Quiz2...

Date:
Page:

Q1. $\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ $\rho_{13} = 0$

$$\rho_{12.3} = \frac{\rho_{12} - \cancel{\rho_{13}\rho_{23}}}{\sqrt{(1-\cancel{\rho_{13}^2})(1-\rho_{23}^2)}} = \frac{\rho_{12}}{\sqrt{1-\rho_{23}^2}} \neq \rho_{12}.$$

Therefore, <u>FALSE</u>.

Also, a lack of correlation b/w X_1 & X_3 <u>doesn't</u> imply a lack of correlation b/w X_2 & X_3.

$\therefore$ We <u>can't</u> assume $\rho_{23} = 0$.

Date:
Page: 8/11

Wk2 - Quiz2. Q.2.

$$\Sigma = \begin{bmatrix} 177 & 40 & -14 & 4 \\ 40 & 98 & -37 & -4 \\ -14 & -37 & 314 & 5 \\ 4 & -4 & 5 & 1 \end{bmatrix} \qquad \vec{\mu} = ?$$

(a) Partial coeff of X_1 & X_2, given X_3 & X_4.

$$\rho_{12\cdot3,4} = \frac{\rho_{12\cdot4} - \rho_{13,4}\,\rho_{23,4}}{\sqrt{(1-\rho^2_{13,4})(1-\rho^2_{23,4})}}$$

First subdivide Σ to get conditional dist of $\vec{X}_{(1)} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ given $\vec{X}_{(2)} = \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}$.

$$\Sigma_{11} = \begin{pmatrix} 177 & 40 \\ 40 & 98 \end{pmatrix} \quad \Sigma_{12} = \begin{pmatrix} -14 & 4 \\ -37 & -4 \end{pmatrix}$$

$$\Sigma_{21} = \begin{pmatrix} -14 & -37 \\ 4 & -4 \end{pmatrix} \quad \Sigma_{22} = \begin{pmatrix} 314 & 5 \\ 5 & 1 \end{pmatrix}$$

$$\therefore \Sigma_{12\cdot34} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$
$$= \begin{pmatrix} 177 & 40 \\ 40 & 98 \end{pmatrix} - \begin{pmatrix} -14 & 4 \\ -37 & -4 \end{pmatrix}\begin{pmatrix} 314 & 5 \\ 5 & 1 \end{pmatrix}^{-1}\begin{pmatrix} -14 & -37 \\ 4 & -4 \end{pmatrix}$$
$$= \begin{pmatrix} 157 & 54 \\ 54 & 81 \end{pmatrix} = \begin{pmatrix} \sigma_{11.34} & \sigma_{12\cdot34} \\ \sigma_{21.34} & \sigma_{22\cdot34} \end{pmatrix}$$

$$\therefore \rho_{12\cdot34} = \frac{\sigma_{12\cdot34}}{\sqrt{\sigma_{11\cdot34}}\sqrt{\sigma_{22\cdot34}}} = \frac{54}{\sqrt{157}\sqrt{81}} \approx 0.4788521 \checkmark$$

Quiz 2...

Date:
Page: 8/11

2.(b) Find multiple correlation coefficient b/w X_1 & rest of $\vec{X}$.

$$R = \sqrt{\frac{\vec{\sigma}_0^T \cdot C^{-1} \vec{\sigma}_0}{\sigma_{YY}}}$$

$$\vec{\sigma}_0 = \begin{pmatrix} 40 \\ -14 \\ 4 \end{pmatrix} \quad C = \begin{pmatrix} 98 & -37 & -4 \\ -37 & 314 & 5 \\ -4 & 5 & 1 \end{pmatrix}$$

$$= \sqrt{\frac{\begin{pmatrix} 40 & -14 & 4 \end{pmatrix} \begin{pmatrix} 98 & -37 & -4 \\ -37 & 314 & 5 \\ -4 & 5 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 40 \\ -14 \\ 4 \end{pmatrix}}{177}}$$

$$\sigma_{YY} = 177$$

$$= \sqrt{\frac{56}{177}} \approx 0.5624804.$$

(c) PCA

Steps:

1. Find eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$.
2. Find eigenvectors $\vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3, \vec{\alpha}_4$.
3. Each PC is 'just' $\vec{\alpha}_1^T \cdot \vec{X}$, $\vec{\alpha}_2^T \cdot \vec{X}$, etc.
   → i.e. a linear combination of X_1, X_2, X_3, X_4.
4. Each variance is 'just' λ_1, λ_2, etc.
5. Total variance = sum of all the eigenvalues λ_i.
6. Contribution of each PC is $\frac{\lambda_i}{\text{Sum of } \lambda_i\text{s}} \times 100\%$

7. 90% variance explained: Stop when PCs exceed 90% contribution to total variance.

8. Kaiser's rule: Only select PCs whose $\lambda_i >$ average λ.

QUIZ 2 - QUESTION 2

Amanjit Gill

November 8, 2022

```
library(ggm)
```

```
## Warning: package 'ggm' was built under R version 4.2.2
```

```
Sigma <- matrix(
    c(
        177, 40, -14, 4,
        40, 98, -37, -4,
        -14, -37, 314, 5,
        4, -4, 5, 1
    ),
    4, 4
)

# PART A

(r12.34 <- pcor(c(1,2,3,4), Sigma))
```

```
## [1] 0.4788521
```

```
# PART A manually

Sig11 <- Sigma[1:2, 1:2]
Sig12 <- Sigma[1:2, 3:4]
Sig21 <- Sigma[3:4, 1:2]
Sig22 <- Sigma[3:4, 3:4]

Sig12.34 <- Sig11 - Sig12 %*% solve(Sig22) %*% Sig21

(r12.34 <- Sig12.34[1,2] / sqrt(Sig12.34[1,1] * Sig12.34[2,2]))
```

```
## [1] 0.4788521
```

```
# PART B

C <- Sigma[2:4, 2:4]

sig0 <- Sigma[1,2:4]

sigYY <- Sigma[1,1]

(r1.234 <- sqrt(t(sig0) %*% solve(C) %*% sig0 / sigYY))
```

```
##           [,1]
## [1,] 0.5624804
```

```r
# PART C

# perform eigendecomposition
eig <- eigen(Sigma)

# extract variances and PC vectors
p <- nrow(Sigma)
sum_vars <- sum(eig$values)

vars <- c()
PCs <- list()
var_conts <- c()

for (i in 1:p) {
    vars <- c(vars, eig$values[i])
    PCs[[i]] <- eig$vectors[,i]
    var_conts <- c(var_conts, vars[i]/sum_vars*100)
}

PCs
```

```
## [[1]]
## [1]  0.14338438  0.18544166 -0.97201274 -0.01560591
##
## [[2]]
## [1] 0.92368901 0.32727505 0.19840428 0.01803896
##
## [[3]]
## [1]  0.35356243 -0.92514658 -0.12528191  0.05832539
##
## [[4]]
## [1] -0.03511620  0.05105128 -0.01146382  0.99801263
```

```r
# choose k using 90% variance explained
var_conts.c <- cumsum(var_conts)
(min(which(var_conts.c >= 90)))
```

```
## [1] 3
```

```r
# choose k using Kaiser's rule
# since Sigma isn't standardised (yet), use an alternative definition
(max(which(vars > mean(vars))))
```

```
## [1] 2
```

```r
# REDO PCA WITH STANDARDISED SIGMA

# standardised Sigma is same as correlation matrix
Sigma.std <- cov2cor(Sigma)

# now redo PCA
eig.std <- eigen(Sigma.std)

# compute variance contribution for each variable
sum_vars.std <- sum(eig.std$values)
```

```
vars.std <- c()
var_conts.std <- c()

for (i in 1:p) {
    vars.std <- c(vars.std, eig.std$values[i])
    var_conts.std <- c(var_conts.std, vars.std[i]/sum_vars.std)
}

var_conts.std
```

```
## [1] 0.40166487 0.31773379 0.20167561 0.07892574
```