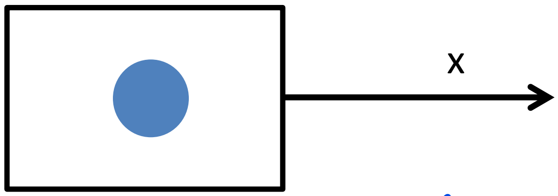
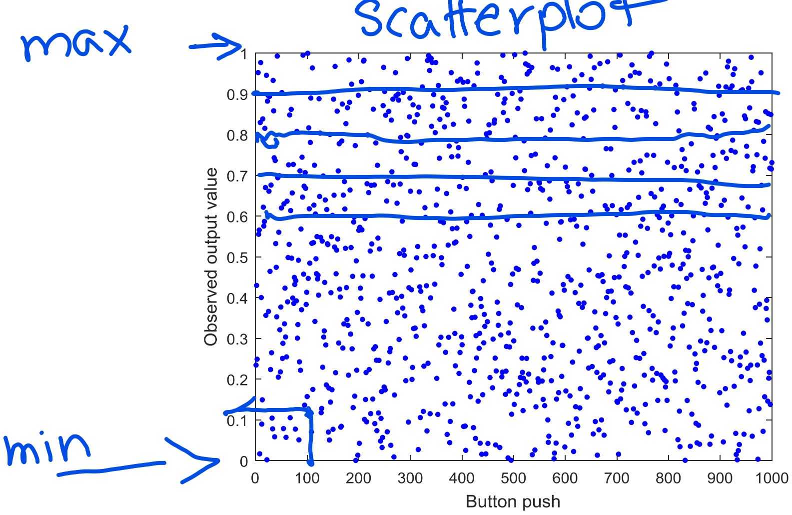
Introduction to STAT 351-Probability & Statistics for ECE

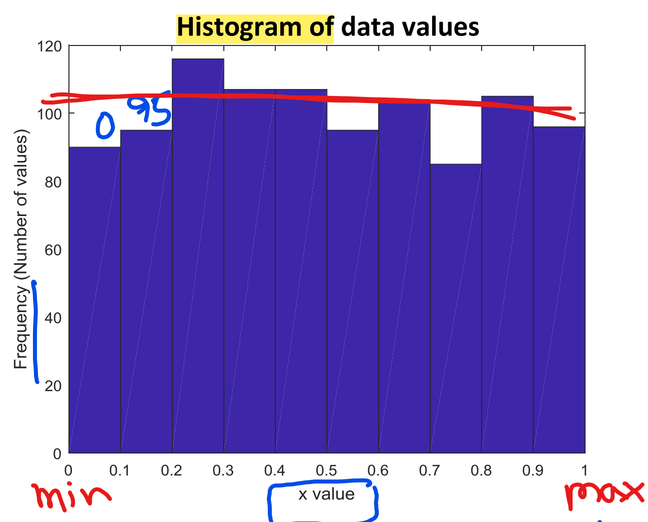
Statistics Use observations (data) to fit models that describes variation (pattern) in observations and predic- a new observation base on the model.

Number Machine: Press the button to get output value for x.



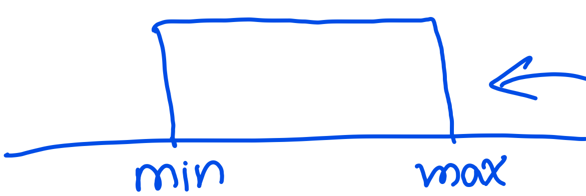
Scatterplot





There is No pattern.

Histogram summarizes observed values.
This istogram shows approxim tely even distribution between 0 and 1.
This is called uniform distribution.

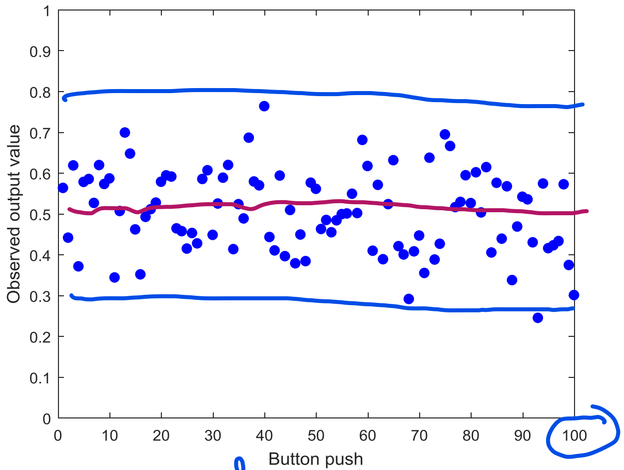


MATLAB code:

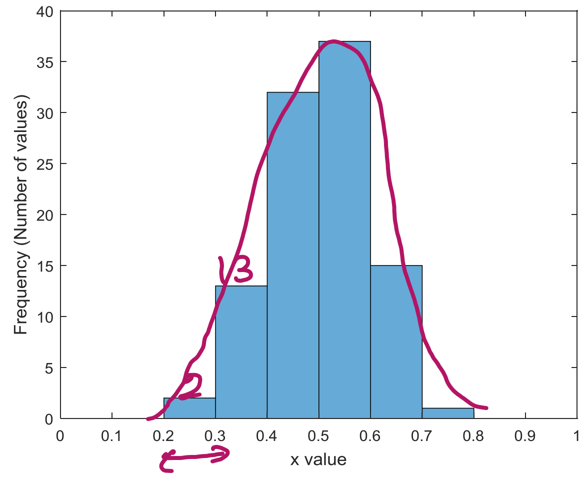
```
% Generate and plot some random sample;
N=1000;
%Uniform distribution
x=rand(1,N);
figure(1); plot(x,'b.', 'MarkerSize',10)
ylabel('Observed output value');
xlabel('Button push');
```

```
% Create a histogram
numbins=10;
figure(2);hist(x,numbins);
xlim([0 1])
ylabel('Number (Frequency) of values');
xlabel('x value');
```

Another machine output:



Histogram of data values with Frequency

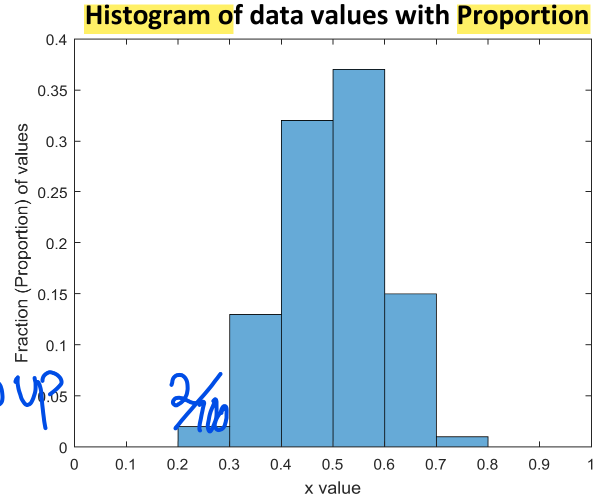


Less spread out.

Centered at about 0.5

Histogram shows approximatel bell-Shape (normal or Gaussian) distribution.

Histogram of data values with Proportion



$$\text{Proportion} = \frac{\text{Frequency in each group}}{\text{Total Frequency}}$$

MATLAB code:

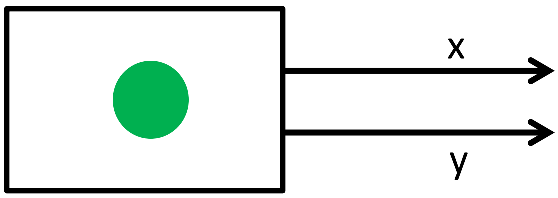
```
% Generate and plot some random sample;
N=100;
% Normal distribution
x=0.5+0.1*randn(1,N);
figure(3); plot(x, 'b.', 'MarkerSize',20);
ylim([0 1]);
ylabel('Observed output value');
xlabel('Button push');
```

Here, we use Monl Carlo simulation to generate data randomly.

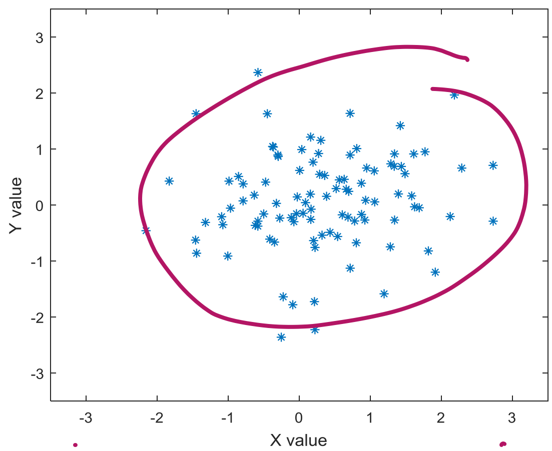
```
% Create a histogram with Frequency
figure(4);
histogram(x,'BinWidth',0.1); xlim([0 1]);
ylabel('Frequency (Number of values)');
xlabel('x value');
```

```
% Create a histogram with Proportion
figure(5);
histogram(x,'Normalization',
'probability','BinWidth',0.1); xlim([0 1]);
ylabel('Fraction (Proportion) of values');
xlabel('x value');
```

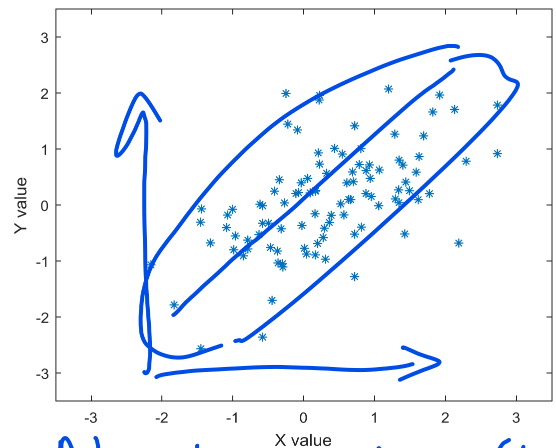
Number Machine with two outputs: Press the button to get output values for x and y.



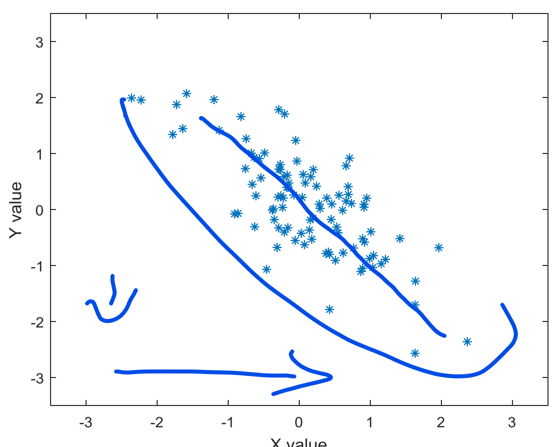
The following are different scatterplots of two output values obtained from different machines. Is there a pattern? Is there a relationship between x and y values? What does number of one value tells about the number of the other value?



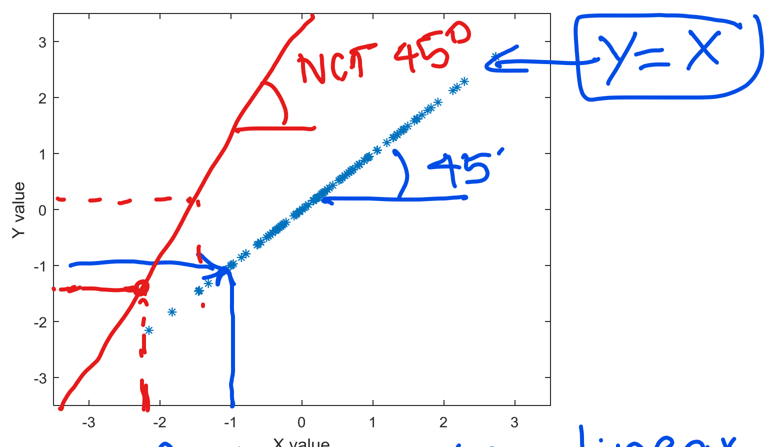
X, Y seems to be unrelated.



About positive (increasing) linear relationship.



About dereasing (inverse or negative) linear relationship.



Perfect increasing linear relationship. This does not exist in statistics.

```
%% Generate correlated two variables
with normal distributions;
mu1     = 0;    sigma1 = 0.5;
mu2     = 0;    sigma2 = 0.5;
a3      = 0;    b3     = 1;
n = 100;
Rho = [1.0  0.1  0.5;
       0.1 1.0  -0.8;
       0.5 -0.8  1.0];
Z = mvnrnd([0 0 0], Rho, n);
```

```
% Plot data;
figure(8); plot(Z(:,1), Z(:,2),'*');xlim([-3.5 3.5]);
ylim([-3.5 3.5]);ylabel('Y value'); xlabel('X value');

figure(9); plot(Z(:,1), Z(:,3),'*');xlim([-3.5 3.5]);
ylim([-3.5 3.5]);ylabel('Y value'); xlabel('X value');

figure(10); plot(Z(:,2), Z(:,3),'*');xlim([-3.5 3.5]);
ylim([-3.5 3.5]);ylabel('Y value'); xlabel('X value');

figure(11); plot(Z(:,1), Z(:,1),'*');xlim([-3.5 3.5]);
ylim([-3.5 3.5]);ylabel('Y value'); xlabel('X value');
```

The house of Statistics is built on the foundation of Probability Theory.

Probability

Computer memories:

- Suppose you are designing a computer memory to hold k-bit words.
- To increase system reliability, you employ an error-correcting-code system.
- With this system, instead of storing just the k data bits, you store an additional n bits (which are functions of the data bits).
- When reading back the $(k+n)$-bit word, if at least m bits are read out correctly, then all k data bits can be recovered (the value of m depends on the code).
- To characterize the quality of the computer memory, we compute the probability that at least m bits are correctly read back.
- You will be able to do this after you study the **binomial** random variable.

Count number of correctly read bits out of $(k+n)$ bits.

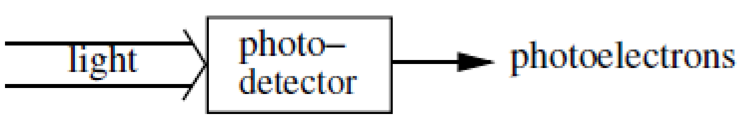
$0, 1, 2, 3, \dots, (k+n)$

Optical communication systems:

- Optical communication systems use photodetectors (see Figure below) to interface between optical and electronic subsystems.

$0, 1, 2, 3, \dots$

- When these systems are at the limits of their operating capabilities, the number of photoelectrons produced by the photodetector is well-modeled by the **Poisson** random variable you will study later.
- In deciding whether a transmitted bit is a zero or a one, the receiver counts the number of photoelectrons and compares it to a threshold.
- System performance is determined by computing the probability that the threshold is exceeded.



Wireless communication systems:

- In order to enhance weak signals and maximize the range of communication systems, it is necessary to use amplifiers.
- Unfortunately, amplifiers always generate thermal noise, which is added to the desired signal.
- As a consequence of the underlying physics, the noise has a **Gaussian (normal)** distribution.
- Hence, the Gaussian density function, which you will learn later, plays a prominent role in the analysis and design of communication systems.
- When noncoherent receivers are used, e.g., noncoherent frequency shift keying, this naturally leads to the Rayleigh, chi-squared, noncentral chi-squared, and Rice density functions that you may meet later.