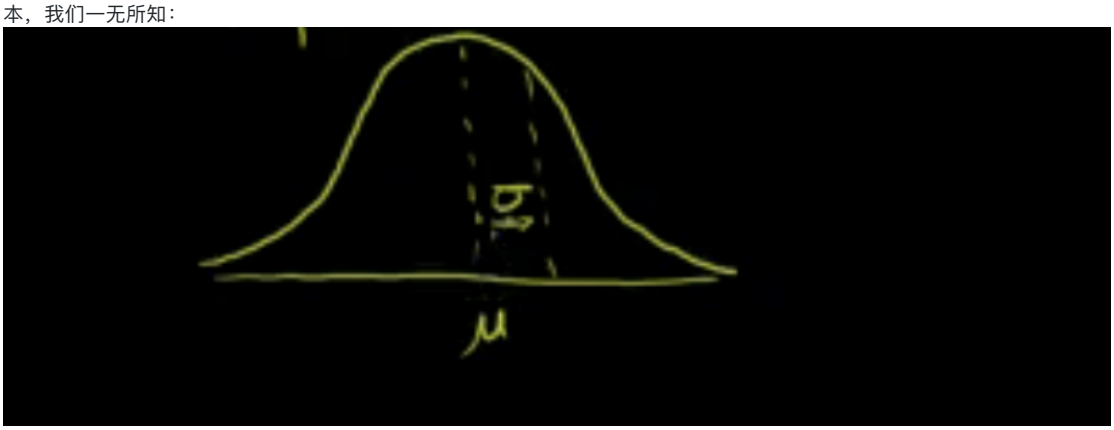
当样本容量很小时，样本均值抽样分布不应该采用正态分布，而应采用t分布。t分布与正态分布很相似，只是它有肥尾。

例如：7个患者在服用新药3个月后测量血压，血压上升分别为：1.5, 2.9, 0.9, 3.9, 3.2, 2.1, 1.9 。为总体中所有病人的血压升高的期望值建立一个95%的置信区间。

这里存在某种总体分布。因为是生物过程，有理由相信它是正态的。这相当于将药品给到所有存在过的患者，会得到一个血压升高均值，然后还会得到一定的标准差。这是大量随机事件的和，而大量随机事件的和服从正态分布。而对于总体的分布，除了样本，我们一无所知：



般情况下，我们可以先求出样本的各种统计量：

$$\bar{X} = \frac{1.5+2.9+0.9+3.9+3.2+2.1+1.9}{7} = 2.34$$

$$S^2 = \frac{(1.5-2.34)^2+(2.9-2.34)^2+(0.9-2.34)^2+(3.9-2.34)^2+(3.2-2.34)^2+(2.1-2.34)^2+(1.9-2.34)^2}{7-1}$$

$$= 1.086$$

$$S = 1.04$$

使用样本标准差来估计总体标准差：

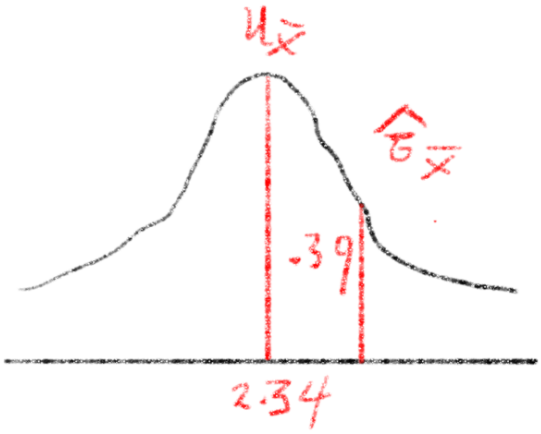
$\sigma \approx S \approx 1.04$ 因为，样本容量太少了，此时，这个估计值不能说很好。这里分布不能像原来那样认为是正态分布，可以认为它是t分布。

> n小于30通常被认为是糟糕的估计。

t分布的标准差：

$$\hat{\sigma}_{\bar{X}} = \frac{\sigma}{\sqrt{7}} = \frac{S}{\sqrt{7}} = \frac{1.04}{\sqrt{7}} = 0.393$$

图形如下：



我们要求95%的置信空间，也就是求均值左右包含95%面积的区间。t分布有t表格（t-table）。

我们这里的分布关于中轴对称，所以求的是双侧。又因为我们的抽样采用的是7个采样点，所以自由度是6。查表：

One Sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two Sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781

也就是2.447个单位的标准差，即：$2.447 \times 0.393 = 0.96$

随机抽样的均值为2.34，表示有95%的概率：2.34在总体均值周围0.96范围内。形成的置信区间为：1.38~3.3。也就是90%的可能，实际均值在这个区间范围内。