STA135 homework10

Zhen Zhang

March 8, 2016

8.3

```r
sigma = diag(c(2, 4, 4))
```

The eigenvalues are 2, 4 and 4, and the eigenvectors are [1, 0, 0], [0, 1, 0] and [0, 0, 1] or [1, 0, 0], [0, 0, 1] and [0, 1, 0], since the eigenvalues are identical for the second and the third one.

8.11

```r
census <- read.table("T8-5.DAT")
census$V5 <- 10 * census$V5
# covariance matrix
census_cov <- cov(census)
eigen(census_cov)
```

```
## $values
## [1] 108.271939  43.139674  31.267127   4.598098   2.347868
##
## $vectors
##             [,1]        [,2]        [,3]        [,4]        [,5]
## [1,]  0.03762881 -0.06230915 -0.03997936 -0.55553173  0.82733777
## [2,] -0.11892964 -0.24930105  0.26052476  0.76839232  0.51517455
## [3,]  0.47967291 -0.75967654 -0.43064872  0.02807896 -0.08098582
## [4,] -0.85891177 -0.31639989 -0.39364417 -0.06867379 -0.04989847
## [5,] -0.12893518 -0.50670427  0.76818907 -0.30895506 -0.20262977
```

```r
# the first two component analysis
eigen(census_cov)$vectors[, 1:2]
```

```
##             [,1]        [,2]
## [1,]  0.03762881 -0.06230915
## [2,] -0.11892964 -0.24930105
## [3,]  0.47967291 -0.75967654
## [4,] -0.85891177 -0.31639989
## [5,] -0.12893518 -0.50670427
```

```r
# total proportion explained
sum(eigen(census_cov)$values[1:2]) / sum(eigen(census_cov)$values)
```

```
## [1] 0.7984804
```

```r
census_y1 <- t(eigen(census_cov)$vectors[, 1] %*% t(census))
census_y2 <- t(eigen(census_cov)$vectors[, 2] %*% t(census))
cor(cbind(census_y1, census_y2), census)
```

```
##              V1         V2         V3        V4         V5
## [1,]  0.2124404 -0.3978987  0.6692133 -0.946998 -0.2376782
## [2,] -0.2220491 -0.5264862 -0.6690038 -0.220200 -0.5895939
```

The correlation table indicates, the first component, is mostly dependent on the first variable, goverment employment, and second component are influenced by the second, third and fifth variable. Variable 5, median home value, has little impact in the first component, though large influence on the second component.

9.1

```r
S <- matrix(c(1, 0.63, 0.45, 0.63, 1, 0.35, 0.45, 0.35, 1), nrow = 3)
total_product <- (0.63 * 0.45 * 0.35)^0.5
```

so $l_1 = \frac{0.315}{0.35} = 0.9$, $l_2 = \frac{0.315}{0.45} = 0.7$, $l_1 = \frac{0.315}{0.63} = 0.5$

```r
L <- c(0.9, 0.7, 0.5)
(L_square <- L %*% t(L))
```

```
##      [,1] [,2] [,3]
## [1,] 0.81 0.63 0.45
## [2,] 0.63 0.49 0.35
## [3,] 0.45 0.35 0.25
```

```r
(phi <- S - L_square)
```

```
##      [,1] [,2] [,3]
## [1,] 0.19 0.00 0.00
## [2,] 0.00 0.51 0.00
## [3,] 0.00 0.00 0.75
```

9.2

```r
(h1_square <- 0.9^2)
```

```
## [1] 0.81
```

```r
(h2_square <- 0.7^2)
```

```
## [1] 0.49
```

```
(h3_square <- 0.5^2)
```

```
## [1] 0.25
```

Since $Corr(Z_i, F_1) = l_{i1}$, so $Corr(Z_1, F_1) = 0.9$, $Corr(Z_2, F_1) = 0.7$, $Corr(Z_3, F_1) = 0.5$, so Z_1 has the largest correlation thus carry the most weight.

9.3

```
(L <- sqrt(1.96) * c(0.625, 0.593, 0.507))
```

```
## [1] 0.8750 0.8302 0.7098
```

Slightly different

```
1.96/3
```

```
## [1] 0.6533333
```

So it explains 65.3% total variance,