Date:

Page:

01.	R	0	(χ_{\prime})	lis	10
			Xz	113	
			(X3)		

Therefore, FALSE.

Mso, a lack of correlation blu X, & X3 doesn't imply a lack of correlation blu X2 Vi X3.

. . We can't assume for = 0.

W/2- Quit 2. Q.2.

Date: 8 1

(a) Pathal well of X, K Xz, given X3 & X4.

$$\frac{2.3.4}{\sqrt{1-2.3.4}} = \frac{2.3.4}{\sqrt{1-2.3.4}} = \frac{2.3.4}{\sqrt{1-2.3.4}}$$

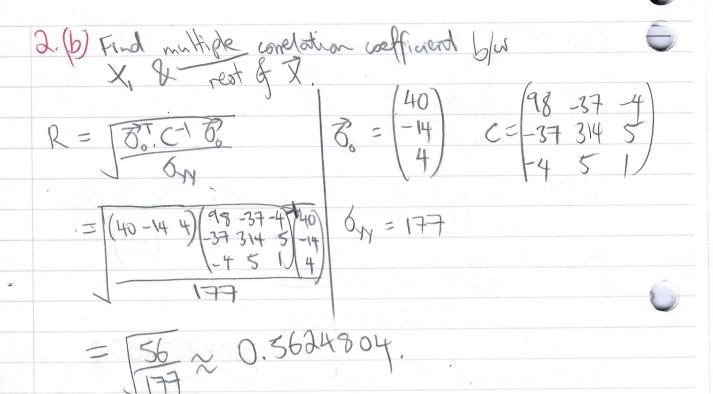
First subdivide & to get conditional dist of $X_{ij} = (x_i)$ given

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

$$2n = \begin{pmatrix} -14 & -34 \end{pmatrix} \times 2n = \begin{pmatrix} 314 & 5 \\ 4 & -4 \end{pmatrix}$$

aniz 2...





(C) .PCA

54995:

1. Find eigenvalues 2, 2, 2, 2, 24.

3. Each PC is just 2T. X, 27. X, etc.

sie a linear combination of X, X2, X3, X4.

4. Each variance is just 2, 2, etc.

5. Total variance = sum of all the eigenvalues Ai.

6. Contination of each PC is his x 100%.

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

8. Kaise's rule: Only select PCs Hose A; > average).

QUIZ 2 - QUESTION 2

Amanjit Gill

November 8, 2022

```
library(ggm)
## Warning: package 'ggm' was built under R version 4.2.2
Sigma <- matrix(</pre>
    с(
        177, 40, -14, 4,
        40, 98, -37, -4,
        -14, -37, 314, 5,
        4, -4, 5, 1
    ),
    4, 4
)
# PART A
(r12.34 \leftarrow pcor(c(1,2,3,4), Sigma))
## [1] 0.4788521
# PART A manually
Sig11 <- Sigma[1:2, 1:2]
Sig12 \leftarrow 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 \leftarrow Sig12.34[1,2] / sqrt(Sig12.34[1,1] * Sig12.34[2,2]))
## [1] 0.4788521
# PART B
C <- Sigma[2:4, 2:4]</pre>
sig0 <- Sigma[1,2:4]
sigYY <- Sigma[1,1]</pre>
(r1.234 <- sqrt(t(sig0) %*% solve(C) %*% sig0 / sigYY))</pre>
##
              [,1]
## [1,] 0.5624804
```

```
# PART C
# perform eigendecomposition
eig <- eigen(Sigma)</pre>
# extract variances and PC vectors
p <- nrow(Sigma)</pre>
sum_vars <- sum(eig$values)</pre>
vars <- c()
PCs <- list()</pre>
var_conts <- c()</pre>
for (i in 1:p) {
    vars <- c(vars, eig$values[i])</pre>
    PCs[[i]] <- eig$vectors[,i]</pre>
    var_conts <- c(var_conts, vars[i]/sum_vars*100)</pre>
}
PCs
## [[1]]
## [1] 0.14338438 0.18544166 -0.97201274 -0.01560591
## [[2]]
## [1] 0.92368901 0.32727505 0.19840428 0.01803896
## [[3]]
## [1] 0.35356243 -0.92514658 -0.12528191 0.05832539
##
## [[4]]
## [1] -0.03511620  0.05105128 -0.01146382  0.99801263
# choose k using 90% variance explained
var_conts.c <- cumsum(var_conts)</pre>
(min(which(var_conts.c >= 90)))
## [1] 3
# choose k using Kaiser's rule
{\it \# since Sigma isn't standardised (yet), use an alternative definition}
(max(which(vars > mean(vars))))
## [1] 2
# REDO PCA WITH STANDARDISED SIGMA
# standardised Sigma is same as correlation matrix
Sigma.std <- cov2cor(Sigma)</pre>
# now redo PCA
eig.std <- eigen(Sigma.std)
# compute variance contribution for each variable
sum_vars.std <- sum(eig.std$values)</pre>
```

```
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</pre>
```

[1] 0.40166487 0.31773379 0.20167561 0.07892574