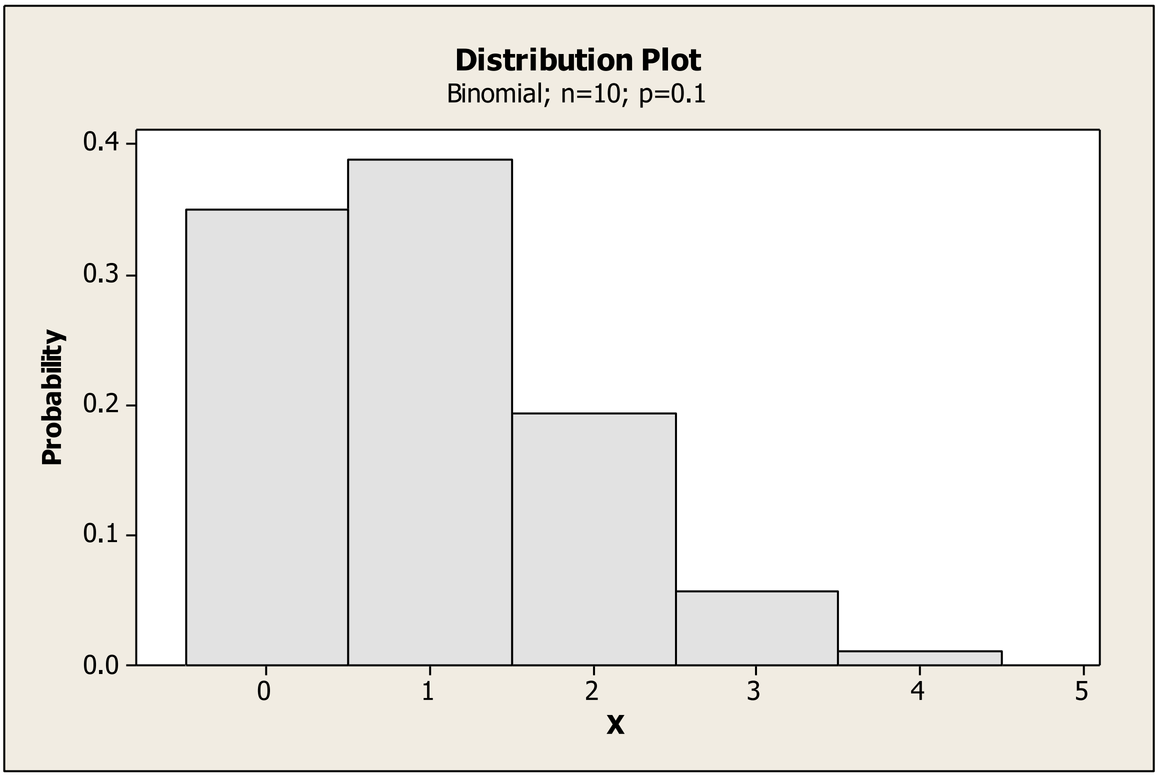
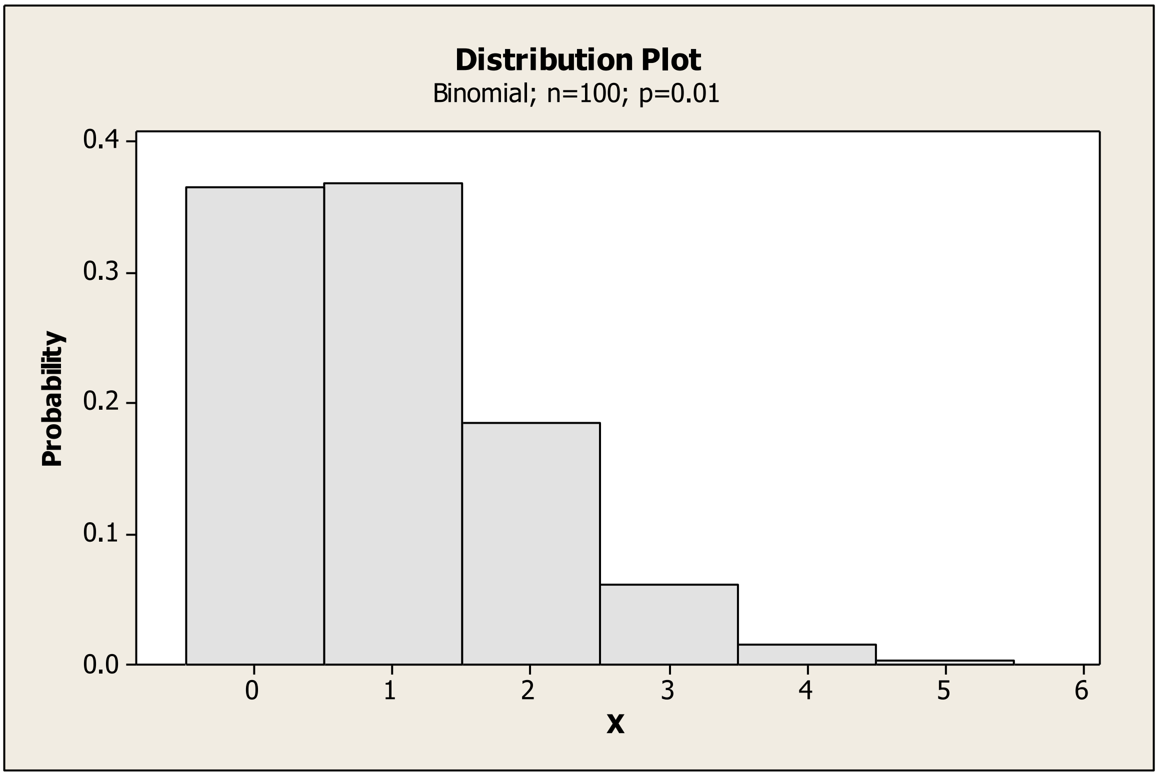
Normal Approximation to the Binomial and Poisson Distributions

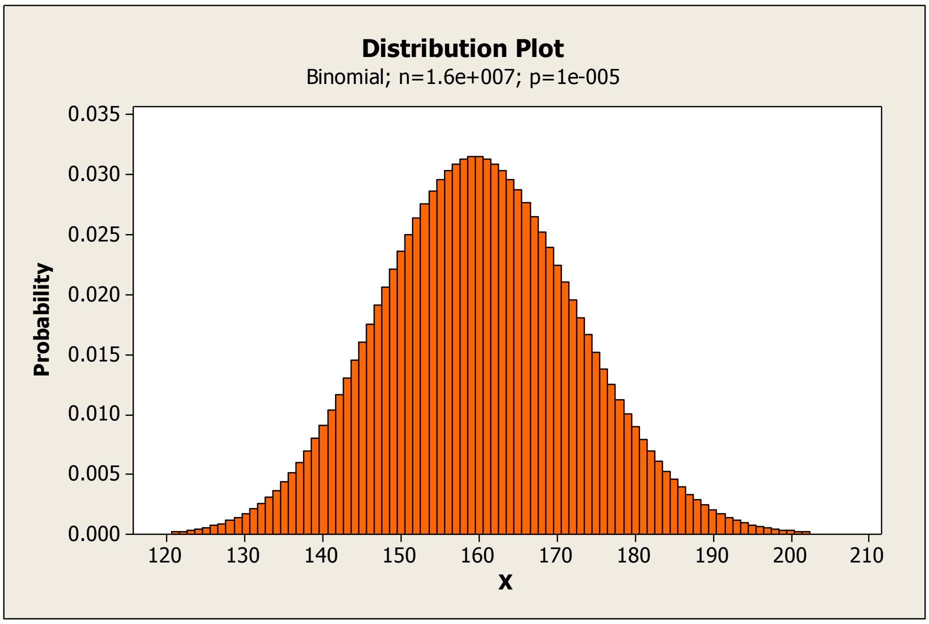




Highly skewed. But if n >> ?

In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable, and assume that a bit is received in **error is $1x10^{-5}$**.

If 16.000.000 bits (16 million) are transmitted, what is the probability that 150 or fewer errors occur?



Let the random variable X denote the number of errors. Then X is a binomial random variable and

$$P(X \leq 150) = \sum_{x=0}^{150} \binom{16000000}{x} (1*10^{-5})^x (1-1*10^{-5})^{16000000-x}$$

Clearly computation of this probability is difficult.

```
MTB > cdf 150 ;
SUBC> binomial 16000000 0.00001.
```

Cumulative Distribution Function

```
Binomial with n = 16000000 and p = 0.00001

  x   P( X <= x )
150      0.228031
```

Normal Approximation to the Binomial Distribution

If X is a binomial random variable with parameters n and p

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately a standard normal random variable.

The approximation is good for

$np > 5$ ***and*** $n(1-p) > 5$

To approximate a binomial probability with a normal distribution a continuity correction is applied as follows

$$P(X \leq x) = P(X \leq x + 0.5) = P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{np(1-p)}} \right)$$

$$P(X \geq x) = P(X \geq x - 0.5) = P\left(Z \geq \frac{x - 0.5 - np}{\sqrt{np(1-p)}} \right)$$

The digital communication problem in the previous example is solved as follows

$$P(X \le 150) = P(X \le 150.5) = P\left(Z \le \frac{150.5-160}{\sqrt{160(1-10^{-5})}} \right)$$

$$= P(Z \le -0.75) = 0.227$$

$np = 16000000(10^{-5}) = 160$

and

$n(1-p) = 16000000(1-10^{-5}) = 159998400$ **is much larger, the approximation is expected to work well.**

Again, consider the transmission of bits. To judge how well the normal approximation works, assume only n=50 bits are to be transmitted and that the probability of an error is p=0.1.

The exact probability that 2 or less errors occur is

$$P(X \leq 2) = \binom{50}{0} 0.90^{50} + \binom{50}{1} 0.1)(0.9^{49}) + \binom{50}{2} 0.1^{2}(0.9^{48}) = 0.112$$

```
MTB > cdf 2;
SUBC> Binomial 50 0.1.
```

Cumulative Distribution Function

```
Binomial with n = 50 and p = 0.1

x  P( X <= x )
2       0.111729
```

Based on the normal approximation

$$P(X \leq 2) = P(X \leq 2.5) = P\left(Z \leq \frac{2.5-5}{\sqrt{50(0.1)(0.9)}} \right)$$

$$= P(Z \leq -1.18) = 0.119$$

```
MTB > cdf -1.18
```

Cumulative Distribution Function

```
Normal with mean = 0 and standard
deviation = 1
     x   P( X <= x )
-1.18       0.119000
```

What is the probability P(X=5)?
Exact probability

$$P(X=5)=\binom{50}{5}(0.1^{5})(0.90^{45})=0.1849$$

```
MTB > pdf 5;
SUBC> Binomial 50 0.1 .
```

Probability Density Function

```
Binomial with n = 50 and p = 0.1
x   P( X = x )
5     0.184925
```

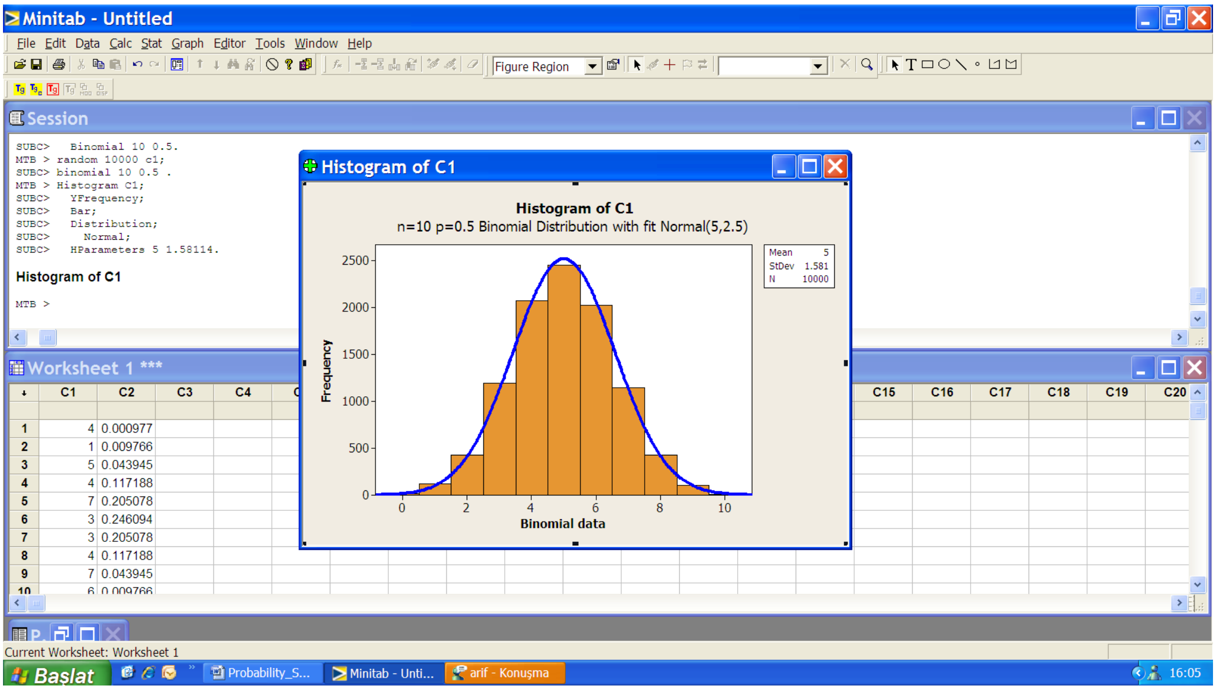
Approximate probability with normal distribution

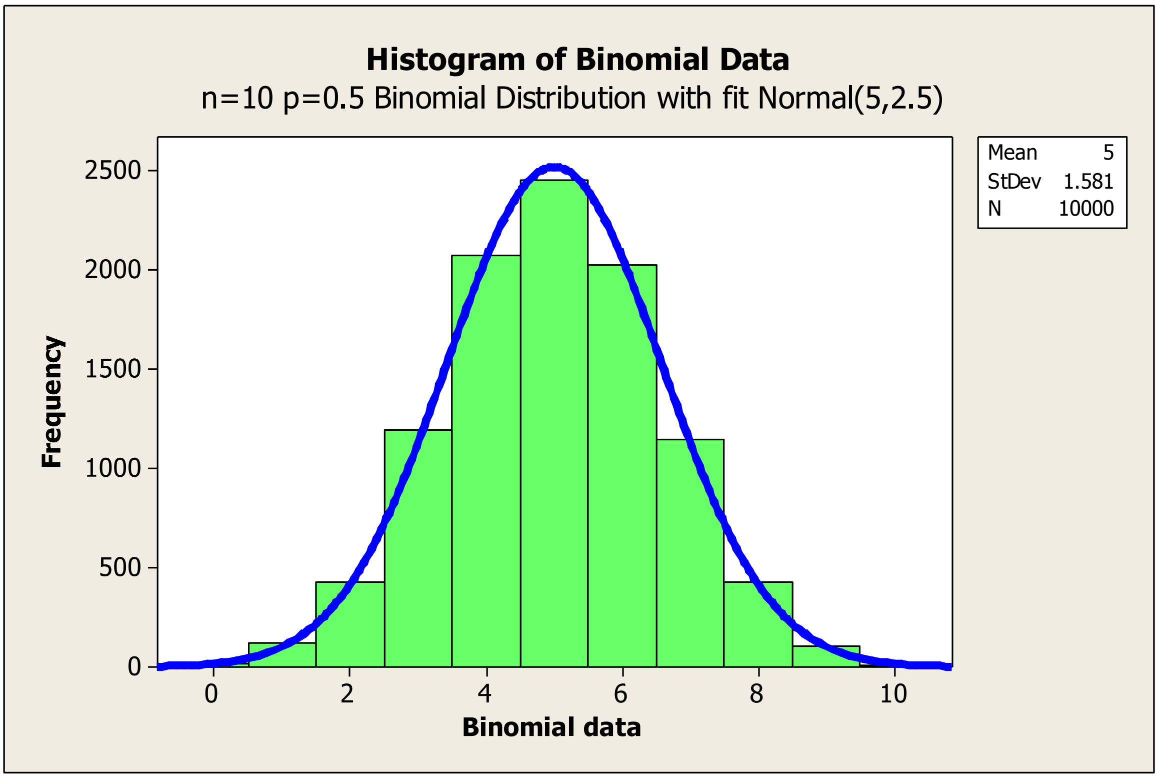
$$P(X=5)=P(4.5<X<5.5)=$$
$$P\left(\frac{5-0.5-5}{\sqrt{50(0.1)(0.9)}}<Z<\frac{5+0.5-5}{\sqrt{50(0.1)(0.9)}}\right)$$
$$=P(-0.24<Z\leq 0.24)=0.189670$$

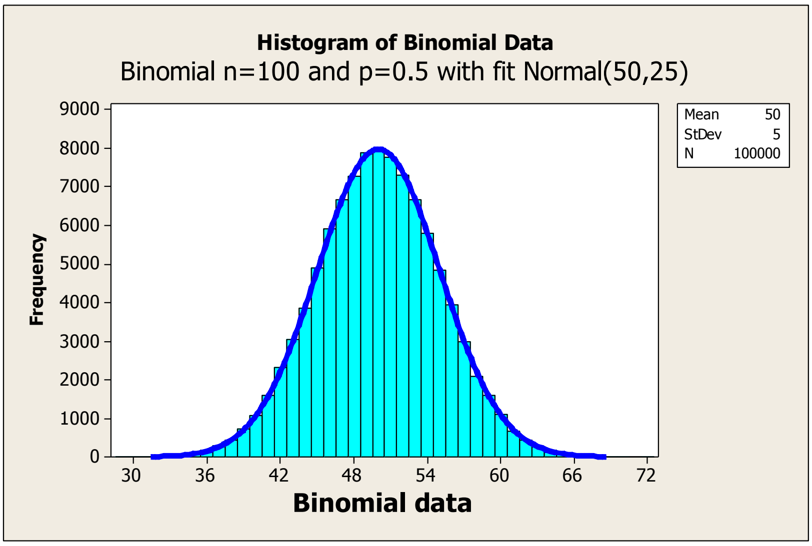
```
MTB > cdf 0.24 k1
MTB > cdf -0.24 k2
MTB > let k3=k1-k2
MTB > print k3
```

Data Display

```
K3    0.189670
```

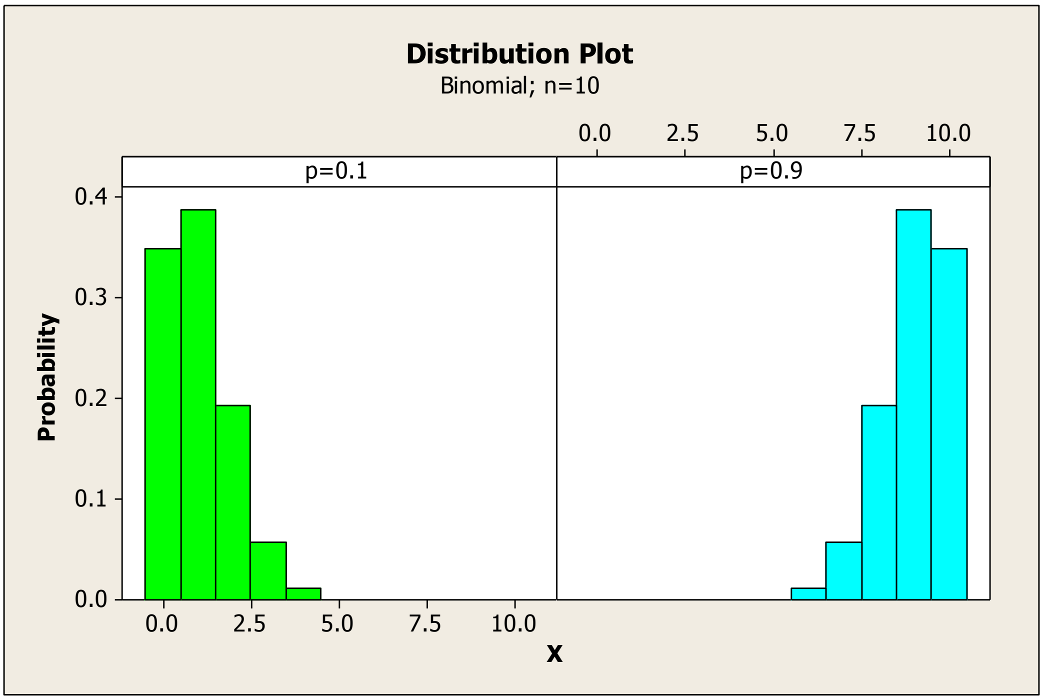

Minitab - Untitled
File Edit Data Calc Stat Graph Editor Tools Window Help
Session
Histogram of C1
Histogram of C1
n=10 p=0.5 Binomial Distribution with fit Normal(5,2.5)
Frequency
Binomial data
Worksheet 1 ***
Current Worksheet: Worksheet 1


Histogram of Binomial Data
n=10 p=0.5 Binomial Distribution with fit Normal(5,2.5)
Mean 5
StDev 1.581
N 10000
Frequency
2500
2000
1500
1000
500
0
0
2
4
6
8
10
Binomial data



The correction factor is used to improve the approximation. However, if

np ***or*** $n(1-p)$ **is small, <u>the binomial distribution is quite skewed and the symmetric normal distribution is not a good approximation</u>.**



Normal Approximation to the Poisson Distributions

We know that the Poisson distribution was developed as the limit of binomial distribution as the number of trials increased to infinity. Consequently, it should not be surprising to find that the normal distribution can also be used to approximate probabilities of a Poisson random variable.

If X is a Poisson random variable with E(X)=λ and V(X)=λ,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is approximately a standard normal random variable.

The approximation is good for

$$\lambda > 5$$

Example: Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000.

If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

This probability can be expressed exactly as

$$P(X \leq 950) = \sum_{x=0}^{950} \frac{e^{-1000} 1000^x}{x!} = ?$$

The computational difficulty is clear.

```
MTB > cdf 950;
SUBC> poisson 1000.
* ERROR * One or more arguments too large.
        * Subcommand ignored.
```

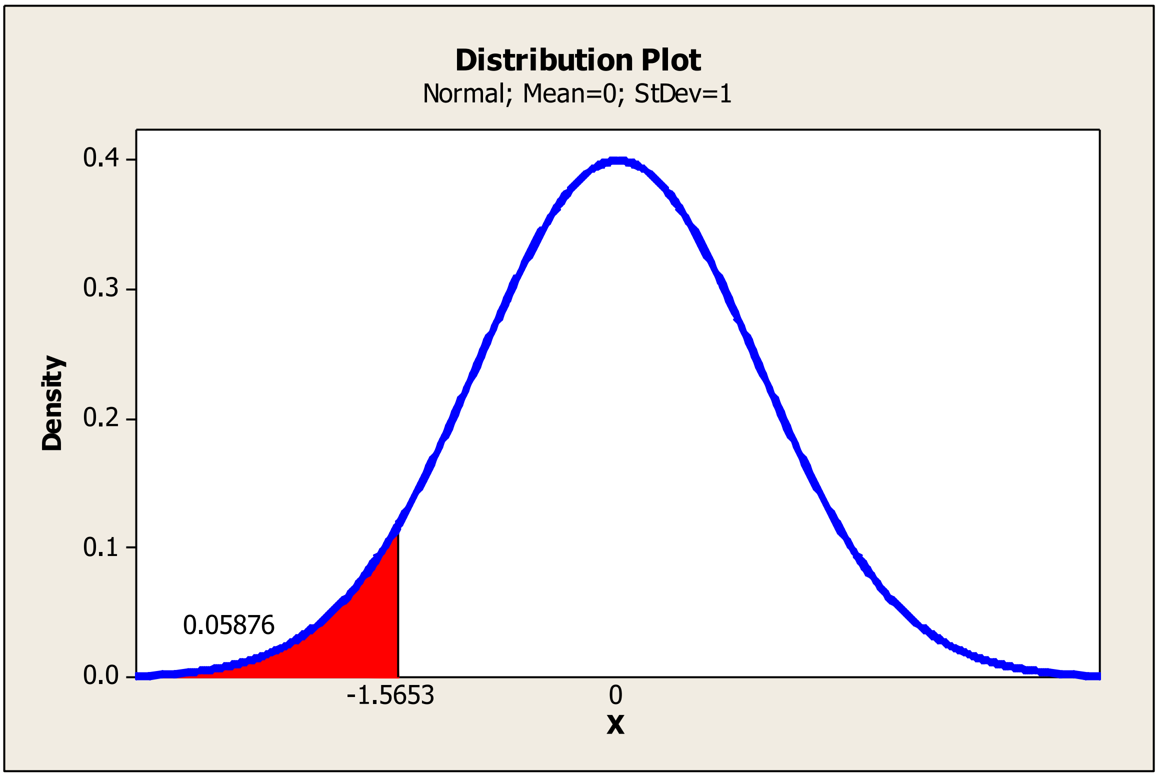
The probability can be approximated as

$$P(X \le 950) = P\left(Z \le \frac{950.5 - 1000}{\sqrt{1000}} \right)$$

$$= P(Z \le -1.5653) = 0.05875$$

```
MTB > cdf -1.5653
 Cumulative Distribution Function
Normal with mean = 0 and standard
deviation = 1

      x   P( X <= x )
-1.5653     0.0587563
```



Poisson distribution with a mean of 100.

If a squared meter of dust is analyzed, what is the probability that 95 or fewer particles are found?

```
MTB > cdf 95;
SUBC> poisson 100.
```

Cumulative Distribution Function

```
Poisson with mean = 100

 x   P( X <= x )
95      0.331192
```

The probability can be approximated as

$$P(X \leq 95) = P\left(Z \leq \frac{95.5 - 100}{\sqrt{100}}\right)$$

$$= P(Z \leq -0.45) = 0.326355$$

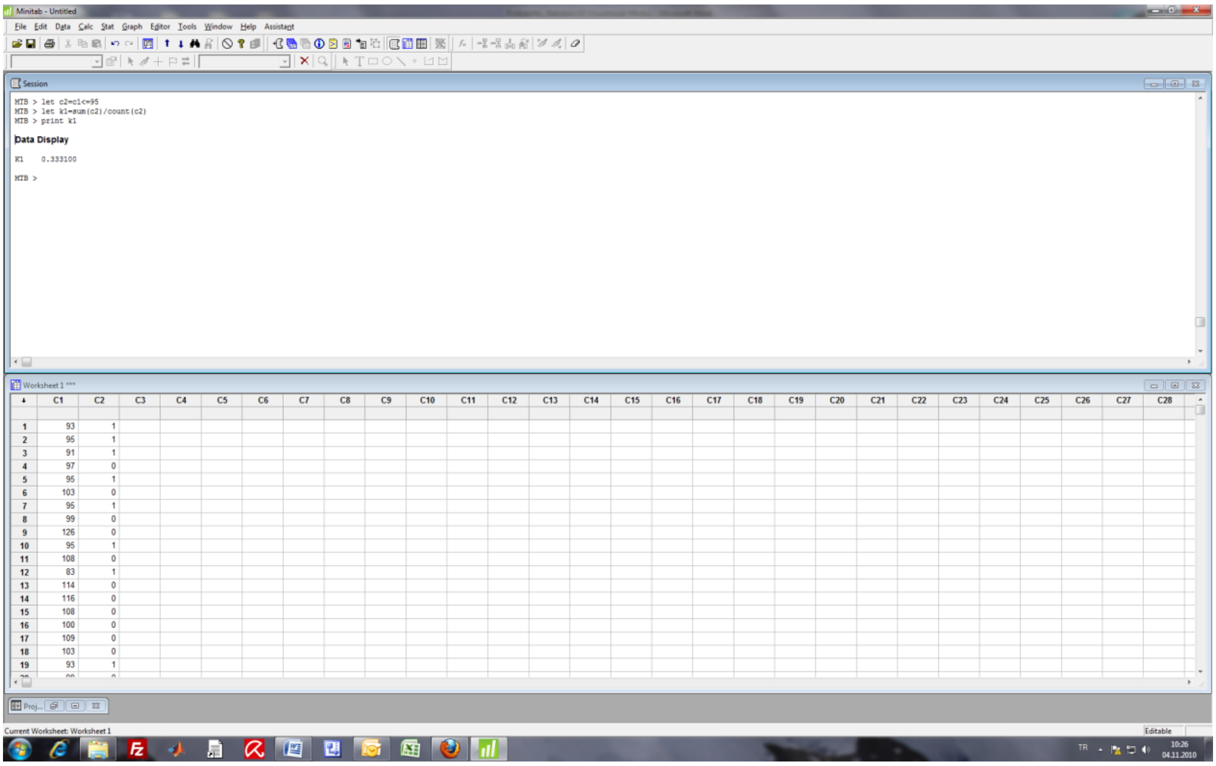
```
MTB > cdf -0.45
```

Cumulative Distribution Function

```
Normal with mean = 0 and standard
deviation = 1

    x   P( X <= x )
-0.45      0.326355
```

Solution by Simulation



```
MTB > Random 10000 c1;
SUBC>   Poisson 100.
MTB > let c2=c1<=95
MTB > let k1=sum(c2)/count(c2)
MTB > print k1
```

Data Display

```
K1      0.333100
```

compare results

```
MTB > cdf 9.5;
SUBC> poissson 10.
```

Cumulative Distribution Function

```
Poisson with mean = 10

  x   P( X <= x )
9.5       0.457930
```

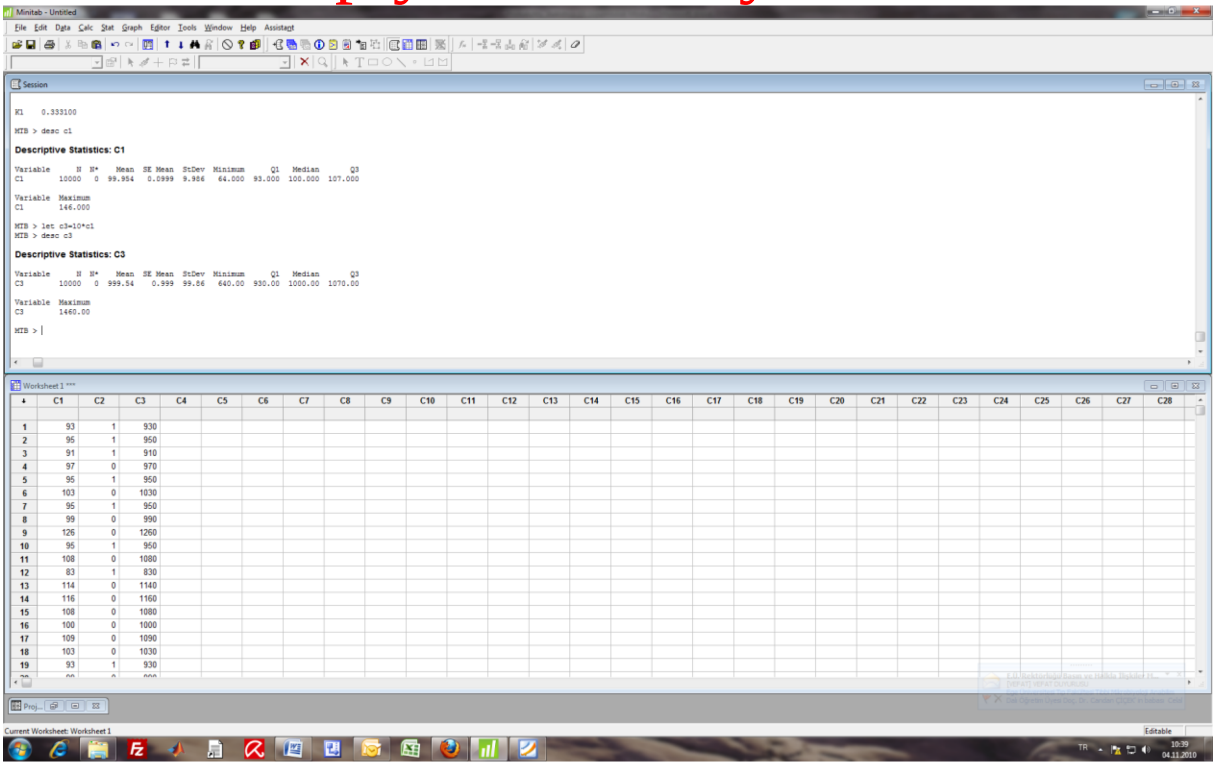
```
MTB > cdf 95;
SUBC> poisson 100.
```

Cumulative Distribution Function

```
Poisson with mean = 100

 x   P( X <= x )
95       0.331192
```

If we multiply the data by 10



Descriptive Statistics: C1

```
Variable        N  N*    Mean  SE Mean  StDev  Minimum      Q1
Median        Q3
C1          10000   0  99.954   0.0999  9.986   64.000  93.000
100.000  107.000

Variable  Maximum
C1        146.000
```

Mean and variance are equal but now !!!

```
MTB > let c3=10*c1
MTB > desc c3
```

Descriptive Statistics: C3

```
Variable        N  N*    Mean  SE Mean  StDev  Minimum      Q1
Median        Q3
C3          10000   0  999.54    0.999  99.86   640.00  930.00
1000.00  1070.00

Variable  Maximum
C3        1460.00
```

Different mean and variance

So we don't solve the previous problem dividing the lambda by 10.