Discrete
Probability Distributions

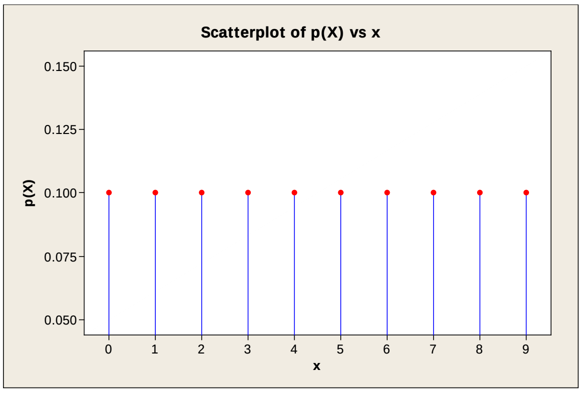
The Discrete Uniform (Integer) Distribution

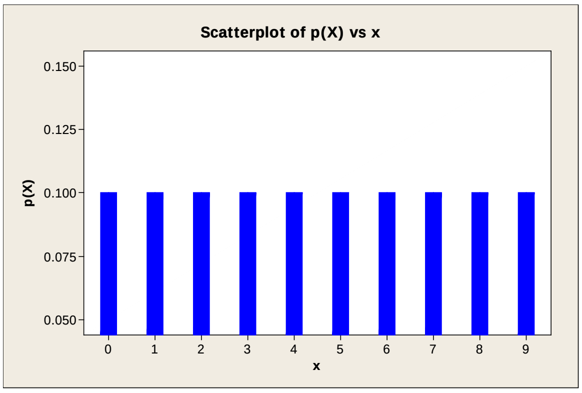
The simplest discrete random variable is one that assumes only a finite number of possible values, each with equal probability. A random variable X that assumes each of the values $x_1, x_2, \ldots, x_n$ with equal probability 1/n, is frequently of interest.

Definition

A random variable X has a discrete uniform distribution if each of the n values in its range, say $x_1, x_2, \ldots, x_n$ has the equal probability. Then,

$$P(x_i) = 1/n$$





Mean and Variance

Suppose X is a discrete uniform random variable on the consecutive integers

$$a, a+1, a+2, \ldots, b, \quad for\ a \le b$$

The mean of X is

$$\mu = E(X) = \frac{a+b}{2}$$

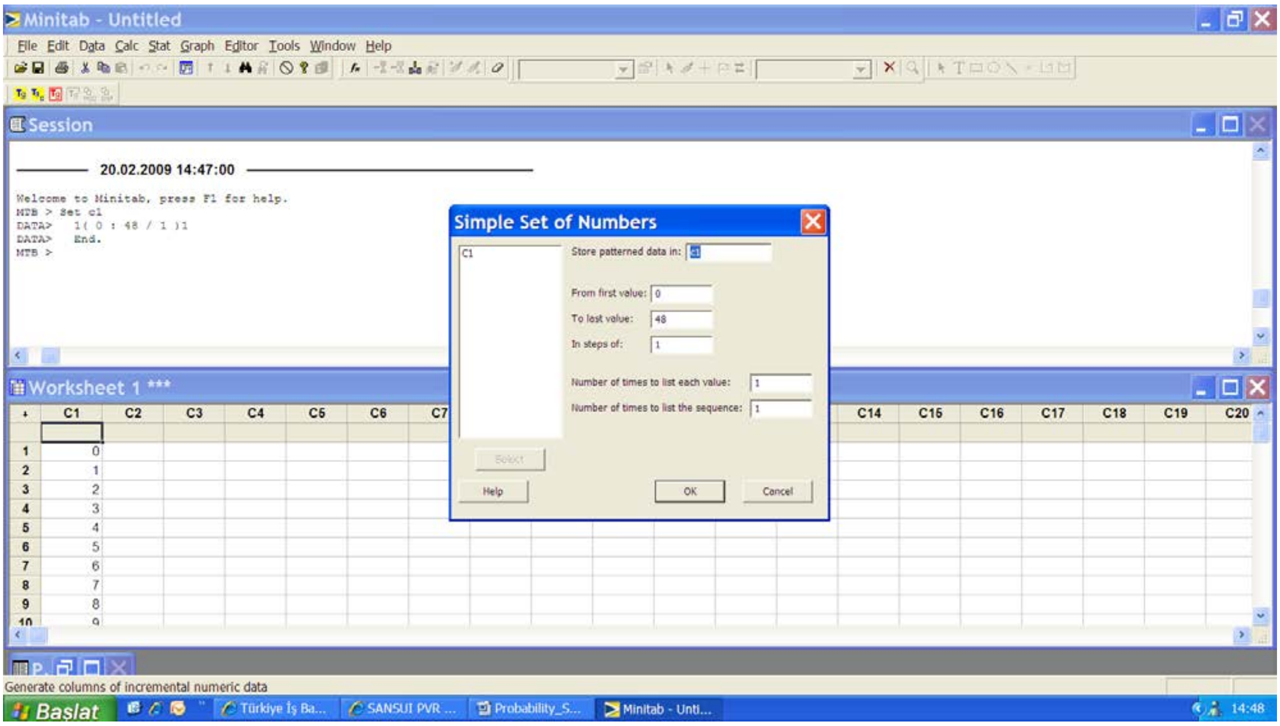
The variance of X is

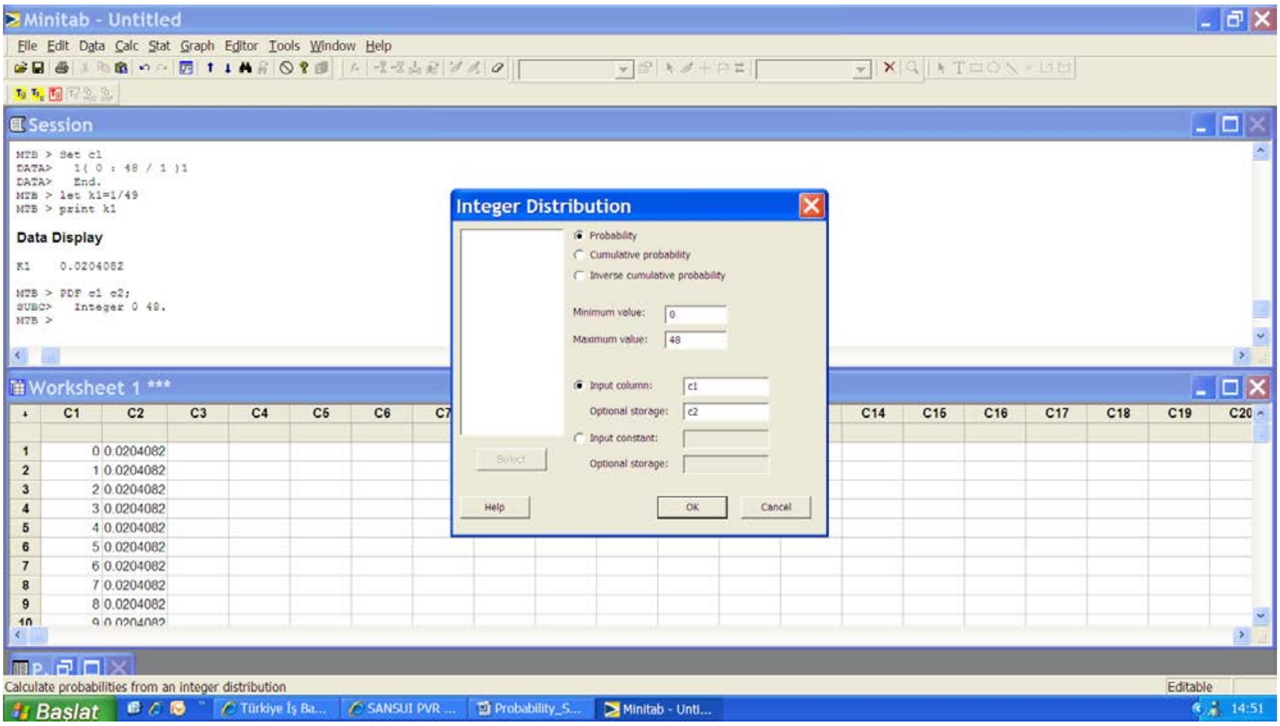
$$\sigma^2 = V(X) = \frac{(b-a+1)^2 - 1}{12}$$

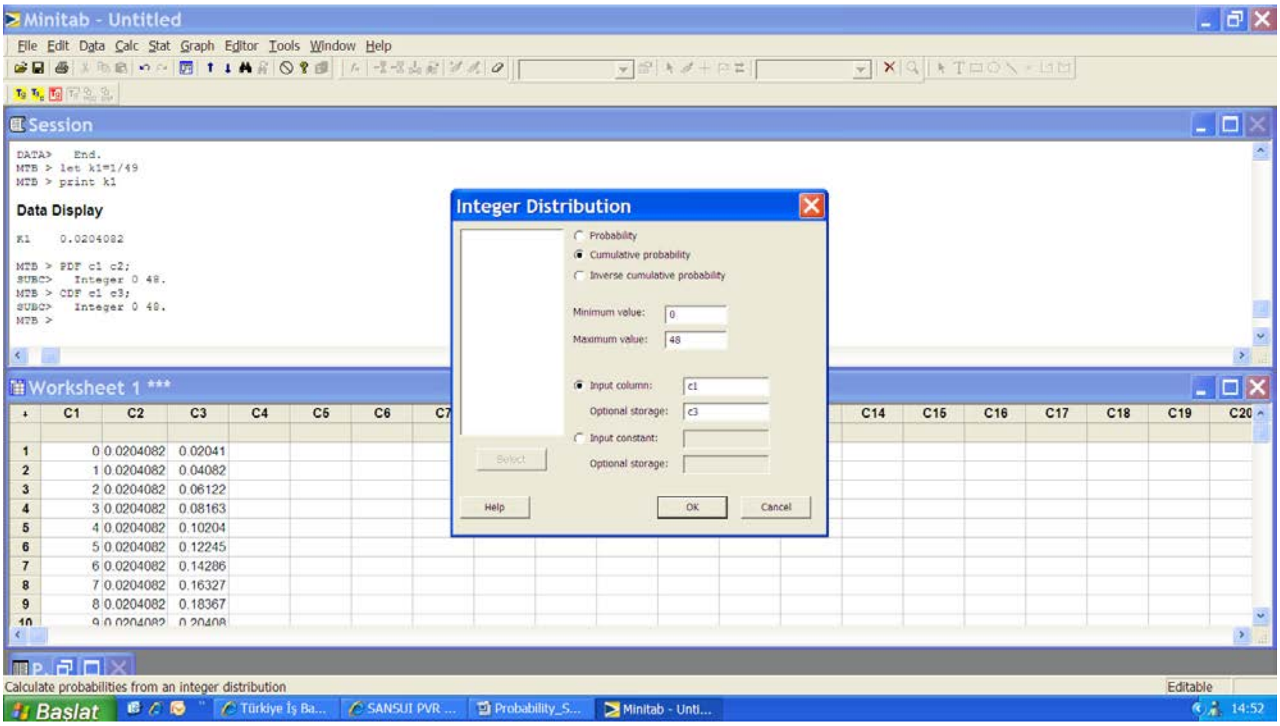
Example: A voice communication system for a business contains 48 external Lines. Let the random variable X denote the number of the voice lines that are in use a particular time. Assume that X is a discrete uniform random variable with a range of 0 to 48. Find mean and standard deviation of the random variable X.

$$\mu = E(X) = \frac{a+b}{2} = \frac{48+0}{2} = 24$$

$$\sigma = \left(\frac{(b-a+1)^2 - 1}{12} \right)^{1/2} = \{48-0+1)^2 - 1]/12\}^{1/2} = 14.14$$





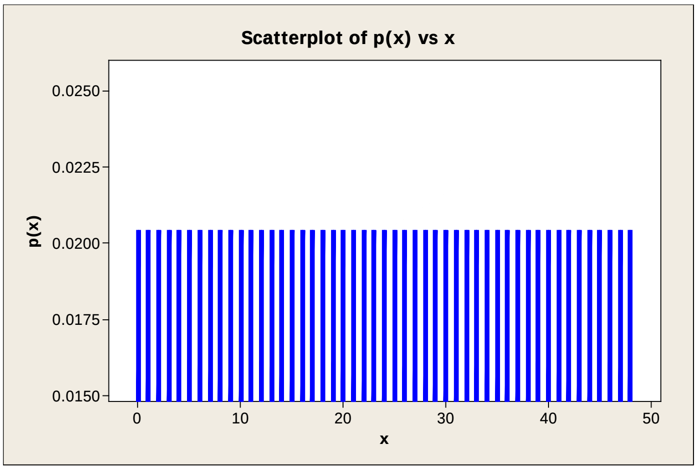


```
MTB > pdf c1 c2;
SUBC> integer 0 48.
MTB > cdf c1 c3;
SUBC> integer 0 48.
MTB > pdf c1 c2;
SUBC> integer 0 48.
MTB > cdf c1 c3;
SUBC> integer 0 48.
MTB > print c1-c3
```

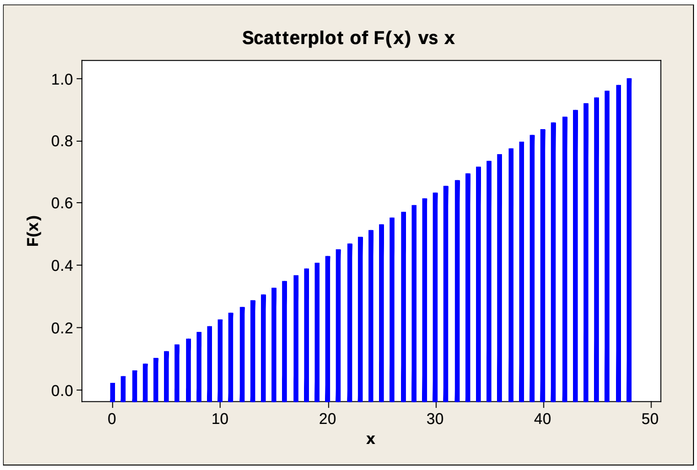
Data Display

Row	x	pdf	cdf
1	0	0.0204082	0.02041
2	1	0.0204082	0.04082
3	2	0.0204082	0.06122
4	3	0.0204082	0.08163
5	4	0.0204082	0.10204
6	5	0.0204082	0.12245
7	6	0.0204082	0.14286
8	7	0.0204082	0.16327
9	8	0.0204082	0.18367
10	9	0.0204082	0.20408
11	10	0.0204082	0.22449
12	11	0.0204082	0.24490
13	12	0.0204082	0.26531
14	13	0.0204082	0.28571
15	14	0.0204082	0.30612
16	15	0.0204082	0.32653
17	16	0.0204082	0.34694
18	17	0.0204082	0.36735
19	18	0.0204082	0.38776
20	19	0.0204082	0.40816
21	20	0.0204082	0.42857
22	21	0.0204082	0.44898
23	22	0.0204082	0.46939
24	23	0.0204082	0.48980
25	24	0.0204082	0.51020
26	25	0.0204082	0.53061
27	26	0.0204082	0.55102
28	27	0.0204082	0.57143
29	28	0.0204082	0.59184
30	29	0.0204082	0.61224
31	30	0.0204082	0.63265
32	31	0.0204082	0.65306
33	32	0.0204082	0.67347
34	33	0.0204082	0.69388
35	34	0.0204082	0.71429
36	35	0.0204082	0.73469
37	36	0.0204082	0.75510
38	37	0.0204082	0.77551
39	38	0.0204082	0.79592
40	39	0.0204082	0.81633
41	40	0.0204082	0.83673
42	41	0.0204082	0.85714
43	42	0.0204082	0.87755
44	43	0.0204082	0.89796
45	44	0.0204082	0.91837
46	45	0.0204082	0.93878
47	46	0.0204082	0.95918
48	47	0.0204082	0.97959
49	48	0.0204082	1.00000

```
MTB >
```



Probability Distribution Function



Cumulative Distribution Function

The cumulative distribution function is

$$F(x) = P[X \leq x] = \sum_{x_k \leq x} p(x_k)$$

$$F(x;a,b) = \frac{x-a+1}{b-a+1}$$

and is also useful to be able to compute. Note that

$$F(x_k) - F(x_{k-1}) = p(x_k)$$

```
MTB > cdf c5 c6;
SUBC> integer 0 48.
```

6	**0.142857**
7	**0.163265**
8	**0.183673**
9	**0.204082**
10	**0.224490**
11	**0.244898**
12	**0.265306**
13	**0.285714**

$$P(6 \le X \le 13) = CDF(13) - CDF(6) = 0.285714 - 0.142857 = 0.142857$$

$$P(X > 13) = ?$$

$$P(X \le 6) = ?$$

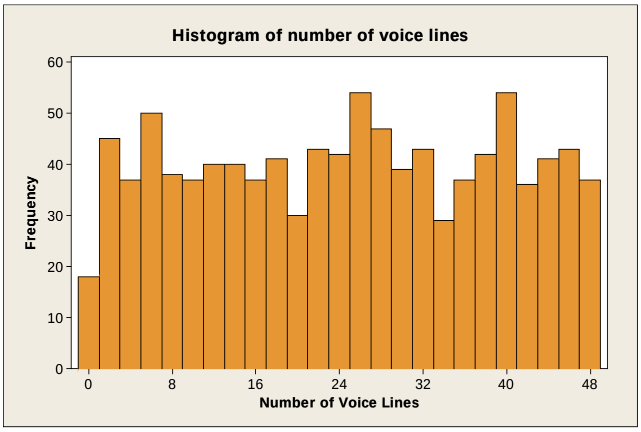
$$P(16 \le X \le 35) = ?$$

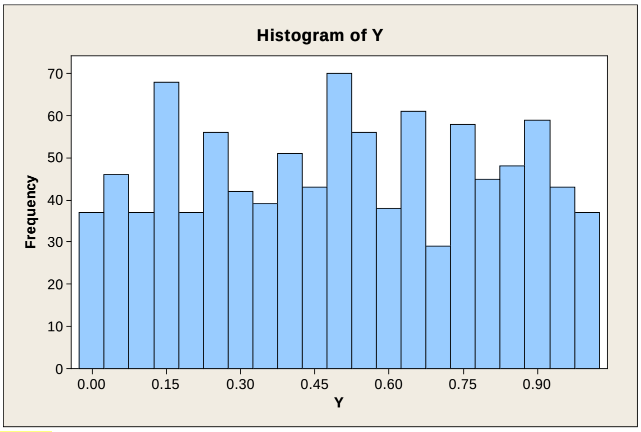
Let the random variable Y denote the proportion of the 48 voice lines that are in use at a particular time, and X denotes the number of lines that are in use at a particular time. Then Y=X/48. Therefore,

$$E(Y) = E(X)/48 = 24/48 = 0.5$$

$$V(Y) = V(X)/48^2 = (14.14)^2/48^2 = 0.087$$

$$STD(Y) = \sqrt{0.087} = 0.29496$$






```
MTB > random 1000 c1;
SUBC> integer 0 48.
MTB > hist c1
Histogram of C1
MTB > let Y=X/48
MTB > DESC X Y
 Descriptive Statistics: X; Y
Variable     N  N*     Mean  SE Mean    StDev  Minimum       Q1   Median
X         1000   0   24.047    0.447   14.126    0.000   12.000   25.000
Y         1000   0  0.50098  0.00931  0.29429  0.00000  0.25000  0.52083

Variable        Q3  Maximum
X           37.000   48.000
Y          0.77083  1.00000
```

The Binomial Probability Distribution

Consider the following random experiments and random variables:

- **Flip a coin 10 times. Let X=number of heads obtained.**
- **A worn machine tool produces 1% defective parts. Let X=number of defective parts in the next 25 parts produced.**
- **Each sample of air has a 10% chance of containing a particular rare molecule. Let X=the number of air samples that contain the rare molecule in the next 18 samples analyzed.**
- **Of all bits transmitted through a digital transmission channel, 10% are received in error. Let X=the umber of bits in error in the next five bits transmitted.**
- **A multiple choice test contains 10 questions, each with four choices, and you guess at each question. Let X= the number of questions answered correctly.**
- **In the next 20 births at a hospital, let X=the number of female births.**
- **Of all patients suffering a particular illness, 30% experience improvement from a particular medication. In the next 100 patients administered the medication, let X= the number of patients who experience improvement.**

These examples illustrate that a general probability model that includes these experiments as particular cases would be very useful.

Definition:

A trial with only <u>two possible outcomes</u> is used so frequently as a building block of random experiment random that is called a <u>Bernoulli trial.</u> It is usually assumed that the trials that constitute random experiment are <u>independent</u>.

$$f(k;p) = \begin{cases} p & \text{if } k = 1, \\ 1-p & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Example:

The chance that a **<u>bit</u>** **<u>transmitted through a digital transmission channel</u>** is received in error is **0.1**. Also assume that the transmission trials are independent. **Let X=the number of bits in error in the next four bits transmitted.**

Determine P(X=2)

Determine E(X) and V(X)

Let the letter E denotes a <u>bit in error</u>.

Let the letter O denotes that the <u>bit is okay</u>.

We can represent the outcomes of this experiment as a list of four letters hat indicate the bits that are in error and those that are okay. **For example OEOE indicates that the <u>second and fourth bits are in error</u> and the other two bits are okay. The corresponding values for x are**

Outcome	x
OOOO	0
OOOE	1
OOEO	1
OOEE	2
OEOO	1
OEOE	2
OEEO	2
OEEE	3
EOOO	1
EOOE	2
EOEO	2
EOEE	3
EEOO	2
EEOE	3
EEEO	3
EEEE	4

The event that X=2 consist of six outcomes:

$$\{EEOO, EOEO, EOOE, OEEO, OEOE, OOEE\}$$

Using the assumption that the trials are independent, the probability of $\{EEOO\}$ **is**

$$P(EEOO) = P(E)P(E)P(O)P(O) = (0.1)^2(0.9)^2 = 0.0081$$

Also, any one of the six mutually outcomes for which X=2 as the same probability of occurring. Therefore

$$P(X = 2) = (6)(0.0081) = 0.0486$$

Definition:

A random experiment consist of n Bernoulli trials such that

1. ***The trials are independent***
2. ***Each trial results in only two possible outcomes, "success" and "failure"***
3. ***The probability of success in each trial, denoted as p, remains constant.***

The random variable X that equals the number of trials that result in a success has a binomial random variable with parameter 0 < p < 1, the probability of k success in n trials is

$$P(X=k)=C_k^n p^k (1-p)^{n-k}, k=0,1,\ldots,n.$$

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The name of the distribution is obtained from the binomial expansion.

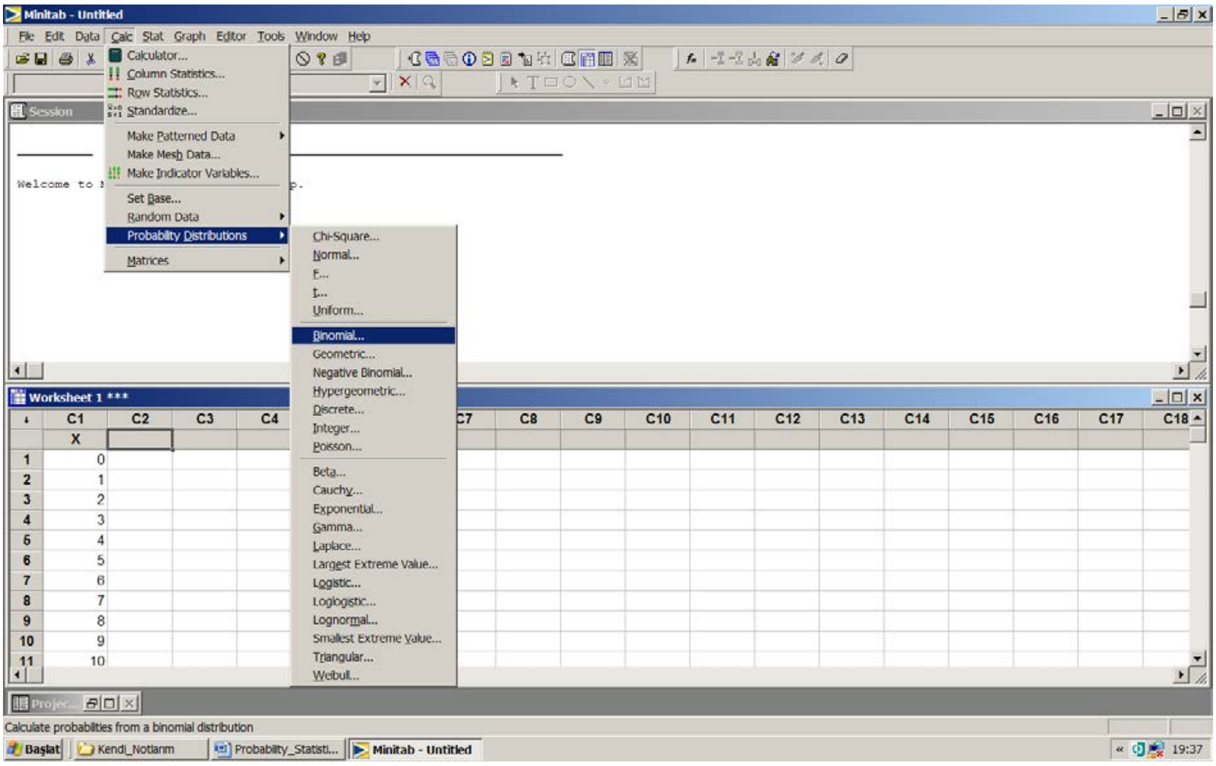
$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

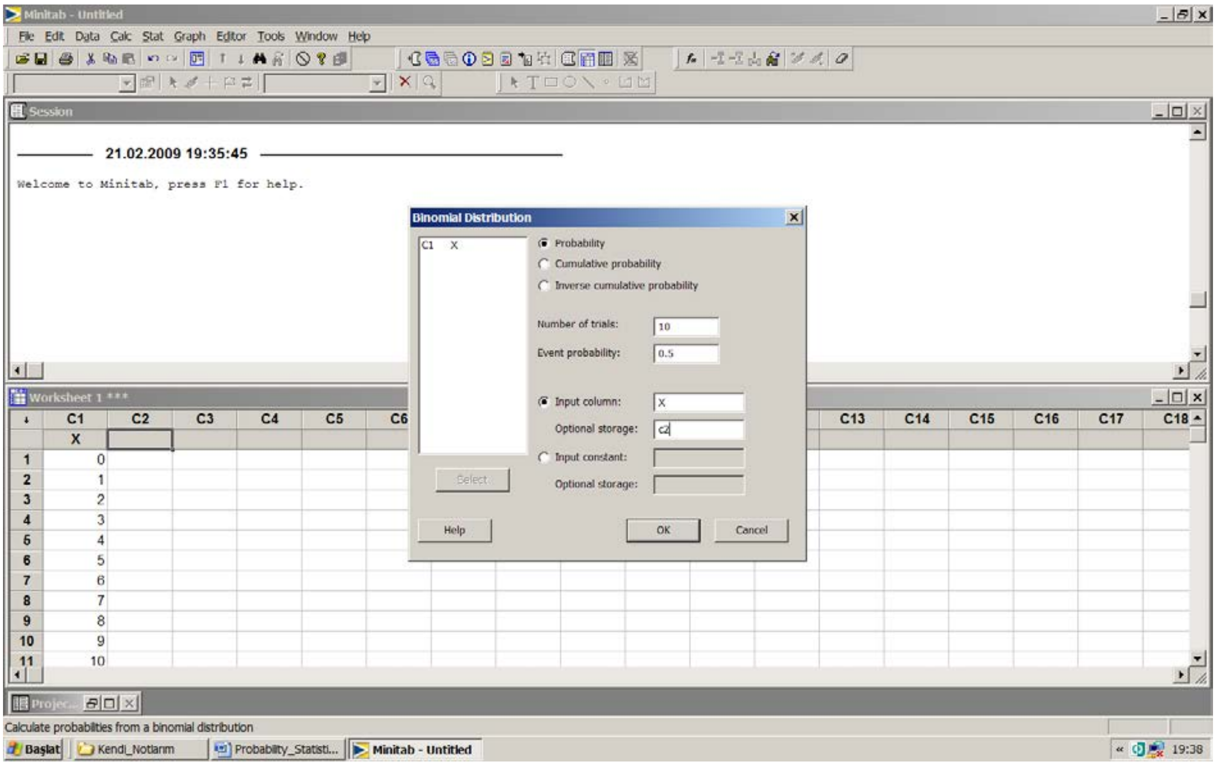
Before we obtain

$$P(X = 2) = (6)(0.0081) = 0.0486$$

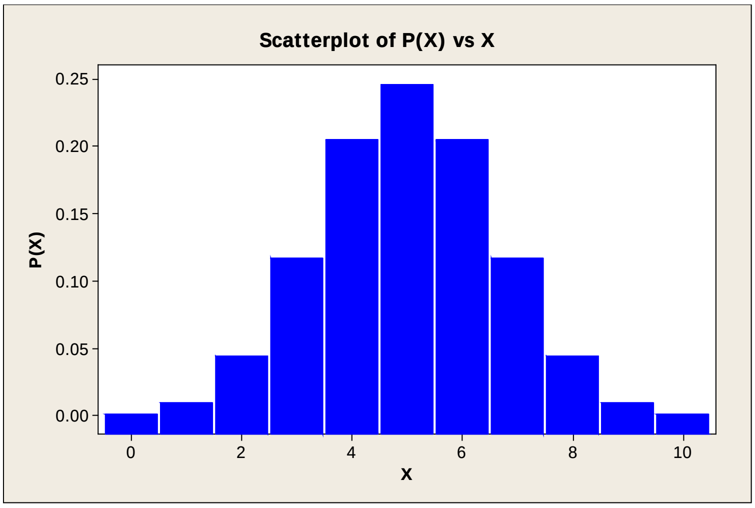
$$P(X = k) = C_k^n p^k (1 - p)^{n-k}, k = 0,1,\ldots,n.$$

$$P(X = 2) = C_2^4 (0.1)^2 (1 - 0.1)^2 = 0.0486$$

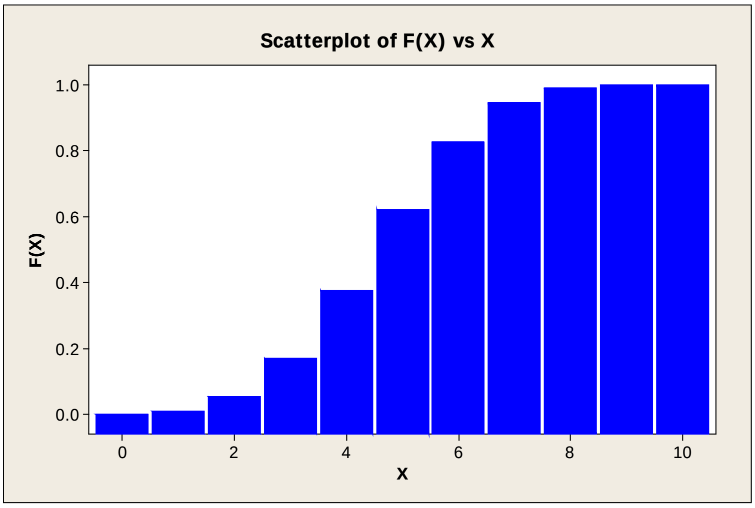

Minitab - Untitled
File Edit Data Calc Stat Graph Editor Tools Window Help
Calculator...
Column Statistics...
Row Statistics...
Standardize...
Make Patterned Data
Make Mesh Data...
Make Indicator Variables...
Set Base...
Random Data
Probability Distributions
Matrices
Chi-Square...
Normal...
F...
t...
Uniform...
Binomial...
Geometric...
Negative Binomial...
Hypergeometric...
Discrete...
Integer...
Poisson...
Beta...
Cauchy...
Exponential...
Gamma...
Laplace...
Largest Extreme Value...
Logistic...
Loglogistic...
Lognormal...
Smallest Extreme Value...
Triangular...
Weibull...
Worksheet 1 ***
Calculate probabilities from a binomial distribution
Başlat
Kendi_Notlarım
Minitab - Untitled
19:37


Minitab - Untitled
File Edit Data Calc Stat Graph Editor Tools Window Help
Session
21.02.2009 19:35:45
Welcome to Minitab, press F1 for help.
Binomial Distribution
C1 X
Probability
Cumulative probability
Inverse cumulative probability
Number of trials: 10
Event probability: 0.5
Input column: X
Optional storage: c2
Input constant:
Optional storage:
Select
Help
OK
Cancel
Worksheet 1 ***
Calculate probabilities from a binomial distribution
Başlat
Kendi_Notlarım
Minitab - Untitled
19:38

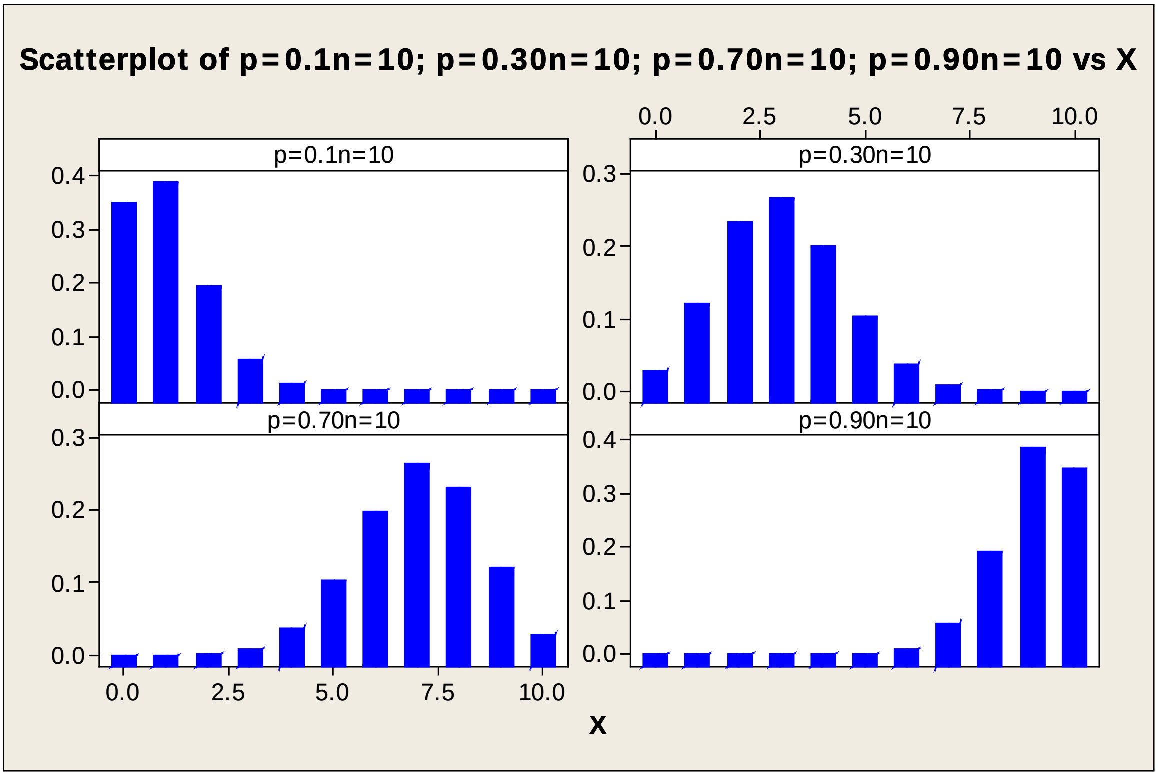
X	P(X)	F(X)
0	0.000977	0.00098
1	0.009766	0.01074
2	0.043945	0.05469
3	0.117188	0.17187
4	0.205078	0.37695
5	0.246094	0.62305
6	0.205078	0.82813
7	0.117188	0.94531
8	0.043945	0.98926
9	0.009766	0.99902
10	0.000977	1.00000



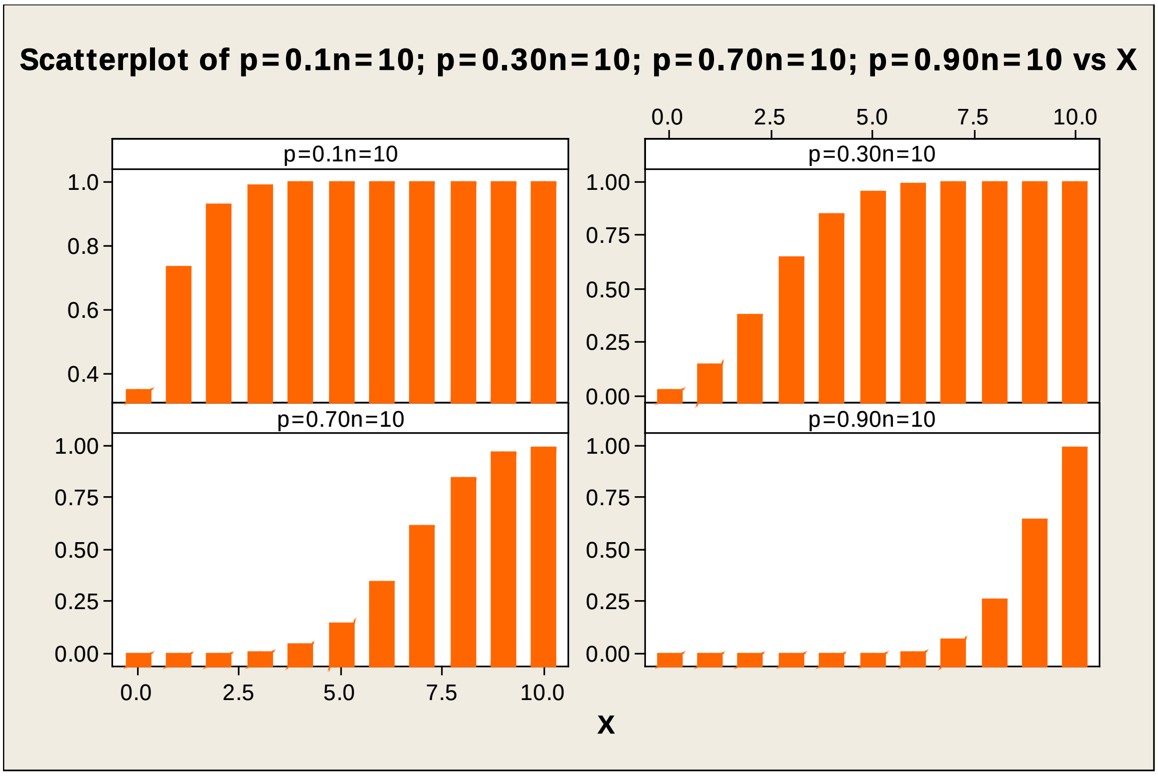
Binomial Distribution for n=10 p=0.5



Binomial Cumulative Distribution for n=10 p=0.5



Binomial Distribution for n=10 and different p.



Binomial Cumulative Distribution for n=10 and different p.

Example:

Find P(X=2) for a binomial random variable with n=10 and p=0.1.

P(X=2) is the probability of observing 2 successes and 8 failures in a sequence of 10 trials

Two successes first

S,S,F,F,F,F,F,F,F,F has probability $p^2(1-p)^8$

However many other sequences also result in X=2 successes.

$$P(X=2)=C_2^{10}0.1^2(1-0.1)^{10-2}=0.1937$$

```
MTB > pdf 2;
SUBC> Binomial 10 0.1.
```

Probability Density Function

```
Binomial with n = 10 and p = 0.1

x  P( X = x )
2    0.193710
```

Binomial probabilities and cumulative binomial probabilities for n=10 and p=0.1.

X	P(X)	F(X)
0	0.348678	0.34868
1	0.387420	0.73610
2	0.193710	0.92981
3	0.057396	0.98720
4	0.011160	0.99837
5	0.001488	0.99985
6	0.000138	0.99999
7	0.000009	1.00000
8	0.000000	1.00000
9	0.000000	1.00000
10	0.000000	1.00000

Find the probabilities of these events

1. Exactly three successes
2. Three or more successes
3. At most two successes

1. $P(X=3)= 0.057396$
2. $P(X\geq 3)=1-P(X<3)=1-P(X\leq 2)$

   $=1-0.92981=0.07019$
3. $P(X\leq 2)=0.92981$

Mean and Variance for the Binomial Random Variable

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

$$\mu = E(X) = np$$

and

$$\sigma^2 = V(X) = np(1-p)$$

Example:

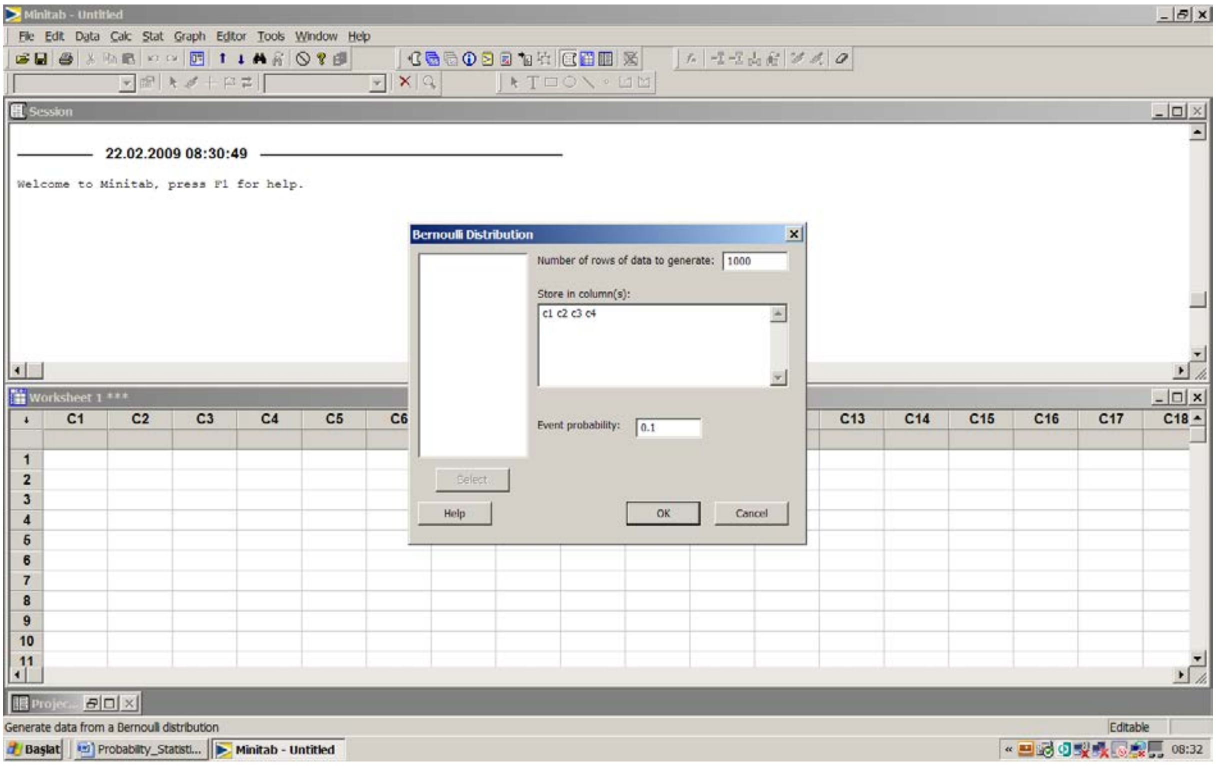
For the number of transmitted bits received in error, n=4 and p=0.1, so

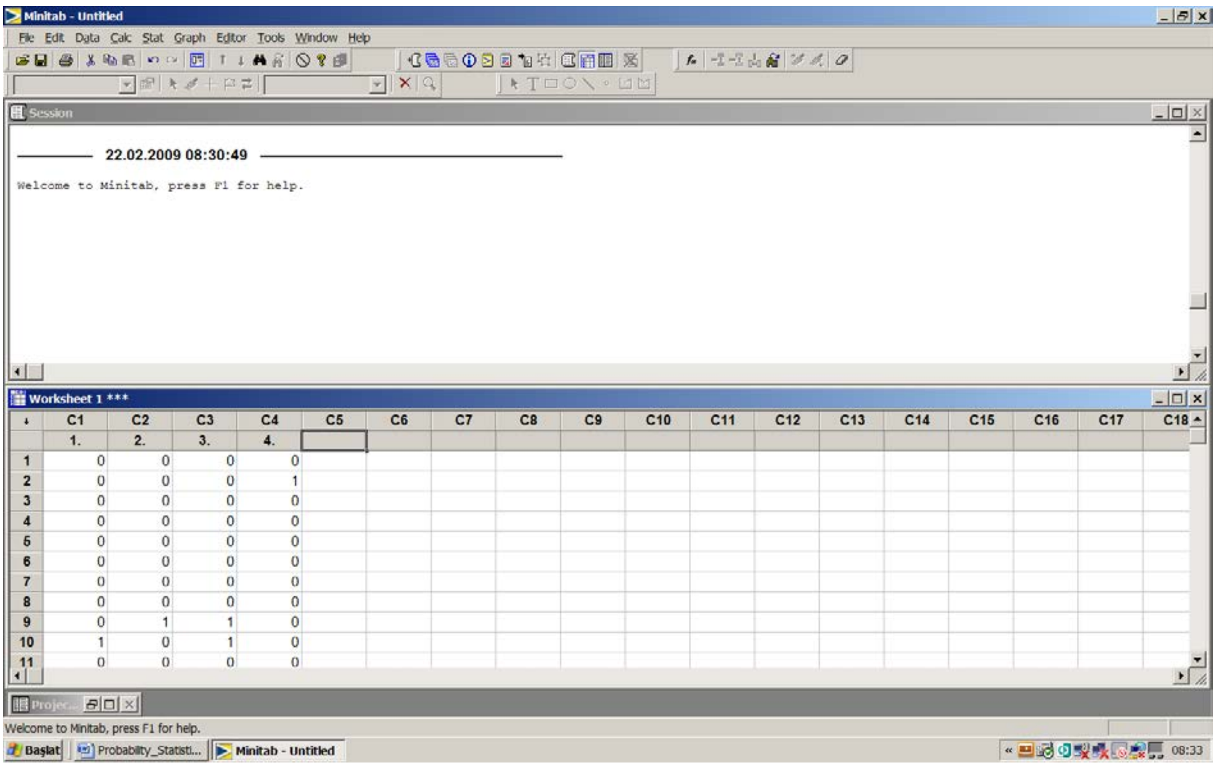
$$\mu = E(X) = np = (4)(0.1) = 0.4$$

$$\sigma^2 = V(X) = np(1-p) = (4)(0.1)(0.9) = 0.36$$

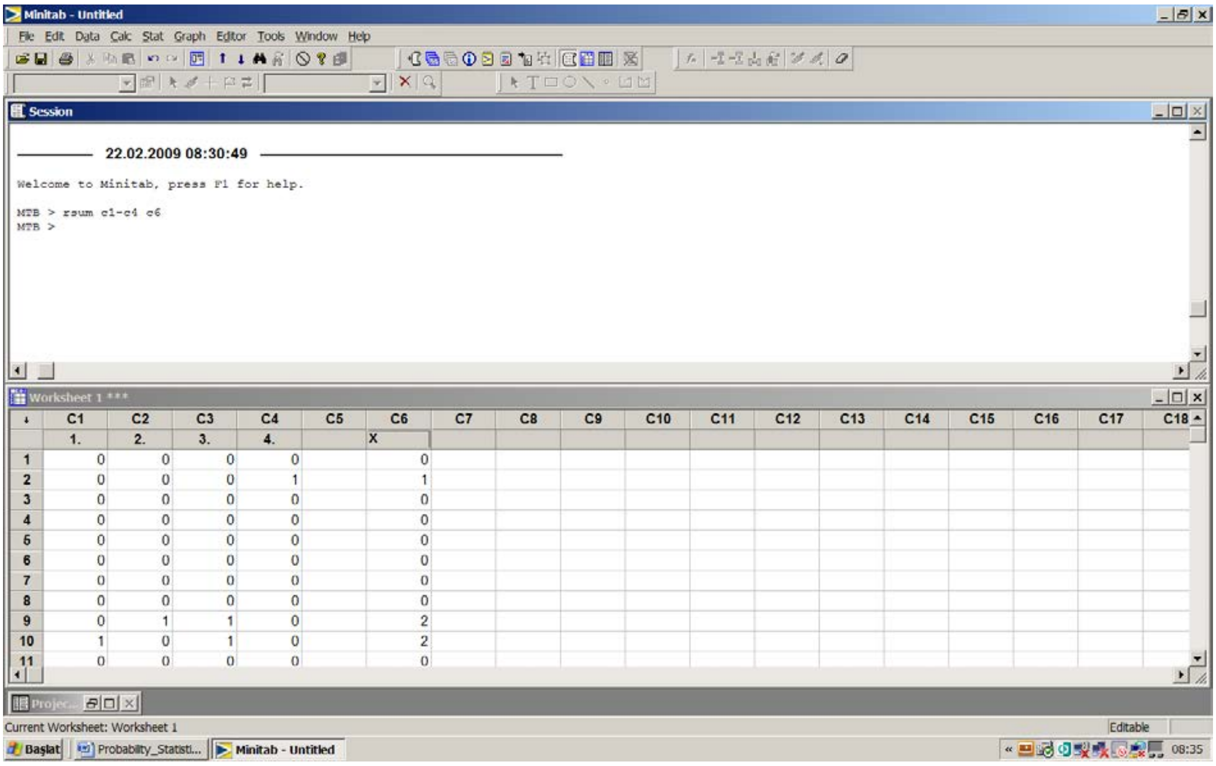
Example:

First generate four Bernoulli trials with probability 0.1.





Sum results in the different column.



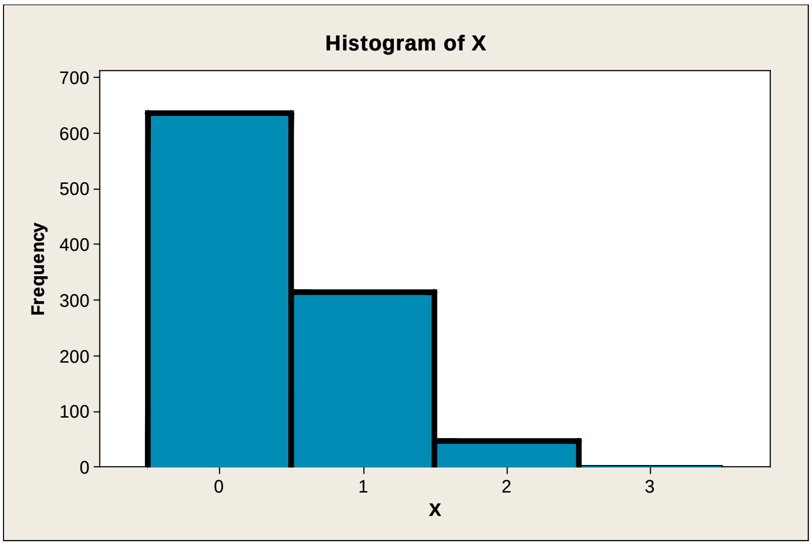

```
MTB > tall c6;
SUBC> all.
```

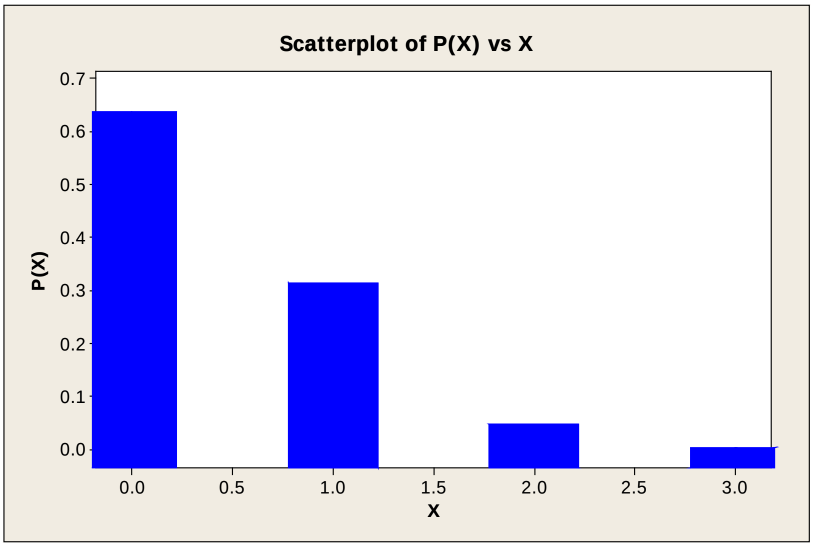
Tally for Discrete Variables: X

```
  X   Count   CumCnt   Percent   CumPct
  0     636      636     63.60    63.60
  1     314      950     31.40    95.00
  2      47      997      4.70    99.70
  3       3     1000      0.30   100.00
N=     1000
```

Compare this with the result before.

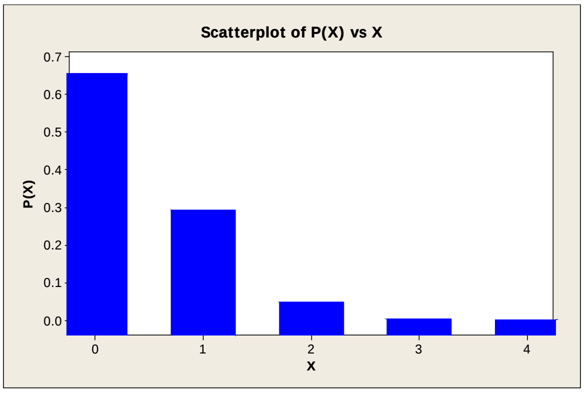
Histogram of the results

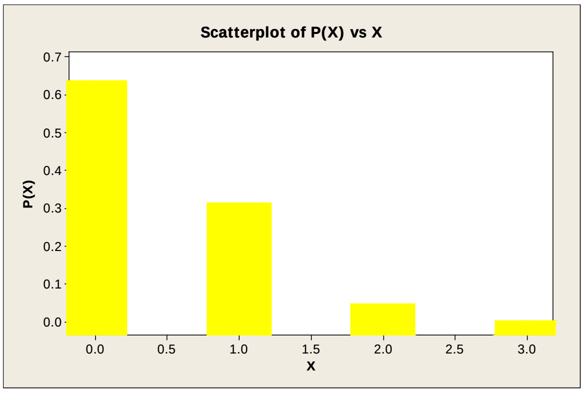




Theoretical results

X	P(X)	F(X)
0	0.6561	0.6561
1	0.2916	0.9477
2	0.0486	0.9963
3	0.0036	0.9999
4	0.0001	1.0000





Theoretical **Experimental**

Experimental mean and variance

```
MTB > desc c6

Descriptive Statistics: x

Variable      N  N*    Mean  SE Mean    StDev  Minimum       Q1
Median       Q3
x          1000   0  0.4170   0.0189   0.5962   0.0000   0.0000
0.0000   1.0000

Variable  Maximum
x          3.0000
```

Theoretical mean and variance (obtained before)

$$\mu = E(X) = np = (4)(0.1) = 0.4$$

$$\sigma^2 = V(X) = np(1-p) = (4)(0.1)(0.9) = 0.36$$

$$\sigma = 0.6$$