| | |
|---|---|
| **Started on** | Monday, 31 October 2022, 10:24 PM |
| **State** | Finished |
| **Completed on** | Tuesday, 8 November 2022, 12:23 PM |
| **Time taken** | 7 days 13 hours |
| **Grade** | **8.00** out of 8.00 (**100**%) |

Information

## Declaration

The University regards plagiarism as a form of academic misconduct, and has very strict rules regarding plagiarism. See UNSW policies, penalties, and information to help you avoid plagiarism, as well as the guidelines in the online ELISE tutorials for all new UNSW students.

By submitting this assessment item, you declare that this assessment item is your own work, except where acknowledged, and acknowledge that the assessor of this item may, for the purpose of assessing this item:

1. Reproduce this assessment item and provide a copy to another member of the University; and/or
2. Communicate a copy of this assessment item to a plagiarism checking service (which may then retain a copy of the assessment item on its database for the purpose of future plagiarism checking)

You certify that you have read and understood the University Rules with respect to Student Conduct.

Information

Unless otherwise stated, multiple choice and True/False questions in the following section have a *guessing penalty*. This means that a correct answer gains you 1 mark, but an incorrect answer *loses* you around $1/(\#\text{ choices} - 1)$ marks. This means that if guessing randomly, your *expected* mark will be 0.

Do **not** let it prevent you from answering questions, because unless you are completely randomly guessing, you are far more likely to gain marks by answering than to lose them.

True or False? Suppose that for a random vector $\boldsymbol{X} \in \mathbb{R}^3$, correlation $\rho_{13} = 0$ (i.e., $X_1$ and $X_3$ are uncorrelated). Then the partial correlation $\rho_{12.3}$ equals the ordinary correlation $\rho_{12}$.

Select one:

- True
- False ✔

Your answer is correct.

The correct answer is: False

**Information**

## Instructions: Please read carefully

The following instructions apply to all CodeRunner questions.

1. Into the space provided, write R code that (ultimately) assigns to each of these variables the answer to the question.
   - You may either implement your answer in the space provided, or you may work elsewhere, and only provide your final answers as constants.
   - Note that not all R packages may be available in the testing environment.
2. Click "Precheck" to check that your answers are formatted correctly, and address any issues.
   - Passing tests get marked with ✔, failing tests with ✘.
   - **IMPORTANT:** If your code fails the first pre-check (i.e., code produces errors when executing), your assessment **cannot** be marked, and you may receive 0 **for the whole question**.
3. The code will be submitted when you submit the quiz.
4. When you submit the quiz, you **may** get a weird error message along the lines of "Expected 0 test results, got 1." Please disregard it.

Suppose that random vector $\boldsymbol{X} \in \mathbb{R}^4$ has mean $\boldsymbol{\mu}$ and variance-covariance matrix

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

For your convenience, you can construct it in R by running the following code:

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

## (a)

Calculate the partial correlation coefficient of $X_1$ and $X_2$ given $X_3$ and $X_4$. Store its value in variable `ans_a`.

## (b)

Calculate the multiple correlation coefficient between $X_1$ and the other elements of $\boldsymbol{X}$. Store its value in variable `ans_b`.

## (c)

Now, consider a principal component analysis of these variables (without scaling them).

### (i)

Obtain the coefficient vectors for the first two principal components of $\boldsymbol{X}$. Store them in vectors `ans_c_i_1` and `ans_c_i_2`, respectively.

### (ii)

How many principal components do you need to explain at least 90% of the variation in the data? Store your answer in variable `ans_c_ii`.

### (iii)

How many principal components should you keep according to the Kaiser's Rule? Store your answer in variable `ans_c_iii`.

### (iv)

Now, perform the scaled principal component analysis. Store the proportions of variation explained by each principal component in a vector `ans_c_iv`.

You may see a message "Your code failed one or more hidden tests" after submitting your assessment. Please disregard it. It's an artefact of some internal computation.

**Answer:** (penalty regime: 100 %)

Reset answer

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

ans_a <- 0.4788521

ans_b <- 0.5624804

ans_c_i_1 <- c(0.14338438, 0.18544166, -0.97201274, -0.01560591)
ans_c_i_2 <- c(0.92368901, 0.32727505, 0.19840428, 0.01803896)

ans_c_ii <- 3

ans_c_iii <- 2

ans_c_iv <- c(0.40166487, 0.31773379, 0.20167561, 0.07892574)


```

| | Criterion | Result | Mark | |
|---|---|---|---|---|
| ✔ | Quiz results available? | Yes | 0.00/0.00 | ✔ |

▲ Back to Top

| | | | | |
|---|---|---|---|---|
| ✓ | (b) is correct | Yes | 0.50/0.50 | ✓ |
| ✓ | (c)i PC1 is correct | Yes | 1.00/1.00 | ✓ |
| ✓ | (c)i PC2 is correct | Yes | 1.00/1.00 | ✓ |
| ✓ | (c)ii is partially correct | Yes | 0.50/0.50 | ✓ |
| ✓ | (c)ii is completely correct | Yes | 0.50/0.50 | ✓ |
| ✓ | (c)iii is at least partially correct | Yes | 0.50/0.50 | ✓ |
| ✓ | (c)iii is completely correct | Yes | 0.50/0.50 | ✓ |
| ✓ | (c)iv is at least partially correct | Yes | 0.50/0.50 | ✓ |
| ✓ | (c)iv is completely correct | Yes | 0.50/0.50 | ✓ |

Passed all tests! ✓

I will go over the details in lecture, but here is an outline of how to approach the questions. Your specific checklist will also provide some directions.

## (a)

The easiest way to solve this was to use `pcor` from the `ggm` package. Alternatively, you could regress $X_1$ and $X_2$ (individually) on $X_3$ and $X_4$ using ordinary multiple regression, and then take a correlation of the residuals.

## (b)

Again, several ways to do this. You could use one of the formulas in lecture, you could use `lm()` to fit a linear model and take the square root of $R^2$, or you could take the fitted values from that regression and correlate them with the observed values.

## (c)

You just needed to take the decomposition of $\Sigma$ here. Then, look at how many eigenvalues did you need to add up to 90% of their total, and for Kaiser's rule, look at how many were higher than the average eigenvalue. (A common mistake was to compare them to 1—that makes sense only for scaled PCA based on the correlation matrix.)

For the scaled PCA version in (iv), the R function `cov2cor()` is a convenient way to convert a covariance matrix into a correlation matrix, which is equivalent to scaling all variables to have variance of 1.

**Question author's solution (R):**

```r
# NOTE: These are for validation purposes only. These are *not* t

Sigma <- matrix(c(177, 40, -14, 4, 40, 98, -37, -4, -14, -37, 314

ans_a <- 0

ans_b <- 0.4321134

ans_c_i_1 <- c(0.84777016, -0.06933528, -0.52197521, -0.06340545)
ans_c_i_2 <- c(-0.3761048,  0.1698356, -0.7036727,  0.5783993)

ans_c_ii <- 3

ans_c_iii <- 2

ans_c_iv <- c(0.3865910, 0.3320694, 0.1780888, 0.1032507)

```

Correct

Marks for this submission: 7.00/7.00.

▲ Back to Top

Previous Activity

Next Activity

Back to Top