Events associated with numbers; Random Variables

Events associated with numbers; Random Variables

Dichotomous (Bernoulli): $X = 0$ or 1

Events associated with numbers; Random Variables

Dichotomous (Bernoulli): $X = 0$ or 1

$$P(X=1) = p$$

$$P(X=0) = 1-p$$

e.g. Suppose that 80% of the villagers should be vaccinated. What is the probability that at random you choose a vaccinated villager?

e.g. Suppose that 80% of the villagers should be vaccinated. What is the probability that at random you choose a vaccinated villager?

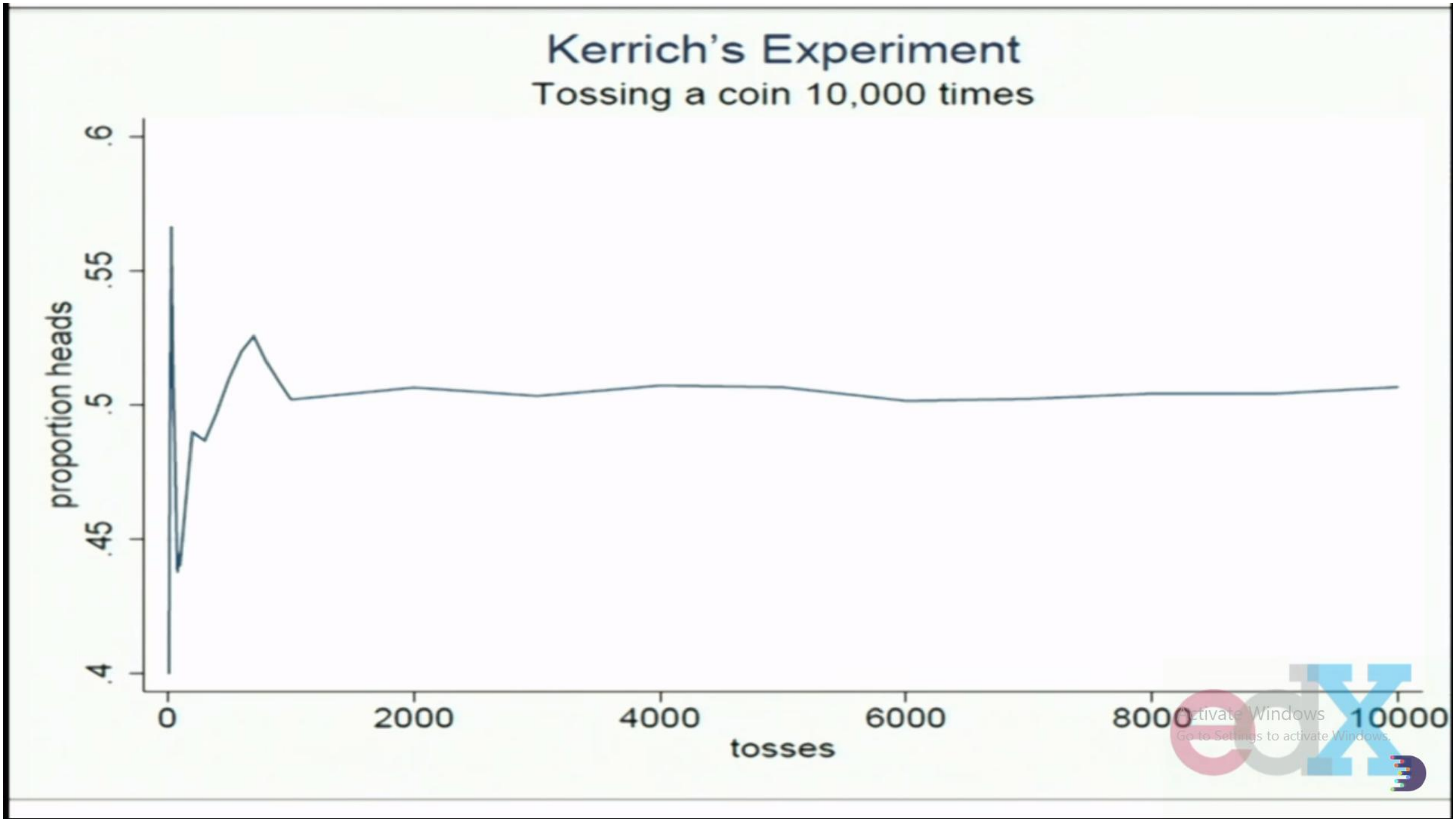
$1 \equiv$ success (vaccinated person)
$0 \equiv$ failure (unvaccinated person)

e.g. Suppose that 80% of the villagers should be vaccinated. What is the probability that at random you choose a vaccinated villager?

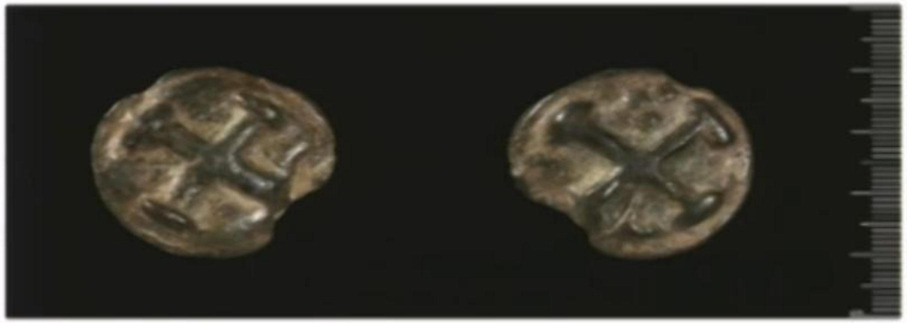
$1 \equiv$ success (vaccinated person)
$0 \equiv$ failure (unvaccinated person)

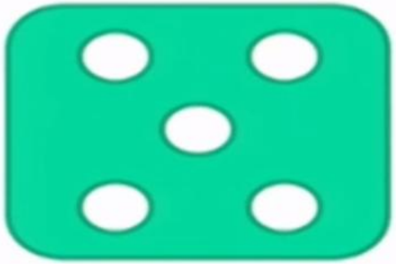
1 Trial

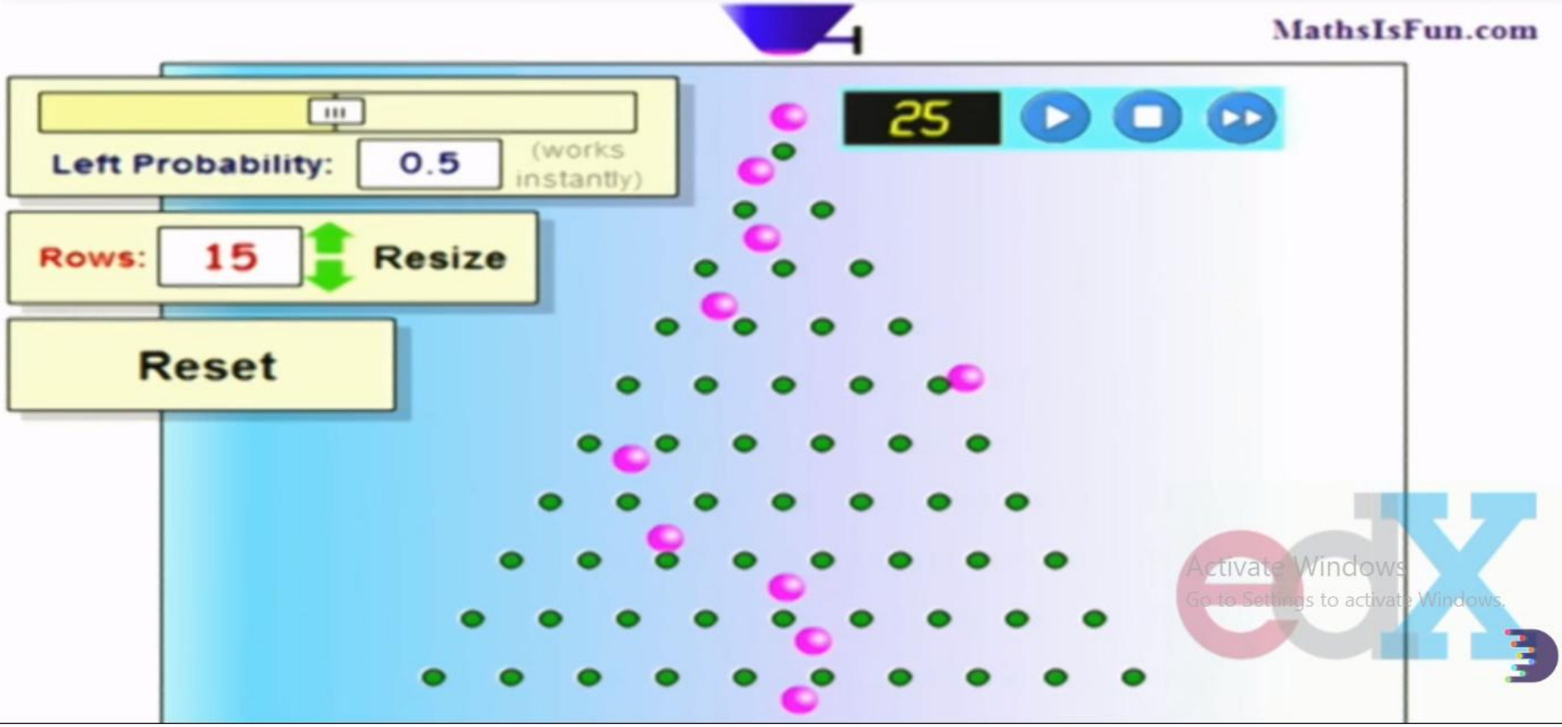
$P(0) = 1\text{-}p = 0.2$
$P(1) = p \quad = 0.8$


Kerrich's Experiment
Tossing a coin 10,000 times
proportion heads
.4
.45
.5
.55
.6
0
2000
4000
6000
8000
10000
tosses

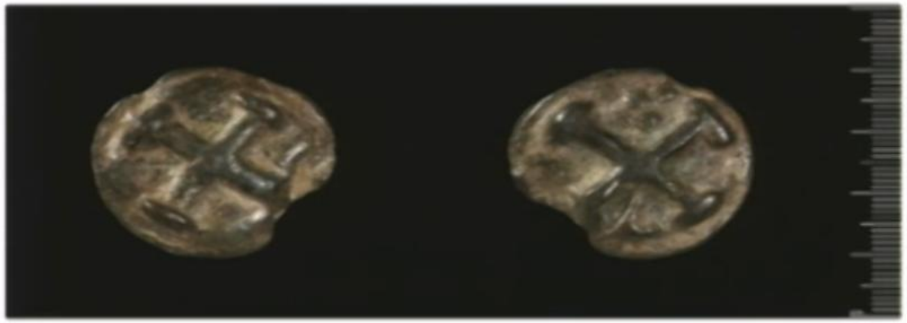
Quincunx

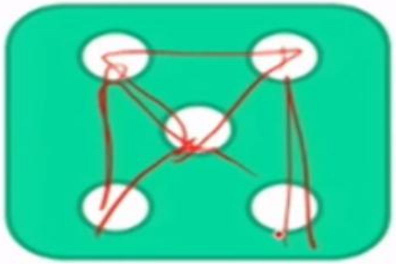


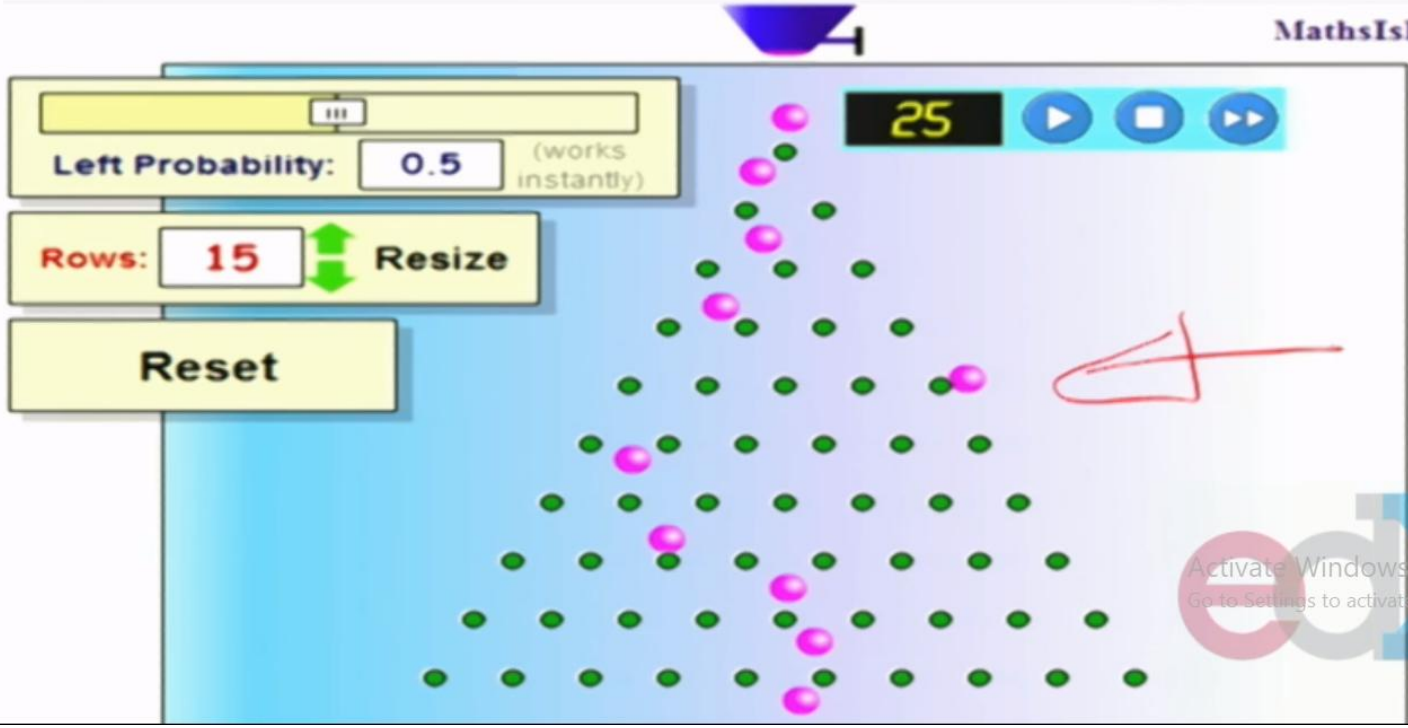



MathsIsFun.com
25
Left Probability:
0.5
(works instantly)
Rows:
15
Resize
Reset

Quincunx

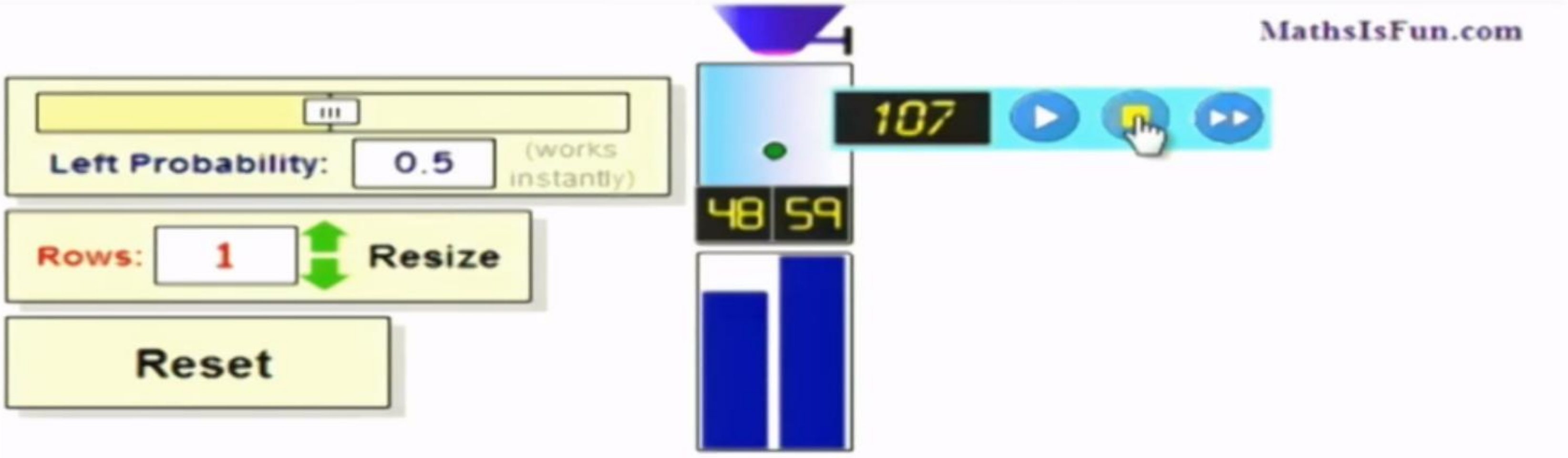





MathsIsFun.com
25
Left Probability: 0.5
(works instantly)
Rows: 15
Resize
Reset

Quincunx

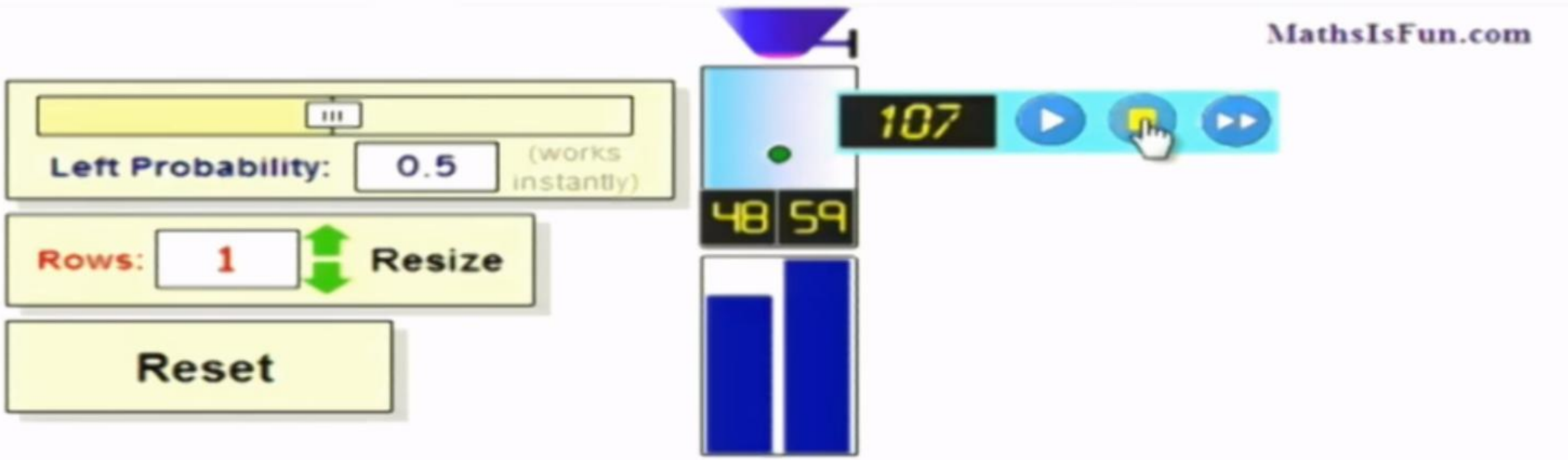
The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



http://www.mathsisfun.com

Quincunx

The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



$$\frac{48}{107} \quad \frac{59}{107}$$

http://www.mathsisfun.com

Mendelian Segregation

Gregor Mendel
1823-1884

Sperm

Egg

	A 1/2	a 1/2
A 1/2	AA 1/4	Aa 1/4
a 1/2	Aa 1/4	aa 1/4

Probabilities

Gregor Mendel
1823-1884

Mendelian Segregation

Sperm

Egg

	A 1/2	a 1/2
A 1/2	AA 1/4	Aa 1/4
a 1/2	Aa 1/4	aa 1/4

Probabilities

Gregor Mendel
1823-1884

Mendelian Segregation

(A,a) ⟷ (A,a)

↓

(A,A)	1/4
(A,a)	1/2
(a,a)	1/4

Gregor Mendel
1823-1884

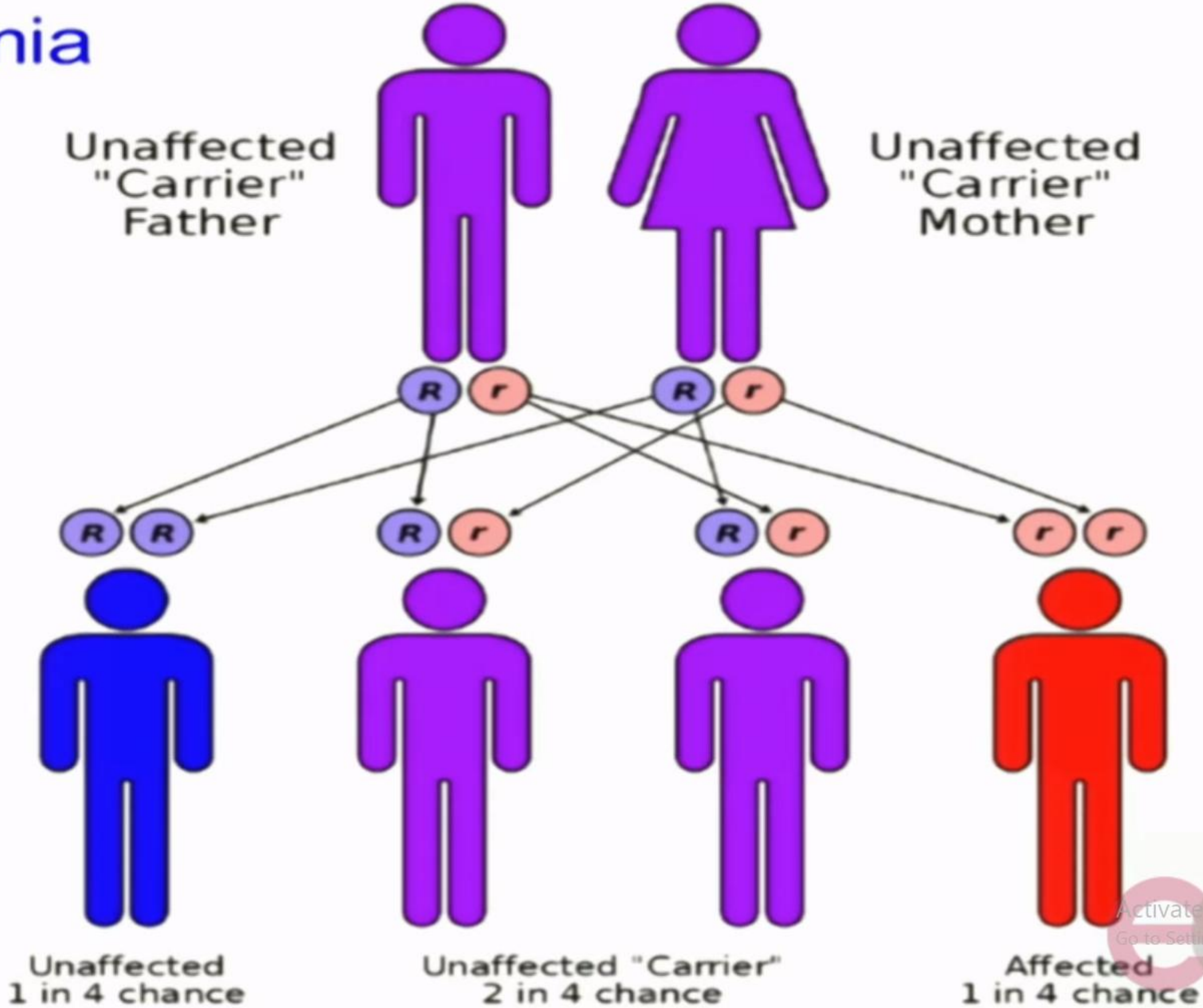
Mendelian Segregation

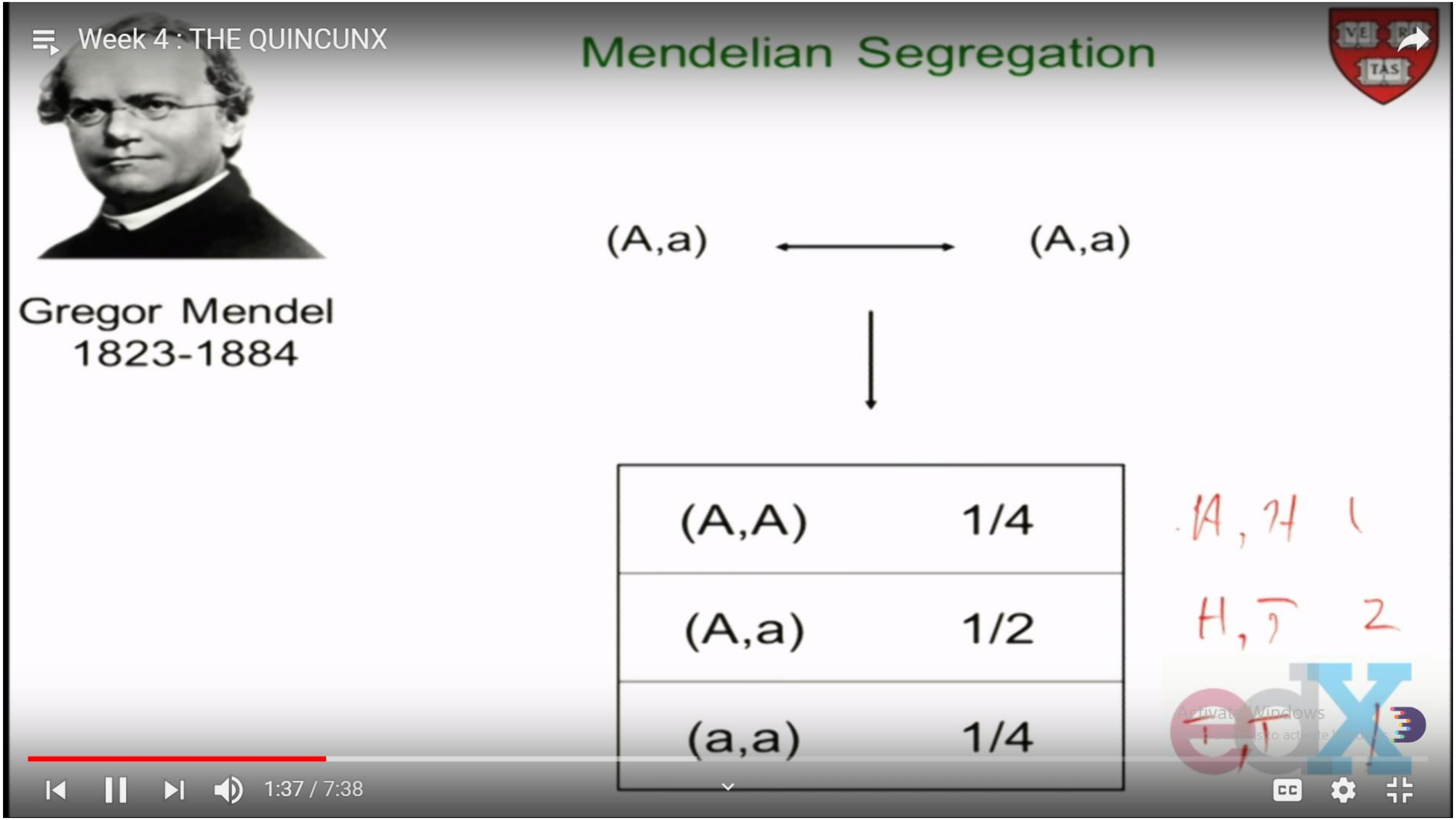
(A,a) $\longleftrightarrow$ (A,a)

$\downarrow$

(A,A)	1/4	H, H
(A,a)	1/2	H, T
(a,a)	1/4	T, T

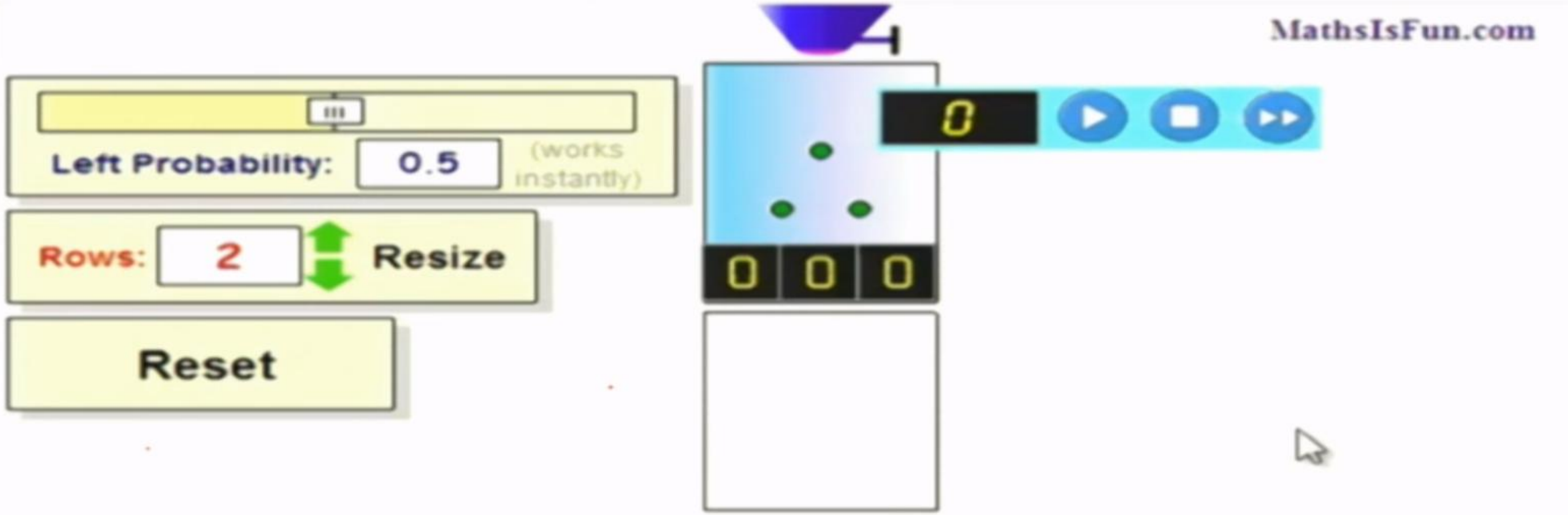
Thalassemia




Week 4 : THE QUINCUNX
Mendelian Segregation
Gregor Mendel
1823-1884
(A,a)
(A,a)
(A,A) 1/4
(A,a) 1/2
(a,a) 1/4
H, T 2
1:37 / 7:38

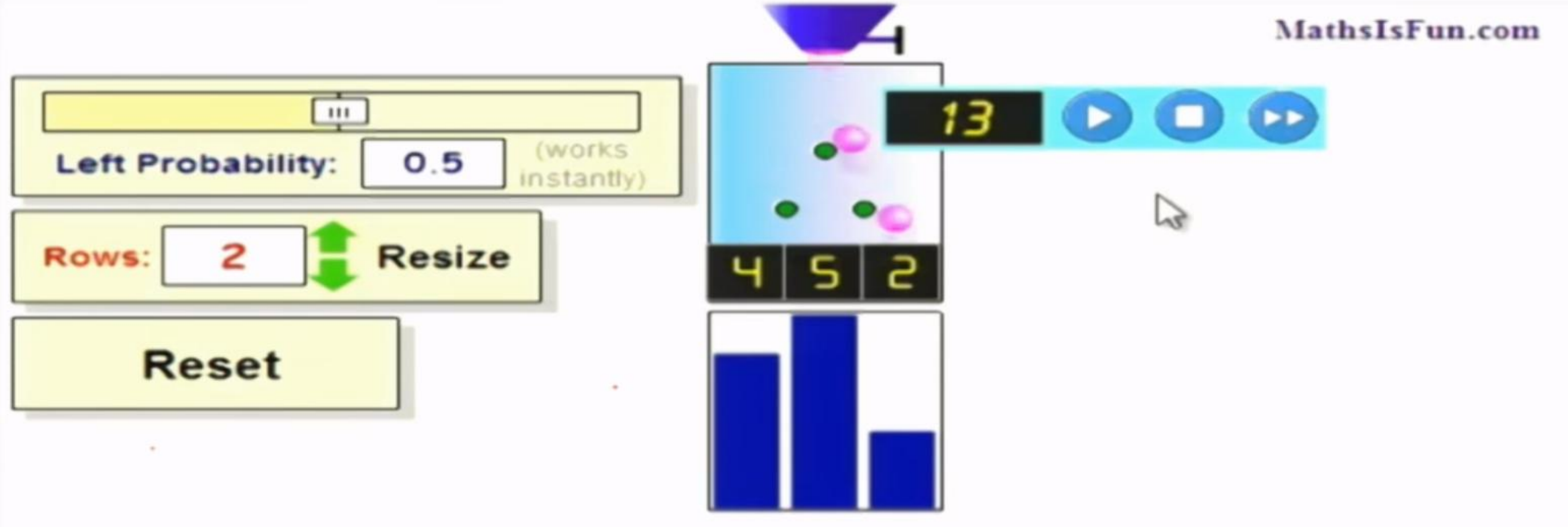
Quincunx

The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



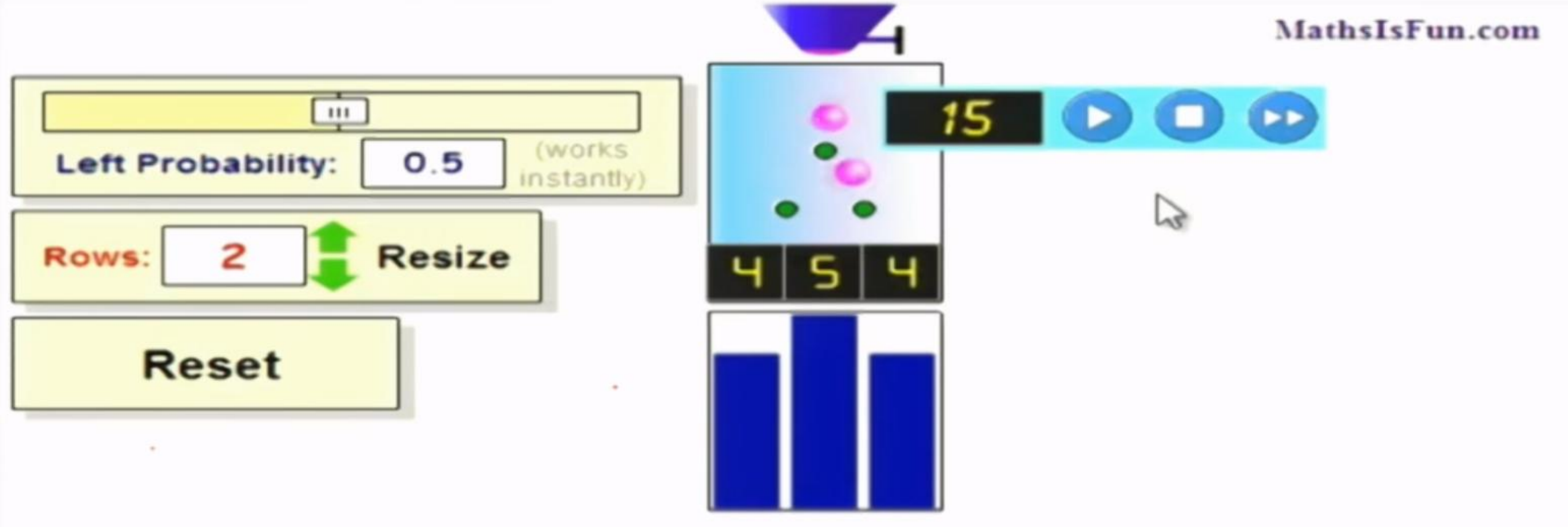
Quincunx

The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



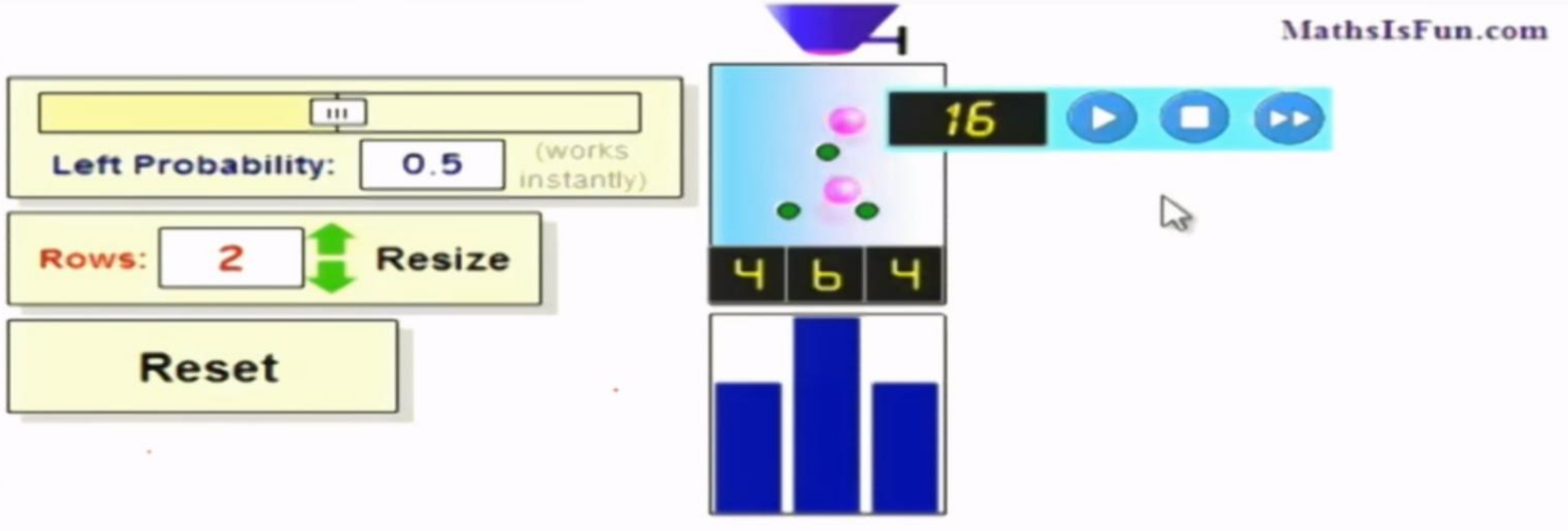
Quincunx

The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



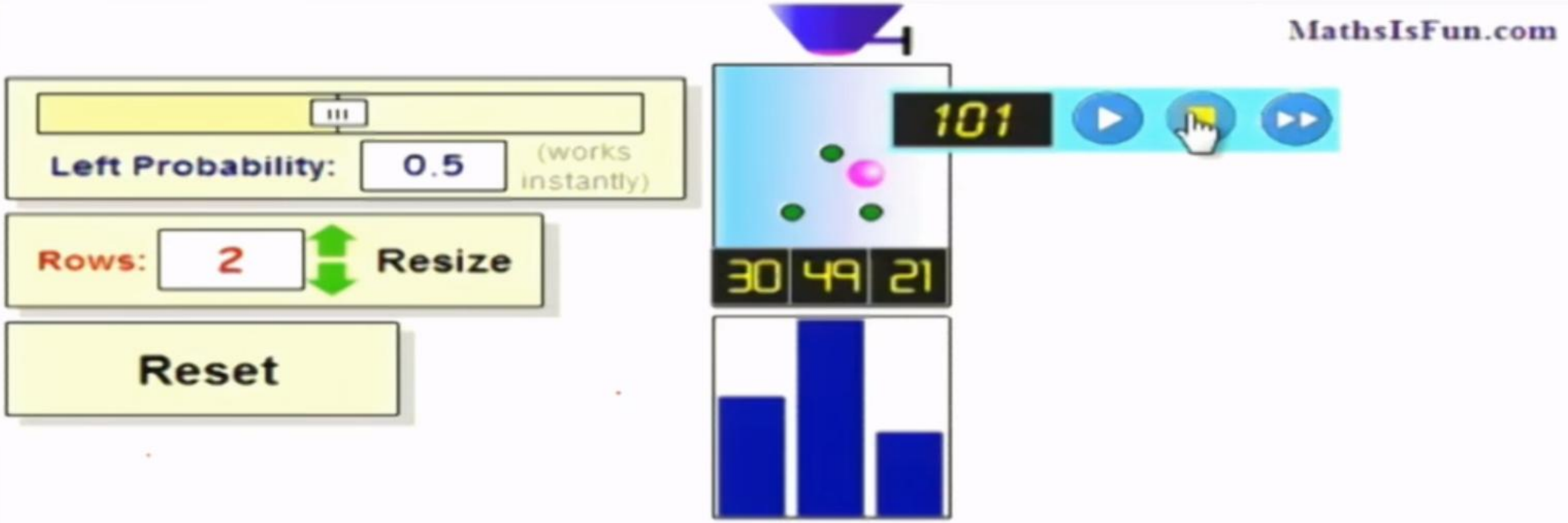
Quincunx

The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



Quincunx

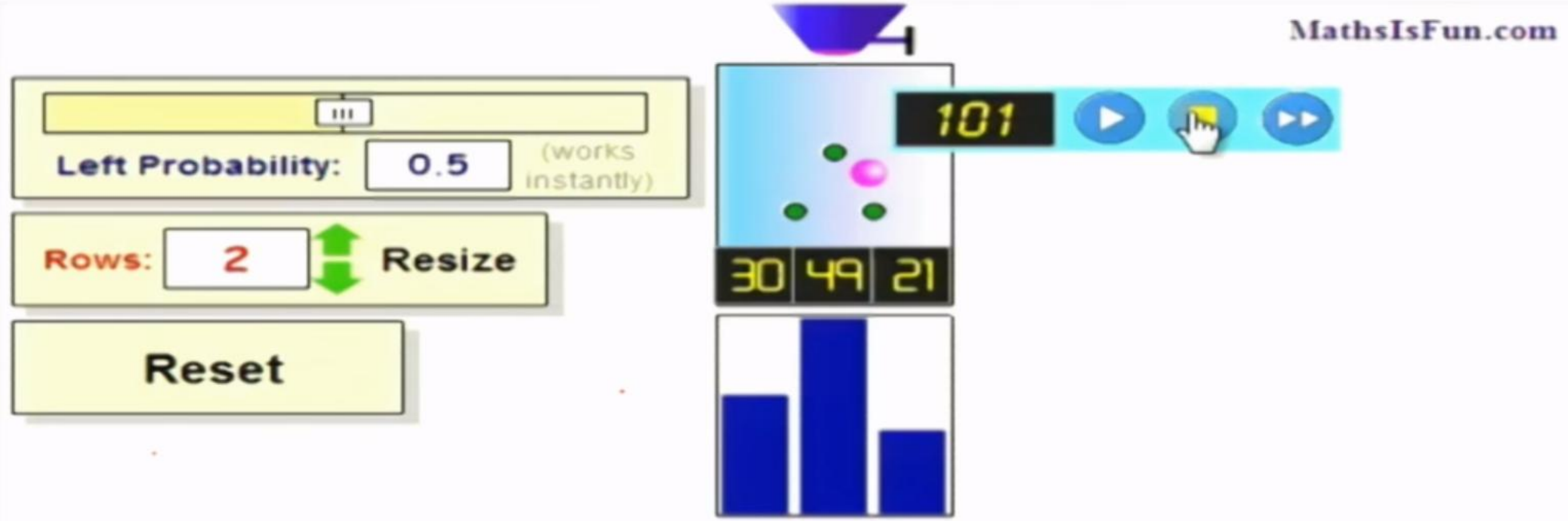
The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



1 2 1

Quincunx

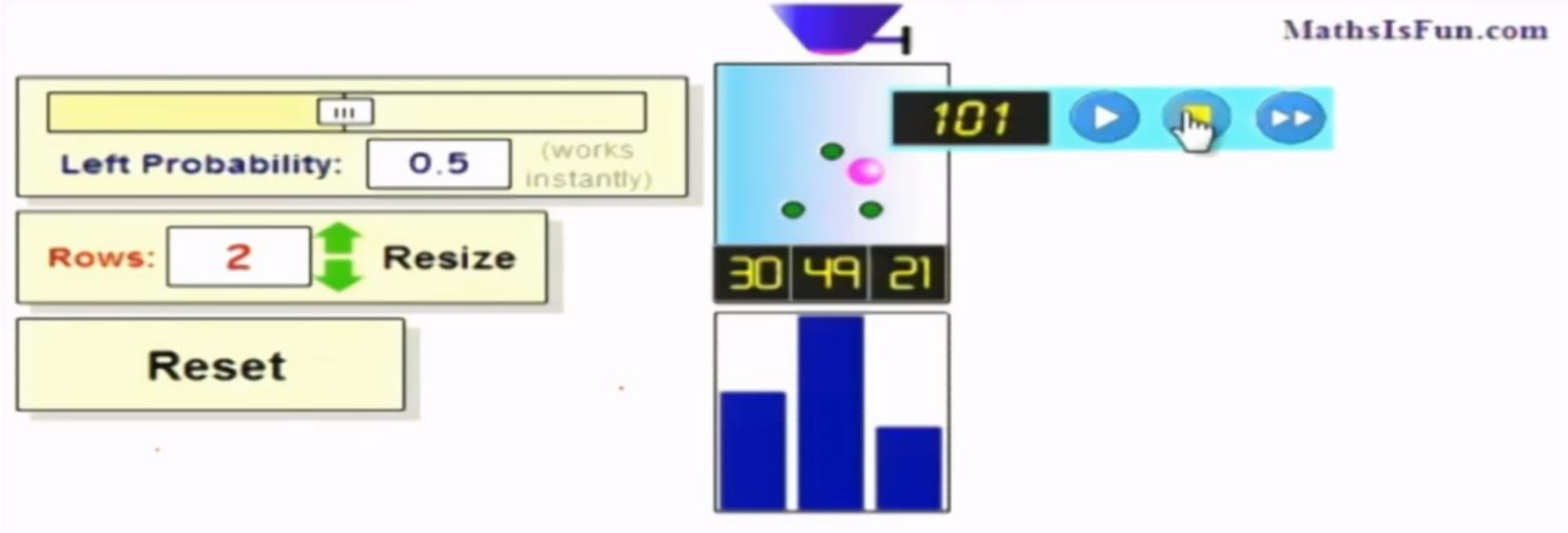
The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



$\frac{1}{4}$ $\frac{2}{4}$ 1

Quincunx

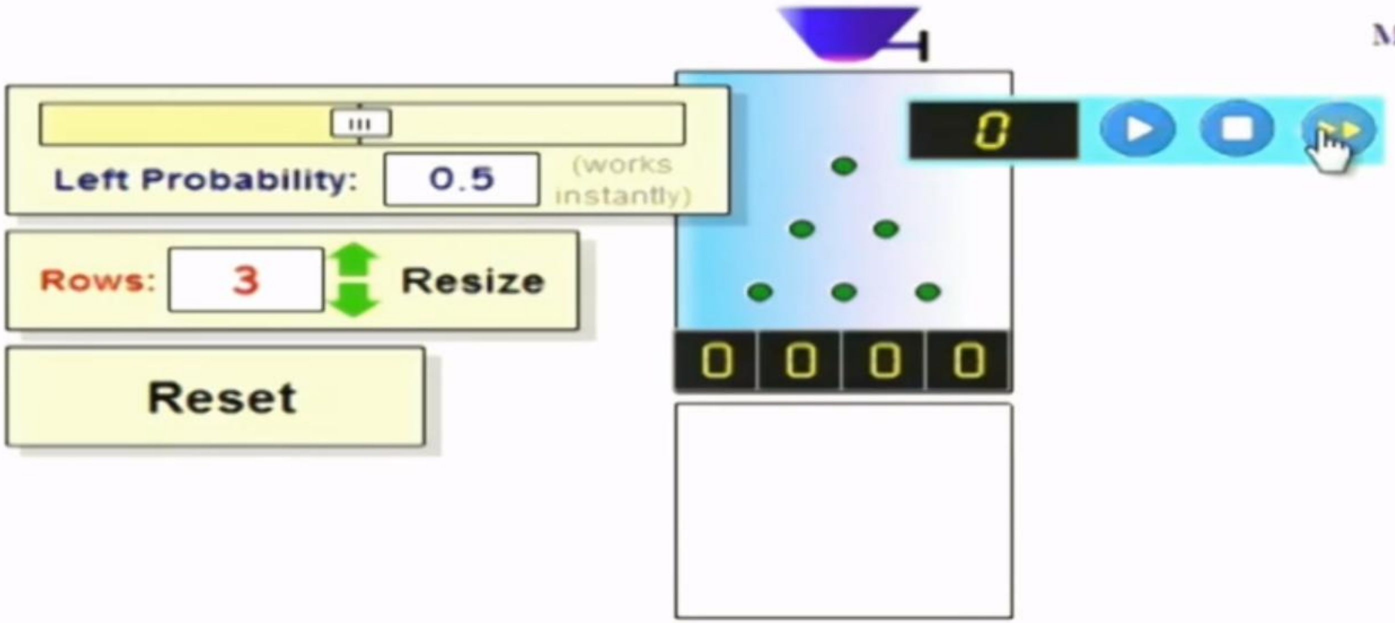
The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



$\frac{1}{4}$ $\frac{2}{4}$ $\frac{1}{4}$

Quincunx

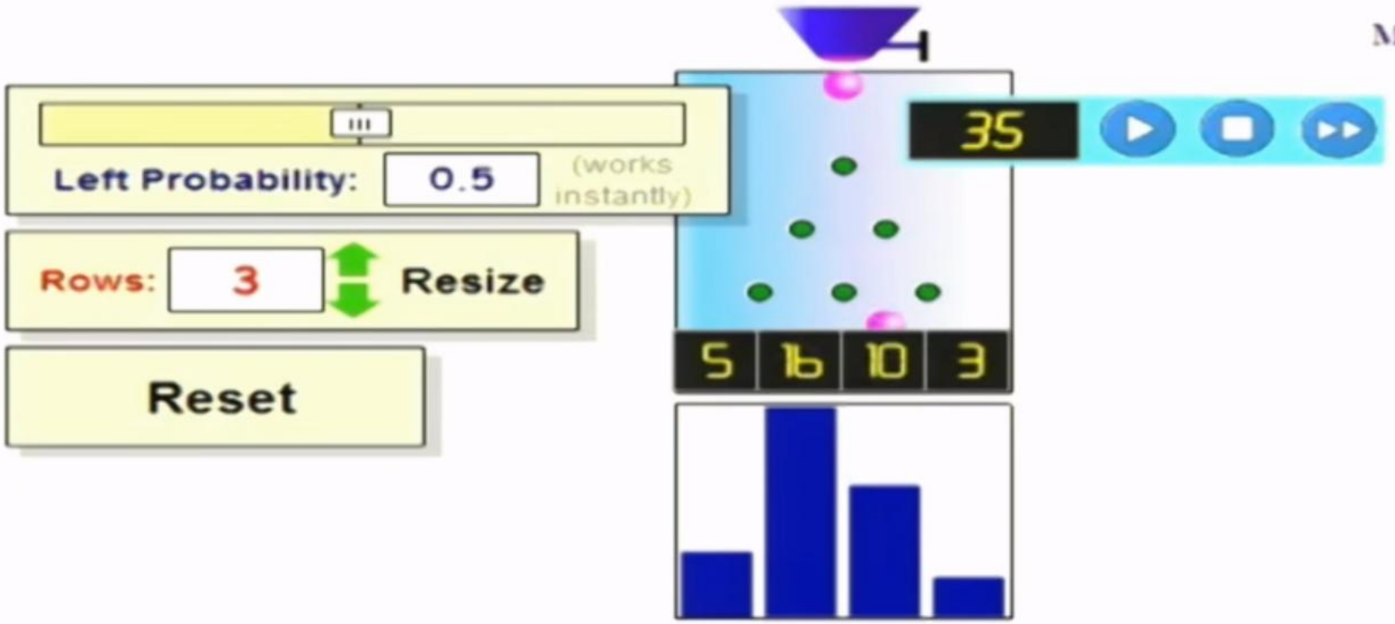
The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



Quincunx

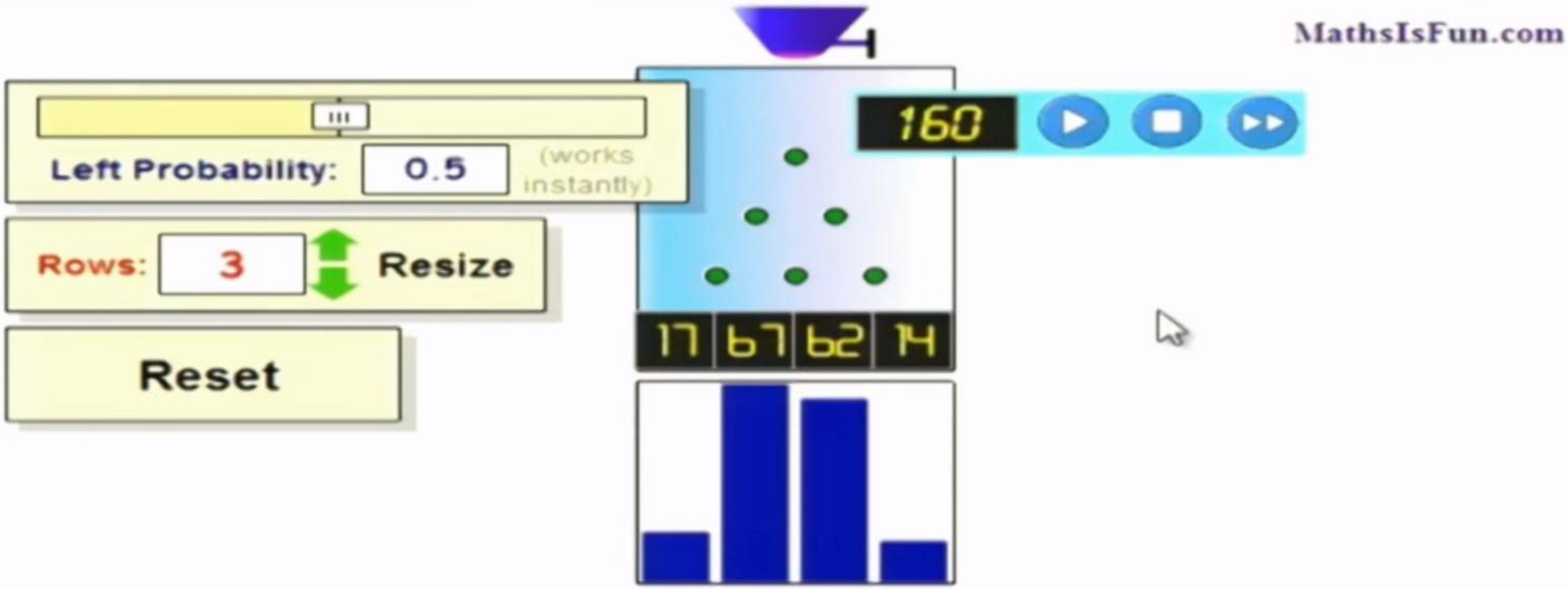
The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.

MathsIsFun.com



Quincunx

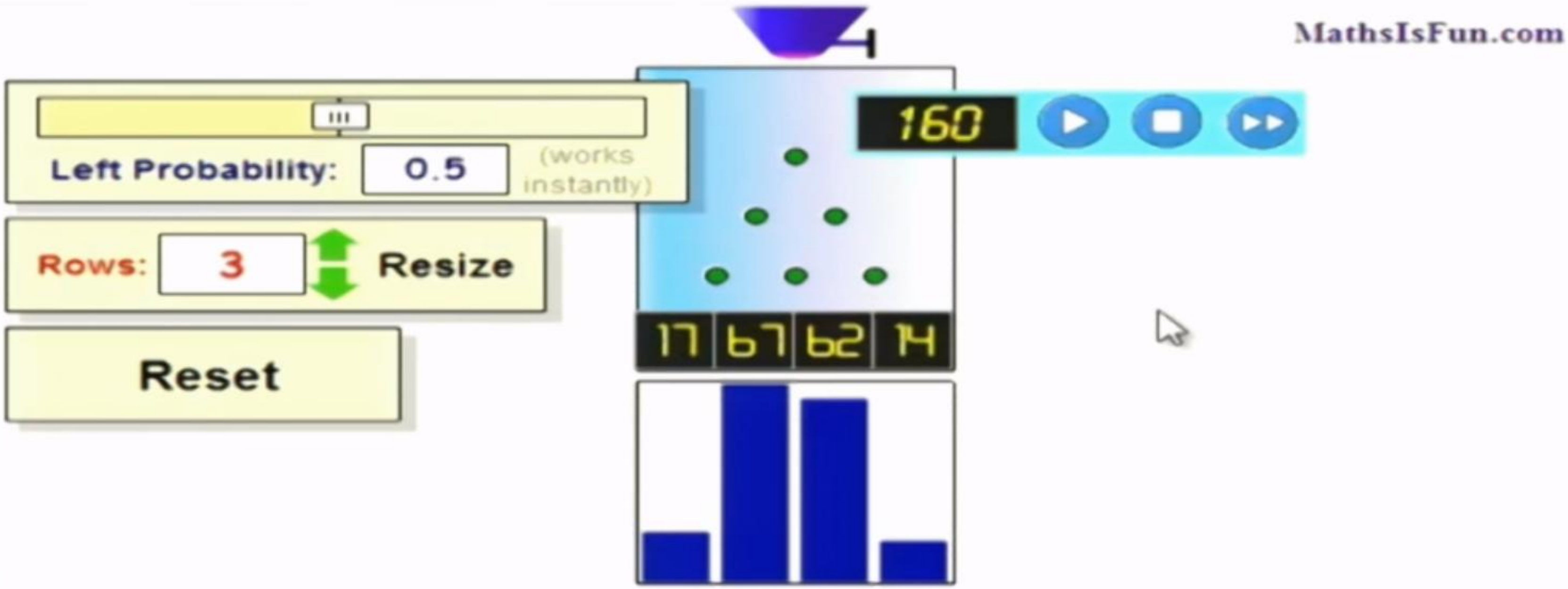
The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



1 3 3 1

Quincunx

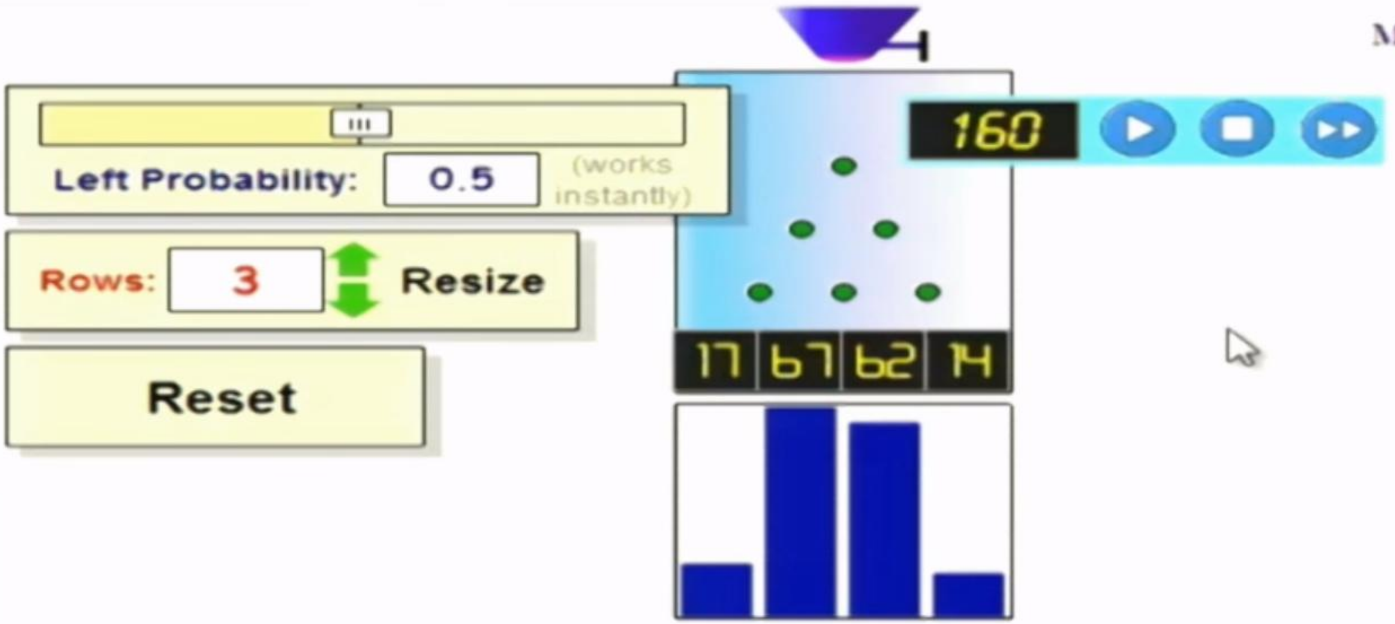
The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



$\frac{1}{8}$ $\frac{3}{8}$ $\frac{3}{8}$ $\frac{1}{3}$

Quincunx

The quincunx is an amazing machine. Pegs and balls and probability!
Have a play, then read the Quincunx Explained.



$\frac{1}{8}$ $\frac{3}{8}$ $\frac{3}{8}$ $\frac{1}{8}$

1

1 1

1 2 1

1 3 3 1

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

$(a+b)^1$

$(a+b)^2 =$

1

1 1 $(a+b)^1$

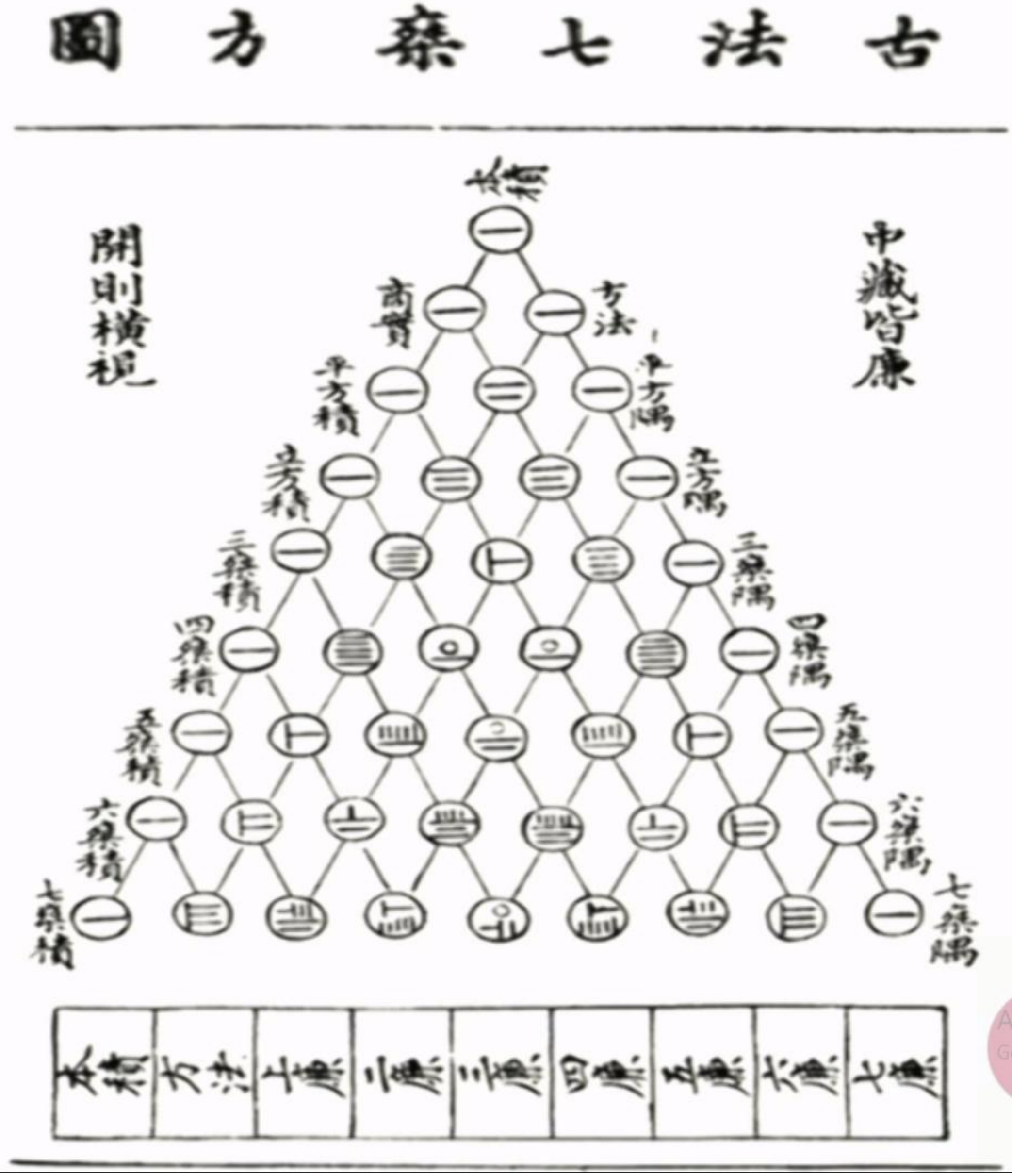
1 2 1 $(a+b)^2 = a^2+2ab+b^2$

1 3 3 1 $(a+b)^3$

1 4 6 4 1 $a+b$

1 5 10 10 5 1

1303 Zhu Shijie
1260-1320
in *Si Yuan Yu Jian*



Factorial notation:

$$1 \times 2 = 2!$$

$$1 \times 2 \times 3 = 3!$$

$$1 \times 2 \times 3 \times 4 = 4!$$

$$\vdots$$

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

Binomial Coefficients :

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!} \qquad n = 1, 2, \ldots$$

$$x = 0, 1, \ldots, n$$

Binomial Coefficients :

$$\boxed{\binom{n}{x}} = \frac{n!}{x!\,(n-x)!} \qquad \boxed{n = 1, 2, \ldots}$$

$$x = 0, 1, \ldots, n$$

Binomial Coefficients :

$$\boxed{\binom{n}{x}} = \frac{n!}{x!\,(n-x)!} \qquad \boxed{n = 1, 2, \ldots}$$

$$x = 0, 1, \ldots, n$$

$\binom{3}{2} = 3$

a b c

a b

a c

b c.

Jakob Bernoulli
1654-1705

Binomial Distribution

A sequence of *independent* *Bernoulli* trials (n) with *constant* probability of success at each trial (p) and we are interested in the total number of successes (x).

Jakob Bernoulli
1654-1705

Binomial Distribution

A sequence of *independent* *Bernoulli* trials (n) with *constant* probability of success at each trial (p) and we are interested in the total number of successes (x).

e.g. In the 4^{th} quarter of 1988 in Mass, of 21,835 births, 60 tested positive for HIV antibodies.

Jakob Bernoulli
1654-1705

Binomial Distribution

A sequence of *independent Bernoulli* trials (n) with *constant* probability of success at each trial (p) and we are interested in the total number of successes (x).

e.g. In the 4^{th} quarter of 1988 in Mass, of 21,835 births, 60 tested positive for HIV antibodies.

How many are infected?

Possible model: binomial with $p \approx 0.25$ and n=60.

Binomial Distribution

X = number of successes

$$P(X) = \binom{n}{X} p^X (1-p)^{n-X} \qquad X = 0, 1, 2, \ldots, n$$

$$n = 1, 2, \ldots$$

Parameters:

p = probability of success
n = number of trials

Binomial Distribution

60

X = number of successes

$$P(X) = \boxed{\binom{n}{X} p^{X}(1-p)^{n-X}} \qquad \boxed{X = 0,1,2,\ldots,n}$$

$$n = 1,2,\ldots$$

.25

Parameters:

p = probability of success

n = number of trials

```
. gen p = binomialp(10,x,.5)

. list x p
```

	x	p
1.	0	.0009766
2.	1	.0097656
3.	2	.0439453
4.	3	.1171875
5.	4	.2050781
6.	5	.2460938
7.	6	.2050781
8.	7	.1171875
9.	8	.0439453
10.	9	.0097656
11.	10	.0009766

```
. gen p = binomialp(10,x,.5)

. list x p
```

n p

	x	p
1.	0	.0009766
2.	1	.0097656
3.	2	.0439453
4.	3	.1171875
5.	4	.2050781
6.	5	.2460938
7.	6	.2050781
8.	7	.1171875
9.	8	.0439453
10.	9	.0097656
11.	10	.0009766

```
. gen p = binomialp(10,x,.5)

. list x p
```

n p

	x	p
1.	0	.0009766
2.	1	.0097656
3.	2	.0439453
4.	3	.1171875
5.	4	.2050781
6.	5	.2460938
7.	6	.2050781
8.	7	.1171875
9.	8	.0439453
10.	9	.0097656
11.	10	.0009766

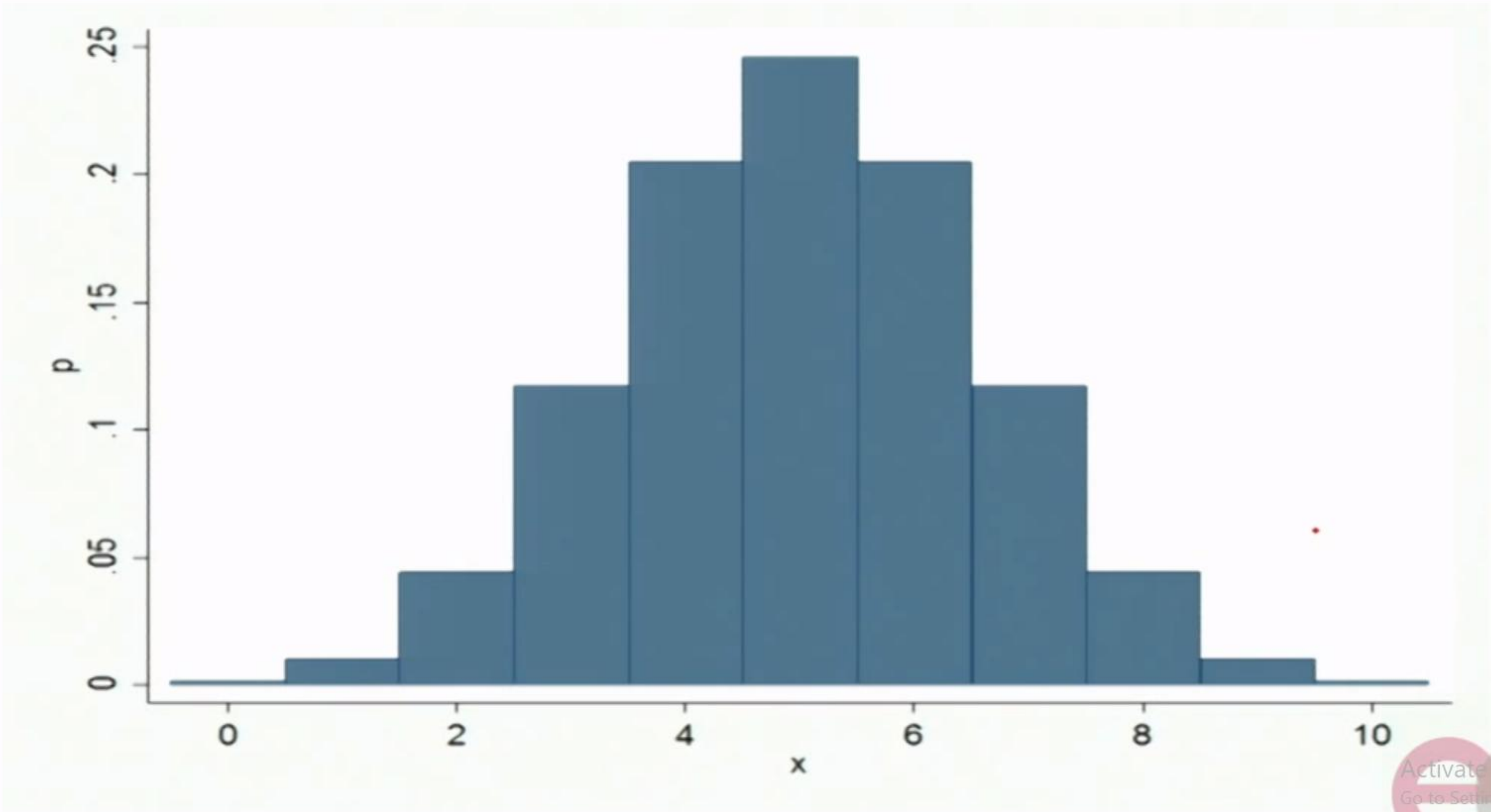
```
. gen p = binomialp(10,x,.5)

. list x p
```

n

p

	x	p
1.	0	.0009766
2.	1	.0097656
3.	2	.0439453
4.	3	.1171875
5.	4	.2050781
6.	5	.2460938
7.	6	.2050781
8.	7	.1171875
9.	8	.0439453
10.	9	.0097656
11.	10	.0009766

Binomial with n=10 and p=0.5



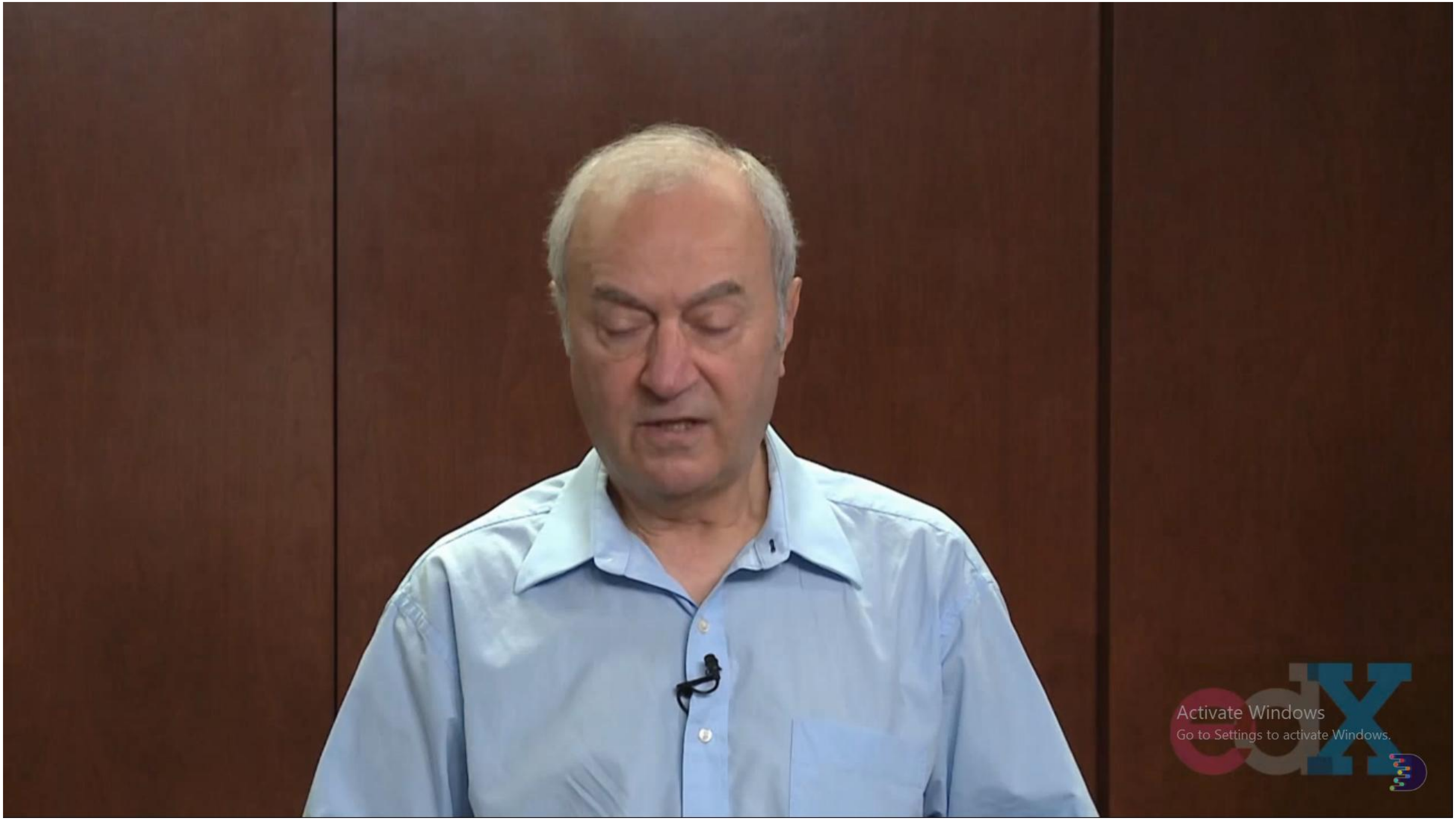
```
. gen p = binomialp(10,x,.2)

. list x p
```

	x	p
1.	0	.1073742
2.	1	.2684354
3.	2	.3019899
4.	3	.2013266
5.	4	.0880804
6.	5	.0264241
7.	6	.005505
8.	7	.0007864
9.	8	.0000737
10.	9	4.10e-06
11.	10	1.02e-07

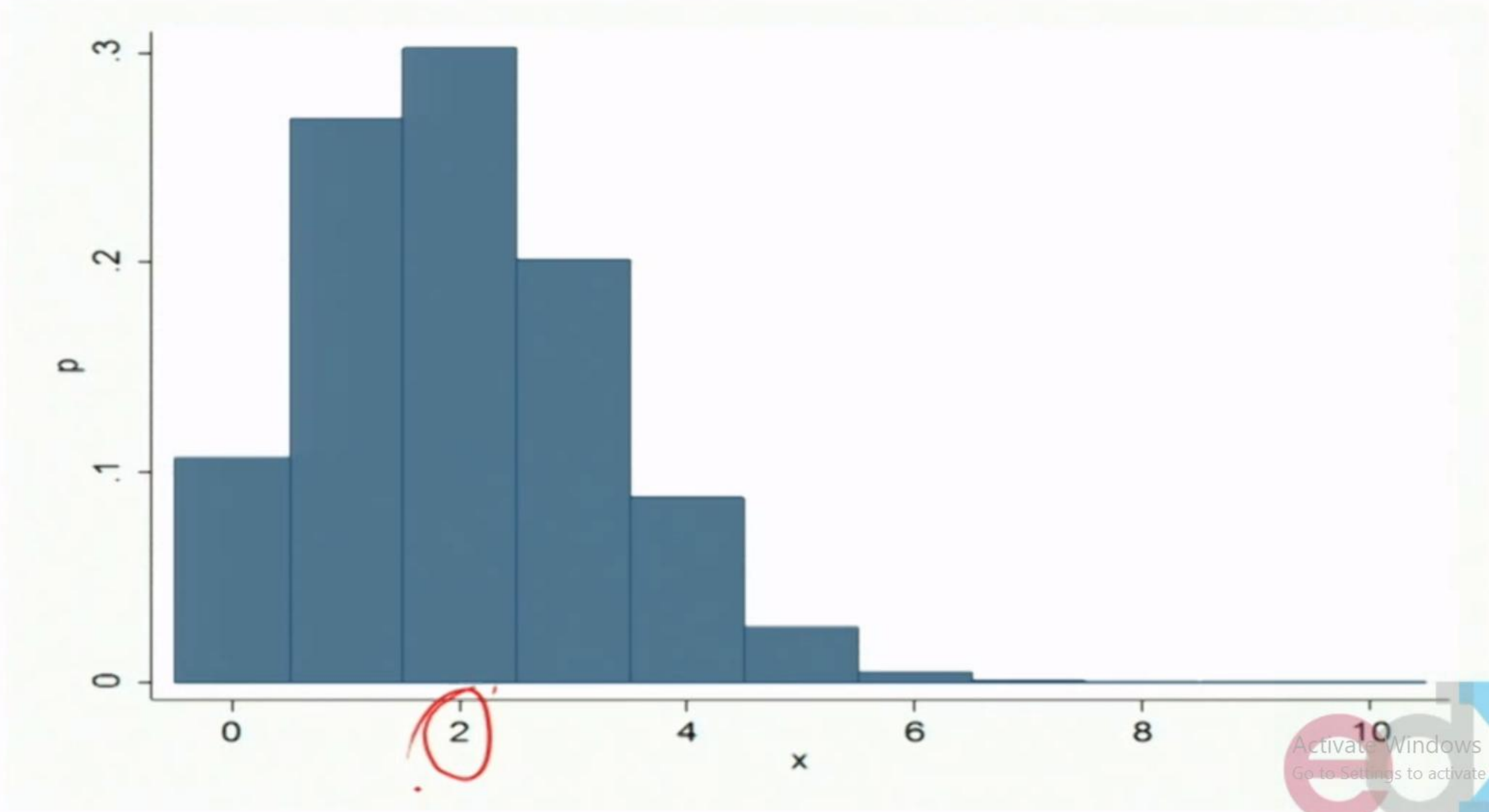
```
. gen p = binomialp(10,x,.2)

. list x p
```

	x	p
1.	0	.1073742
2.	1	.2684354
3.	2	.3019899
4.	3	.2013266
5.	4	.0880804
6.	5	.0264241
7.	6	.005505
8.	7	.0007864
9.	8	.0000737
10.	9	4.10e-06
11.	10	1.02e-07

$10 \times .2 =$

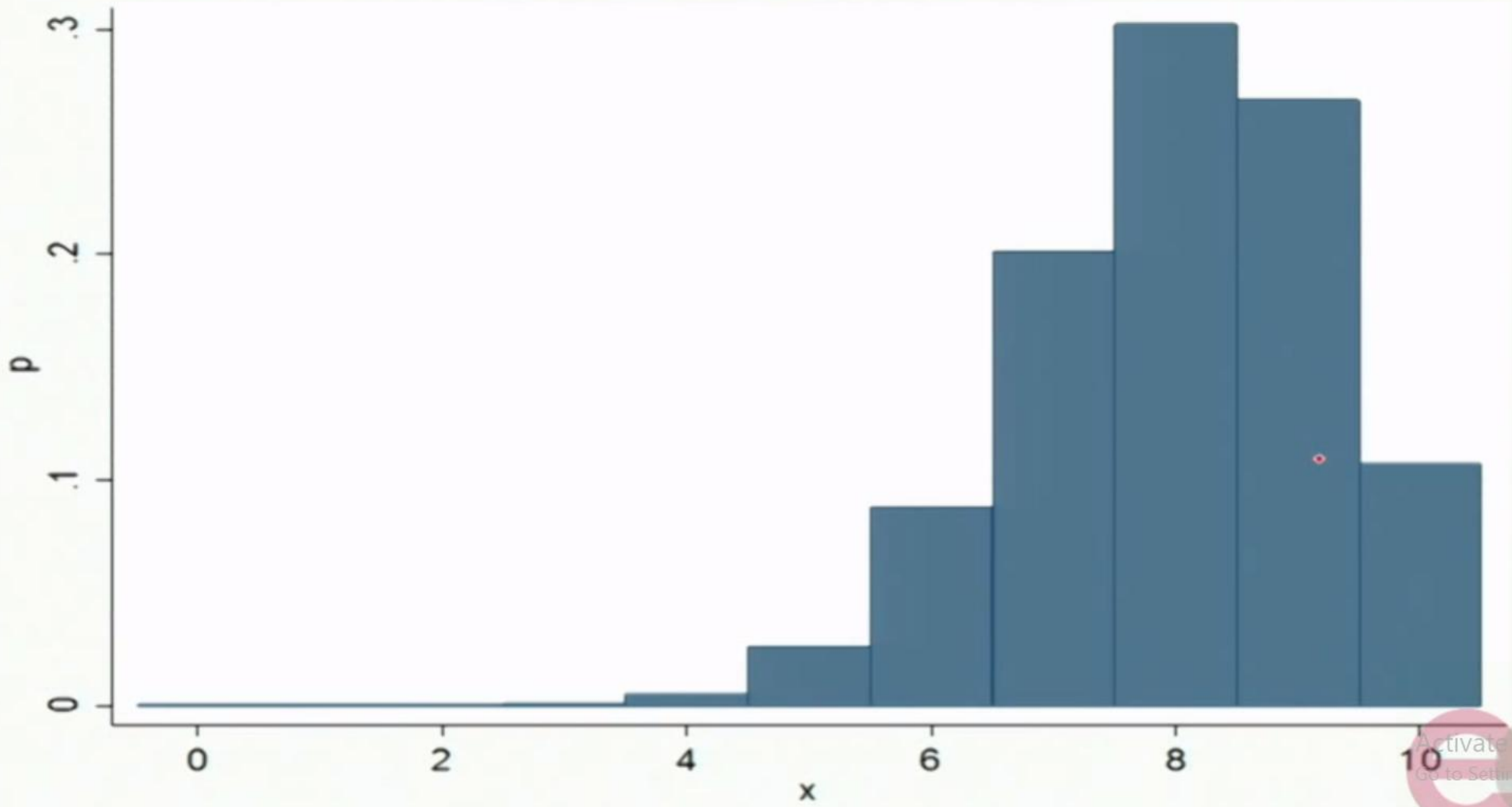


Binomial with n=10 and p=0.2

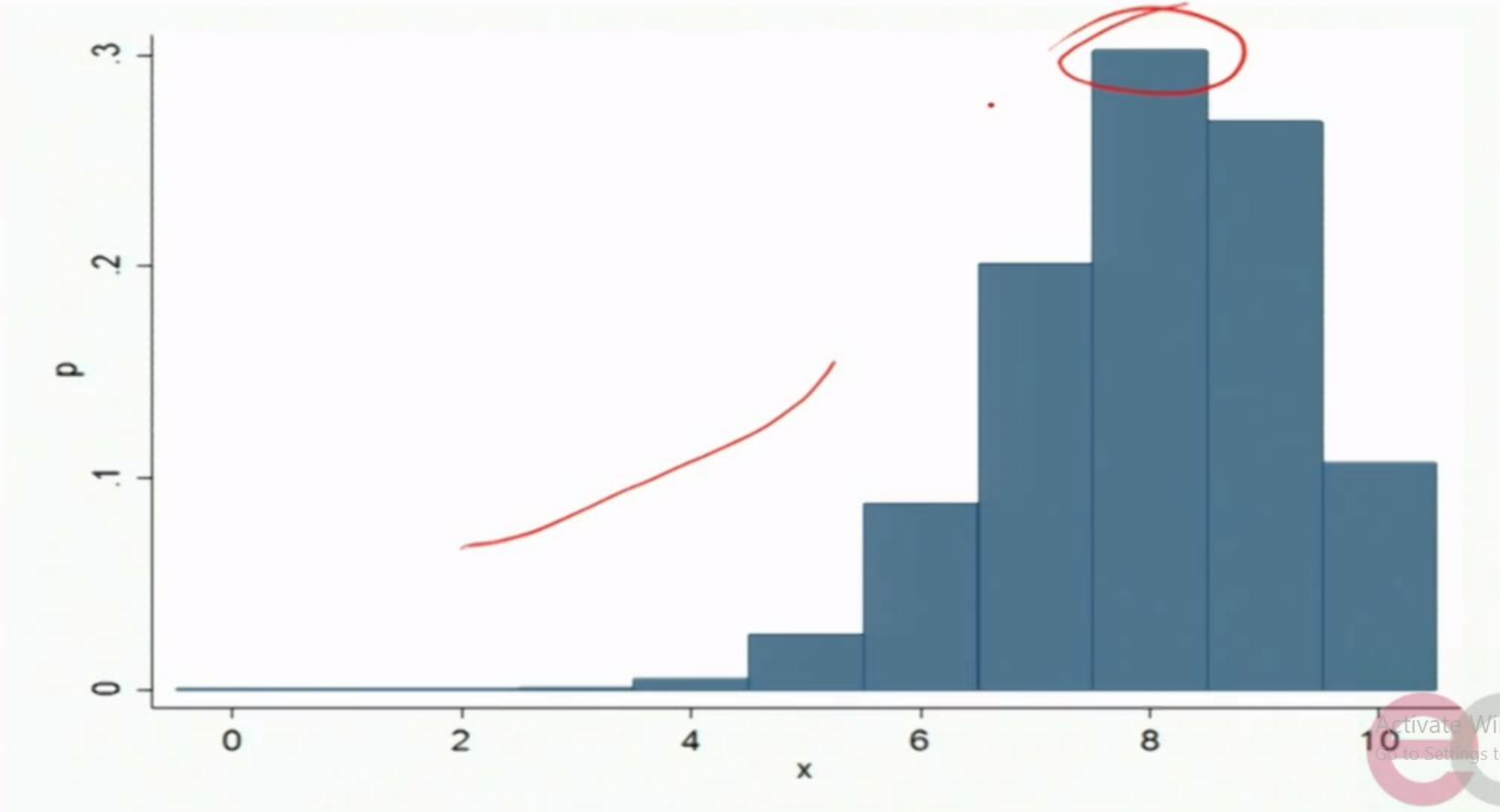


```
. gen p = binomialp(10,x,.8)

. list x p

     +----------------+
     |  x           p |
     |----------------|
  1. |  0    1.02e-07 |
  2. |  1    4.10e-06 |
  3. |  2    .0000737 |
  4. |  3    .0007864 |
  5. |  4     .005505 |
     |----------------|
  6. |  5    .0264241 |
  7. |  6    .0880804 |
  8. |  7    .2013266 |
  9. |  8    .3019899 |
 10. |  9    .2684354 |
     |----------------|
 11. | 10    .1073742 |
     +----------------+
```

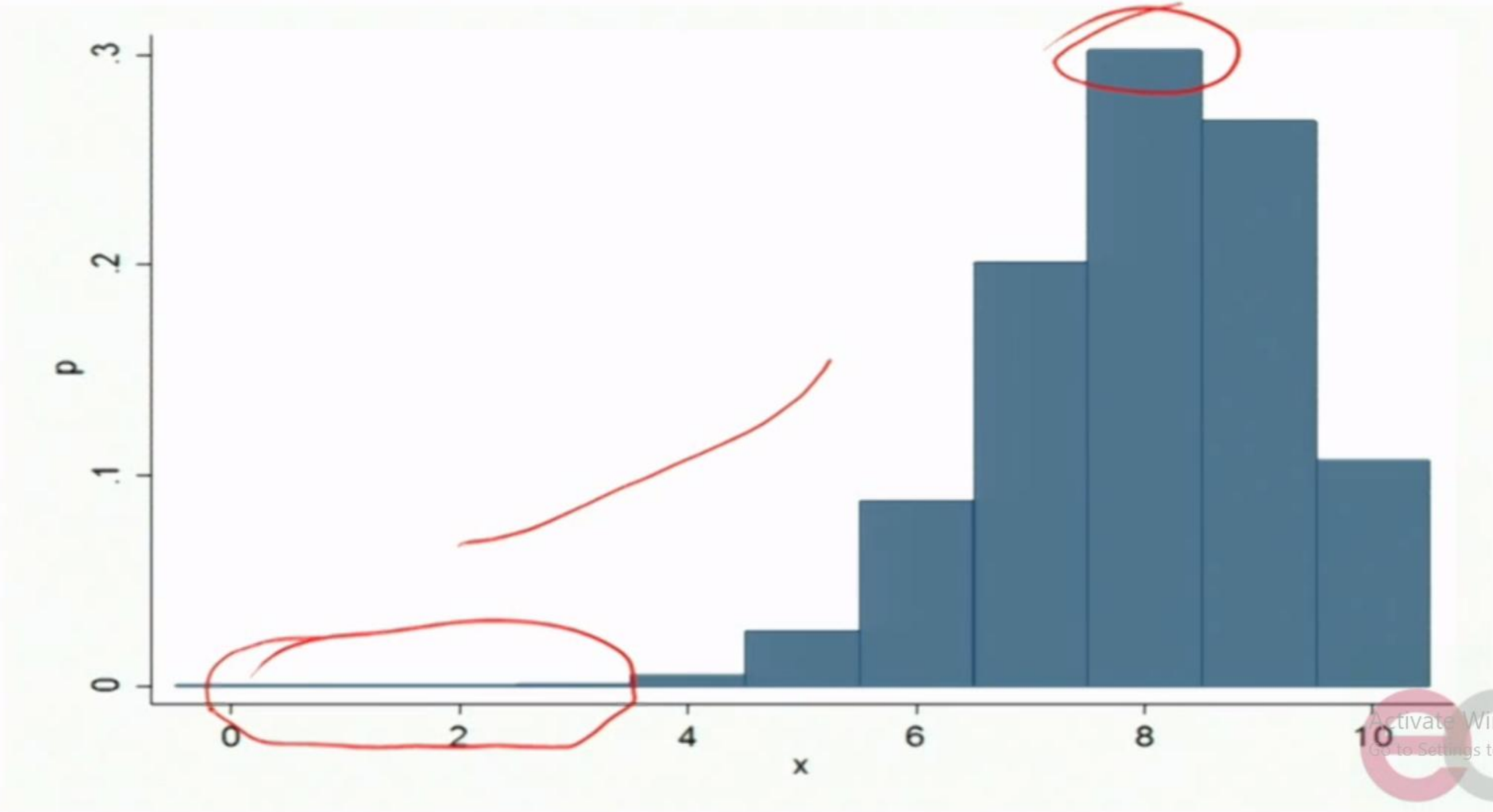
Binomial with n=10 and p=0.8



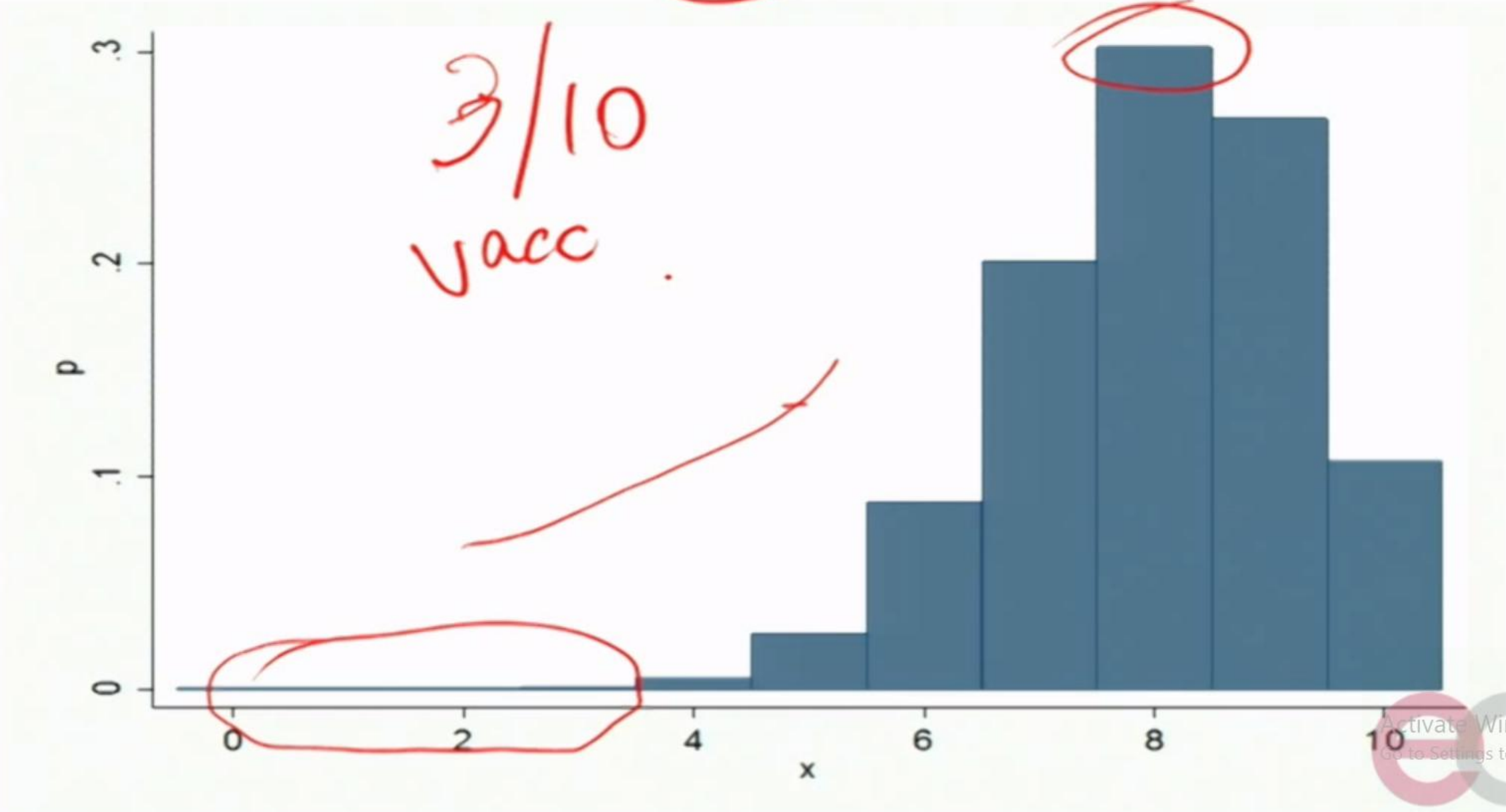
Binomial with n=10 and p=0.8



Binomial with n=10 and p=0.8



Binomial with n=10 and p=0.8



For Binomial with n & p

Then

Mean = np

Stand. Dev. = $\sqrt{np(1-p)}$

e.g. $p=0.5 \quad n=10$

$$\text{Mean} = np = 10 \times 0.5 = 5$$

$$\text{Stand. Dev.} = \sqrt{10 \times 0.5 \times 0.5} = 1.6$$

$$5 \pm 1.6 = (3.4, 6.6)$$

$$5 \pm 3.2 = (1.8, 8.2)$$

e.g. p=0.5 n=10

$$\text{Mean} = np = 10 \times 0.5 = 5$$

$$\text{Stand. Dev.} = \sqrt{10 \times 0.5 \times 0.5} = 1.6$$

$$5 \pm 1.6 = (3.4, 6.6)$$

$$5 \pm 3.2 = (1.8, 8.2)$$

For Binomial with n & p

Then

Mean = np

Stand. Dev. = $\sqrt{np(1-p)}$ ← $\frac{\sqrt{n}}{2}$

e.g. $\boxed{p=0.25 \quad n=60}$

$$\text{Mean} = np = 60 \times 0.25 = 15$$

$$\text{Stand. Dev.} = \sqrt{60 \times 0.25 \times 0.75} = 3.4$$

$$15 \pm 3.4 = (11.6, 18.4)$$

$$15 \pm 6.8 = (8.2, 21.8)$$

Binomial Distribution

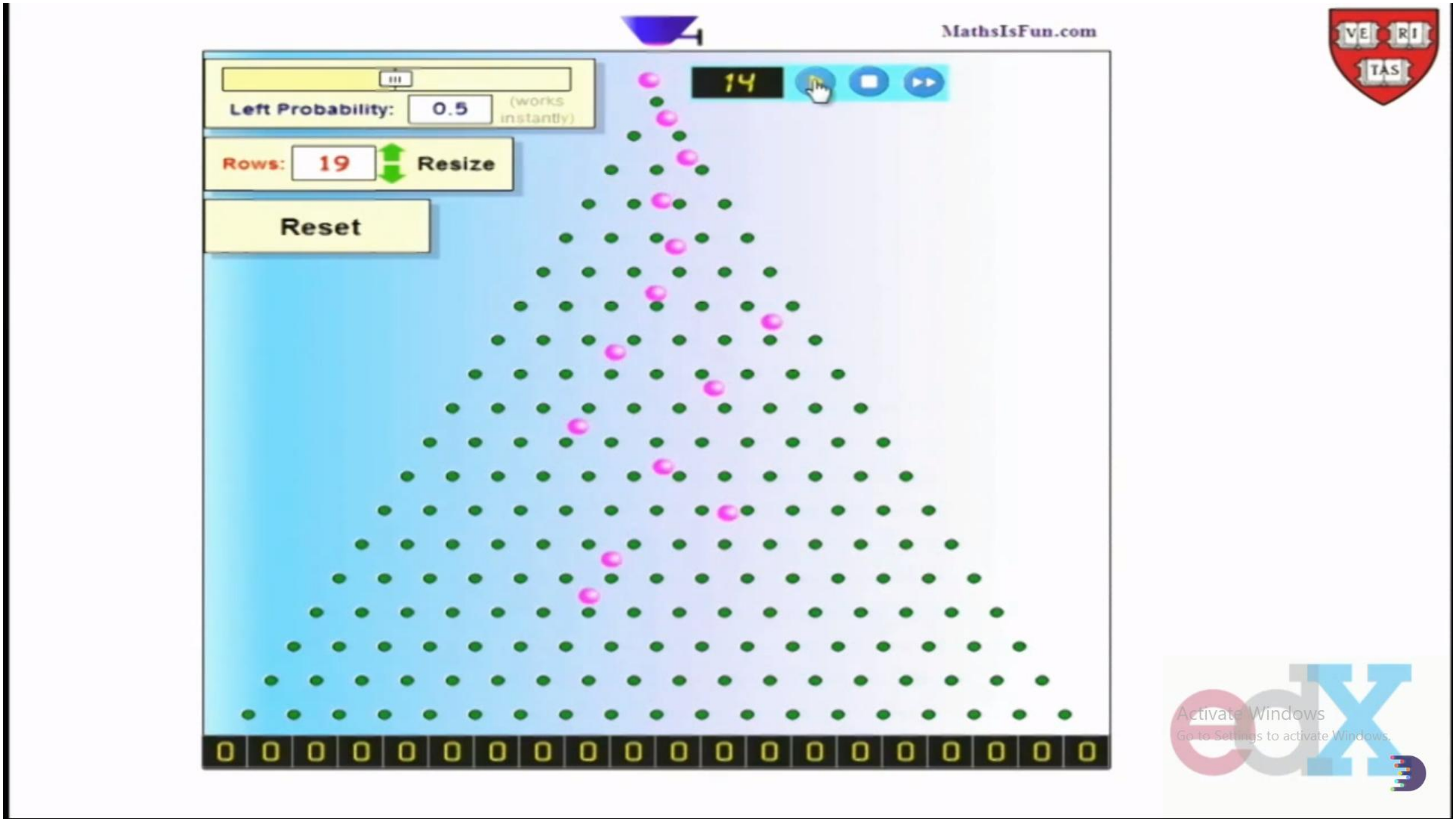
X = number of successes

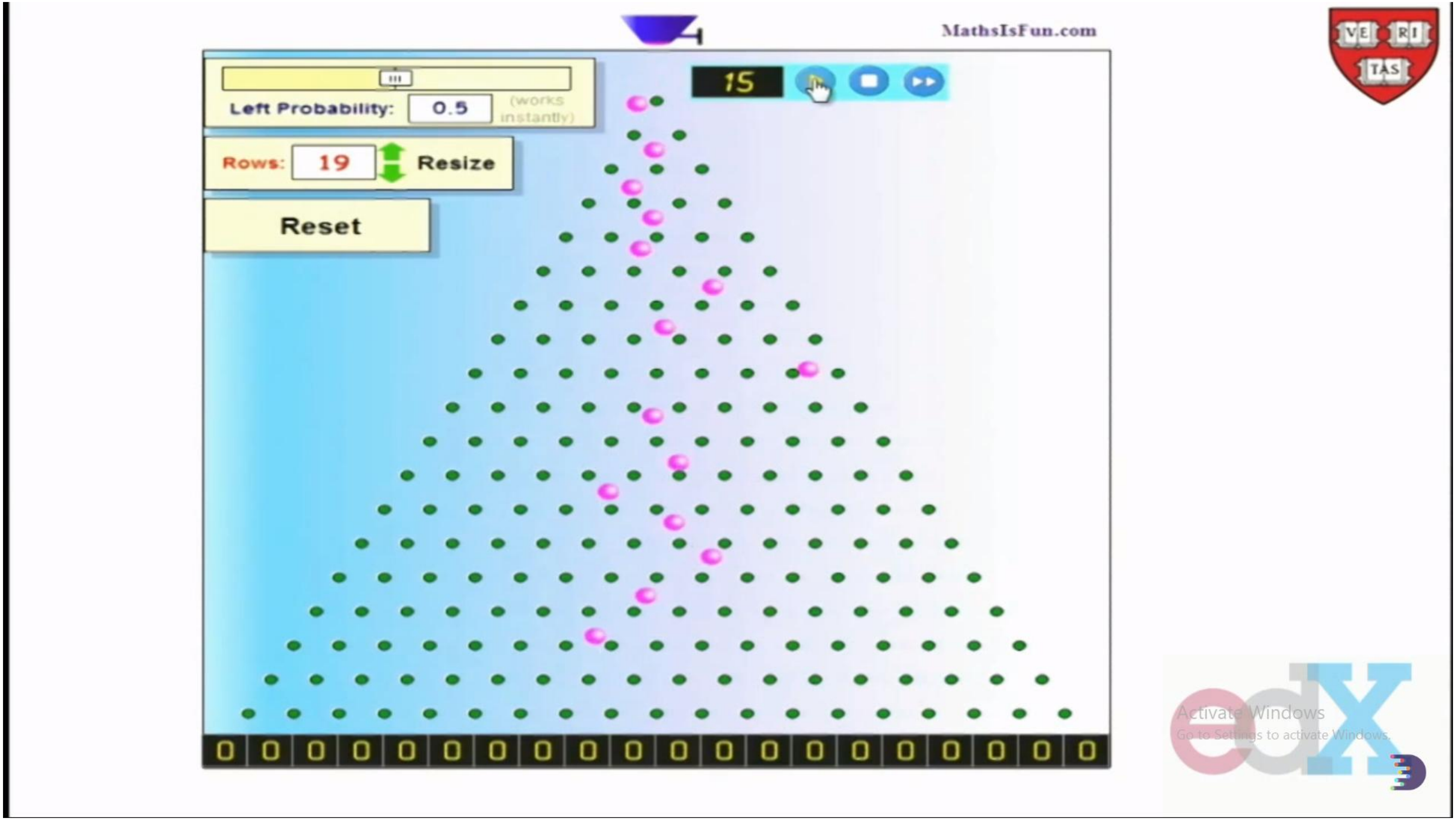
$$P(X) = \binom{n}{X} p^X (1-p)^{n-X} \qquad X = 0, 1, 2, \ldots, n$$

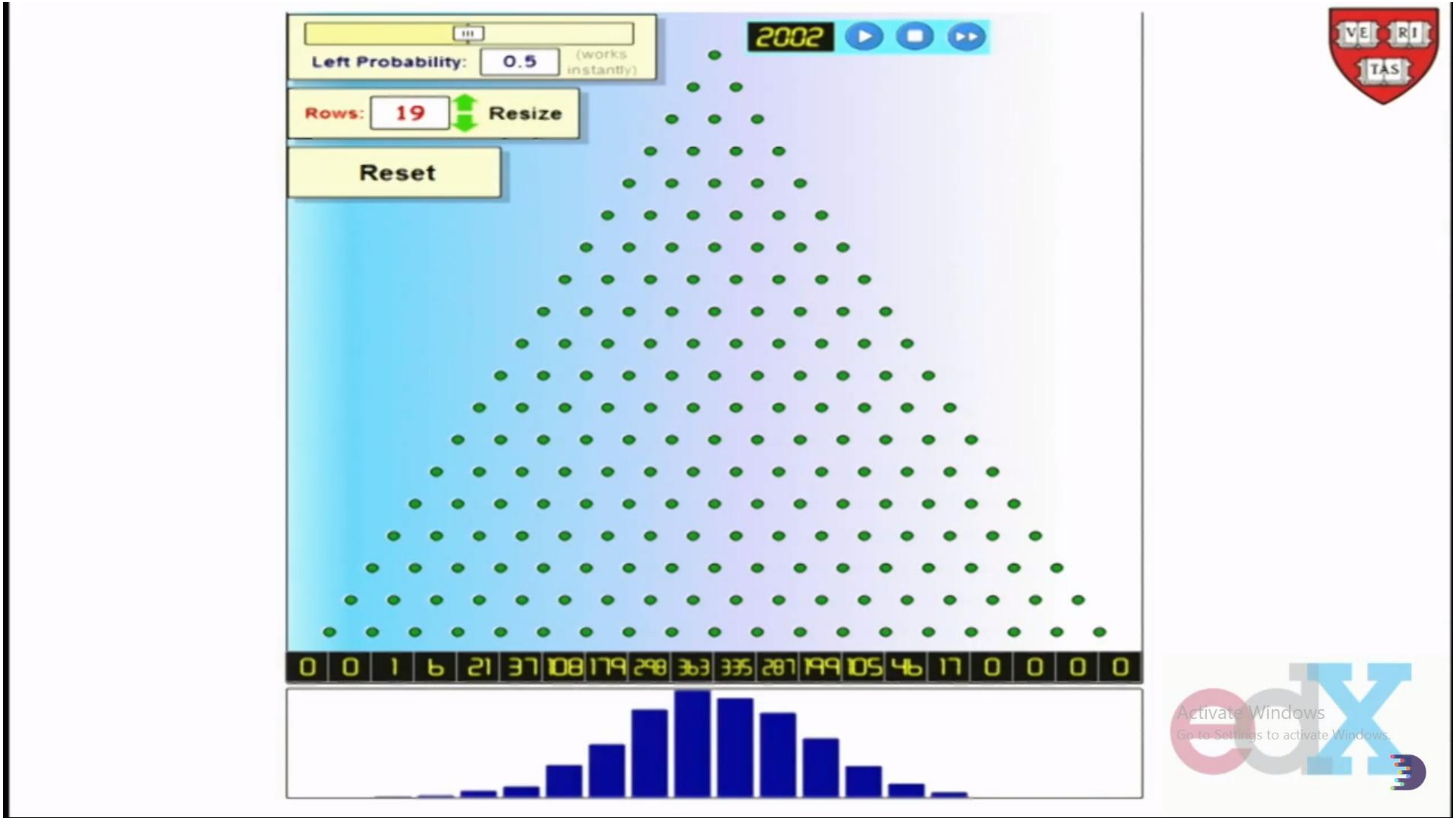
$$n = 1, 2, \ldots$$

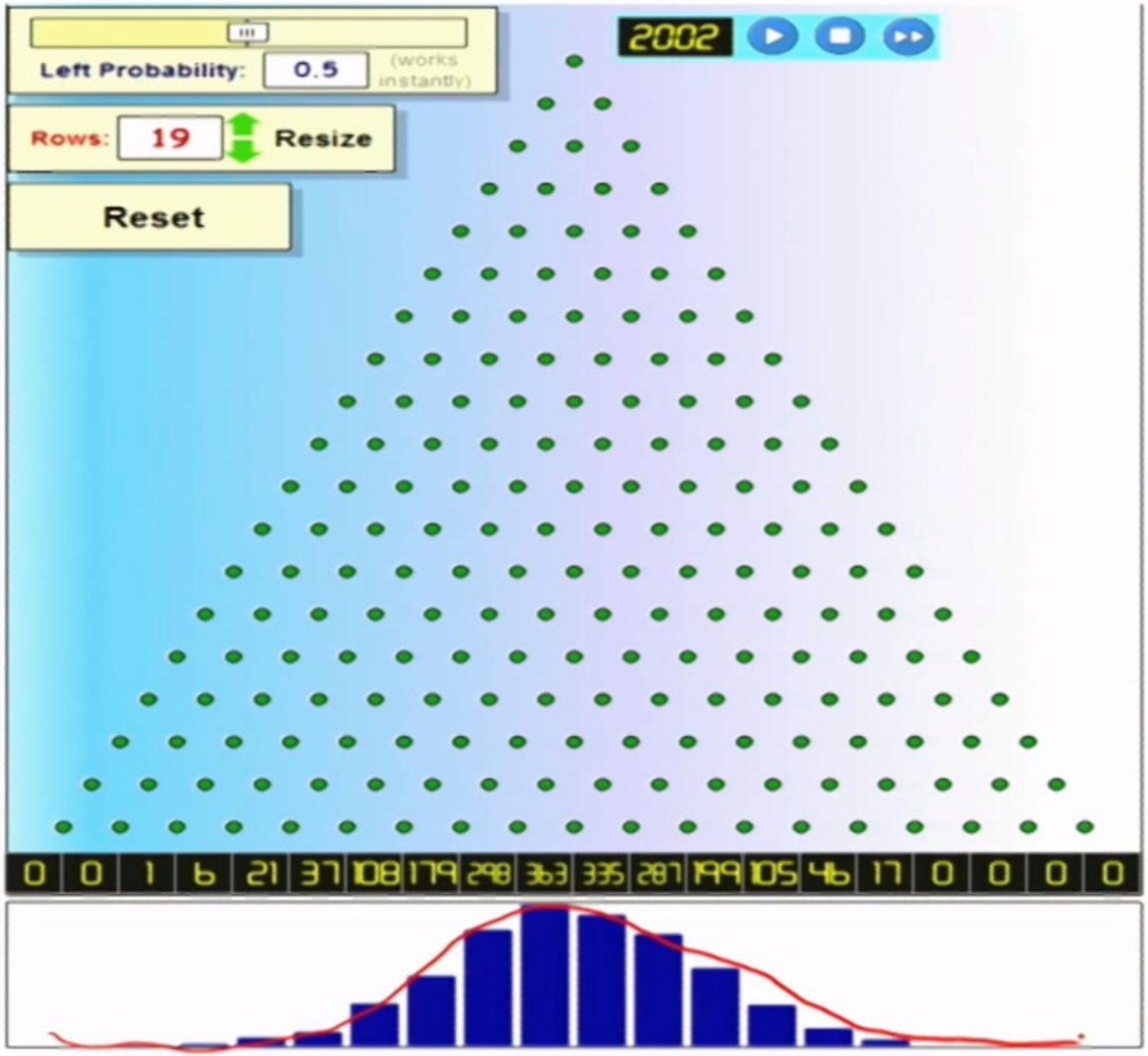
Parameters:

p = probability of success
n = number of trials


MathsIsFun.com
Left Probability: 0.5
(works instantly)
Rows: 19
Resize
Reset
14
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
VE RI TAS


MathsIsFun.com
Left Probability: 0.5
(works instantly)
Rows: 19
Resize
Reset
15
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
VE RI TAS


Left Probability: 0.5 (works instantly)
Rows: 19
Resize
Reset
2002
0 0 1 6 21 37 108 179 298 363 335 287 199 105 46 17 0 0 0 0


2002
Left Probability:
0.5
(works instantly)
Rows:
19
Resize
Reset
0 0 1 6 21 37 108 179 298 363 335 281 199 105 46 17 0 0 0 0

Binomial Distribution

X = number of successes

$$P(X) = \binom{n}{X} p^X (1-p)^{n-X} \qquad X = 0,1,2,\ldots,n$$

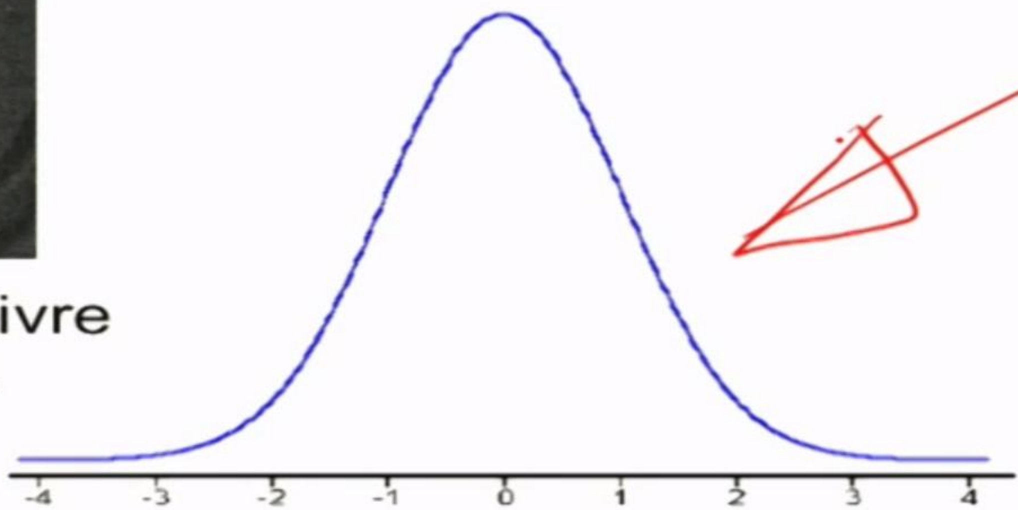
$$n = 1,2,\ldots$$

$$\frac{X-\mu}{\sigma} = \frac{X-np}{\sqrt{np(1-p)}} \text{ approx. Normal}$$

Continuous Random Variables

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \qquad -\infty < x < \infty$$

Abraham de Moivre
1667 – 1754

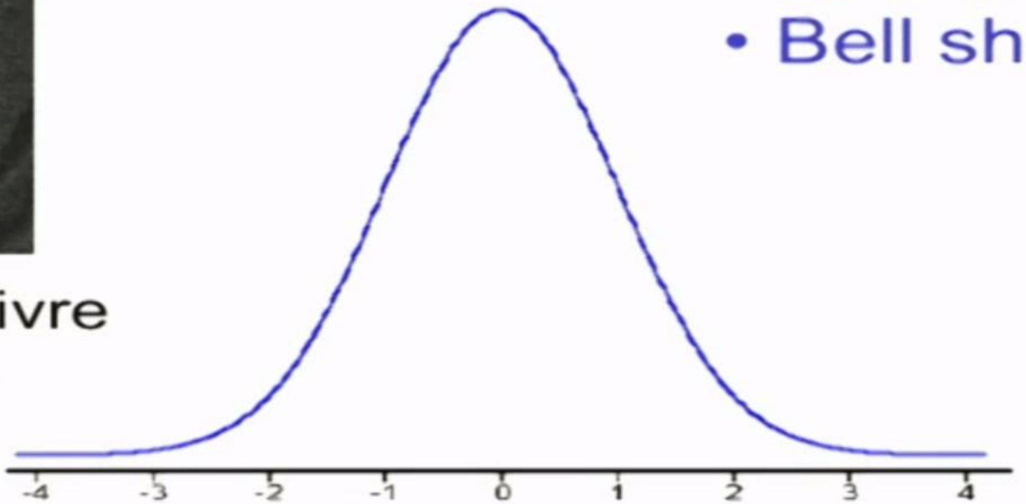


Continuous Random Variables

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \qquad -\infty < x < \infty$$

Abraham de Moivre
1667 – 1754

- Normal Distribution
- Gaussian (De Moivre)
- Law of errors
- Bell shaped

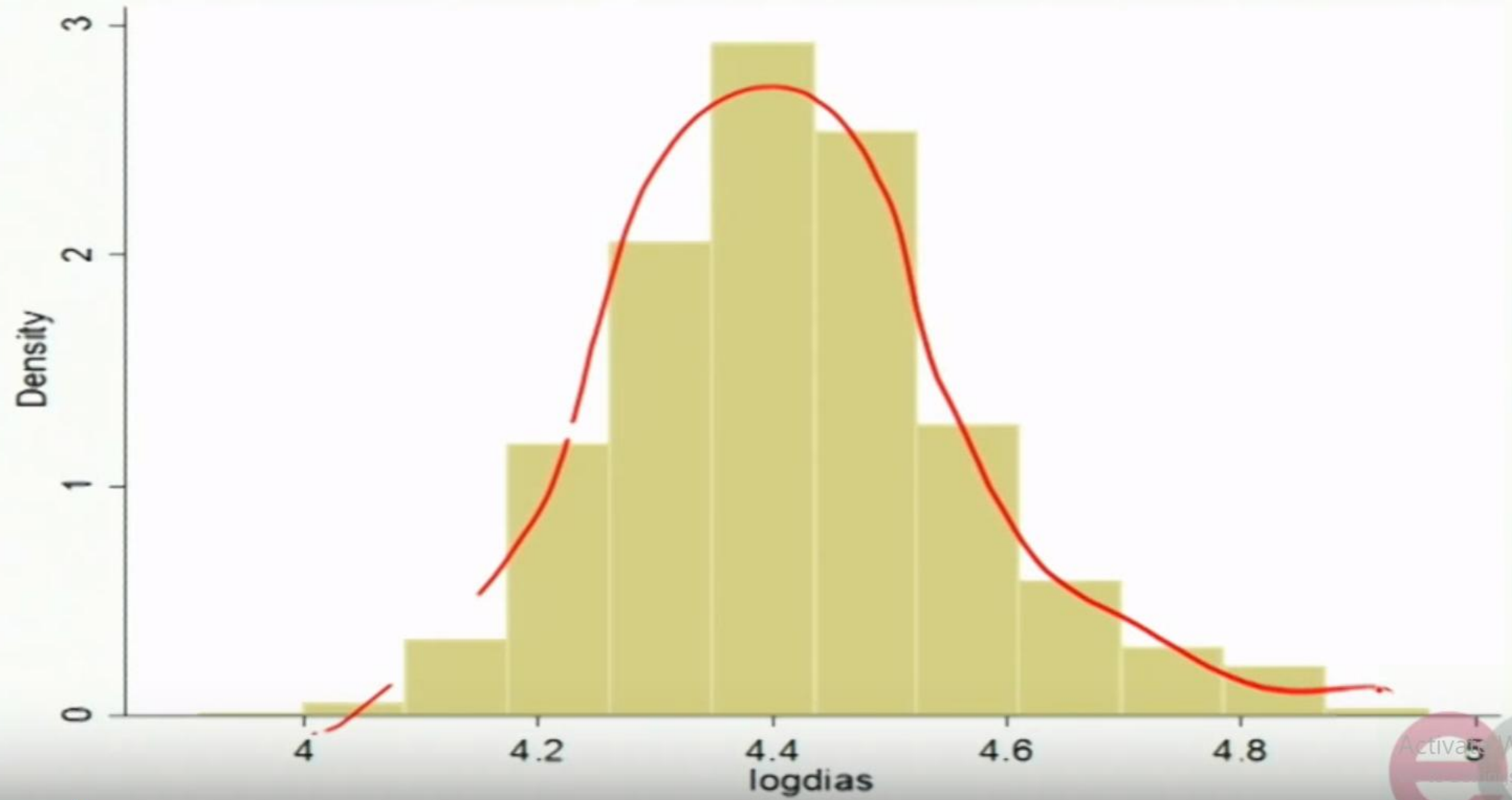


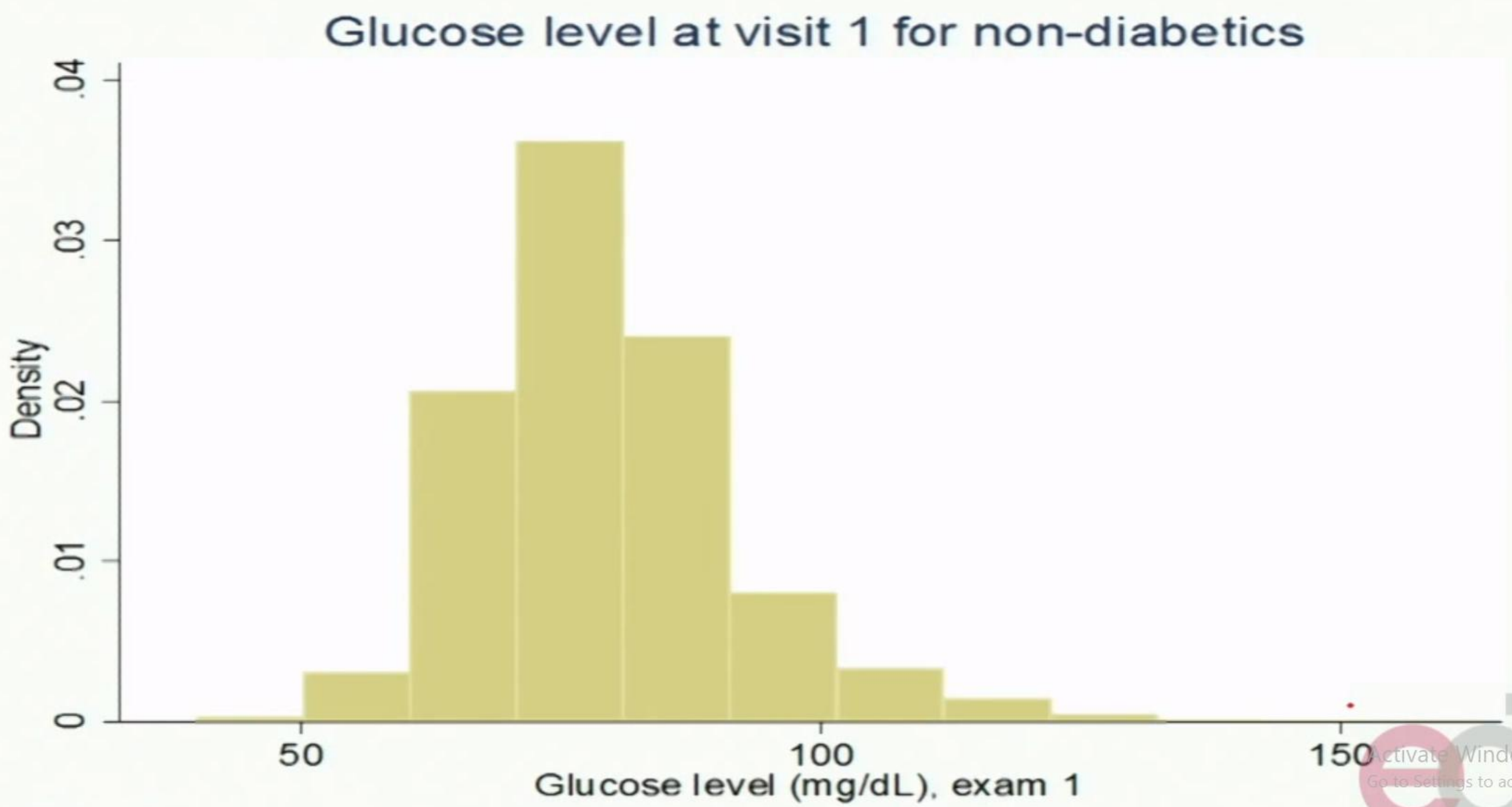

GD9674175N9
DEUTSCHE BUNDESBANK
Banknote
10
1777–1855 Carl Friedr. Gauß
ZEHN DEUTSCHE MARK
GD9674175N9
Deutsche Bundesbank
Frankfurt am Main
1.Oktober 1993

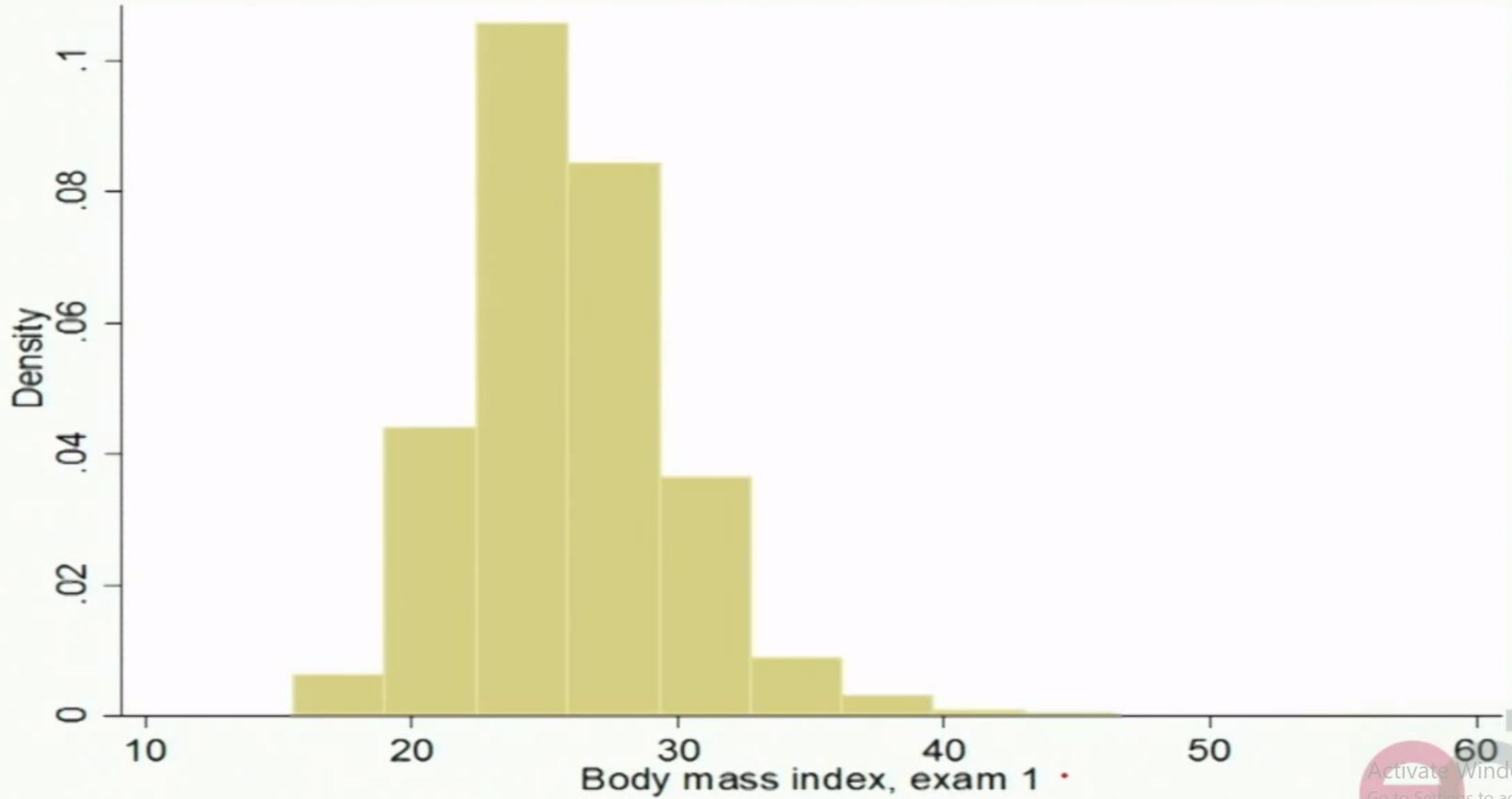

GD9674175N9
DEUTSCHE BUNDESBANK
Banknote
10
1777–1855 Carl Friedr. Gauß
ZEHN DEUTSCHE MARK
10
GD9674175N9
Deutsche Bundesbank
Frankfurt am Main
1.Oktober 1993

Week 4 : THE NORMAL DISTRIBUTION

Histogram of logarithm of diastolic blood pressure at visit one (fhs) (bins = 12)



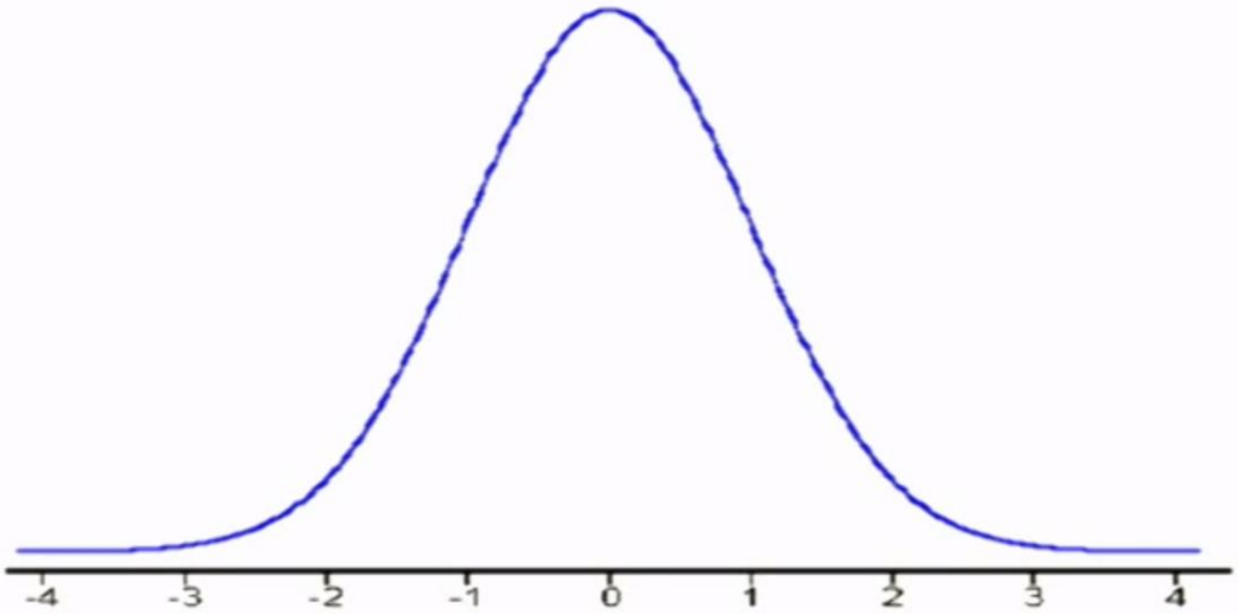

Glucose level at visit 1 for non-diabetics
Density
0
.01
.02
.03
.04
50
100
150
Glucose level (mg/dL), exam 1



Density
0
.02
.04
.06
.08
.1
10
20
30
40
50
60
Body mass index, exam 1

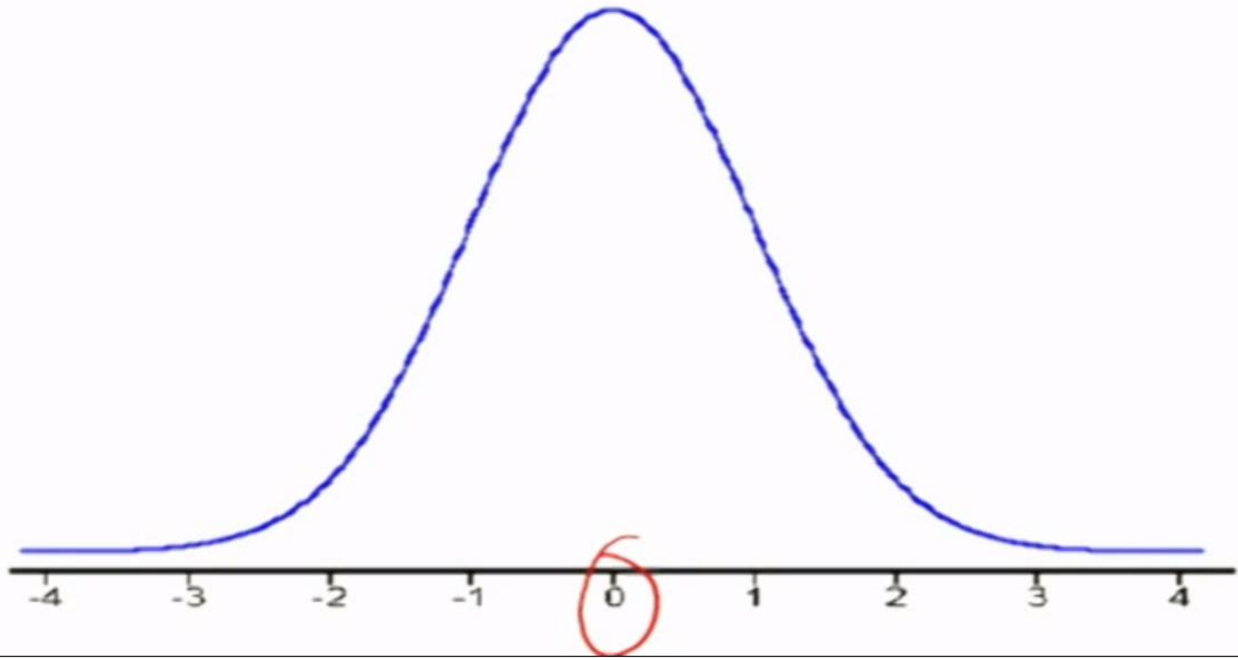
Normal Distribution

$$\boxed{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}} \quad -\infty < x < \infty$$



Normal Distribution

$$\boxed{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}} \quad -\infty < x < \infty$$

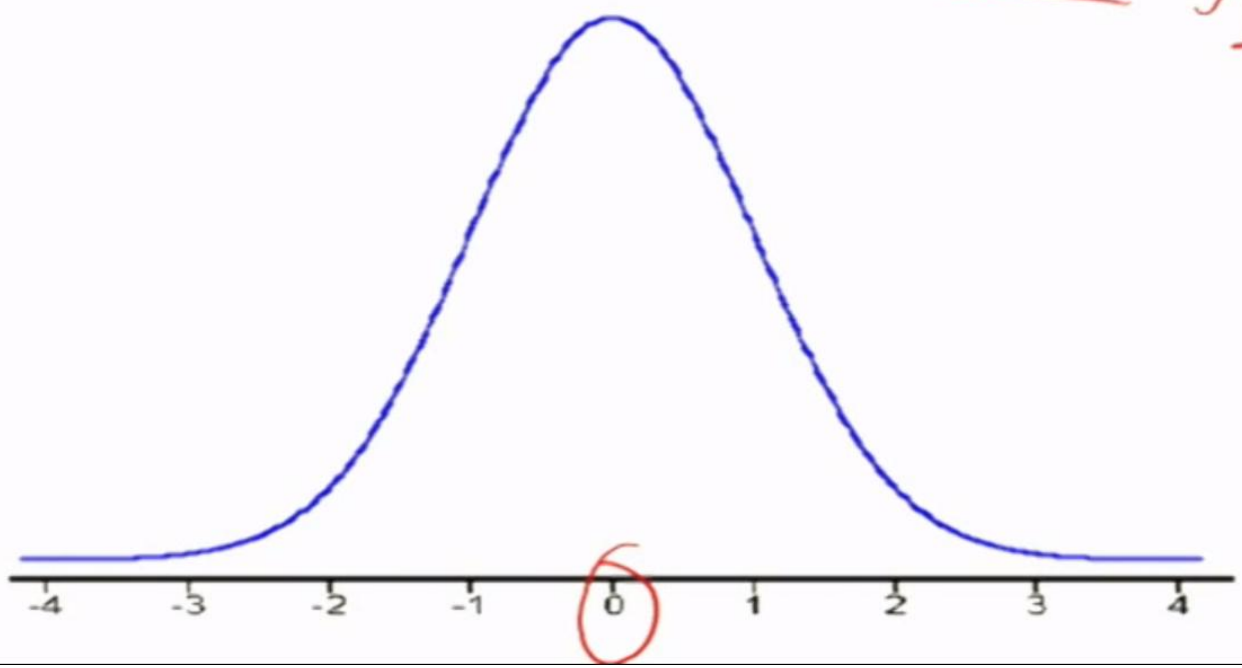


Normal Distribution

$$\boxed{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}} \qquad -\infty < x < \infty$$

0

1



Properties:

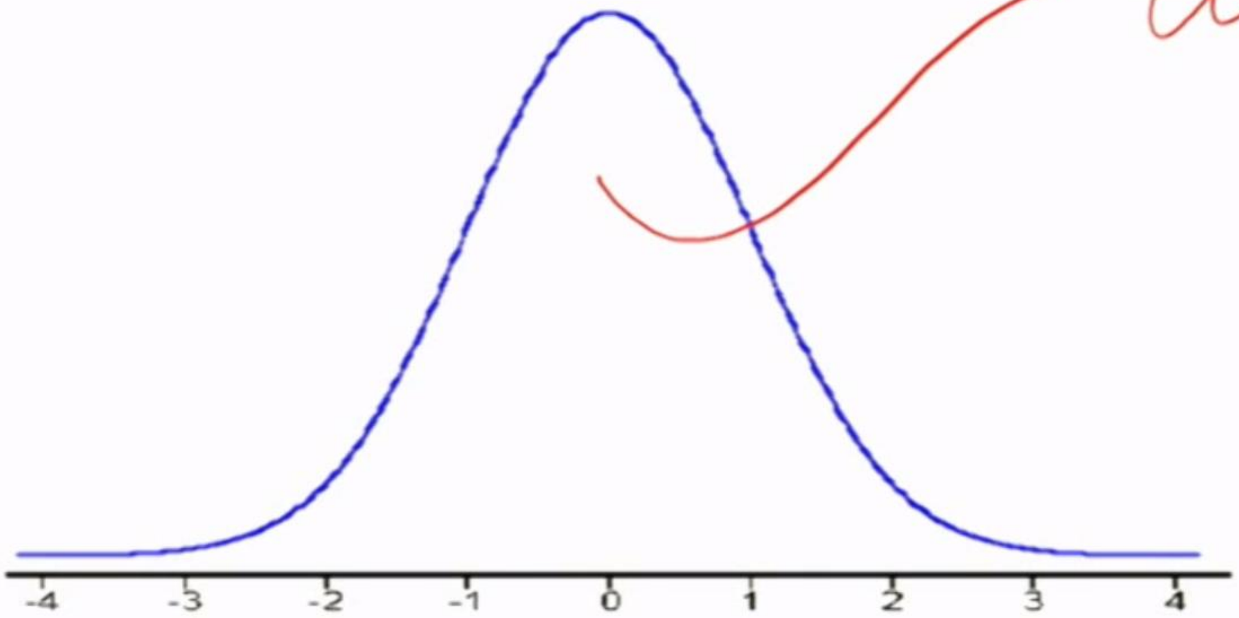
- symmetric about μ
- spread determined by σ
- "Standard Normal", with $\mu = 0$ and $\sigma = 1$, has been tabulated.

For example, with z=1.96 the area to the right, or probability, is 0.025.

Normal Distribution

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \qquad -\infty < x < \infty$$

area =

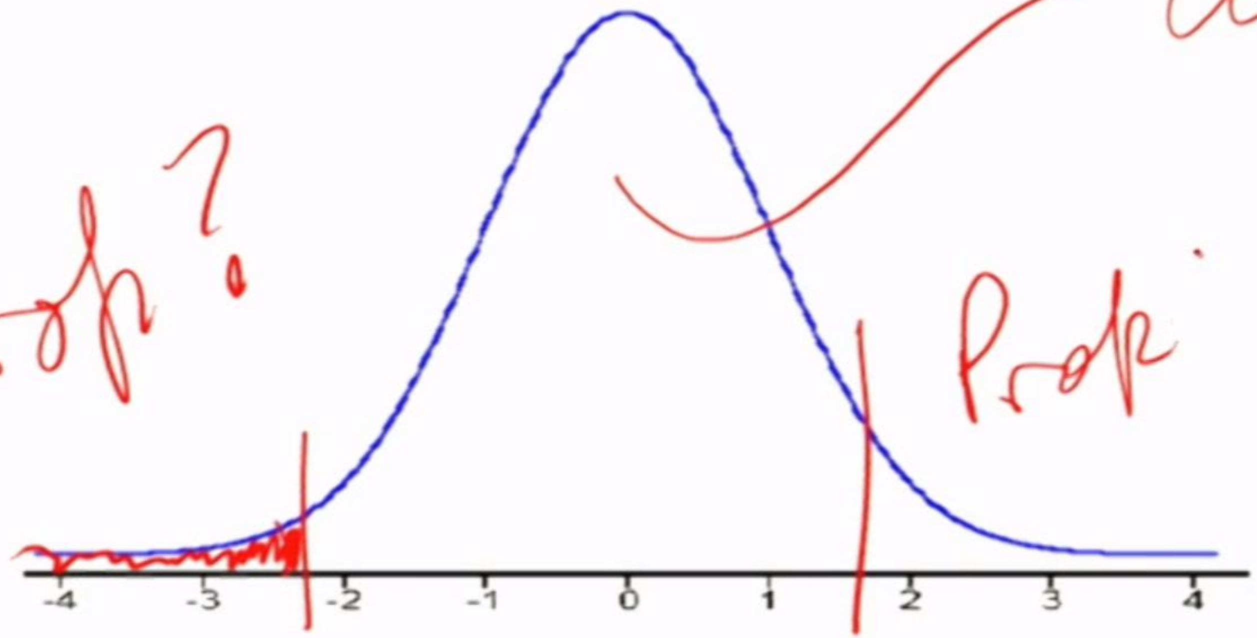


Normal Distribution

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \qquad -\infty < x < \infty$$

area = 1

Prob?



Prob

Properties:

- symmetric about μ
- spread determined by σ
- "Standard Normal", with $\mu = 0$ and $\sigma = 1$, has been tabulated.

For example, with z=1.96 the area to the right, or probability, is 0.025.

(2.5%

Properties:

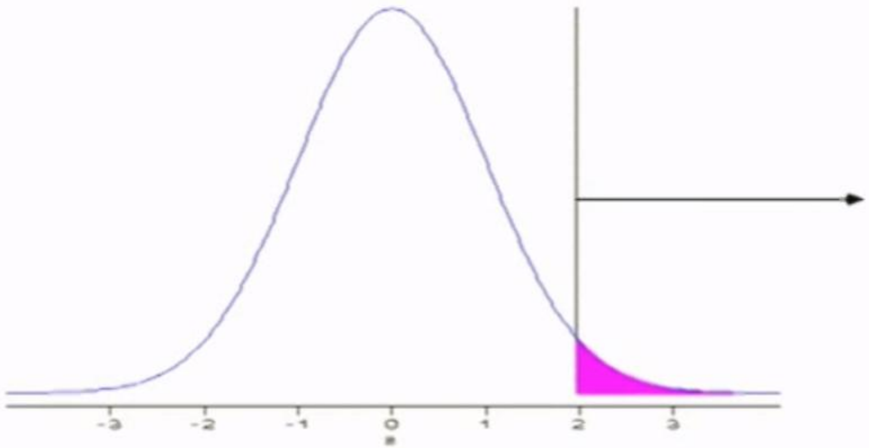
- symmetric about μ
- spread determined by σ
- "Standard Normal", with $\mu = 0$ and $\sigma = 1$, has been tabulated.

For example, with z=1.96 the area to the right, or probability, is 0.025.

≈ 2

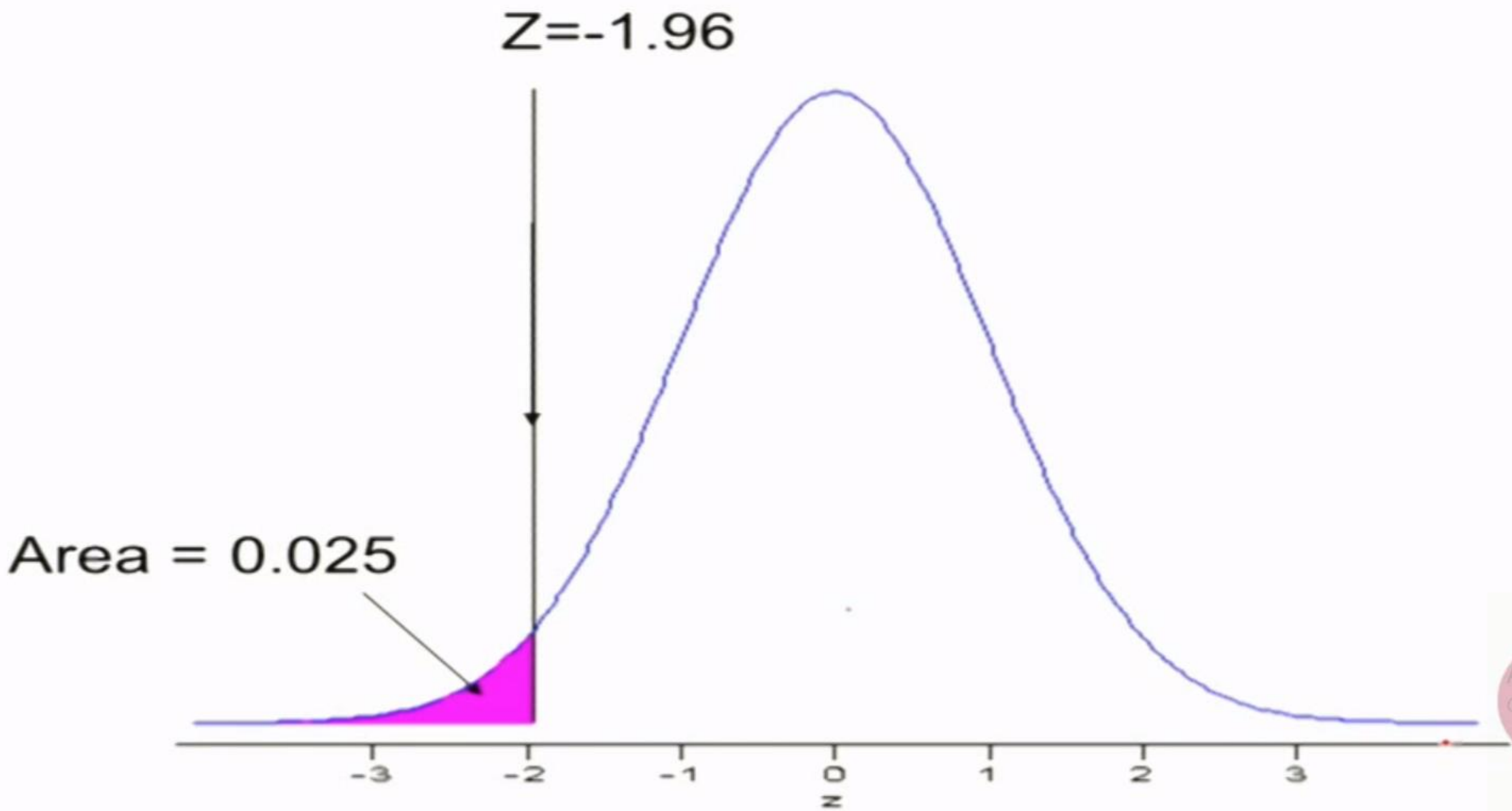
2.5%

Standard Normal

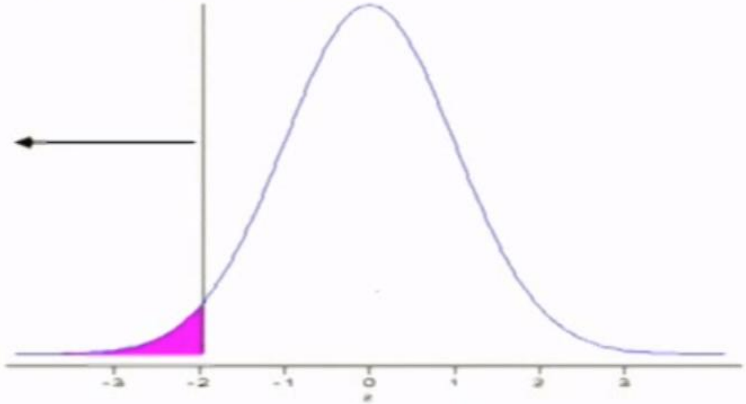


z	area
0	0.5
1.65	0.049
1.96	0.025
2.58	0.005
3	0.001

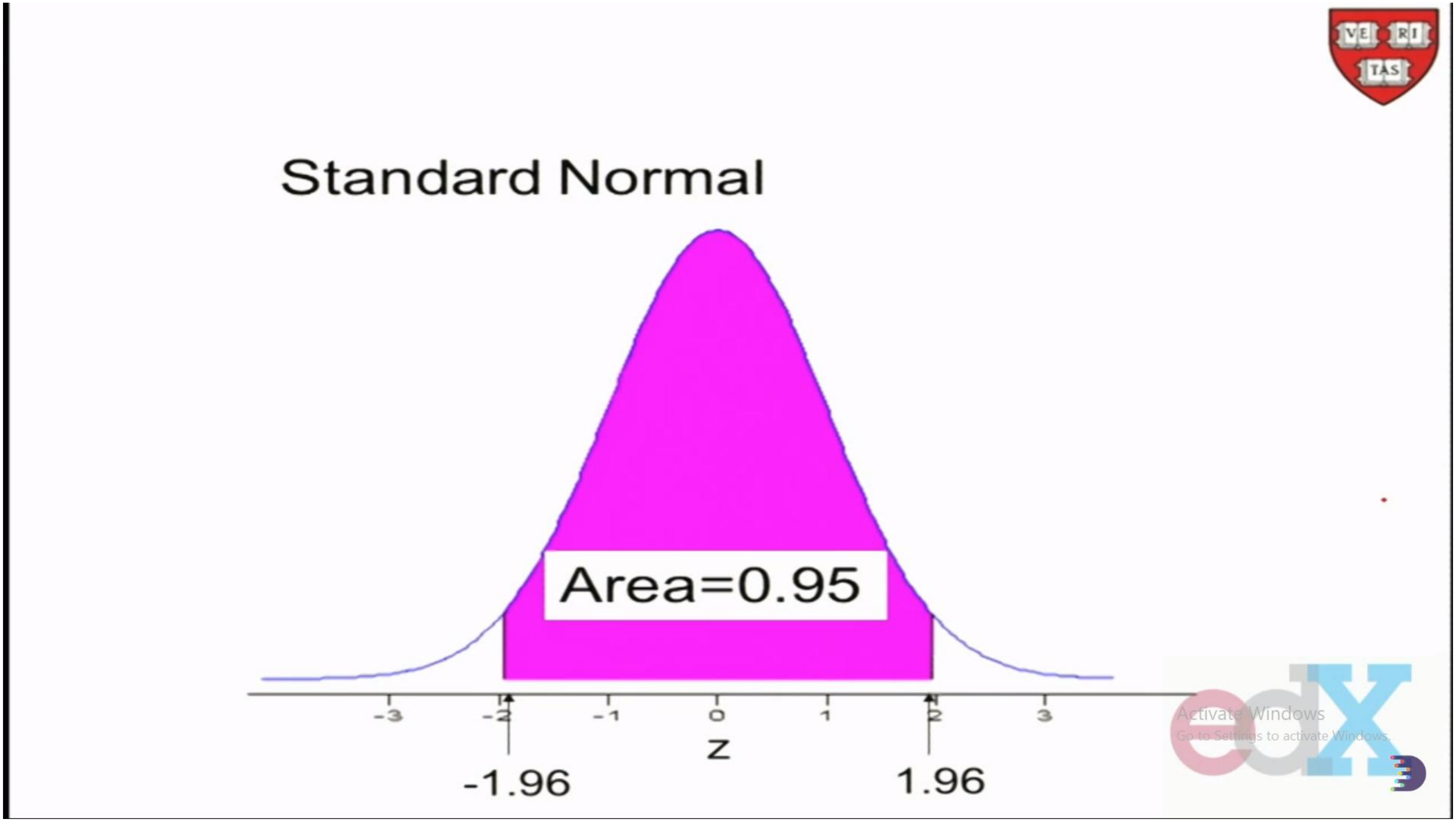
Standard Normal

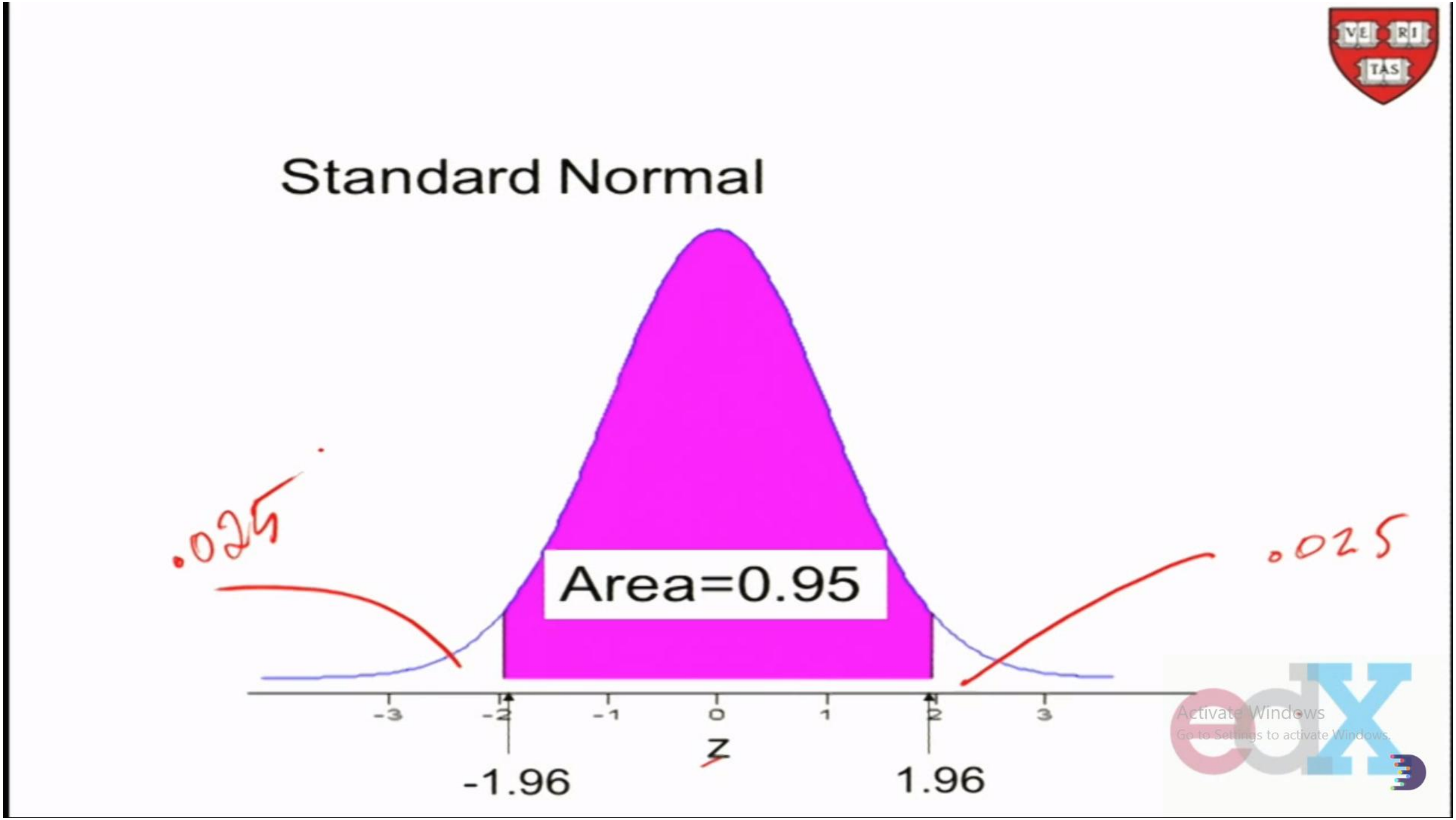


Standard Normal

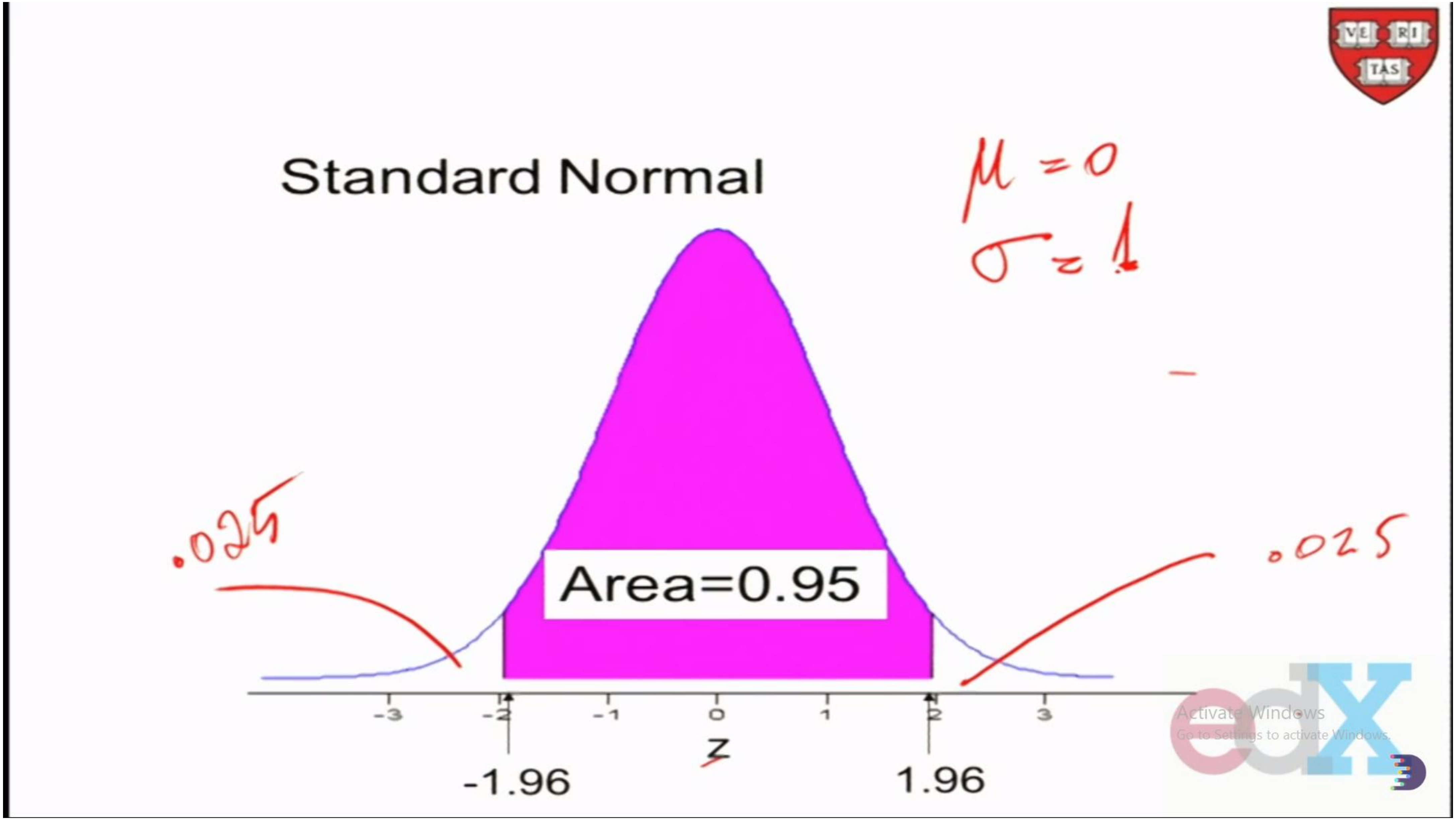


z	area	di normal(z)
0	0.5	0.5
-1.65	0.049	0.04947147
-1.96	0.025	0.0249979
-2.58	0.005	0.00494002
-3	0.001	0.0013499

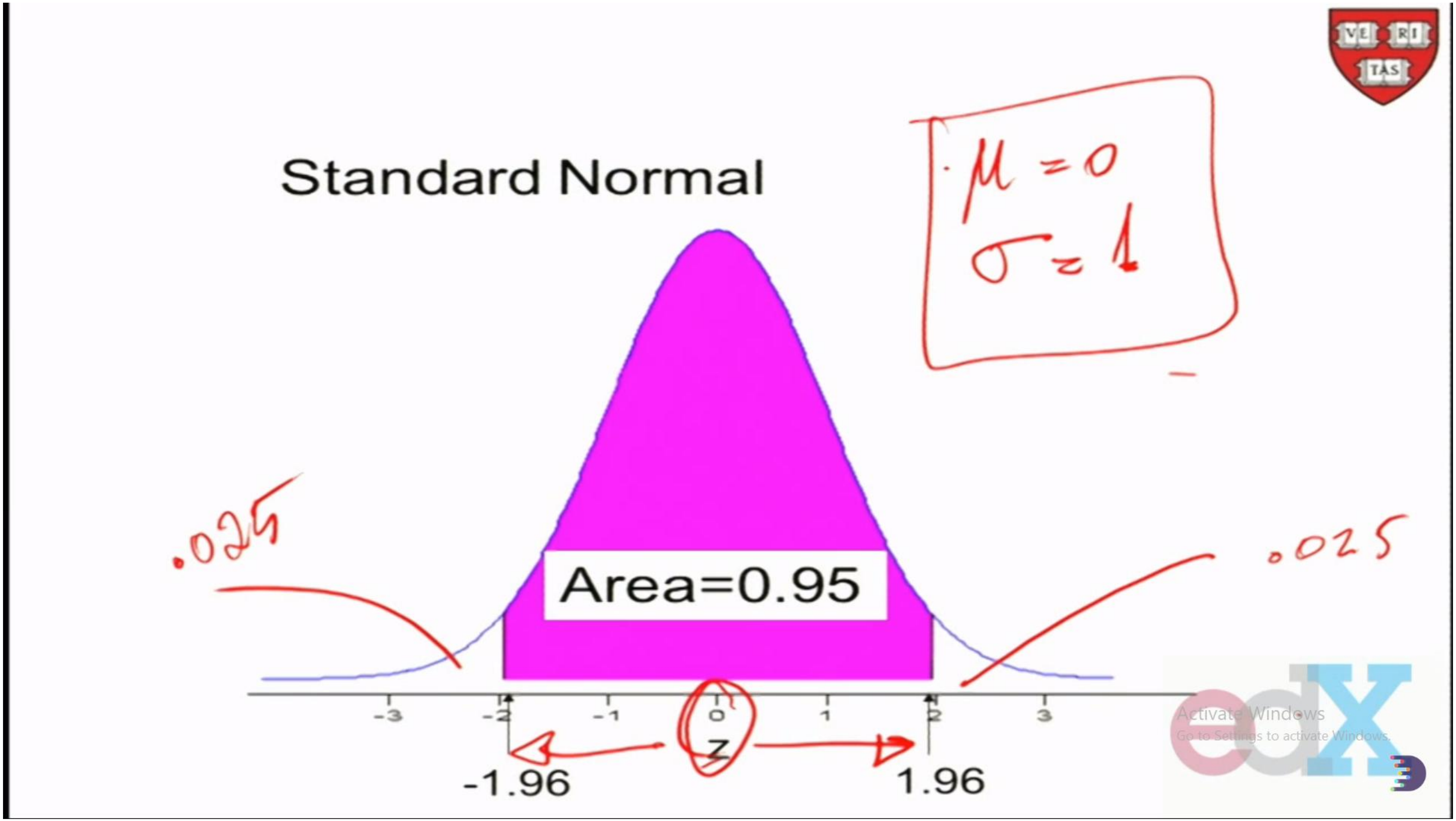

Standard Normal
Area=0.95
-3
-2
-1
0
1
2
3
z
-1.96
1.96

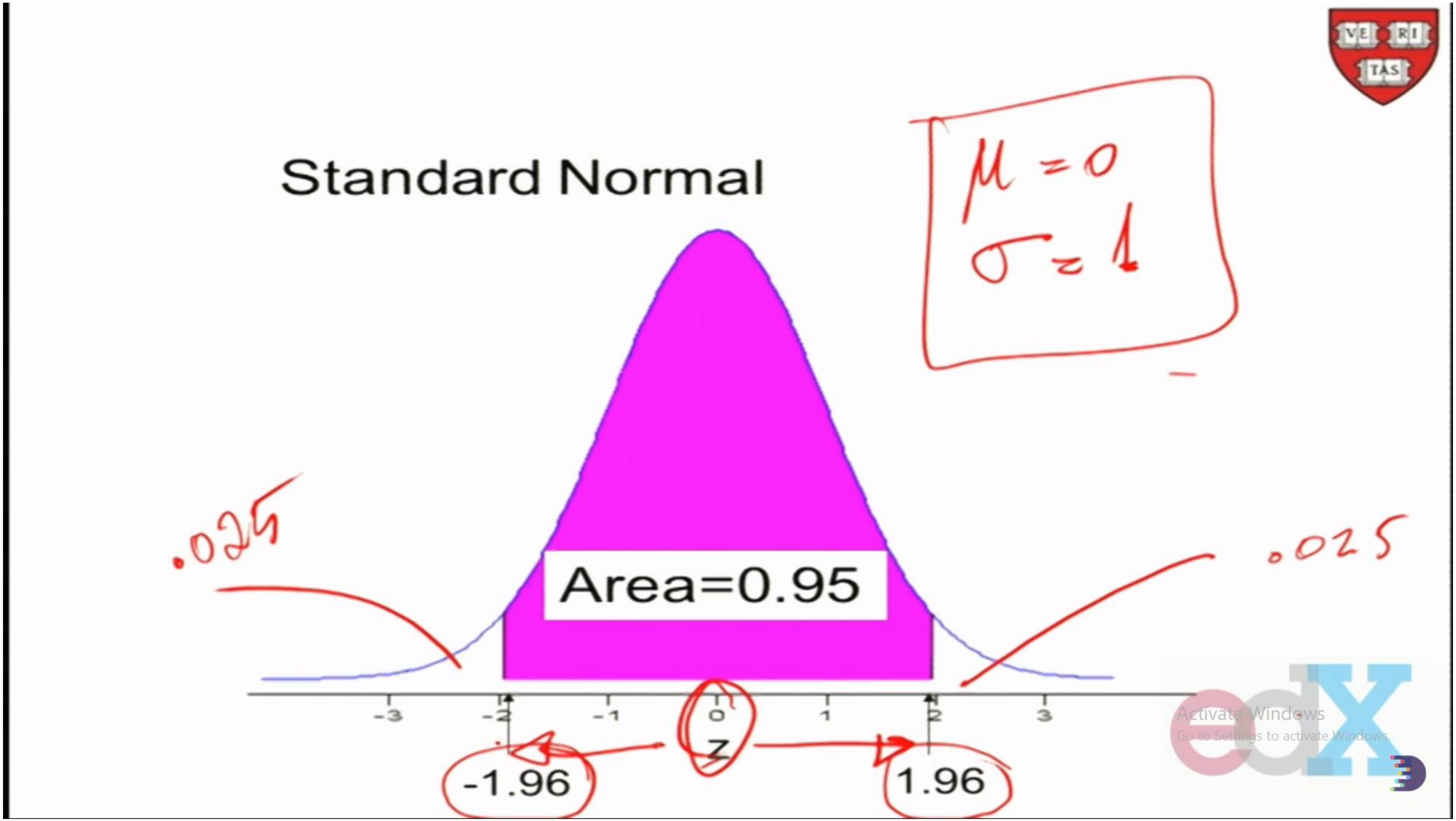


Standard Normal
Area=0.95
.025
.025
-3
-2
-1
0
1
2
3
z
-1.96
1.96



Standard Normal
μ = 0
σ = 1
Area=0.95
.025
.025
-3
-2
-1
0
1
2
3
z
-1.96
1.96


Standard Normal
μ = 0
σ = 1
.025
.025
Area=0.95
-3
-2
-1
0
1
2
3
Z
-1.96
1.96


Standard Normal
$\mu = 0$
$\sigma = 1$
.025
.025
Area=0.95
-3
-2
-1
0
1
2
3
z
-1.96
1.96

General Normal

Suppose X is a normal random variable with mean μ and standard deviation σ, then

$$Z = \frac{X - \mu}{\sigma}$$

is a standard normal (mean zero, standard deviation one).

Predictive Interval

95% of the time:

$$-1.96 \leq Z \leq 1.96$$

Predictive Interval

95% of the time:

$$-1.96 \leq Z \leq 1.96$$

n

$$-1.96 \leq \frac{X - \mu}{\sigma} \leq 1.96$$

Predictive Interval

95% of the time:

$$-1.96 \le Z \le 1.96$$

$$-1.96 \le \frac{X - \mu}{\sigma} \le 1.96$$

$$-1.96\sigma \le X - \mu \le 1.96\sigma$$

$$\mu - 1.96\sigma \le X \le \mu + 1.96\sigma$$

e.g. If X denotes systolic blood pressure, then approximately normal. For 18-74-year-old men in US the mean is 129 mm Hg and the stand. dev. is 19.8 mm Hg.

e.g. If X denotes systolic blood pressure, then approximately normal. For 18-74-year-old men in US the mean is 129 mm Hg and the stand. dev. is 19.8 mm Hg.

$\mu = 129$
$\sigma = 19.8$

So

$$Z = \frac{X - 129}{19.8}$$

is standard normal.

e.g. If X denotes systolic blood pressure, then approximately normal. For 18-74-year-old men in US the mean is 129 mm Hg and the stand. dev. is 19.8 mm Hg.

$\mu = 129$
$\sigma = 19.8$

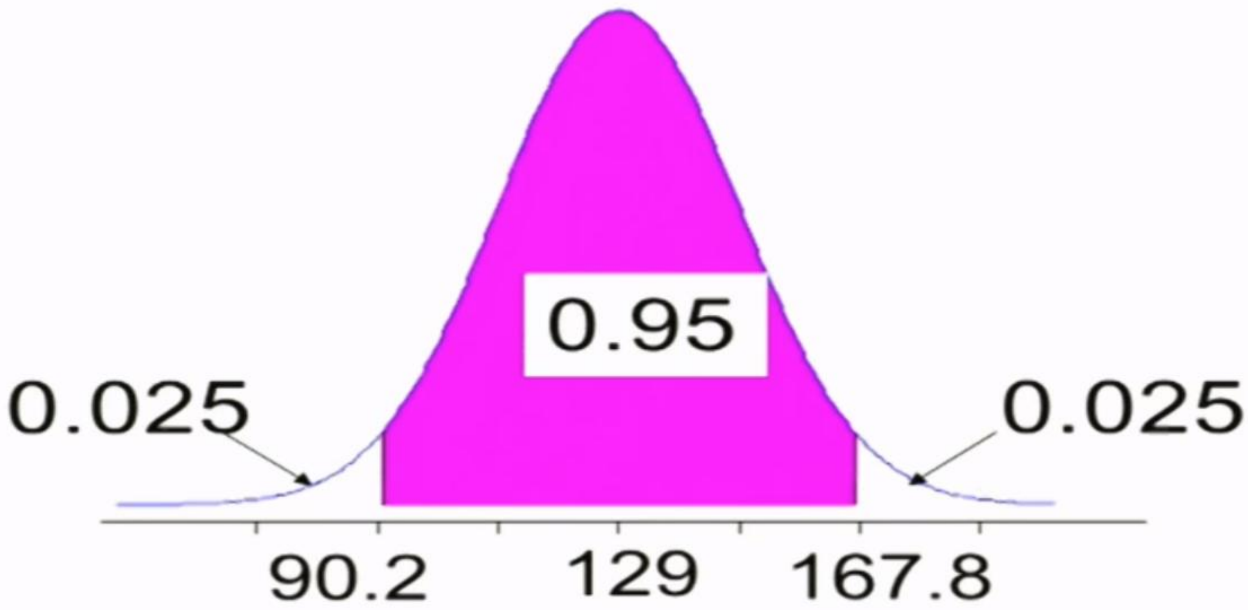
So

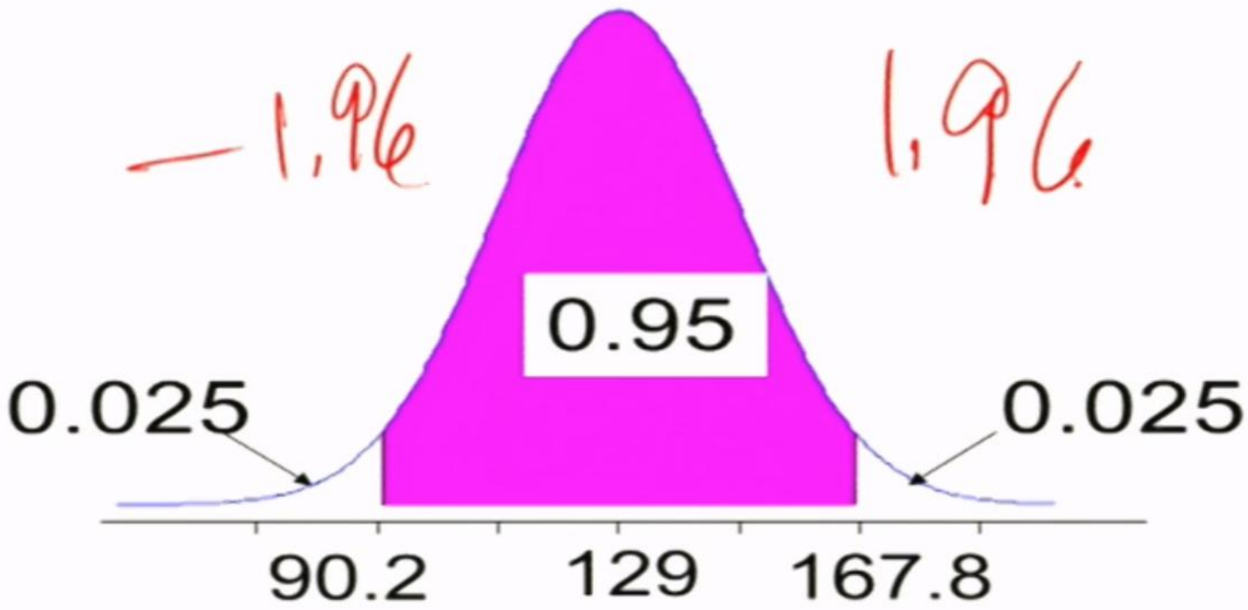
$$Z = \frac{X - 129}{19.8}$$

is standard normal. So if

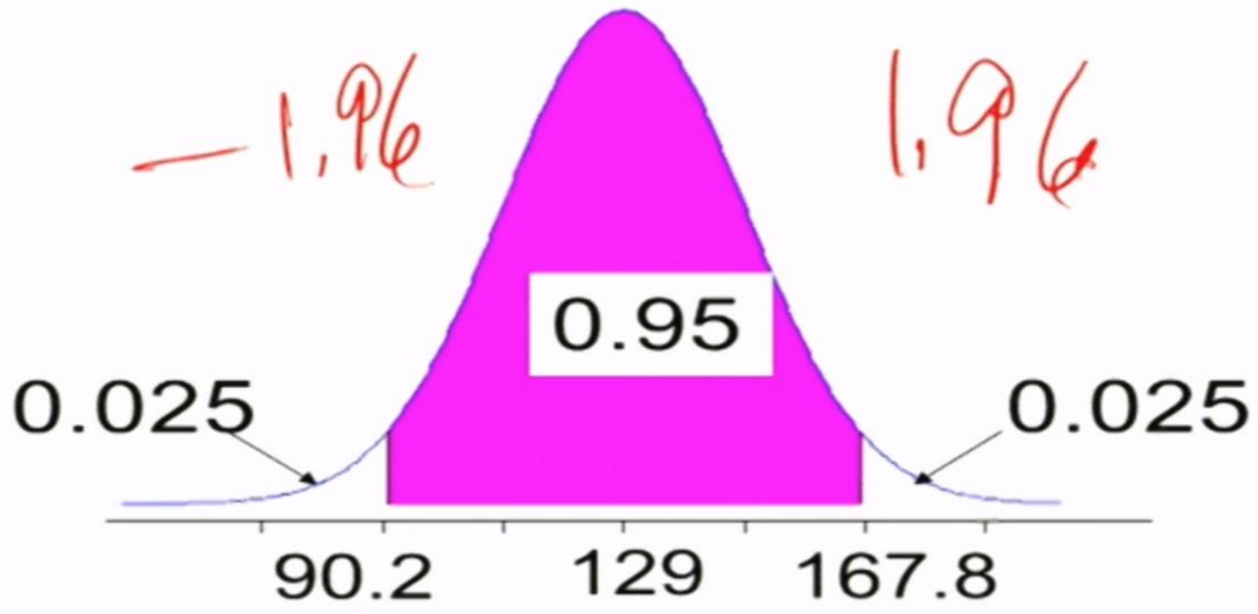
$$1.96 = \frac{X - 129}{19.8}$$

then $X = 167.8$


0.95
0.025
0.025
90.2
129
167.8



-1.96
1.96
0.95
0.025
0.025
90.2
129
167.8



If we choose a person at random from this population, the probability is 0.975 that the person has systolic blood pressure less than 167.8.

How many have blood pressure above 150 mm Hg.?

$$Z = \frac{150 - 129}{19.8} = 1.06$$

Stata: > di normal(1.06)

> .8554277

So, approximately 14.5% of men in the US between the ages of 18 and 74 have systolic blood pressure above 150 mm Hg.

How many have blood pressure above 150 mm Hg.?

$$Z = \frac{150 - 129}{19.8} = 1.06$$

Stata: > di normal(1.06)

> .8554277

So, approximately 14.5% of men in the US between the ages of 18 and 74 have systolic blood pressure above 150 mm Hg.

How many have blood pressure above 150 mm Hg.?

$$Z = \frac{150 - 129}{19.8} = 1.06$$

Stata: > di normal(~~1.06~~) (150-129)/19.8)

> .8554277

So, approximately 14.5% of men in the US between the ages of 18 and 74 have systolic blood pressure above 150 mm Hg.

Poisson Distribution

Siméon Poisson
1781--1840

1. The probability an event occurs in the interval is proportional to the length of the interval.

2. An infinite number of occurrences are possible.

3. Events occur independently at a rate λ.

Siméon Poisson
1781--1840

Poisson Distribution

1. The probability an event occurs in the interval is proportional to the length of the interval.

2. An infinite number of occurrences are possible.

3. Events occur independently at a rate λ.

$$P(X = x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} \qquad x = 0, 1, 2, \ldots$$

For the Poisson one parameter: λ

$$\text{Mean} = \lambda = np$$

$$\text{Variance} = \lambda = np(1-p)$$

$$\approx np$$

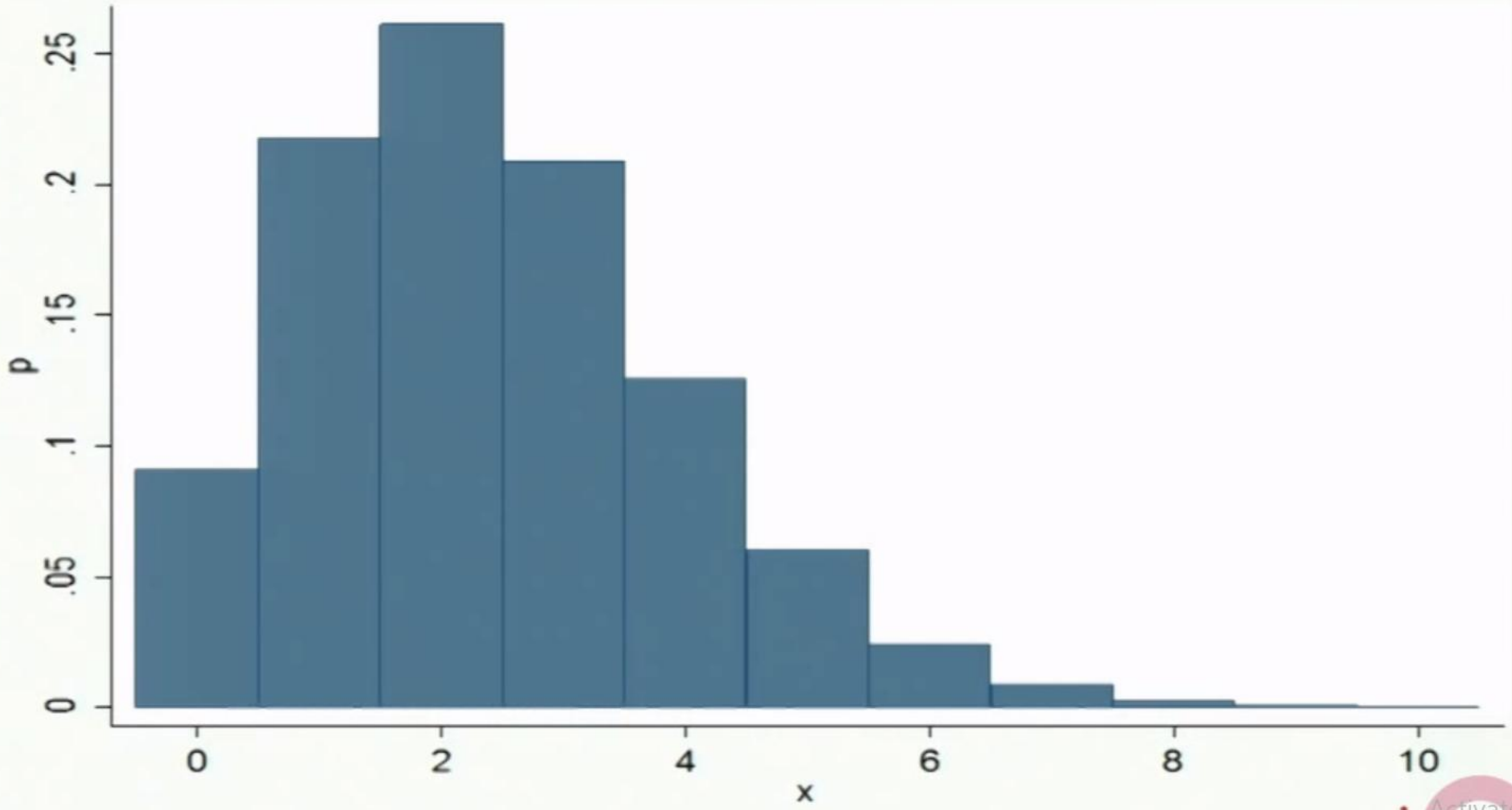
e.g. Probability of an accident in a year is 0.00024. So in a town of 10,000, the rate

$$\lambda = np$$
$$= 10{,}000 \times 0.00024 = 2.4$$

$$P(X=0) = \frac{e^{-2.4}(2.4)^0}{0!} = 0.0907$$

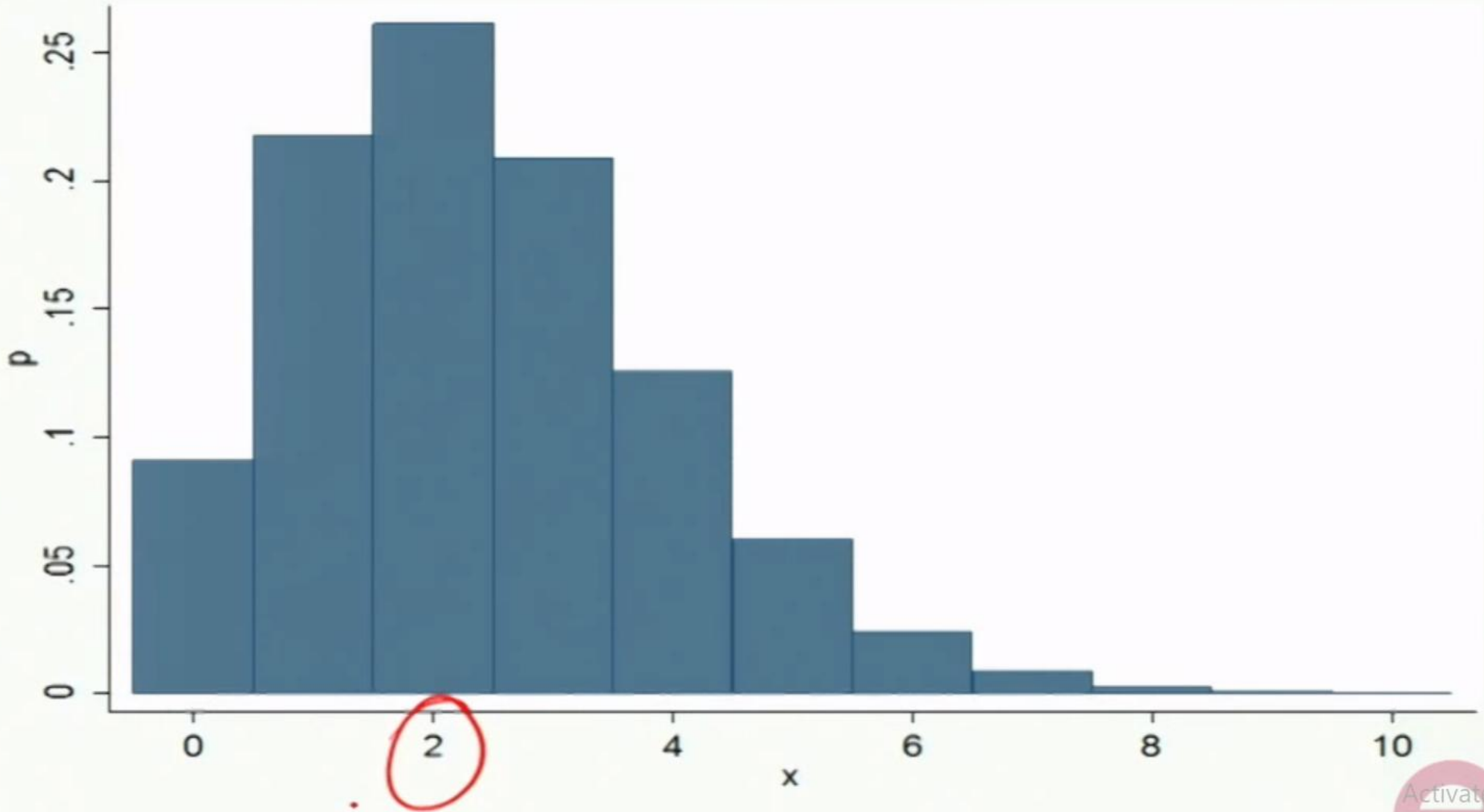
$$P(X=1) = \frac{e^{-2.4}(2.4)^1}{1!} = 0.2177$$

Poisson distribution with mean (λ) = 2.4



Probability of x accidents per year

Poisson distribution with mean (λ) = 2.4



Probability of x accidents per year

```
. gen p = poissonp(2.4,x)

. list x p
```

	x	p
1.	0	.090718
2.	1	.2177231
3.	2	.2612677
4.	3	.2090142
5.	4	.1254085
6.	5	.0601961
7.	6	.0240784
8.	7	.0082555
9.	8	.0024766
10.	9	.0006604
11.	10	.0001585