Lab 7 – Statistics I

This lab will cover how to do statistical calculations in excel using formulas. (Note that your version of excel may have additional formulas to calculate statistics, but these formulas work across all excel versions.)

Lab 7 Quick View

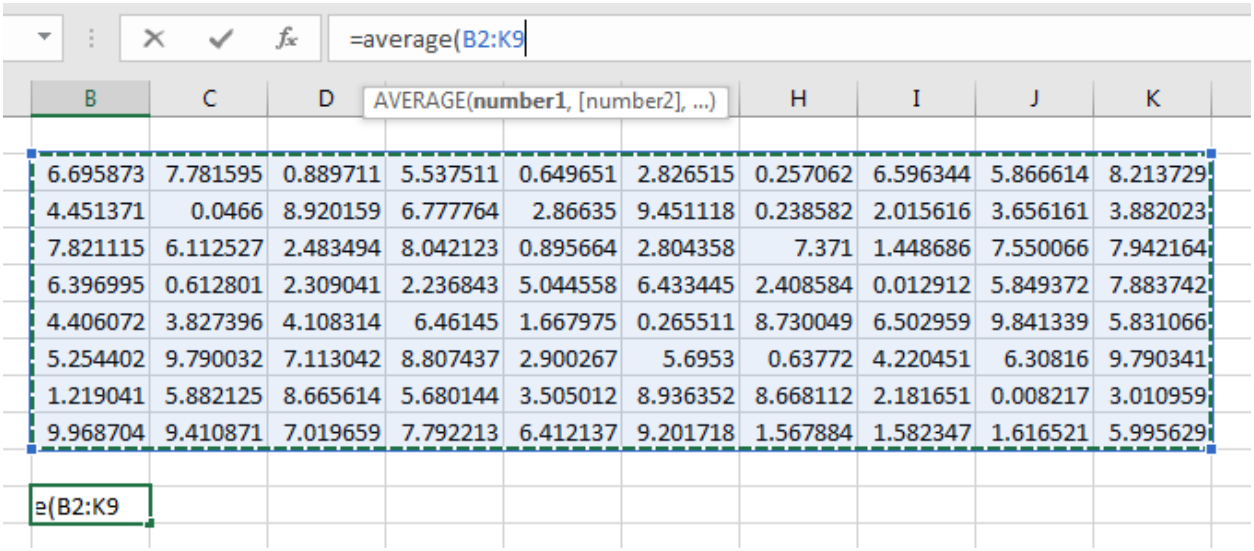
- In all of these formulas, (X,Y) will make the formula apply to all cells in a box between cell X in the top left to cell Y in the bottom right. The formula can be filled by either typing in the X and Y cells, or highlighting the cells when you are typing in the formula.
- The number of observations in your data set can easily be calculated using the formula **=count(X,Y)**.
- The arithmetic mean, or average, can be calculated as **=average(X,Y).**
- Variance is calculated as **=Var(X,Y)** . (Or **=Var.s(X,Y)**)
- Standard deviation is calculated as **=Stdev(X:Y).** (Or **=Stdev.s(X:Y)**)
- The minimum and maximum value in a set of data can be found using the formulas **=Min(X,Y)** and **=Max(X,Y)**.
- The median can be calculated as **=Median(X,Y).**
- The K-th percentile is calculated as **=percentile(X:Y,k)**, where K is entered as a decimal (the 75-th percentile will have k=0.75).

- To create a Histogram or Distribution Graph of data:
  - Highlight the data you would like to plot (including the title) as a distribution and under the "Insert" menu, choose "Pivot Chart".
  - Drag the name of the data set to both "Rows" and "∑ Values". Click on the arrow beside the data set under "∑ Values" and select "Value Field Settings" to change "Sum" to "Count"
  - Back in your spreadsheet, highlight all the entries under "Row Labels", then under "PivotTable Tools", select "Analyze", then "Group Selection"
  - Choose a starting value below the minimum and an ending value above the maximum (another reason the =Max and =Min commands are important) and an appropriate range for each category.
  - Lastly, still under "PivotTable Tools" and "Analyze", select "Pivot Chart" and select "Column" for your selected chart to show a histogram.
  - If you have ranges that have NO data points, highlight all row labels, right-click and choose "Field Settings", then make sure "Show items with no data" has a checkmark.

Written by Lorne Priemaza

A) Selecting cells

In all of these formulas, (X,Y) will make the formula apply to all cells in a box between cell X in the top left to cell Y in the bottom right. The formula can be filled by either typing in the X and Y cells, or highlighting the cells when you are typing in the formula. For example, =average(B2, K9) will take the average of all cells in the box between cell B2 and cell K9 as seen below:



B) Number of observations

The number of observations in your data set can easily be calculated using the formula =count(X,Y).

C) Average/Mean

The arithmetic mean, or average, can be calculated as =average(X,Y)

D) Variance and Standard Deviation

Variance and standard deviation are a measure of dispersion of data; of how spread out the data is. Variance is useful in many calculations and standard deviation can easily be shown in a distribution such as a bell curve. Variance is calculated as =Var(X,Y) and standard deviation is calculated as =Stdev(X:Y).

Note: Some versions of Excel do not accept the compatibility functions **Var** and **Stdev**. In these cases, you can use **=Var.s(X.Y)** and **=Stdev.s(X.Y)**. These commands use identical background formulas. The **.s** indicates that we are using sample data, rather than population data.

Written by Lorne Priemaza

E) Minimum and maximum

The minimum and maximum value in a set of data can be found using the formulas =Min(X,Y) and =Max(X,Y). If you have a large data set, these are useful formulas to ensure there are no errors in your data. (For example, if your data is grades in percentage, and the maximum value is 378%, there is an error in your data.)

F) Median

The middle of a data set is referred to as the MEDIAN. By definition, half (50%) of the data set will be below the median, and half (50%) of the data set will be above the median. The median can be calculated as =Median(X,Y) and is also referred to as the 50th percentile.

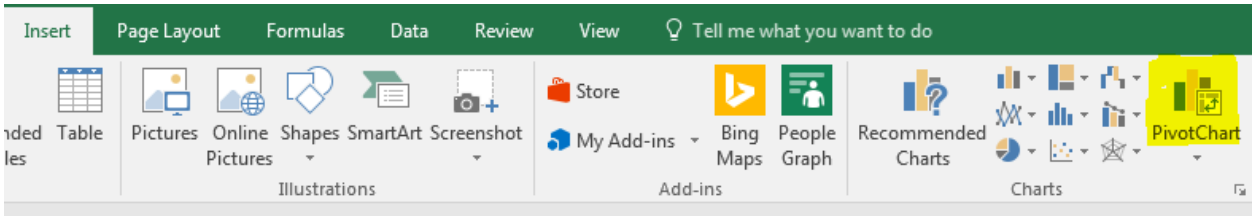
G) Percentiles

The X-th is the number that is higher than exactly X% of the data. For example, often scholarships want to know if a student is in the top 10% of the class. To figure this out, the 90% percentile has to be calculated, where 90% of the class will be below that value and only 10% of the class will be above that percentile. The K-th percentile is calculated as =percentile(X:Y,k), where K is entered as a decimal (the 75-th percentile will have k=0.75).

H) Histograms (distribution graphs) of data sets

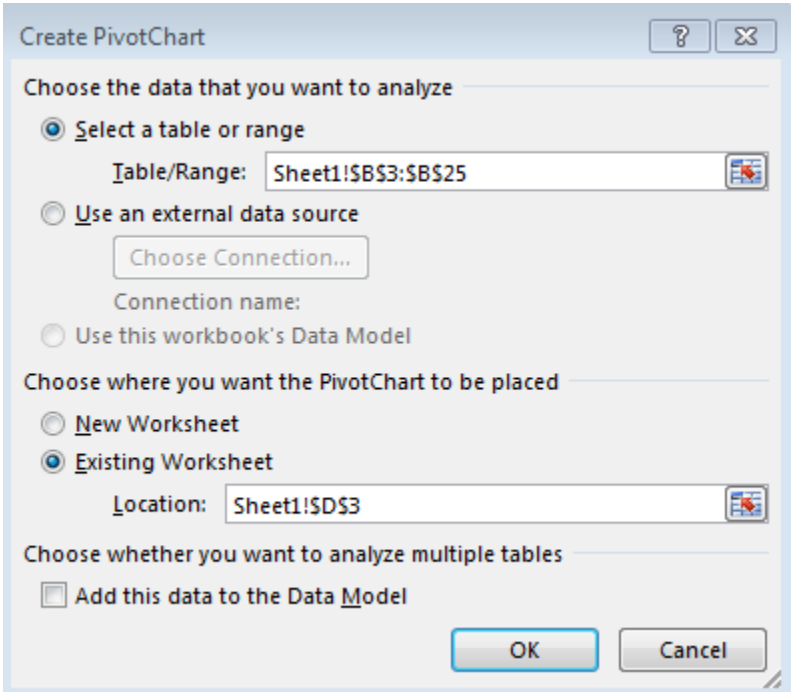
One common method of analyzing data is by looking at its DISTRIBUTION (for example, is our data normally or close to normally distributed?) We will learn one way to examine distributions using histograms and pivot charts.

- Highlight the data you would like to plot (including the title) as a distribution and under the "Insert" menu, choose "Pivot Chart".

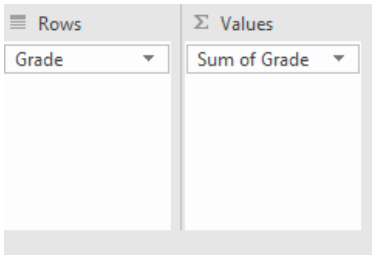


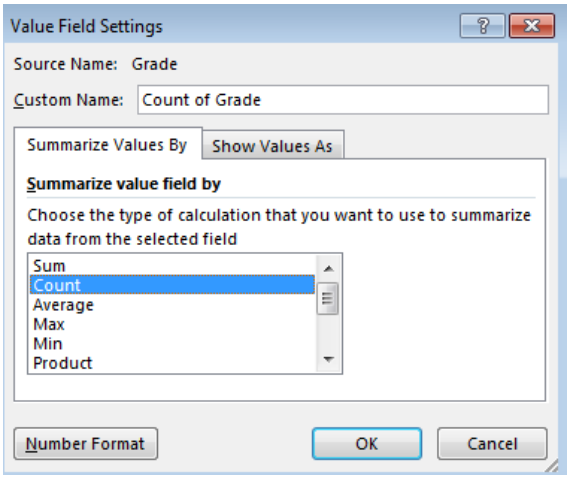
Written by Lorne Priemaza

- Choose to create the chart in a new worksheet (tab) or on the current page and click “OK”

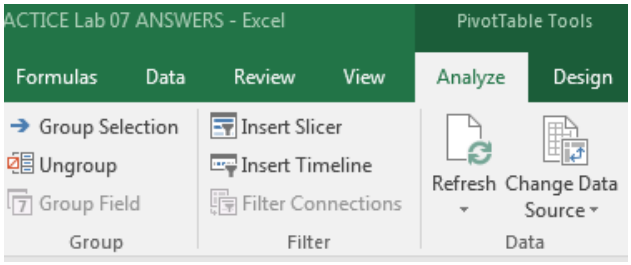


- On the right, drag the name of the data set to both “Rows” and “∑ Values”. Click on the arrow beside the data set under “∑ Values” and select “Value Field Settings” to change “Sum” to “Count”

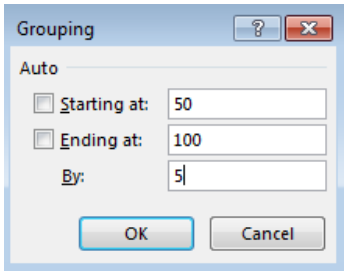




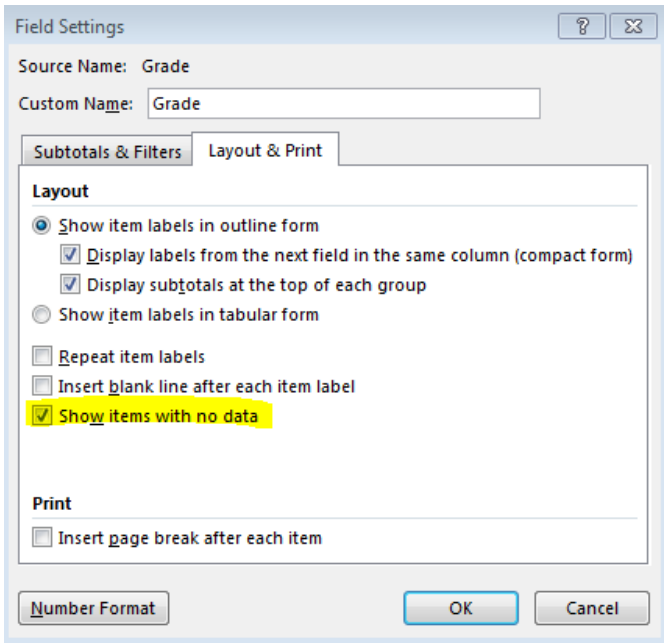
- Back in your spreadsheet, highlight all the entries under “Row Labels”, then under “PivotTable Tools”, select “Analyze”, then “Group Selection”



- Choose a starting value below the minimum and an ending value above the maximum (another reason the =Max and =Min commands are important) and an appropriate range for each category. (If you are examining grades, a 5% or 10% window is common).

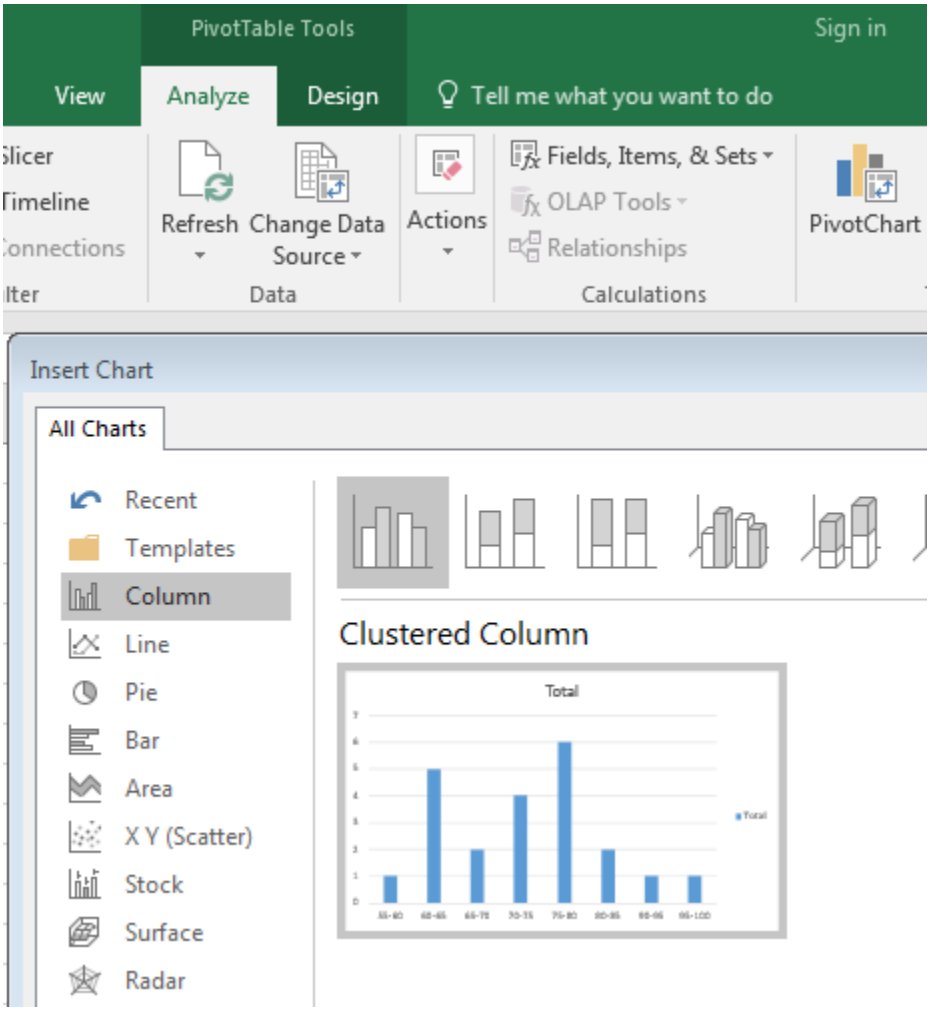


- You will see that your spreadsheet will now count how many data points are in each range.
- If you have ranges that have NO data points, highlight all row labels, right-click and choose “Field Settings”, then make sure “Show items with no data” has a checkmark.



Written by Lorne Priemaza

- Lastly, still under “PivotTable Tools” and “Analyze”, select “Pivot Chart” and select “Column” for your selected chart to show a histogram.



- Edit the titles and colours if your chart as seen in a previous lab.

Written by Lorne Priemaza

Econ 299 Practice Lab 7:

A) Copy or download the following grades from a class of Econ 298 after the midterm:

	A	B
1	**Econ 298 Student Grades**	
2		
3	**Student**	**Grade**
4	Will Smith	77.63
5	Liu Bei	64.02
6	Guan Yu	70.38
7	Zhang Fei	78.45
8	Dave Duncan	77.59
9	Naomi Novak	55.30
10	Caitlin Stark	80.60
11	Megan Grimes	82.30
12	Iron Man	71.97
13	Black Widow	79.90
14	Get Daved	74.95
15	Shen Yyr	60.39
16	Quill Eighteen	77.09
17	Ddear Jake	60.68
18	Terra Mystica	68.84
19	Small World	69.91
20	Catan Settler	70.51
21	Zee Garcia	64.43
22	Sam Healey	75.84
23	Tom Vassel	64.34
24	Doctor Who	94.80
25	Doctor Donna	95.70
26		

B) Use a formula to calculate the number of students in this class. Create subtitle for this and all other statistics you calculate.

C) Calculate the average, variance, standard deviation, minimum and maximum of this class.

D) Assume that the top 50% of this class didn't have to write a paper. Find the grade that cuts the class in half (the median or 50^{th} percentile).

E) Due to budget cuts, the bottom 20% of the class has to drop the class. Find the grade that cuts off the bottom 20% of the class (the 20^{th} percentile).

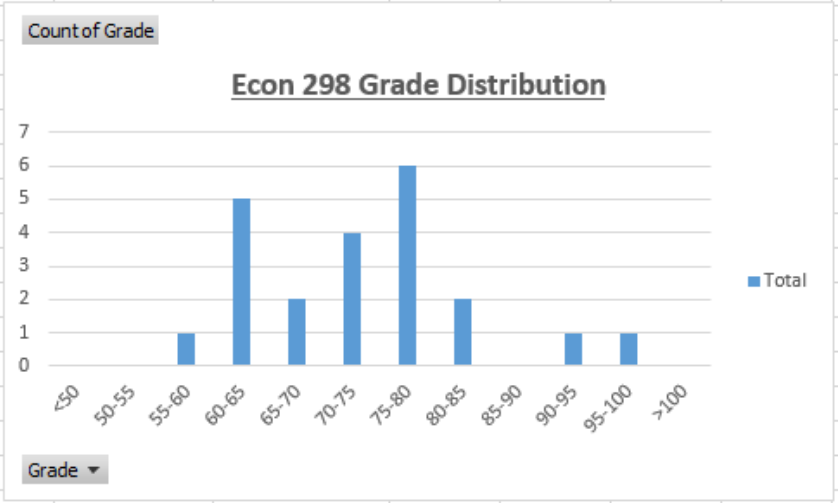
F) A scholarship is available for the top 10%. Find the grade that cuts off the top 10% of the class (the 90^{th} percentile.)

Written by Lorne Priemaza

G) Create a histogram of the class' distribution starting at 50% and with 5% intervals. (Make sure to show empty ranges.) Label the distribution appropriately.

Lab 7 Practice Answers:

	A	B	C	D	E	F	G	H	I
1	**Econ 298 Student Grades**								
2									
3	**Student**	**Grade**		**Row Labels**	**Count of Grade**				
4	Will Smith	77.63		<50					
5	Liu Bei	64.02		50-55					
6	Guan Yu	70.38		55-60	1				
7	Zhang Fei	78.45		60-65	5				
8	Dave Duncan	77.59		65-70	2				
9	Naomi Novak	55.30		70-75	4				
10	Caitlin Stark	80.60		75-80	6				
11	Megan Grimes	82.30		80-85	2				
12	Iron Man	71.97		85-90					
13	Black Widow	79.90		90-95	1				
14	Get Daved	74.95		95-100	1				
15	Shen Yyr	60.39		>100					
16	Quill Eighteen	77.09		**Grand Total**	**22**				
17	Ddear Jake	60.68							
18	Terra Mystica	68.84							
19	Small World	69.91							
20	Catan Settler	70.51							
21	Zee Garcia	64.43							
22	Sam Healey	75.84							
23	Tom Vassel	64.34							
24	Doctor Who	94.80							
25	Doctor Donna	95.70							
26									
27	# of Students	22.00							
28	Average	73.44							
29	Variance	103.49							
30	Standard Dev.	10.17							
31	Min	55.30							
32	Max	95.70							
33	Median	73.46							
34	20th Percentile	64.36							
35	90th Percentile	82.13							



Written by Lorne Priemaza