The Sampling Distribution of The Sample Mean

- **If a random sample of n measurements is selected from a population with mean μ and standard deviation σ, the sampling distribution of the sample mean $\bar{x}$ will have mean μ and standard deviation* $\frac{\sigma}{\sqrt{n}}$.**
- **If the population has a <u>normal distribution</u>, the sampling distribution of $\bar{x}$ will be exactly normal distributed, <u>regardless of the sample size n</u>.**
- **If the population distribution is <u>nonnormal</u>, the sampling distribution $\bar{x}$ will be approximately normal distributed for <u>large samples</u> (by the Central Limit Theorem).**

***When repeated samples of size n are randomly selected from a <u>finite population with N elements</u> whose mean is μ and whose variance σ^2, the standard deviation of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.**

When N is large relative to the sample size n, $\sqrt{(N-n)/(N-1)}$ is <u>approximately equal to 1</u>, and the standard deviation of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}$.

Standard Error

***Definition:* The standard deviation of a statistic used as an estimator of a population parameter is also called the *standard error of the estimator* (abbreviated SE) because it refers to the precision of the estimator.**

Therefore, the standard deviation of $\bar{x}$ -given by

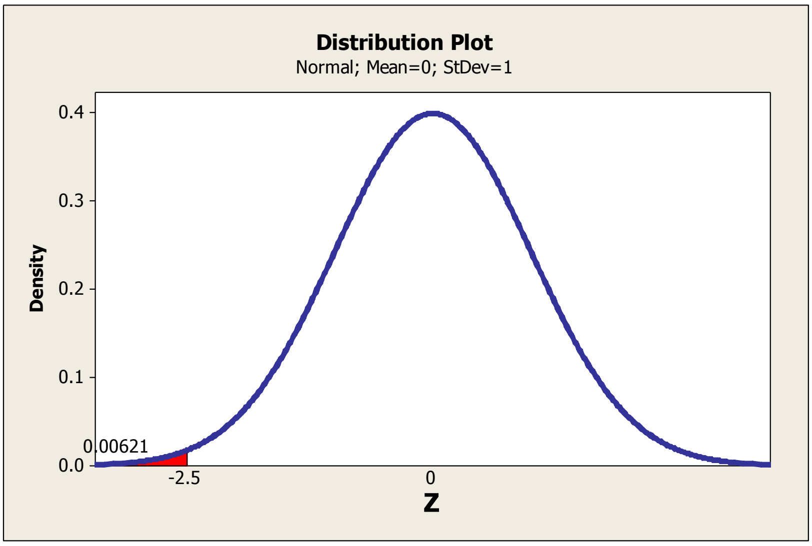
$$\frac{\sigma}{\sqrt{n}}$$

-is referred to as the standard error of the mean.

Example: An electronics company manufactures resistors that have a mean resistance of 100 ohms and standard deviation of 10 ohms. The distribution of resistors is normal. **Find the probability that a random sample of n=25 resistors will have an average resistance less than 95 ohms.**

Note that the sampling distribution of $\bar{x}$ is normal, with mean $\mu_{\bar{x}} = 100$ and a standard error of $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{10}{\sqrt{25}} = 2$.

$$P(\bar{X} < 95) = P(Z < \frac{95-100}{2})$$
$$= P(Z < -2.5) = 0.0062$$



```
MTB > random 10000 c1-c25;
SUBC> normal 100 10.
MTB > rmean c1-c25 c27
MTB > let c28=c27<95
MTB > sum c28
```

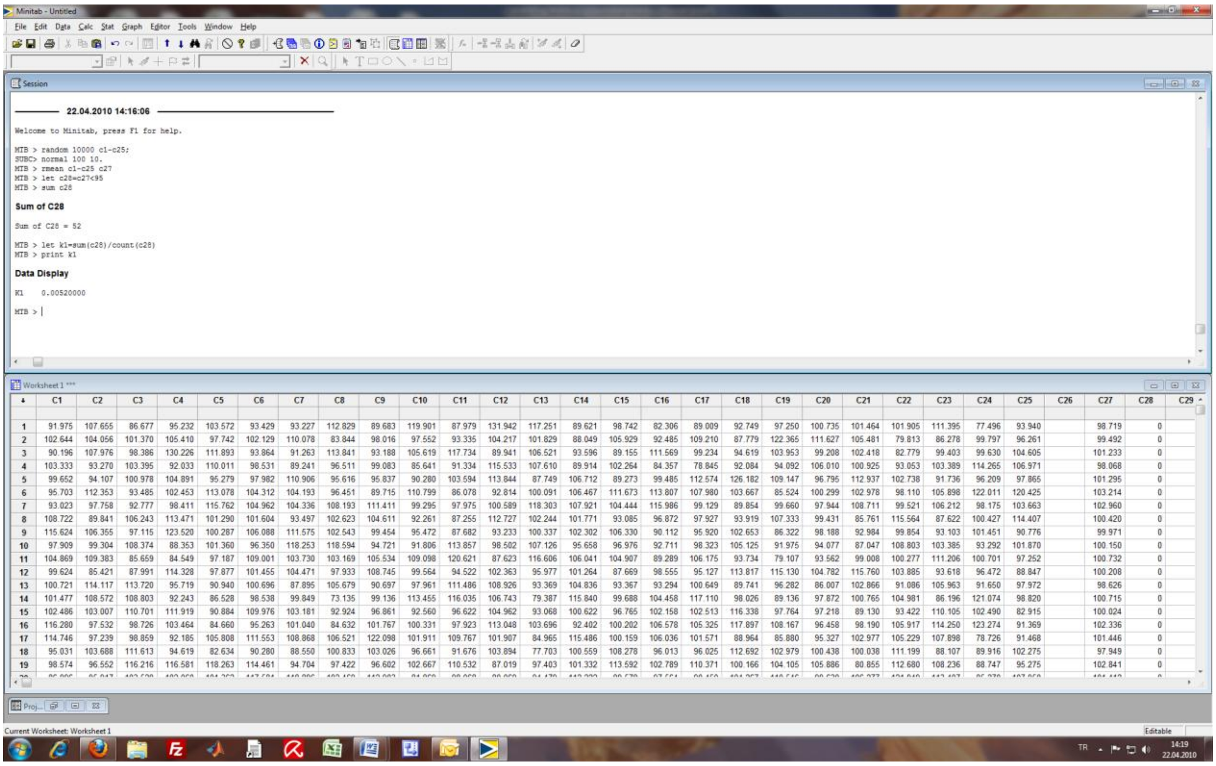
Sum of C28

```
Sum of C28 = 52
```

```
MTB > let k1=sum(c28)/count(c28)
MTB > print k1
```

Data Display

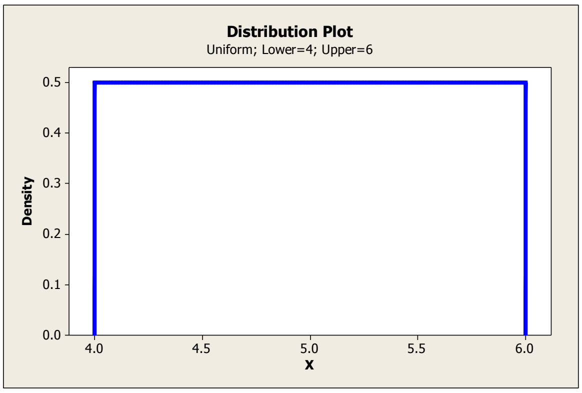
```
K1    0.00520000
```



Example: Suppose that a random variable X has a continuous uniform distribution

$$f(x) = \begin{cases} 1/2 & 4 \le x \le 6 \\ 0 & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size n=40.



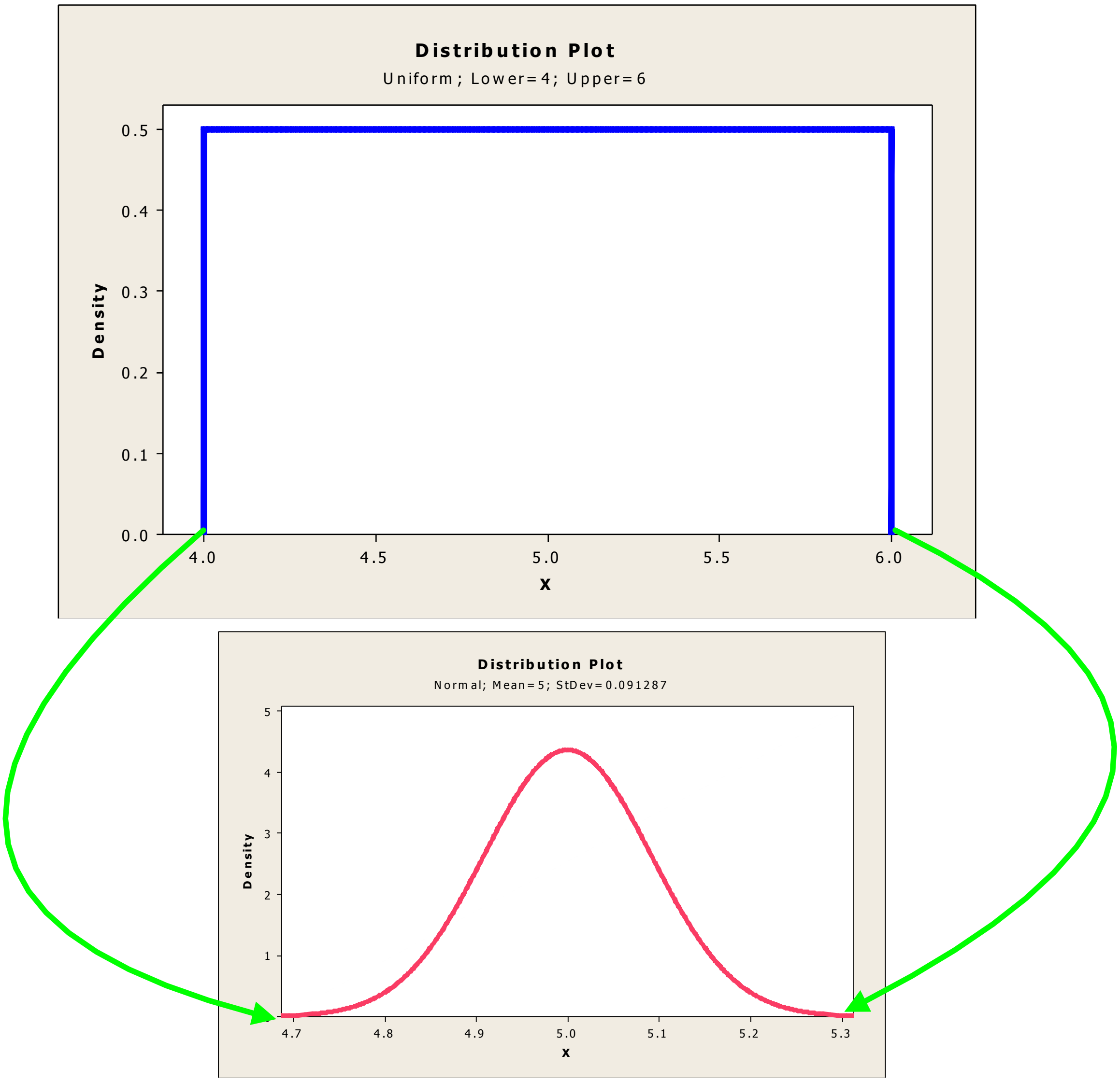
The mean and variance of X are

$\mu = 5$ **and** $\sigma^2 = (6-4)^2/12 = 1/3.$

The central limit theorem indicates that the distribution of the sample mean is approximately normal with mean $\mu_{\bar{x}} = 5$ **and variance**

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{1}{3(40)} = 1/120 = 0.0083333.$$

$$\sigma_{\bar{x}} = \sqrt{0.00833333} = 0.091287092$$



A Sampling Application: Statistical Process Control

A control Chart for the Process Mean: The $\bar{X}$ Chart

When dealing with a quality characteristic that can be expressed as a measurement, it is customary to monitor both the mean value of the quality characteristic and its variability. Control over the average quality is exercised by the control chart for averages, usually called $\bar{X}$ **Chart.**

Suppose that the process mean and standard deviation μ and σ are known and that we can assume that the quality characteristic has a normal distribution. We can use μ as the center line for the control chart, and we can place the upper and lower 3-sigma limits at

$$UCL = \mu + 3\sigma/\sqrt{n}$$

$$LCL = \mu - 3\sigma/\sqrt{n}$$

$$CL = \mu$$

When the parameters μ and σ are unknown, we usually estimate them on the basis of preliminary samples, taken when the process is thought to be in control.

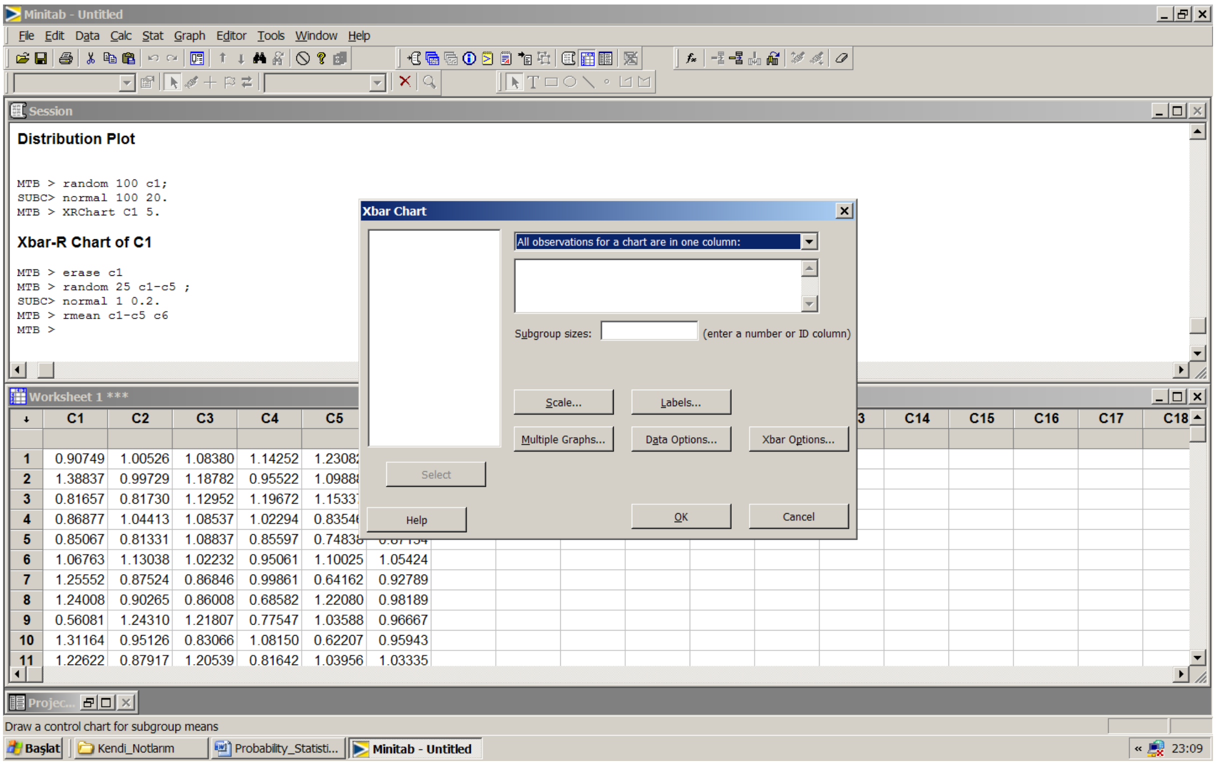
Example:

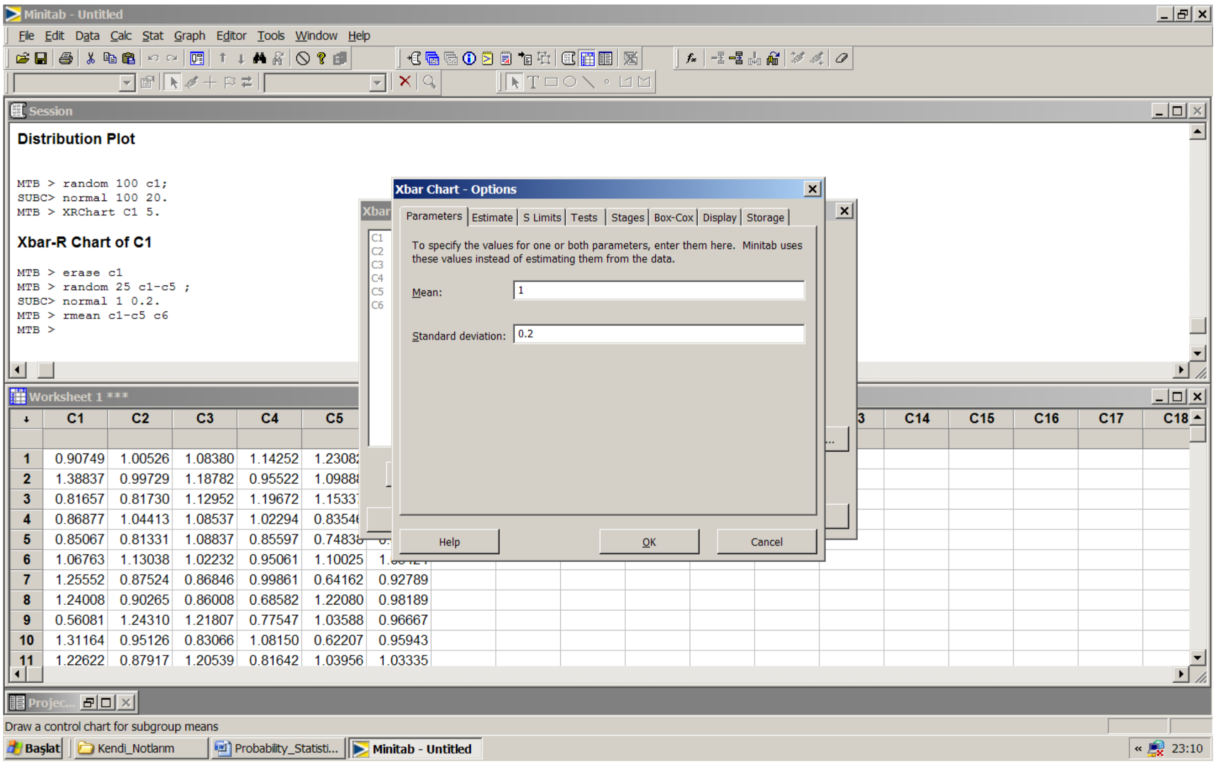
Suppose that the process mean and standard deviation **μ=1 and σ=0.2** and that we can assume that the quality characteristic has a normal distribution.

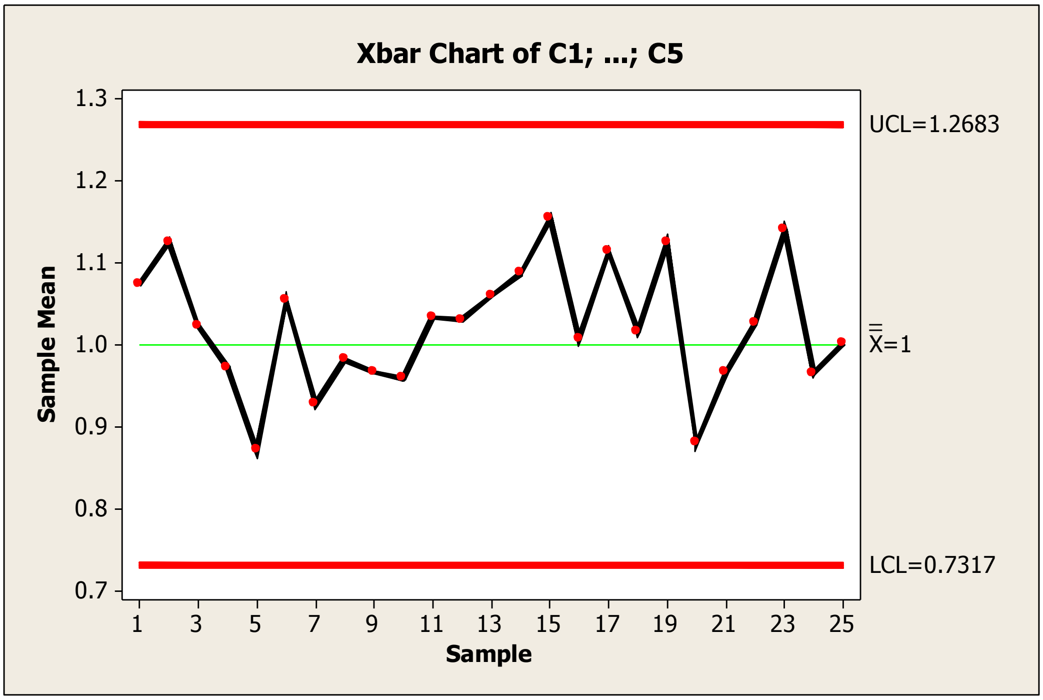
25 different samples with n=5

Sample No						**Sample Mean**
1	0.90749	1.00526	1.08380	1.14252	1.23082	1.07398
2	1.38837	0.99729	1.18782	0.95522	1.09888	1.12552
3	0.81657	0.81730	1.12952	1.19672	1.15337	1.02270
4	0.86877	1.04413	1.08537	1.02294	0.83546	0.97133
5	0.85067	0.81331	1.08837	0.85597	0.74838	0.87134
6	1.06763	1.13038	1.02232	0.95061	1.10025	1.05424
7	1.25552	0.87524	0.86846	0.99861	0.64162	0.92789
8	1.24008	0.90265	0.86008	0.68582	1.22080	0.98189
9	0.56081	1.24310	1.21807	0.77547	1.03588	0.96667
10	1.31164	0.95126	0.83066	1.08150	0.62207	0.95943
11	1.22622	0.87917	1.20539	0.81642	1.03956	1.03335
12	1.15023	1.50672	0.95409	0.94753	0.59443	1.03060
13	1.01910	1.25452	1.17502	0.77601	1.07422	1.05977
14	0.98906	0.78773	1.12180	1.25016	1.28602	1.08695
15	1.03531	1.43301	1.21064	0.89277	1.19646	1.15364
16	0.80729	0.83433	1.05260	1.11564	1.22582	1.00714
17	1.14338	1.34432	0.75482	1.19389	1.13077	1.11344
18	0.53231	1.24181	1.30629	0.79913	1.20193	1.01629
19	0.90349	1.32646	1.08458	1.30120	1.00706	1.12456
20	0.96802	0.91451	0.80393	0.50912	1.20517	0.88015
21	0.92469	1.23052	1.21457	0.56130	0.89812	0.96584
22	0.65908	0.96796	0.97453	1.23083	1.29607	1.02569
23	1.31454	1.16998	1.12052	1.15885	0.93641	1.14006
24	1.08584	0.76206	0.98833	1.15747	0.83338	0.96542
25	0.92148	1.18888	0.79026	0.99808	1.10959	1.00166

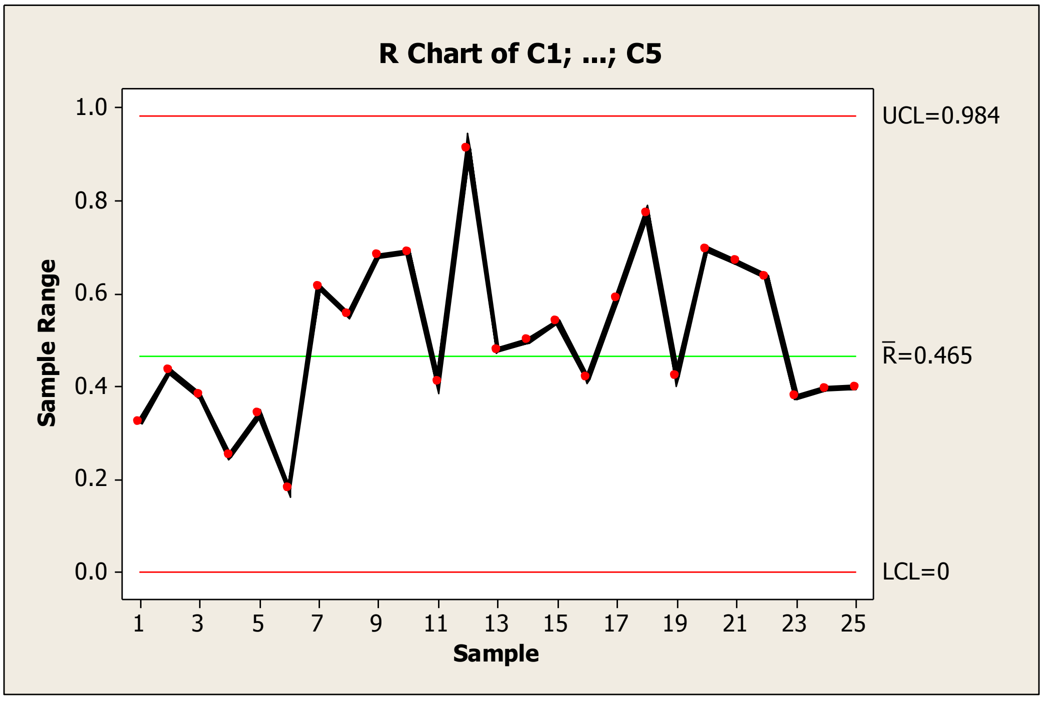
Draw $\overline{X}$ Chart.



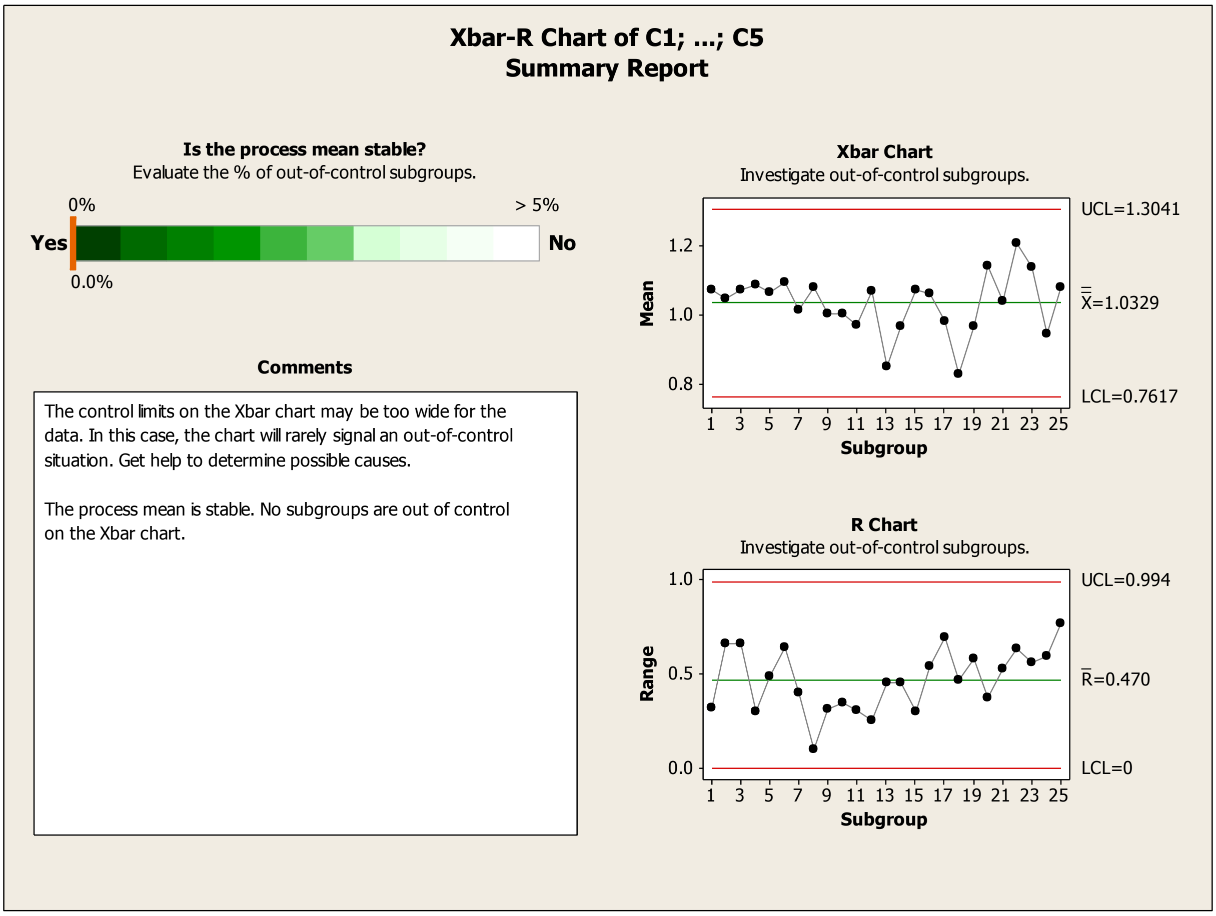




Can we say that the process is under control?



Process variability can be controlled by either a **range chart**(**R Chart**) or a standard deviation chart(**S chart**), depending on how the population standard deviation is estimated.


Xbar-R Chart of C1; ...; C5
Summary Report
Is the process mean stable?
Evaluate the % of out-of-control subgroups.
0%
> 5%
Yes
No
0.0%
Comments
The control limits on the Xbar chart may be too wide for the data. In this case, the chart will rarely signal an out-of-control situation. Get help to determine possible causes.
The process mean is stable. No subgroups are out of control on the Xbar chart.
Xbar Chart
Investigate out-of-control subgroups.
UCL=1.3041
X̿=1.0329
LCL=0.7617
Mean
Subgroup
R Chart
Investigate out-of-control subgroups.
UCL=0.994
R̄=0.470
LCL=0
Range
Subgroup

Sampling Distribution of a Difference in Sample Means

Now consider the case in which we have two independent distributions.

- **Let the first population have mean μ_1 and variance σ_1^2.**
- **The second population have mean μ_2 and variance σ_2^2.**
- **Suppose that both populations are normally distributed.**

Then the sampling distribution of $\overline{X}_1 - \overline{X}_2$ is normal with mean

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$$

and variance

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If the two populations are not normally distributed and if both sample sizes n_1 and n_2 are greater than 30, we may use central limit theorem and assume $\overline{X}_1 - \overline{X}_2$ follow approximately normal distribution.

Approximate Sampling Distribution of a Difference in Sample Means

Definition: If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

is approximately standard normal, if the conditions of central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

```
>  random 1000 c1-c50;
SUBC> exponential 1.
MTB > rmean c1-c25 c51
MTB > rmean c26-c50 c52
MTB > desc c51-c52
```

Descriptive Statistics: C51; C52

```
Variable      N  N*     Mean  SE Mean    StDev  Minimum       Q1
Median
C51        1000   0  0.99595  0.00620  0.19600  0.53058  0.86307
0.98592
C52        1000   0   1.0007  0.00639   0.2019   0.5522   0.8509
0.9826

Variable       Q3  Maximum
C51       1.11458  1.69254
C52        1.1421   1.7724
```

$1/\sqrt{25}=0.2$

$0.2/\sqrt{1000}=0.00632455$

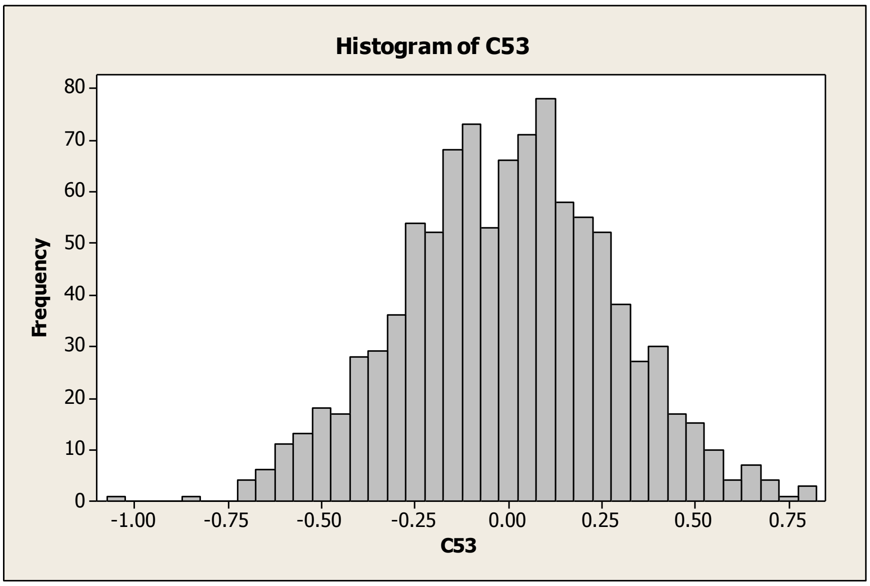
```
MTB > let c53=c51-c52
MTB > desc c53
```

Descriptive Statistics: C53

```
Variable      N  N*      Mean  SE Mean    StDev   Minimum
Q1    Median
C53        1000   0  -0.00480  0.00891  0.28170  -1.06471  -
0.19405  0.00288

Variable       Q3  Maximum
C53       0.18668  0.81055
```

$1/\sqrt{25}+1/\sqrt{25}=0.4$



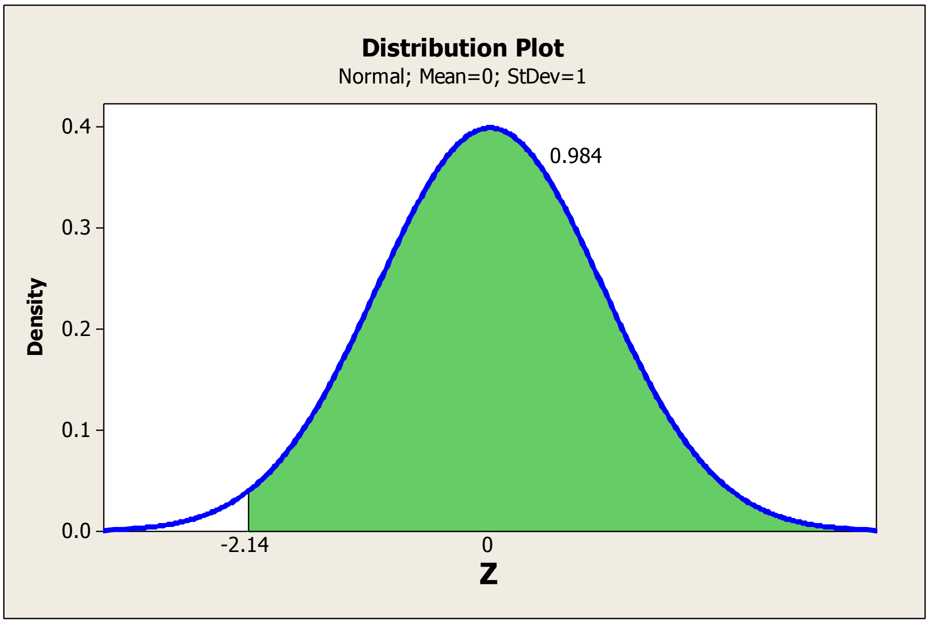
Example: The effective life of a component is a random variable with **mean 5000 hours** and **standard deviation 40 hours.** The distribution of effective life is fairly close to a normal distribution. The manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life **to 5050 hours** and decreases the standard deviation to **30 hours. Suppose that a random sample of n_1=16 components is selected from the old process and a random sample of n_2=25 components is selected from the "improved" process.**

Assume that the old and improved process can be regarded as independent populations.

What is the probability that the difference in the two sample means $\overline{X}_2 - \overline{X}_1$ is at least 25 hours?

The distribution of $\bar{X}_2 - \bar{X}_1$ **is normal with mean** $\mu_2 - \mu_1 = 5050 - 5000 = 50$ **hours and the variance** $\dfrac{\sigma_2^2}{n_2} + \dfrac{\sigma_1^2}{n_1} = \dfrac{30^2}{25} + \dfrac{40^2}{16} = 136$ **hours²**.

$$P(\bar{X}_2 - \bar{X}_1 \geq 25) = P(Z \geq \frac{25-50}{\sqrt{136}})$$
$$= P(Z \geq -2.14) = 0.9838$$



```
MTB > random 1000 c1-c16;
SUBC> normal 5000 40.
MTB > random 1000 c17-c41;
SUBC> normal 5050 30.
MTB > rmean c1-c16 c42
MTB > rmean c17-c41 c43
MTB > let c44=c43-c42
MTB > hist c43
```

Histogram of C43

```
MTB > hist c44
```

Histogram of C44

```
MTB > let c45=c43>=25
MTB > let k1=sum(c45)/count(c45)
MTB > print k1
```

Data Display

```
K1     1.00000  ??????
```