

Exercise 1

Task a

$$\begin{aligned} \text{Var}(X_1 + X_2 + X_3 + X_4) &= \sum_{i=1}^4 \sum_{j=1}^4 s_{ij} \\ &= 0.25 + 0.77 + 3.68 + 1.18 + 2(0.12 + 0.66 + 0.74 + 0.36 + 0.35 + 0.16) = 10.66 \end{aligned}$$

Task b

Calculate the estimated correlation matrix from the above covariance matrix

$$r_{x_1, x_2} = \frac{\text{cor}(x_1, x_2)}{\sigma_{x_1} \sigma_{x_2}}$$

$$r_{x_1, x_2} = \frac{0.12}{\sqrt{(0.25)}\sqrt{0.77}} = 0.274$$

$$r_{x_1, x_3} = \frac{0.36}{\sqrt{(0.25)}\sqrt{3.68}} = 0.375$$

$$r_{x_1, x_4} = \frac{0.16}{\sqrt{(0.25)}\sqrt{1.18}} = 0.295$$

$$r_{x_2, x_3} = \frac{0.16}{\sqrt{(0.77)}\sqrt{3.68}} = 0.392$$

$$r_{x_2, x_4} = \frac{0.35}{\sqrt{(0.77)}\sqrt{1.18}} = 0.367$$

$$r_{x_3, x_4} = \frac{0.74}{\sqrt{(3.68)}\sqrt{1.18}} = 0.355$$

Correlation matrix

$$\begin{bmatrix} 1 & 0.274 & 0.375 & 0.294 \\ 0.274 & 1 & 0.392 & 0.367 \\ 0.375 & 0.392 & 1 & 0.355 \\ 0.295 & 0.367 & 0.355 & 1 \end{bmatrix}$$

Task c

Overall, there is a positive but weak linear relationship between the pairs of all the variables. This suggest even though they all have the same scoring direction, each of the items may be measuring different things

Task d

To better understanding the relationship between the variables, we would need to have some statistics of the data which was used to arrive at the co-variance table. The distribution of values in this sample may provide some vital information about the population. It nearly impossible to describe the data since the correlation is dependent on degree of freedom within the sample which we do not know.

Exercise 2

Item	1	2	3	4	5	6	7
p	0.30	0.20	0.94	0.69	0.52	0.69	0.52
$1-p$	0.70	0.80	0.06	0.31	0.48	0.31	0.48
$p(1-p)$	0.21	0.16	0.06	0.21	0.25	0.21	0.25

$$\sum_{j=1}^m Var(X_j) = 1.35$$

$$\alpha_c = \frac{m}{m-1} \left[1 - \frac{\sum_{j=1}^m Var(X_j)}{\sigma_Y^2} \right]$$

where $Var(X) = p(1-p)$

and $m = 7$

$$\sigma_Y^2 = 3.67$$

$$\text{which implies that } \alpha_c = \frac{7}{7-1} \left[1 - \frac{1.35}{3.67} \right] = 0.74$$

Exercise 3

Part 1

Score	Item
1	Never
2	Less than once a month
3	Less than once a week but at least once a month
4	At least once a week but not everyday
5	Every day

Part 2

Score	Item
1	I have never heard of this
2	I have heard about this but I would not be able to explain what it is really about.
3	I know something about this and I could explain the general issue.
4	I am familiar with this and I would be able to explain this well.

Part 3

Gender

Score	Item
1	Female
2	Male
3	Other

Favorite colour

Score	Item
1	Blue
2	Green
3	Red
4	Yellow

Which country has the friendliest people

Score	Item
1	Denmark
2	Finland
3	Iceland
4	Norway
5	Sweden

Task a

A ratio scale is suitable for recoding the possible responses to the items in Part 1 of the questionnaire. This is because in this case, the order of the variables is important, the difference between them is clearly specified as well as providing information on the value of a true zero. Thus, the magnitude of the numbers assigned to these response items is meaningful and zero (0) clearly implies that an individual never uses his or her email for instance.

In recoding the response items in Part 2, it is proper to use an ordinal scale. An ordinal scale is used in situations where order of values matters but not their differences. If a respondent chooses "I know something about this and could explain the general issue", clearly one can determine that this respondent has more knowledge on the issue than one who chooses "I have heard about this but I would not be able to explain what it is really about". However, it is impossible to measure the amount of knowledge the first respondent has over the second one.

A nominal scale is appropriate for the items in Part 3. This is because the order on the items in this category the variables are named with not specific order. Assigning a number to the gender of an individual, his preferred colour and choice of country does not imply any level of importance. For instance, assigning 2 to Male and 1 to Female does not imply that Male is greater than Female. Thus, in this part, the numbers assigned to the items here are only tags and do not have any quantitative meaning since they do not follow a natural order of ranking.

Task b

Summary Statistics of selected Items

	Mean	Standard Deviation	Median
Use of Email	4.88	0.33	5
Nuclear Waste	2.77	0.48	3
What is your gender	1.65	0.49	2

Mean and median are measures of central tendencies.

The mean and median derived from the item “Use of Email” are 4.88 and 5 respectively on the scale of 1 to 5. This is an indication that almost all the respondents use email daily. A mean and median of 2.77 and 3 respectively on how well-informed respondents are on Nuclear Waste is an indication that most people have heard about nuclear waste and can explain what it means.

Task c

Construct a composite score for the first two parts of the questionnaire.

Person	Part 1 Comp. score	Part 2 Comp. score
1	4.14	3
2	4.29	2.86
3	4.29	2.71
4	4.29	2.86
5	3.43	2.86
6	4	2.43
7	4.57	3.29
8	3.86	3.14
9	3.43	3.71
10	4	2
11	4	3
12	4.14	3.14
13	3.29	3.43
14	3.29	2.86
15	4.71	3.14
16	4.29	3.14
17	3.71	3.14

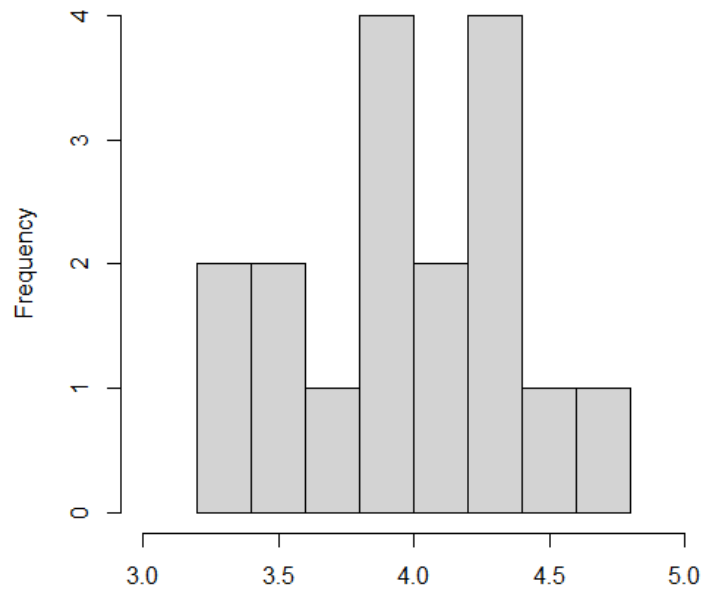
The mean as a composite score in this questionnaire seems to be unbiased as it takes all the values from the various respondents into consideration. In part 1, an average composite score of 4 is an indication that an individual uses email, internet and office applications at least once a week but not every day. In part 4, given a scale of 1 to 4, a composite score of 3 is an indication that an individual knows something about environmental issues and he or she could explain general issues on relevant topics.

Task d

Graphical representation of the item composite score

A histogram is more appropriate for the composite scores since they have continuous values.

Histogram of the Composite Score of Part 1



Histogram of the Composite Score of Part 2

