Alternative ways to create histograms in Excel Phys290 Fall 2010

Excel was not intended for scientific data analysis but it and its clones (like OpenOffice Spreadsheet) are generally available, so we use them when necessary. However, on some recent versions of Excel, Microsoft apparently eliminated the useful "Data Analysis Toolpack" that formed the basis for the previously posted set of instructions to create a histogram.

Since histograms are so important for data analysis, we offer some alternative ways of using Excel to create and plot them. Google will find other examples of instructions to create histograms.

All the functions you need to create a histogram are already built in to all versions of Excel, so you can use them in sequence to tabulate data, create histograms and then plot the result in various ways. Here is a step-by-step set of instructions for Windows, Excel 2003. The screen formats will look different in other versions but the functions are the same.

In this example, we will use the function "COUNTIF" to create the actual histogram and then plot the result. How COUNTIF works will be explained later.

Procedure:

1. Open Excel

2. Enter the data for the histogram. Here, I'm using the first 100 entries in the Atrium_all_2008.txt file which I opened using the **File menu.** On the next page, I show the result. The data are in the A column.

3. In the B column, type min, max, avg and sigma into the four top cells.

4. In the C column, find the minimum, maximum, average and standard deviation of the first 100 entries in column A

5. To do this, enter into cells C1-C4 (The "=" says the cell contains a formula) =MIN(A1:A100) =MAX(A1:A100) =AVERAGE(A1:A100) =STDEV(A1:A100)

N	licrosoft E	xcel - A	trium	_all_200	8.txt				
:1	<u>E</u> ile <u>E</u> dit	⊻iew	Insert	F <u>o</u> rmat	<u>T</u> ools	<u>D</u> ata	<u>W</u> indow	Help	Ado <u>b</u> e PDF
10	💕 🖬 🛛	3 3		۶ 📖 ۱	6 🗈 🕻	L • 🥥	r 19 +	(* - 8	SΣ → AZ↓
:	12 22 2	1 🔁 1		5 3 3) VV R	eply with	Changes	. E <u>n</u> d Review
	B5	-	fx	1					
	A	В		С	D		E	F	G
1	1490	min		1180					
2	1570	max		1570					
3	1320	avg		1371.77					
4	1570	sigma	7	2.85531					
5	1400								
6	1500								
7	1399								
8	1570								
9	1410								
10	1320								
11	1470								
12	1410								
13	1320								
14	1440	-							
15	1260								
16	1440								
17	1550								
18	1440								
19	1350								
20	1390								
21	1460								
22	1457								
	1 100			1					1

6. Select a suitable X-axis range for the histogram. Note that the minimum and maximum are 1180 and 1570. A convenient range for the histogram might be 1100 to 1600, with steps of 10.

7. Create a column of values in steps of 10 beginning at 1100 and continuing to 1600. To do this, enter 1100 in one cell and 1110 in the next. This sets the pattern for the column. Select both of these cells. Note that there is a small square dot on the lower right corner of the two boxed cells, like this: (think of it as a pull tab)

× N	Aicrosoft E	xcel - A	trium_	_all_200)8.txt			
:2	<u>Eile E</u> dit	⊻iew	Insert	Format	<u>T</u> ools	<u>D</u> ata	<u>W</u> indow	, <u>H</u> elp
1	📂 🖬 🕻	16	ABC	P 🛍)	6 🗅 🛍	- 🛷	1-17 -	(* + §
1		1 🗠 🖄		3 3	B (2)	₩¢ Re	eply with	i <u>C</u> hanges
	E2	-	fx	1100				
	A	В		С	D		E	F
1	1490	min		1180		bins		
2	1570	max		1570			1100	
3	1320	avg		1371.77			1110	
4	1570	sigma	72	2.85531				
5	1400							
6	1500							
7	1399							
8	1570							
9	1410							
10	1320							
11	1470							
12	1410							
13	1320							
14	1440							
15	1260							
4.0	1.1.40							

8. Grab the square dot with the mouse/cursor and pull it down to create the list of BIN VALUES. See

1	📂 🖬 🕻		l 🧐 🛍 🐰		l • 💞 I 🔊 • (
	220		1501	5 p	Reply with C
E53		•			
	A	В	C	D	E
1	1490	min	1180		bins
2	1570	max	1570		1100
3	1320	avg	1371.77		1110
4	1570	sigma	72.85531		1120
5	1400	100			1130
6	1500				1140
7	1399				1150
8	1570				1160
9	1410				1170
10	1320				1180
11	1470				1190
12	1410				1200
13	1320				1210
14	1440				1220

8. Next, we create the actual histogram. Use the function COUNTIF.

COUNTIF(A,B) counts the number of occurrences of a value in the range A that satisfy the criterion B. For example, COUNTIF(A1:A100, 1) returns zero, because there are no values equal to 1 in cells A1 to A100. We want to know how many values fall in a range of, say 1100 to 1109, which are the bin values in column E. One way to do this is to use the function COUNTIF twice in one cell, first to count the values that are >=1100 and subtract them from the number that are >=1110.

9. In Column F, cell 2, type the formula

=COUNTIF(\$A\$1:\$A\$100, ">="&E2) -COUNTIF(\$A\$1:\$A\$100,">="&E3)

10. Copy cell F2 down the column to F52 and your spreadsheet should look like this:

1 N	Aicrosoft E	xcel - Atrii	um_all_200	8.txt						
:2)	<u>Eile E</u> dit	<u>V</u> iew <u>I</u> ns	ert F <u>o</u> rmat	<u>T</u> ools <u>D</u>	ata <u>W</u> indov	w <u>H</u> elp A	do <u>b</u> e PDF			
	😂 🖬 🖁		1 🍄 🛍 🛛 🐰	- Ca 😤	- 🥑 🔊 -	CI + 18	Σ - <u>A</u> ↓	ZI I III A	100%	- (
: (3)	(4 1 4 1 (2		SXIB	助商	Y Reply with	n Channes	End Review		-	
	F2	-	£ =COUNT	TE(SAS1.S	:A\$100 ">='	'&E2)-COUI		\$A\$100 ">	="&F3)	
-	Δ	B		D	F	F	G G	. ф/ ф100, >	1	
1	1490	min	1180	U	hins	count	~			-
2	1570	max	1570		1100	n n				-
3	1320	avo	1371.77		1110					-
4	1570	siama	72.85531		1120	Ō				1
5	1400				1130	0				-
6	1500				1140	0				
7	1399				1150	0				
8	1570				1160	0				
9	1410				1170	0				
10	1320				1180	1				
11	1470		1		1190	0				
12	1410				1200	0				
13	1320				1210	0				
14	1440				1220	0				
15	1260				1230	0				
16	1440				1240	0				
17	1550				1250	1				
18	1440				1260	2				
19	1350				1270	5				1
20	1390				1280	2				
21	1460				1290	1				
22	1457				1300	4				
23	1460				1310	0				
24	1386				1320	7				
25	1270				1330	10				
26	1386				1340	10				
27	1360				1350	2				
28	1333				1360	Q				

Columns E and F are your histogram! All you need to do now is to plot it. Here is an example of the many, many ways to do this:

11. Select Insert Chart from the menu. Bar/Columns are fine

3 N	icrosoft E	xcel - A	trium_all_200	8.txt									
2	Eile Edit	View	Insert Format	Tools	Data Windo	v <u>H</u> elp	Ado <u>b</u> e PC	F					
D			a. 1 49 10 1)	Ca 🙉	- J .	(Σ-		100% -		rial	- 1	0 - B
-	dia dia 15			R. alt	Weineshund	h Changes	Fed Do					total (the Party of
-	un un	-			I I A Kophy and	n Strenges v	rilla veo						
	112 A	-		D	E	E	G			1	L V	1	M
	1/90	min	1180	U	hinc	count	0			J	N	L	111
2	1430	max	1570		1100	Count		-					
3	1320	970	1371 77		1110	0							
4	1570	sigma	72 85531		1120	0							
5	1400	Signia	12.00001		1130	0							
6	1500				1140	0		Chart Wizard	- Step 1 of 4	- Chart	Туре		?
7	1399				1150	0					a delan		
8	1570				1160	0		Standard Types	Custom Type	ts			
9	1410				1170	C C		Charthan		Charle	a de la cara a		
10	1320				1180	1		Chart type:		Unart	sub-gype:	- 1	
11	1470				1190	0		Column	^				
12	1410				1200	0		EBar					
13	1320				1210	C		Line Line					
14	1440				1220	0		Pie					
15	1260				1230	0	(XY (Scatter)				
16	1440				1240	0		Area					
17	1550				1250	1		🙆 Doughnut	_				
18	1440				1260	2		Radar 🔬		1			
19	1350				1270	5		Surface					
20	1390				1280	2		Subble	~				
21	1460				1290	1							
22	1457				1300	4				Cluste	red Column. C	ompares val	les across
23	1460				1310	0				categ	ories.		
24	1386				1320	7							
25	1270				1330	10				-			
26	1386				1340	10					Press and Holi	d to <u>V</u> iew Sar	nple
27	1360				1350	2							
28	1333				1360	8			Cancel	< Ba	ck Ne:	d >	Einish
29	1370				1370	5					-		
30	1450				1380	8							
31	1430				1390	5							

•

12. The next chart option is **"Source Data"**, with options **"Data Range"** or **"Series"**. Either will work, but "Data Range" was selected as shown next.

	Aicrosoft E	xcel - Atr	ium_all_200	8.txt				
:101	<u>Eile E</u> dit	<u>V</u> iew Ir	nsert F <u>o</u> rmat	Tools Da	ata <u>W</u> indow	<u>H</u> elp Ad	o <u>b</u> e PDF	
10			1 2 13	608.	- (*) +	0-18	$\Sigma = \frac{A}{Z} \begin{bmatrix} Z \\ A \end{bmatrix}$	10 - B Z
100	1 40 80 2		153	助商门	Reply with	Changes I	ind Review	
-	F2	-	fx.					
	A	В	C	D	E	F	G H	HIJKLM
7	1399				1150	Ū,		
8	1570				1160	0		
9	1410				11/0(Source D	Data 🔹 🚺
10	1320				1100		Data Bas	DOD Cavita
12	1470		-		1200	0	Data Rai	ige series
13	1320				1210	ō	12	
14	1440				1220	0	10	
15	1260				1230	0		
16	1440				1240	0		Series 1
17	1550		_		1250	1	°	
18	1440				1260	- 21		
20	1300		_		1270	5		
20	1460				1200	1	0	1 4 7 1 1 1 1 2 2 2 3 3 3 4 4 4 4
22	1457				1300	4		
23	1460				1310	0		
24	1386				1320	71	Data ran	nge: =Atrium_all_2008!\$F\$2:\$F\$52
25	1270				1330	10	Series in	1: O Power
26	1386				1340	10		Column
27	1360				1350	- 21	_	Cojumis
20	1333		-		1300	5	_	
30	1450				1380			
31	1430				1390	5		
32	1270				1400	3		
33	1372				1410	6		
34	1363				1420	0]		Cancel < Back Next > Einish
35	1360				1430	3		
36	1350				1440	3		
38	1363				1450	21		
39	1395				1470	2		
40	1400				1480	õ		
41	1404				1490	1		
42	1385				1500	1		
43	1410		_		1510	0		
44	1410				1520	<u> </u>		
45	1280				1530			
40	1365				1540	11		
48	1380				1560			
49	1370				1570	31		
50	1360				1580	0		
51	1435				1590	0		
52	1435				1600	0		
53	1388					100		

13. To select the data range, click the mouse in the "Data Range" box and then with the mouse, outline the **Count** column as shown above.

14. Select "Next" and give the data range a name like "Frequency"

15. "Finish" and the chart will look like the prototype. *However the X axis is mislabeled!*

16. so, **Relabel the X-axis**.

To do this, select the chart with the mouse and select "**Source Data**". In the "**Category X-axis Labels**" box, use the mouse/cursor to select the **Bin Values** F2:F52

The chart should look like this:



17. Click Finish.

18. Now you can add the "Calculated" column, which is a Gaussian curve or Normal Distribution with the same mean, total of observations and bin range.

FCALC = TOTAL*BIN RANGE*NORMDIST(X, MEAN, SIGMA,CUMULATIVE)

In our case, this becomes

FCALC=100*10*NORMDIST(bin, 1371.8, 72.85, FALSE)

Because there are 100 observations, the bin size is 10 and the mean and standard deviation are entered as derived from the data. We use FALSE for the last parameter because we do not want to plot the cumulative distribution.

Enter a new column labeled "calcd" and enter the above formula in G2. The spreadsheet looks like this:

:	<u>Eile E</u> dit	<u>V</u> iew <u>I</u> n:	sert F <u>o</u> rmat	<u>T</u> ools <u>D</u>	ata <u>W</u> indov	w <u>H</u> elp A	Ado <u>b</u> e PDF	
1	📂 🖬 🕻	3 🖪 🖪	1 🥙 🔣 🛛	6 🖬 🛍	- 🍼 🔊 -	- (* - 🧕	Σ - <u>A</u> ↓	X 🛄 🛷
:		1 🗠 🖄 I	5010		💖 Reply wit	h ⊆hanges	End Review.	- 📮 i 🔁 🛃
	G2	•	<i>f</i> ∗ =100*10	*NORMDI	ST(E2,1371	.8,72.85,FA	ALSE)	
	A	В	C	D	E	F	G	Н
1	1490	min	1180		bins	count	calcd	
2	1570	max	1570		1100	0	0.005197	
3	1320	avg	1371.77		1110	0		
4	1570	sigma	72.85531		1120	0		
5	1400				1130	0		
6	1500				1140	0		
7	1399				1150	0		
8	1570				1160	0		
9	1410				1170	0		
10	1320				1180	1		
11	1470				1190	0		
12	1410				1200	0		
13	1320				1210	0		12 -
14	1440				1220	0		
15	1260				1230	0		
16	1440				1240	0		10

19. Copy that cell down to G52, to get a table of expected counts. It won't look much like what you observed but the calculated counts should peak at about the same place.

X N	Aicrosoft E	xcel - A	trium_	all_200	8.txt					
:	<u>E</u> ile <u>E</u> dit	⊻iew	Insert	F <u>o</u> rmat	Tools	<u>D</u> ata <u>W</u> indov	w <u>H</u> elp 4	Ado <u>b</u> e PDF		
: 🗅	💕 🖬 🛛	3 3	à 🍣		(🗅 🛍	- 🍼 🔊 -	- (2 - 1 😣	Σ·Ž↓	X. 1	B
1	1 2 2 2	1 🔁 🕅		3 3	B (2)	₩9 Reply wit	h <u>⊂</u> hanges	End Review.	🚽 🕅	
	G2	•	fx =	=100*10	*NORMD	IST(E2,1371	.8,72.86,F/	ALSE)		
	А	В		C	D	E	F	G	Н	
1	1490	min		1180		bins	count	calcd	2	
2	1570	max		1570		1100	0	0.005207		
3	1320	avg	1	371.77		1110	0	0.008606		
4	1570	sigma	72	.85531		1120	0	0.013961		
5	1400					1130	0	0.022224		
6	1500					1140	0	0.034717		
7	1399					1150	0	0.053222		
8	1570	-				1160	0	0.080067		
9	1410					1170	0	0.118206		
10	1320					1180	1	0.171254		
11	1470					1190	0	0.243479		
12	1410					1200	0	0.339704		
13	1320	1				1210	0	0.465115	1	_
14	1440					1220	0	0.62494	-	
15	1260					1230	0	0.824015		
16	1440	-				1240	0	1.06623		
17	1550					1250	1	1.353898		
18	1440					1260	2	1.687097		
19	1350					1270	5	2.063065		
20	1390					1280	2	2.47574		
21	1460	1				1290	1	2.91552	1	_
22	1457					1300	4	3.369349	-	_
23	1460					1310	0	3.821158		
24	1386					1320	7	4.252684		
25	1270					1330	10	4.644619		
26	1386					1340	10	4.978015		
27	1360					1350	2	5.235778		
28	1333					1360	8	5.404124		
29	1370					1370	5	5.473793		_

20 Now add this column to the chart. **Select the chart**, select **"Source Data"** and add the **Series** G2:G52



21. The default data representation is bar columns. If you like, you can change the "Chart Type" to X,Y:



Keep in mind that there are other ways to do all this, and that most of these instructions will work for OpenOffice, which is an open-source Excel clone.