

FIT 1006 Assignment 1

Tutor: Anthony Wong / Tutorial Time: Tuesday

Name: Rui Qin / ID: 30874157

Introduction

The purpose of this report is to analyze the total assets of the sample one year after investing 1000 dollars in 8 different groups. We are going to find the best group with maximum income with low risk. The final result is group 7 has the investment data, the data will be analyzed with a statistics table, box plot, and histograms. The outcome is worked by Excel and Systat 13.

	(\$)	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
Number of investment		64	83	128	103	57	16	45	93
Minimum		993.65	785.32	558.41	729.83	664.12	1035.09	1111.96	739.16
Maximum		2,794.28	3151.4	1483.59	2282.96	5369.66	1125.14	1499.37	1540.96
Median		1,120.85	1091.12	1004.305	955	1009.76	1062.26	1367.02	1211.72
Range		1800.63	2366.08	925.18	1553.13	4705.54	90.05	387.41	801.8
Mean		1136.5174	1101.4453	996.81416	965.8741	1040.2761	1067.0191	1363.3336	1201.8186
10% Trimmed Mean		1120.5281	1091.457	1011.453	961.11077	1013.1737	1067.845	1367.8441	1206.9186
Standard Deviation		216.86699	230.30502	173.58454	185.3357	580.80574	23.17215	56.292881	107.94729
Variance		47031.292	53040.403	30131.593	34349.321	337335.31	536.94852	3168.8885	11652.618
Q1		1089.1675	1080.605	908.1275	889.35	987.85	1051.38	1341.74	1155.22
Q3		1150.1175	1098.91	1086.18	1003.49	1040.05	1074.2975	1386.88	1249.84
Q2		1120.85	1091.12	1004.305	955	1009.76	1062.26	1367.02	1211.72
IQR		60.95	18.305	178.0525	114.14	52.2	22.9175	45.14	94.62

Table 1: Statistics created by Excel

Analyzation of Statistics

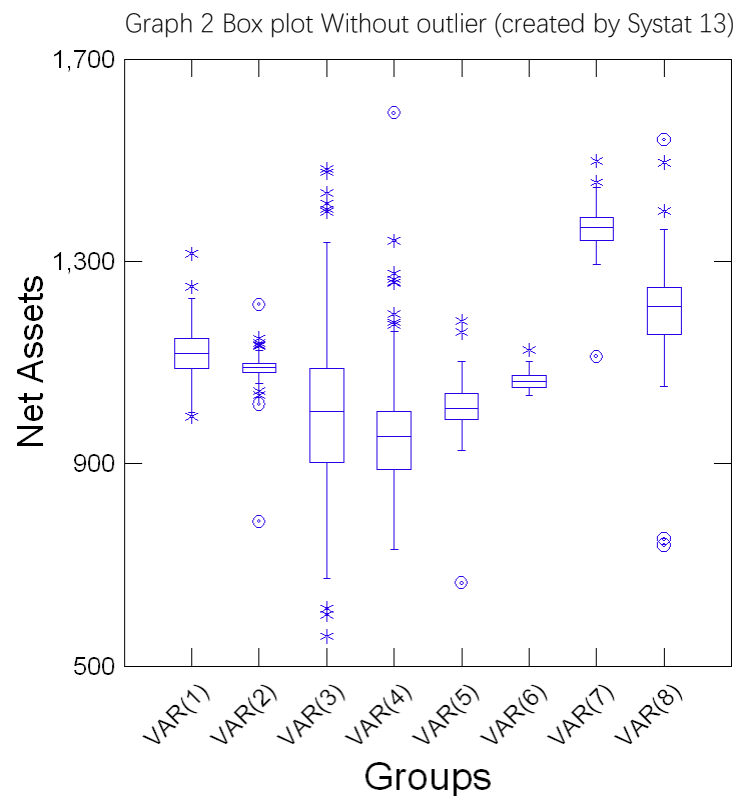
This table includes descriptive statistics of every group. In this table, we can observe the data difference in these 8 groups.

We can easily find out the number of samples has a great difference. Group 6 only gets 16 samples which are in the smallest sample size of entire groups, which may increase the inaccuracy in other statistics calculation (Selvananthan & Keller 2021, p. 36), because according to the central limit theorem, a large sample size should be 30.

Inside the maximum and minimum numbers, the too-large number and too small number may lead to the inaccuracy of the mean (Selvananthan & Keller 2021, p. 138). The maximum

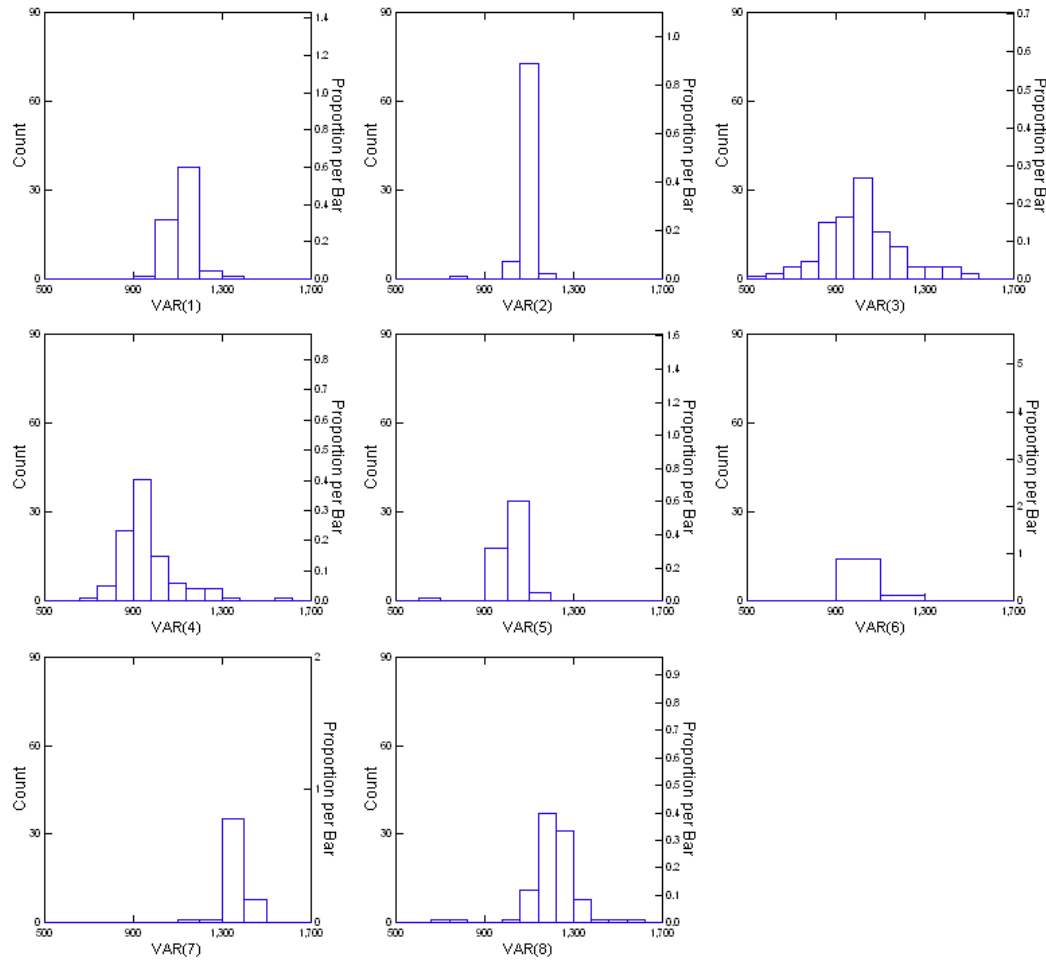
of group 5 is more than 5000 which may cause a strong increase of mean in group 5, and also the greatest number of groups 1, group 2, and group 4 can be considered the outlier with the help of a box plot graph (with outlier). The minimum of these groups does not create so much outlier. According to this information, we can create a box plot without an outlier, a maximum limit of 1700, and a minimum limit of 500 for stretching the graph.

Due to so much outlier, the median is more reliable, because it cannot affect by the outlier. According to the table, it can be seen the group 7 and 8 medians is more than 1200 dollars.



Analysis of the box plot

Now we pay attention to the box plot, it is obvious the group 7 has the highest median with a small IQR. High median corresponding with the value of the reward concentrates with, small IQR means the distribution is very small, most of the assets is concentrated close to the median. Meanwhile, group 7 has the largest minimum number which still over 1100 dollars. High median, small IQR, and largest minimum number can show the group 7 has a high reward with the small risk. On the opposite side, group 3 is the worst, it has a high risk with high distribution (high IQR: 178) and the smallest minimum, group 4 also has risk with the lowest median and large IQR. Group 2 and 6 have small IQR but the medians are not high enough, group 1 and 5 performance are ordinary, even group 1 has not extreme outlier and median is high, but it cannot compare with group 8 and 7 which have higher median and better data performance.



Histograms Graph with limited maximum and minimum

Analysis of histograms

Graph 3 includes all group's histograms with all numbers, but we need to remove the outliers to make the graph available for analysis. It can be seen that most of the histograms show the normal distribution because we can see the magnitude of values get further away from the mean decrease equally in left and right direction, even group 1,2,4,5 has outlier and group 7 has more sample between 1400 to 1500 dollars which make their graph are not unimodal. In this case, the negative distribution is the graph skew to the left and more sample heap up in the right side, the positive distribution skew to the right (Selvananthan & Keller 2021, p. 95). A negative histogram means more samples can earn more in investment, which is better than a positive and normal histogram. By looking at all graphs we can see group 6 is right-skewed, groups 1, 2, 4, 5 are normal if we ignore the outlier. Groups 3, 7, and 8 perform left-skewed which are better for investment, but in the sample of group 3, 500 to 1000 dollars occupy most of the data which means most of the investors lost money in group 3. The range that data focus on is important, the sample of group 7 concentrates from 1200 to 1400 dollars, which is better than group 8 (1100 to 1300 dollars) and group 3(800 to 1200 dollars).

Conclusion

With the help of graphs and tables, we can easily find out group 7 is the best choice for investment, even in all group they have outliers, but group 7 have the largest minimum outlier and the best median. Group 4 and 3 are the worst with high risk or low median, some investors even occur loss. Other performances are normal, group 8 is great but still cannot compare with group 7.

References:

Selvenanthan & Keller (2021) *Business statistics abridged: 8th*, Australia, Melbourne.

Appendix

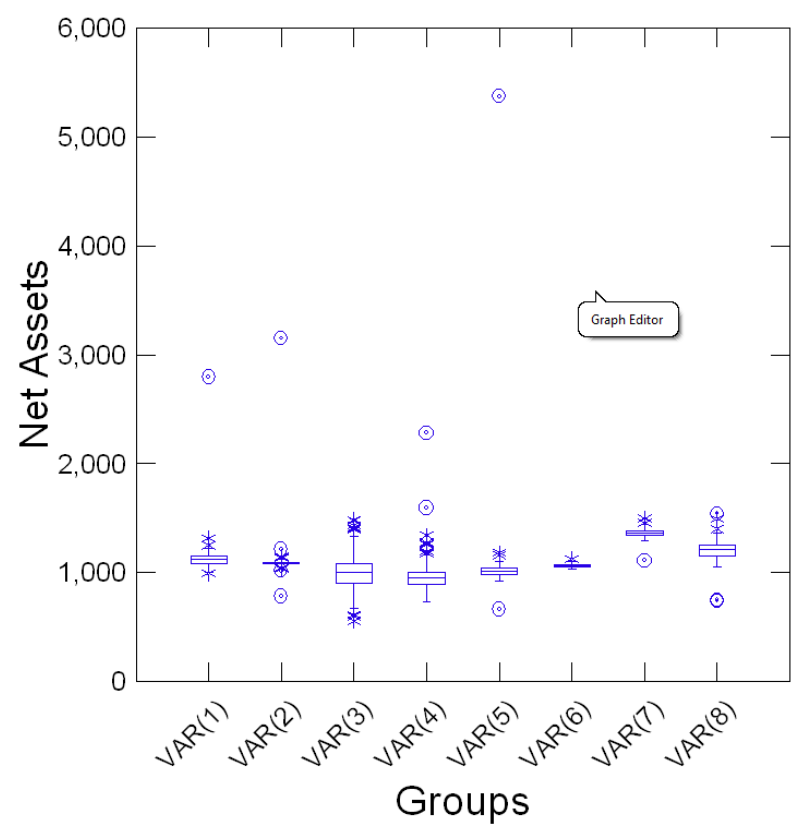


Figure 3 Box plot with outlier

Graph 4 histograms with the total number

