**Practice RAT for Week 07, Monday**

**Answers at the end**

|  |  |
| --- | --- |
| 1 | Imagine you were asked to compute correlations for each of the following datasets:  Macintosh HD:Users:kiaangloo:Desktop:Screen Shot 2015-03-05 at 12.39.22 AM.pngMacintosh HD:Users:kiaangloo:Desktop:Screen Shot 2015-03-05 at 12.39.07 AM.pngMacintosh HD:Users:kiaangloo:Desktop:Screen Shot 2015-03-05 at 12.39.16 AM.png  A  C  B  Which of the following options provides the most plausible correlation for each dataset? |
|  |
| (a) | A: 0.94, B: -0.77, C: 0.59 |
| (b) | A: 1, B: 0.8, C: 0 |
| (c) | A: -0.8, B: 0.77, C: -0.6 |
| (d) | A: 0.94, B: 0.83, C: 0.5 |

|  |  |
| --- | --- |
| 2 | As part of the course review process, the QR professors correlated students’ mid-term scores with their SAT scores for math. After running the correlation analysis, the QR professors discovered that Canvas had systematically miscounted students’ mid-term scores. This issue was rectified by adding 2 marks to each person’s total scores. (Note: This is a fictitious scenario!)  The correlation between mid-term scores and SAT scores will: |
|  |
| (a) | Increase by 2. |
| (b) | Increase by an amount proportionate to 2. |
| (c) | Decrease by 2. |
| (d) | Remain unchanged. |

|  |  |
| --- | --- |
| 3 | Michael is a researcher who has collected data on the age, weight, and income of 1000 healthy adults aged between 20 and 60. He found that age and weight had a correlation of 0.5, and income and weight had a correlation of -0.5.  What can he conclude (assuming that the conditions for using a correlation were met)? |
|  |
| (a) | There is an association between age and weight, but no association between income and weight. |
| (b) | The association between age and weight is stronger than the association between income and weight. |
| (c) | The numerical strength of association between age and weight and that of income and weight cannot be compared. |
| (d) | The association between age and weight is as strong as the association between income and weight. |

|  |  |
| --- | --- |
| 4 | Researchers from the West Virginia State University, conducted an observational study about the relationship between the use of chewing tobacco and longevity. They recruited a cohort of 3000 West Virginians aged 50–59 in 1980 and asked them to estimate how much chewing tobacco they used during a week. The researchers then checked the Registry of Births and Deaths every year to see which of their cohort had died. In 2020, they used that data to calculate the correlation of age at death with chewing tobacco use and obtained a correlation coefficient r = - 0.6.  Which of the following statements best describes their findings? |
|  |
| (a) | Chewing tobacco use reduces the risk of death |
| (b) | Chewing tobacco use increases the risk of death |
| (c) | There is an association between chewing tobacco use and age at death |
| (d) | Chewing tobacco use is independent of mortality |

|  |  |
| --- | --- |
| 5 | Imagine you are asked to conduct an analysis of a series of numeric variables stored in a data frame called X. As part of that analysis, you are asked to calculate the correlations for each pair of variables stored in X. After importing and inspecting X, however, you realise that the columns occasionally contain missing values. You decide that when calculating the correlations, you wish to exclude observations with any missing values and only retain observations with a complete set of values.  Which of the following commands in R will produce the desired output? |
|  |
| (a) | correl(X, use = “complete.obs”) |
| (b) | correl(X, use = “pairwise.complete.obs”) |
| (c) | cor(X, use = “complete.obs”) |
| (d) | cor(X, use = “pairwise.complete.obs”) |

**Answers:**

1. ***Answer: A.*** *Graphs A and C must have a positive sign, with a stronger correlation for A than for C. Graph B has to have a negative sign.*
2. ***Answer: D.*** *A uniform shift of marks changes the mean of the marks, but not the distance of any mark from the mean. Z-scores are not affected, and therefore it will not affect the correlation.*
3. ***Answer: D.*** *The strength of an association is not affected by the sign. Numerically, age and weight have as strong a linear relation as income and weight.*
4. ***Answer: C.*** *The negative correlation between chewing tobacco use and age at death itself does not imply a causal relationship between chewing tobacco use and longevity. It would thus be incorrect to conclude from the correlation alone that chewing tobacco increases the risk of death.*
5. ***Answer: C.*** *There is no command* correl() *in base R (though this is the command you would use to calculate a correlation in MS Excel). The correct option for the argument* use *is* “complete.obs”*.*