

## MTH245 Unit 3 Module 3 Modeling Randomness

This module covers the mechanics of efficiently running multiple simulations of an experiment to estimate probabilities. So it is an extension of Unit 3 Module 2, and once we have this process down we will move on to probability distributions in Unit 3 Module 4 to analyze our results. So there will not be any theory in this module, just a “how to”.

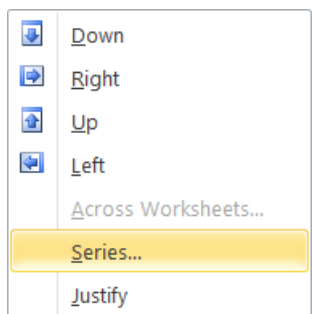
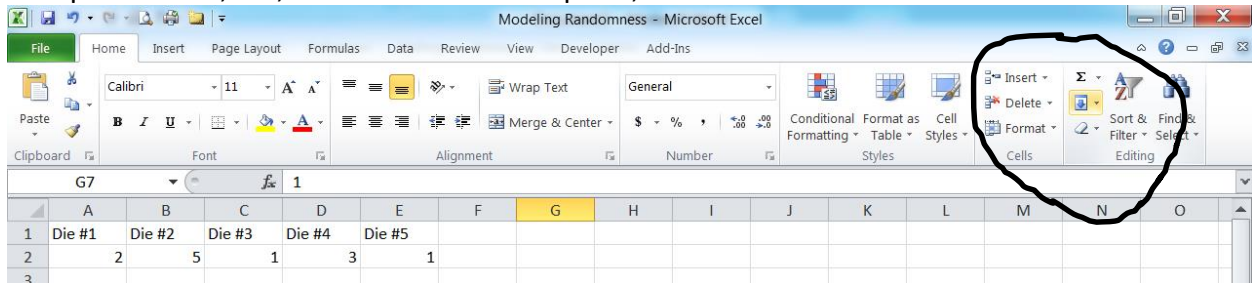
Our goal is generate a table that shows all possible outcomes of an experiment and the probability of each of those outcomes occurring. To do this we will:

- Set up an experiment
- Run the experiment a large number of times to generate data
- Collect the data
- Counting how often each outcome occurred
- Calculate the percentage of times each outcome occurred

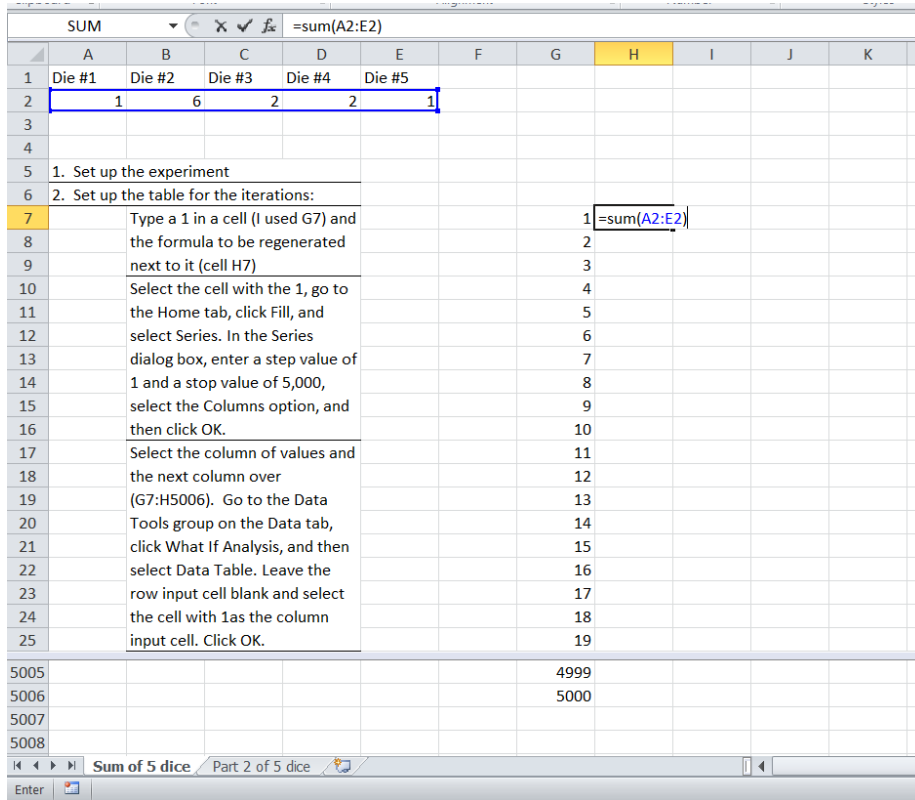
Our experiment will be to roll 5 fair dice and find the sum. Our possible outcomes would be all integers from 5 to 30...that would be 26 possible outcomes!

We know how to use `=randbetween(1,6)` to simulate a die, and we can set up 5 cells with `randbetween` and another cell to sum the values. Dragging down would work, but it would not be tidy. With 26 possible outcomes we really need a large number of iterations to get a good sample. Rather than regenerating every cell, we just want to regenerate the sum of the 5 dice.

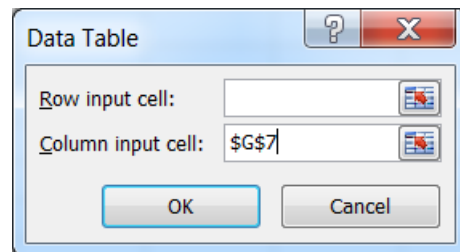
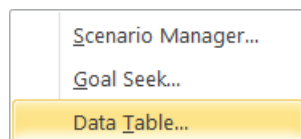
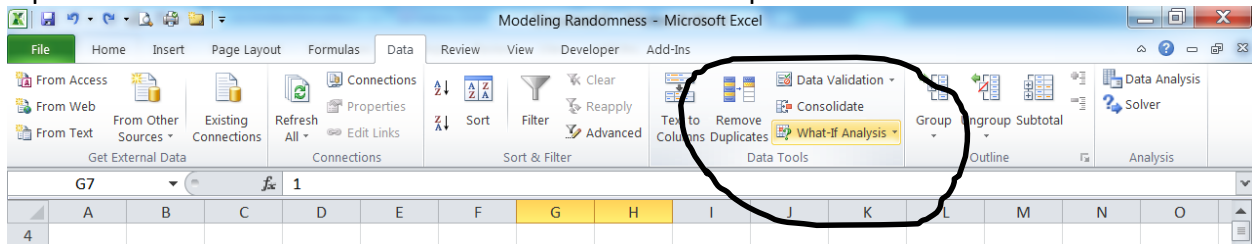
First we are going to set up a column to keep track of how many simulations we have run. A nice way to do this is to use the Fill Series tool in Excel. Type a 1 in a cell and select the cell. Go to the Home tab, click Fill, and select Series. In the Series dialog box, enter a step value of 1 and a stop value of 5,000, select the Columns option, and then click OK.



This will give us a column of values from 1 to 5000. Next to the cell with a 1, we can type in the formula we wish to recalculate.

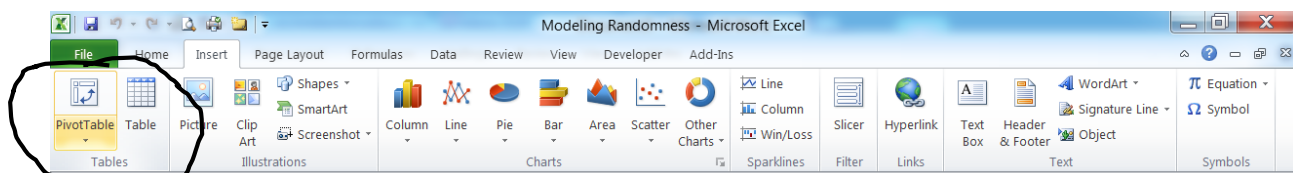


We do not want to just copy down the formula—try it! We will just get the same sum 5000 times. Instead we are going to create a data table which will recalculate the sum 5000 times. Select the counting column of values and the next column over (G7:H5006). Go to the Data Tools group on the Data tab, click What If Analysis, and then select Data Table. Leave the row input cell blank and select the cell with 1 as the column input cell. Click OK.



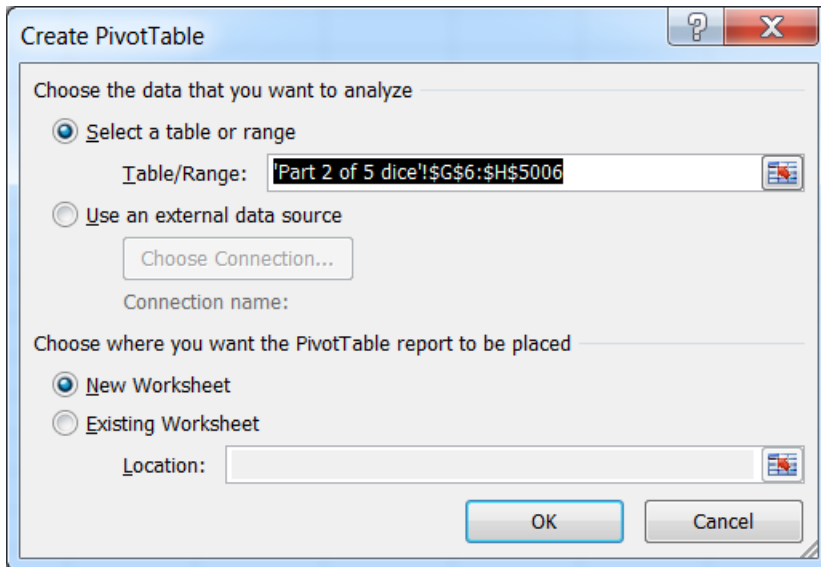
	A	B	C	D	E	F	G	H	I	J	
4											
5		1. Set up the experiment									
6		2. Set up the table for the iterations:									
7		Type a 1 in a cell (I used G7) and					1	20			
8		the formula to be regenerated					2	17			
9		next to it (cell H7)					3	21			
10		Select the cell with the 1, go to					4	23			
11		the Home tab, click Fill, and					5	14			
12		select Series. In the Series					6	18			
13		dialog box, enter a step value of					7	11			
14		1 and a stop value of 5,000,					8	15			
15		select the Columns option, and					9	22			
16		then click OK.					10	16			
17		Select the column of values and					11	14			
18		the next column over					12	17			
19		(G7:H5006). Go to the Data					13	14			
20		Tools group on the Data tab,					14	14			
21		click What If Analysis, and then					15	16			
22		select Data Table. Leave the					16	19			
23		row input cell blank and select					17	16			
24		the cell with 1 as the column					18	19			
25		input cell. Click OK.					19	20			
26						20	17				
27						21	19				
28						22	9				
5005							4999	18			
5006							5000	13			
5007											
5008											

Now we have 5000 sums of the 5 dice. More data than we would want to tally up by hand, so we will get Excel to do it. This will involve a Pivot Table. Click anywhere on the table and go to the Insert tab. The PivotTable tool is on the left end of the ribbon.

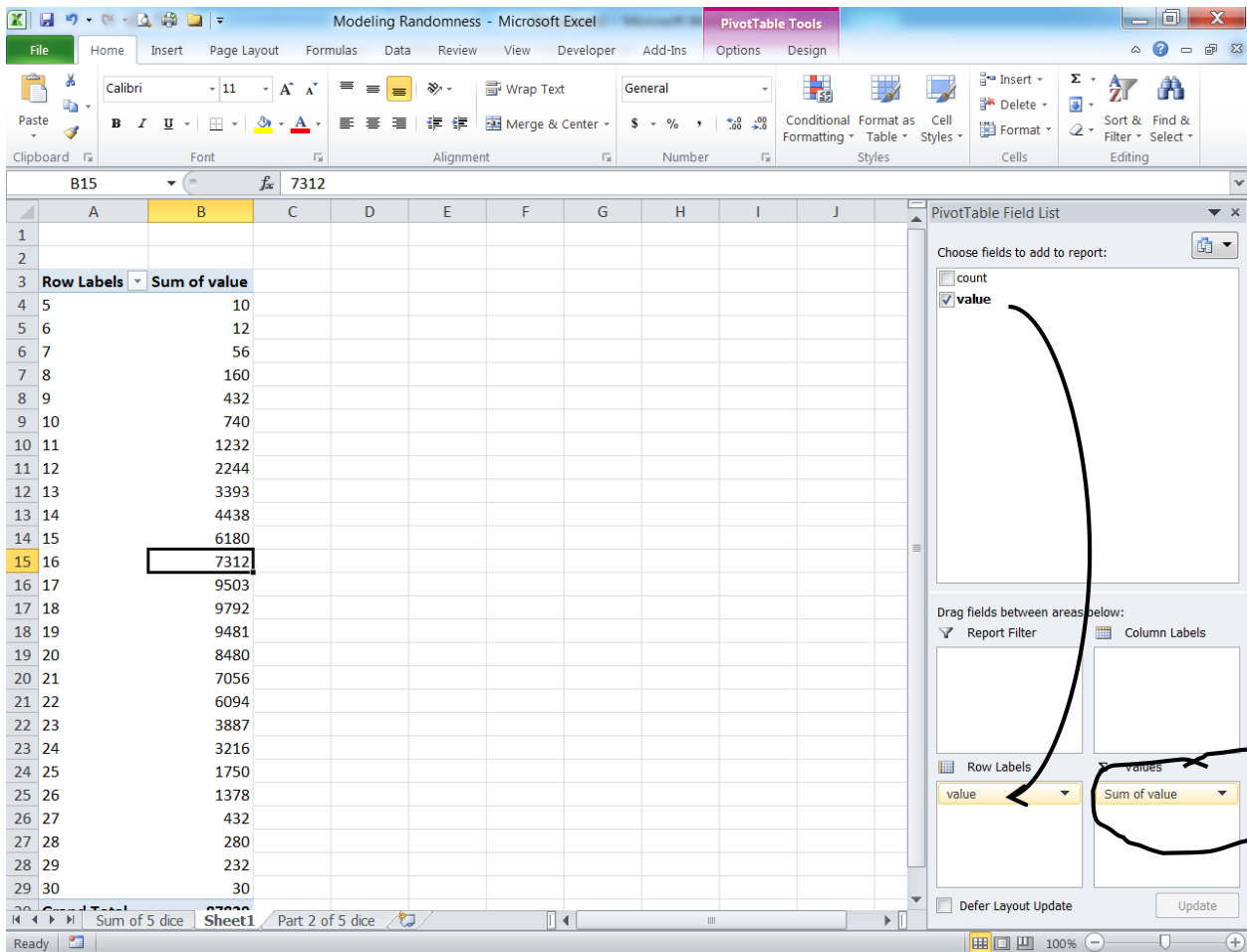


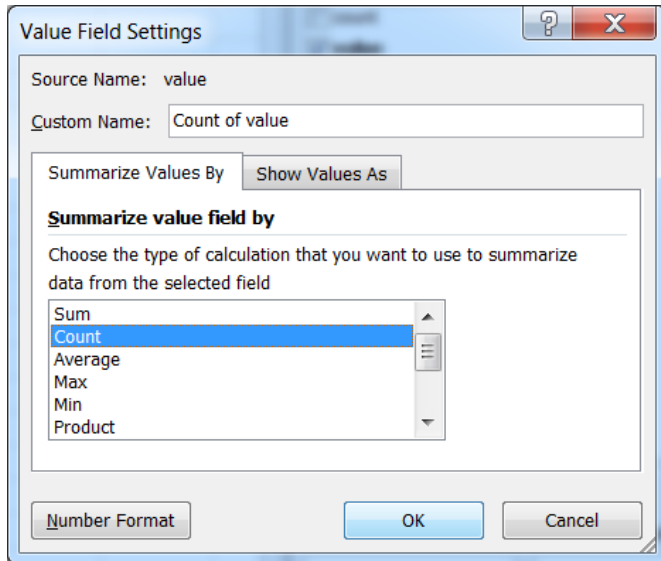
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	dice 1	dice 2	dice 3	dice 4	dice 5										
2	6	3	6	5	6										
3															
4	Once we have generated our data, we need a summary														
5	Click anywhere in the table and go to the Insert tab														
6	Click on Pivot Table and the data should be														
7	automatically selected. Leave as the default that the														
8	PivotTable will be placed on a new sheet.														
9	On the new sheet, click on the value field and drag it														
10	down to row labels AND change Sum of Values to														
11	Count of Values.														
12							count	value							
							1	26							
							2	18							
							3	18							
							4	19							
							5	19							
							6	16							

Click on Pivot Table and the data should be automatically selected. Leave as the default that the PivotTable will be placed on a new sheet.



On the new sheet, click on the value field and drag it down to row labels AND change Sum of Values to Count of Values by clicking on the drop down menu and selecting Value Field Settings.

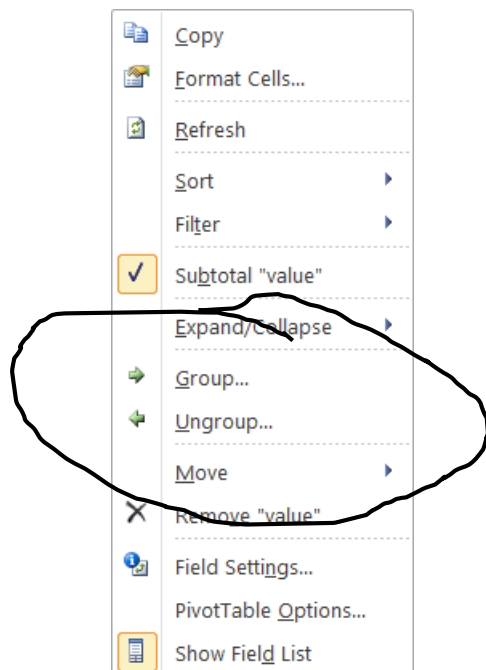




When you bring up the Value Field Settings select "Count".

We now have how many times each outcome occurred in our 5000 experiments:

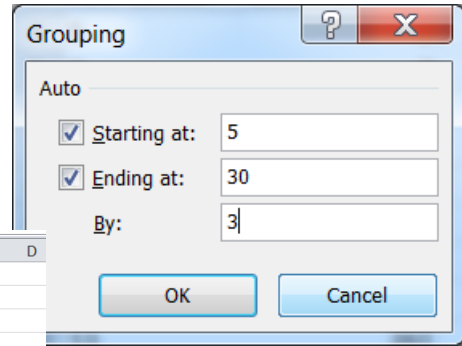
Row Labels	Count of value
2	2
3	2
4	8
5	20
6	48
7	74
8	112
9	187
10	261
11	317
12	412
13	457
14	559
15	544
16	499
17	424
18	336
19	277
20	169
21	134
22	70
23	53
24	16
25	10
26	8
27	1
28	
29	
30	
<b>Grand Total</b>	<b>5000</b>



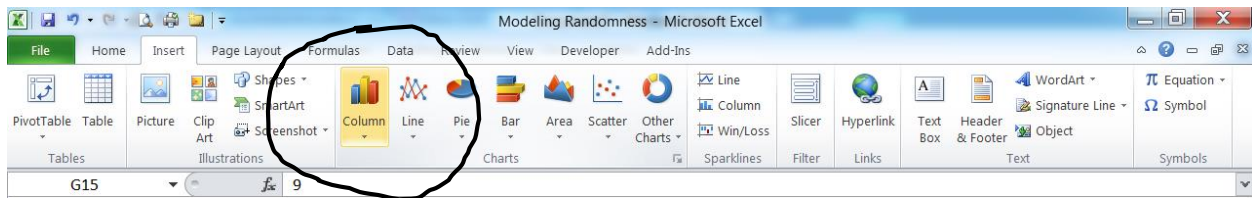
To change the way the data is grouped, right click on a row label and select Group.

The starting and ending values can be adjusted. For example, if we had not actually observe a sum of 5 in our data set, we might want to include it in our table (and histogram) as it was a possible value. The “By” category determines the “width” of each group. With a starting value of 5 and a width of 3 our first group would include values of 5, 6, and 7. The next group would include values of 8, 9 and 10.

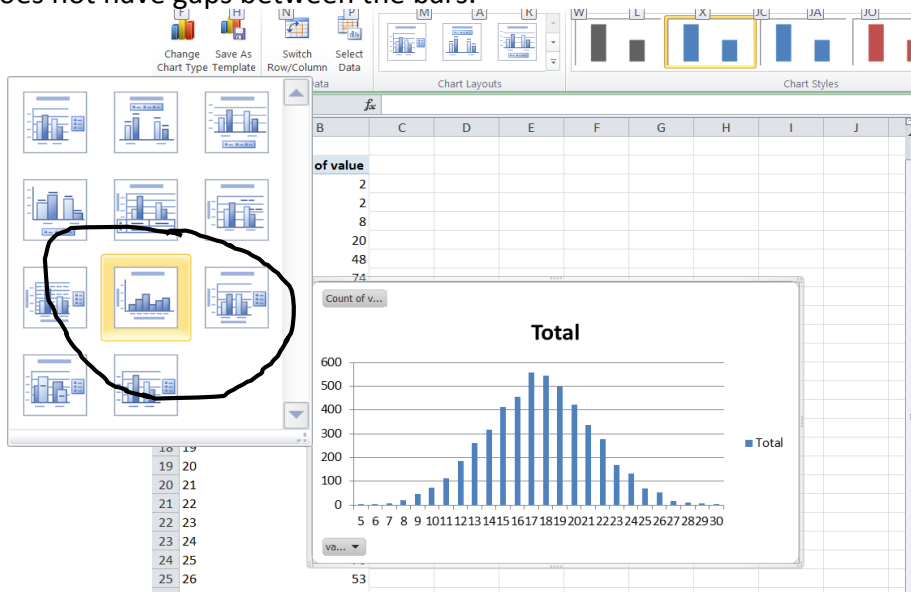
	A	B	C	D
2				
3	Row Labels	Count of value		
4	5-7	12		
5	8-10	142		
6	11-13	560		
7	14-16	1186		
8	17-19	1602		
9	20-22	1037		
10	23-25	373		
11	26-28	79		
12	29-31	9		
13	Grand Total	5000		
14				

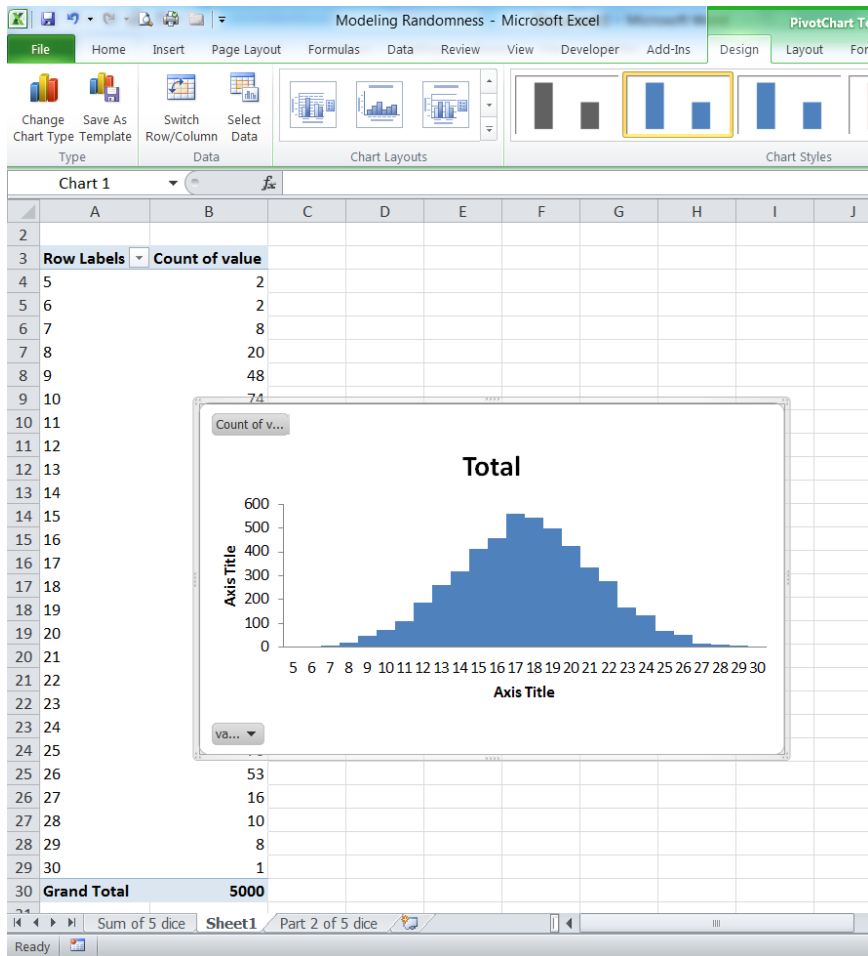


The next step will be to generate a histogram—a visual summary of our table. To create a histogram, click on a cell in the PivotTable, go to the Insert tab and click on Column.

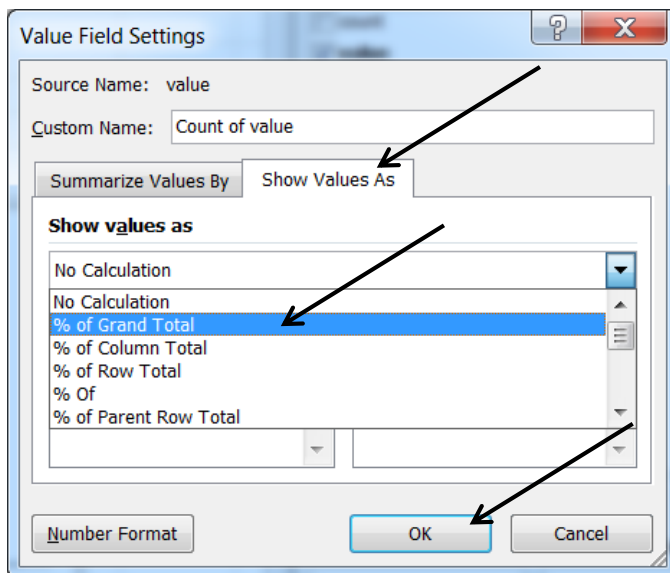


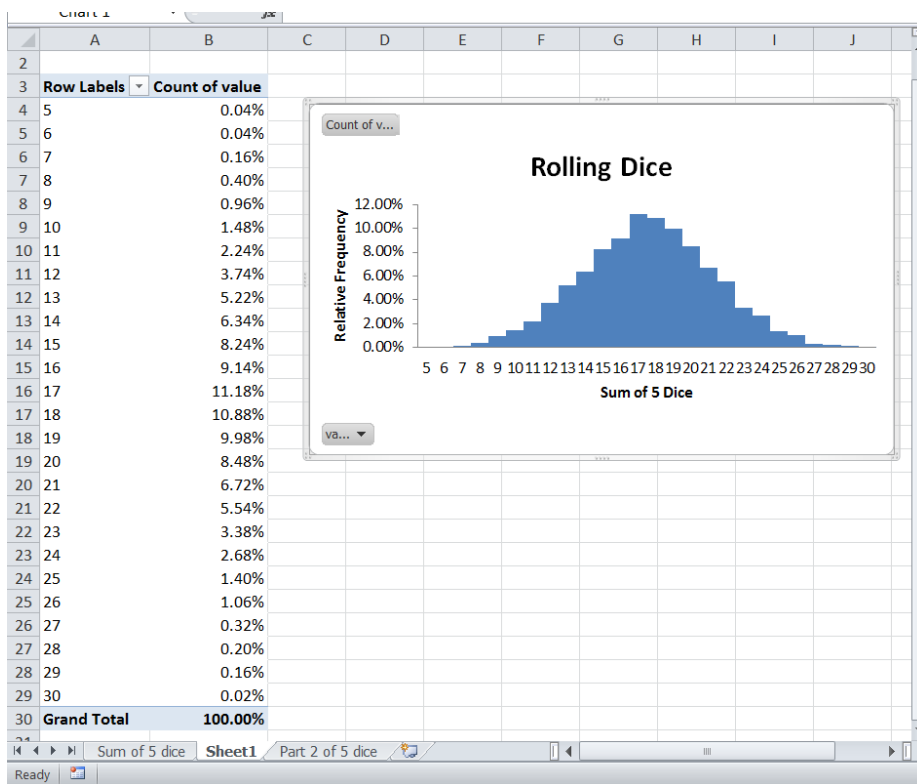
The default is not a nice looking graph, so click on Chart Layouts and select the histogram that does not have gaps between the bars.





What we have created is a Frequency Chart and Histogram. Often we will want a Relative Frequency table and histogram. Click on the drop down menu next to count of values and select values field settings. Click on Show Values As, select % of Grand Total and click OK.





The table displays the possible outcomes, and the percentage of times each outcome was observed in our 5000 trials. These percentages will vary slightly but, with our large sample size, will be very close to the theoretical probabilities of observing each outcome. What we have created is an empirical probability distribution table. The histogram can be labeled and gives a nice visual representation of the table, it is easy to see which outcomes are more likely to be observed and which are rarer events.