Day 4 assessment

1.Children of three ages are asked to indicate their preference for three photographs of adults.

Do the data suggest that there is a significant relationship between age and photograph preference? What is wrong with this study?

Photograph:

Age of child A B C

5-6 years: 18 22 20

7-8 years: 2 28 40

9-10ears: 20 10 40

(i) Use cov() to calculate the sample covariance between B and C.

(ii) Use another call to cov() to calculate the sample covariance matrix for the preferences.

(iii)Use cor() to calculate the sample correlation between B and C.

(iv)Use another call to cor() to calculate the sample correlation matrix for the preferences.

CODE:

```
> photographs <- matrix(c(18, 2, 20, 22, 28, 10, 20, 40, 40), nrow = 3, by
row = TRUE)
> cov(photographs["B",], photographs["C",])
> cov(photographs)
     [,1]     [,2]     [,3]
[1,]    4  26.0000 -10.0000
[2,]   26 377.3333 143.3333
[3,]  -10 143.3333 233.3333
```

2. Gopal travels daily from his house located at santhom to his office located at OMR road by his car and he wants know how much time he spends on travel. He does record the time taken

to reach the off from his home for about a week and has the following value: 46.45, 34.34, 30,

56,12,44.67,43,36.45,48, 35.67, 37.23,32.7,39.20,40.01,45.02,34.12,33.19. Help Gopal to

analyse the time data using skewness and kurtosis and give your interpretation.

CODE:

```
library(moments)

times <- c(46.45, 34.34, 30, 56.12, 44.67, 43, 36.45, 48, 35.67, 37.23, 32.7, 39.2, 40.01, 45.02, 34.12, 33.19)

skewness(times)

kurtosis(times)
```

3(i). Generate a sample of 5000 random numbers and create a normal distribution with a mean value of 70 and respectively fix the Standard deviation to

(ii). Calculate the skewness of the normal distribution along with kurtosis and interpret your results.

(iii)Write suitable R code to compute the median of the following values.

12,7,3,4.2,18,2,54, -21,8, -5

(iv) A student recorded her scores on weekly math quizzes that were marked out of a possible 10 points. Her scores were as follows:

8, 5, 8, 5, 7, 6, 7, 7, 5, 7, 5, 5, 6, 6, 9, 8, 9, 7, 9, 9, 6, 8, 6, 6, 7

What is the mode of her scores on the weekly math quizzes?

CODE:

```
set.seed(123) # set seed for reproducibility

sample <- rnorm(5000, mean = 70, sd = 10)

library(moments)

skewness(sample)

kurtosis(sample)

values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

median(values)

library(DescTools)
```

```
> scores <- c(8, 5, 8, 5, 7, 6, 7, 7, 5, 7, 5, 5, 6, 6, 9, 8, 9, 7, 9, 9, 6, 8, 6, 6, 7)
> mode(scores)
[1] "numeric"
```

4. The following table of grouped data represents the weight (in kg) of 100 students. Calculate

the mean weight for a student.

Weight (pounds) Number of Student

21kg 8

30kg 25

56kg 45

73kg 18

110kg 4

CODE:

4.To find the mean weight for a student, we need to calculate the weighted mean. We can do this by multiplying each weight by its corresponding frequency, summing the results, and dividing by the total number of students.

Using the given data, we first need to convert the weights from kilograms to pounds:

Weight (kg) Weight (lb) Number of Students

21 46.3 8

30 66.1 25

56 123.5 45

73 160.9 18

110 242.5 4

Now we can calculate the weighted mean as follows:

weighted mean = (46.38 + 66.125 + 123.545 + 160.918 + 242.5*4) / (8+25+45+18+4)

= 101.4 kg

Therefore, the mean weight for a student is approximately 101.4 kg.

5. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in

increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40,

45, 46, 52, 70. Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?

CODE:

5.Yes, we can find the first quartile (Q1) and the third quartile (Q3) of the data.

There are 26 data points, so the median is the average of the 13th and 14th value, which is:

median = (25 + 25) / 2 = 25

To find the first quartile (Q1), we need to find the median of the data below the overall median. These values are:

13, 15, 16, 16, 19, 20, 20, 21, 22, 22

The median of this set is:

Q1 = (20 + 20) / 2 = 20

To find the third quartile (Q3), we need to find the median of the data above the overall median. These values are:

30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70

The median of this set is:

Q3 = (40 + 45) / 2 = 42.5

Therefore, the first quartile (Q1) is approximately 20 and the third quartile (Q3) is approximately 42.5.

6. Suppose a hospital tested the age and body fat data for 18 randomly selected adults with the

following result

age 23 23 27 27 39 41 47 49 50

%fat 9.5 26.5 7.8 17.8 31.4 25.9 27.4 27.2 31.2

age 52 54 54 56 57 58 58 60 61

%fat 34.6 42.5 28.8 33.4 30.2 34.1 32.9 41.2 35.7

a. Calculate the standard deviation of age and %fat.

b. Calculate the Variance of age and %fat for the above dataset.

CODE:

```
age <- c(23, 23, 27, 27, 39, 41, 47, 49, 50, 52, 54, 54, 56, 57, 58, 58, 6
0, 61)
> fat <- c(9.5, 26.5, 7.8, 17.8, 31.4, 25.9, 27.4, 27.2, 31.2, 34.6, 42.5,
28.8, 33.4, 30.2, 34.1, 32.9, 41.2, 35.7)
> sd(age)
[1] 13.21862
> sd(fat)
[1] 9.254395
> var(age)
[1] 174.732
> var(fat)
[1] 85.64382
```

7. Find the H.M of the values 20.0, 35.5, 40.0 and 37.0 with their respective weights 7.0, 8.5,

3.0 and 6.0

CODE:

7)To find the harmonic mean (H.M) of values with respective weights, we can use the formula:

H.M = (sum of weights) / (sum of weights / values)

So, for the given values and weights, we have:

Values: 20.0, 35.5, 40.0, 37.0

Weights: 7.0, 8.5, 3.0, 6.0

Sum of weights = 7.0 + 8.5 + 3.0 + 6.0 = 24.5

Sum of weights / values = (7.0/20.0) + (8.5/35.5) + (3.0/40.0) + (6.0/37.0) = 0.3507042

H.M = (sum of weights) / (sum of weights / values) = 24.5 / 0.3507042 = 69.92088

Therefore, the harmonic mean of the values 20.0, 35.5, 40.0, and 37.0 with their respective weights 7.0, 8.5, 3.0, and 6.0 is approximately equal to 69.92088.

8. The demand for a product on each of 20 days was as follows, (in units). 3, 12, 7, 17, 3, 14,

9, 6, 11, 10, 1, 4, 19, 7, 15, 6, 9, 12, 12, 8 Calculate arithmetic mean

CODE:

8.To calculate the arithmetic mean of the demand for the product on each of the 20 days, we can use the formula:

Arithmetic Mean = (sum of all values) / (number of values)

In this case, we have:

Demand: 3, 12, 7, 17, 3, 14, 9, 6, 11, 10, 1, 4, 19, 7, 15, 6, 9, 12, 12, 8

Number of values: 20

Sum of all values: 3 + 12 + 7 + 17 + 3 + 14 + 9 + 6 + 11 + 10 + 1 + 4 + 19 + 7 + 15 + 6 + 9 + 12 + 12 + 8 = 180

Arithmetic Mean = 180 / 20 = 9

Therefore, the arithmetic mean of the demand for the product on each of the 20 days is 9 units.