# CAT QUANTITATIVE METHODS

**Student Name:** Kevin Omenda

**Student number:** L123/25915/2022

**Course:** QUANTITATIVE METHODS

**Lecturer:** David Mwadzaya

**Date: 05/03/2023**

**Question One (30 marks) (Compulsory)**

Consider the following scores marks of students in a common statistics exam from two group of students in both the day and evening classes.

Required:

a) Classify this information into six (6) equal classes using the exclusive method with a class interval of 15 and one of the upper-class limits of either of the distribution is 50. Ensure to sure your tally marks
(**8 Marks**)

The first step in this question was to find the range of our dataset where we subtracted the smallest value which was 8 from the largest value 94 in the Scores by Evening Students.

Therefore, range = 94 – 8 = 86

After finding the range we then divided it with the required number of classes which is six to get an approximate class interval value.

Therefore, class interval = 86 / 6 giving us an interval of 14.33

We however rounded off by a whole number since the question required a class interval of 15.

The third step now involved getting our lower and upper limits where we subtracted 0.5 * the class interval from our smallest value in the column and rounded it off to the nearest.

Lower limit is therefore = 8 (our smallest value) – 0.5 * 15 to get 0.5 rounded to the next integer to become 1.

Exclusive method basically means that the upper limit of the previous class excluded.

Table 1.1: Scores table by day class Students

| Scores by Day Class Students | | | |
|---|---|---|---|
| Class | Frequency | Tally | Percentage |
| 1 - 15 | 4 | \|\|\|\| | 7% |
| 16 - 30 | 8 | \|\|\|\|\|\|\|\| | 13% |
| 31 - 45 | 16 | \|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\| | 27% |
| 46 - 60 | 14 | \|\|\|\|\|\|\|\|\|\|\|\|\|\| | 23% |
| 61 - 76 | 11 | \|\|\|\|\|\|\|\|\|\|\| | 18% |
| 76 - 100 | 7 | \|\|\|\|\|\|\| | 12% |
| **Total** | 60 | | |

Table 1.2: Scores table by Evening Class student

| Scores by Evening Class Students | | | |
|---|---|---|---|
| Class | Frequency | Tally | Percentage |
| 1 - 15 | 3 | \|\|\| | 7% |
| 16 - 30 | 5 | \|\|\|\|\| | 13% |
| 31 - 45 | 16 | \|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\| | 27% |
| 46 - 60 | 15 | \|\|\|\|\|\|\|\|\|\|\|\|\|\|\| | 23% |
| 61 - 76 | 12 | \|\|\|\|\|\|\|\|\|\|\|\| | 18% |
| 76 - 100 | 9 | \|\|\|\|\|\|\|\|\| | 12% |
| **Total** | 60 | | |

b) Based on the values of the median, lower quartile and upper quartile comment as to which group had the better performance in each clearly explain your answer (**10 Marks**)

To get the quartiles for the day students we first sorted the entire dataset in ascending order:

13, 14, 15, 18, 18, 20, 20, 24, 25, 30, 30, 34, 35, 35, 38, 38, 38, 40, 40, 40, 42, 45, 45, 45, 47, 49, 50, 50, 55, 58, 59, 60, 60, 60, 61, 63, 64, 64, 65, 65, 68, 69, 70, 70, 72, 79, 85, 88, 90, 90, 92, 94

We the added the two middle values (48+ 49) and divided by two to get our median value which was 48.5.

For Q1 we got by adding the two values in the lower half and dividing them by two

Q1 = (35 + 36)/2 = 35.5

Q3 = (64 + 64)/2 = 64

Therefore, the quartiles for the day student are Q1 = 35.5, Q2 = 48.5 and Q3 = 64

To get the quartiles for the evening students we first, ordered the dataset in ascending order:

8 10 14 16 18 19 20 28 33 33 34 34 35 36 38 40 40 40 42 45 45 45 46 48 48 49 49 50 50 53 55 60 60 60 64 65 65 65 68 68 68 70 77 78 80 85 88 90 90 94 94

Next, we looked for the median (Q2) vale. In this case, the dataset has an even number of observations, so we took the average of the two middle values: (49 + 50) / 2 = 49.5

To find the lower quartile (Q1), we took the median of the lower half of the dataset. To do this, we split the dataset into two halves: 8 10 14 16 18 19 20 28 33 33 34 34 35 36 38 40 40 40 42 45 45 45 46 48 48 49. The median of this half is (38 + 40) / 2 = 39, so Q1 = 39.

To find the upper quartile (Q3), we took the median of the upper half of the dataset. To do this, we split the dataset into two halves: 50 53 55 60 60 60 64 65 65 65 68 68 68 70 77 78 80 85 88 90 90 94 94. The median of this half is 68.

Therefore, the quartiles for the evening student's dataset are Q1 = 39.5, Q2 = 49.5, and Q3 = 68.

In conclusion, we observe that the evening students had a higher third quartile score as compared to that of the day students. This means that evening students had the higher percentage of students who scored high grades as compared to day students. The evening students also had a higher range in terms of score as compared to that of the day students.

c) Based on Pearson's first coefficient of skewness, compare and comment on skewness of the two groups, explaining what this implies with reference to the performance of each group. (**12 Marks**)

We used the Pearson formula below to check the skewness:

$$Sk_2 = \frac{3(\bar{X} - Md)}{s}$$

Where $\bar{X}$ = the mean, Md = the median and s = the standard deviation for the sample.

To get our standard deviation we used the excel function " =stdev.s() "

Skweness_day = 3(49.65 – 48.5)/21.5 = 0.1948

Skewness_evening = 3(52.08 – 49.5) / 21.52 = 0.1198

We observe a positive skewness in the two groups where the skewness value in both groups is greater than zero. Even though the values are almost the same, we observe a lower value in the evening class data. This indicates that students who attend class during the day have a more pronounced skewness towards higher scores as compared to those of evening class. This in terms of performance implies that the day students have a larger proportion of students who performed highly as compared to the evening students.

**QUESTION TWO (20 Marks)**

a. As part of an investigation into levels of overtime working, a company decides to tabulate the number of orders received weekly and compare this with the total weekly overtime worked to give the provided dataset.

**Requirement:**

Using the method of least squares obtain the regression equation of total overtime on orders received and predict the level of total overtime necessary for 100 orders (**8 Marks**)

The first step was to calculate the means of the variables orders received and that of the total overtime.

Oders_received_mean = (83 + 22 + 107 + 55 + 48 + 92 + 135 + 32 + 67 + 122) / 10 = 76.3

Total_overtime_mean = (38 + 9 + 42 + 18 + 11 + 30 + 48 + 10 + 29 + 51) / 10 = 28.6

Then we looked for the deviations in the variable orders received and the total overtime, which we again subtracted from their mean values.

Orders_received_deviation:

= [83, 22, 107, 55, 48, 92, 135, 32, 67, 122] – 76.3

From that we get [6.7, -54.3, 30.7, -21.3, -28.3, 15.7, 58.7, -44.3, -9.3, 45.7]

Total_overtime_deviations:

[38, 9, 42, 18, 11, 30, 48, 10, 29, 51] - 28.6

From that we get: [9.4, -19.6, 13.4, -10.6, -17.6, 1.4, 19.4, -18.6, 0.4, 22.4]

Then the third step involved multiplying the deviations of the two variables and summing them up.

Sum_product_deviation = (10.9 * 9.4) + (-50.1 * -19.6) + (34.9 * 13.4) + (-17.1 * -10.6) + (-24.1 * -17.6) + (19.9 * 1.4) + (62.9 * 19.4) + (-40.1 * -18.6) + (-5.1 * 0.4) + (49.9 * 22.4)

= 5267.2

Then in the fourth step we calculated the sum of the squared deviation for the variable orders received.

Sum_squared_deviations = (10.9) ^2 + (-50.1) ^2 + (34.9) ^2 + (-17.1) ^2 + (-24.1) ^2 + (19.9) ^2 + (62.9) ^2 + (-40.1) ^2 + (-5.1) ^2 + (49.9) ^2

= 13020.09

To get the regression coefficient we then divided the sum of product deviation by the sum of the squared deviation for orders received

Regression_coefficient = Sum_product_deviation / Sum_squared_deviations

Regression_coefficient = 5267.2/ 13020.09

Regression_coefficient = 0.405

Intercept = mean of dependent variable – regression coefficient * mean of independent variable

Intercept = 28.6 – 0.405 * 76.3 = -2.267

Finally, we write our regression equation as indicated below and which can be used to predict the Total_overtime

Total_overtime = a + b * Orders received

From the above equation, the Total_overtime is our dependent variable, the Orders received is our independent variable, a is the intercept value which can be the value of Total_overtime in the case orders received is equal to zero or the intercept. The b represents the Regression_coefficient that we have calculated.

Hence, in a situation we want to predict the Total overtime that will be required for 100 orders we can predict using the equation below:

Total_overtime = a + b * Orders received

‘a = -2.267

‘b = 0.405

Total_overtime = -2.267 + 0.405 * Orders received

Total_overtime = -2.267 + 0.405 * 100 = **38.23 hours**

b. The provided table gives the distribution of wages in kE in three branches of Keses factory.

   i) Based on the mean wage, which branch pays the higher average wage?

Table 2.1: Number of workers and wages frequency table

| Daily wages Kes | Number of Workers | | |
|---|---|---|---|
| | Branch A | Branch B | Branch C |
| 10-14 | 14 | 15 | 12 |
| 15-19 | 18 | 22 | 20 |
| 20-24 | 20 | 26 | 32 |
| 25-29 | 23 | 10 | 11 |
| 30-34 | 10 | 12 | 10 |

In finding the mean wage for the different branches we will use the weighted mean formula below:

$$\textbf{Weighted Mean} = \frac{\sum_{i=1}^{n}(xi * wi)}{\sum_{i=1}^{n} wi}$$

Where W represents the weight

∑ represents the summation

And x s the value

The first step in our calculation involved finding the mean wage for each class under the branches.

We started by multiplying the number of works in each class by the class midpoint and then summing all of them up for each branch.

Branch A = (12 x 14) + (17 x 18) + (22 x 20) + (27 x 23) + (32 x 10) = 1004

Branch B = (12 x 15) + (17 x 22) + (22 x 26) + (27 x 10) + (32 x 12) = 974

Branch C = (12 x 12) + (17 x 20) + (22 x 32) + (27 x 11) + (32 x 10) = 1120

After multiplying with the mid-points, we then take the sum of all workers in each branch and divide with the number we got in the previous calculation

Branch_A_Mean = 1,004 / (14 + 18 + 20 + 23 + 10) giving us a value of 14.18

Branch_B_Mean = 974 / (15 + 22 + 26 + 10 + 12) giving us a mean of 15.67

Branch_C_Mean = 1,120 / (12 + 20 + 32 + 11 + 10) giving us a mean value of 13.176

To now get the weighted wage average for all employees we multiply the mean wage for each branch to the sum of workers in the given branch and then add for the branches before finally dividing with the sum of all workers in the company.

Weighted_Average = (((14.18 x 85) + (15.67 x 85) + (13.176 x 85)) / (85+ 85 + 85))

1205.3+1331.95+1119.96 /255 = 14.342

In conclusion, we observe that the Branch which we named Branch_B or the second column in the number of workers is the one that pays the highest average wage to its employees with a mean wage of Kes 15.67 then followed by the column we named Branch_A with a mean wage of 14.18 Kes and finally the column we named Branch C with a mean wage of Kes 13.176. In overall, the factory paid its workers an average salary of KES 14.342.

ii) Which of the three branches has greater variability in wages? (**12 Marks**)

Branch A mean = 14.18

Branch B mean = 15.67

Branch C mean = 13.176

Formula we will use to calculate the variance:

$$s^2 = \frac{\Sigma f(X - \overline{X})^2}{\Sigma f - 1}$$

We had already calculated the mean for the variables. To calculate for variance, we followed the following steps for each branch.

X = to our mid-point values.

x̄ = group mean

fx = here we multiplied the frequency by the mid-point values

X - x̄ = Then we subtracted the group mean from the mid-point value

(X - x̄) ^2 = Then we squared the subtracted group mean from the mid-point value.

f((X - x̄) ^2) = here we multiplied the frequency to the squared value we found when we subtracted the group mean from the mid-point value.

Lastly, we calculated the sum of f((X - x̄) ^2) and the sum of total frequencies subtract 1 (∑ f – 1) and divided the two as in our variance formula above to get the group variance.

Table 2.2: The variance calculation table for Branch A

| Branch A | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | f | X | x̄ | fx | X - x̄ | (X - x̄) ^2 | f((X - x̄) ^2) | Variance |
| 10-14 | 14 | 12 | 14.18 | 168 | -2.18 | 4.75 | 66.53 | 99.86 |
| 15-19 | 18 | 17 | 14.18 | 306 | 2.82 | 7.95 | 143.14 | 99.86 |
| 20-24 | 20 | 22 | 14.18 | 440 | 7.82 | 61.15 | 1223.05 | 99.86 |
| 25-29 | 23 | 27 | 14.18 | 621 | 12.82 | 164.35 | 3780.11 | 99.86 |
| 30-34 | 10 | 32 | 14.18 | 320 | 17.82 | 317.55 | 3175.52 | 99.86 |
| **∑ Total** | 85 | 110 | 70.9 | 1855 | 39.1 | | 8388.35 | |

Table 2.3: The variance calculation table for Branch B

| Branch B | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | f | X | x̄ | fx | X - x̄ | (X - x̄) ^2 | f((X - x̄) ^2) | Variance |
| 10-14 | 15 | 12 | 15.67 | 180 | -3.67 | 13.47 | 202.03 | 68.65 |
| 15-19 | 22 | 17 | 15.67 | 374 | 1.33 | 1.77 | 38.92 | 68.65 |
| 20-24 | 26 | 22 | 15.67 | 572 | 6.33 | 40.07 | 1041.79 | 68.65 |
| 25-29 | 10 | 27 | 15.67 | 270 | 11.33 | 128.37 | 1283.69 | 68.65 |
| 30-34 | 12 | 32 | 15.67 | 384 | 16.33 | 266.67 | 3200.03 | 68.65 |
| **∑ Total** | 85 | 110 | 78.35 | 1780 | 31.65 | | 5766.46 | |

Table 2.4: The variance calculation table for Branch C

| Branch C | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | f | X | x̄ | fx | X - x̄ | (X - x̄) ^2 | f((X - x̄) ^2) | Variance |
| 10-14 | 12 | 12 | 13.67 | 144 | -1.67 | 2.79 | 33.47 | 92.74 |
| 15-19 | 20 | 17 | 13.67 | 340 | 3.33 | 11.09 | 221.78 | 92.74 |
| 20-24 | 32 | 22 | 13.67 | 704 | 8.33 | 69.39 | 2220.44 | 92.74 |
| 25-29 | 11 | 27 | 13.67 | 297 | 13.33 | 177.69 | 1954.58 | 92.74 |
| 30-34 | 10 | 32 | 13.67 | 320 | 18.33 | 335.99 | 3359.89 | 92.74 |
| **∑ Total** | 85 | 110 | 68.35 | 1805 | 41.65 | | 7790.16 | |

From our calculation of variance, we can observe the highest variability in workers working in Branch A then followed by workers working in Branch C and lastly workers who work in Branch B. These results show that the wages for employees in the factory in Branch A and Branch C are widely spread from the mean wage as compared to the employees who work in Branch B.