Chapter 3: Characterizing and Displaying Multivariate Data

DA 410

Marjorie Blanco

Problem 3.10 (a) and (c)

Use the calcium data in Table 3.3:

Y =

Table 1: Table 3.3. Calcium in Soil and Turnip Greens

Location Number	y1	y2	y3
1	35	3.5	2.80
2	35	4.9	2.70
3	40	30.0	4.38
4	10	2.8	3.21
5	6	2.7	2.73
6	20	2.8	2.81
7	35	4.6	2.88
8	35	10.9	2.90
9	35	8.0	3.28
10	30	1.6	3.20

$\overline{y}$ =

Table 2: Mean vector y

	mean
y1	28.100
y2	7.180
y3	3.089

(a) Calculate S using the data matrix Y as in (3.29).

S =

Table 3: Sample covariance matrix

	y1	y2	y3
y1	140.544444	49.680000	1.9412222
y2	49.680000	72.248444	3.6760889
y3	1.941222	3.676089	0.2501211

(c) Find R using (3.37).

R =

Table 4: Sample correlation matrix

	y1	y2	y3
y1	1.0000000	0.4930154	0.327411
y2	0.4930154	1.0000000	0.864762
y3	0.3274110	0.8647620	1.000000

Problem 3.14 (a) OR (b)

For the variables in Table 3.3, define z = $3y_1$ - y_2 + $2y_3$ = (3, -1, 2) y. Find $\overline{z}$ and s_z^2 in two ways:

(a) Evaluate z for each row of Table 3.3 and find $\overline{z}$ and s_z^2 directly from z_1, z_2,$\ldots$, z_{10} using (3.1) and (3.5).

Sample mean of z

z =

	mean
z1	107.10
z2	105.50
z3	98.76
z4	33.62
z5	20.76
z6	62.82
z7	106.16
z8	99.90
z9	103.56
z10	94.80

```
z_bar <- sum(z) * (1/nrow(calcium))
```

$\overline{z}$ = 83.298

```
a <- c(3,-1, 2)
s2z<- t(a) %*% as.matrix(S) %*% a
```

s_z^2 = 1048.65924

(b) Use $\overline{z}$ = a' $\overline{y}$ and s_z^2 = a'Sa, as in (3.54) and (3.55).

```
y_bar <- colMeans(calcium)
a <- matrix( c(3,-1, 2))
```

Sample mean of z

```
z_bar <- t(a) %*% as.matrix(y_bar)
```

$\overline{z} = 83.298$

Covariance matrix

```
S <- cov(calcium)
```

S =

	y1	y2	y3
y1	140.544444	49.680000	1.9412222
y2	49.680000	72.248444	3.6760889
y3	1.941222	3.676089	0.2501211

Sample variance of z_1**,** z_2**,. . . ,** z_n

```
s2z<- t(a) %*% as.matrix(S) %*% a
```

$s_z^2 = 1048.65924$

Problem 3.21 (a) and (b)

The data in Table 3.7 consist of head measurements on first and second sons (Frets 1921). Define y_1 and y_2 as the measurements on the first son and x_1 and x_2 for the second son.

Table 7: Table 3.7. Measurements on the First and Second Adult Sons in a Sample of 25 Families

y1	y2	x1	x2
191	155	179	145
195	149	201	152
181	148	185	149
183	153	188	149
176	144	171	142
208	157	192	152
189	150	190	149
197	159	189	152
188	152	197	159
192	150	187	151
179	158	186	148
183	147	174	147
174	150	185	152
190	159	195	157

y1	y2	x1	x2
188	151	187	158
163	137	161	130
195	155	183	158
186	153	173	148
181	145	182	146
175	140	165	137
192	154	185	152
174	143	178	147
176	139	176	143
197	167	200	158
190	163	187	150

(a) Find the mean vector for all four variables and partition it into $\binom{y}{x}$ as in (3.41).

```r
y <- data.frame(mean = colMeans(bones))
```

$\binom{y}{x}$ =

	mean
y1	185.72
y2	151.12
x1	183.84
x2	149.24

(b) Find the covariance matrix for all four variables and partition it into

$$\mathbf{S} = \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}$$

as in (3.42).

S =

	y1	y2	x1	x2
y1	95.29	52.87	69.66	46.11
y2	52.87	54.36	51.31	35.05
x1	69.66	51.31	100.81	56.54
x2	46.11	35.05	56.54	45.02

S_{yy} =

	y1	y2
y1	95.29	52.87
y2	52.87	54.36

S_{yx} =

	x1	x2
y1	69.66	46.11
y2	51.31	35.05

$S_{xy} =$

	y1	y2
x1	69.66	51.31
x2	46.11	35.05

$S_{xx} =$

	x1	x2
x1	100.81	56.54
x2	56.54	45.02

```
Syy <- matrix(c(cov(bones$y1, bones$y1), cov(bones$y1, bones$y2), cov(bones$y2, bones$y1), cov(bones$y2

Sxx <- matrix(c(cov(bones$x1, bones$x1), cov(bones$x1, bones$x2), cov(bones$x2, bones$x1), cov(bones$x2

Syx <- matrix(c(cov(bones$y1, bones$x1),  cov(bones$y1, bones$x2), cov(bones$y2, bones$x1), cov(bones$y

Sxy <- t(Syx)

S <- cbind(rbind(Syy, Sxy), rbind(Syx, Sxx))
```

S =

```
##       [,1]  [,2]   [,3]  [,4]
## [1,] 95.29 52.87  69.66 46.11
## [2,] 52.87 54.36  51.31 35.05
## [3,] 69.66 51.31 100.81 56.54
## [4,] 46.11 35.05  56.54 45.02
```