

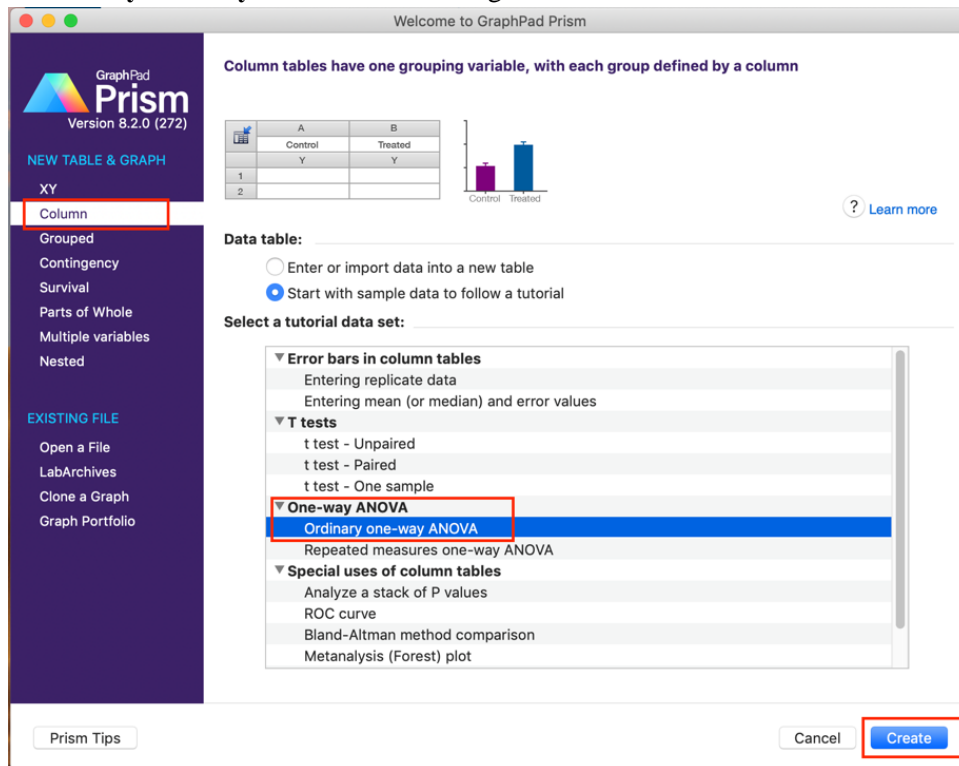
Descriptive Statistics

Qinlu (Claire) Wang
BCBB/OCICB/NIAID

This descriptive statistics lab explains how to analyze columns of numbers to compute descriptive statistics, compare the mean or median to a hypothetical value, and test for normality.

1. Descriptive Statistics

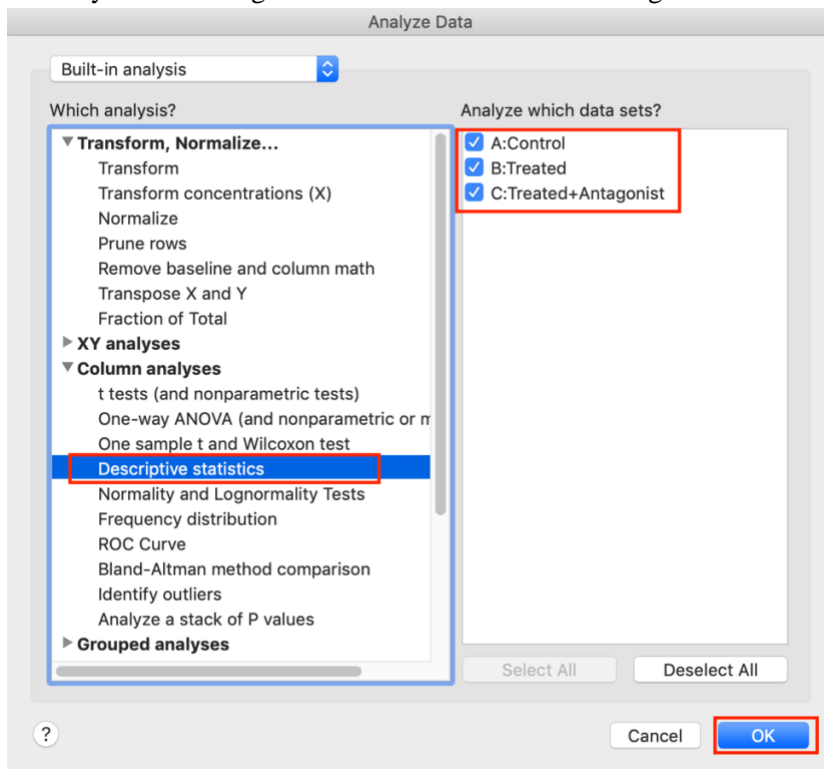
- 1) Descriptive statistics could be chosen on XY, column and Grouped data tables. In this case, let's take a column data table as an example. Open Prism. Select “Column” on the left then select “Ordinary one-way ANOVA” on the right. Click “Create”.



- 2) Once the data is opened. Click “Analyze”.

	Group A	Group B	Group C
	Control	Treated	Treated+Antagonist
	Y	Y	Y
1	54	87	45
2	23	98	39
3	45	64	51
4	54	77	49
5	45	89	50
6	47		55
7			
8			
9			
10			

It pops up an Analyze Data dialog. Select **“Descriptive Statistics”** on the left then **all datasets** to be analyzed on the right. Click **“OK”** on the bottom right.



Then choose whatever analysis you want. In this case, we keep the default selection: “Mean, SD, SEM” and “Minimum and Maximum, range” then **click “OK”**.

Parameters: Descriptive Statistics

Basics

☒ Mean, SD, SEM ☒ Minimum and maximum, range

☐ Column sum ☐ Quartiles (Median, 25th and 75th percentile)

Advanced

☐ Coefficient of variation ☐ Geometric mean

☐ Skewness and kurtosis ☐ Harmonic mean

☐ Percentile 90 ☐ Quadratic mean

Confidence intervals

☐ CI of the mean ☐ CI of harmonic mean

☐ CI of geometric mean ☐ CI of quadratic mean

☐ CI of median

Confidence level 95%

Subcolumns

☒ Average the replicates in each row, and then perform the calculation for each column

☐ Perform the calculation for each subcolumn separately

☐ Treat all the values in all subcolumns as one set of data

Output

Show this many significant digits: 4

☐ Make these choices the default for future analyses.

Cancel OK

The analysis checklist has been listed below:

Value	Meaning
Minimum	The smallest value.
25th Percentile	25% of values are lower than this.
Median	Half the values are lower; half are higher.
75th Percentile	75% of values are lower than this.
Maximum	The largest value.
Mean	The average.
Standard Deviation	Quantifies variability or scatter.
Standard Error of Mean	Quantifies how precisely the mean is known.
95% Confidence Interval	Given some assumptions, there is a 95% chance that this range includes the true overall mean.
Coefficient of Variation	The standard deviation divided by the mean.
Geometric Mean	Compute the logarithm of all values, compute the mean of the logarithms, and then take the antilog of that mean. It is a better measure of central tendency when data follow a lognormal distribution (long tail).
Harmonic Mean	Compute the reciprocal of all values, compute the mean of the reciprocals, and then take the reciprocal of that mean.
Quadratic Mean	Compute the square of all values, compute the mean of the squares, and then take the square root of that mean.
Skewness	Quantifies how symmetrical the distribution is. A distribution that is symmetrical has a skewness of 0.
Kurtosis	Quantifies whether the tails of the data distribution matches the Gaussian distribution. A Gaussian distribution has a kurtosis of 0.

Then Prism generates the descriptive statistics results as well as the graphs:

Q Search		Descriptive statistics			A	B	C
					Control	Treated	Treated+Antagonist
					Y	Y	Y
▼ Data Tables	»	1	Number of values		6	5	6
▼ Info	»	2					
▼ Results	»	3	Minimum		23.00	64.00	39.00
▼ Graphs	»	4	Maximum		54.00	98.00	55.00
▼ Layouts	»	5	Range		31.00	34.00	16.00
		6					
		7	Mean		44.67	83.00	48.17
		8	Std. Deviation		11.40	12.98	5.529
		9	Std. Error of Mean		4.652	5.805	2.257
		10					
		11					

Depending on the types of variables, authors should present the appropriate descriptive statistics. For numerical variables, if the variable is normally distributed, the mean and standard deviation (SD) are presented. In the text, this is reported as mean (SD = value), for example, “the mean age was 46.5 (SD = 3.0).” Whenever the variable is not normally distributed, the median and inter-quartile range (IQR) are reported instead.

For the categorical variable, count (n) and percentage (%) are presented. In addition, authors must report the group size and total sample size, written as $n = \text{size}$ in the table headers and the table description, respectively. The use of a capital N in place of n must be avoided as it refers to population size instead of sample size.

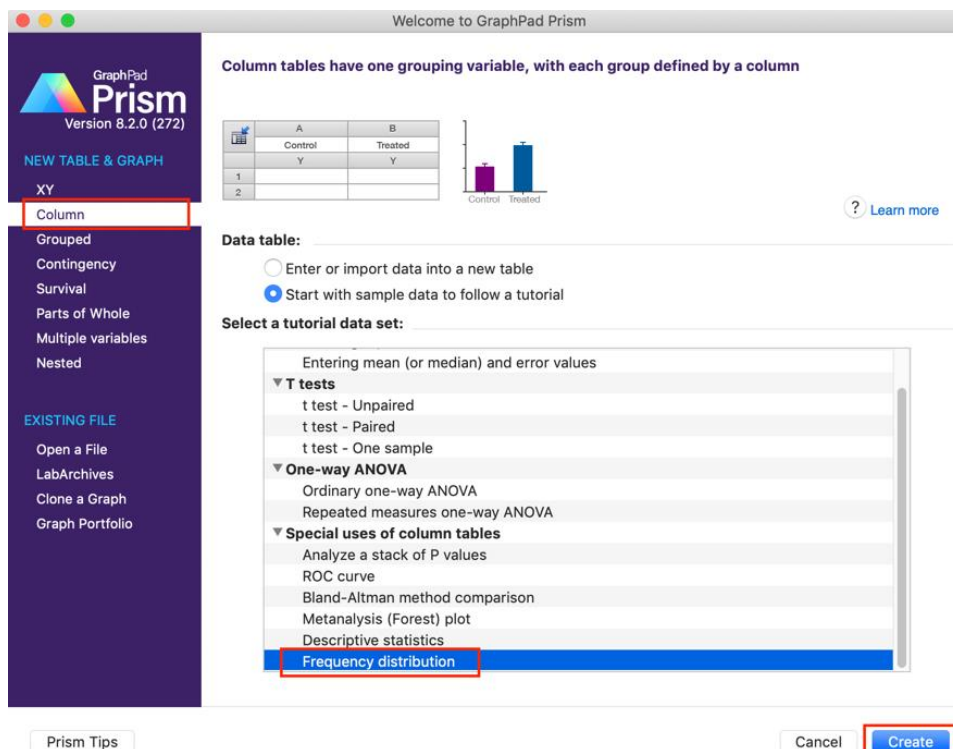
More details of how to report statistical results in medical journals could be found at:

[Reporting Statistical Results in Medical Journals](#)

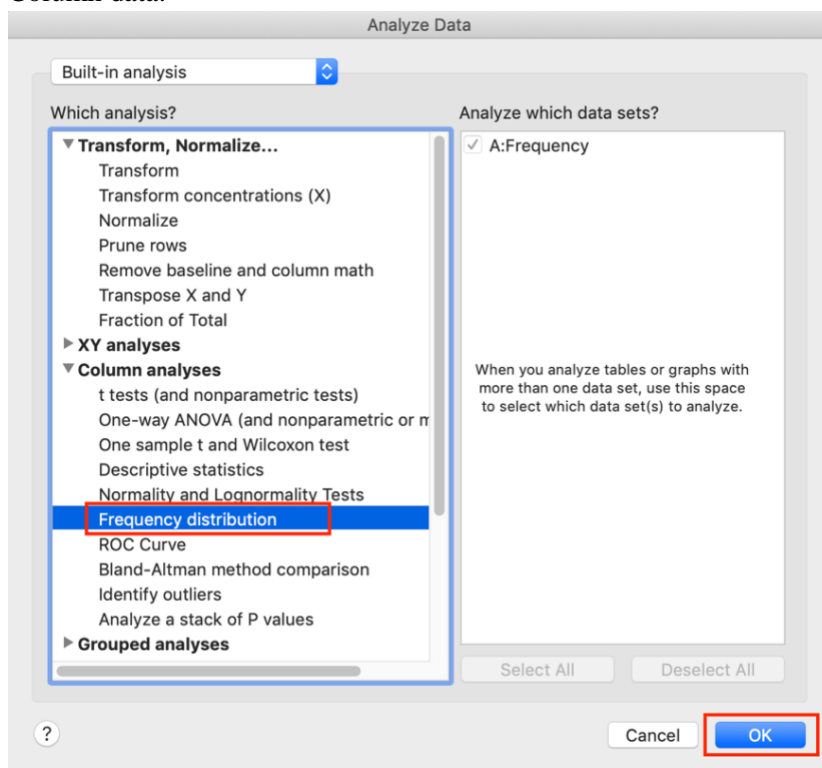
2. Frequency Distributions

This section explains how to create and plot a frequency distribution from a column of numbers.

- 1) Open Prism 8. Select “Column” table on the left then “Frequency distribution” on the right. Then click “Create”.



- 2) Click “Analyze” and then choose “Frequency distribution” from the list of analyses for Column data.



Choose analysis options in the pop-up window. In this case, we use the default settings. More details about the other analysis options could be found at [How to: Frequency distribution](#)

Parameters: Frequency Distribution

Create

☒ Frequency distribution
☐ Cumulative frequency distribution

Tabulate

☒ Number of values
☐ Relative frequency (fractions)
☐ Relative frequency (percentages)

Bin range

Center of first bin: ☒ Auto
☐ 0

Center of last bin: ☒ Auto
☐ 90

Bin width

☒ Choose automatically
☐ Bin width 10
☐ No bins. Tabulate exact cumulative frequency

Replicates

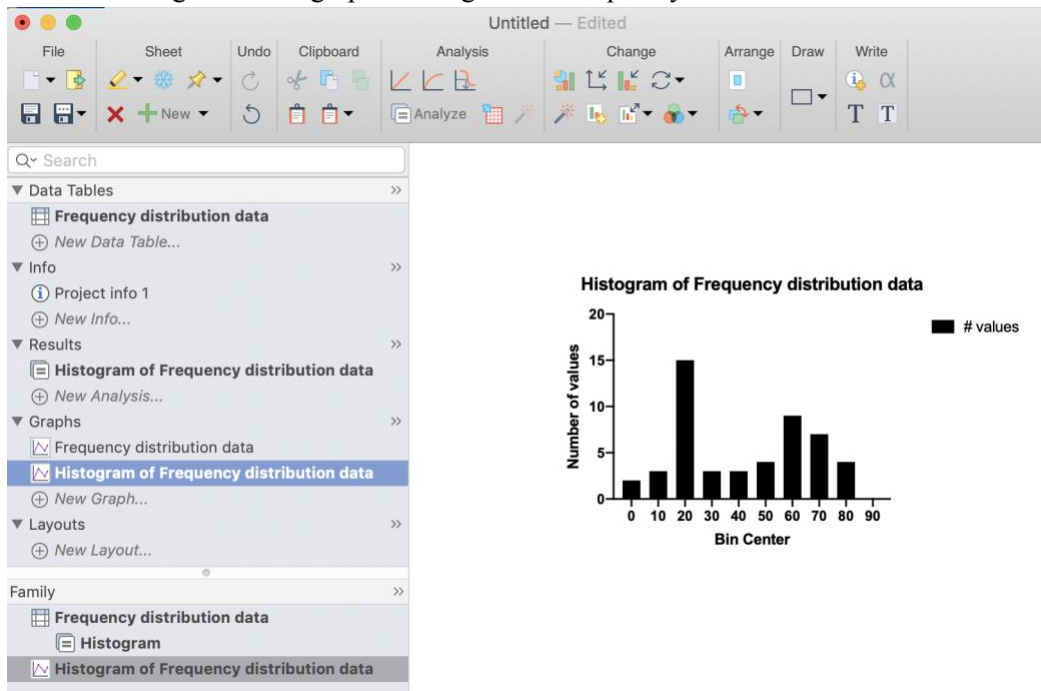
☐ Bin each replicate
☐ Bin only means

New graph

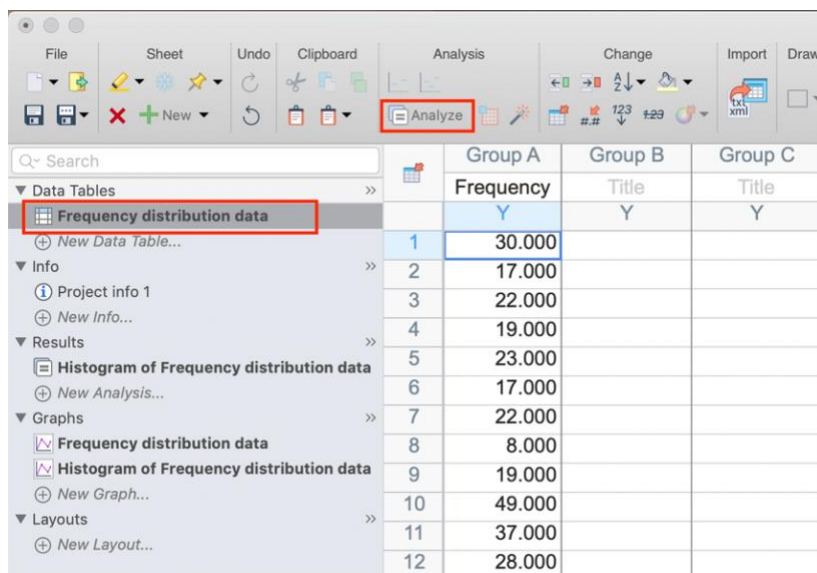
☒ Create a graph of the results
Graph type: Bar graph

? Cancel OK

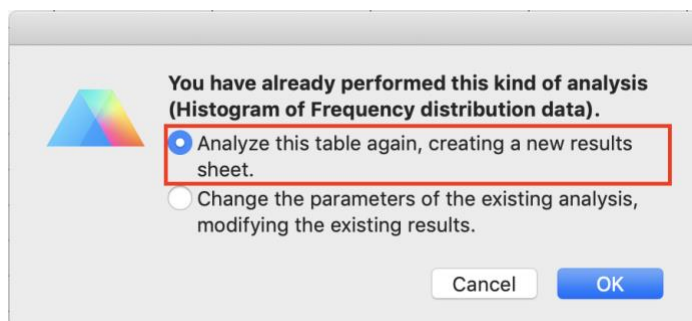
Then Prism 8 generates a graph “Histogram of Frequency distribution data”



- 3) Similarly, we could create another graph for cumulative frequency distribution. **Go back to the data table “Frequency distribution data”.** Then click “Analyze”.



Then Prism 8 would ask if you would like to create a new results sheet or change the existing analysis. In this case, we **choose “Analyze this table again, creating a new results sheet”**. Then **click “OK”**.



In the pops-up window “Parameters: Frequency Distribution”, this time we **choose “Cumulative frequency distribution”** then **click “OK”**.

Parameters: Frequency Distribution

Create

☐ Frequency distribution

☒ Cumulative frequency distribution

Tabulate

☒ Number of values

☐ Relative frequency (fractions)

☐ Relative frequency (percentages)

Bin range

Center of first bin: ☒ Auto ☐ 0

Center of last bin: ☒ Auto ☐ 90

Bin width

☒ Choose automatically

☐ Bin width 10

☐ No bins. Tabulate exact cumulative frequency

Replicates

☐ Bin each replicate

☐ Bin only means

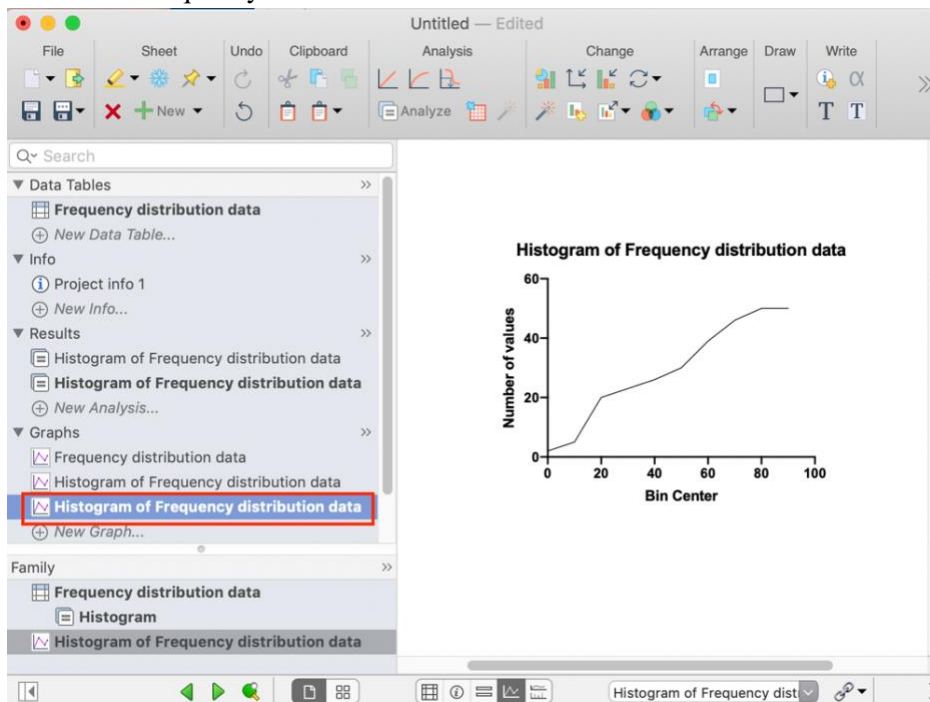
New graph

☒ Create a graph of the results

Graph type XY graph, Points

Cancel OK

Then Prism generates another graph “Histogram of Frequency distribution data” for the cumulative frequency distribution.



Conclusion

If you have any question about descriptive statistics, please feel free to contact me (qinlu.wang@nih.gov) or our BCBB (bioinformatics@niaid.nih.gov)

Reference

[Statistics with Prism 8](#)