Solutions to Assignment #3

1. (a) Define s_{ii} to be the diagonal elements of S. Then

$$s_{ii} = \psi_{ii} + \sum_{j=1}^{r} \ell_{ij}^2$$

Therefore,

$$\psi_{ii} = s_{ii} - \sum_{j=1}^{r} \ell_{ij}^2$$

(b) The diagonal elements of D are 0 by definition; moreover, if we define $D^* = LL^T - S$, D and D^* have the same off-diagonal elements and so

$$\sum_{i=1}^{p}\sum_{j=1}^{p} d_{ij}^2 \le \sum_{i=1}^{p}\sum_{j=1}^{p} (d_{ij}^*)^2$$

where d_{ij}^* are the diagonal elements of D^*. Now

$$D^* = LL^T - S = V\Lambda^* V^T - V\Lambda V^T$$

where Λ^* is a $p \times p$ diagonal matrix with elements $\lambda_1, \cdots, \lambda_r, 0, \cdots, 0$. Thus

$$D^* = -\sum_{j=r+1}^{p} \lambda_j \boldsymbol{v}_j \boldsymbol{v}_j^T.$$

The i-th row of D^* is

$$\boldsymbol{d}_i^T = -\sum_{j=r+1}^{p} \lambda_j v_{ij} \boldsymbol{v}_j^T$$

and

$$\boldsymbol{d}_i^T \boldsymbol{d}_i = \sum_{j=r+1}^{p} \lambda_j^2 v_{ij}^2.$$

Therefore,

$$\sum_{i=1}^{p}\sum_{j=1}^{p} (d_{ij}^*)^2 = \sum_{i=1}^{p} \boldsymbol{d}_i^T \boldsymbol{d}_i = \sum_{j=r+1}^{p} \lambda_j^2 \sum_{i=1}^{p} v_{ij}^2 = \lambda_{r+1}^2 + \cdots + \lambda_p^2$$

(c) The result in part (b) says that $LL^T + \Psi$ is a good approximation to S if $\lambda_{r+1}^2 + \cdots + \lambda_p^2$ is small. Therefore, we might choose r so that

$$\frac{\lambda_1^2 + \cdots + \lambda_r^2}{\lambda_1^2 + \cdots + \lambda_p^2}$$

is close to 1.

2. (a) The R output for the single factor model is given below:

```
> r <- factanal(~mec+vec+alg+ana+sta,factors=1)
> r
Call:
factanal(x = ~mec + vec + alg + ana + sta, factors = 1)
Uniquenesses:
  mec   vec   alg   ana   sta
0.641 0.555 0.158 0.403 0.476
Loadings:
    Factor1
mec 0.599
vec 0.667
alg 0.917
ana 0.772
sta 0.724
               Factor1
SS loadings      2.766
Proportion Var   0.553

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 8.65 on 5 degrees of freedom.
The p-value is 0.124
```

The fit of the one factor model is not too bad – the p-value for the test of the null hypothesis that the one factor model holds is 0.124.

(b) The graphical dependence model discussed in lecture suggested the only the algebra mark was directed connected to the other four marks. In the single factor model, the loading for algebra is highest, which suggests that it is the most important variable in this factor. If we were to infer that this factor is largely driven by the algebra mark (which may be a stretch) then the graphical dependence model is somewhat consistent with the single factor model. in lecture?

(c) The results for the 2 factor model using the varimax rotation are given below:

```
> r
Call:
factanal(x = ~mec + vec + alg + ana + sta, factors = 2)
Uniquenesses:
  mec   vec   alg   ana   sta
0.466 0.419 0.189 0.352 0.431
Loadings:
    Factor1 Factor2
mec 0.265   0.681
```

```
vec 0.356   0.674
alg 0.740   0.514
ana 0.738   0.322
sta 0.696   0.290
               Factor1 Factor2
SS loadings      1.774   1.370
Proportion Var   0.355   0.274
Cumulative Var   0.355   0.629

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 0.07 on 1 degree of freedom.
The p-value is 0.785
```

The 2 factor model does seem to fit somewhat better than the single factor model (although the single factor model is OK). Both the "promax" and the "none" rotation yield somehat sparser (i.e. more 0s) loadings although interpretation does not appear obvious.

3. (a) The R output is below:

```
>  r <- lda(group~FL+RW+CL+CW+BD, CV=T)
> sum(r$class!=group)
[1] 10
```

The estimated misclassification rate is $10/200 = 0.05$
(b) The R output is below:

```
> r <- lda(group~FL+RW+CL+CW+BD)
> newdata <- data.frame(FL=18.7,RW=15.0,CL=35.0,CW=40.3,BD=16.6)
> predict(r,newdata)
$class
[1] 4
Levels: 1 2 3 4
$posterior
             1            2            3         4
1 2.433387e-15 9.97961e-10 3.277912e-06 0.9999967
$x
        LD1       LD2       LD3
1 -4.393598 -2.095638 -1.433924
```

The predicted class is 4 (orange female) with a posterior probability very close to 1.
(c) The R output is below:

```
> r1 <- qda(group~FL+RW+CL+CW+BD, CV=T)
```

```
> sum(r1$class!=group)
[1] 13
> r2 <- qda(group~FL+RW+CL+CW+BD)
> predict(r2,newdata)
$class
[1] 4
Levels: 1 2 3 4
$posterior
            1            2            3 4
1 1.81922e-32 1.460731e-35 1.179623e-14 1
```

The estimated misclassification rate for QDA is $13/200 = 0.065$, slightly higher than for LDA. (This is not too surprising — the model for QDA is more complicated than that for LDA and the variability in estimating additional parameters may lead to QDA having a higher misclassification rate.) The classification for the crab with is the same as for LDA.

4. (a) First of all,

$$\begin{aligned}\sum_{i=1}^{n}(g_i - \beta_0 - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 &= \sum_{i=1}^{n}(g_i - \beta_0 - \bar{\boldsymbol{x}}^T\boldsymbol{\beta} - (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T\boldsymbol{\beta})^2 \\ &= \sum_{i=1}^{n}(g_i - \beta_0^* - (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T\boldsymbol{\beta})^2\end{aligned}$$

The latter function is minimized at $\widehat{\beta}_0^* = (g_1 + \cdots + g_n)/n$ and

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T\right)^{-1}\left(\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})g_i\right)$$

(b) This is very very tricky. We need to show that $S\widehat{\boldsymbol{\beta}} = k(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_0)$ for some constant k. First of all, some tedious algebra yields

$$\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_0 = \frac{1}{n\lambda(1-\lambda)}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})g_i$$

and so

$$\left(\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T\right)\widehat{\boldsymbol{\beta}} = n\lambda(1-\lambda)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_0)$$

In addition,

$$\begin{aligned}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T &= \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{g_i} + \bar{\boldsymbol{x}}_{g_i} - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{g_i} + \bar{\boldsymbol{x}}_{g_i} - \bar{\boldsymbol{x}})^T \\ &= (n-2)S + \sum_{i=1}^{n}(\bar{\boldsymbol{x}}_{g_i} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{g_i} - \bar{\boldsymbol{x}})^T\end{aligned}$$

Now $\bar{\boldsymbol{x}} - \bar{\boldsymbol{x}}_1 = (1-\lambda)(\bar{\boldsymbol{x}}_0 - \bar{\boldsymbol{x}}_1)$ and $\bar{\boldsymbol{x}} - \bar{\boldsymbol{x}}_0 = \lambda(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_0)$ and so

$$\sum_{i=1}^{n}(\bar{\boldsymbol{x}}_{g_i} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{g_i} - \bar{\boldsymbol{x}})^T = n\lambda(1-\lambda)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_0)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_0)^T$$

Hence

$$\left(\sum_{i=1}^{n}(\bar{\boldsymbol{x}}_{g_i}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{g_i}-\bar{\boldsymbol{x}})^T\right)\widehat{\boldsymbol{\beta}}=\left(n\lambda(1-\lambda)(\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_0)^T\widehat{\boldsymbol{\beta}}\right)(\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_0)$$

and so

$$S\widehat{\boldsymbol{\beta}}=\frac{1}{n-2}\left\{n\lambda(1-\lambda)-n\lambda(1-\lambda)(\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_0)^T\widehat{\boldsymbol{\beta}}\right\}(\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_0).$$