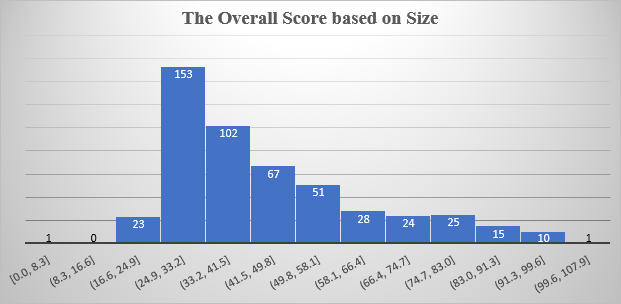
**Problem 1**

1. **) Overall Scaled Score i. Use suitable summary statistics (e.g., mean, min, max, mode, median, 1st and 3rd quartile) and chart(s) to analyze the distribution of the overall score.**

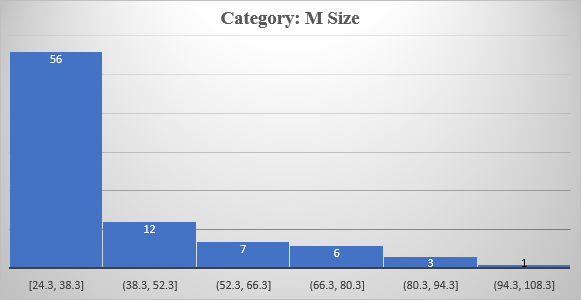
|  |  |
| --- | --- |
| *Summary Statistics* | |
| Mean | 44.6194 |
| Median | 38.55 |
| Mode | 29.9 |
| Minimum | 24.2 |
| Maximum | 100 |
| Largest(3) | 98.5 |
| Smallest | 24.2 |

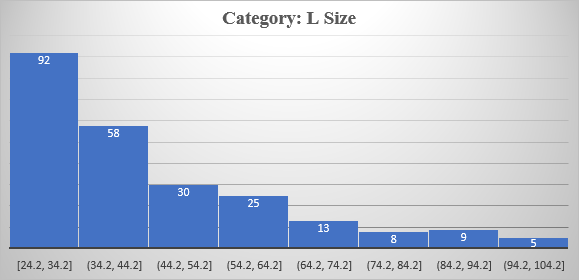


The histogram shows that the overall distribution score is skewed to the right, lacking normality. Also, the average between the maximum (100) and minimum (24.2) gives us 62.1. Comparing this value to the mean (44.6194) proves skewness.

1. **Use suitable summary statistics (e.g., mean, min, max, mode, median, 1st and 3rd quartile) and**

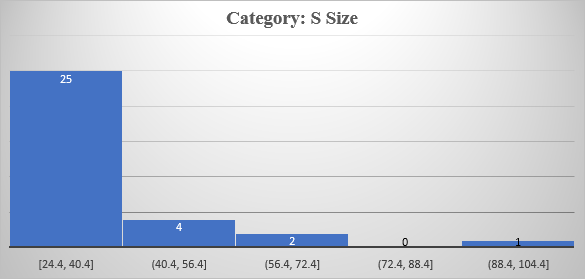
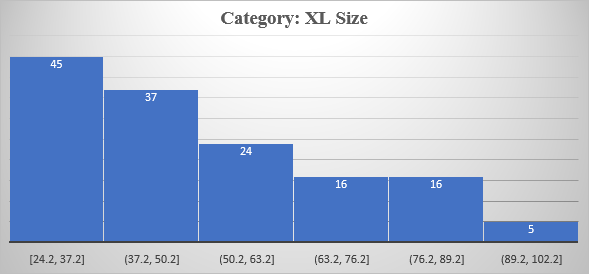
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | M | L | S | XL |
| Mean | 45 | 44.2 | 36.3 | 49.6 |
| Max | 100 | 98.8 | 97 | 95 |
| Min | 24.5 | 24.2 | 24.4 | 24.2 |
| Median | 35.4 | 38.4 | 30.3 | 45.4 |
| 1st Quartile | 28.3 | 29.875 | 27.975 | 34.2 |
| 3rd Quartile | 45.7 | 54.225 | 38.9 | 63.65 |



*Summary Statistics*

In the M size category table, the average of the min and max values is 62, and the mean is 40.5 showing the data is skewed to the right, lacking normality.

The distribution of the L-size category is still skewed to the right, as displayed by the chart. The average of the min and max is 61.5, whereas the mean is 44.2, which further proves the right skewness.

The data is positively skewed from the chart and contain outlier from the gap.

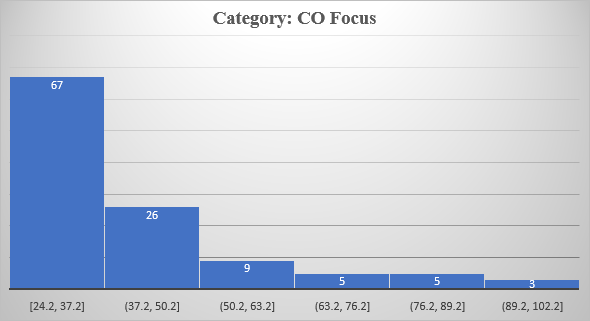
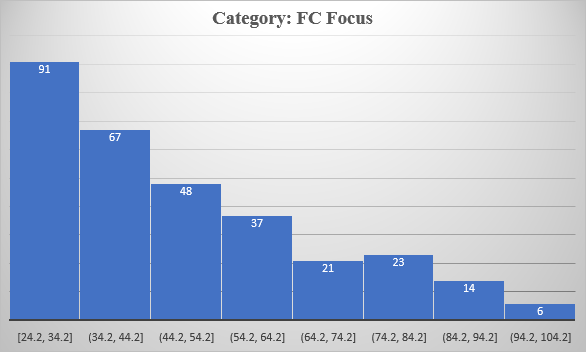
The XL Size category is also skewed to the right, as displayed by the above chart.

1. **Repeat the analysis in (ii) to compare the overall score based respectively on Subject Range (Focus), Research Intensity (Res.), and Age.**

**Subject Range (Focus)**

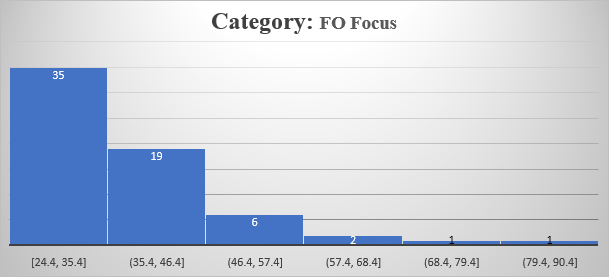
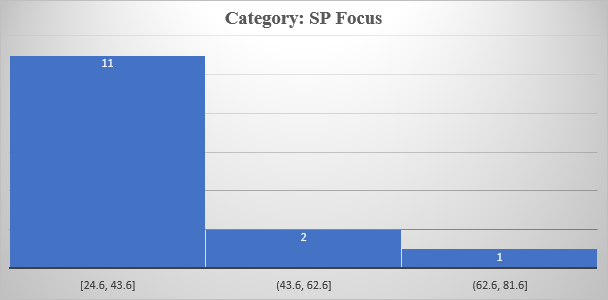
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CO | FC | FO | SP |
| Mean | 40.4 | 48.2 | 37.4 | 34.5 |
| Max | 100.0 | 98.8 | 89.2 | 72.3 |
| Min | 24.2 | 24.2 | 24.4 | 24.6 |
| Median | 34.7 | 43.6 | 34.3 | 29.5 |
| Mode | 26.3 | 81.5 | 36.1 | 24.8 |
| 1st Quartile | 28.15 | 31.65 | 29.1 | 26.375 |
| 3rd Quartile | 45.15 | 59.3 | 41.4 | 36.325 |

Looking at the tables, for the CO Focus type, the mode is less than the median, which is also less than the mean, indicating that the data is skewed to the right. This can be seen in the below histogram:

In the FC type, the average of the min and max is 61.5, which is greater than the mean (48.2), showing right skewness. The chart below shows the results.

For the FO type, the average of the min and max is 56.8, which is greater than the mean (37.4), showing positive skewness. This is displayed in the below graph:

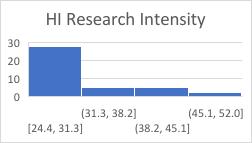
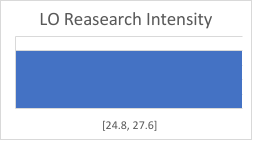
For the SP type, the average of the min and max is 48.45, which is greater than the mean (34.5), showing positive skewness. This is displayed below:

**Research Intensity**

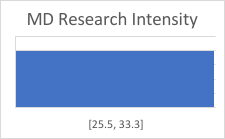
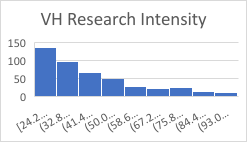
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | HI | LO | MD | VH |
| Mean | 30.8 | 25.5 | 28.6 | 46.1 |
| Min | 24.4 | 24.8 | 25.5 | 24.2 |
| Max | 50.8 | 26.2 | 33.3 | 100.0 |
| Mode | 24.9 | #N/A | #N/A | 30.2 |
| Median | 28.1 | 25.5 | 27.7 | 40.0 |
| 1st Quartile | 25.375 | 25.15 | 26.1 | 30.7 |
| 3rd Quartile | 35.425 | 25.85 | 30.15 | 56.2 |

The HI type is right skewed because the mode is less than the median, which is less than the mean. The graph that follows demonstrates this.

In the LO type, there is no mode, or the values have the same number of counts. The average between the min and max gives us 25.5, equal to the mean showing normality. This is because there are only two data points for the LO type. This is indicated by the below graph.

In the MD type, the result will be the same as the LO type due to the smaller number of data points.

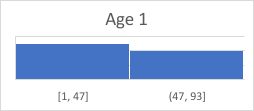
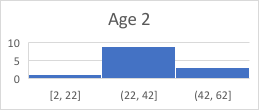
For the VH type, the mode is less than the median, which is also less than the mean showing the right skewness in the data.

**Age**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ages** | **1** | **2** | **3** | **4** | **5** |
| Mean | 55.3 | 35.0 | 38.5 | 41.5 | 50.8 |
| Min | 27.6 | 24.2 | 24.4 | 24.3 | 24.4 |
| Max | 83.8 | 55.9 | 92.7 | 90.1 | 100.0 |
| Median | 55.7 | 32.7 | 33.7 | 37.2 | 46.8 |
| Mode | #N/A | #N/A | 33.7 | 29.8 | 54.5 |
| 1st Quantile | 37.475 | 27.35 | 28 | 29.8 | 34.825 |
| 3rd Quantile | 71.775 | 41.925 | 42.05 | 48.4 | 62.15 |
| Mean | 55.3 | 35.0 | 38.5 | 41.5 | 50.8 |

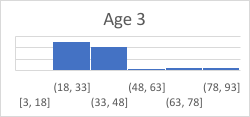
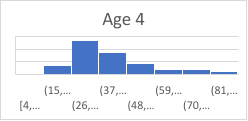
**Age 1**

The age 1 type will tend to display normality due to a small number of data points.

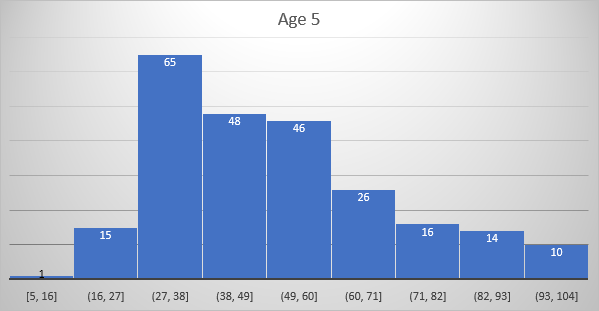
The same case applies to Age 2.

In Age 3, the average of the Min and Max for Age 3 is 58.7, which is greater than the mean (38.5), indicating positive skewness.

In Age 4, the average of the Min and Max for Age 3 is 57.2, which is greater than the mean (41.5), indicating positive skewness.

In Age 5, the average of the Min and Max for Age 3 is 62.2, which is greater than the mean (50.8), indicating positive skewness.



**v. Determine confidence intervals for the mean overall score by Size, Subject Range (Focus), Research Intensity (Res.), and Age. Soln.** The confidence interval is represented by Upper and Lower bounds**.**

**Size**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confidence Interval | M | L | S | XL |
| alpha | 0.05 | 0.05 | 0.05 | 0.05 |
| stdev | 17.89074 | 18.46144 | 14.96243 | 18.95913 |
| n | 85 | 240 | 32 | 143 |
| Sample mean | 40.5 | 44.3 | 36.3 | 49.6 |
| Marginal error | 3.803356 | 2.335651 | 5.184121 | 3.107409 |
| Upper bound | 44.3 | 46.6 | 41.5 | 52.7 |
| Lower bound | 36.7 | 42.0 | 31.1 | 46.5 |

**Subject Range**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confidence Interval | CO | FC | FO | SP |
| alpha | 0.05 | 0.05 | 0.05 | 0.05 |
| stdev | 17.36852 | 13.6674 | 12.43242 | 13.33109 |
| n | 115 | 307 | 64 | 14 |
| Sample mean | 40.4 | 48.2 | 37.4 | 34.5 |
| Marginal error | 3.174403 | 1.52885 | 3.045888 | 6.983125 |
| Upper bound | 43.6 | 49.7 | 40.4 | 41.4 |
| Lower bound | 37.3 | 46.6 | 34.3 | 27.5 |

**Research Intensity**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confidence Interval | HI | LO | MD | VH |
| alpha | 0.05 | 0.05 | 0.05 | 0.05 |
| stdev | 6.720172 | 0.989949 | 3.522783 | 18.87543 |
| n | 40 | 2 | 4 | 453 |
| Sample mean | 30.8 | 25.5 | 28.6 | 46.1 |
| Marginal error | 2.082565 | 1.371975 | 3.452264 | 1.738184 |
| Upper bound | 32.9 | 26.9 | 32.0 | 47.8 |
| Lower bound | 28.7 | 24.1 | 25.1 | 44.4 |

**Age**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confidence Interval | 1 | 2 | 3 | 4 |
| alpha | 0.05 | 0.05 | 0.05 | 0.05 |
| stdev | 21.7416 | 9.989824 | 16.07533 | 15.8862 |
| n | 8 | 12 | 55 | 137 |
| Sample mean | 55.3 | 35.0 | 38.5 | 41.5 |
| Marginal error | 15.06588 | 5.652171 | 4.248412 | 2.66016 |
| Upper bound | 70.3 | 40.7 | 42.8 | 44.2 |
| Lower bound | 40.2 | 29.4 | 34.3 | 38.9 |

**v. Using appropriate hypothesis tests, help BUG finds out whether there is a significant difference in overall score between:**

**1. New (Less than ten years old) and Historic (100 years old and more) Universities**

**2. Small (Fewer than 5,000 students) and Extra Large (More than 30,000 students) Universities**

H0: µS = µXL

H1: µS ≠ µXL

|  |
| --- |
| t-Test: Two-Sample Assuming Unequal Variances |

|  |  |  |
| --- | --- | --- |
|  | *Small 1* | *Extra Large 2* |
| Mean | 36.275 | 49.63846 |
| Variance | 223.8742 | 359.4486 |
| Observations | 32 | 143 |
| Hypothesized Mean Difference | 0 |  |
| df | 56 |  |
| t Stat | -4.33347 |  |
| P(T<=t) one-tail | 3.08E-05 |  |
| t Critical one-tail | 1.672522 |  |
| P(T<=t) two-tail | 6.16E-05 |  |
| t Critical two-tail | 2.003241 |  |

The p-value is less than our significance level of 0.05; hence we reject the null hypothesis and conclude that there is a significant difference between the overall score means of Small and Extra-Large Universities.

3. Specialist (2 or fewer faculty areas) and Full comprehensive (All 5 faculty areas + medical school) Universities

H0: µsp = µfc

H1: µsp ≠ µfc

|  |  |  |
| --- | --- | --- |
|  | Specialist | Full Comprehensive |
| Mean | 34.4571429 | 48.15732899 |
| Variance | 177.718022 | 380.9599706 |
| Observations | 14 | 307 |
| Hypothesized Mean Difference | 0 |  |
| df | 16 |  |
| t Stat | -3.670051 |  |
| P(T<=t) one-tail | 0.0010346 |  |
| t Critical one-tail | 1.74588368 |  |
| P(T<=t) two-tail | 0.0020692 |  |
| t Critical two-tail | 2.1199053 |  |

The p-value is less than 0.05; hence we reject the null hypothesis and conclude that the overall score means of Special and Full Comprehensive universities are different.

**b) Scaled Overall Score by continents. Use suitable charts to compare the overall score by Continent.**

The American Continent has the highest overall score, followed by the Europe continent. The Continent with the least overall score is Oceania.

**c) Relationships between variables. i. Calculate all correlation coefficients where appropriate.**

|  |
| --- |
| Correlation Coefficient of Academic Reputation vs Overall Score |

|  |  |  |
| --- | --- | --- |
|  | *Academic Reputation Score* | *Overall Score* |
| *Academic Reputation Score* | 1 |  |
| *Overall Score* | 0.895281648 | 1 |

The correlation coefficient for the above variables is 0.8953, which shows a strong correlation; hence there is a strong relationship between Academic Reputation Score and Overall Score.

|  |
| --- |
| The correlation coefficient of Employer Reputation Score vs Overall Score |

|  |  |  |
| --- | --- | --- |
|  | *Overall Score* | *Employer Reputation Score* |
| *Overall Score* | 1 |  |
| *Employer Reputation Score* | 0.775902923 | 1 |

The correlation coefficient for the above two variables is 0.776, which shows a strong correlation; hence there is a strong relationship between the Employer Reputation Score and the Overall Score.

|  |
| --- |
| The correlation coefficient of Faculty Student Score vs Overall Score |

|  |  |  |
| --- | --- | --- |
|  | *Overall Score* | *Faculty Student Score* |
| *Overall Score* | 1 |  |
| *Faculty Student Score* | 0.322265175 | 1 |

The correlation coefficient for the above two variables is 0.0.322which shows a weak correlation; hence there is a fragile relationship between the Faculty Student Score and Overall Score.

The correlation coefficient for the above two variables is 0.502, which shows a relatively strong correlation; hence there is a strong relationship between the Score per Faculty and the Overall Score.

The correlation coefficient of International Faculty Score vs Overall Score

|  |  |  |
| --- | --- | --- |
|  | *Overall Score* | *International Faculty Score* |
| *Overall Score* | 1 |  |
| *International Faculty Score* | 0.350472555 | 1 |

The correlation coefficient for the above two variables is 0.0.3505, which shows a weak correlation; hence there is a weak relationship between the International Faculty Score and the Overall Score.

**d) i. Develop a model using all variables. Comment on your results and potential issues.**

The adjusted R-Squared score is 0.99999, which shows a very good prediction.

1. **Develop a model using only significant variables. Comment on your results and potential issues.**

The adjusted R-Squared score is 0.99998, which shows a very good prediction.

1. **Develop a model using only the variables not used to calculate the overall score. Comment on your results**

The adjusted R-Squared score is 0.42784, which shows low performance.

**Problem 2**

1. **What kind of model was developed? What was the aim(s) of the model?**

Model: Decision Analysis model

Aim: It was used to discuss medical and orthopaedic decisions, such as the cost-effectiveness of the antibiotic-loaded bone cement for THA, the surgical treatment of the opposite hip in a patient with a slipped capital femoral epiphysis, and the cost-effectiveness of noncompartmental arthroplasty versus TKA. Briefly describe the possible alternatives, outcomes, and criteria for selecting the best decision.

Possible Decision Alternatives

* Open Irrigation/Debridement
* Single Stage Exchange
* Two-Stage Exchange

Outcomes

The outcomes were based on Quality of life and Toll.

|  |  |
| --- | --- |
| Utilities | Quality of life |
| Open Irrigation/Debridement | 0.86 |
| One-stage exchange | 0.82 |
| Two-stage exchange | 0.82 |
| Resection arthroplasty | 0.60 |

|  |  |
| --- | --- |
| Disutility of Revision | Toll |
| Open Irrigation/Debridement | -0.1 |
| One-stage exchange | -0.15 |
| Two-stage exchange | -0.2 |

Criteria for selecting the best TKA

It was determined that the disutility following an open debridement was -0.1, similar to the number used following a major THA. The disutility at the time of expected reimplantation during a two-stage exchange was estimated to be -0.2, double the value of initial THA, to account for two surgeries and an antibiotic spacer gap in between. A one-stage transaction that was assumed to be between these two values was given a value of -0.15 in this model.

Although no studies have assessed the utilities for the particular end health states in this model, the value obtained from the largest series in the Clinical Research and Health Policy Research database for an uncomplicated primary THA was 0.86. A disutility toll was added to the final quality-of-life estimate to account for this additional therapy since repeat surgery to address infection results in a poorer quality of life than an uncomplicated THA (i.e., due to a repeat procedure, duration of treatment, and loss of income). They estimated that the utility following a successful open debridement would be equivalent to that for a simple THA (0.86). Still, due to the repeat operation, a disutility toll of 0.1 was deducted from the estimated quality of life of an uncomplicated THA (to account for the morbidity of a second procedure). Therefore, 0.86 (utility of a simple THA) minus 0.1 (disutility toll for debridement operation), or 0.76, was the utility estimate (quality of life) following a debridement for THA infection. A patient undergoing a revision THA reportedly had a 0.82 quality-of-life value evaluation [32]. The researchers could not find any studies that specifically looked at the utility values of one-stage exchanges vs two-stage exchanges. Their model assigned a score of 0.82 to the utility following a one-stage or two-stage revision THA for sepsis. They projected that a two-stage exchange would ultimately lead to a lower quality of life than a one-stage exchange because of the second treatment and the intermediate period without a suitable hip prosthesis. This discrepancy in the final quality-of-life outcome estimations between one-stage and two-stage exchanges was accounted for by subtracting the disutility toll from the final estimate of the quality-adjusted life year after each procedure. The usefulness of resection arthroplasty was determined to be 0.60.

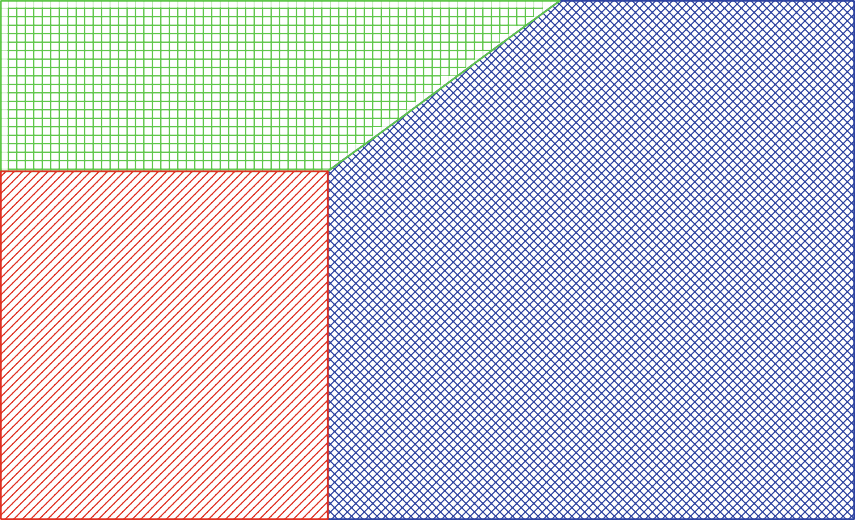
**Using a suitable package such as Precision Tree, draw, for this problem, a decision tree such as the one in Figure 1. Include all relevant parameters.**

1. **Fold back the tree and determine the best option. Is it the same as in the paper? Justify why or why not.**

No. The measurements in the paper are precise as compared to the software being used for this analysis.

1. **Undertake the same one-way and two-way sensitivity analyses as in the paper, as well as any additional sensitivity analysis as you see fit (justify). You might use graphics of your choice to display results. Comment on your results, in particular on: - how they compare with those in the paper.**

 Open Irrigation/Debridement  Single Stage Exchange  Two-Stage Exchange



0.000 0.099 0.198 0.297 0.396 0.495 0.594 0.693 0.792 0.891 0.

The results are almost similar to the ones on paper, but there are some differences in the prediction of the treatment.

**Problem 3**

**1 a) Formulate a linear programming (LP) model algebraically to assist the company in minimizing the annual variable cost of meeting a demand for air conditioners. Solve using Excel Solver.**

**Solution**

I then calculated the cost of production in each region by the estimated monthly demand of each conditioner and got the totals.

I then used Excel Cover to minimize the annual variable cost and got the following result.

1. **The monthly fixed cost of operating a factory in each City is shown in the following table. Revise your formulation to minimize the monthly total (variable + fixed) cost of meeting the demand for air conditioners. Solve using Excel Solver.**

In the below table, I added the monthly factory cost with the variable costs.

I then calculated the costs for the expected number of conditioners in each region and got the totals. Then created another column where I added the expenses of conditioners with fixed factory costs.

I then solved the problem using Excel solver to get the following:

1. **Additionally, at least 5,000 units of North demand must come from either City 1 or City 2. Revise your formulation in (b) to incorporate this constraint and solve using Excel Solver**

Multiplying the 5000 by the cost for City 1, which is 230, you get 1,150,000. Looking at the production cost of CityCity, the value of 2645000 is greater than the minimum value required; hence the solver is optimal. For City 2, the 5000 units account for 1,105,000, which is also less than 2,541,000; hence the solver is optimal.

1. **The company has gathered historical data of the monthly demand for air conditioners over the past 9 years 9 months (since 2013)- see file Data\_Pb3.xlsx.**

**i. Exploring the time series for each region.**

**1. Plot the data for each region as time series**

The time series plot for East Region exhibits some trend when the trend line is looked upon closely. The trend line tends to decrease as time increases. There is some seasonality exhibited, but it seems to be low.

The time series plot for the South region exhibits an increasing trend. There is also some pattern (increasing and decreasing) which proves seasonality.

There is no trend in the data as the line is constant. There is, however, significant seasonality identified by increasing and decreasing patterns.

There is no trend as the trend line is constant, but there is some seasonality displayed by the pattern.

**Seasonal Indexes**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | East | South | North | West |
| Average |  | 0429.04 | 16380.98 | 11005.98 | 8920.051 |
| Total Average | | 46736.06 |  |  |  |
| Accumulative Average | | 11684.01 |  |  |  |
| Seasonal Indexes | | 0.892591 | 1.401999 | 0.941969 | 0.763441 |

**4. State with justification for each region what would be appropriate forecasting methods to forecast monthly demand until September 2023.**

East Region – The Random Walk model would be the best here due to non-stationarity.

South Region – The Random Walk model will be suitable here due to non-stationarity.

North Region – The best model here is the Autoregressive model since there is no stationarity, and also, there seems to be some clear pattern between the years showing some correlation.

West Region – The best model here is the Moving Average since there is no stationarity and the values are not that predictable. Although there is some pattern, it cannot be identified well, meaning the values are not highly correlated.

**I. Apply two appropriate forecasting methods for each region to forecast monthly demand between October 2022 and September 2023.**

**1. For each region, compare the performance of the two methods using MAE, RMSE, and MAPE.**

Using Random Walk Method

Using Moving Average

Appendix

