**Report of Assignment 5**

Group: Not At All

Jun Ma 001080796

Yu Chen 001385622

Qiushi Zhang 001057253

In our views, the key of solving this problem is the maximum price and the minimum price. For the target price, it is suitable for the model because the sales price has an average distribution around the target price at most time.

Although we handle the target price with arithmetic mean, there are still many differences among the random generate data. Errors shown in some of data, whose range is beyond -5%-5% which is required in the guidance. Therefore, we thought that the error may occurs from the maximum price and the minimum price of product. The origin Data is generating by this role:

minPrice = cheapestProduct + rand.nextInt(mostExpensiveProduct-cheapestProduct);

maxPrice = minPrice + minRange +rand.nextInt(maxRange - minRange);

targetPrice = minPrice + (maxPrice - minPrice)/2;

Acording to group discussion, we decided to change the maxPrice generate model to this formula:

maxPrice = minPrice + rand.nextInt((int)(aNumber\*minPrice**)**)+minRange;

To evaluate the performance of this formula, we regenerate and analysis data. A better example was generated in the computer as two pairs of csv. The result of the analysis is showed in a .txt file.

As we see, by decreasing the range of the price, error rate will be decreased remarkable. We set aNumber=0.05 and afterwards we got two table like following.

The origin Data analysis by program as follow table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product id | Ave price | quantities | difference | target | error |
| beyond the target: | | | | | |
| 19 | 96.66 | 1657 | 1.66 | 95 | -0.02 |
| 15 | 200.03 | 1423 | 1.03 | 199 | -0.01 |
| 11 | 116.54 | 1468 | 0.54 | 116 | -0.00 |
| 17 | 194.43 | 1301 | 0.43 | 194 | -0.00 |
| 13 | 116.17 | 1517 | 0.17 | 116 | -0.00 |
| Below the target | | | | | |
| 0 | 120.95 | 1482 | -0.05 | 121 | 0.00 |
| 12 | 93.90 | 1547 | -0.10 | 94 | 0.00 |
| 7 | 142.66 | 1376 | -.034 | 143 | 0.00 |
| 14 | 188.62 | 1373 | -0.38 | 189 | 0.00 |
| 2 | 208.42 | 1565 | -0.58 | 209 | 0.00 |
| 16 | 153.19 | 1585 | -0.81 | 154 | 0.01 |
| 9 | 73.03 | 1273 | -0.97 | 74 | 0.01 |
| 4 | 162.02 | 1439 | -0.98 | 163 | 0.01 |
| 3 | 189.44 | 1309 | -1.56 | 191 | 0.01 |
| 18 | 75.20 | 1544 | -1.80 | 77 | 0.02 |
| 10 | 122.09 | 1256 | -1.91 | 124 | 0.02 |
| 5 | 113.70 | 1499 | -2.30 | 116 | 0.02 |
| 1 | 192.88 | 1484 | -3.12 | 196 | 0.02 |
| 6 | 58.59 | 1197 | -3.41 | 62 | 0.06 |
| 8 | 155.12 | 1697 | -3.88 | 159 | 0.03 |

The modify Data analysis result as follow:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product id | Ave price | quantities | difference | target | error |
| beyond the target: | | | | | |
| 5 | 95.20 | 1473 | 0.20 | 95 | -0.00 |
| 9 | 45.01 | 1476 | 0.01 | 45 | -0.00 |
| below the target: | | | | | |
| 2 | 204.95 | 1453 | -0.05 | 205 | 0.00 |
| 10 | 89.93 | 1226 | -0.07 | 90 | 0.00 |
| 11 | 93.93 | 1298 | -0.07 | 94 | 0.00 |
| 3 | 167.82 | 1594 | -0.18 | 168 | 0.00 |
| 17 | 187.80 | 1560 | -0.20 | 188 | 0.00 |
| 16 | 147.73 | 1340 | -0.27 | 148 | 0.00 |
| 0 | 121.72 | 1657 | -0.28 | 122 | 0.00 |
| 12 | 80.67 | 1491 | -0.33 | 81 | 0.00 |
| 13 | 87.66 | 1177 | -0.34 | 88 | 0.00 |
| 14 | 180.66 | 1338 | -0.34 | 181 | 0.00 |
| 15 | 180.54 | 1653 | -0.46 | 181 | 0.00 |
| 6 | 39.47 | 1233 | -0.53 | 40 | 0.01 |
| 19 | 72.46 | 1632 | -0.54 | 73 | 0.01 |
| 8 | 111.46 | 1386 | -0.54 | 112 | 0.00 |
| 18 | 52.46 | 1767 | -0.54 | 53 | 0.01 |
| 7 | 120.41 | 1276 | -0.59 | 121 | 0.00 |
| 1 | 158.28 | 1804 | -0.72 | 159 | 0.00 |
| 4 | 132.21 | 1384 | -0.79 | 133 | 0.01 |

From these tables, we can easily find that the change is useful to decrease the difference and the error. However, decreasing the difference and the error by limiting the range is not practical enough in real examples. To improve, we decide to find the best aNumber which can consider both range and error.

When aNumber is beyond 1, the error in average is below the origin and the max range. In fact, the error in average is beyond the origin and the minimal range is not less than the origin too much.