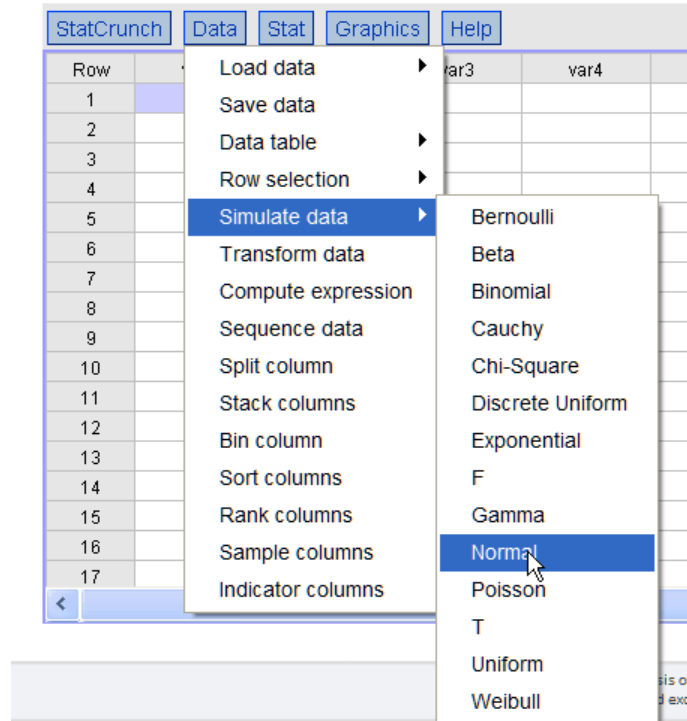


WHAT IS A CONFIDENCE INTERVAL

click on ...

DATA / SIMULATE DATA / NORMAL

this will allow you to create some data by sampling from a normal distribution with a mean and standard deviation that you specify



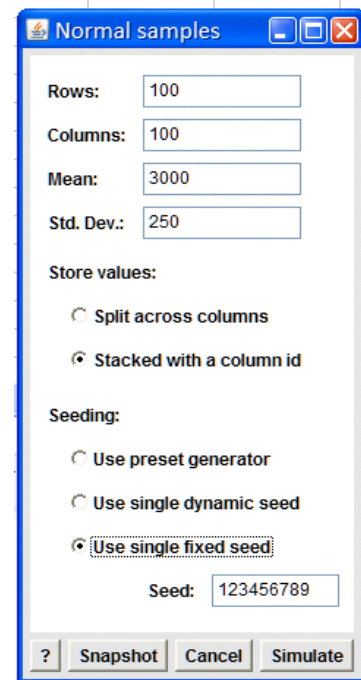
fill in the boxes on the next screen as shown on the right

what you are doing is creating 100 samples (columns = 100) of size 100 (rows = 100) and you are sampling from a population with a known mean (3000) and standard deviation (250)

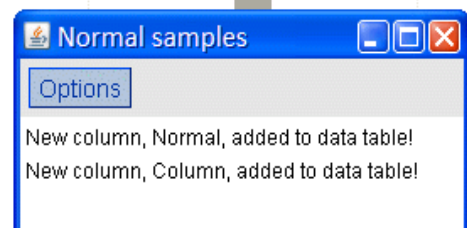
selecting the 'stack' option makes it easier to do one of the following steps, compute the mean and 95% confidence interval of each of the 100 samples

selecting the 'single fixed seed' will allow you to produce the same results that you see in this example

you can 'play' later with different size samples (rows), different number of sample (columns), different data attributes (mean, standard deviation), and different ways of generating the samples (seeding options)



you should see the screen on the right after you click on SIMULATE



WHAT IS A CONFIDENCE INTERVAL

you should see a data table that look like the one on the right

hopefully you see the SAME DATA (you should if you followed the instructions on the previous page)

Row	Normal	Column	var3
1	2972.1965	1	
2	3431.995	1	
3	2919.608	1	
4	2988.4502	1	
5	3099.9678	1	
6	2988.4834	1	
7	3083.969	1	
8	2806.0442	1	
9	3006.117	1	
10	3105.0505	1	
11	2960.6355	1	
12	3046.5657	1	
13	2943.293	1	
14	2920.3667	1	
15	2941.6199	1	
16	2937.1475	1	
17	2687.634	1	

now click on ...

STAT / T STATISTICS / ONE
SAMPLE / WITH DATA

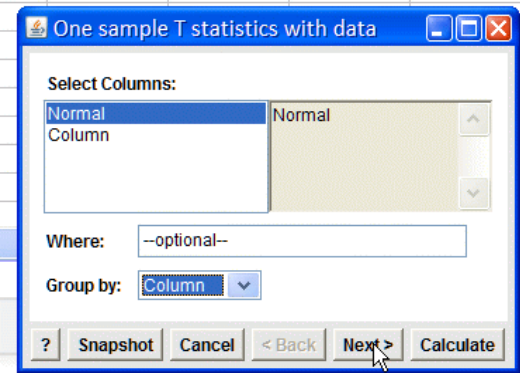
this will allow you to compute
the mean and a confidence
interval for each of the 100
samples in the data table

the variable "column" ranges
from 1 to 100 and changes for
each of the samples

Row	Normal	Summary Stats	var4	var5	var6
1	2972.1965	Tables			
2	3431.995	Z statistics			
3	2919.608	Proportions			
4	2988.4502	T statistics			
5	3099.9678	One sample			
6	2988.4834	Two sample			
7	3083.969	Paired			
8	2806.0442	Variance			
9	3006.117	Regression			
10	3105.0505	ANOVA			
11	2960.6355	Nonparametrics			
12	3046.5657	Goodness-of-fit			
13	2943.293	Control Charts			
14	2920.3667	Calculators			
15	2941.6199	Resample			
16	2937.1475				
17	2687.634				

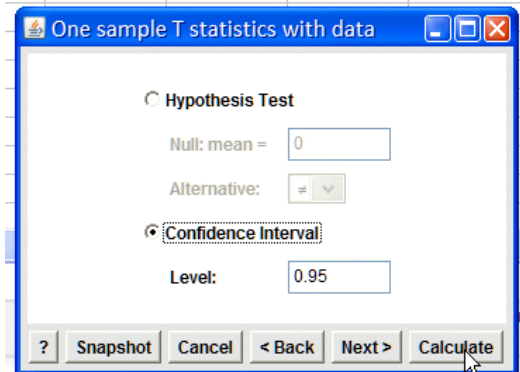
WHAT IS A CONFIDENCE INTERVAL

fill in the next screen as shown on the right and then click NEXT (do NOT click on CALCULATE)



select CONFIDENCE INTERVAL and leave the level at 0.95

click on CALCULATE



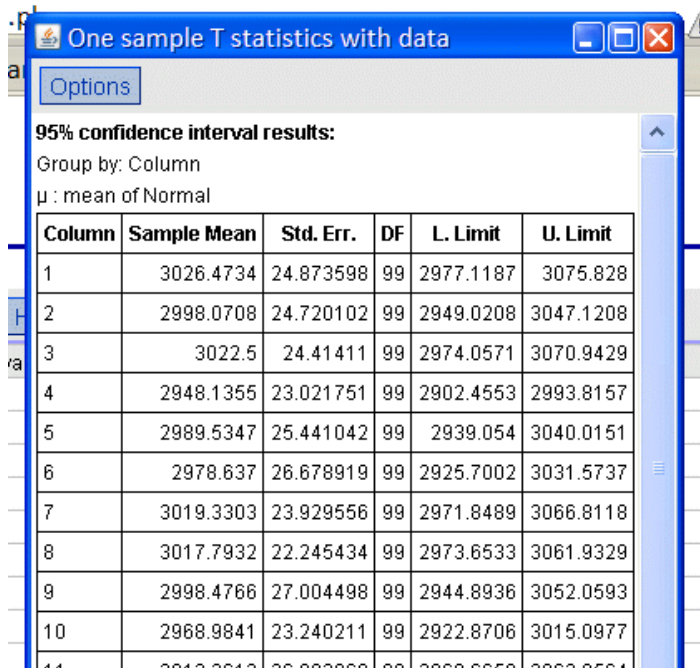
you should see the screen on the right

it will have 100 entries, a set of statistics for each of the 100 samples

if you look at the right side of the table you will see the 95% confidence interval

95% of those intervals should contain the value 3000 (remember we sampled from a population with a known mean of 3000) ... look at sample 4 ... the 95% confidence interval is 2902 to 2993

remember, in "real life" if you were trying to find the mean of a population, you would most likely take 1 sample, not 100... if you took sample 4 you would say that the mean was 2948 with a 95% confidence interval of 2902 to 2993 ... that DOES NOT contain the true population mean of 3000



Column	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
1	3026.4734	24.873598	99	2977.1187	3075.828
2	2998.0708	24.720102	99	2949.0208	3047.1208
3	3022.5	24.41411	99	2974.0571	3070.9429
4	2948.1355	23.021751	99	2902.4553	2993.8157
5	2989.5347	25.441042	99	2939.054	3040.0151
6	2978.637	26.678919	99	2925.7002	3031.5737
7	3019.3303	23.929556	99	2971.8489	3066.8118
8	3017.7932	22.245434	99	2973.6533	3061.9329
9	2998.4766	27.004498	99	2944.8936	3052.0593
10	2968.9841	23.240211	99	2922.8706	3015.0977
11	3012.2812	26.002980	99	2980.6650	3083.9564

if you look at results for all 100 samples, there should be approximately 5 confidence intervals (5 out of 100, 5%) that do NOT contain the population mean of 3000 ... that is the meaning of a confidence interval ... upon repeated sampling from a population, 95% of the confidence intervals will contain the true mean (or 5% will not contain the mean)

WHAT IS A CONFIDENCE INTERVAL

this plot (done with SAS, not Statcrunch) shows that in 7 samples the true mean of 3000 was NOT within the 95% confidence interval and that's pretty close to the 5% that we expect ... if any of those 7 samples had been the one that you took, your 95% confidence interval would not contain 3000

