EXAMPLES - SAMPLING DISTRIBUTION

EXCEL INSTRUCTIONS

This exercise illustrates the process of the sampling distribution as stated in the Central Limit Theorem. Enter the actual data in Column A in MICROSOFT EXCEL. The data represents a survey of the number of hours that the commuters spent in their cars each day. Use the EXCEL - Tools -Data Analysis. You may need to Add-in the Analysis Tool Pak every time you use the network. This project is to generate 100 observations of n = 5.

Step 1: Enter the actual data and labels

Step 2: Select **Sampling** from **Tools - Data Analysis**. Highlight the **Input Range** with only the data your Column A (do not include your labels or titles)

Type in 100 in the box corresponding to the **Random Number of Samples**

Type in the **Output Range** B2:B102

Select **OK**

You will notice on your spreadsheet that EXCEL has generated a column of numbers randomly from your original set of data. If you compare your results in Column B with the person next to you, your sample values may not be the same as your neighbor.

Step 3: Go to Column C. Use the function AVERAGE to calculate the mean values of n = 5. Column C is the column of the means.

Step 4: Delete the selected cells in the Column C. Notice the cells in the AVERAGE function do not contain any data values.

Step 5: Select **Histogram** from **Tools - Data Analysis**. You will be making two different histograms. The first histogram is your actual data in Column A. So highlight your actual data in the **Input Range**. Type in Histogram for Actual Data under **New Worksheet**.

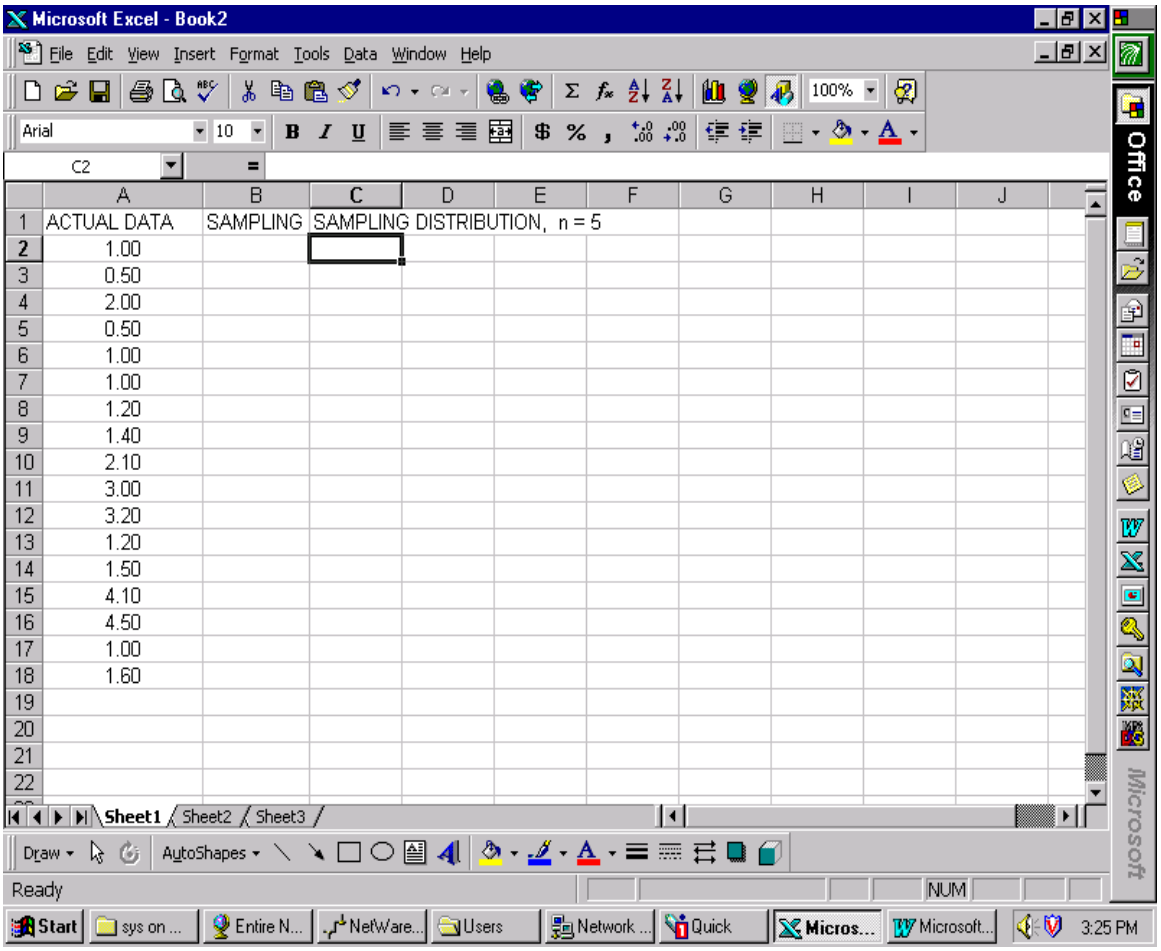
The second histogram is your distribution of the means in Column C. Highlight the data in the **Input Range**. Type in Histogram for Sampling under **New Worksheet**.

Step 6: Use the Chart Wizard to plot out these two histograms.

Step 7: Save your file

EXCEL PROJECT (CONT.)

Step 1. Type in the data in the ACTUAL DATA column



EXCEL PROJECT (CONT.)

Step 2: Select **Sampling** from **Tools - Data Analysis**. Highlight the **Input Range** with only the data your Column A (do not include your labels or titles)

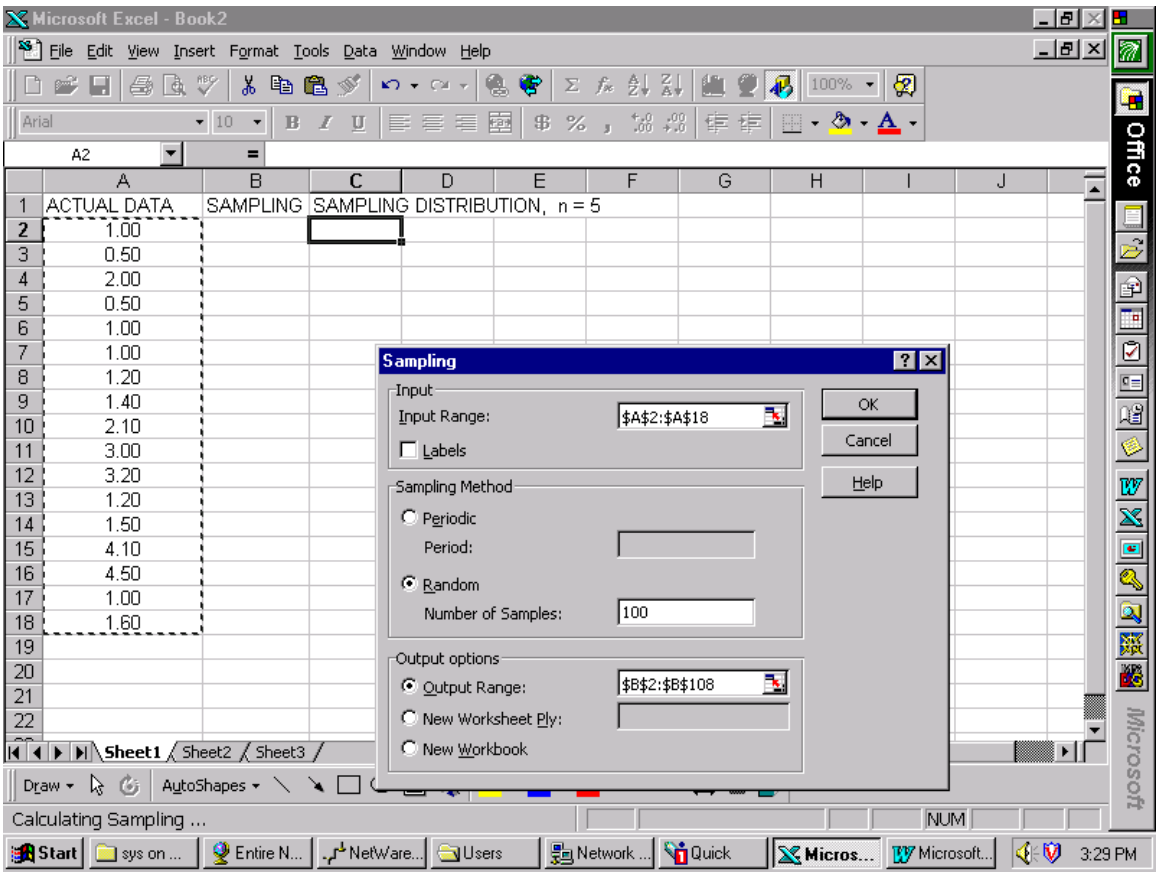
Type in 100 in the box corresponding to the **Random Number of Samples**

Type in the **Output Range** B2:B108

Select **OK**

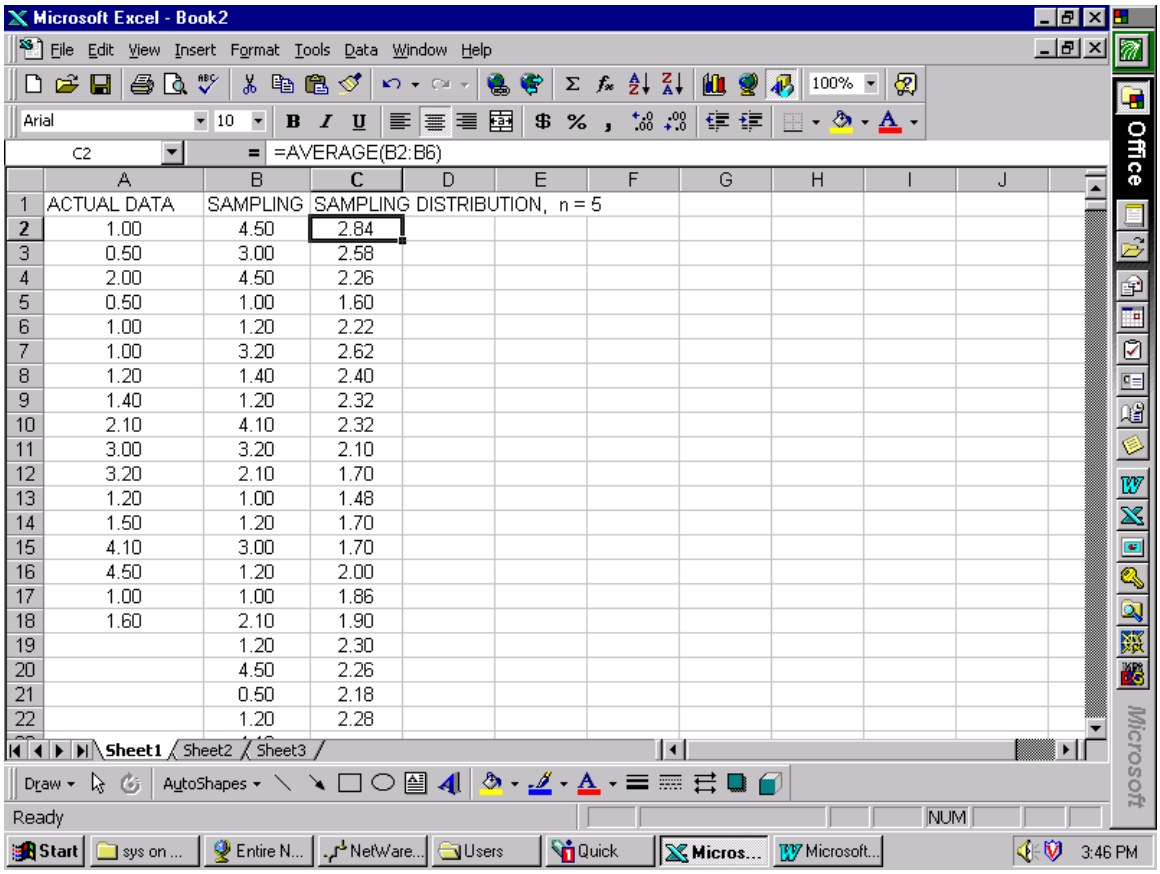
You will notice on your spreadsheet that EXCEL has generated a column of numbers randomly from your original set of data. If you compare your results in Column B with the person next to you, your sample values may not be the same as your neighbor.

However, the data in the SAMPLING and SAMPLING DISTRIBUTION columns in your spreadsheet may not be identical to what is printed here.



EXCEL PROJECT (CONT.)

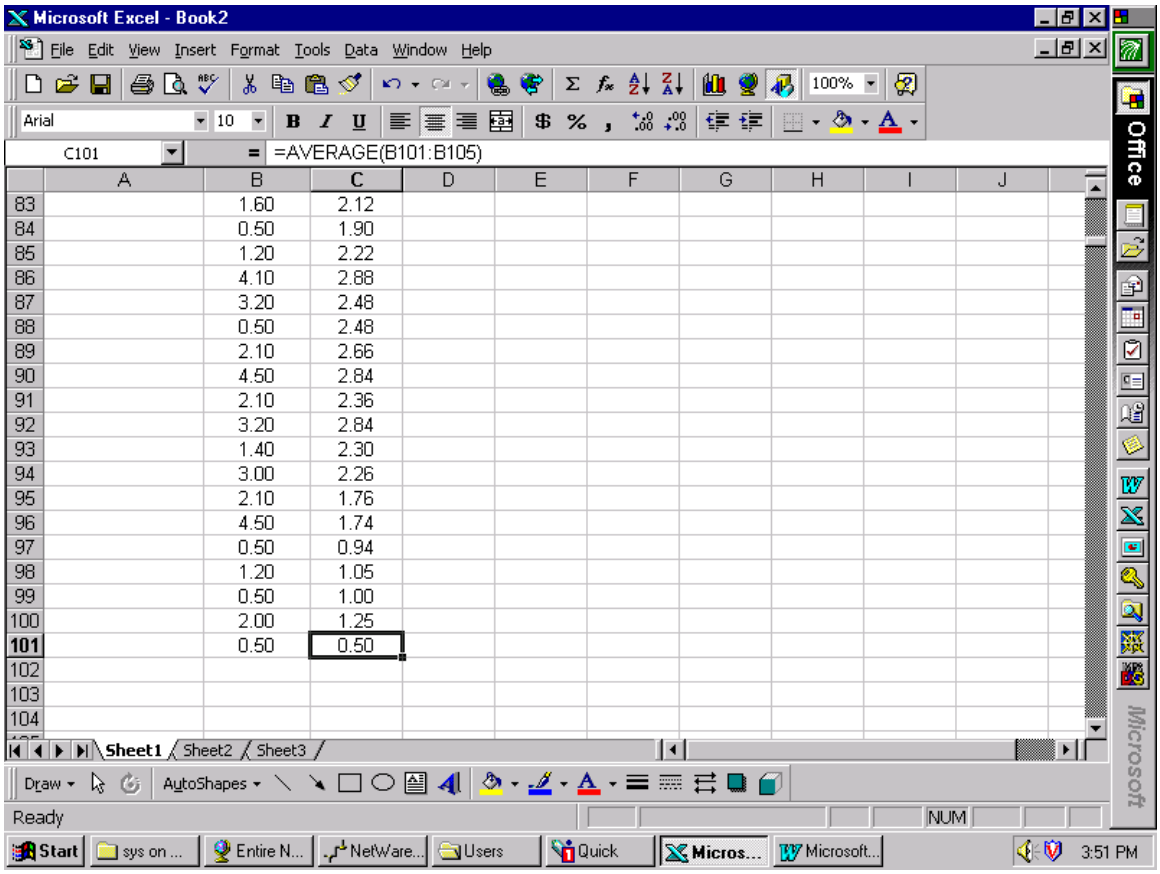
Step 3: Go to Column C. Use the function AVERAGE to calculate the mean values of n = 5. Column C is the column of the means.



	A	B	C	D	E	F	G	H	I	J
1	ACTUAL DATA	SAMPLING	SAMPLING DISTRIBUTION, n = 5							
2	1.00	4.50	2.84							
3	0.50	3.00	2.58							
4	2.00	4.50	2.26							
5	0.50	1.00	1.60							
6	1.00	1.20	2.22							
7	1.00	3.20	2.62							
8	1.20	1.40	2.40							
9	1.40	1.20	2.32							
10	2.10	4.10	2.32							
11	3.00	3.20	2.10							
12	3.20	2.10	1.70							
13	1.20	1.00	1.48							
14	1.50	1.20	1.70							
15	4.10	3.00	1.70							
16	4.50	1.20	2.00							
17	1.00	1.00	1.86							
18	1.60	2.10	1.90							
19		1.20	2.30							
20		4.50	2.26							
21		0.50	2.18							
22		1.20	2.28							

EXCEL PROJECT (CONT.)

Step 4: Delete the selected cells in the Column C. Notice the cells in the AVERAGE function do not contain any data values.

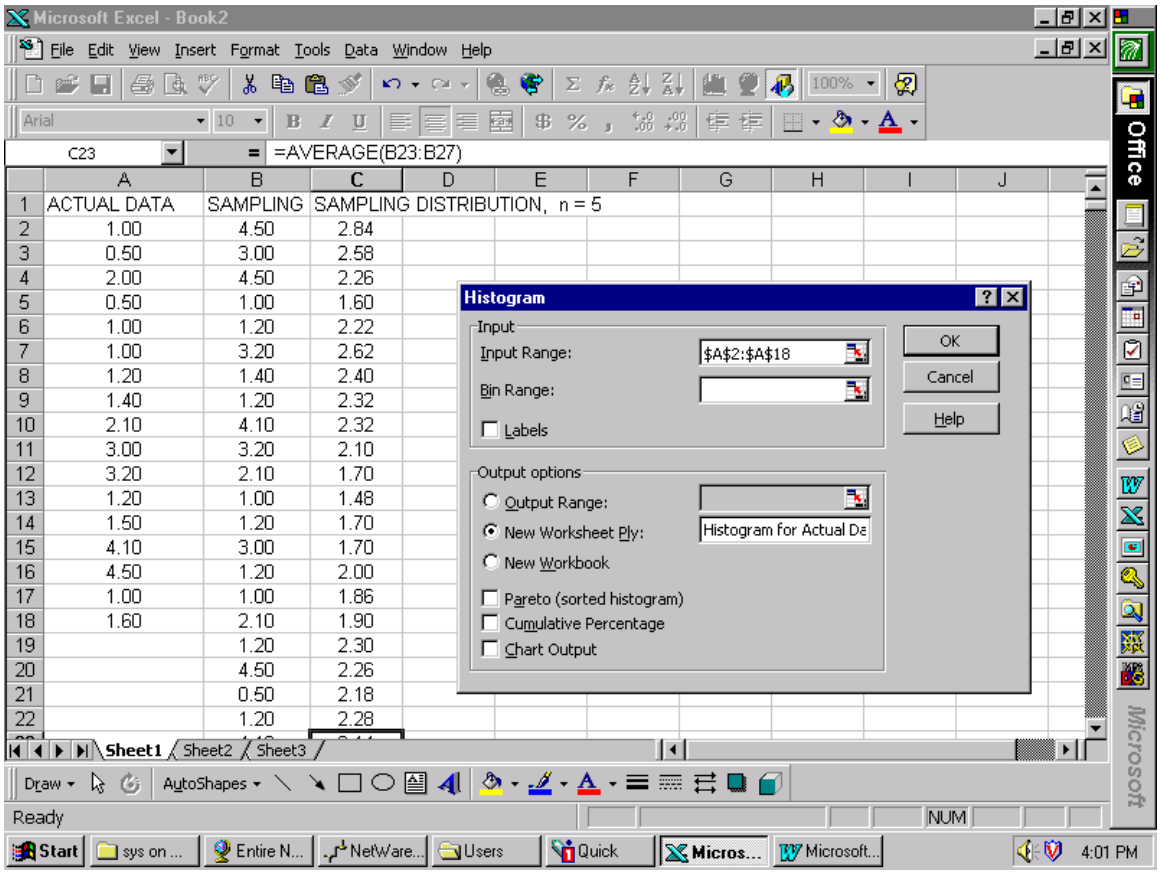


	A	B	C	D	E	F	G	H	I	J
83		1.60	2.12							
84		0.50	1.90							
85		1.20	2.22							
86		4.10	2.88							
87		3.20	2.48							
88		0.50	2.48							
89		2.10	2.66							
90		4.50	2.84							
91		2.10	2.36							
92		3.20	2.84							
93		1.40	2.30							
94		3.00	2.26							
95		2.10	1.76							
96		4.50	1.74							
97		0.50	0.94							
98		1.20	1.05							
99		0.50	1.00							
100		2.00	1.25							
101		0.50	0.50							
102										
103										
104										

EXCEL PROJECT (CONT.)

Step 5: Select **Histogram** from **Tools - Data Analysis**. You will be making two different histograms. The first histogram is your actual data in Column A. So highlight your actual data in the **Input Range**. Type in Histogram for Actual Data under **New Worksheet**.

The second histogram is your distribution of the means in Column C. Highlight the data in the **Input Range**. Type in Histogram for Sampling under **New Worksheet**.

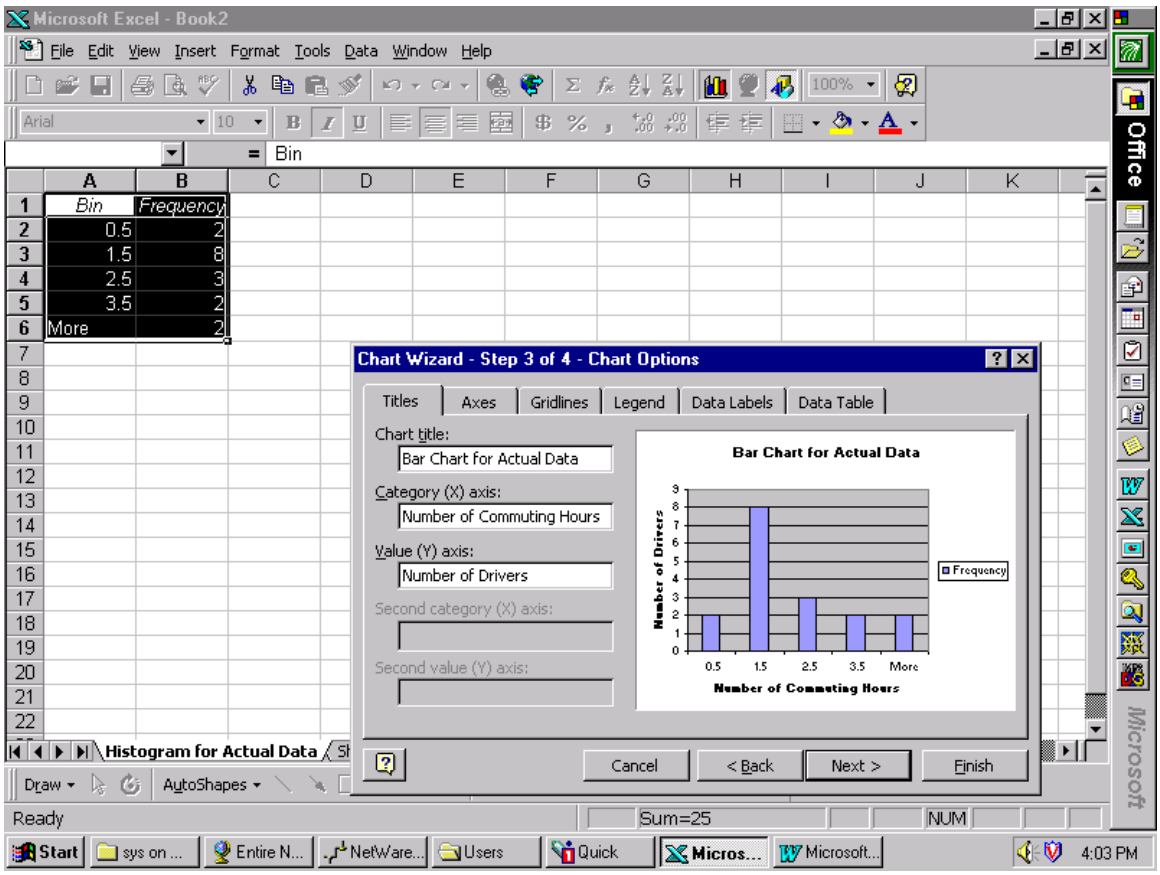


	A	B	C
1	ACTUAL DATA	SAMPLING	SAMPLING DISTRIBUTION, n = 5
2	1.00	4.50	2.84
3	0.50	3.00	2.58
4	2.00	4.50	2.26
5	0.50	1.00	1.60
6	1.00	1.20	2.22
7	1.00	3.20	2.62
8	1.20	1.40	2.40
9	1.40	1.20	2.32
10	2.10	4.10	2.32
11	3.00	3.20	2.10
12	3.20	2.10	1.70
13	1.20	1.00	1.48
14	1.50	1.20	1.70
15	4.10	3.00	1.70
16	4.50	1.20	2.00
17	1.00	1.00	1.86
18	1.60	2.10	1.90
19		1.20	2.30
20		4.50	2.26
21		0.50	2.18
22		1.20	2.28

EXCEL PROJECT (CONT.)

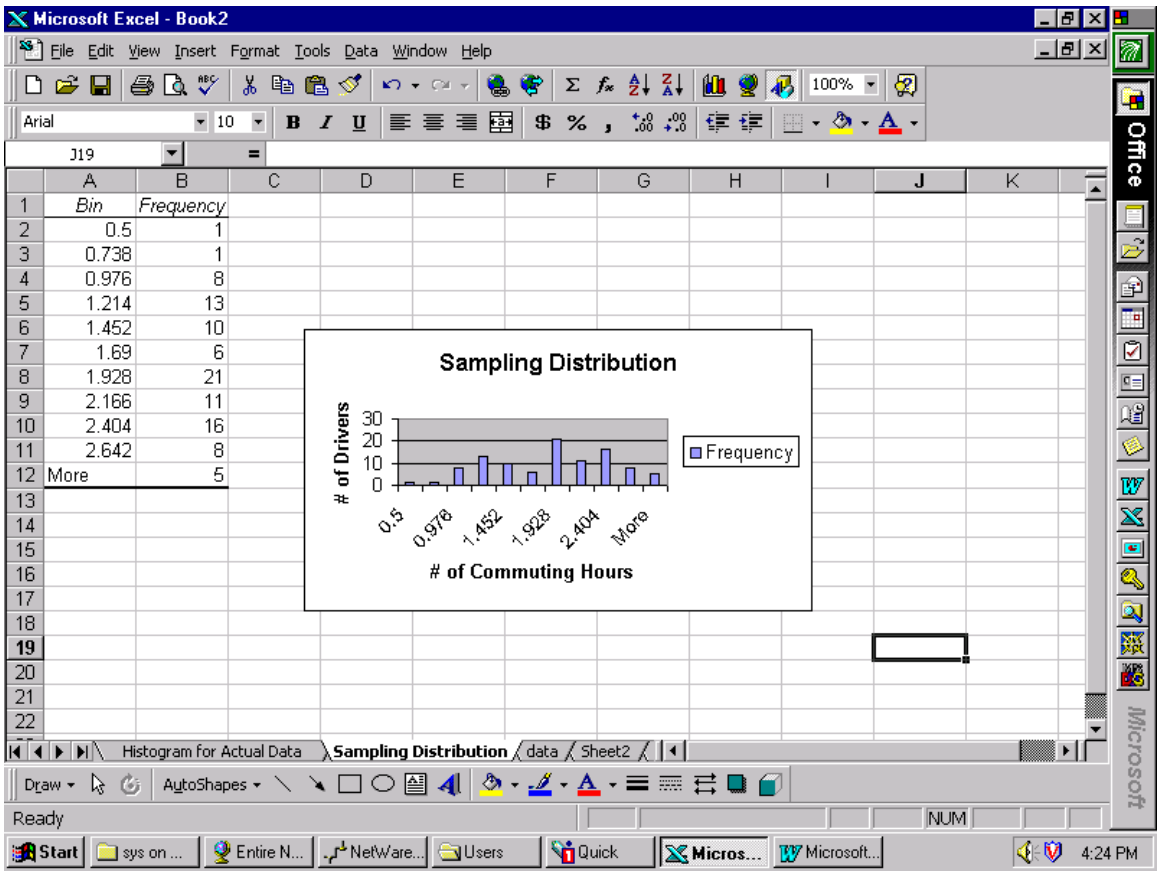
Step 6: Use the Chart Wizard to plot out these two histograms. The histogram for the Actual

Data is shown here.



EXAMPLES - SAMPLING DISTRIBUTION (CONT.)

Step 6: Use the Chart Wizard to plot out these two histograms. The histogram for the Sampling Distribution is shown here. Your sampling distribution will be different from the chart below.



	A	B
1	*Bin*	*Frequency*
2	0.5	1
3	0.738	1
4	0.976	8
5	1.214	13
6	1.452	10
7	1.69	6
8	1.928	21
9	2.166	11
10	2.404	16
11	2.642	8
12	More	5

Income Analysis Project

Based on past experience, the annual pay for a person with a Bachelor degree in one Business school with less than three year of experience from the sample taken by a recruiting officer is as follows:

78000, 65000, 58500, 67000, 51250, 54000, 69500, 58000, 72000, 76850,
73500, 79000, 72000, 68000, 61000, 66000, 64750, 61500, 75500, 64000.

Use EXCEL to generate 250 random numbers from this sample. Form the distribution of means with size = 10.

(i) Print the entire sheet with the actual data and the numerical data of the sampling distribution using the font size of 5. What is the mean value and the standard deviation.

(ii) Print the histogram of the actual data and the column chart (the bar graph)

(iii) Print the histogram of the sampling distribution and the column chart (the bar graph).

(iv) With the mean value and standard deviation obtained in (i), apply the command **NORMINV(probability,mean,standard_dev)** to generate 20 normal distributed numbers with two decimal places. Based on the generated numbers, do (i) (ii) and (iii) accordingly.