Discrete Random Variables and Their Probability Distributions

Counting Techniques

Multiplication rule (for counting techniques)

If an operation can be described as a sequence of k steps, and

- ***If the number of ways of completing step 1 is n_1, and***
- ***If the number of ways of completing step 2 is n_2, for each way of completing step 1, and***
- ***If the number of ways of completing step 3 is n_3, for each way of completing step 2, and so forth, the total number of ways of completing operations is***

$$n_1 * n_2 * \ldots . * n_k.$$

if n_1=4, n_2=3 and n_3=3

From the multiplication rule (4)(3)(3)= 36 different designs are possible.

Permutation

The number of permutation of <u>n different elements</u> is n! where

$$n! = n(n-1)(n-2)(n-3)\ldots\ldots(2)(1)$$

Consider a set of elements, such as S={a, b, c}. A <u>permutation</u> of the elements is an ordered sequence of the elements. For example

abc, acb, bac, bca, cab, cba

are all of the permutations of the element of S.

Permutation of Subsets

The number of <u>permutations</u> of subsets of r element selected from a set of n different elements is

$$P_r^n = \frac{n!}{(n-r)!}$$

Consider a set of elements, such as S={a, b, c, d} and r=2 the permutation of subsets are

$$P_2^4 = \frac{4!}{(4-2)!} = 12$$

ab, ac, ad,

ba, bc, bd,

ca, cb, cd,

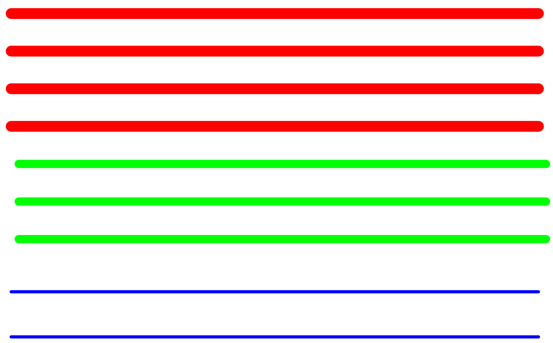
da, db, dc.

Permutation of Similar Objects

The number of <u>permutations</u> of subsets of $n=n_1+n_2+\ldots+n_r$ objects of which n_1 are of one type, n_2 are of a second type,..., and n_r are of an rth type is

$$\frac{n!}{n_1!\,n_2!\ldots n_r!}$$

Example: Bar Codes A part is labeled by printing with <u>four thick lines</u>, <u>three medium lines</u> and <u>two thin lines.</u> If each ordering of the nine lines represents a different label , how many different labels can be generated by using this scheme?



The number of possible part labels is

$$\frac{9!}{4!\ 3!\ 2!}=1260$$

Combinations

The number of combinations, subsets of size r that can be selected from a set of n elements, is denoted as

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

The number combinations and the number of permutations are related

$$C_r^n = \frac{P_r^n}{r!}$$

Consider a set of elements, such as S={a, b, c, d} and r=2 the combination subsets are

$$C_2^4 = \frac{4!}{2!(4-2)!} = 6$$

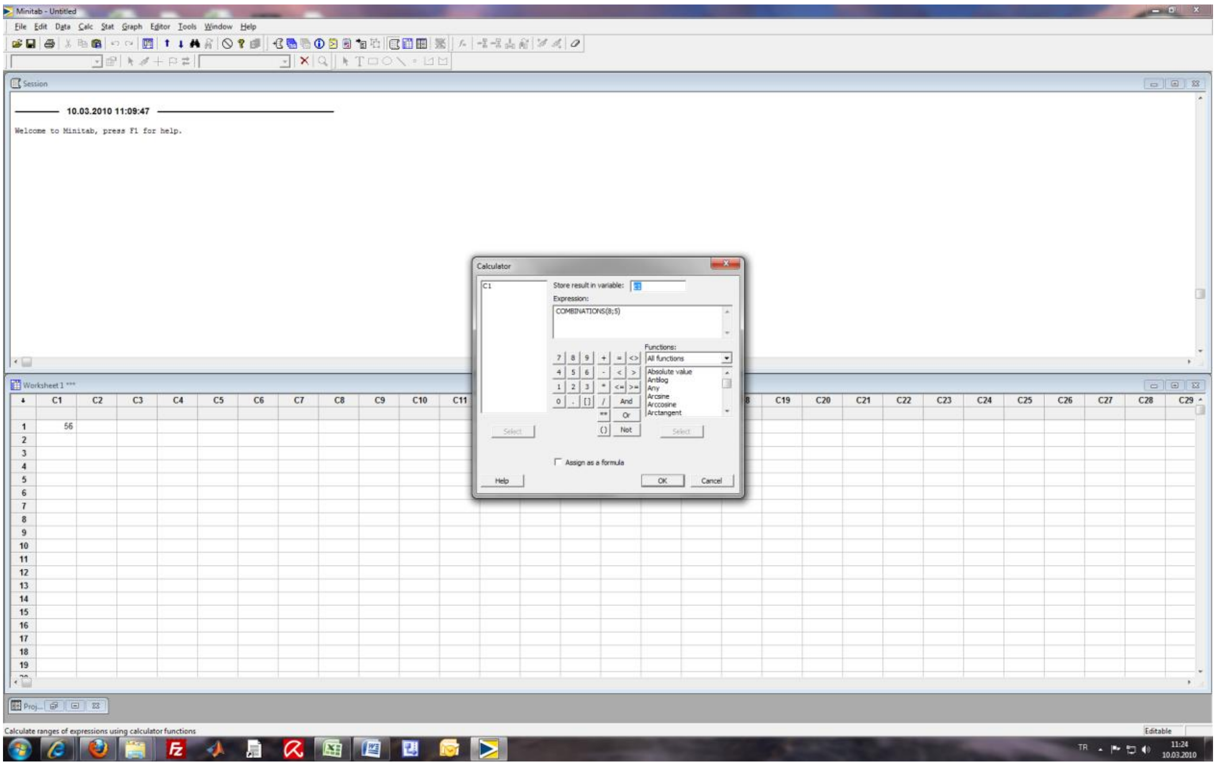
ab, ac, ad, bc, bd, cd

Example: A printed circuit board has **eight different locations** in which a component can be placed. **If five identical components are to be placed on the board, how many different designs are possible?**

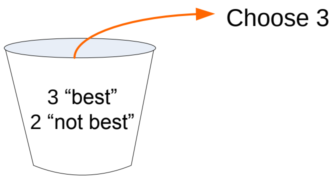
The number of possible design is

$$\frac{8!}{5!\,(8-5)!} = \frac{8!}{5!\ \ 3!} = 56.$$

In Minitab, you can use the Calculator to find the number of combinations of n things taken k at a time.



Example: **Five** manufacturers produce a certain electronic device, whose quality varies from manufacturer to manufacturer. If you were to select **three manufacturers at random,** what is the chance that the selection would contain **exactly two of the best three**?



The total number of simple events N can be counted as the number of ways to choose three of the five manufacturers, or $N = C_3^5 = \frac{5!}{3!\ 2!} = 10$.

Since the manufacturers are selected at random, any of these 10 simple events will be **equally likely with probability 1/10.** But how many of these simple events result in the event

A: Exactly two of the “best” three.

We can count n_A, the number of events in A, in two steps because event A will occur when we select two of the “best” three and one of the two “not best”.

There are $C_2^3 = \frac{3!}{2!\ 1!} = 3$ ways to accomplish the **first stage** and, $C_1^2 = \frac{2!}{1!\ 1!} = 2$ ways to accomplish the **second stage.** Applying **multiplication rule**, we find there are $n_A=(3)(2)=6$ of the simple events in **event A** and

$$P(A) = \frac{n_a}{N} = \frac{6}{10} = 0.6$$ >>> **?** ***Distribution***

$$P(X=k)=\frac{C_k^M C_{n-k}^{N-M}}{C_n^N}$$

Probability Distributions

We learned how to construct the *relative frequency distribution* for a set of numerical measurements on a variable x. The distribution gave this information about x:

- ***What values of x occurred***
- ***How often each value of x occurred***

Definition: ***The probability distribution for a discrete random variable is a***

- ***Formula,***
- ***Table, or***
- ***Graph***

that gives the possible values of x, and the probability p(x) associated with each value of x.

Requirements for a Discrete Probability Distribution

- $0 \leq p(x) \leq 1$

Each probability must lie between 0 and 1.

- $\sum p(x) = 1$

The sum of the probabilities for all simple events in space S equals 1.

Example: Toss **two fair coins** and let x equal **the number of heads** observed. Find the probability distribution for x.

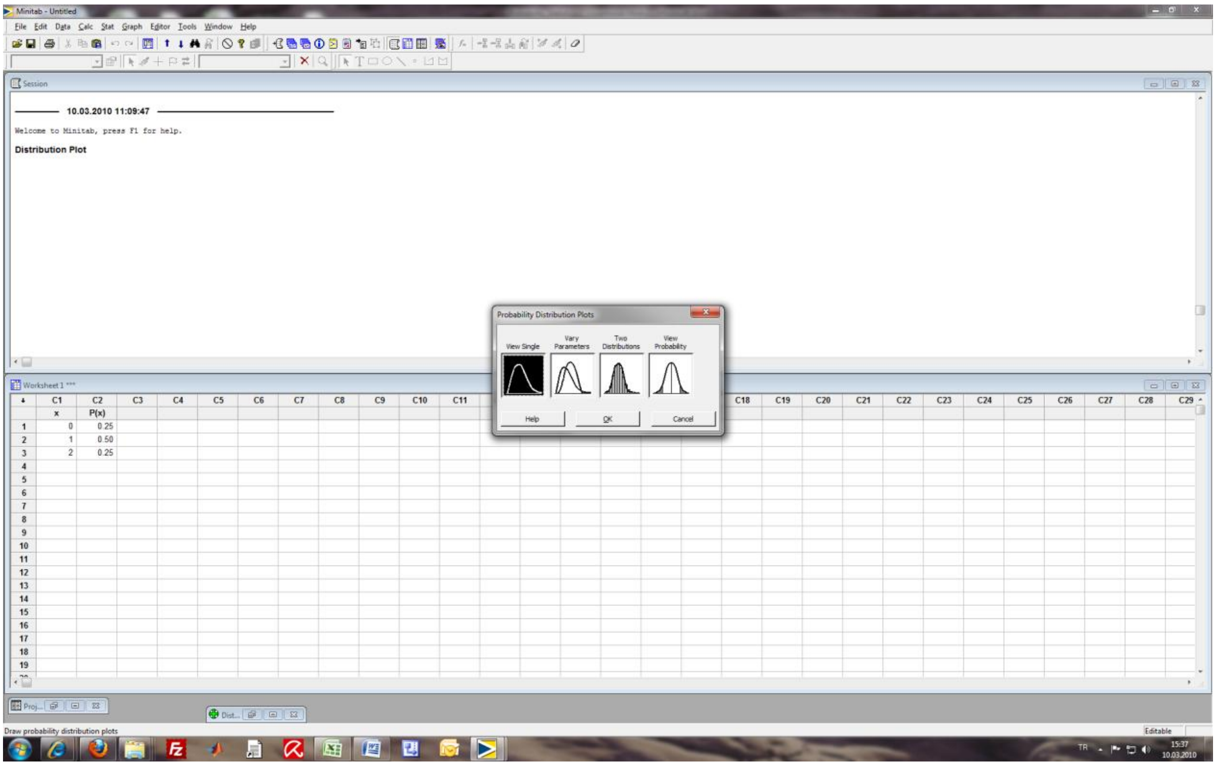
Simple Events and Probabilities

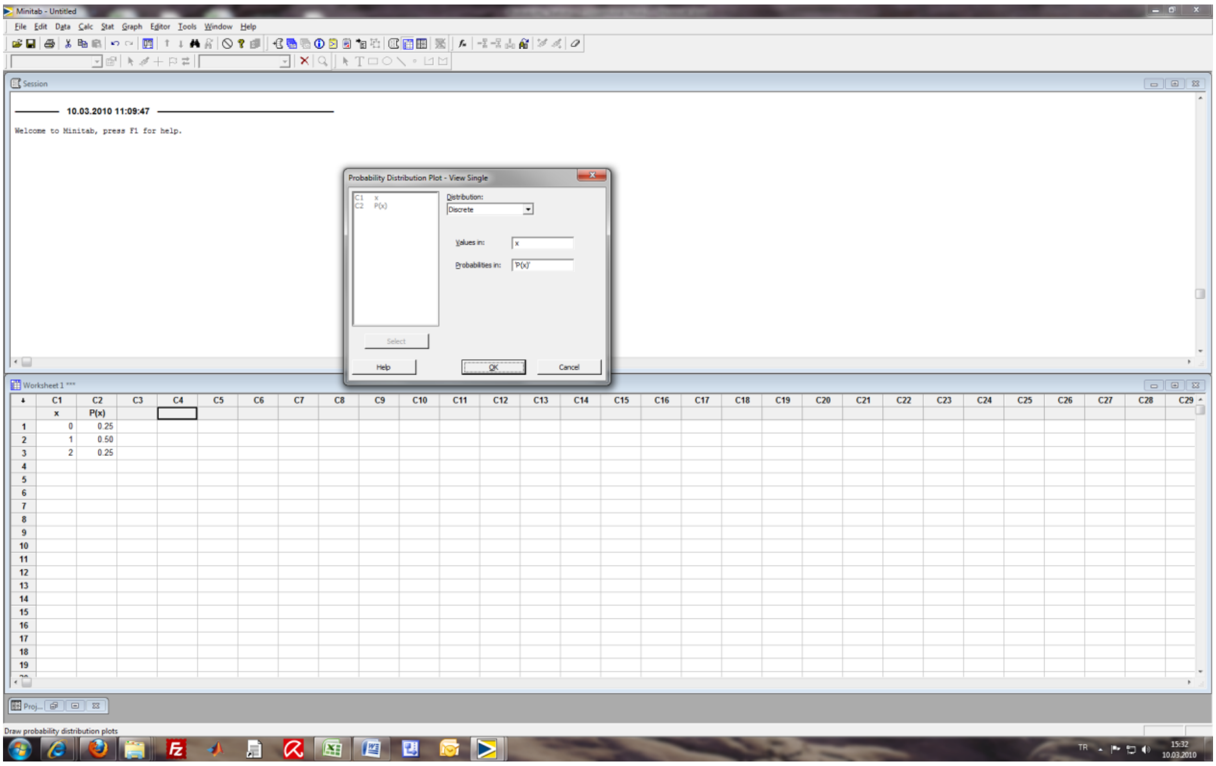
Simple Event	Coin 1	Coin 2	P(E)	x
E_1	H	H	¼	2
E_2	H	T	¼	1
E_3	T	H	¼	1
E_4	T	T	¼	0

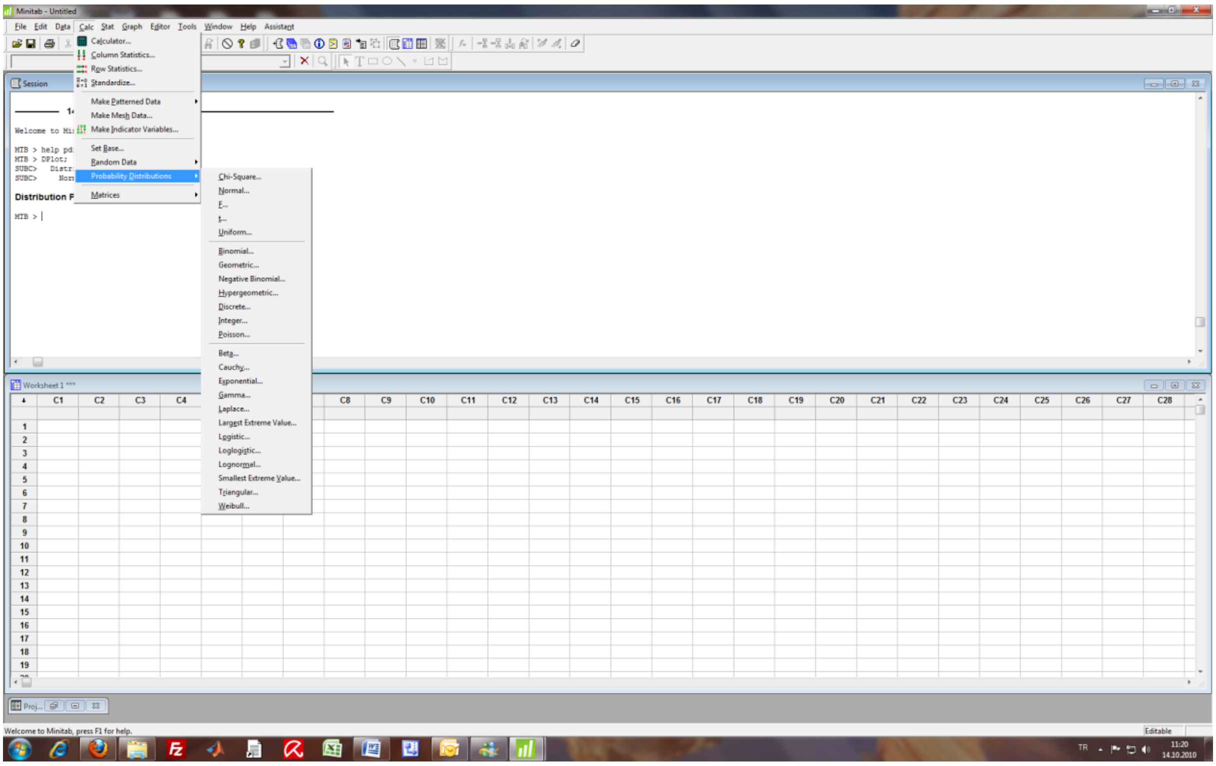
Probability Distribution for x (x=Number of Heads)

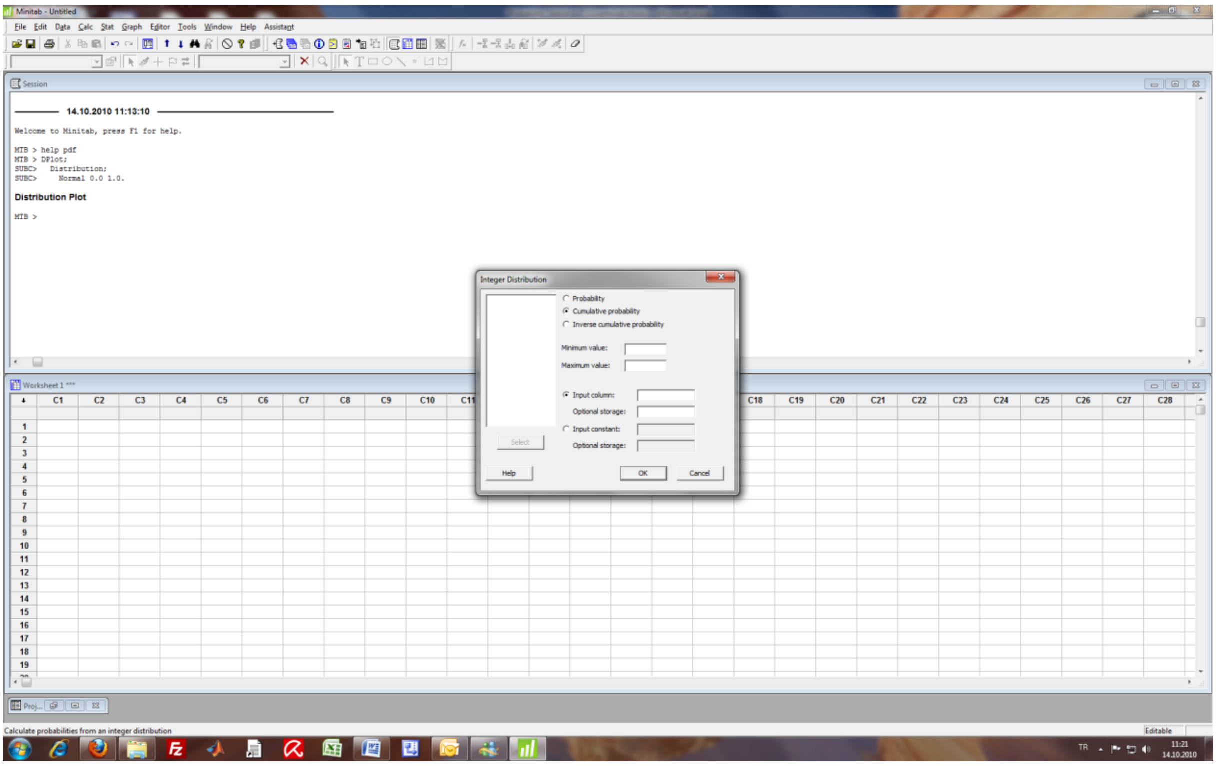
x	Simple Events in x	P(x)
0	E_4	¼
1	E_2,E_3	½
2	E_1	¼

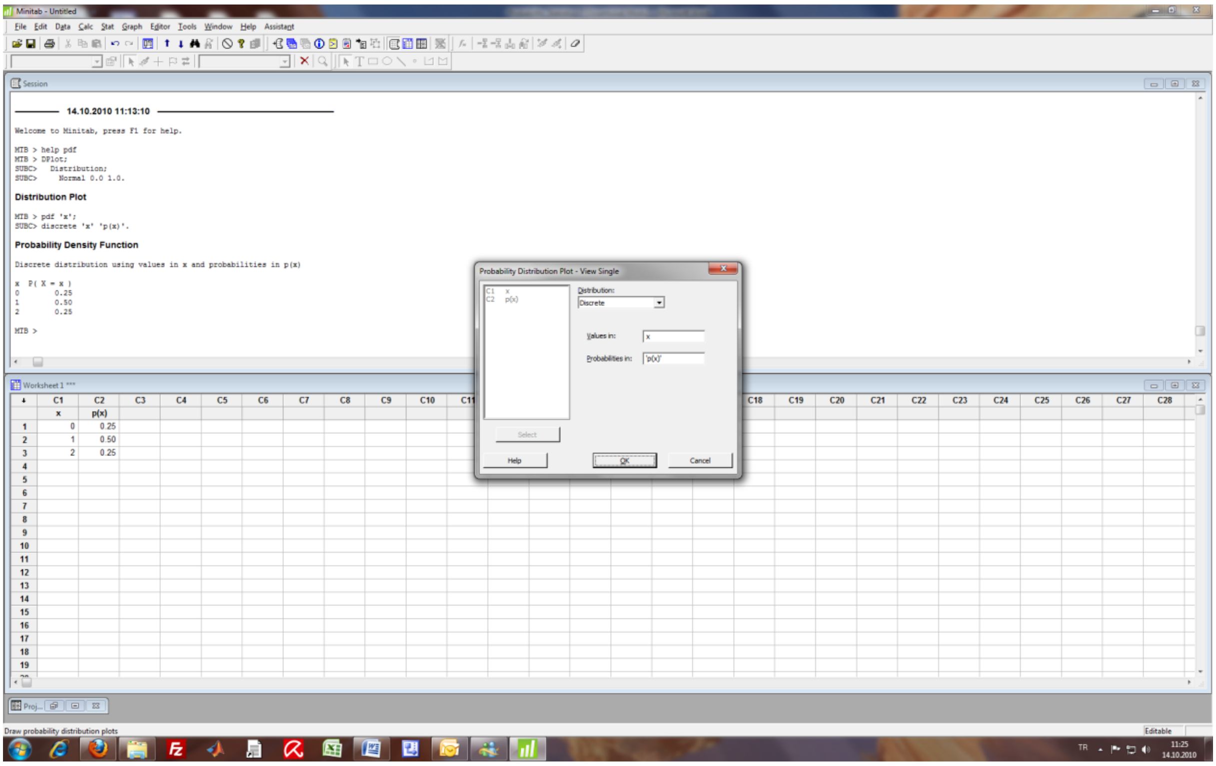
Probability Histogram

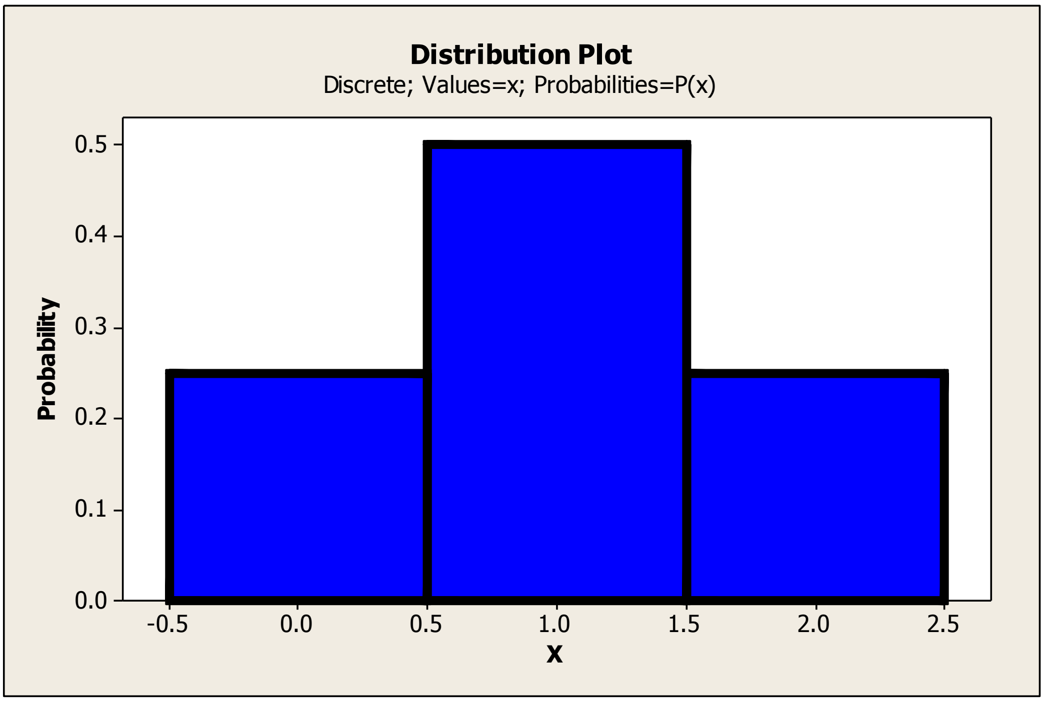












Since the width of each bar is 1, the area under the bar is the probability of observing the particular value of x and <u>the total area equals 1</u>.

```
MTB > PDF 'x';
SUBC> Discrete 'x' 'p(x)'.
```

Probability Density Function

```
Discrete distribution using values in x and
probabilities in p(x)

x      P( X = x )
0         0.25
1         0.50
2         0.25
```

PDF E [E]	**Calculates density values or probabilities for the specified values in E from a standard normal distribution or another specified distribution and stores in E**
CHISQUARE K	Specifies distribution with degrees of freedom = K
NORMAL [K [K]]	Specifies distribution. Generates data from a standard normal; optionally specify mean = K, and standard deviation = K
F K K	Specifies distribution, with numerator degrees of freedom = K and denominator degrees of freedom = K
T K	Specifies distribution, with degrees of freedom = K
UNIFORM [K K]	Specifies distribution. Generates data using lower endpoint = 0.0 and upper endpoint = 1.0. Optionally, specify lower endpoint = K and upper endpoint = K
BINOMIAL K K	Specifies distribution, with number of trials = K and event probability = K
GEOMETRIC K	Specifies distribution, with event probability = K

NONEVENT	Models the number of nonevents before the first event occurs.
TOTAL	Models the total number of trials needed to produce one event.
NEGBINOMIAL K K	Specifies distribution, with event probability = K and number of events needed = K
NONEVENT	Models the number of nonevents before the specified number of events occurs.
TOTAL	Models the total number of trials needed to produce the specified number of events.
HYPERGEOMETRIC K K K	Specifies distribution, with population size = K, event count in population = K, and sample size = K
DISCRETE C C	Specifies distribution, with values in C and probabilities in C
INTEGER K K	Specifies distribution, with discrete uniform on integers from minimum value = K to maximum value = K
POISSON K	Specifies distribution, with mean = K
BETA K K	Specifies distribution, with first shape parameter = K and second shape parameter = K
CAUCHY [K [K]]	Specifies distribution. Generates data using location = 0.0 and scale = 1.0. Optionally, specify location = K and scale = K
EXPONENTIAL [K [K]]	Specifies distribution. Generates data using mean = 1.0 and threshold = 0.0. Optionally, specify mean = K and threshold = K
GAMMA K K [K]	Specifies distribution, with shape = K, scale = K, and optionally, threshold = K
LAPLACE [K [K]]	Specifies distribution. Generates data using location = 0.0 and scale = 1.0. Optionally, specify location = K and scale = K
LEXTREME [K [K]]	Specifies distribution. Generates data using location = 0.0 and scale = 1.0. Optionally, specify location = K and scale = K
LOGISTIC [K [K]]	Specifies distribution. Generates data using

	location = 0.0 and scale = 1.0. Optionally, specify location = K and scale = K
LLOGISTIC [K [K [K]]]	Specifies distribution. Generates data using location = 0.0, scale = 1.0, and threshold = 0.0. Optionally, specify location = K, scale = K, and threshold = K
LNORMAL [K [K [K]]]	Specifies distribution. Generates data using location = 0.0, scale = 1.0, and threshold = 0.0. Optionally, specify location = K, scale = K, and threshold = K
SEXTREME [K [K]]	Specifies distribution. Generates data using location = 0.0 and scale = 1.0. Optionally, specify location = K and scale = K
TRIANGULAR K K K	Specifies distribution, with lower endpoint = K, mode = K, and upper endpoint = K
WEIBULL K K [K]	Specifies distribution, with shape = K, scale = K, and optionally, threshold = K

- For a discrete distribution, the probability distribution function (pdf) calculates probabilities for the specified values (sometimes called the discrete probability distribution function). If you specify a discrete distribution (BINOMIAL, GEOMETRIC, NEGBINOMIAL, HYPERGEOMETRIC, DISCRETE, INTEGER, POISSON), the arguments on the PDF line are optional. If you do not specify arguments, Minitab displays a table of the distribution. If you execute PDF from the menu, you must supply the input columns.

- For a continuous distribution, pdf calculates the continuous probability density function (often called the density function).

- If you do not specify a distribution, results are generated for a normal distribution with mu = 0 and sigma = 1.

Storage is optional. If you specify a storage column, pdf values are stored there and are not displayed in the Session window. If you do not specify a storage column, Minitab displays pdf values.

The Mean and Standard Deviation for a Discrete Random Variable

Definition: ***Let x be a discrete random variable with probability distribution p(x). The mean or expected value of x is given as***

$$\mu = E(x) = \sum xP(x)$$

Where the elements are summed over all values of the random variable x.

Definition: ***Let x be a discrete random variable with probability distribution p(x) and mean μ. The variance of x is given as***

$$\sigma^2 = E[(x-\mu)^2] = \sum (x-\mu)^2 P(x)$$

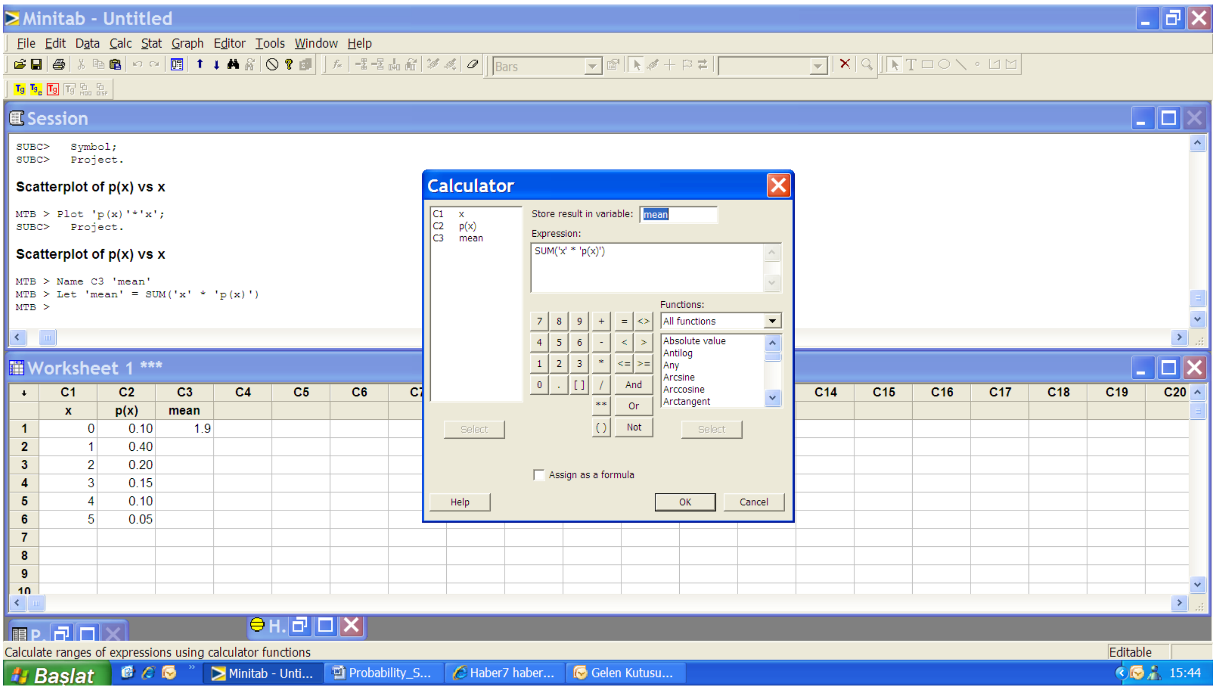
Where the elements are summed over all values of the random variable x.

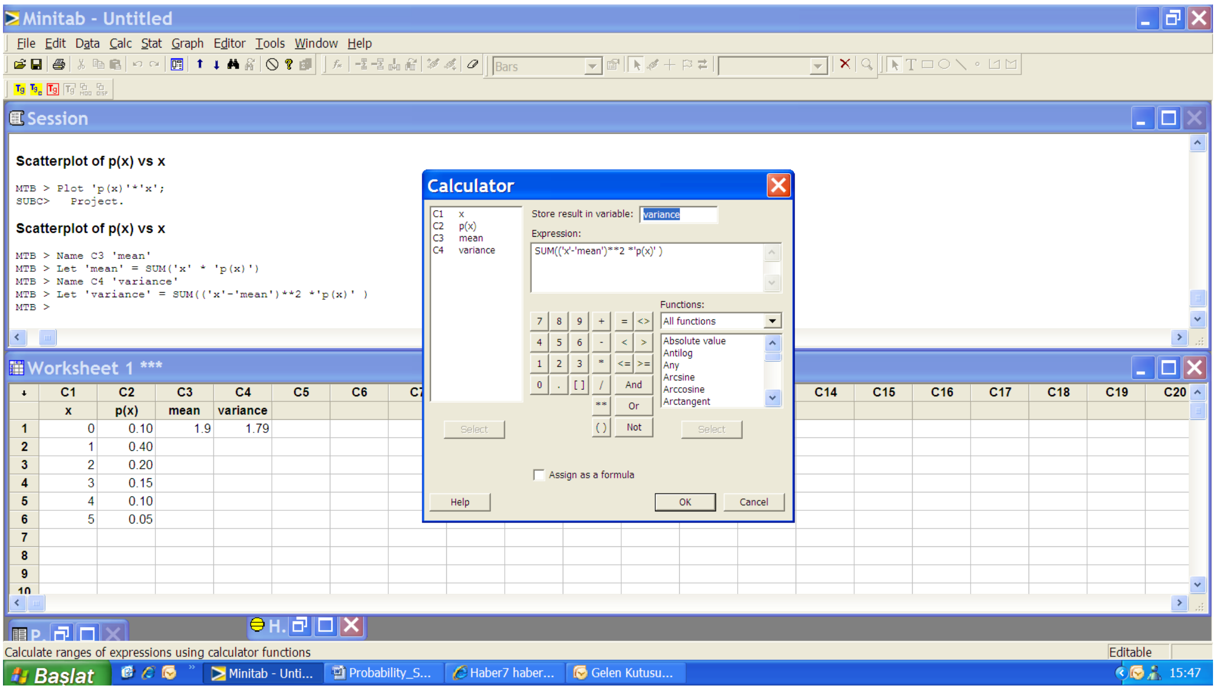
Definition: ***The standard deviation or of a random variable x is equal to the positive square root of its variance.***

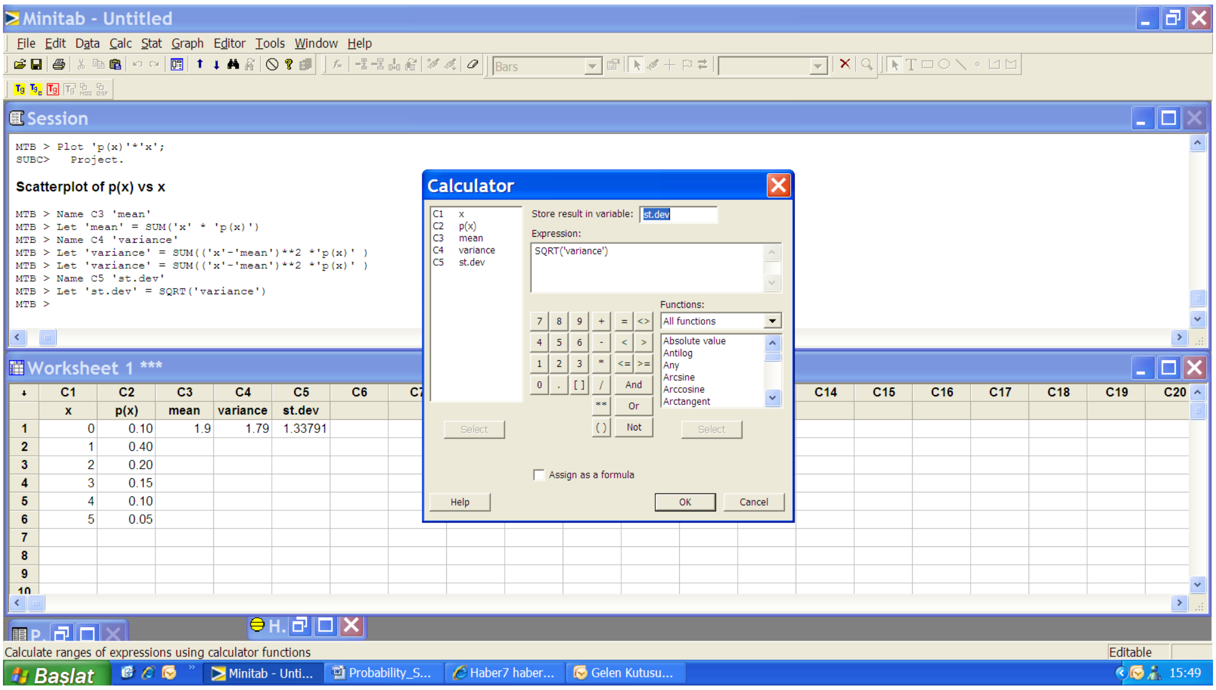
Example: Find the mean, variance, and standard deviation of x for the given probability table below.

Daily demand for the notebook

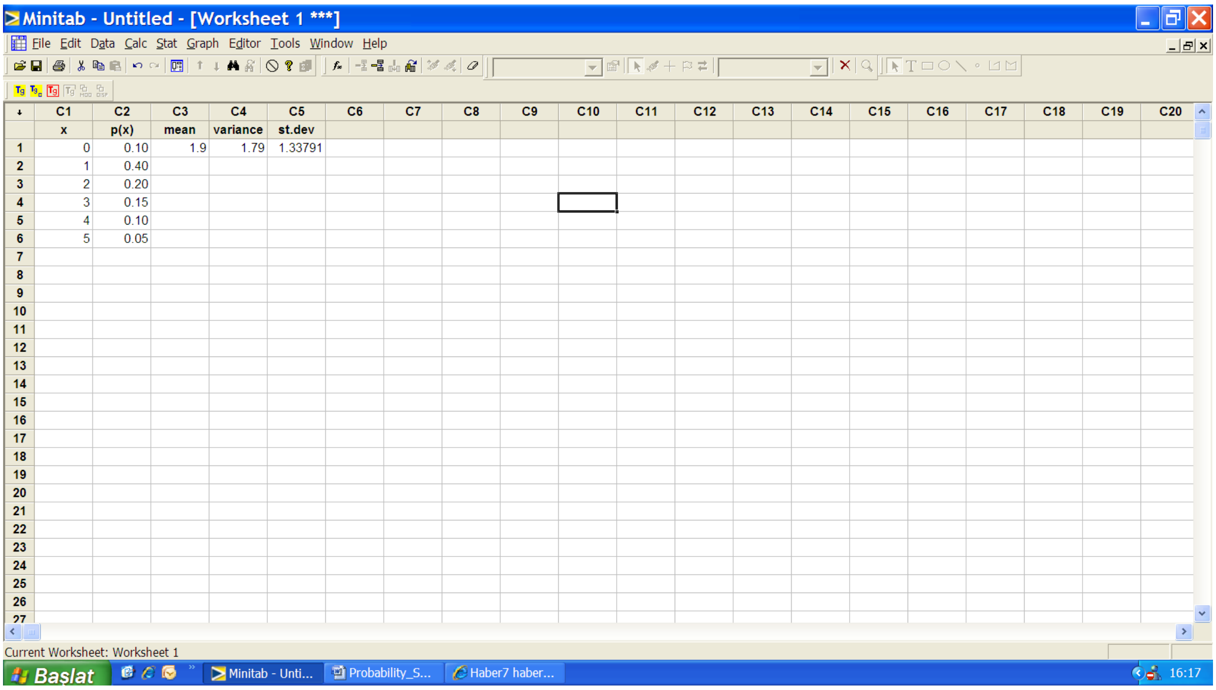
x	0	1	2	3	4	5
P(x)	0.10	0.40	0.20	0.15	0.10	0.05








```
MTB > Name C3 'mean'
MTB > Let 'mean' = SUM('x' * 'p(x)')
MTB > Name C4 'variance'
MTB > Let 'variance' = SUM(('x'-'mean')**2 *'p(x)' )
MTB > Let 'variance' = SUM(('x'-'mean')**2 *'p(x)' )
MTB > Name C5 'st.dev'
MTB > Let 'st.dev' = SQRT('variance')
```



x	P(x)	xP(X)	(x-mean)^2	(x-mean)^{2}P(x)
0	0.10	0.00	3.61	0.3610
1	0.40	0.40	0.81	0.3240
2	0.20	0.40	0.01	0.0020
3	0.15	0.45	1.21	0.1815
4	0.10	0.40	4.41	0.4410
5	0.05	0.25	9.61	0.4805
Total	1.00	1.90		1.79

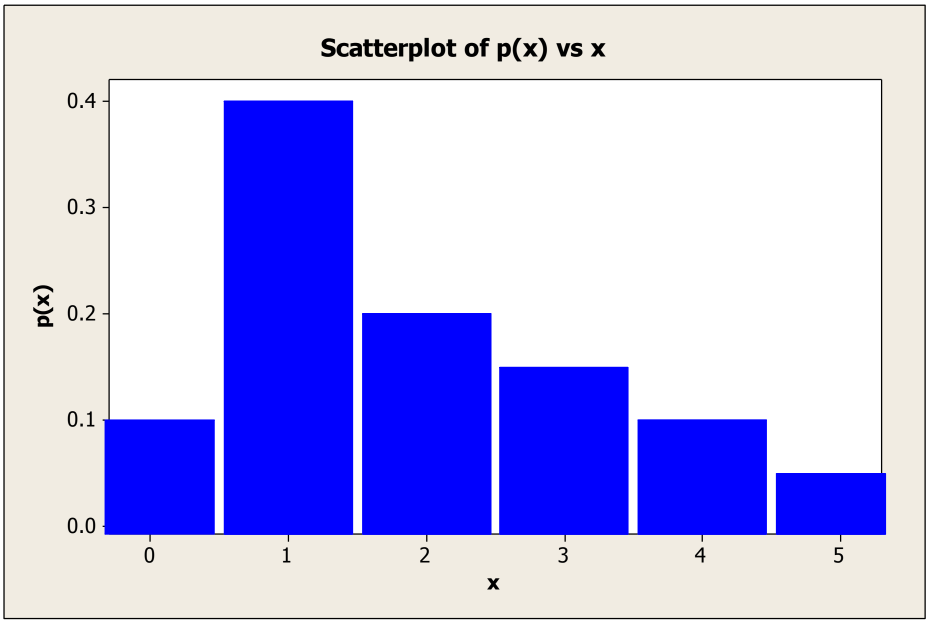
Alternative Computation for variance

x	P(x)	xP(X)	x^{2}P(X)
0	0.10	0.00	0.00
1	0.40	0.40	0.40
2	0.20	0.40	0.80
3	0.15	0.45	1.35
4	0.10	0.40	1.60
5	0.05	0.25	1.25
Total	1.00	1.90	5.4

$$\sigma^2 = E[(x-\mu)^2] = \sum (x-\mu)^2 P(x)$$

$$\sigma^2 = E(x^2) - [E(x)]^2 = \sum x^2 P(x) - [\sum xP(x)]^2$$

$$\sigma^2 = 5.4 - [1.90]^2 = 5.4 - 3.61 = 1.79$$



Probability distribution for the example

Since the distribution is **mound-shaped, approximately** 95% of all measurements should lie within two standard deviations of the mean-that is,

$$\mu \pm 2\sigma = 1.90 \pm 2(1.34)$$ or **-0.78 to 4.58**

Since x=5 lies outside this interval, we can say it is unlikely that five or more customers will want to buy a notebook today. In fact, $P(x \geq 5)$ is exactly 0.05 or 1 time 20.

Cumulative Distribution Function

Definition: ***The cumulative distribution function of a discrete random variable X, denoted F(x), is***

$$F(x) = P(X \le x) = \sum_{x_i \in x} p(x_i)$$

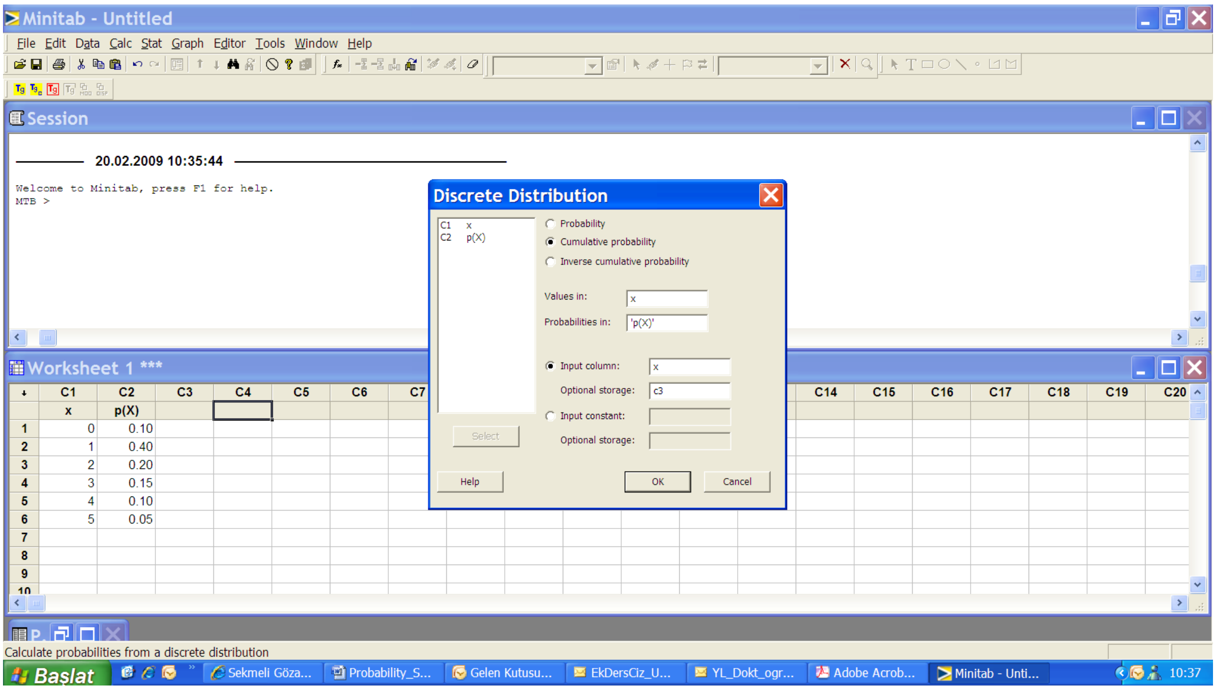
For a discrete random variable X, F(x) satisfies the following properties.

- $F(x) = P(X \le x) = \sum_{x_i \in x} p(x_i)$
- $0 \le F(x) \le 1$
- ***If*** $x \le y$ ***then*** $F(x) \le F(y)$

Like a probability function, a cumulative distribution function provides probabilities.

Example: Find the cumulative distribution function for the given probability table below.

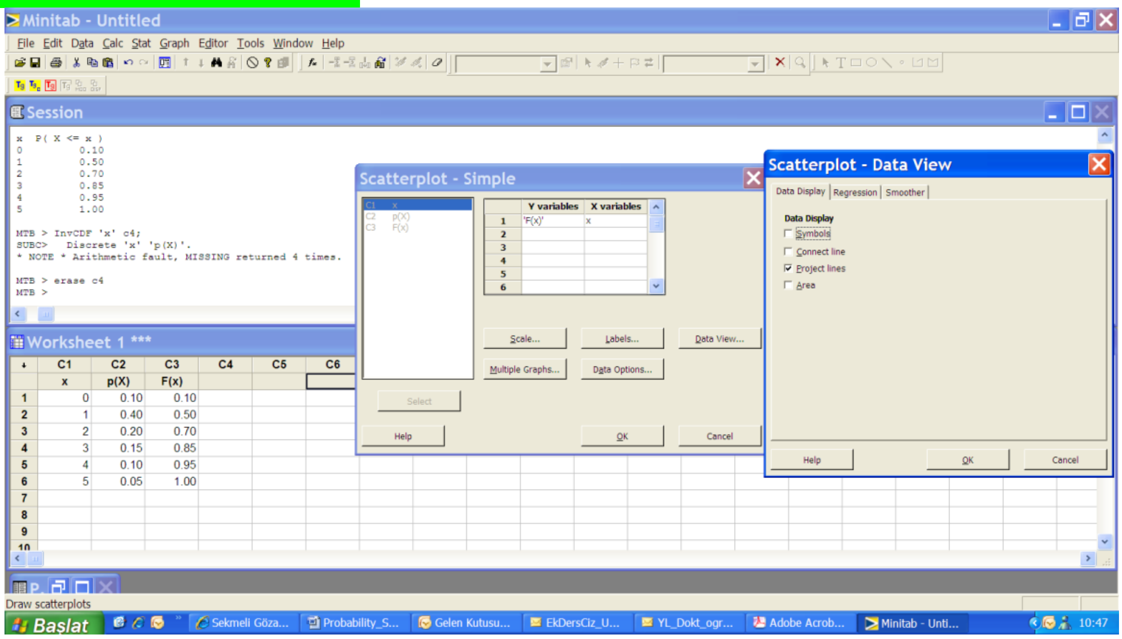
x	P(x)
0	0.10
1	0.40
2	0.20
3	0.15
4	0.10
5	0.05
Total	1.00

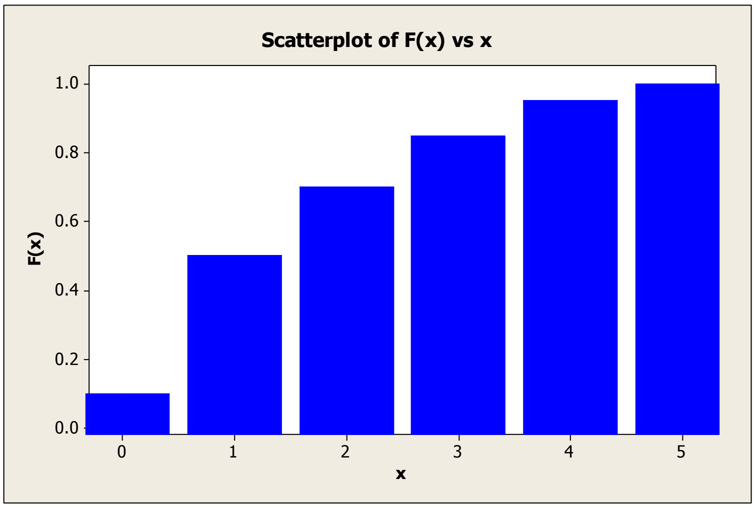


```
MTB > CDF 'x';
SUBC>   Discrete 'x' 'p(X)'.
```

Cumulative Distribution Function

```
Discrete distribution using values in x
and probabilities in p(X)
x   P( X <= x )
0          0.10
1          0.50
2          0.70
3          0.85
4          0.95
5          1.00
```





Cumulative Distribution Function