

Data Summarization in EXCEL

We will show how to visually summarize data and compute Descriptive Statistics in EXCEL2007 via two examples.

1. In a study of job satisfaction, a series of tests were administered to 60 subjects and the following data was obtained (higher score represents higher job satisfaction) (data file Score.xlsx)

Score					
76	86	66	91	53	79
85	82	80	84	66	66
78	69	77	70	75	72
70	62	56	77	75	69
78	85	69	75	83	58
75	69	70	71	76	53
73	84	72	72	80	71
70	82	55	84	51	68
72	101	86	93	87	65
78	99	90	77	79	78

Compute (a) descriptive statistics and (b) graph the histogram for the data.

2. The following table shows economic data for 36 firms in Japan (data file Japan.xlsx)

Short Name	Market_Cap_Yen	Market Cap US \$	Book Equity	Revenues	Net Income	PE
CENTURY21 REAL	14496	\$123.72	1924.79	2452.09	463.24	31.29
BALS CORP	43231.88	\$369.01	3239	18994	464	93.17
MAINICHI COMNET	10723.2	\$91.69	2675	6449	464	23.11
MUTUAL	5990.63	\$51.42	6915	10100	464	12.91
NISSIN SHOJI CO	9728	\$83.18	17129	70529	464	20.97
TOW CO LTD	8030.752	\$68.94	3782	10705	465	17.27
CLEX CO LTD	10105.62	\$86.41	4455	9486	466	21.69
FENWAL						
CONTROLS	6334.975	\$54.37	4357	14930	466	13.59
FUJII SANGYO	8008	\$68.62	11392	46927	466	17.18
MOLITEC STEEL	10760.17	\$92.07	10221	18923	466	23.09
TAKACHIHO						
KOHEKI	19198.13	\$164.18	12108	22318	466	41.2
GMO HOSTING	67081.95	\$571.52	1247	3038	467	143.64
MIURA PRINTING	11011.9	\$94.15	9504	22136	467	23.58
TSUKEN CORP.	9278.147	\$79.65	14182	44581	468	19.83
KING CO LTD	14615.24	\$124.94	16822	22369	469	31.16
NIPPON RESIBON	5926.8	\$50.88	5760	15012	471	12.58
AVAL DATA CORP	10725.84	\$91.76	7682.24	7775.44	471.64	22.74
SOMAR CORP	15885.06	\$135.62	14781	36540	472	33.65
MISUMI CO LTD	8296.68399	\$71.22	10796	48836	473	17.54
NICHIWA SANGYO	11144.49	\$95.32	15779	42124	473	23.56

Data Summarization in EXCEL

HARADA INDUS CO	11388.83	\$97.43	6376	20771	474	24.03
IMV CORP	8478.508	\$72.75	2006	5088	476	17.81
MARUFUJI SHEET P	16108.48	\$137.51	24738	32758	476	33.84
DAIDO SIGNAL CO KANESHITA CONSTR	6954.948	\$59.69	8214	17632	477	14.58
KAKAKU.COM INC	17605.8	\$150.57	23205	19760	477	36.91
MR MAX CORP	71655.67	\$610.72	2223.01	2138.87	478.05	149.89
NAGOYA ELECTRIC	28044.59	\$239.30	28391	86133	480	58.43
ZOA CORP	6755.84	\$57.98	15568	15646	480	14.07
CHUO KAGAKU CO L	6889.5	\$59.12	1629	17589	481	14.32
PALTEK CORP	28088.1	\$239.70	30825	82965	483	58.15
RIX CORP	6150.098	\$52.78	8711	19355	483	12.73
ENSHU	8726.4	\$74.90	4978.53	27043.88	483.76	18.04
TOYAMA BANK LTD	22391.1	\$191.39	6409	35665	484	46.26
WAO CORP	17417.16	\$148.64	22493	7486	484	35.99
OIE SANGYO CO	6733.8	\$57.71	3577	15033	484.0665	NA
	11568.75	\$98.92	9021	52180	485	23.85

- a) Calculate descriptive statistics for each of the 6 variables.
- b) Calculate descriptive statistics for PE by the following categories:
 - Market_Cap_US\$ 50 to 100
 - 100 to 200
 - 200 to 300
 - > 300
- c) Graph Book_equity vs. Revenue.
- d) Calculate the correlation coefficient matrix of the 6 variables.

1(a). Decriptives for data of Example 1

Start with SCORE data in the cells A2 through A61, and click on the sequence

Data/Data Analysis/Descriptive Statistics as shown in Figures 1a - 1b.

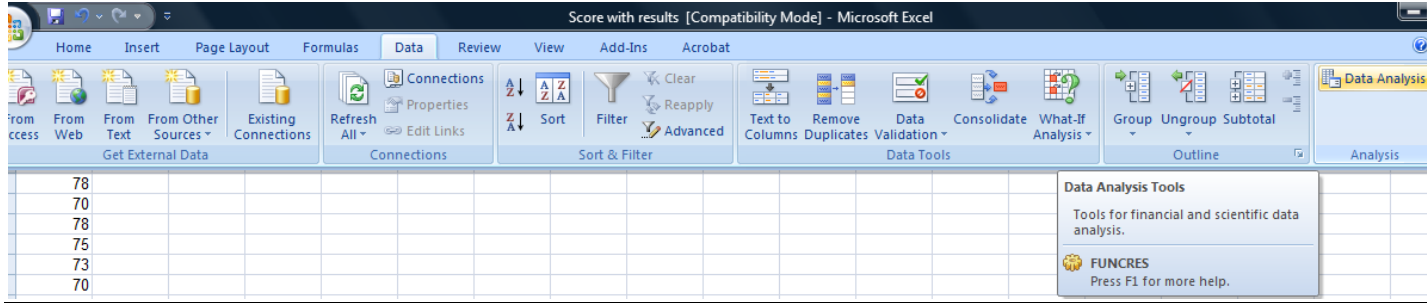


Figure 1(a): Data/Data Analysis

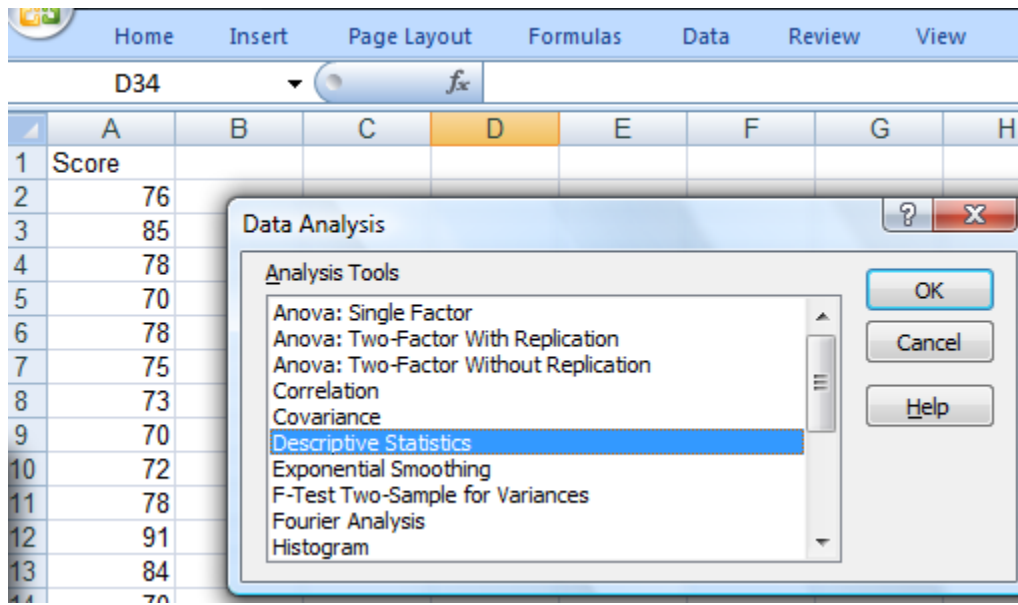


Figure 1b: Descriptive Statistics

This will open the window shown in Figure 1(c); select \$A\$2:\$A\$61 cells in the Input Range, select 'Grouped By Columns', and if you want the results in a New Worksheet, give a name to the worksheet (e.g., RESULTS), check Summary Statistics and enter 95 in the 'Confidence Level for Mean' box.

Data Summarization in EXCEL

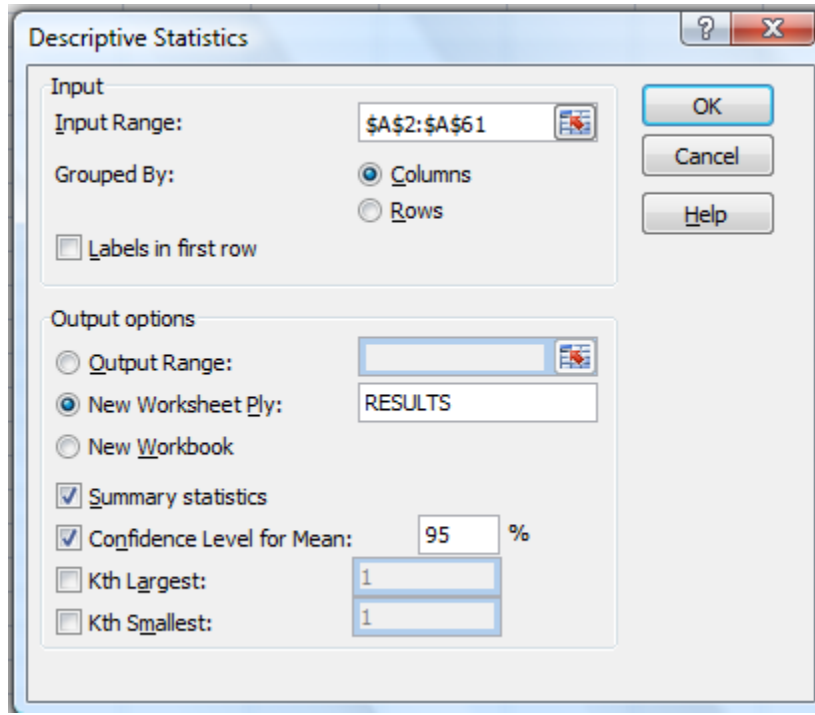


Figure 1(c): selecting input for Descriptive Statistics

This will create a new tab RESULTS in the same excel file with output shown in Figure 1d.

	A	B
1	Column1	
2		
3	Mean	74.883333
4	Standard Error	1.370614
5	Median	75
6	Mode	78
7	Standard Deviation	10.616731
8	Sample Variance	112.71497
9	Kurtosis	0.313634
10	Skewness	-0.071932
11	Range	50
12	Minimum	51
13	Maximum	101
14	Sum	4493
15	Count	60
16	Confidence Level(95.0%)	2.7425923

Figure 1d: Output from Descriptive Statistics

OUTPUT from Descriptive Statistics

Excel is using non-standard terminology here. In standard statistics terminology, confidence level here is 95%, and not 2.7425923.

The last term C = Confidence level (95.0%) equals upper 5% point $t_{n-1, .05}$ from the t-table with degrees of freedom $n-1 = 59$ for Example 1.

95% Confidence Interval for MEAN SCORE is

$$\bar{X} \pm C$$

1(b): Histogram for data of Example 1

The histogram in EXCEL is drawn in two steps:

1) Calculate the frequency distribution using the Data Analysis function HISTOGRAM (which just gives the frequency table, and does not plot the histogram). To use the HISTOGRAM function, you need to compute the BINS which are the end points of the Class Intervals of the Frequency Table, as shown below in Step (i):

Step (i) : Calculate sample minimum, maximum, range, and class width in excel. For example, minimum is calculated by typing

=min(a2:a61)

in the cell D2, if the data is in the cells A2 through A61, range is calculated by typing

= D3 - D2

in the cell D4, and class width h is calculated by dividing the range by the desired number of class intervals (we have selected 5 class intervals for this example) - see Figure 1d.

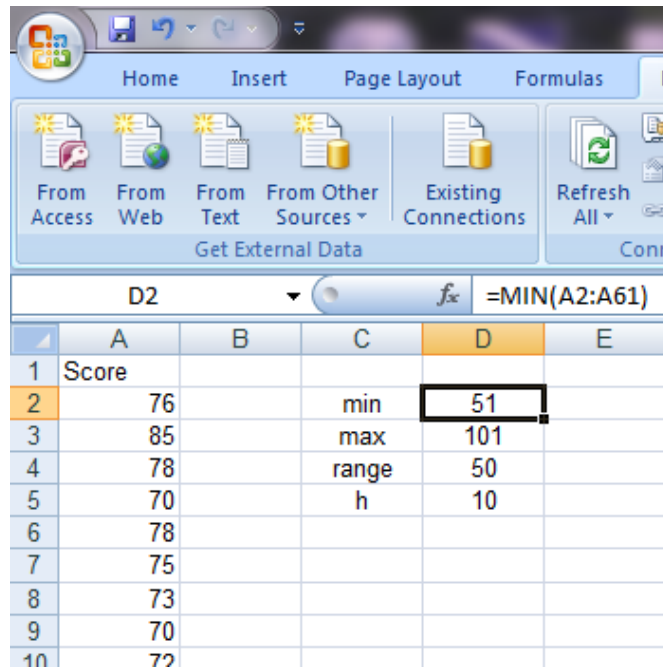


Figure 1d: Calculating min, max, range, class width h

Data Summarization in EXCEL

Next calculate the BINS for the 5 class intervals: $b_1 = 51$, $b_2 = b_1+h = 61$, ..., $b_5 = 91$, as shown in Figure 1e.

fx =F5+\$D\$5			
C	D	E	F
			BIN
min	51		51
max	101		61
range	50		71
h	10		81
			91

Figure 1e: Calculating BIN values

Next click on Data/Data Analysis/Histogram sequence (Figure 1f)

The screenshot shows the Microsoft Excel interface with the 'Data Analysis' dialog box open. The 'Histogram' option is selected in the list. The background spreadsheet shows a list of scores in column A (rows 2-14) and a frequency table in columns M and N (rows 2-6).

Interval	Frequency
0-51	1
51-61	5
61-71	16
71-81	22
81-91	13
91-101	3

Figure 1f: Calculating Frequency Table for SCORE data

Select Input Range - type \$A\$2:\$A\$61 (or click on the red arrow next to Input Range box, highlight cells A2:A61)

select Bin Range - \$F\$2:\$F\$6

select Output Range (we used \$J\$1 here) - see Figure 1g.

The output is shown in cells J1 : K7 (see Figure 1h).

Data Summarization in EXCEL

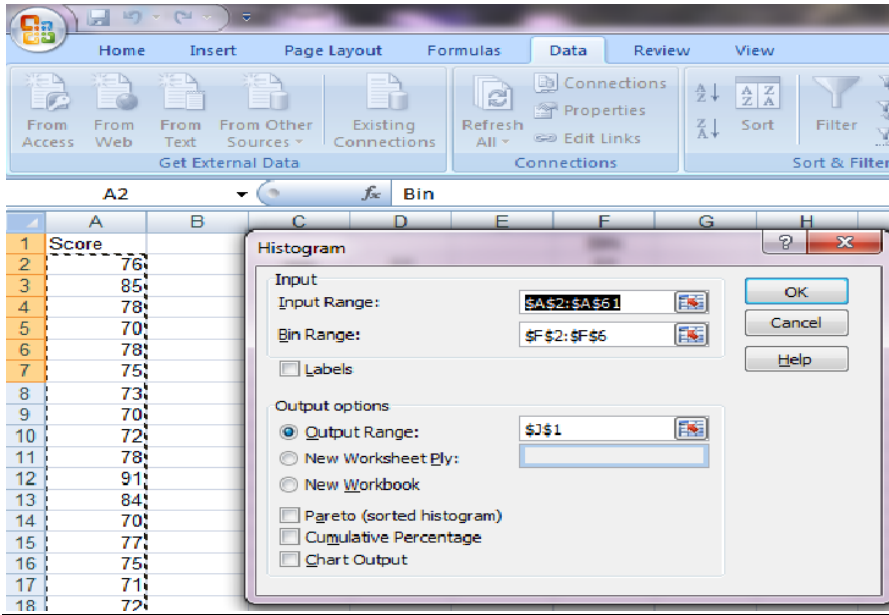


Figure 1g: Select input range of data and BIN, provide Output Range for Histogram

Bin	Frequency
51	1
61	5
71	16
81	22
91	13
More	3

Figure 1h: Frequency Table created by the Function Histogram

Step (ii): Draw the histogram - highlight cells J1:K7, click on INSERT/COLUMN/click on 1st 2D column (Figure 1i) - which results in histogram (figure 1j).

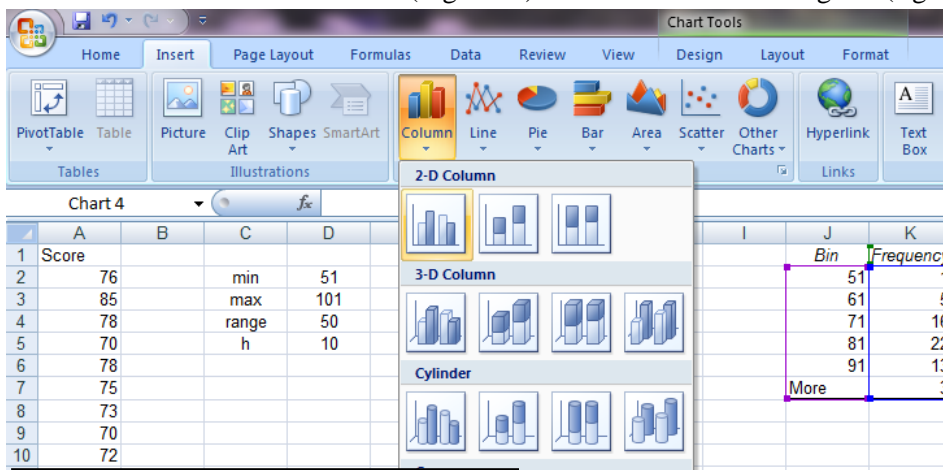


Figure 1i: Drawing Histogram

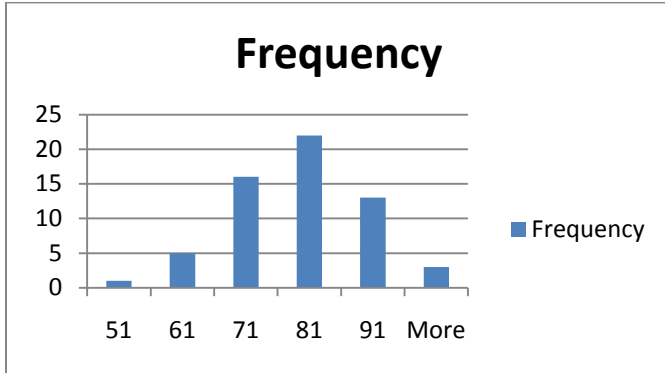


Figure 1j: Excel Histogram for SCORE

Notice that the label 51 in Figure 1j above represents the class interval '0 - 51', label 61 represents the class interval '51 - 61', ..., 91 represents the interval '81 - 91' and label More represents the interval '91 - 101'. Also, in a typical histogram, there is no gap between the bars. To convert the histogram of Figure 3(b) to a standard histogram, follow the following steps:

(i) Create a column Interval with text values '0-51', '51-61', ..., '91-101' by typing '0-51' in cell I2, '51-61' in cell I3, etc., (Figure 1k).

The screenshot shows an Excel spreadsheet with a table. The table has two columns: "Interval" and "Frequency". The data rows are as follows:

Interval	Frequency
0-51	1
51-61	5
61-71	16
71-81	22
81-91	13
91-101	3

Figure 1k: Type Class Interval Column

(ii) Highlight the cells I1:J7, click on Insert/Column/2D Column to get the following histogram (Figure 1l).

Data Summarization in EXCEL

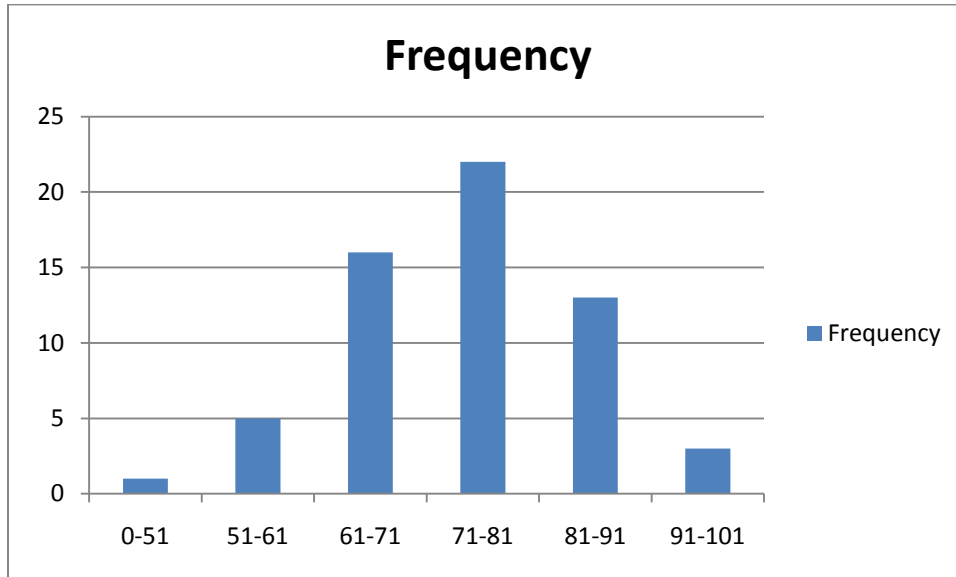
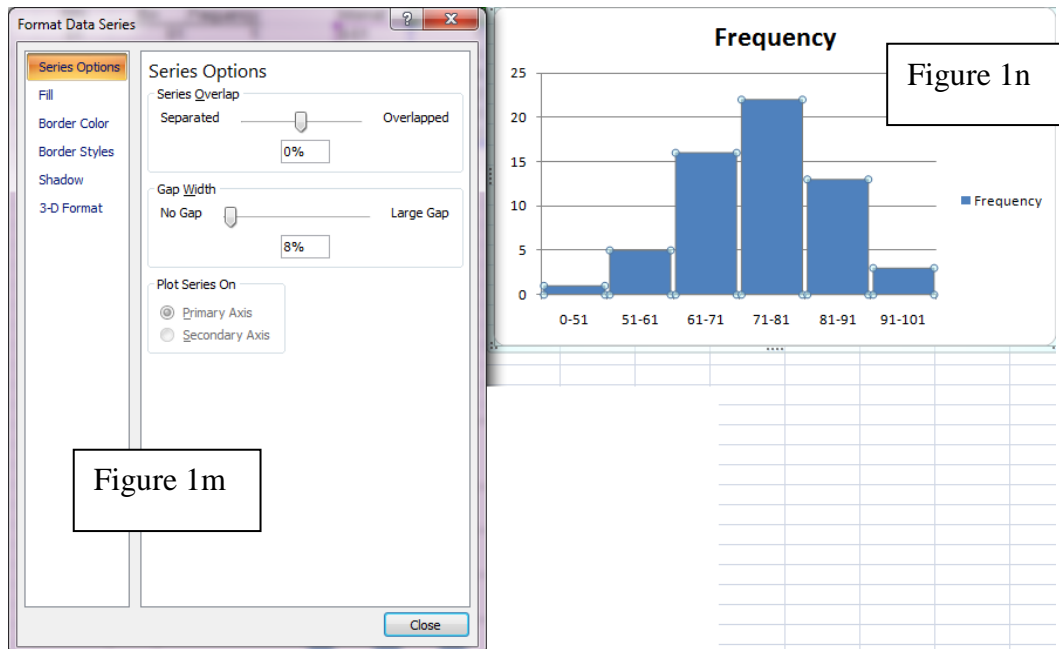


Figure 1l: Basic Histogram for Score

The labels on the x-axis look better, but the bars still have gaps between them. To fix this, right-click on one of the bars, then click on Format Data Series, which opens the Series Option Window (Figure 1m). You can close the gap by moving the 'No Gap' Slider to 0% as shown in Figure 1n.



Summarization of Data of Example 2

2(a). Descriptive Statistics for Each Column

Click on Data/Data Analysis/Descriptive Statistics

Select Input Range (include Title Row in range)

Check 'Labels in first row' box, Grouped by COLUMNS,

Name New Worksheet Ply (Descriptives in this example)

Check 'Summary Statistics' and 'Confidence Level for Mean' boxes, then click OK (see Figure 2a) to get the output of Figure 2b.

B	C	D	E	F	G
Market_Cap_Yen	Market Cap US \$	Book Equity	Revenues	Net Income	P/E
14496	\$123.72	1924.79	2452.09	463.24	31.29
43231.88	\$369.01	3239	18994	464	93.17
10723.2	\$91.69	2675	6449	464	23.11
5990.63	\$51.42	6915	10100	464	12.91
9728	\$83.18	17129	70529	464	20.97
8030.752	\$68.94	3782	10705	465	17.27
10105.62	\$86.41	4455	9486	466	21.69
6334.975	\$54.37	4357	14930	466	13.59
					17.18
					23.09
					41.2
					143.64
					23.58
					19.83
					31.16
					12.58
					22.74
					33.65
					17.54
					23.56
					24.03
					17.81
					33.84
					14.58
					36.91
					149.89
					58.43
					14.07
					14.32
28088.1	\$239.70	30825	82965	483	58.15
6150.098	\$52.78	8711	19355	483	12.73
8726.4	\$74.90	4978.53	27043.88	483.76	18.04
22391.1	\$191.39	6409	35665	484	46.26
17417.16	\$148.64	22493	7486	484	35.99
6733.8	\$57.71	3577	15033	484.0665	
11568.75	\$98.92	9021	52180	485	23.85

Figure 2a: Calculating Descriptive Statistics for Each Column

Data Summarization in EXCEL

A	B	C	D	E	F	G	H	I	J	K	L
Market_Cap_Yen	Market Cap US \$		Book Equity		Revenues		Net Income		PE		
Mean	15986.97339	Mean	136.605	Mean	10364.32139	Mean	25874.11889	Mean	473.1043472	Mean	34.36142857
Standard Error	2537.920033	Standard Error	21.61534242	Standard Error	1319.055993	Standard Error	3575.977824	Standard Error	1.231357085	Standard Error	5.496326537
Median	10743.005	Median	91.915	Median	8462.5	Median	19174.5	Median	472.5	Median	23.11
Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	466	Mode	#N/A
Standard Deviation	15227.5202	Standard Deviat	129.6920545	Standard Deviat	7914.335959	Standard Deviat	21455.86695	Standard Deviat	7.388142511	Standard Deviat	32.51670631
Sample Variance	231877371.3	Sample Varianc	16820.02901	Sample Varianc	62636713.68	Sample Varianc	460354226.4	Sample Varianc	54.58464977	Sample Varianc	1057.336189
Kurtosis	7.517096582	Kurtosis	7.508480838	Kurtosis	0.356424827	Kurtosis	1.630812188	Kurtosis	-1.430056224	Kurtosis	7.241614204
Skewness	2.734339301	Skewness	2.733023356	Skewness	1.029985399	Skewness	1.411894106	Skewness	0.239862517	Skewness	2.700000876
Range	65728.87	Range	559.84	Range	29578	Range	83994.13	Range	21.76	Range	137.31
Minimum	5926.8	Minimum	50.88	Minimum	1247	Minimum	2138.87	Minimum	463.24	Minimum	12.58
Maximum	71655.67	Maximum	610.72	Maximum	30825	Maximum	86133	Maximum	485	Maximum	149.89
Sum	575531.042	Sum	4917.78	Sum	373115.57	Sum	931468.28	Sum	17031.7565	Sum	1202.65
Count	36	Count	36	Count	36	Count	36	Count	36	Count	35
Confidence Level(9	5152.251547	Confidence Level	43.88147775	Confidence Level	2677.826013	Confidence Level	7259.620886	Confidence Level	2.499787765	Confidence Level	11.16987936

Figure 2a: Descriptive Statistics for Each Column of Japan.xlsx data file

2(b). Descriptive Statistics for PE by Market Cap Categories

To calculate descriptive statistics for PE by the following categories,

Market Cap	Market Cap Code
US\$ 50 to 100	1
US\$100 to 200	2
US\$200 to 300	3
> US\$300	4

we first have to create a variable Cap Code in Excel, as follows:

In cell H1, type the variable name Cap Code.

In cell H2, type the formula `=IF(C2<100,1,IF(C2<200,2,IF(C2<300,3,4)))` and copy formula in cell H2 and then paste it into the rows H2:H37. This will create all of the values of Cap Code column.

Clipboard		Font		Alignment		Number			
H2		fx		=IF(C2<100,1,IF(C2<200,2,IF(C2<300,3,4)))					
	A	B	C	D	E	F	G	H	I
1	Short Name	Market_Cap_Yen	Market Cap US \$	Book Equity	Revenues	Net Income	PE	Cap Code	PE_Code
2	CENTURY21 REAL	14496	\$123.72	1924.79	2452.09	463.24	31.29	2	1
3	BALS CORP	43231.88	\$369.01	3239	18994	464	93.17	4	
4	MAINICHI COMNET	10723.2	\$91.69	2675	6449	464	23.11	1	
5	MUTUAL	5990.63	\$51.42	6915	10100	464	12.91	1	
6	NISSIN SHOJI CO	9728	\$83.18	17129	70529	464	20.97	1	
7	TOW CO LTD	8030.752	\$68.94	3782	10705	465	17.27	1	
8	CLEX CO LTD	10105.62	\$86.41	4455	9486	466	21.69	1	
9	FENWAL CONTROLS	6334.975	\$54.37	4357	14930	466	13.59	1	
10	FUJII SANGYO	8008	\$68.62	11392	16927	466	17.18	1	

Figure 2b: Creat Cap Code Column for data of Example 2.

Data Summarization in EXCEL

Next, we calculate the descriptive statistics of PE by the variable Market Cap by using Pivot Table in Excel.

Click on Insert/PivotTable/PivotTable which will open the Pivot Table Wizard (see Figure 2c)

Right click on Cap_Code, click on 'Add to Row Labels' (Figure 2d)

Right click on PE, then click on 'Σ Add to Values' which will result in 'Count of PE'; click on the arrow next to it, and select 'Average' (Figure 2e). Repeat this process 3 more times, changing 'Count of PE' to 'StDev of PE', 'Min of PE', and 'Max of PE', which will result in the output shown in Figure 2f.

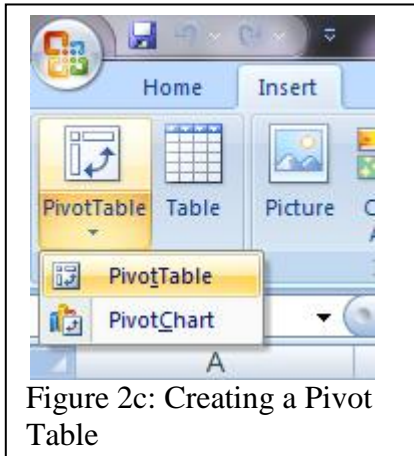


Figure 2c: Creating a Pivot Table

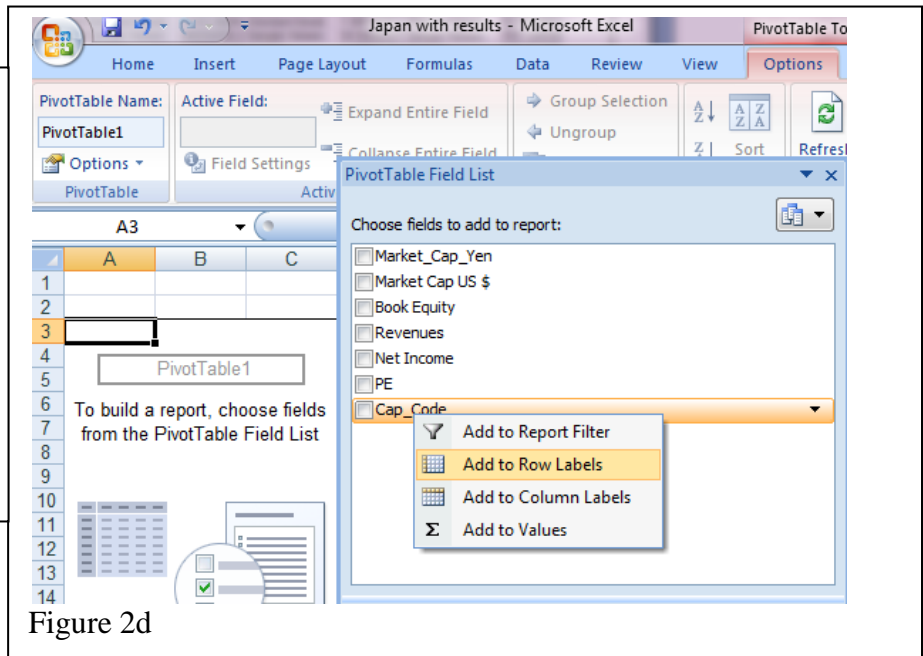


Figure 2d

Data Summarization in EXCEL

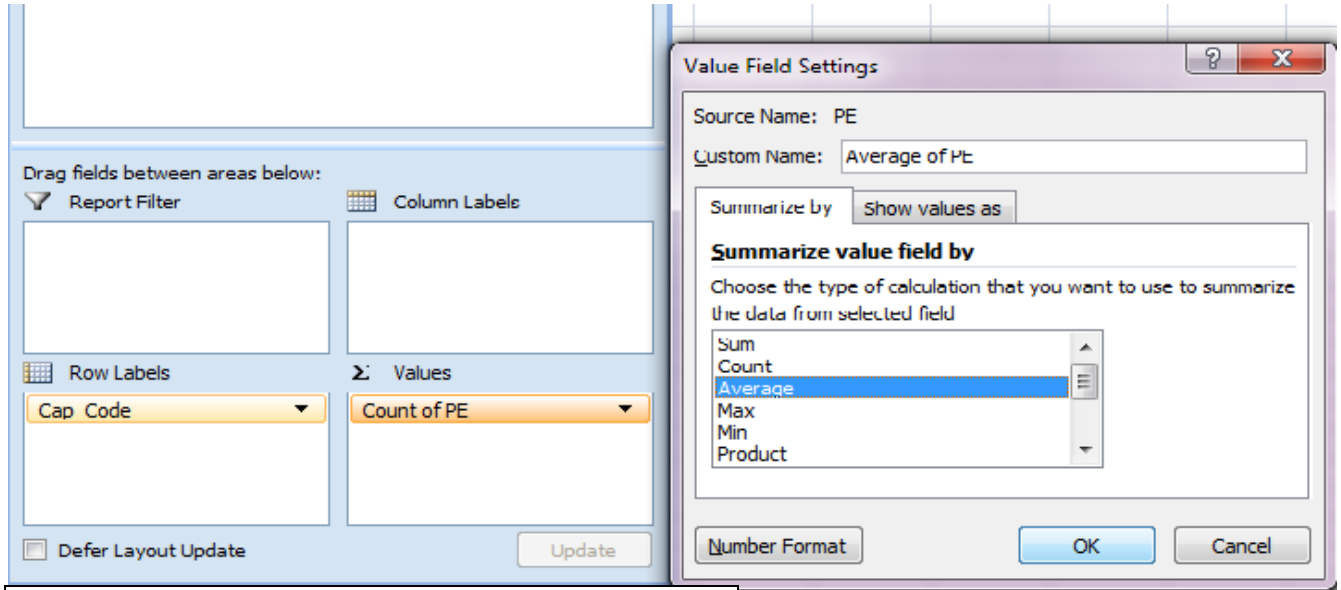


Figure 2e: Change Count of PE to Average of PE.

	Values			
Row Labels	Average of PE	StdDev of PE	Min of PE	Max of PE
1	18.59409091	4.196291627	12.58	24.03
2	100.656	44.49001663	58.15	149.89
3	36.2875	5.186140183	31.16	46.26
Grand Total	34.36142857	32.51670631	12.58	149.89

Figure 2f: Mean, sd, min and max of PE by Cap_Code

2(c). Scatter Plot of Revenues vs. Book Equity

Open the data file JAPAN.xlsx, highlight cells D1:E37 (Revenues and Book Equity Columns, including the titles), then INSERT/SCATTER/click on the first icon (see Figure 2g) to get the graph of Figure 2h.

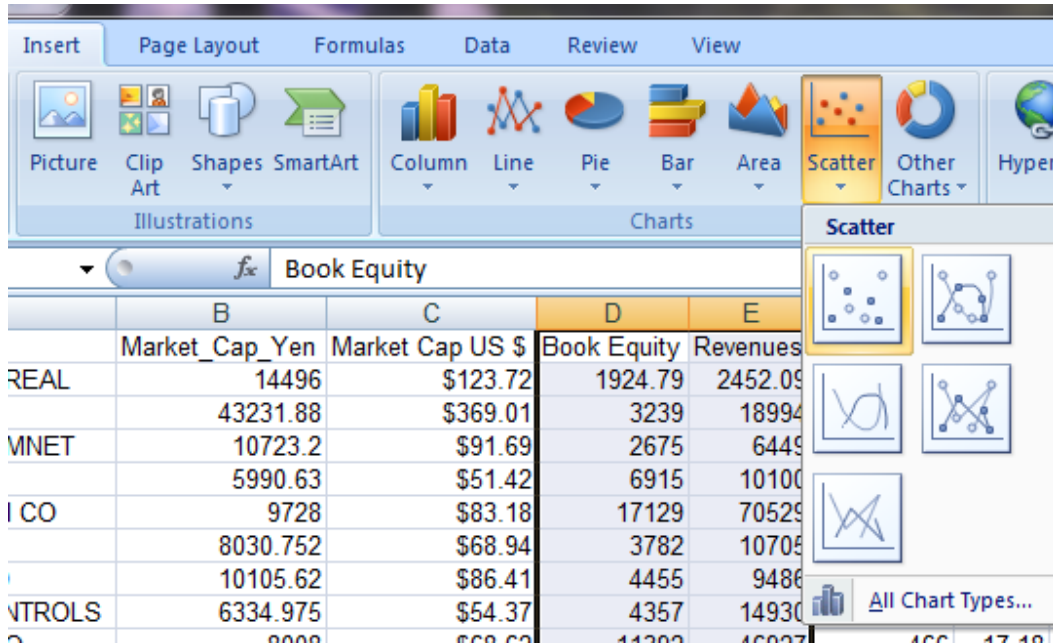


Figure 2g: Drawing Scatter Plot of Book Equity vs. Revenues in EXCEL

Data Summarization in EXCEL

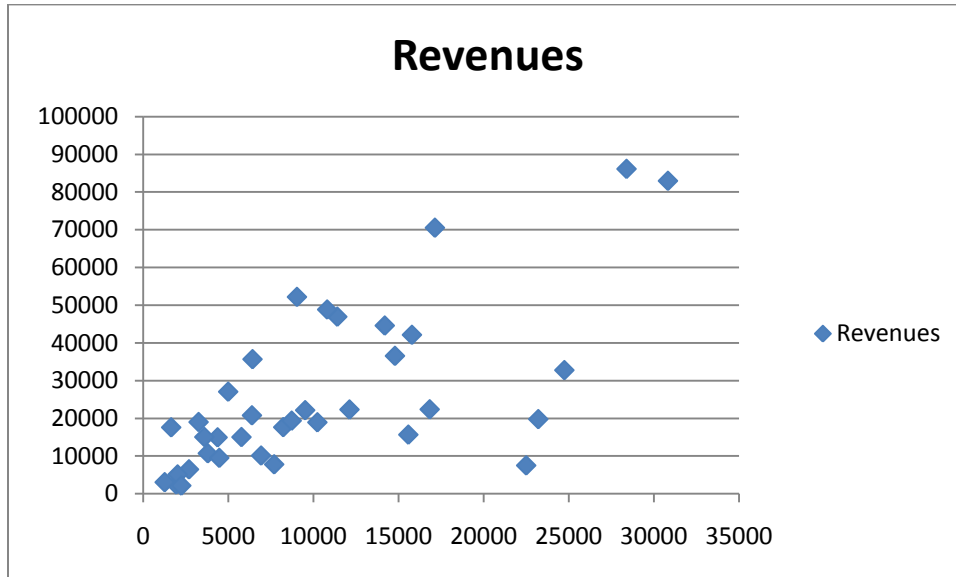


Figure 2h: Scatter Plot of Book Equity vs. Revenues from EXCEL

To add X-axis and Y-axis titles in EXCEL, click anywhere on the graph, then click on Layout/Axis Titles/Primary Horizontal Axis Title/Title Below Horizontal Axis, (see Figure 2i), type Book Equity as your x-axis title. Repeat for y-axis title, delete the legend Revenues on the right in the chart area to get the final Scatter Plot of Figure 2j.

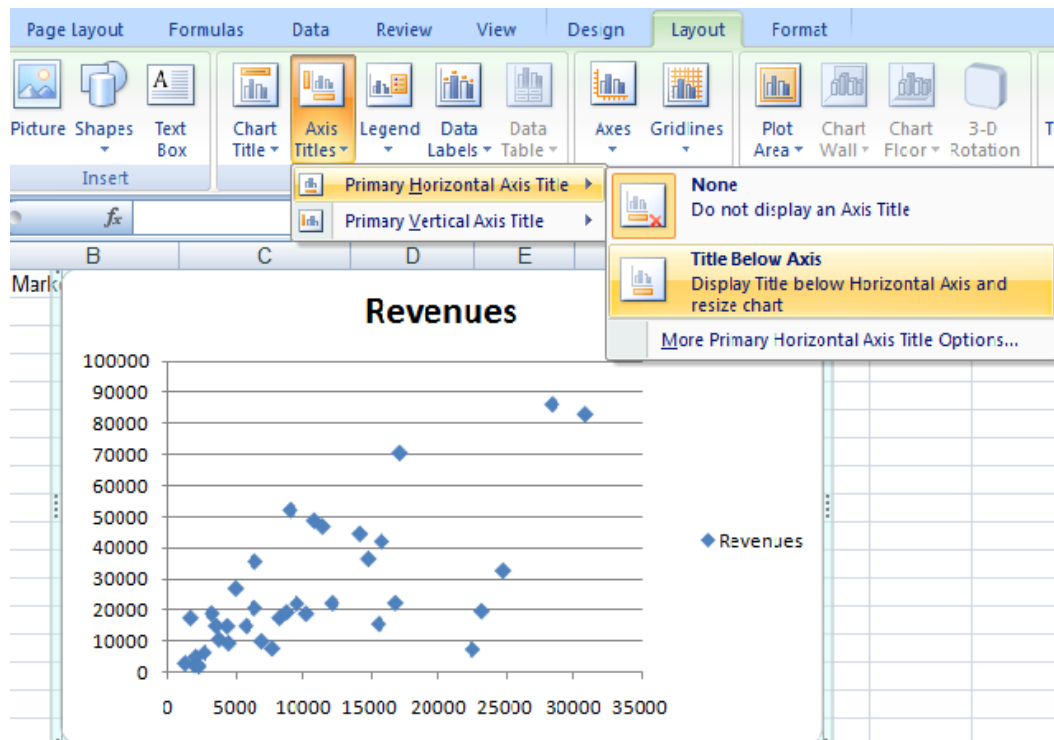


Figure 2i: Adding x-axis title in EXCEL

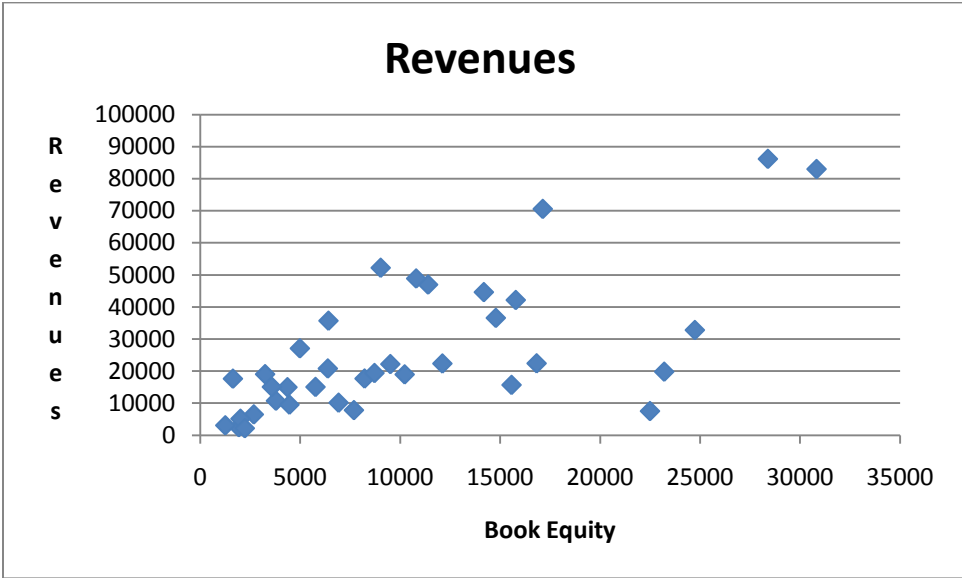


Figure 2j: Final Scatter Plot of Book Equity vs. Revenues from EXCEL

2(d) Correlation Matrix

To compute the correlation matrix for all of the variables in the datafile JAPAN.xlsx, click on:

Data/Data Analysis/Correlation, select Input Rang, Grouped By Columns, enter a name for New Worksheet, click on OK (see Figure 2k) to obtain the correlation matrix shown in table 1 on page 14.

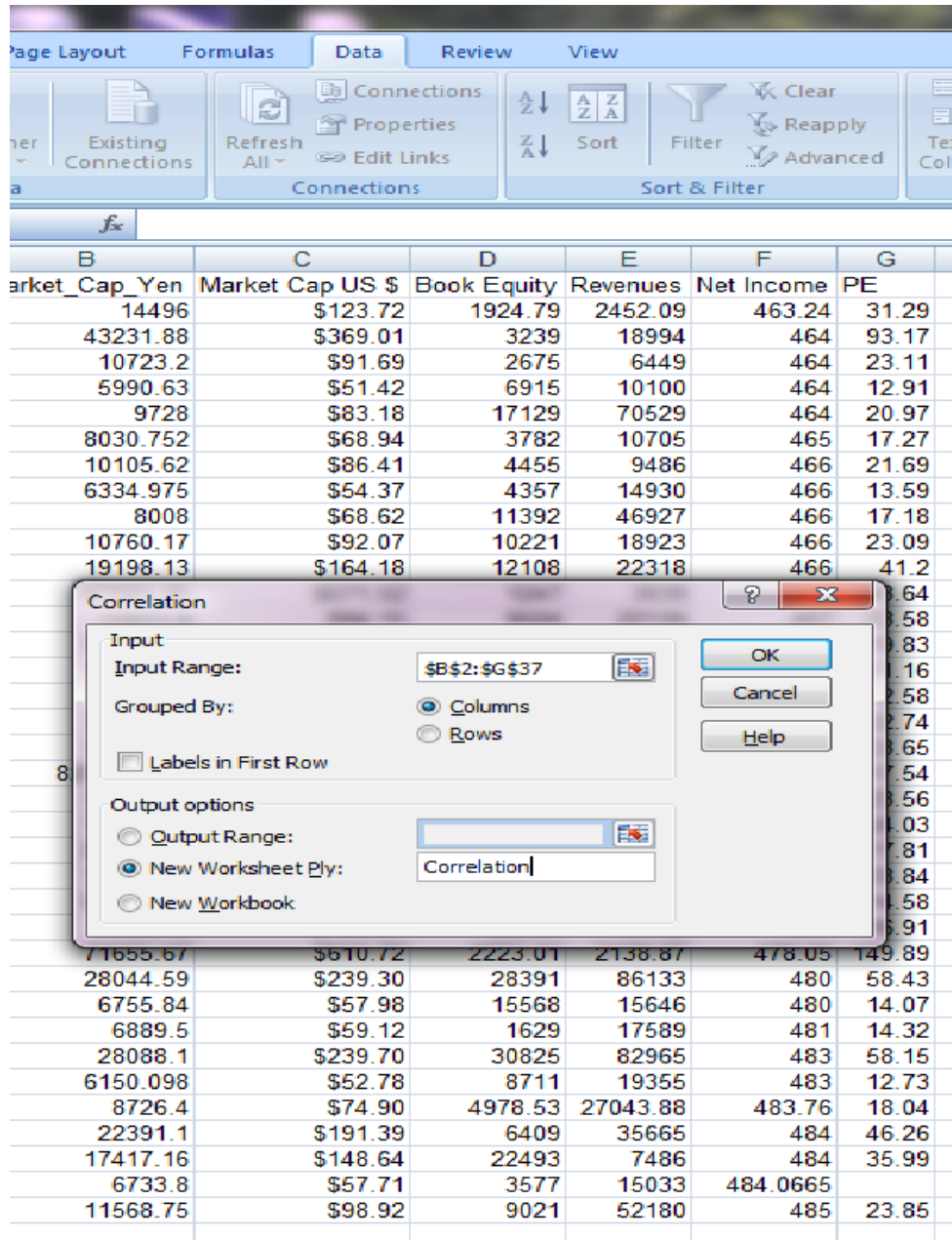


Figure 2k: Computing Correlation Matrix for data of Example 2.

Data Summarization in EXCEL

Table 1: Correlation Matrix for data of Example 2

	<i>Column</i> 1	<i>Column</i> 2	<i>Column</i> 3	<i>Column</i> 4	<i>Column</i> 5	<i>Column</i> 6
Column 1	1					
Column 2	1	1				
Column 3	0.04479	-0.04479	1			
Column 4	0.06351	-0.06338	0.675578	1		
Column 5	0.01068	0.010747	0.275668	0.223344	1	
Column 6	0.99978	0.999779	-0.06929	-0.07978	0.022446	1