/results with hypotheses, test result(s) and conclusion(s)/interpretation(s). **

Hide

```
sids.lda <- lda(Group~ ., data = sids, prior=c(0.5,0.5))
sids.lda
```

Call:

```
lda(Group ~ ., data = sids, prior = c(0.5, 0.5))
```

Prior probabilities of groups:

```
  1  2
0.5 0.5
```

Group means:

```
      HR      BW  Factor68  Gesage
1 129.2408 3437.857 0.3108163  40.00
2 132.9500 2964.688 0.4018125  39.25
```

Coefficients of linear discriminants:

```
      LD1
HR      0.001144709
BW     -0.001148097
Factor68 9.966188556
Gesage  -0.137896906
```

Hide

```
classobs <- rbind(c(100, 3000, 0.30, 40))
dimnames(classobs) <- list(NULL,c('HR','BW', 'Factor68', 'Gesage'))
classobs <- data.frame(classobs)
classobs
```

| HR <dbl> | BW <dbl> | Factor68 <dbl> | Gesage <dbl> |
|--------------------|--------------------|--------------------------|------------------------|
| 100 | 3000 | 0.3 | 40 |

1 row

PREDICT GROUP

Hide

```
predict(sids.lda,newdata=classobs)$class
```

```
[1] 1
Levels: 1 2
```

With the prediction we can say that a baby with HR = 100, BW=3000, Factor68=0.3, and Gesage=40 would classify with the sids of group 1.

POSTERIOR PROBABILITIES

[Hide](#)

```
predict(sids.lda,newdata=classobs)$posterior
```

```
      1      2
1 0.6570845 0.3429155
```

With the likley-hood of prior probability and new evidence we can suggest that with the current baby prediction, 65.71% of sids babies will fall into group 1, while the other 34.29% would fall into group 2. We have further analysis of misclassification and cross-validation rates below.

MISCLASSIFICATION RATE

[Hide](#)

```
group<-predict(sids.lda, sids, method='plug-in')$class
table(group, sids$Group)
```

```
group  1  2
      1 41  3
      2  8 13
```

Plug-in missclassification: $(3+8) / (41+3+8+13) = 11/65 = 0.169$.

CROSS VALIDATION

[Hide](#)

```
correct<-rep(0,times=nrow(sids) )
for (j in 1:nrow(sids) ) {
  mydis<-lda(grouping=sids$Group[-j], x=sids[-j,2:5], prior=c(0.5,0.5))
  mypred<-predict(mydis,newdata=sids[j,2:5])$class
  correct[j] <- (mypred==sids$Group[j])
}
cv.misclass <- 1-mean(correct)
cv.misclass
```

```
[1] 0.2
```

Cross-validation misclassification rate for LDA is 0.20.