Principles of Measurement - Assignment 1

**Exercise 1**

1. The covariance values between each item range from 0.12 to 0.74 since it depends on the unit of measurements. The correlation coefficients, on the other hand, are units independent, which show that if the four items are linearly related to each other, there is mild positive relationship between each other around 0.3.

* We need further information about the items themselves, such as the type of variables (quantitative or categorical) and the measurement scales.
* We can then calculate the mean, median, and mode of each item across individuals to see the central tendency and find out the average item score and most people’s option.
* We can also use scatter plots to identify the relationship between each item. Since we cannot tell if two variables have linear relationship or not based on the
* By calculation the variance of total score and the variance of the score of each item, we can identify the internal consistency between four items.

**Exercise 2**

**Exercise 3**

1. The first two parts include ordered-category items. Each item has at least four response options that have a natural order. I would like to use Likert scales to assign integer scores to the response categories.

|  |  |  |  |
| --- | --- | --- | --- |
| Never | 0 | I have never heard of this | 0 |
| Less than once a month | 1 | I have heard about this but I would not be able to explain what it is really about | 1 |
| Less than a week but at least once a month | 2 | I know something  about this and could explain the general issue | 2 |
| At least once a week but not everyday | 3 | I am familiar  with this and I would be able to explain this well | 3 |
| Every day | 4 |  |  |

Part 3 has three unordered categorical items which do not possess a natural order. In this case there is no rational way to score each item. For the gender item, there is no ‘other’ option chosen by respondents, we can assign 0 to male, 1 for female, and 2 for other. Then we can compare the attributes difference between male and female. While for the color preference and country preference, it is meaningless to assign number to each response options unless we have a study design to investigate one specific option. For exam, if we are interested in people who favor green color, the response can be coded as 1 for green and 0 for the remaining categories, while the code cannot be treated as a score.

1. (1) Summary statistic of real-time discussions on the internet (Part 1, Q7)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Min | 1st Qu | Median | Mean | 3rd Qu. | Max | Mode | s.d. |
| 0 | 1 | 3 | 2.588 | 4 | 4 | 4 | 0.56 |

The mode represents most people (6 out of 17) have online conference or chat every day. The average frequency of using this function is at least once a week. The standard deviation is small showing that there is not too much variation among the data.

(2) Summary statistic of knowledge on extinction of plants and animals (Part 2, Q6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max | Mode |
| 1 | 2 | 2 | 2.118 | 2 | 3 | 2 |

More than half of the sampled people (11 of 17) have knowledge of extinction of plants and animals and they can explain the general issue. No one has never heard of this issue, while there are 4 persons are very familiar with this topic.

(3) For the item chosen in part 3, a frequency table can be created.

|  |  |  |  |
| --- | --- | --- | --- |
| Blue | Green | Red | Yellow |
| 9 | 5 | 1 | 2 |

The mode for the item is color ‘blue’, so we can know most people like this color. For unordered categorical items, it is meaningless to calculate the mean or median.

1. The first composite score is constructed by adding up all 7 item scores in part 1, which measures people’s usage of internet in their daily life. Every item is given equal weight because it measures one function that people can use from internet. The total score, therefore, is the summation of all items range from 0 to 28. For those who have higher scores, indicating they use internet for different functions more frequent than others.

The second composite score for everyone is constructed by taking the average score of all 7 item scores in part 2, which measures people’s awareness of environment issues. The average score is ranging from 0 to 3. If a person has higher score, representing she or he has better understanding about environment issues in general.

1. From the figure below we can see the distribution of composite scores for internet usage. Basically speaking, people use internet for some reasons at a relatively high frequency, which is consistent with the summary statistics of one specific item measured above. Due to the sample limit, the composite scores are not distributed normally. However, because the target of sample is from one specific group of people who study the same program, generality is defeated here, that is, any patterns found are not applicable to another group with different backgrounds.

Chart, histogram

Description automatically generated

Boxplots are used to demonstrate the distribution of average scores measuring how people aware different environment issues. From the graph, there is no significant difference between male and female towards their knowledge on this topic. However, the range of awareness is larger in male group, and there is one outlier.

Chart, box and whisker chart

Description automatically generated