数据是对于一种感兴趣的现象的数据观察

观察得到的总体叫做population  
观察得到的部分叫做sample

我们通过数字和图形的方式(通常是使用样本数据)来理解这个集合，它可能非常庞大。

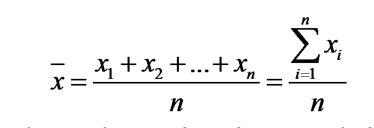
我们用形状shape、离群值outliers、中心center和扩展spread (SOCS)来描述集合。

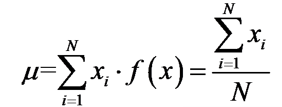
center通过mean来测量

spread通过variance来测量

sample mean:x-bar

如果从一个随机样本里有n次观察，记作 sample mean就是

,注意这个sample总共就n项

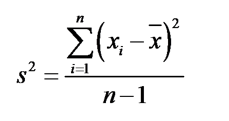
如果从总体population中共进行N次观察，记作population mean就是

比如150万人测身高，我们挑出5个人求平均数，就是sample mean

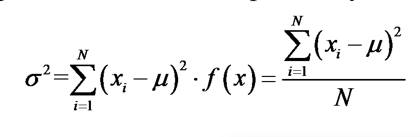
150万人所有人数据加起来 测平均数，就是Population mean

Variance，

n个obeservation 从一个sample，记作x1,x2,xn，sample variance=



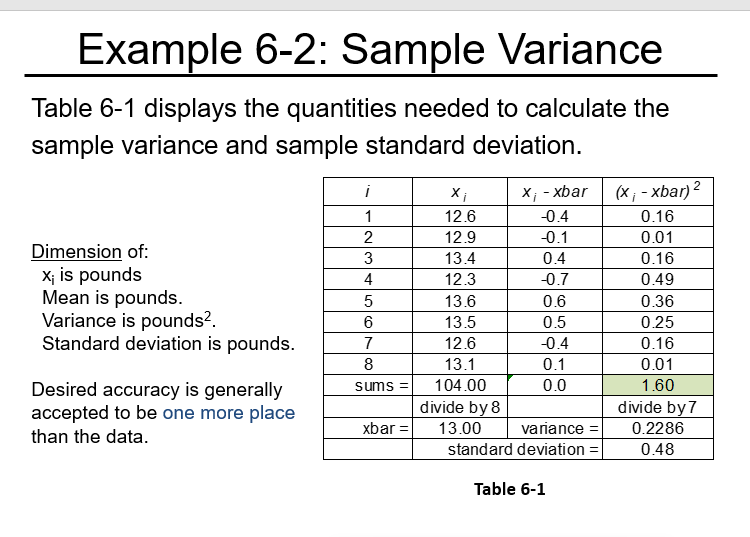
population variance=



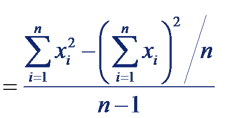
standard deviation

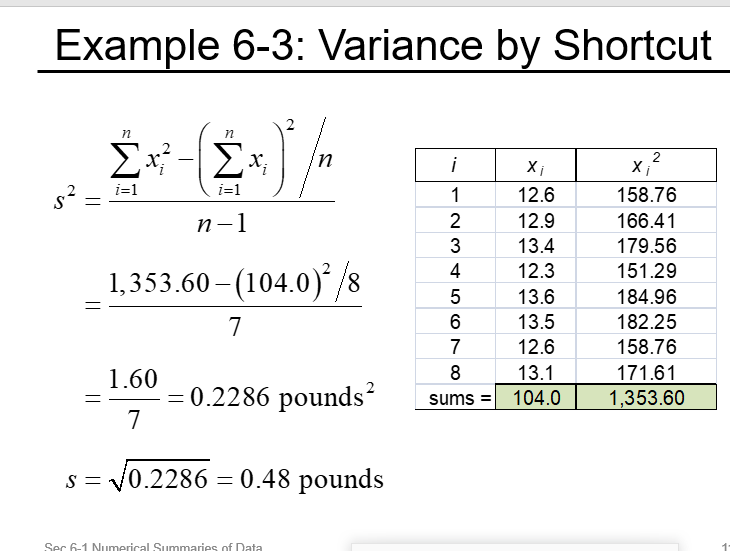
标准差就是variance开根号，

是Population variance的标准差，s是sample variance的标准差



注意sample variance除以的是n-1，

sample variance还有一种算法



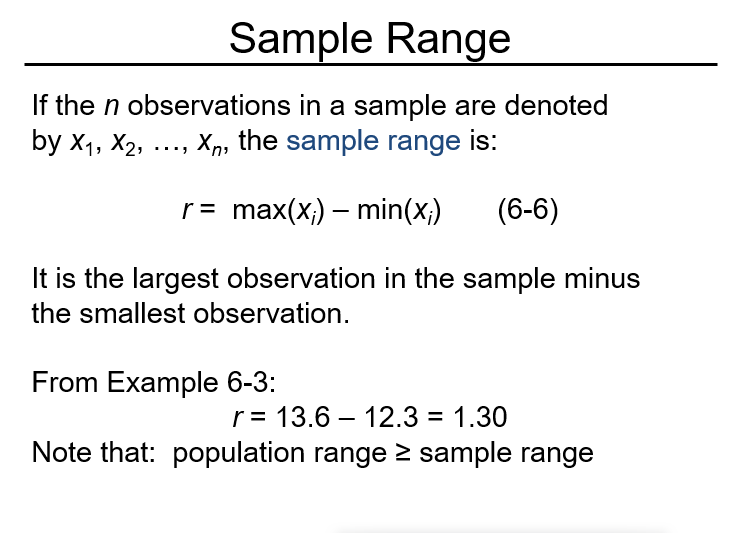
sample variancesample是用数据n-1测量的

这个·n-1 叫做 degrees of freedom

sample range

n次观察 x1, x2 ,...xn

sample range= xi里最大的减xi里最小的



Probability plot

把n次数据观察以递增顺序排列



用得到他的纵坐标

