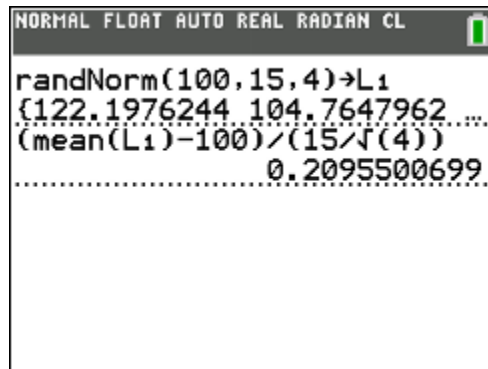


Why t ?

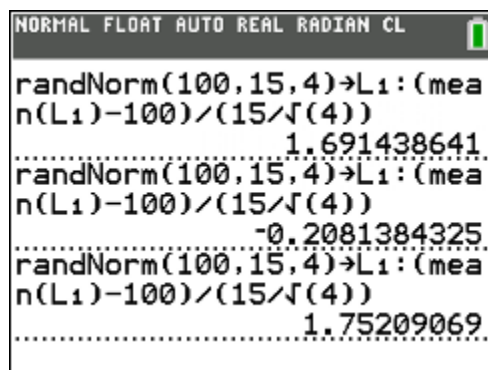
Suppose we sample IQ scores from a population that is known to be approximately normal with $\mu = 100$ and $\sigma = 15$. According to the Central Limit Theorem, the sampling distribution of \bar{x} is approximately normal with $\mu_{\bar{x}} = 100$ and $\sigma_{\bar{x}} = \frac{15}{\sqrt{n}}$.

We begin by taking samples of size 4, calculating \bar{x} , and investigating the behavior of the distribution of the standardized z-score $= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$.



```
NORMAL FLOAT AUTO REAL RADIAN CL
randNorm(100,15,4)→L1
{122.1976244 104.7647962 ...
(mean(L1)-100)/(15/√(4))
0.2095500699
```

To investigate the behavior of the test statistic, we want to repeat these steps many times.



```
NORMAL FLOAT AUTO REAL RADIAN CL
randNorm(100,15,4)→L1:(mea
n(L1)-100)/(15/√(4))
1.691438641
randNorm(100,15,4)→L1:(mea
n(L1)-100)/(15/√(4))
-0.2081384325
randNorm(100,15,4)→L1:(mea
n(L1)-100)/(15/√(4))
1.75209069
```

We should not be surprised to see that these values tend to land in the range from -3.00 to 3.00 since they are standard z-scores.

Now, suppose that σ is unknown (which is generally the case). It would be reasonable for us to replace σ with our best estimate (namely, the standard deviation of the sample s). Let's consider the consequences of such a substitution. We can go back to our simulation on the TI-84 and replace $\sigma = 15$ with the standard deviation of the sample and repeat several times to explore the distribution of this test statistic...

```
NORMAL FLOAT AUTO REAL RADIAN CL
randNorm(100,15,4)→L1:(mean(L1)-100)/(stdDev(L1)/√(4))
-1.279334847
randNorm(100,15,4)→L1:(mean(L1)-100)/(stdDev(L1)/√(4))
4.796955078
```

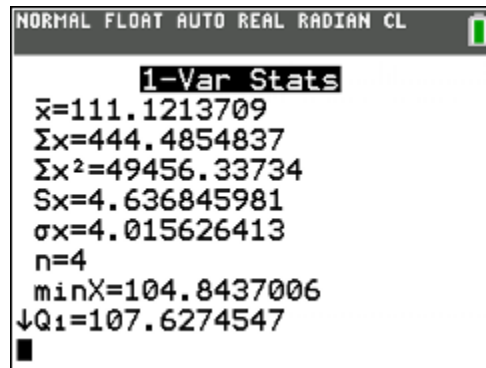
Whoa! How did we get a test statistic that was so large?? These kids must be geniuses! Let's see just how high their IQ's must be to get a test statistic of 4.797...

L1	L2	L3	L4	L5	S
104.84	-----	-----	-----	-----	
114.32					
114.91					
110.41					

L3(1)=

Wait... Recall that $\mu = 100$ and $\sigma = 15$ for this population. Not one of these kids had an IQ that was more than one standard deviation above the mean. What gives?

Recall that we are substituting s for σ . For this sample, the standard deviation appears to be much smaller than 15. In fact,



The mean for this sample is 111.12, which is larger than $\mu = 100$. If we had known that $\sigma = 15$, we would have obtained a test statistic equal to $\frac{111.12 - 100}{15/\sqrt{4}} = 1.48$ (not an unusual z score).

When we substitute $s = 4.637$ for σ , we get a very different test statistic.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

William Gosset discovered the t -distribution while serving as a chemist and statistician for Guinness Brewing Company in 1908.



Further investigation...

Suppose the sample size n is increased from 4 to 9? What happens to the t -distribution? Let's investigate... Simply change the 4's to 9's...

```

NORMAL FLOAT AUTO REAL RADIAN CL
1.249067395
randNorm(100,15,9)→L1:(mea
n(L1)-100)/(stdDev(L1)/√(9
))
-0.4254005528
randNorm(100,15,9)→L1:(mea
n(L1)-100)/(stdDev(L1)/√(9
))
0.1644169609

```

Hitting [ENTER] several times reveals that the distribution seems to “settle” more in the range from -3.00 to 3.00. It seems to follow more of a normal distribution. Why would this be?

Looking at our last sample taken...

NORMAL FLOAT AUTO REAL RADIAN CL						NORMAL FLOAT AUTO REAL RADIAN CL					
L1	L2	L3	L4	L5	Σ	1-Var Stats					
101.01	-----	-----	-----	-----		\bar{x} =100.8847897					
95.051						Σx =907.9631076					
125.74						Σx^2 =93684.73131					
111.19						Sx =16.14413249					
115.54						σx =15.22083408					
89.115						n =9					
69.851						$\min X$ =69.85093287					
102.46						$\downarrow Q_1$ =92.08321365					
98.006											

L3(1)=											

Note that as n increases, s will become less variable and the t -distribution will begin to resemble the standard z -distribution. The t -distribution is actually a family of curves each with a different degree of freedom, where $df = n - 1$.