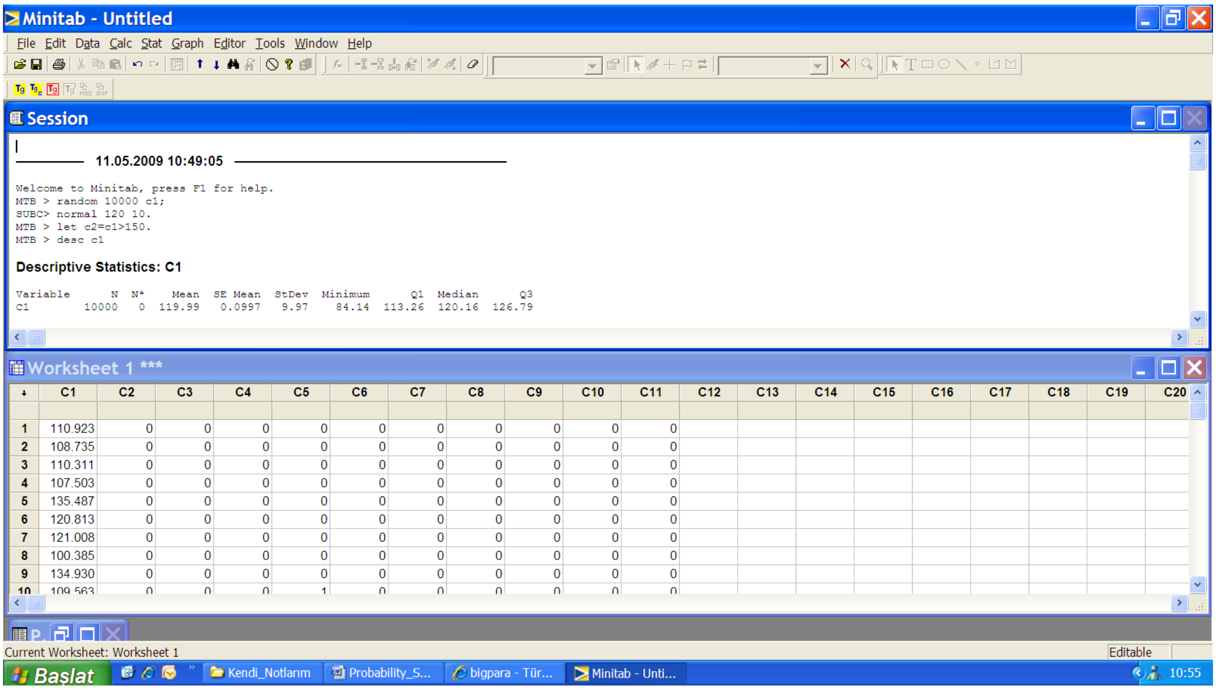
The Sampling Distribution of The Sample Proportion

- **If a random sample of n measurements is selected from a binomial population with parameter p, then the sampling distribution of the sample proportion** $\hat{p} = \frac{x}{n}$ **will have a mean p and a standard deviation** $SE(\hat{p}) = \sqrt{\frac{pq}{n}}$ **where** $q = 1 - p$.
- **When the sample size n is large, the sampling distribution of** $\hat{p}$ **can be approximated by a normal distribution.**
- **The approximation will be adequate if np>5 and nq>5.**

1. **Find the necessary values of n and p.**
2. **Check whether the normal approximation to the binomial distribution is appropriate (np>5 and nq>5).**
3. **Write down the event of interest in terms of** $\hat{p}$**, and locate the appropriate area on the normal curve.**
4. **Convert the necessary values of** $\hat{p}$ **to z-values using** $z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$
5. **Use standard normal table calculate the probability**

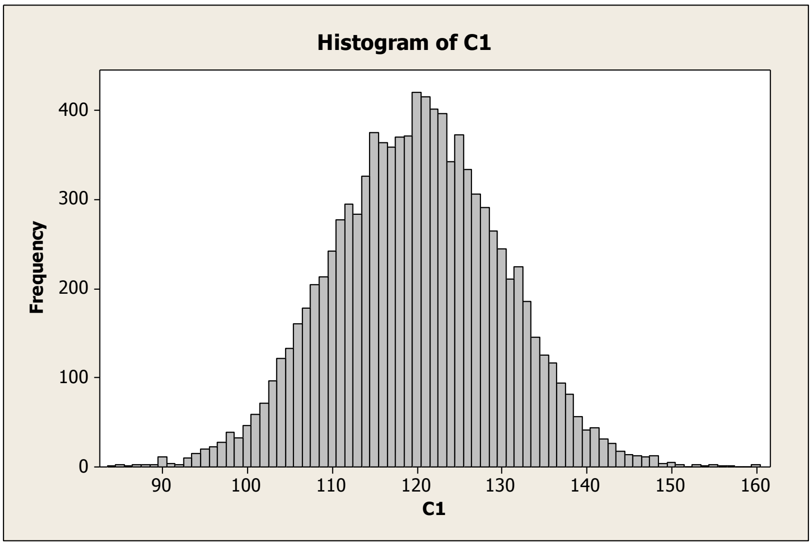


```
MTB > random 10000 c1;
SUBC> normal 120 10.
MTB > desc c1
```

Descriptive Statistics: C1

```
Variable      N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3
C1        10000   0  119.99   0.0997   9.97    84.14  113.26  120.16  126.79

Variable  Maximum
C1         160.18
```



```
MTB > let c2=c1>140
MTB > table c2
```

Tabulated statistics: C2

```
Rows:  C2

        Count
0        9786
1         214
All     10000
MTB > let k1=sum(c2)/10000
MTB > print k1
```

Data Display

```
K1     0.0214000

MTB > sample 100 c2 c3
MTB > sample 100 c2 c4
MTB > sample 100 c2 c5
MTB > sample 100 c2 c6
MTB > sample 100 c2 c7
MTB > sample 100 c2 c8
MTB > sample 100 c2 c9
MTB > sample 100 c2 c10
MTB > sample 100 c2 c11

EXPECTED 100*0.0214=2.14
```

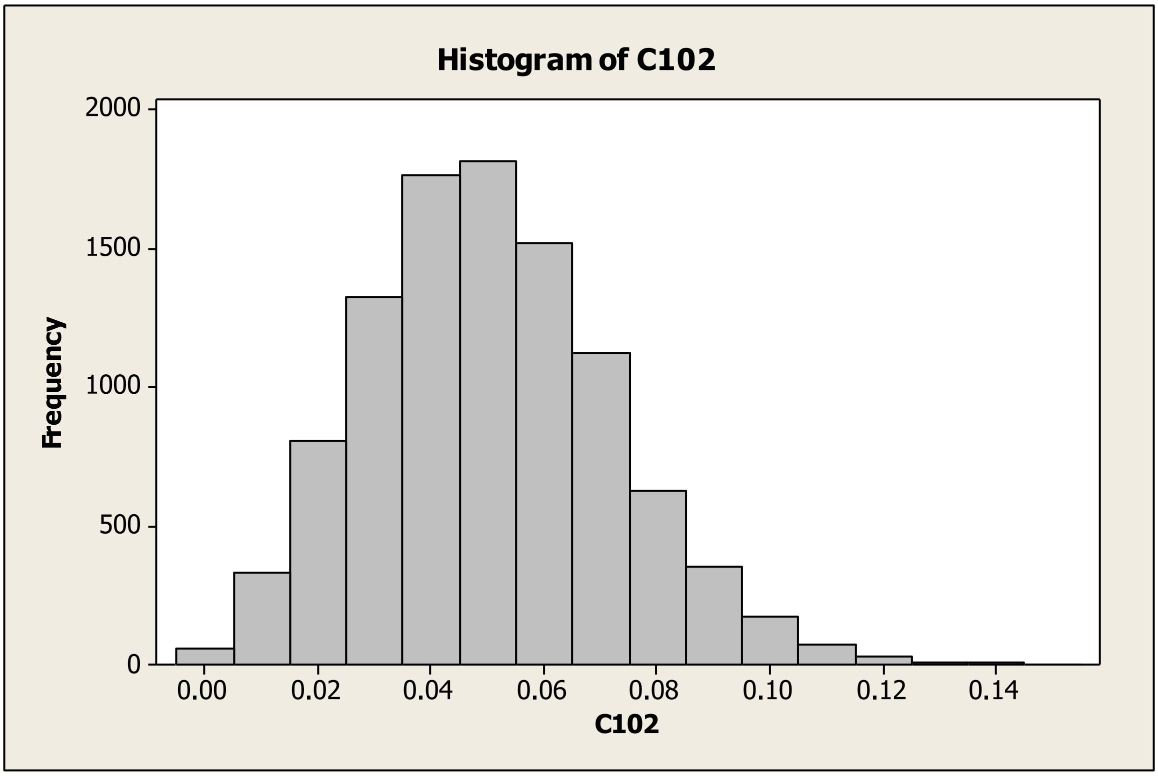
Sum of C3 = 2
Sum of C4 = 0
Sum of C5 = 6
Sum of C6 = 4
Sum of C7 = 3
Sum of C8 = 5
Sum of C9 = 2
Sum of C10 = 1
Sum of C11 = 1

```
MTB > Random 10000 c1-c100;
SUBC>   Bernoulli 0.05.
MTB > rsum c1-c100 c101
MTB > let c102=c101/100
MTB > desc c102
```

Descriptive Statistics: C102

```
Variable        N   N*        Mean    SE Mean       StDev    Minimum
Q1      Median
C102        10000    0   0.050197   0.000218   0.021827   0.000000
0.030000   0.050000

Variable             Q3    Maximum
C102        0.060000   0.150000
```



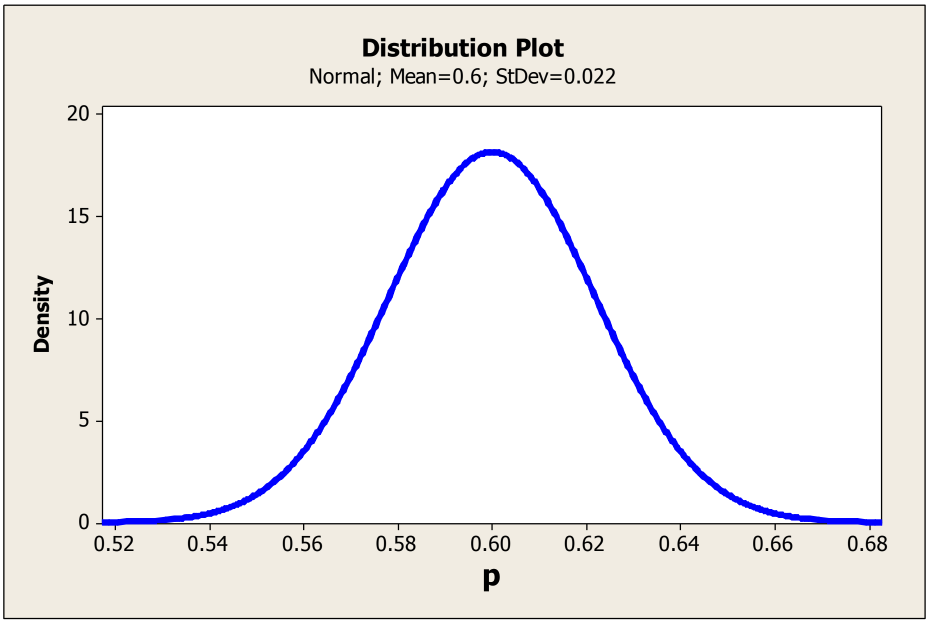
$$SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.05)(0.95)}{100}} = 0.0218$$

Example: In a survey, 500 mothers and fathers were asked about the importance of sports for boys and girls. Of the parents interviewed, 60% agreed that the genders are equal and should have equal opportunities to participate in sports.

Describe the sapling distribution of the sample proportion.

$$\hat{p} = 0.60$$

$$SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.60)(0.40)}{500}} = 0.022$$

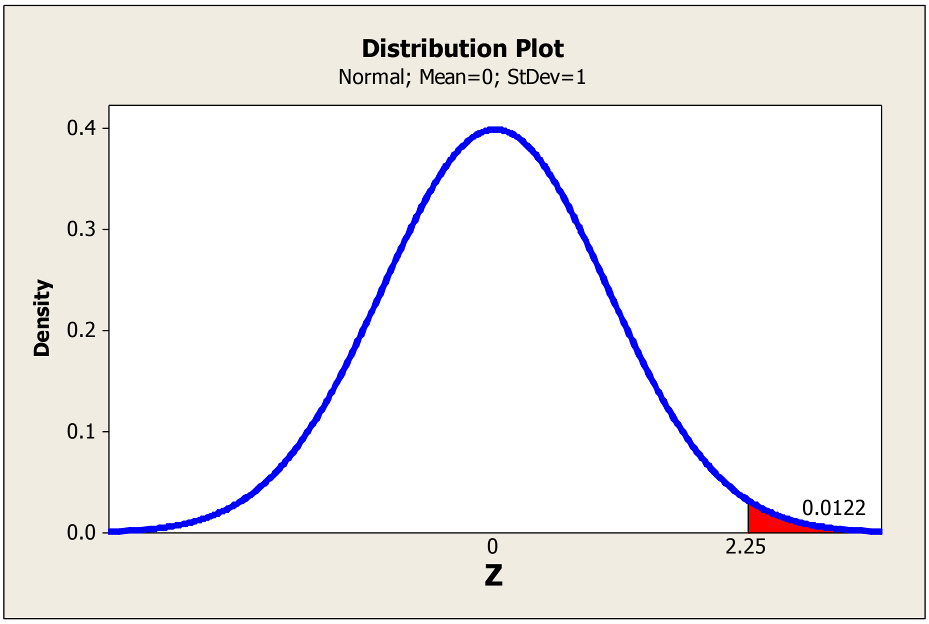


Suppose the proportion p of parents in the population is actually equal to 0.55. What is the probability of observing a sample proportion as large as or larger than the observed value $\hat{p} = 0.60$?

$$p = 0.55$$

$$SE = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.55)(0.45)}{500}} = 0.0222$$

$$P(\hat{p} > 0.60) = P(z > \frac{0.60 - 0.55}{0.0222})$$
$$= P(Z > 2.25) = 1 - 0.9878 = 0.0122$$



Exercise: Random samples of size n=75 were selected from a binomial population with p=0.30. Use the normal distribution to approximate the following probabilities:

$$P(\hat{p} \geq 0.40)$$

$$P(\hat{p} \leq 0.28)$$

$$P(0.25 \leq \hat{p} \leq 0.32)$$

A Control Chart for the Proportion Defective: The P Chart

To monitor a process for defective items, samples of size n are selected at periodic intervals and the sample proportion $\hat{p}$ is calculated. When the process is in control, $\hat{p}$ should fail into the interval $p \pm 3SE$, where p is the proportion of defectives in the population (or the process fraction defective) with standard error

$$SE = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

The process fraction defective is unknown but can be estimated by the average of the k sample proportions

$$\bar{p} = \frac{\sum_{i=1}^{k} \hat{p}_i}{k}$$

The standard error is estimated by

$$SE = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

The centerline for the p chart is located at

$$CL = \bar{p}$$

The upper and lower control limits are

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Example:

Proportions of Defectives in samples of n=1000.

Day	Sample Size	Number of Defective	Proportion
1	1000	18	0.018
2	1000	14	0.014
3	1000	15	0.015
4	1000	17	0.017
5	1000	21	0.021
6	1000	13	0.013
7	1000	12	0.012
8	1000	19	0.019
9	1000	23	0.023
10	1000	14	0.014
11	1000	16	0.016
12	1000	18	0.018
13	1000	21	0.021
14	1000	14	0.014
15	1000	17	0.017
16	1000	19	0.019
17	1000	19	0.019
18	1000	21	0.021
19	1000	21	0.021
20	1000	23	0.023

$$\overline{p} = \frac{\sum_{i=1}^{k} \hat{p}_i}{k} = \frac{0.018 + \ldots + 0.023}{20} = 0.01775$$

$$SE = \sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

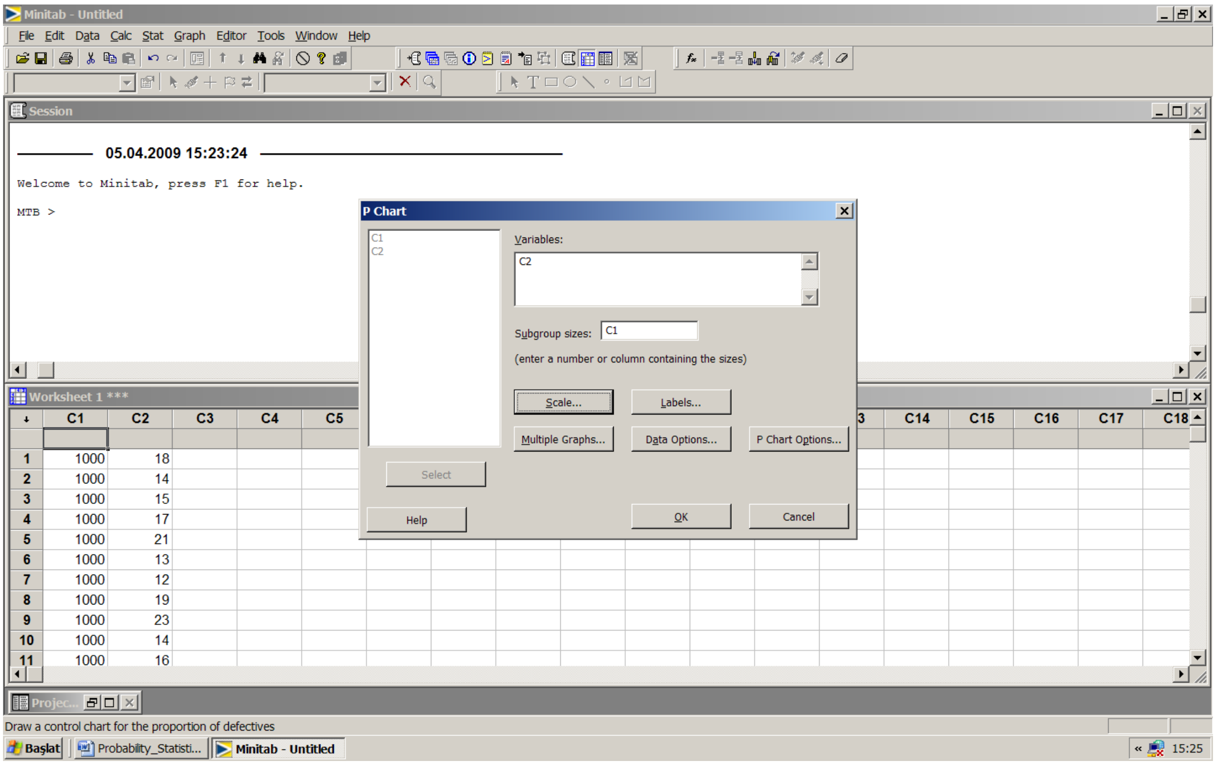
$$= \sqrt{\frac{0.01775(1-0.01775)}{1000}} = 0.00417552$$

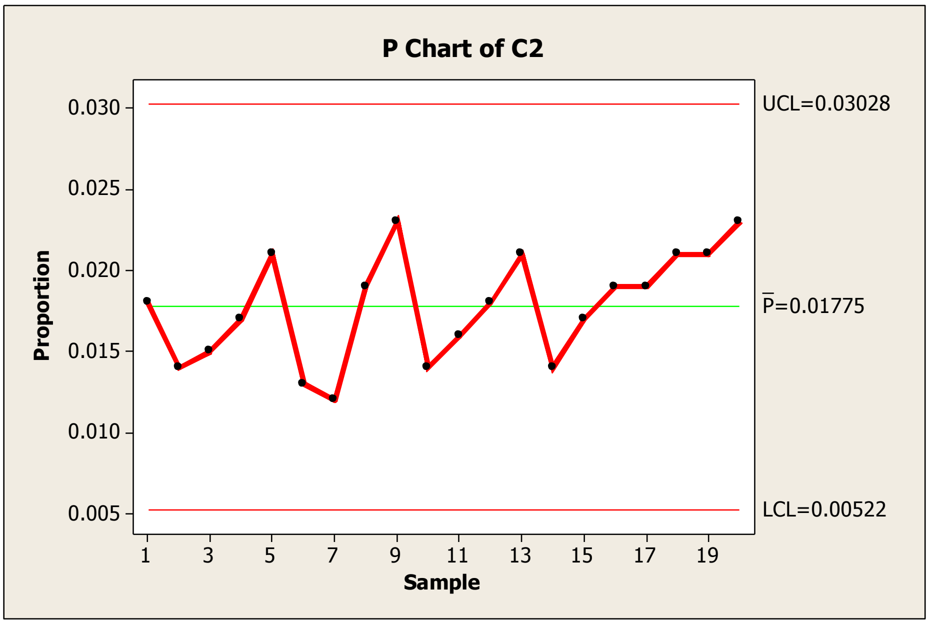
$$UCL = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

$$= 0.01775 + 3(0.00417552) = 0.03028$$

$$LCL = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

$$= 0.01775 - 3(0.00417552) = 0.00522$$


Minitab - Untitled
Session
05.04.2009 15:23:24
Welcome to Minitab, press F1 for help.
MTB >
P Chart
Variables:
C2
Subgroup sizes:
C1
(enter a number or column containing the sizes)
Scale...
Labels...
Multiple Graphs...
Data Options...
P Chart Options...
Select
Help
OK
Cancel
Worksheet 1 ***
Draw a control chart for the proportion of defectives


P Chart of C2
UCL=0.03028
P̄=0.01775
LCL=0.00522
Proportion
Sample