**Assignment 1**

***Task 1***

**a)**

The variance of the sum score of the items from the covariance matrix will be the sum of all values in the covariance matrix.

0.25+0.12+0.36+0.16+0.12+0.77+0.66+0.35+0.36+0.66+3.68+0.74+0.16+0.35+0.74+1.18

= 10.66

**b)**

The formula for finding the Pearson’s correlation based on covariance is:

In the covariance matrix the covariances between the four items can be found, but standards deviation, which the covariance should be divided by to find correlation, has to be calculated. As the standard deviation (σ) is the square root of the variance (σ^2) the standard deviation can be calculated by:

The variances of the items are: 0.25, 0.77, 3.68, 1.18

And the variances, square-rooted, becoming the standard deviations there for are:

The next step is to put he values into the correlation formula:

The correlation between the four items and themselves will always be 1 as they are the same. The correlation matrix will therefore look like this:

**c)**

The correlation between all of the variables is similar, meaning in the range of .27 to .375. The covariances between the variables is rather different and has a larger range.

**d)**

Further information that would be of relevance could be the number of respondents on the items.

**Exercise 2**

The formula:

First, *m* is equal to the number of items. Making the formula:

The estimate variance of the sum score is 3.67, as stated in the exercise-text.

The sum of the variance of the items can be estimated through:

The estimate for the variance of can therefore be calculated through:

With all the estimator plugged into the formula:

The result of the therefore is 0.74.

**Exercise 3**

**a)**

For the first part I chose to score the results from 1 to 5 with 5 being “Every Day” and “Never” being 1. The second part I score it 1 to 4 with four being “I am familiar with this, and I would be able to explain this well” being 4 and “I have never heard of this” as 1. Male is scored as 2, female as 1 and unknown as 0. The countries are score from 1 to 5 and the colors are score 1 to 4. The choice of scoring rule is motivated by items being similar in measuring the constructs and can therefore be similarly scored.

**b)**

I have chosen to look at summary statistics for the variables X2, X1 and Country.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Mode** |  |
| **X2** - Use the Internet to better understand issues related to, for example, your health or illnesses, financial matters, or environmental issues? | 5 | “Every day” |
| **X1** - The increases of greenhouse gases in the atmosphere. | 3 | “I am familiar with this, and I would be able to explain this well” |
| **Country** | 5 | “Norway” |

As the measurement level of each of the items is either nominal or ordinal, using a mean as a summary statistic will not be as useful as other measures. I therefore choose the mode as the summary statistic. The mode shows the response that is most frequently chosen.

**c)**

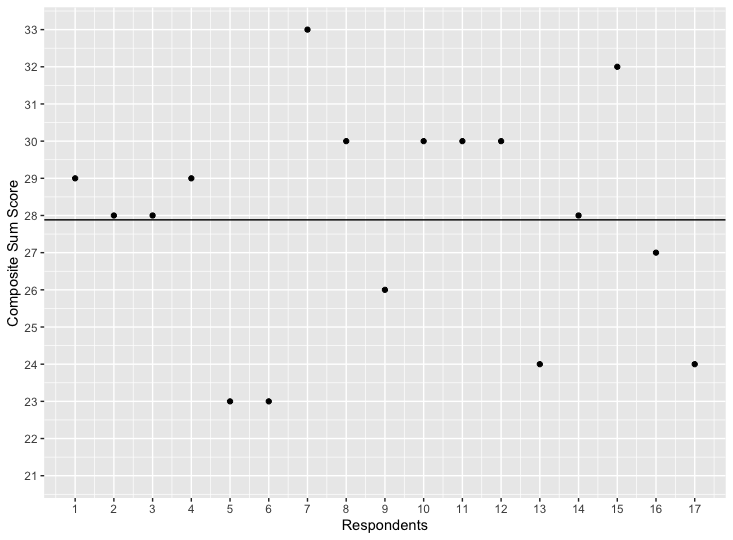
I choose to have the sum of the research unit per part of the test representing the composite score. I choose this as it gives an understanding of the level of understanding/knowledge the respondents have. The consequence of this choice is that a high sum score will indicate a high level of understanding/knowledge of the topic.

**d)**

Figure 1

*Scatterplot Showing Composite Sum Score of Respondents on question*

*regarding internet use.*

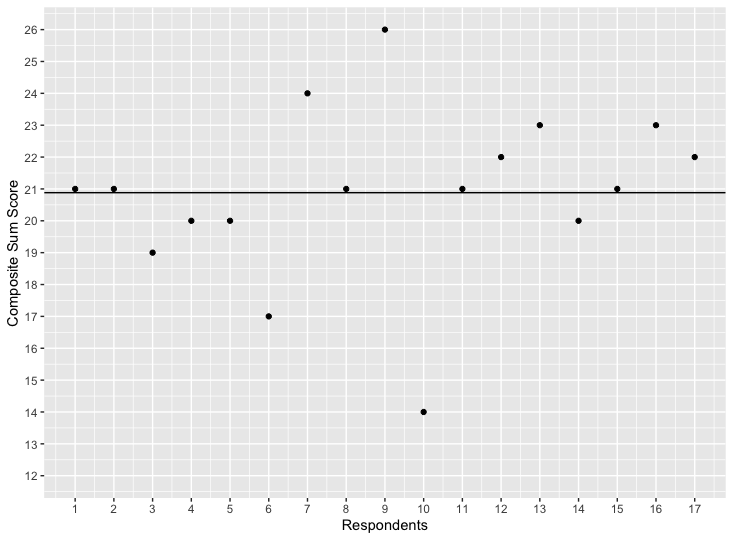
****

*Note.* A line representing the mean of the sum scores is included.

Figure 2

*Scatterplot Showing Composite Sum Score of Respondents on question*

*regarding the environment.*



*Note.* A line representing the mean of the sum scores is included.

**Motivation**

The use of scatterplot is motivated by the breadth that can be shown, all values being shown in the plot. I choose to reduce the y-axes to start from 12 and 21, so as to not have over half the plot be without data. I argue that this does not reduce the integrity of the data I show. To avoid reducing the data integrity I add breaks on every possible value on the y-axes. I choose to add a line representing the mean values of the data. This is done to make it easier for the viewer of the graph to see where the different respondents fall compared to the central tendency.