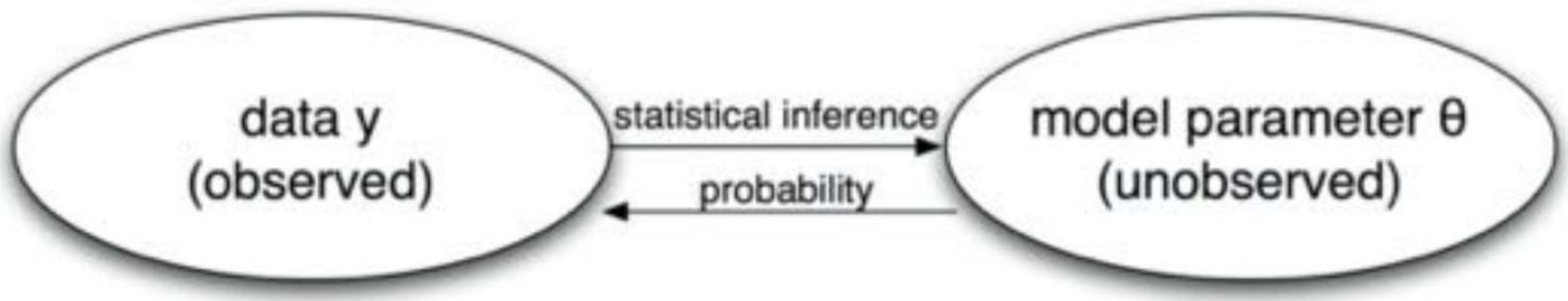
Statistical Models

Instructor, Nero Chan Zhen Yu

Forward School

What is statistical models?

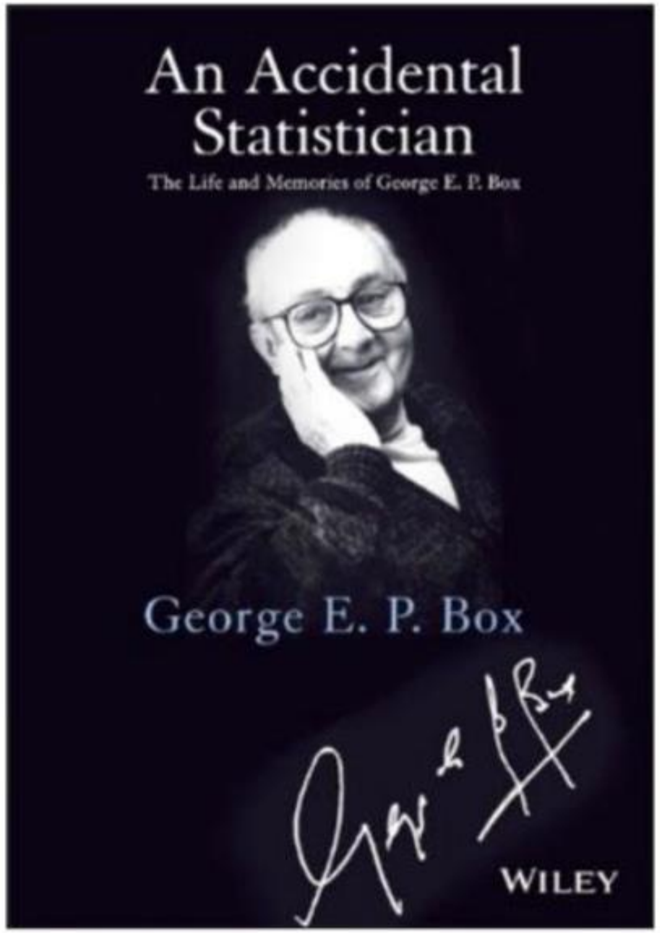
- A family of distributions, indexed by parameters
- Sharpens distinction between data and parameters, and between estimators and estimands.
- Parametric (eg. Based on Normal, Binomial) vs Non-parametric (eg. Methods like bootstrap, KDE)



Forward School

What good is a statistical model?

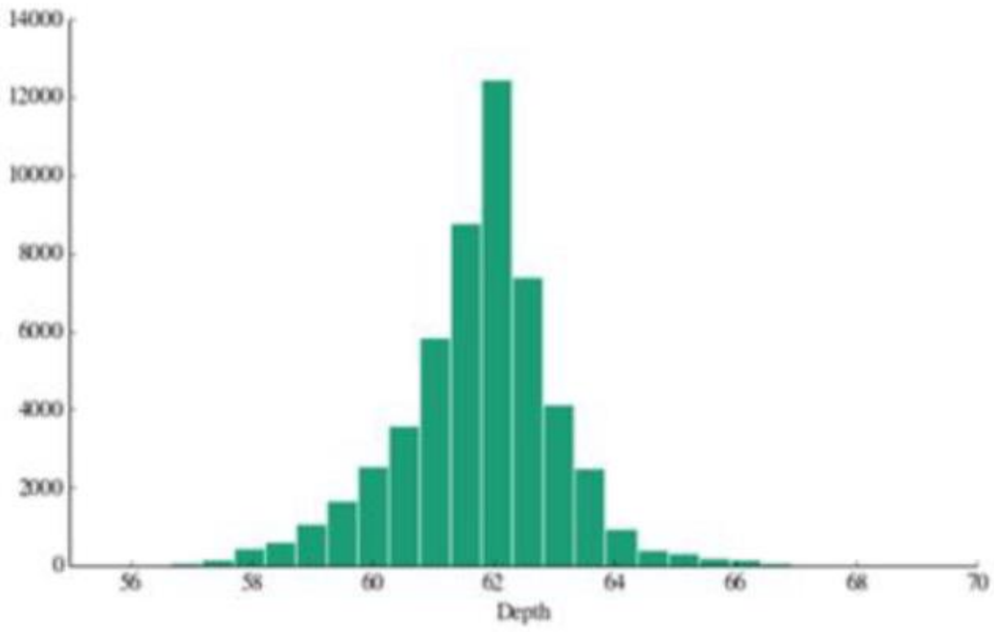
- "All models are wrong, but some models are useful." – George Box (1919-2013)



Forward School

Parametric vs Non-parametric

- Parametric: finite-dimensional parameter space (eg. Mean and Variance for a Normal)
- Non-Parametric: infinite-dimensional parameter space
- Is there anything in between?
- Non-parametric is very general, but no free lunch!
- Remember to plot and explore the data!



Forward School

Statistical Inference

- The basic idea of Statistical Inference:

Sample → Whole population

infer or make decisions

↓ Using sampling distribution theory

Application in

- Confidence Intervals:
  - What is the population value?
- Tests of Significance:
  - Is the population value really what is proposed?

Forward School

Term used in Probability

- Probability:
  - The study of randomness and uncertainty. In other words, probability is a numerical measure of chance for the occurrence of an event.
- Experiment:
  - A repeatable procedure with a well-defined set of possible outcomes. Eg. Tossing a coin, Rolling a die.
- Observation:
  - Data collected only by monitoring what occurs.
- Sample space:
  - The set of all possible outcomes of an experiment.
- Event:
  - A set of outcomes, subset of sample space.

Example

Consider an experiment of rolling a 6-sided die.

Sample space, S: {1, 2, 3, 4, 5, 6}

- A: even number is rolled
  - Equivalently to {2, 4, 6}
  - P(A) = 3/6

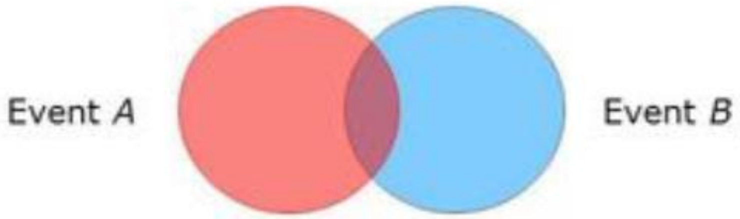
- B: number less than three is rolled
  - Equivalently to {1, 2}
  - P(B) = 2/6

Forward School

Basic Rules of Probability Theory

1) For an event A: $0 \leq P(A) \leq 1$
2) $P(A) + P(A^C) = 1 \Rightarrow P(A) = 1 - P(A^C)$

   A^C: the complement event of event A
3) $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

   if A, B are mutually exclusive, then
   $P(A \cup B) = P(A) + P(B)$ since $P(A \cap B) = 0$
4) $P(A \cap B) = P(A)\,P(B)$ if A and B are independent

   independent : occurrence of A is not depend on B and vice versa

Example

- Lets assume that if we choose a young adult aged 25 to 30 from Malaysia population, the probability that the person chosen has a diploma is 0.46, 0.35 that the person chosen has a bachelor's degree, and 0.13 that the person chosen has both diploma and bachelor's degree.
- What is the probability that a randomly selected Malaysian young adult has a diploma or bachelor's degree?

Example

$P(\text{Diploma}) = 0.46$

$P(\text{Bachelor}) = 0.35$

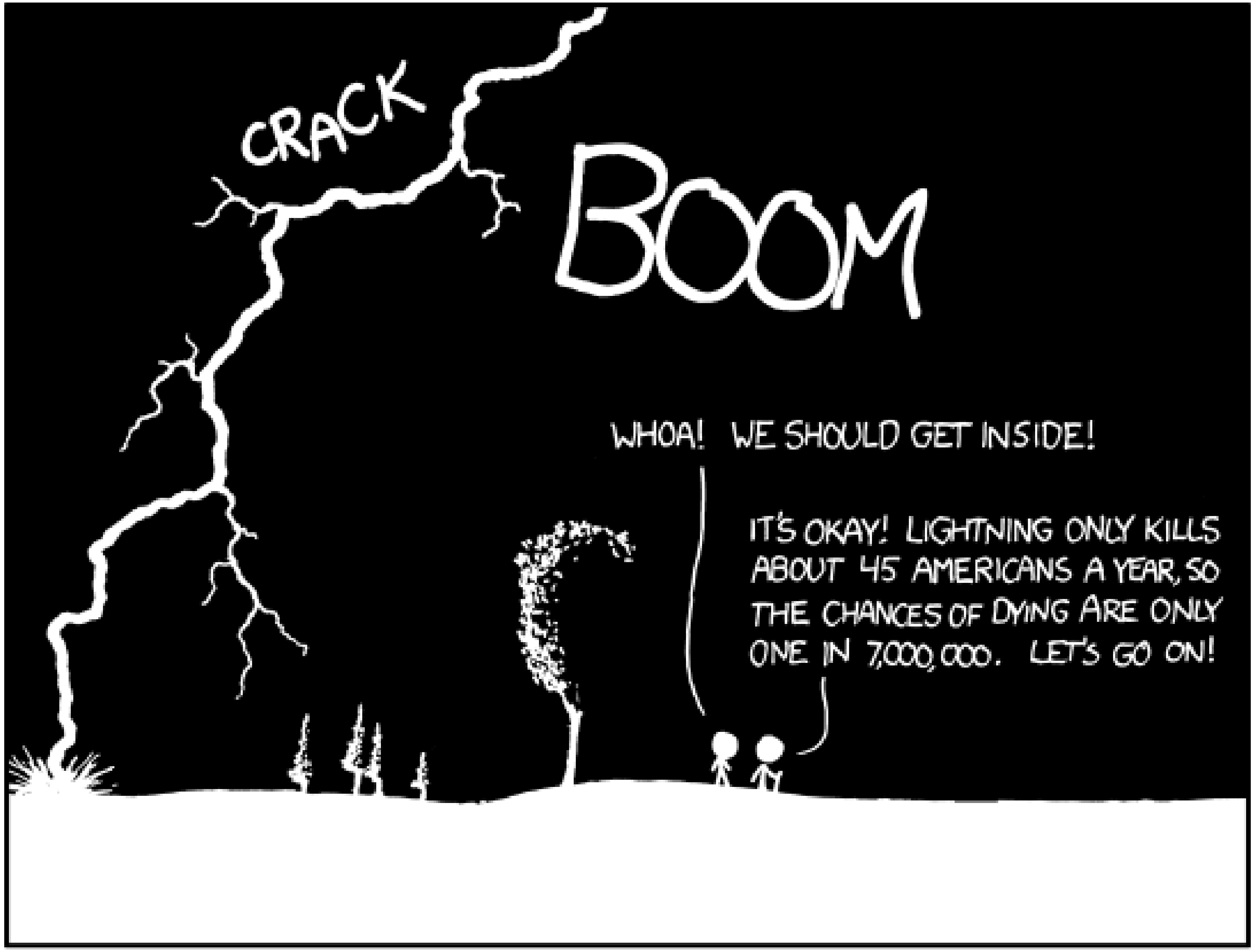
$P(\text{Diploma} \cap \text{Bachelor}) = 0.13$

$P(\text{Diploma} \cup \text{Bachelor}) = P(\text{Diploma}) + P(\text{Bachelor}) - P(\text{Diploma} \cap \text{Bachelor})$

$= 0.46 + 0.35 – 0.13$

$= 0.68$

Forward School



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

Forward School

Conditional Probability

- The probability of an event B occurring when it is known that some event A has occurred is called conditional probability and is denoted by P(B|A).
- For any two events A and B (with P(A) > 0), the conditional probability of B given A, denoted by P(B|A), is defined as

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- If we are interested in the conditional probability of A given B (with P(B)>0), then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Forward School

Bayes' Theorem

$$\underbrace{P(B|A)}_{\text{Posterior}} = \frac{\overbrace{P(A|B)}^{\text{Likelihood}}\ \overbrace{P(B)}^{\text{Prior}}}{\underbrace{P(A)}_{\text{Model Evidence}}}$$

Forward School

Conditional Probability

- $P(B|A) = \dfrac{P(A \cap B)}{P(A)} \Longrightarrow P(A \cap B) = P(B|A)\,P(A)$
- $P(A|B) = \dfrac{P(A \cap B)}{P(B)} \Longrightarrow P(A \cap B) = P(A|B)\,P(B)$

Forward School

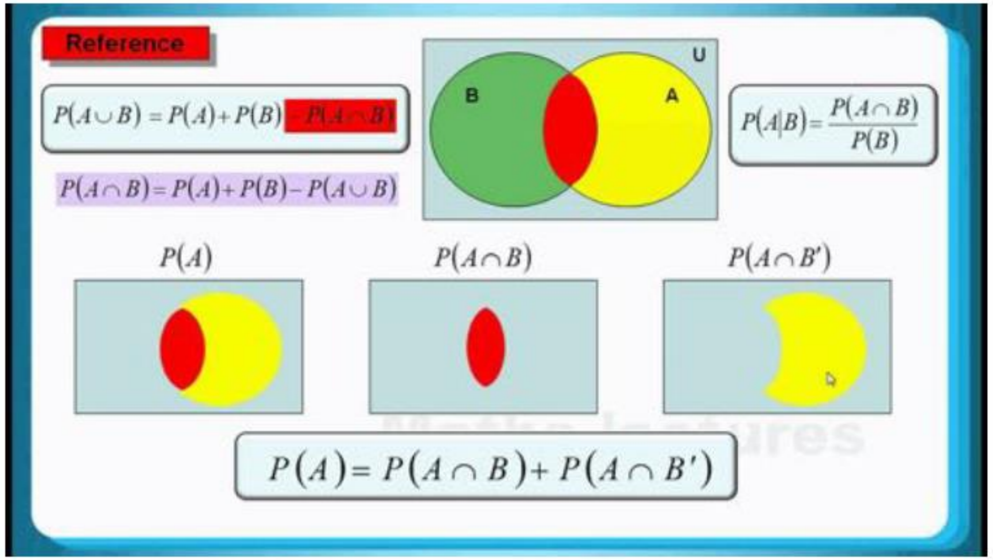
Law of Total Probability

- Suppose there are 2 possible events A and B: What are the total possibilities of B happening?

$$P(B) = P(B \cap A) + P(B \cap A^c)$$

- Since $P(B \cap A) = P(B|A)\,P(A)$
- $P(B \cap A^c) = P(B|A^c)\,P(A^c)$
- **The Law of Total Probability:**

$$\mathbf{P(B) = P(B|A)\,P(A) + P(B|A^c)\,P(A^c)}$$



Forward School

Example

- Draw 2 cards without replacement from a standard pack of cards. What is the probability of getting an Ace in the second draw?
- Pack of cards = 52 cards, of which 4 are Aces

  A = Get an Ace in first draw

  B = Get an Ace in second draw

- P(A) = 4/52= 0.077
- P(B|A) = 3/51 = 0.069
- P(B|A^c) = 4/51 = 0.078

So P(B)

$= P(B \cap A) + P(B \cap A^c)$

$= P(B|A)\,P(A) + P(B|A^c)\,P(A^c)$

$= (0.069)(0.077) + (0.078)(1-0.077)$

$= 0.0053 + 0.072$

$= \mathbf{0.077}$

Forward School

Bayes' Rule

$$
\begin{aligned}
&P(A \mid B) \\
&= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(B \mid A)P(A)}{P(B)} \\
&= \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}
\end{aligned}
$$

- Where the denominator uses **The Law of Total Probability**

Example

- A patient with certain symptoms consulted her doctor to be checked for a disease, and she undergoes a test.
- With this test there is a probability of 0.90 that a woman with the disease shows a positive result, and a probability of only 0.99 that a healthy woman shows a negative result.
- Historical information also suggests that the prevalence of this disease in the population is 2 in 10000.
- Find the probability that a woman has the disease given the test says she does (i.e. does the test diagnose true patient status?)

Forward School

Example

Let D be the event 'woman has the cancer' and + be the event 'biopsy is positive'.

P(D) = 2/10000 = 0.0002 (disease prevalence)

P(+|D) = 0.90 (sensitivity)

P(-|D^C)= 0.99 (specificity)

```
> (0.9)*(0.0002)/((0.9)*(0.0002)+(1-0.99)*(1-0.0002))
0.0176852
```

Need to find the positive predictive value, P(D|+)

$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)}$$

$$= \frac{(0.9)(0.0002)}{(0.9)(0.0002) + (1 - 0.99)(1 - 0.0002)}$$

Forward School

Important Point on Notation:

- A numerical value associated with a <u>population</u> is called a "parameter": (FIXED)
  - population mean: μ(mu)
  - population standard deviation: σ(sigma)
  - population proportion: p

- A numerical value associated with a <u>sample</u> is called a "statistic": (VARIABLE)
  - sample mean: $\bar{x}$(x-bar)
  - sample standard deviation: s
  - sample proportion: $\hat{p}$ (p-hat)

Forward School

Random Variable

- A random variable has values which depend on the outcome of a random experiment
- Random variables are labelled with capital letter. (e.g. X = sum of numbers on 2 throws of a die)
- The values of random variable is labelled with small letters. ($x = 2, 3, 4, \ldots, 12$)
- Random variables can be discrete or continuous.

Discrete Random Variable

- **Discrete random variable** is a **random variable** whose possible values can be listed in a **finite** or an **infinite sequence**
- E.g.

  number of students attend this class,

  number of students in this family,

  number of cars pass through a toll in given day

Forward School

Continuous Random Variable

- **Continuous random variable** are **random variables** that are measured on **a continuous scale.**
- They can usually take on any value over some interval
- E.g.

    heights,

    weights,

    time spending on reading a statistics book

Forward School

Probability Mass Function (PMF)

- This function evaluates the probability of a discrete random variable *X* takes certain value *x*.

$$P(x) = P(X = x)$$

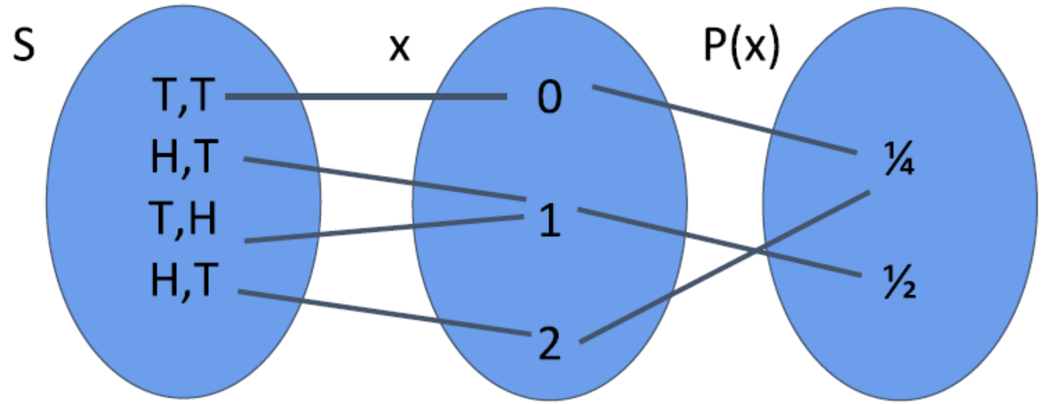
- The following conditions are required for any pmf:

1. p(x) > 0
2. $\sum p(x) = 1$

Forward School

Probability Mass Function (PMF)

Toss of a coin, and Random Variable, x is HEAD



Function P(x): Random Variable → Probability

Probability Mass Function (PMF)

What is the probability of getting:

a) No HEAD

P(x=0) = ¼

b) One HEAD

P(x=1) = P(HT) + P(TH) = 2/4 = ½

c) Two HEADs

P(x=2) = ¼

Forward School

Probability Mass Function (PMF)

- The PMF of a random variable *X* is:

$$p(x) = x/36 \quad \text{for} \quad x = 1, 2, 3, \ldots 8$$

$$P(X = 1) = p(1) = 1/36$$
$$P(X = 2) = p(2) = 2/36$$
$$\vdots$$
$$P(X = 8) = p(8) = 8/36$$

Forward School

Example

x	$P(x)$
1	1/36
2	2/36
3	3/36
4	4/36
5	5/36
6	6/36
7	7/36
8	8/36
Sum =	1

Cumulative Distribution Function (CDF)

- This function evaluates the probability of a random variable *X* takes value less or equal to *x*

$$F(x) = P(X \leq x)$$

For the example in the previous slide:

F(3) = P(X<= 3) = 1/36 + 2/36 + 3/36 = 6/36 = 1/6

Forward School

Probability Density Function (PDF)

- This function associates with the probability of a continuous random variable
- The probability that the continuous random variable takes on a value between *a* and *b* is the area under the curve between *a* and *b.*
- For continuous random variable

$$P(X < a) = P(X \leq a)$$

- That is with or without equal sign makes no difference to the probability, this also apply to greater than

Example

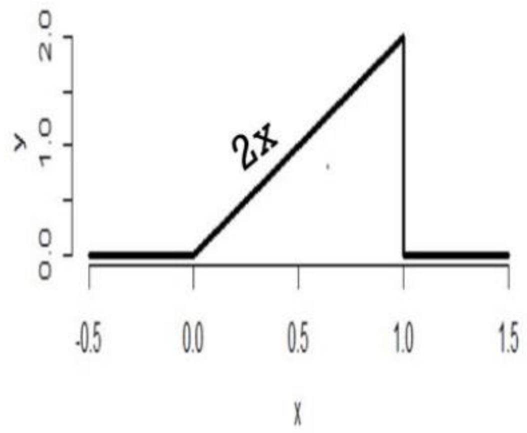
- Suppose that the proportion of help calls that get addressed in a random day by a help line is given by:

$$f(x) = \begin{cases} 2x, 0 < x < 1 \\ 0, otherwise \end{cases}$$

Forward School

Example

- If this is a mathematically valid density, then the total probability (area under curve) should equal to 1

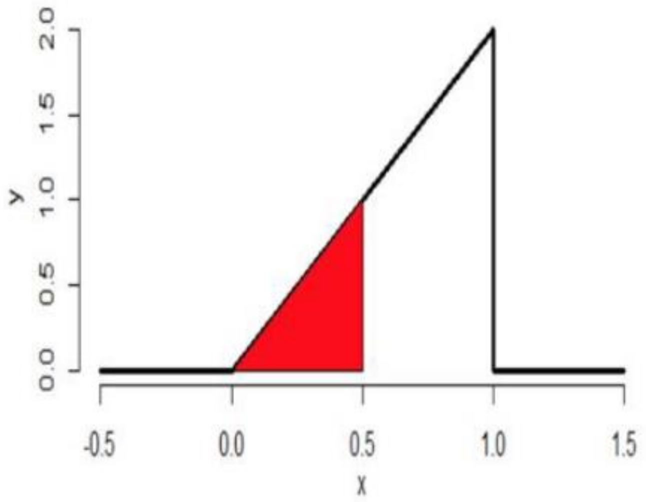


- Area under the curve (triangle) = (1/2)(base)(height)
  = (1/2) (1)(2) = 1
- Total probability = 1, it is valid density

Forward School

Example

- What is the $P(X < 0.5)$?



- Area under the curve (red color triangle) = (1/2)(base)(height) = (1/2)(0.5)(1) = 0.25
- Or if we know that this is a beta distribution with parameter 2 and 1, just type in python

```
from scipy.stats import beta
beta.cdf(0.5,2,1)
[1] 0.25
```

Forward School

Percentile

The αth percentile of a distribution is the value x such that

- The probability that a random variable drawn from the population is less than x is α %

$$P(X < x) = \alpha\%$$

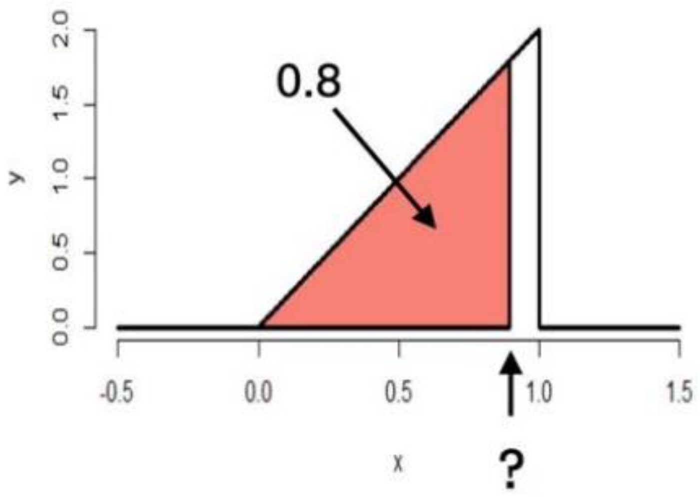
- The probability that a random variable drawn from the population is more than x is $(1 - \alpha)\%$

$$P(X > x) = (1 - \alpha)\%$$

Forward School

Example

- What is the 80th percentile of this pdf?



$P(X < ?) = 0.8$

$\left(\frac{1}{2}\right)(x)(2x) = 0.8 \Longrightarrow x^2 = 0.8 \Longrightarrow x{=}0.89$

Or by using scipy:

```
> beta.ppf(0.8, 2, 1)
```

Forward School

Quartiles

- The first quartile is the 25th percentile
$$P(X < x) = 0.25$$
- The second quartile is the 50th percentile (the median)
$$P(X < x) = 0.5$$
- The third quartile is the 75th percentile
$$P(X < x) = 0.75$$

```
## 1st quartile
beta.ppf(0.25, 2, 1)

## 2nd quartile
beta.ppf(0.5, 2, 1)

## 3rd quartile
beta.ppf(0.75, 2, 1)
```

Forward School

The Population Mean

- Mean, also know as expected value
- Expected value of a random variable *X, E(X) or* μ
- For discrete random variable, if we know the PMF then the mean can be determined
- The mean of a discrete random variable *X* is:

$$\mu = E(X) = \sum xp(x)$$

Forward School

Example

X = face value when we throw a die

Pmf of X:

x	$P(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$$E(X) = \sum xp(x) = (1)(1/6) + (2)(1/6) + \ldots + (6)(1/6)$$

$$= 3.5$$

Forward School

Sample Mean

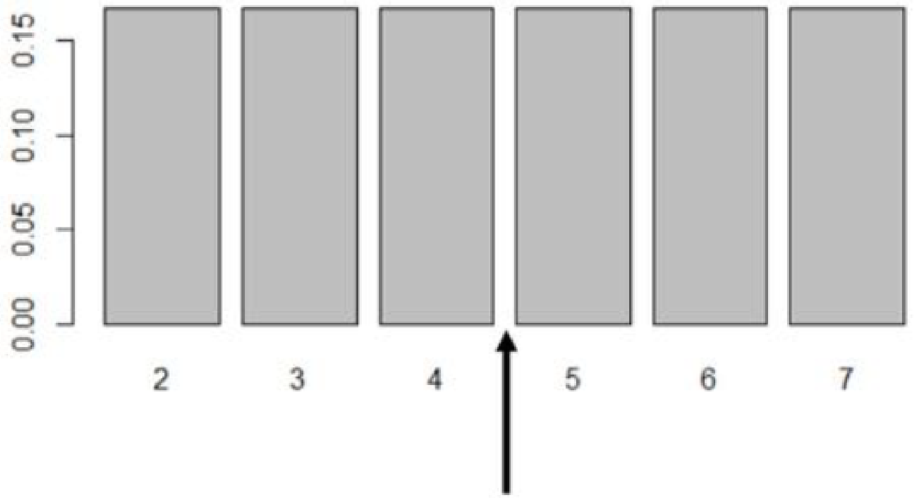
- The mean calculated from sample data
- It estimates the population mean
- Sample mean : $\bar{x}$
- Population mean : μ
- E.g. sample data value: 4, 6, 2, 5, 3, 7

$$\bar{x} = \frac{4+6+2+5+3+7}{6} = 4.5$$

Forward School

Mean

- The sample mean is the center of mass of the sample data
- For the sample data: 4, 6, 2, 5, 3, 7
- Let's assume that they are build of solid bar and we arrange them on a level surface. The mean will be point that we can put a stick that balance these bars.

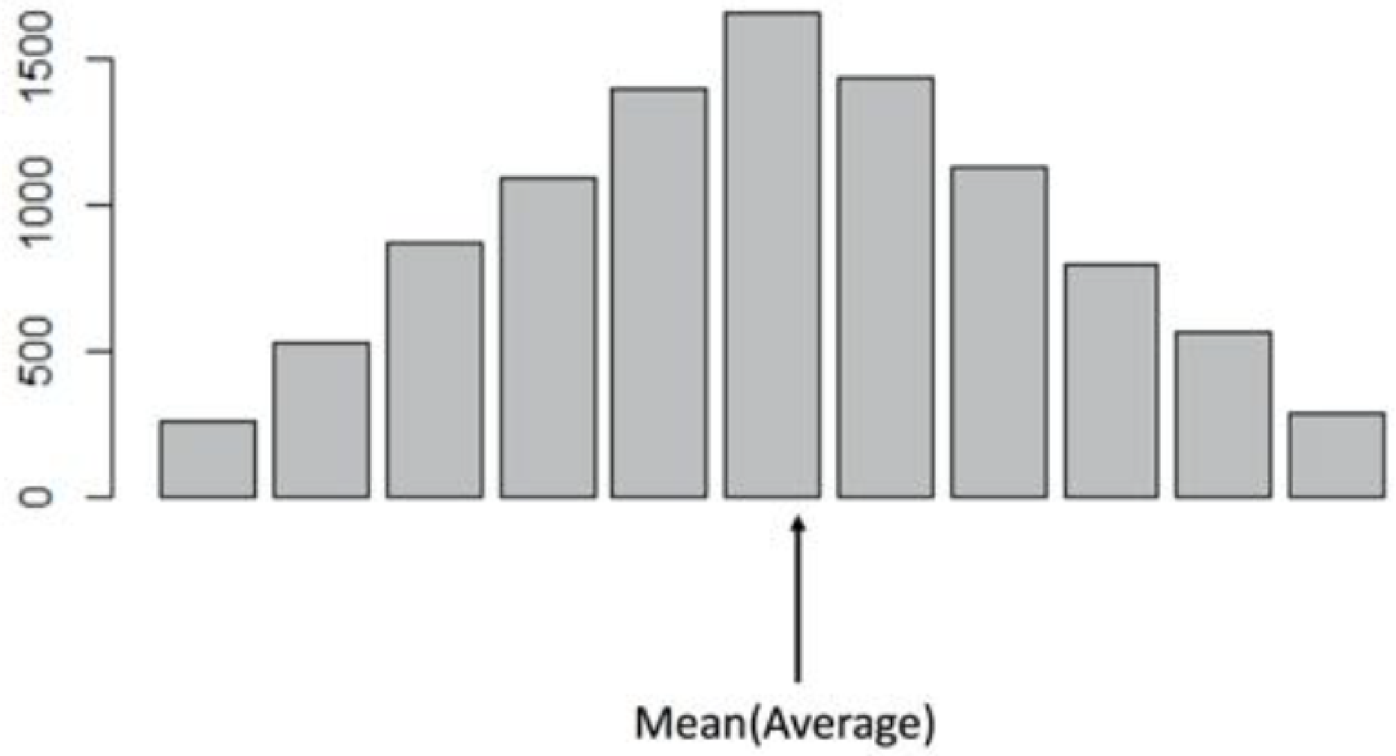


Forward School

Behaviour of Sample Mean

- Sample Mean is a random variable
- When we take different sample from the same population, the sample mean might be different
- Let’s take a look on how these sample mean behave
- Imagine we roll a die 2 times, calculate the mean of the 2 outcomes, now we have 1 sample mean of die roll with sample size 2
- We repeat this procedure for 10000 times, so we will have 10000 sample mean of die roll with sample size 2

Behaviour of Sample Mean



Forward School

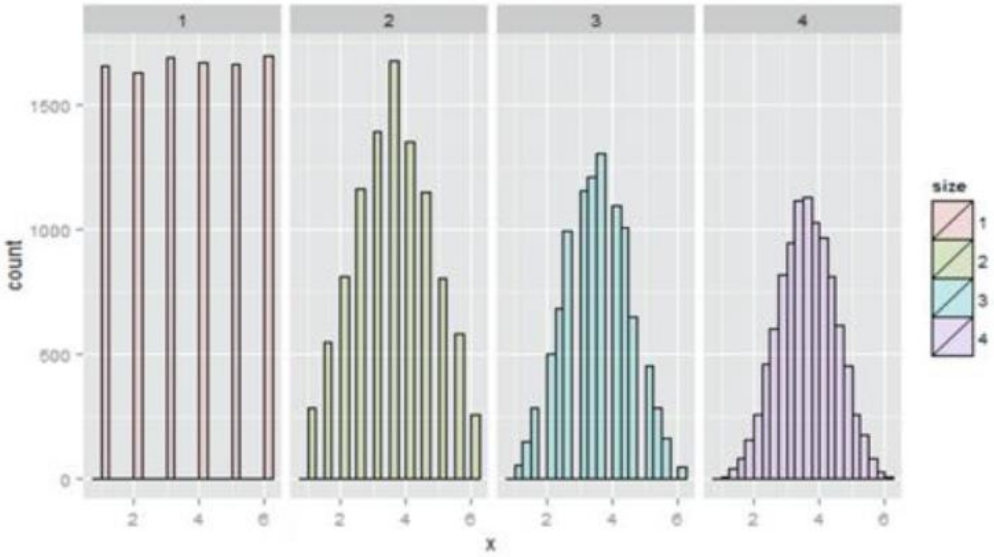
Behaviour of Sample Mean

- Let's repeat the above steps with larger sample size, say *n=3*, and *n=4*
- *n=3:*
  - We roll the die 3 times, calculate the mean of the 3 outcomes, now we have 1 sample mean of die roll with sample size 3
  - We repeat this procedure 10000 times, so we will have 10000 sample mean of die roll with sample size 3
- *n=4:*
  - We roll the die 4 times, calculate the mean of the 4 outcomes, now we have 1 sample mean of die roll with sample size 4
  - We repeat this procedure 10000 times, so we will have 10000 sample mean of die roll with sample size 4

Forward School

Behaviour of Sample Mean

- Plot the distributions of the sample mean side by side:



- When *n* increases:
  - The distribution of the sample mean become more concentrated and more Gaussian (symmetric)
  - The center of mass remain unchanged

Behaviour of Sample Mean

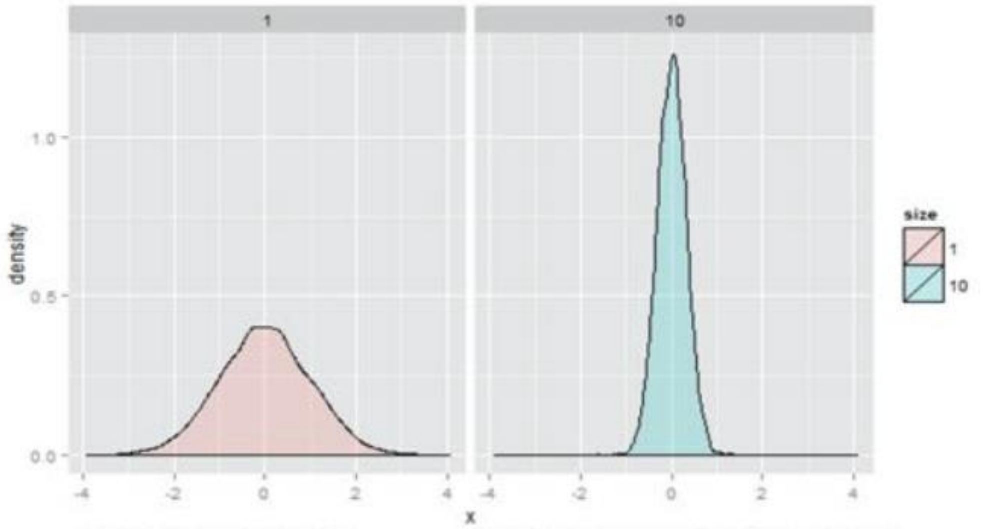
- When *n* increases:
  - The distribution of the sample mean become more concentrated and more Gaussian (symmetric)
  - The center of mass remain unchanged

Center of the mass = population mean

- Sample mean is an unbiased estimator for the population mean

Forward School

Behaviour of Sample Mean



Population follow a standard normal distribution: N(0,1)

Distribution of 10000 sample means (10000 samples size 10 taken from N(0,1))

- The distribution of the sample mean (blue) is Gaussian, more concentrated than the population distribution and has mean = 0

Forward School

Behaviour of Sample Mean

- When the population is normal, the distribution of the sample mean is exactly normal with mean μ and variance $\frac{\sigma^2}{n}$

- When the population is not normal (eg. Toss coin), the distribution of the sample mean is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ if n is large enough

Forward School

Summary

- Expected values are properties of distributions
- The population mean is the center of mass of population
- The sample mean is the center of mass of the observed data (sample)
- The sample mean is an estimate of the population mean
- The sample mean is unbiased
  - The mean of its distribution is the mean that it's trying to estimate
  - The more data that goes into the sample mean *(n increase)*, the more concentrated its density / mass function is around the population mean

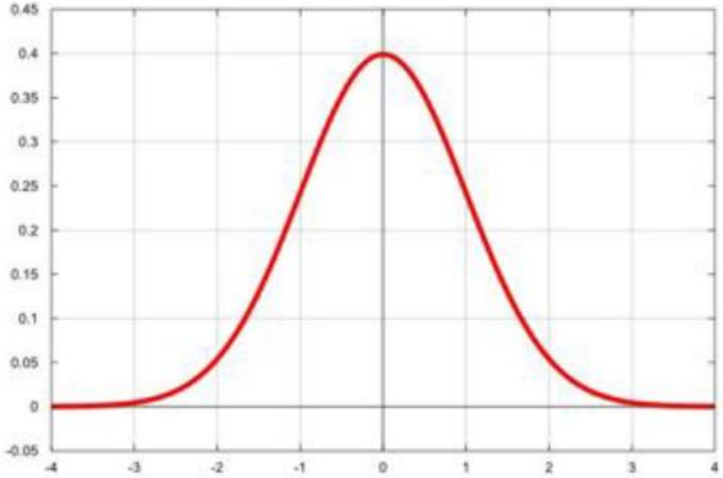
Forward School

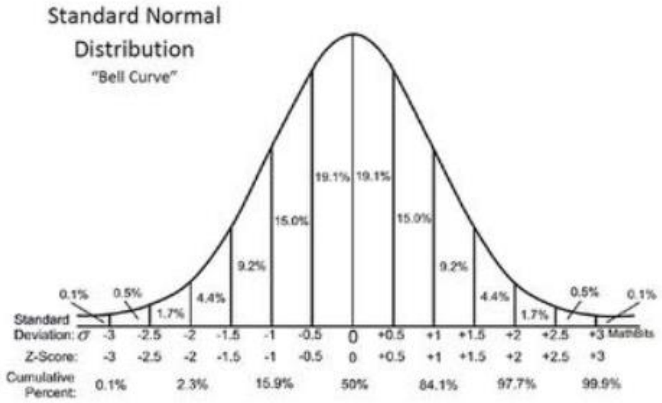
Normal Distribution

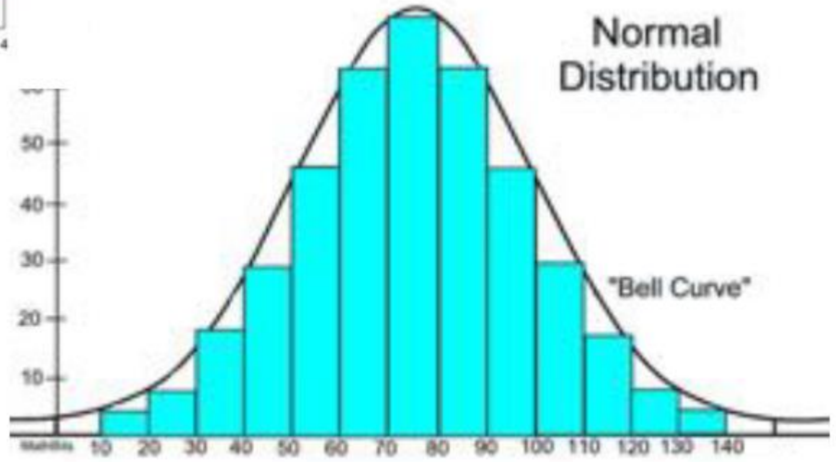
Instructor, Nero Chan Zhen Yu

Forward School

Have you seen this shape before?







Forward
School

The Normal Distribution

- Often times data is described as being “normal” (in statistical sense)
- As known as (or Gaussian or Gauss or Laplace-Gauss)
- The frequency by which some events occur; both natural and man-made:
  - Natural: human height, temperature, blood pressure etc.
  - Man-made: machine products, financial data, sales etc.
- For these measures, the average (mean) tends to be very frequent while measures away from the mean are less and less frequent

Standard Normal Distribution

- Take any Normal Distribution and convert it to The Standard Normal Distribution

Forward School

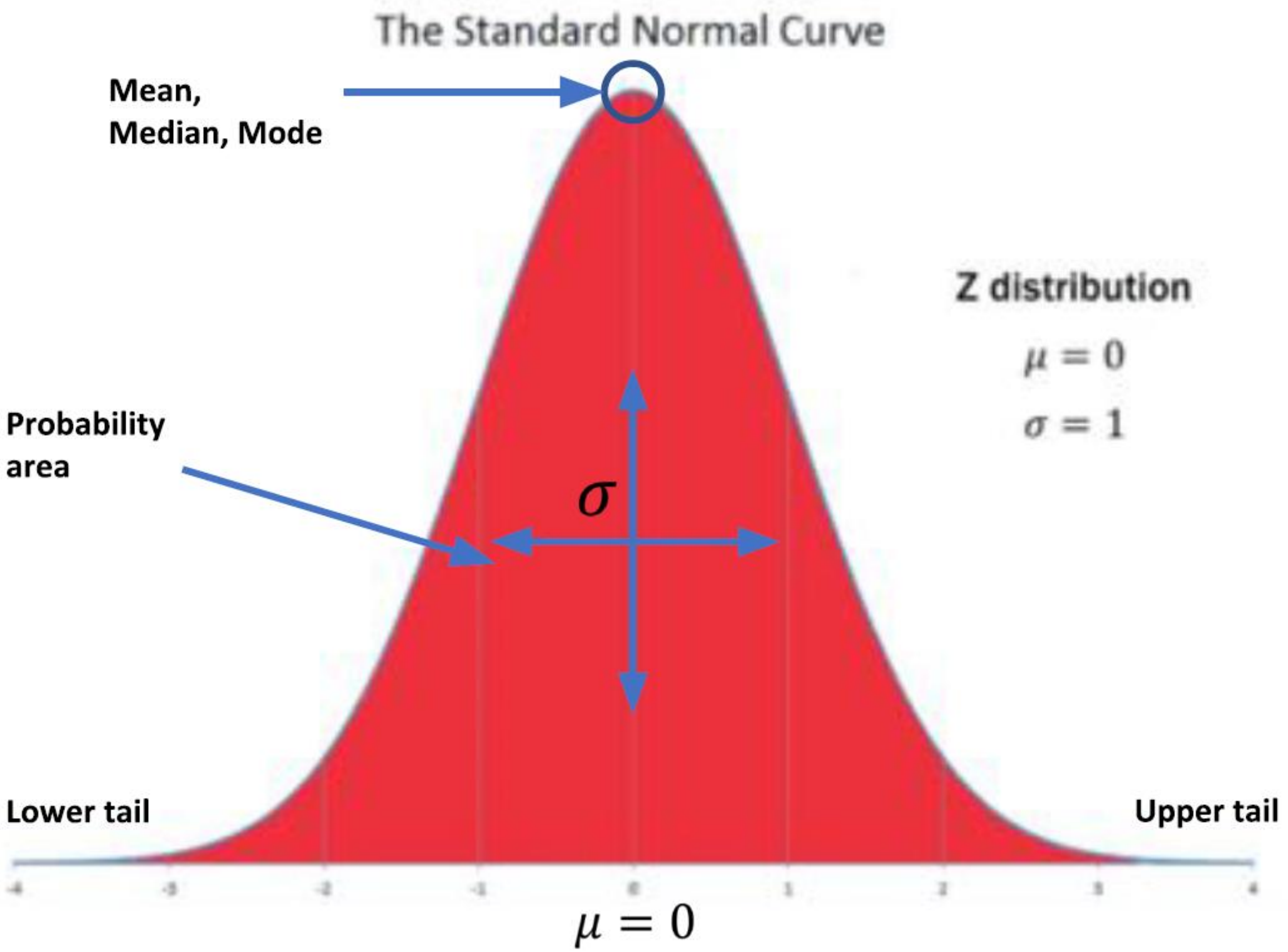
Standard Normal Distribution

- The number of standard deviations from the mean is also called the “Standard Score”, “Sigma” or “z-score”.
- Here is the formula for z-score that we have been using:

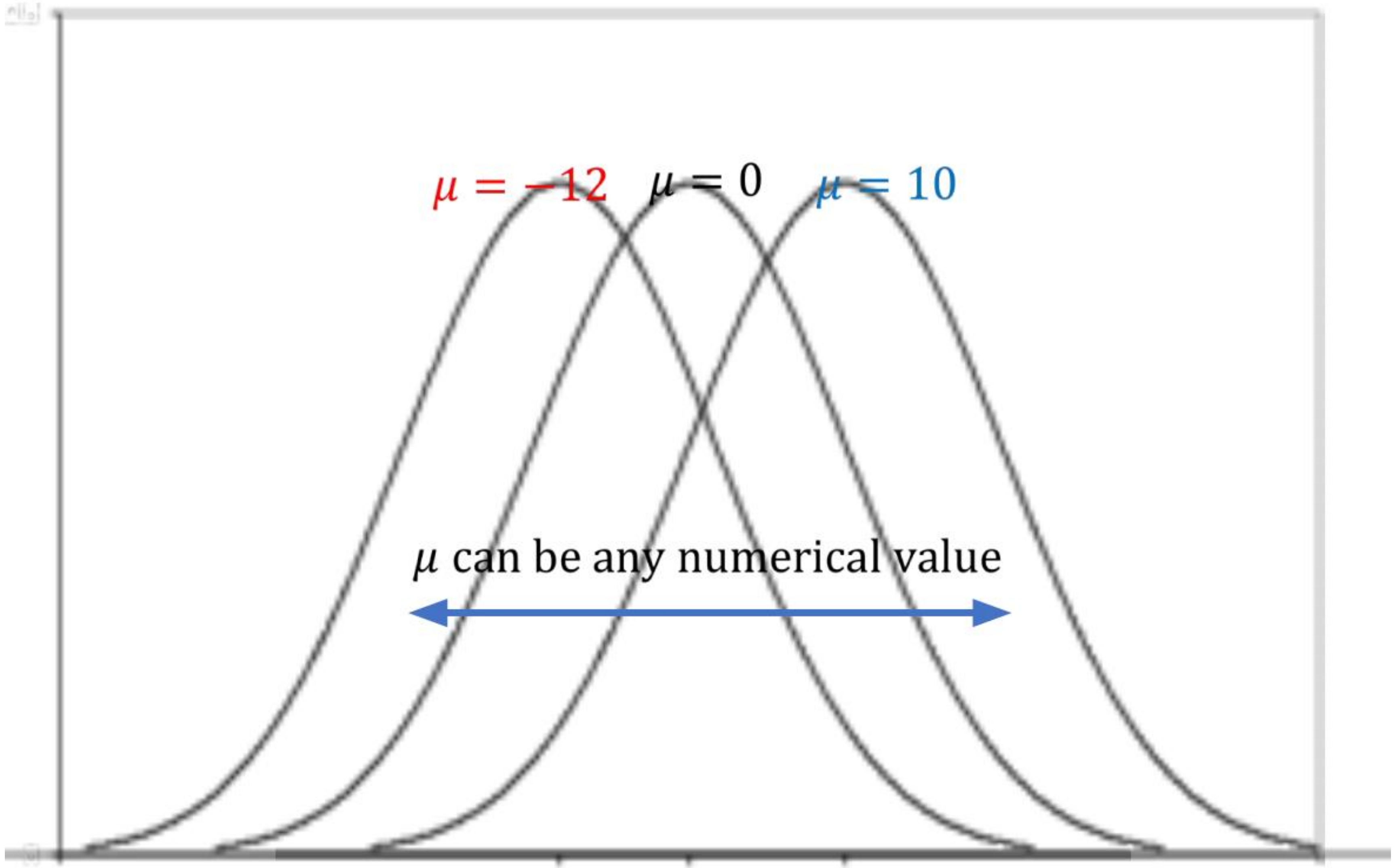
$$z = \frac{x - \mu}{\sigma}$$

- **z** is the "z-score" (Standard Score)
- **x** is the value to be standardized
- **μ** is the mean
- **σ** is the standard deviation

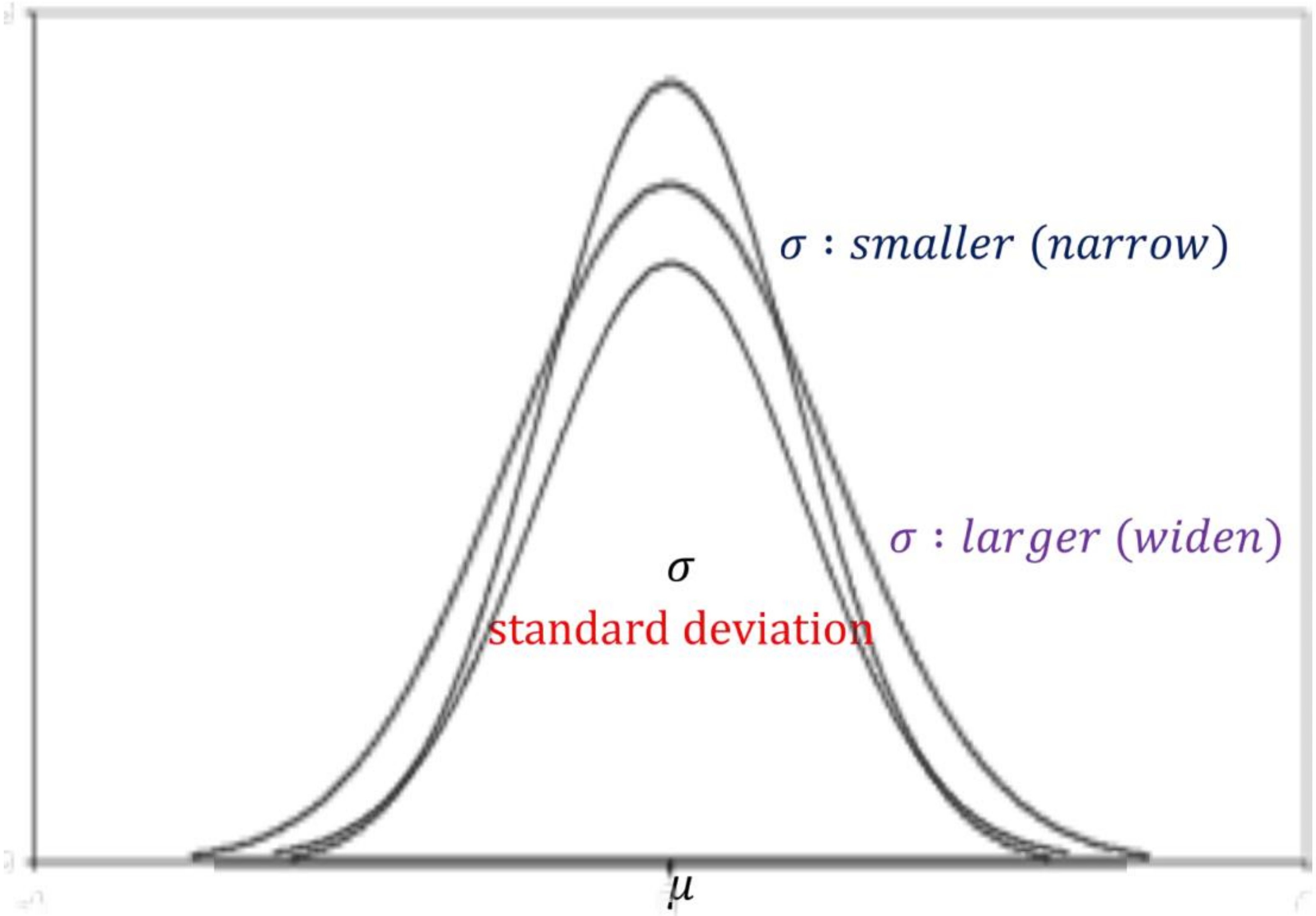
Forward School



Forward
School



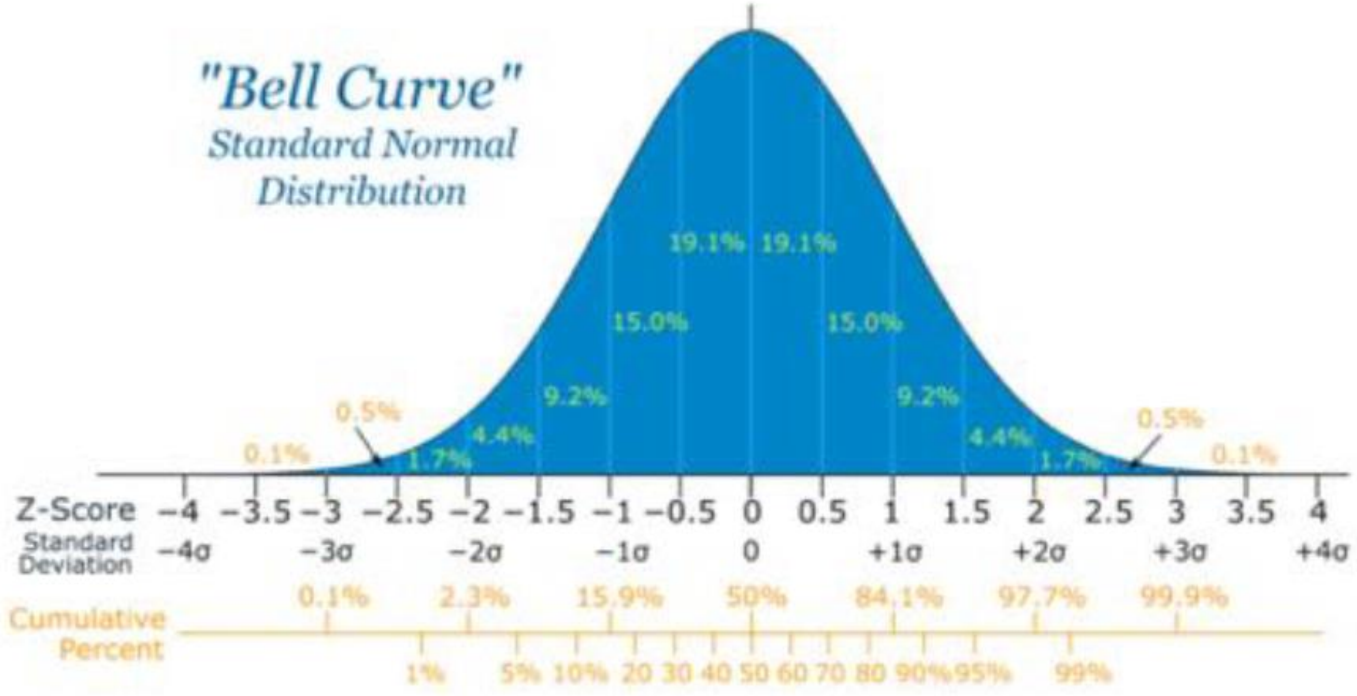
Forward School



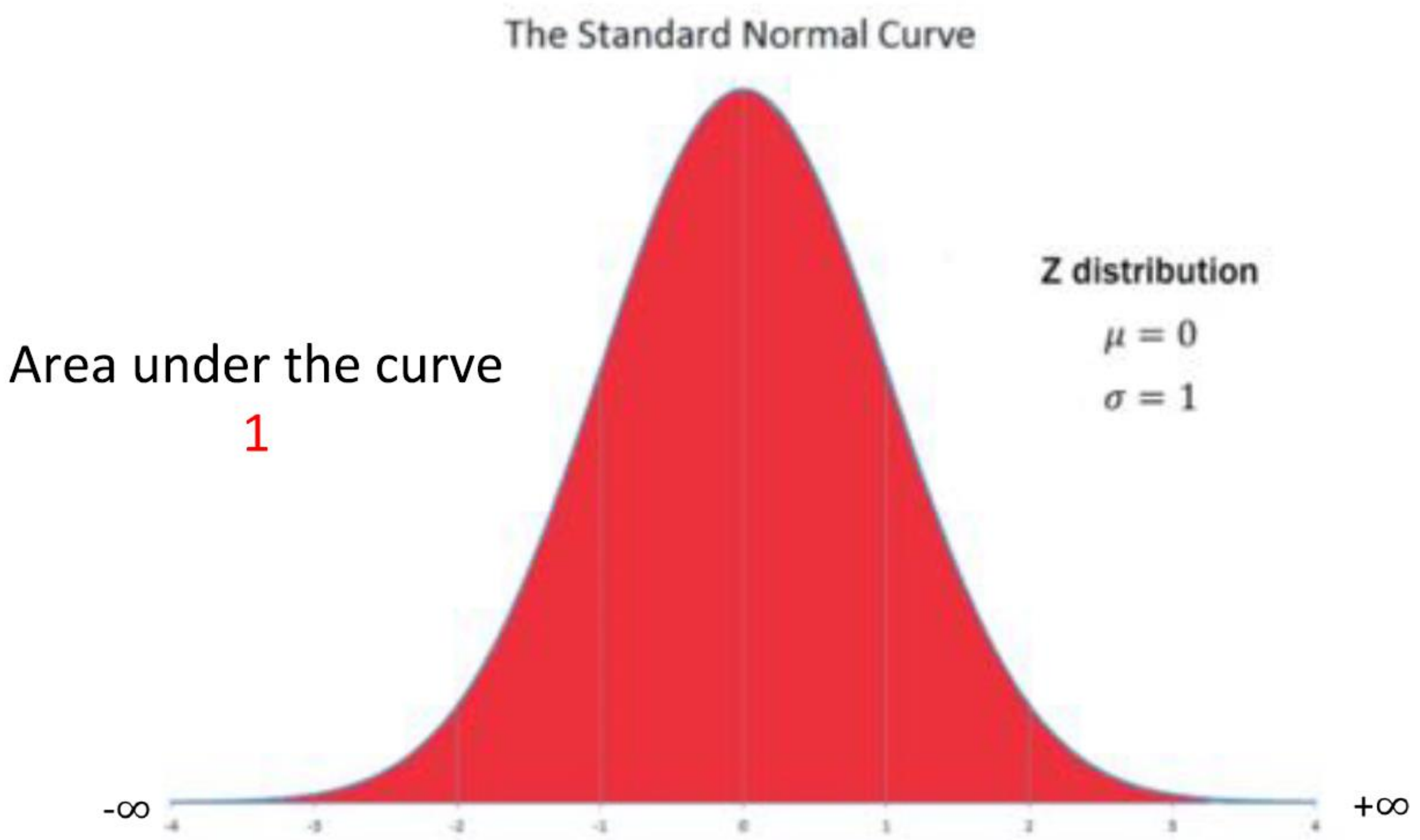
Forward School

Standard Normal Distribution

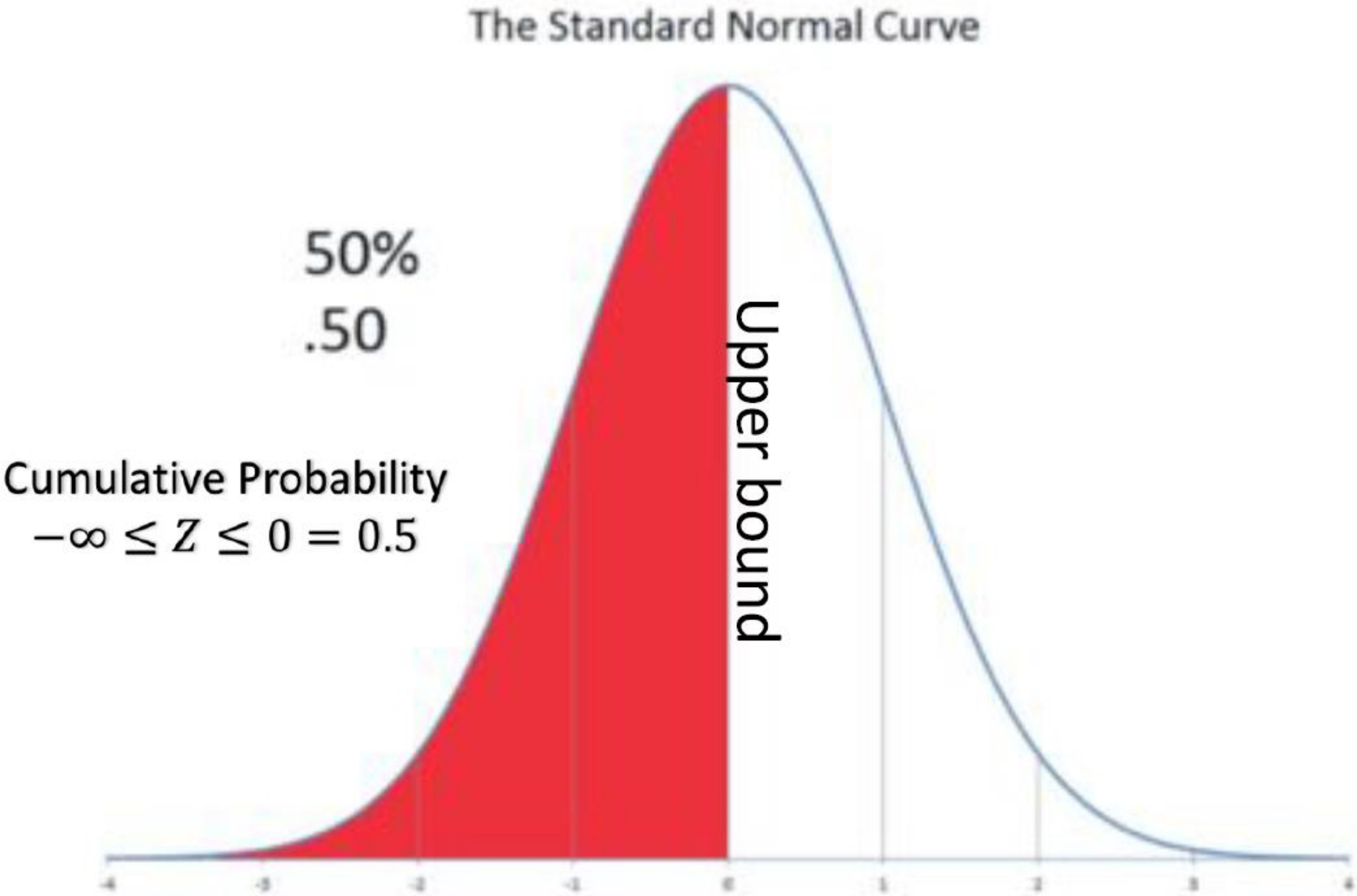
- Here is the Standard Normal Distribution with percentages for every half of a standard deviation and cumulative percentages:



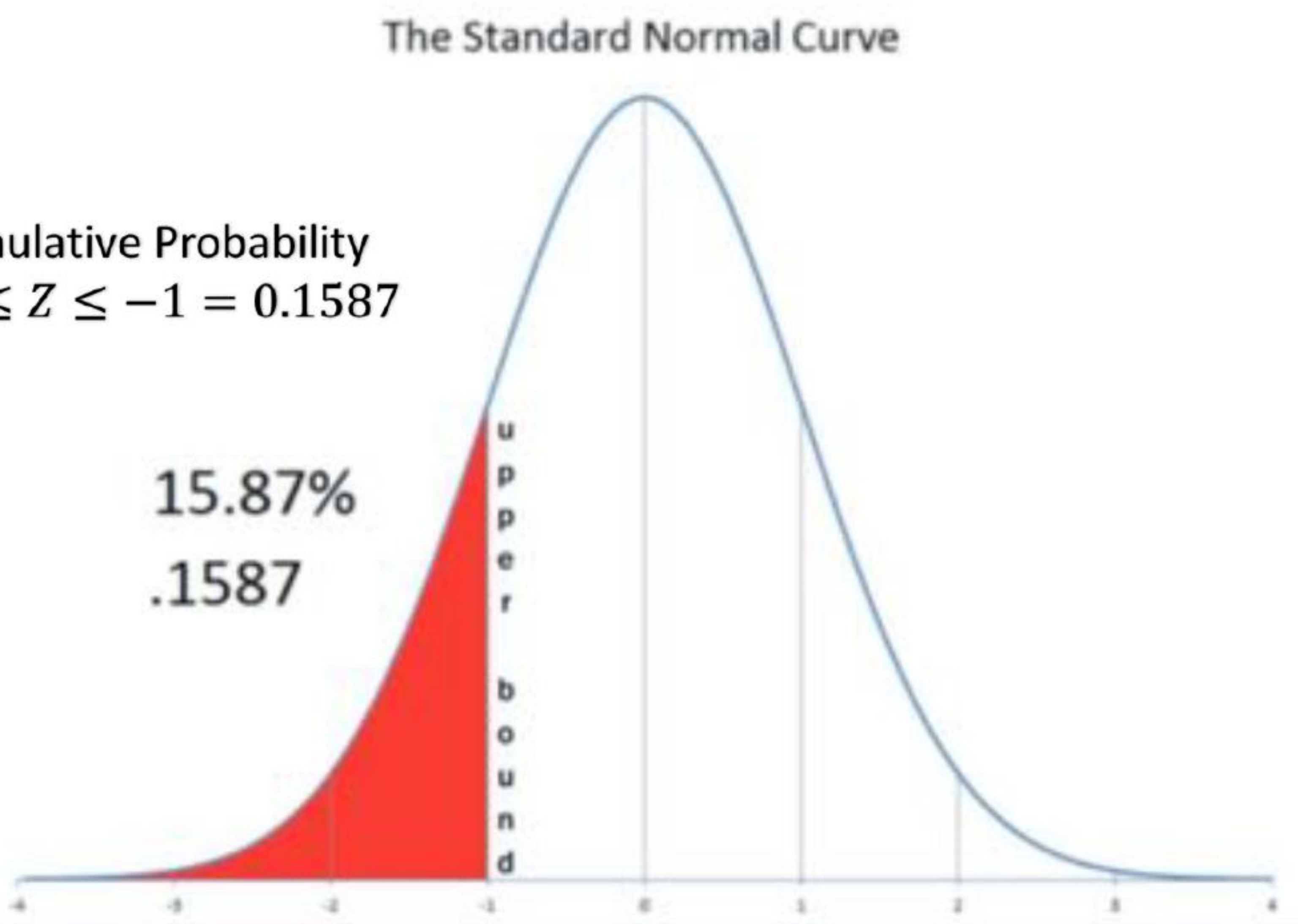
Forward School



Forward School



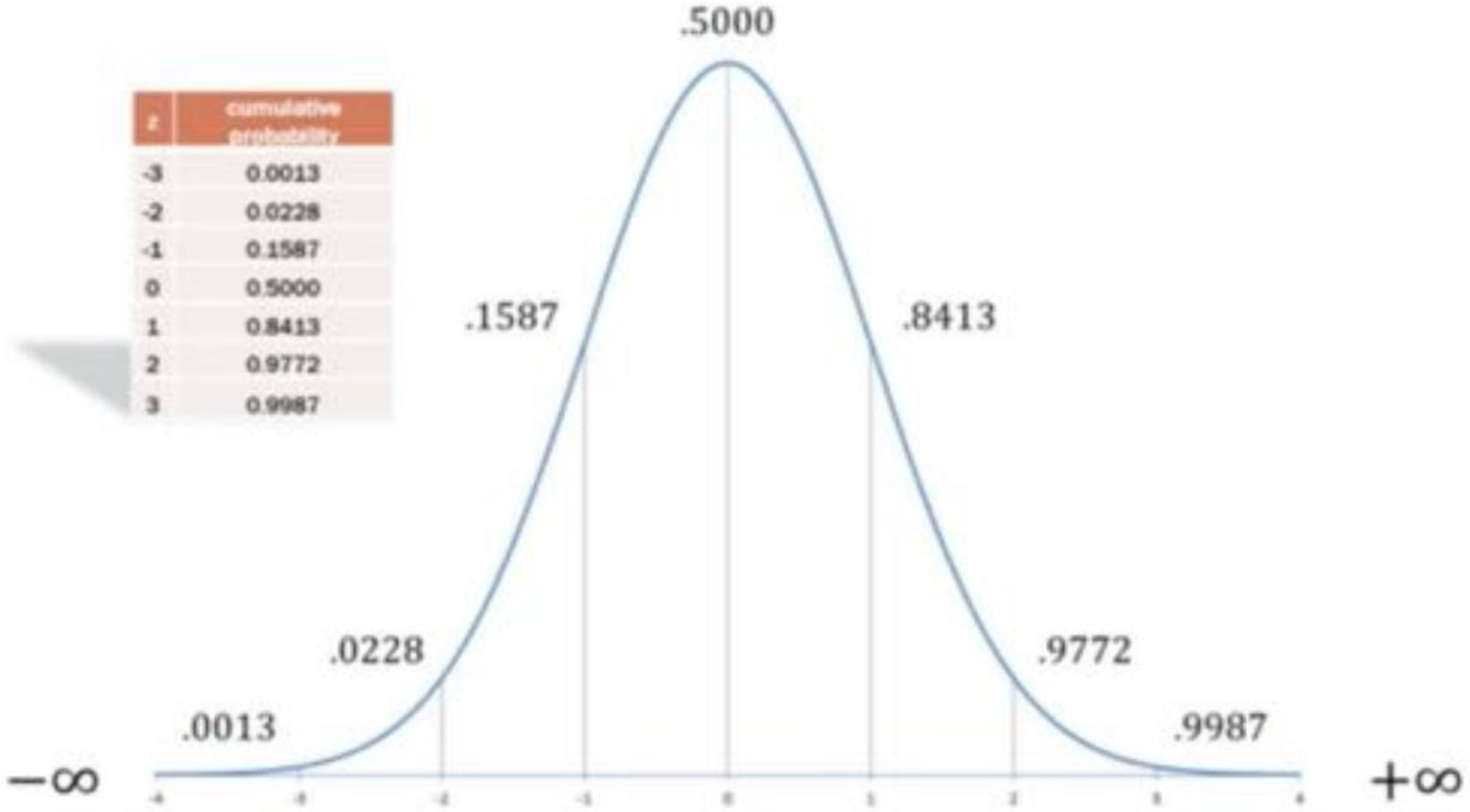
Forward School



Cumulative Probability

$$-\infty \leq Z \leq -1 = 0.1587$$

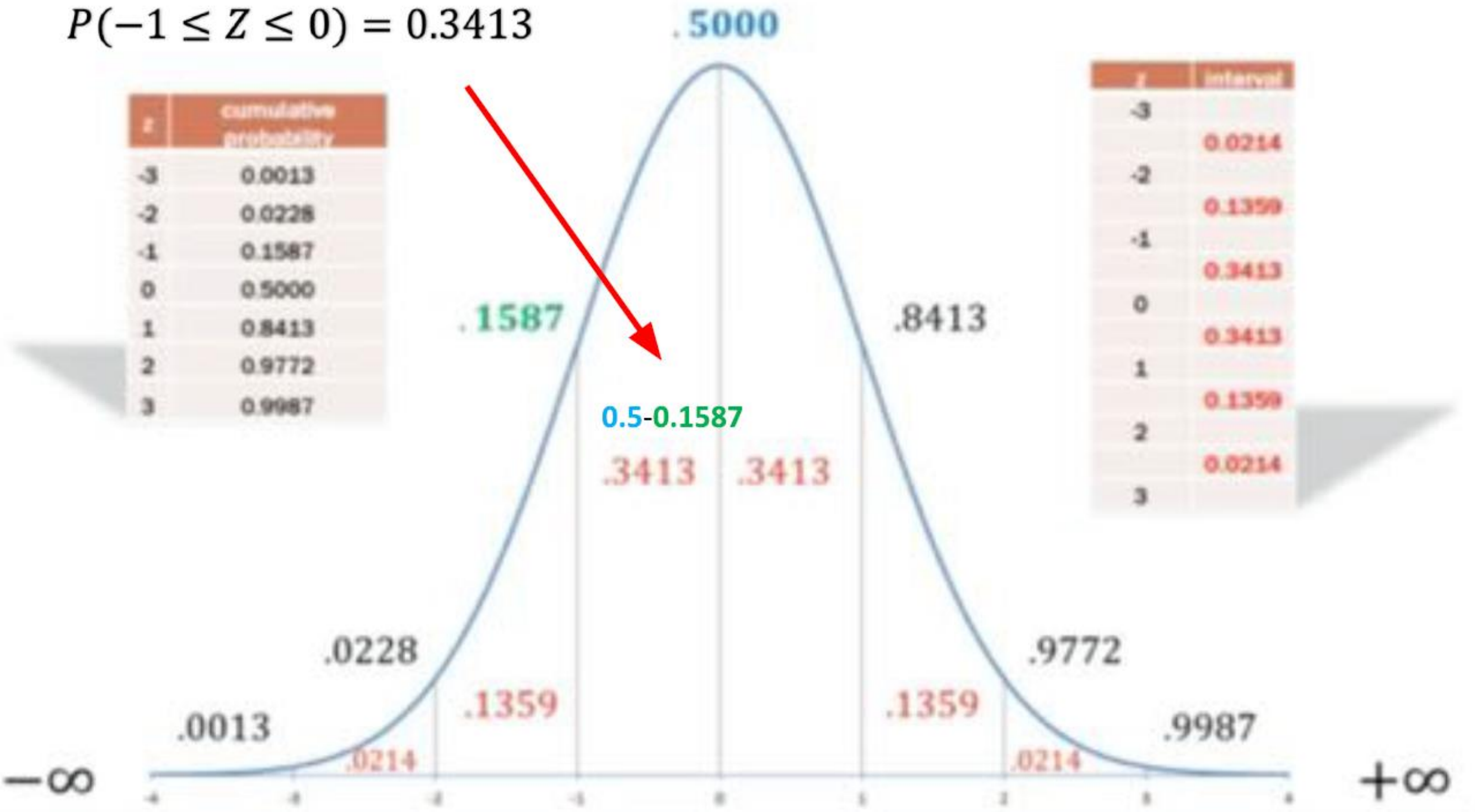
z	cumulative probability
-3	0.0013
-2	0.0228
-1	0.1587
0	0.5000
1	0.8413
2	0.9772
3	0.9987



Forward School

$$P(-1 \leq Z \leq 0) = 0.3413$$

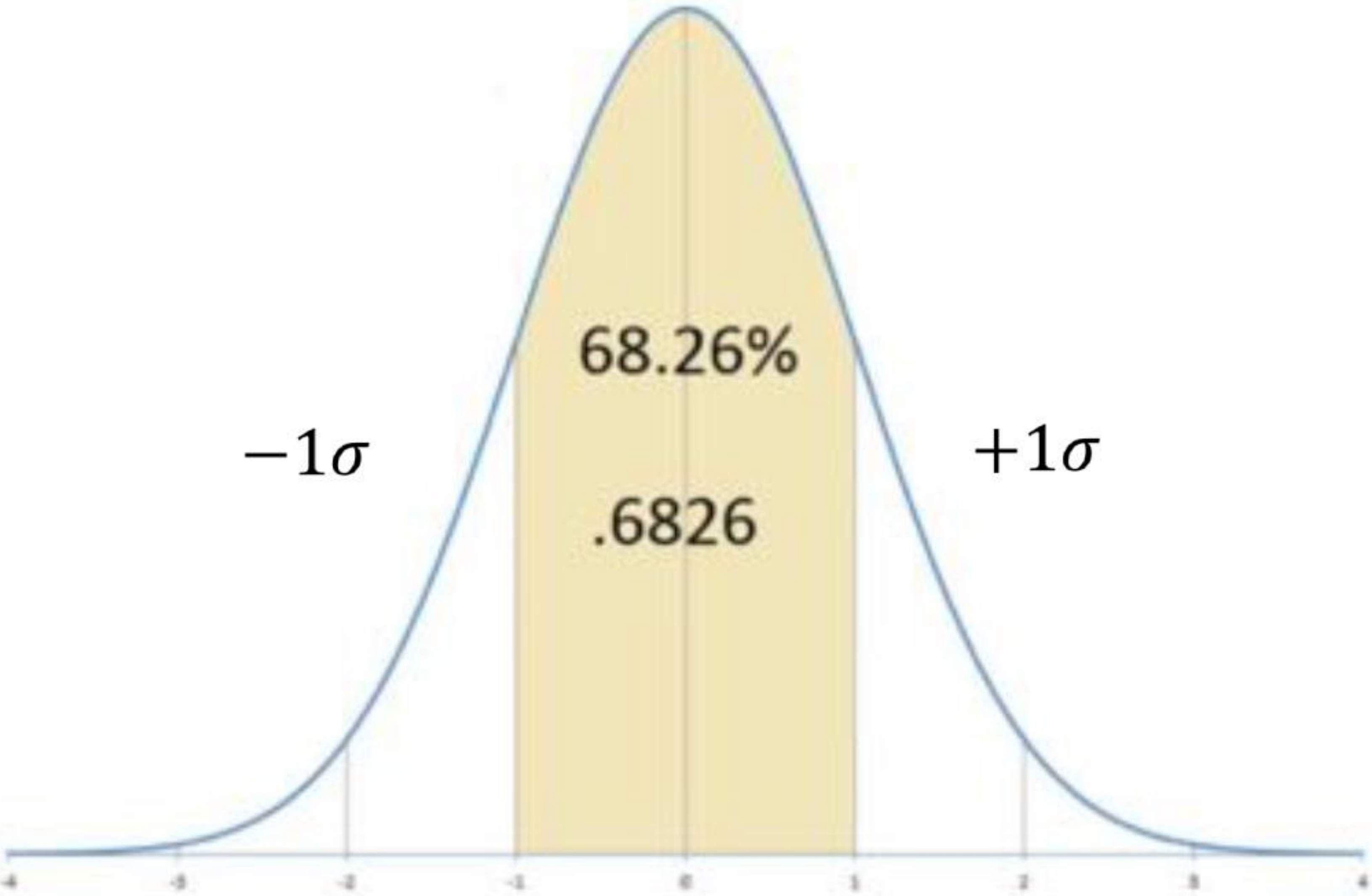
z	cumulative probability
-3	0.0013
-2	0.0228
-1	0.1587
0	0.5000
1	0.8413
2	0.9772
3	0.9987



z	interval
-3	
	0.0214
-2	
	0.1359
-1	
	0.3413
0	
	0.3413
1	
	0.1359
2	
	0.0214
3	

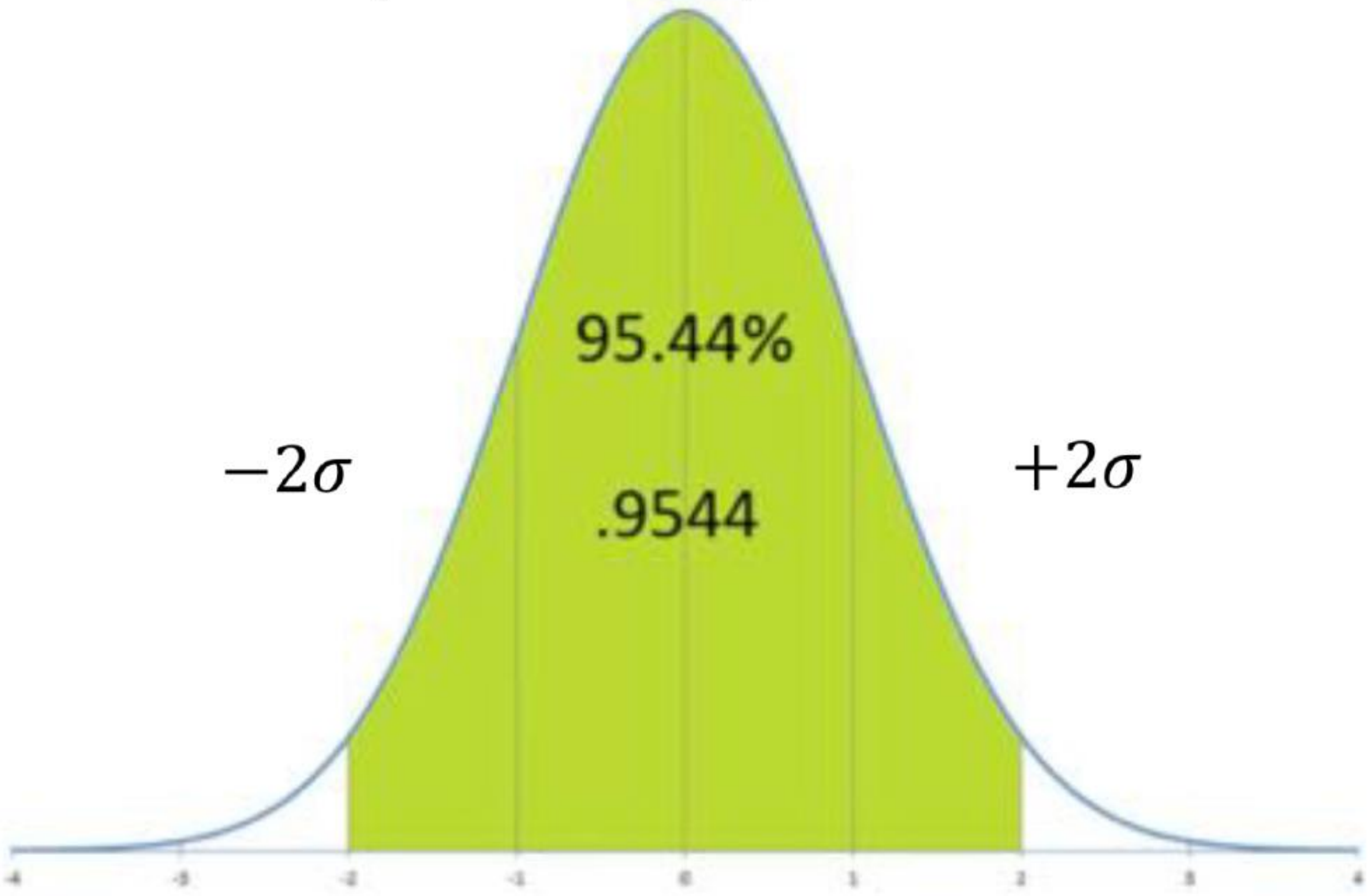
Forward School

$$P(-1 \le Z \le 1) = 0.6826$$



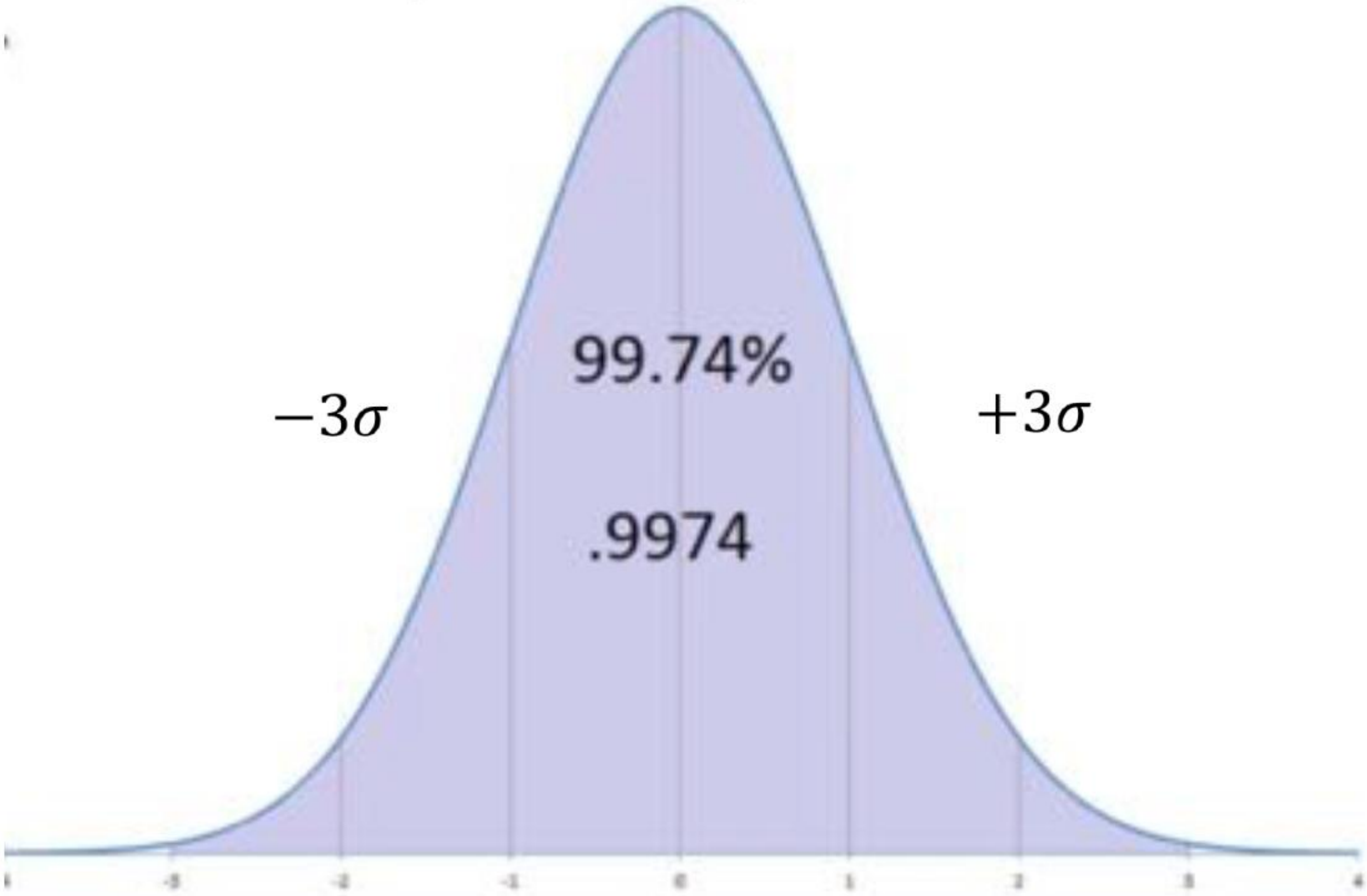
Forward School

$$P(-2 \leq Z \leq 2) = 0.9544$$

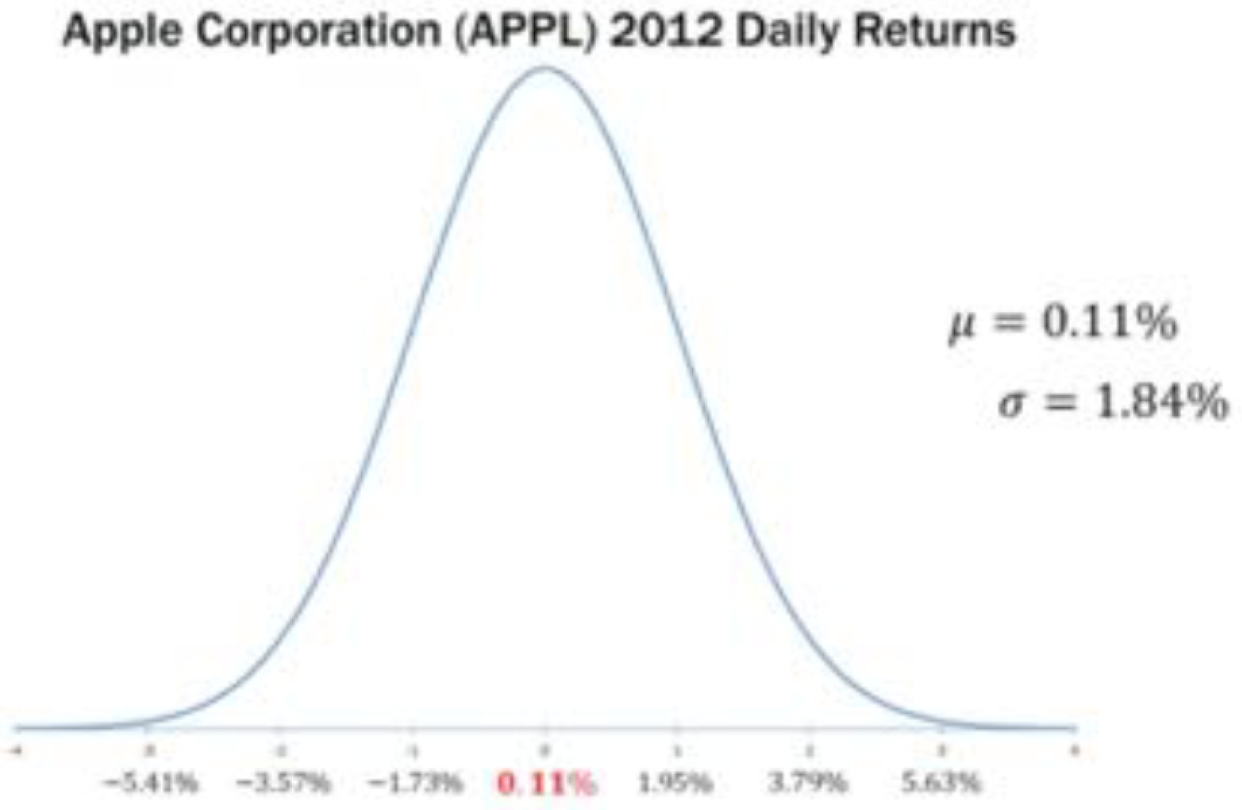


Forward School

$$P(-3 \leq Z \leq 3) = 0.9974$$



Forward School

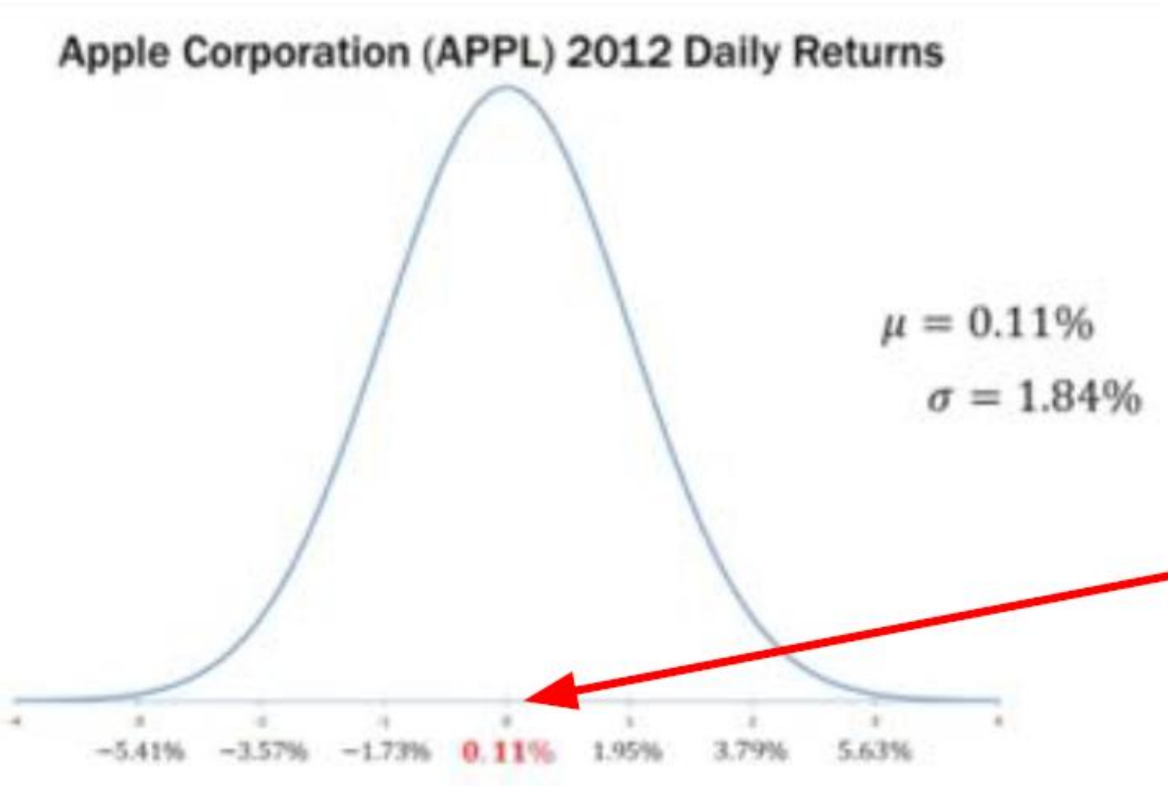


Questions:

1. What is the probability, for any given day, of a return greater than 0.5%?
2. What is the probability, for any given day, of a loss greater than 2%?
3. What is the probability, for any given day, of a return between 0% and 1%?
4. What is the probability, for any given day, of a return OR loss greater than 3%?

Forward School

What is the probability, for any given day, of a return greater than 0.5%?

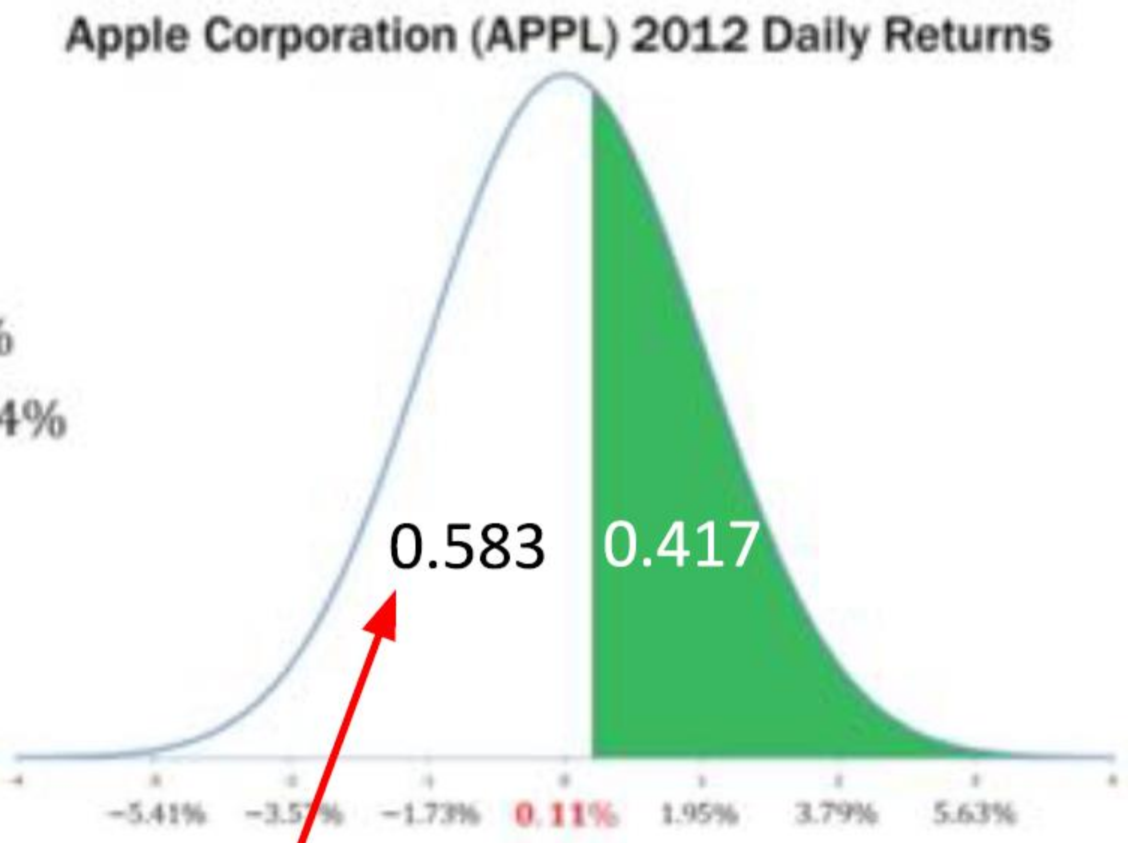


$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{0.5 - 0.11}{1.84}$$

$$z = 0.21$$

Forward School



$\mu = 0.11\%$

$\sigma = 1.84\%$

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{0.5 - .11}{1.84}$$

$$z = .21$$

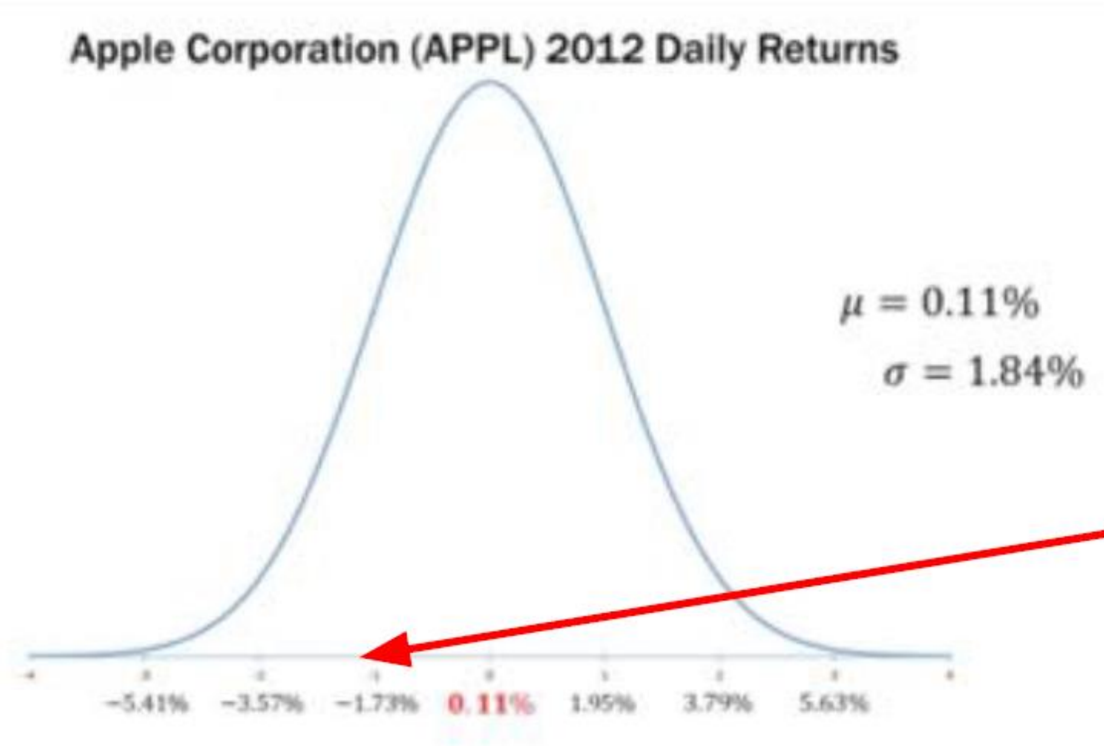
What is the probability, for any given day, of a return greater than 0.5%?

norm.cdf(0.21,0,1)   1-0.583 = 0.417

norm.cdf(0.5, 0.11, 1.84)

42%

What is the probability, for any given day, of a loss greater than 2%?



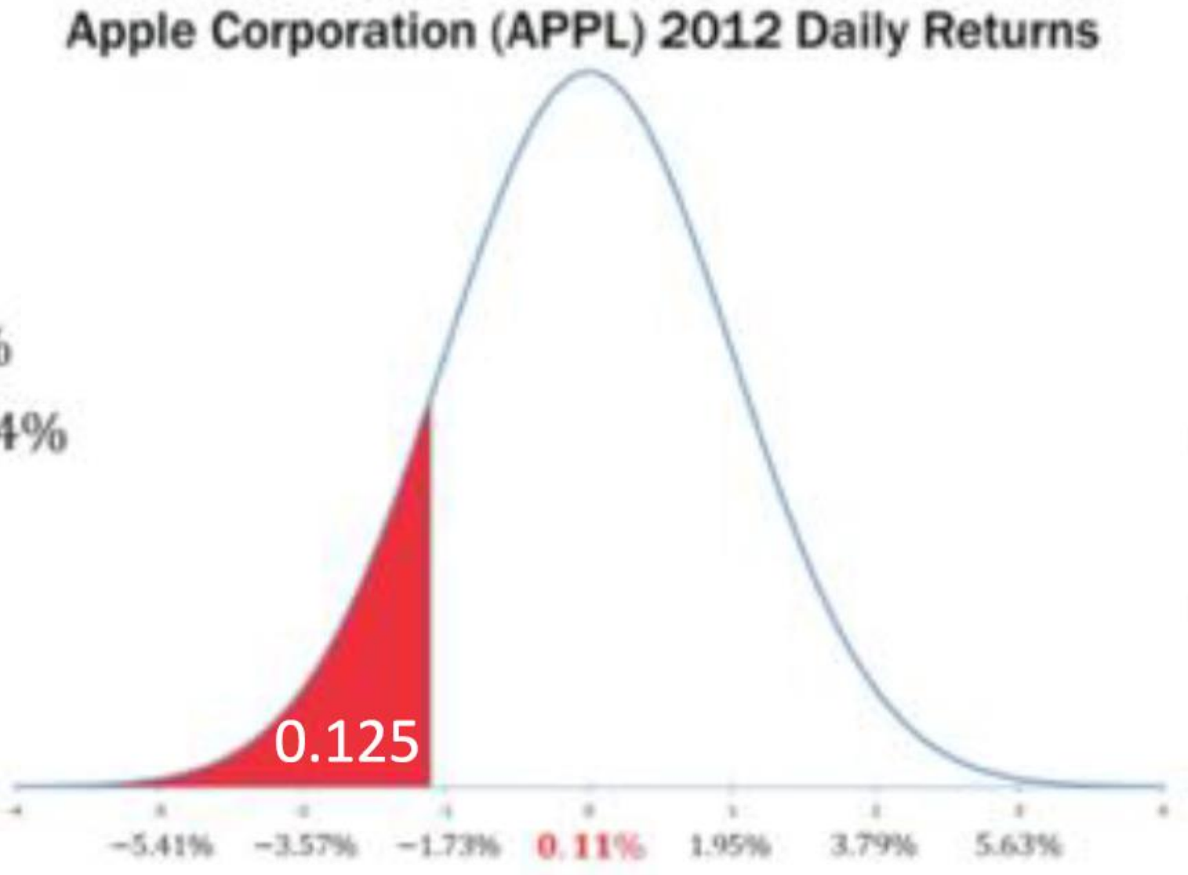
$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{-2 - 0.11}{1.84}$$

$$z = -1.15$$

Forward School

$$\mu = 0.11\%$$
$$\sigma = 1.84\%$$



$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{-2 - .11}{1.84}$$

$$z = -1.15$$

What is the probability, for any given day, of a loss greater than 2%?

norm.cdf(-1.15,0,1) = 0.125

norm.cdf(-2,0.11,1.84)

12.5%

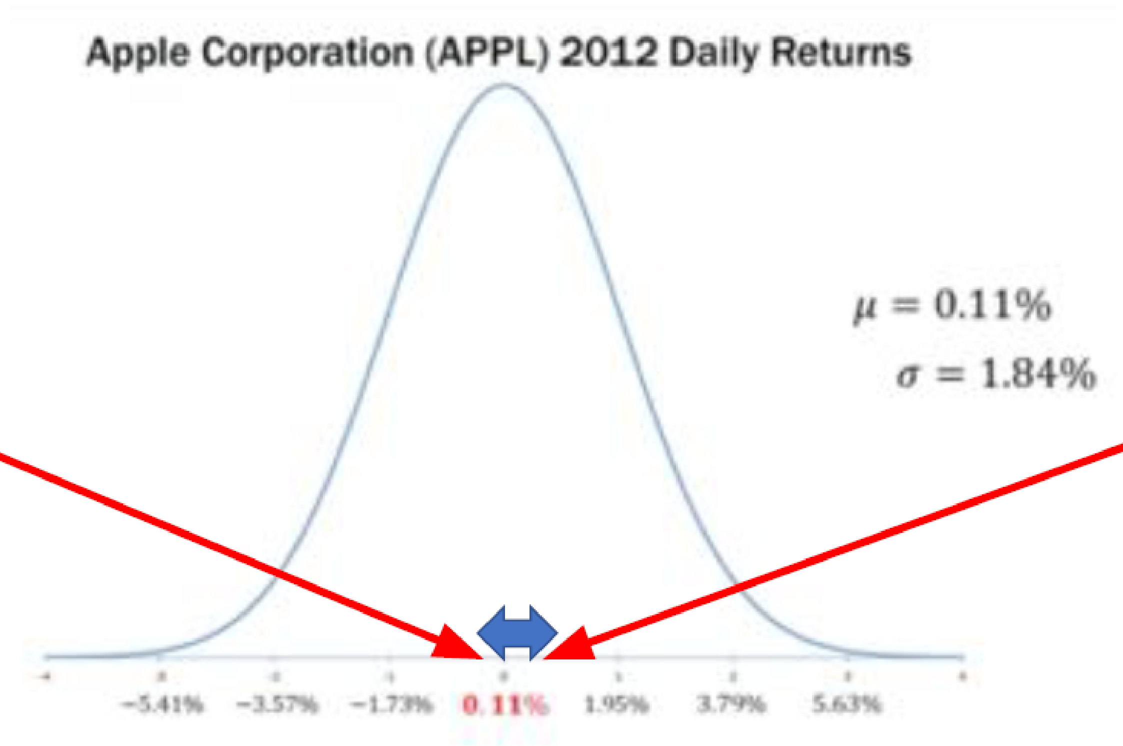
Forward School

What is the probability, for any given day, of a return between 0% and 1%?

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{0 - 0.11}{1.84}$$

$$z = -0.06$$



$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{1 - 0.11}{1.84}$$

$$z = 0.48$$

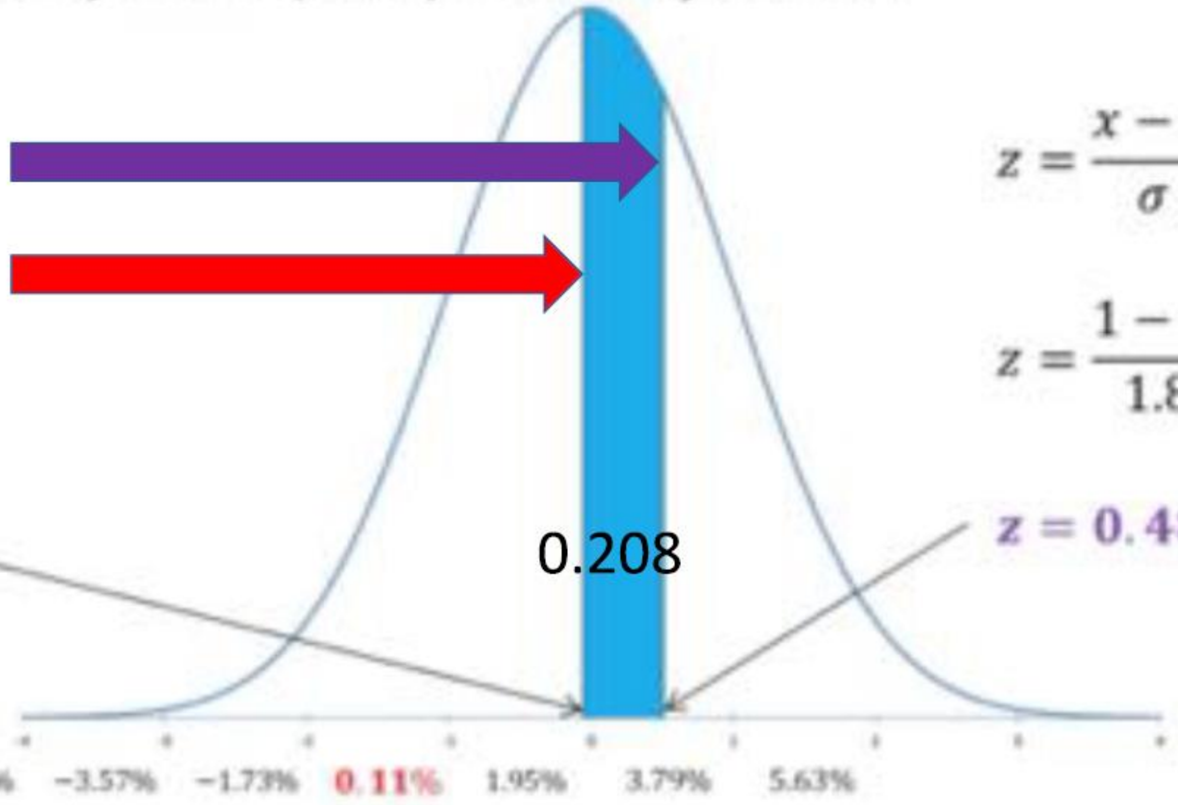
Forward School

Apple Corporation (APPL) 2012 Daily Returns

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{0 - .11}{1.84}$$

$$z = -0.06$$



$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{1 - .11}{1.84}$$

$$z = 0.48$$

What is the probability, for any given day, of a return between 0% and 1%?

norm.cdf(0.48,0,1) = 0.684 - norm.cdf(-0.06,0,1) = 0.476 = 0.208

norm.cdf(1,0.11,1.84)-norm.cdf(0,0.11,1.84)

21%

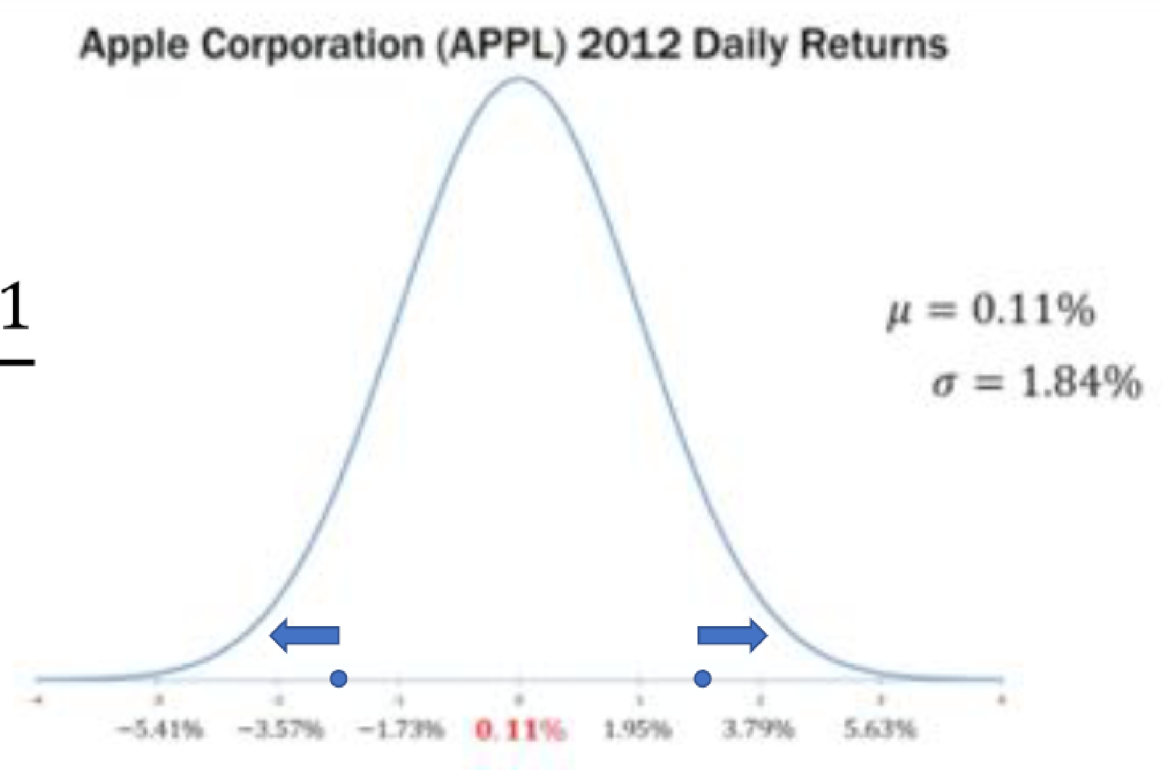
Forward School

What is the probability, for any given day, of a return OR loss greater than 3%?

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{-3 - 0.11}{1.84}$$

$$z = -1.69$$



$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{3 - 0.11}{1.84}$$

$$z = 1.57$$

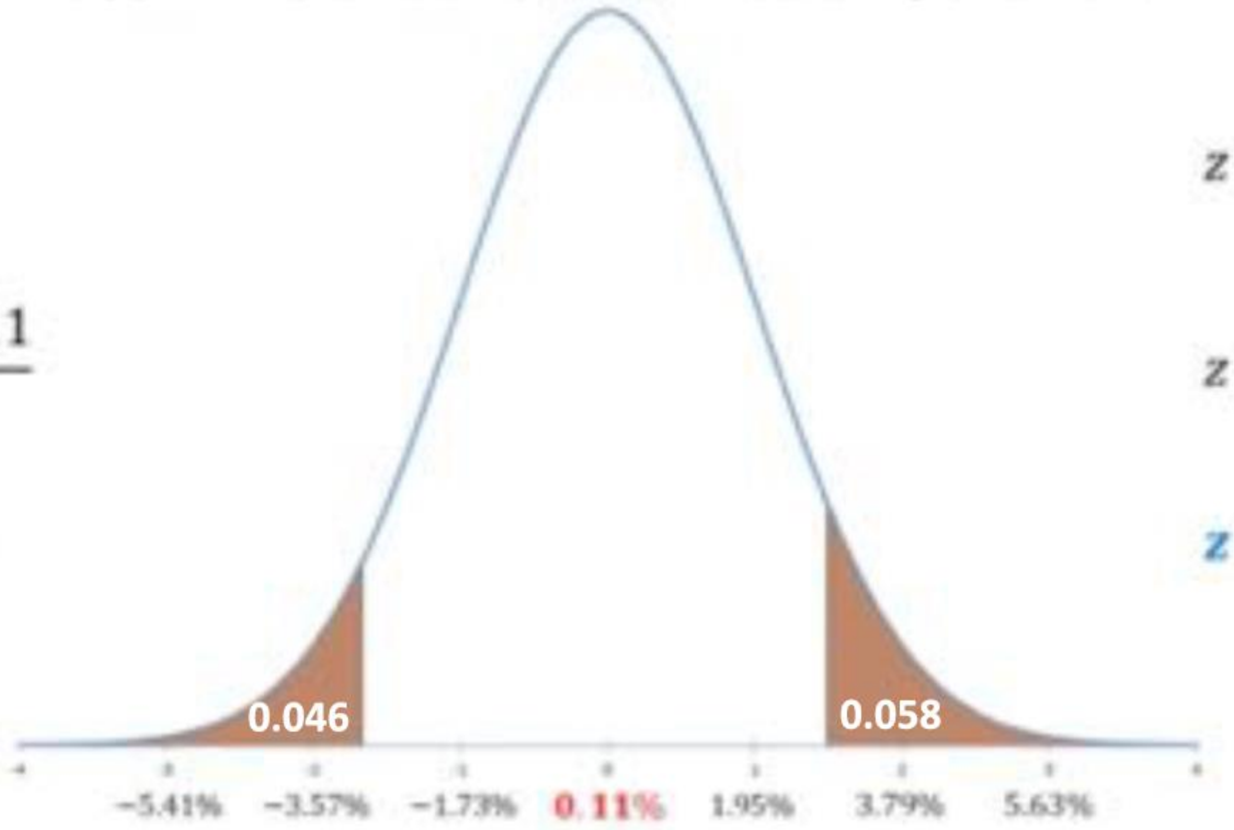
Forward School

Apple Corporation (APPL) 2012 Daily Returns

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{-3 - .11}{1.84}$$

$$z = -1.69$$



$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{3 - .11}{1.84}$$

$$z = 1.57$$

10.4%

What is the probability, for any given day of a return OR loss greater than 3%?

norm.cdf(-1.69,0,1) = 0.046 + 1-norm.cdf(1.57,0,1) = 0.058

0.046+ 0.058 = 0.104 or 10.4%

norm.cdf(-3,0.11,1.84)+(1-norm.cdf(3,0.11,1.84)

Forward School

Normal Distribution with Scipy

- https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html
- The normal distribution is defined by the following probability density function, where μ is the population mean and σ^2 is the variance.
- In particular, the normal distribution with μ = 0 and σ = 1 is called standard normal distribution and denoted as N(0, 1).
- The normal distribution is important because of the Central Limit Theorem, which states that the population of all possible samples of size *n* from a population with mean μ and variance σ^2 approaches a normal distribution with mean μ and σ^2 /n when n approaches infinity.

```
norm.pdf(x, loc = 0, scale = 1) Probability Density Function
norm.cdf(x, loc = 0, scale = 1) Cumulative Density Function
norm.ppf(x, loc = 0, scale = 1) Percent Point Function
norm.rvs(loc=0, scale=1, size=1, random_state=None) Generate random normal numbers
```

Forward School

```
norm.pdf(x, loc = 0, scale = 1)
```

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- The function `pdf` returns the value of the probability density function for the normal distribution given parameters for x, μ, and σ.
- Although x represents the independent variable of the pdf for the normal distribution, it's also useful to think of x as a Z-score.

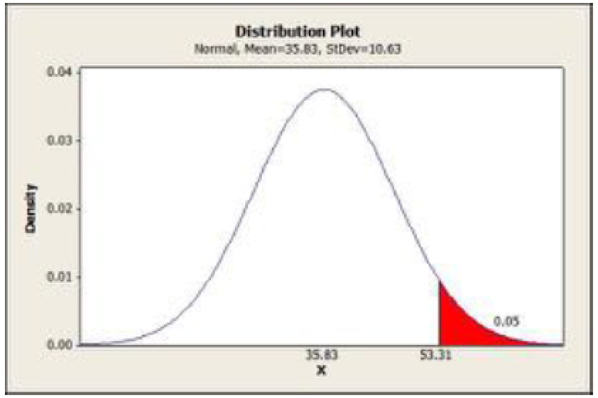
```
norm.pdf(2, loc = 5, scale = 3)
```

Forward School

```
norm.cdf(x, loc = 0, scale = 1)
```

- The function `cdf` returns the integral from $-\infty$ to x of the pdf of the normal distribution where x is a Z-score.
- The `cdf` function also takes the argument lower tail. For upper tail, it is equivalent to `cdf` returns the integral from x to ∞ of the pdf of the normal distribution. Note that upper tail for `norm.cdf(x)` is the same as `1-norm.cdf(x)`

```
norm.cdf(2, 5, 3)
```



`norm.ppf(x, loc = 0, scale = 1)`

- The `norm.ppf` function is simply the inverse of the cdf, which you can also think of as the inverse of `norm.cdf`

```
# What is the Z-score of the 50th quantile of the normal distribution?
> norm.ppf(0.5)
# [1] 0
# What is the Z-score of the 99th quantile of the normal distribution?
> norm.ppf(0.99)
# [1] 2.326348
```

Forward School

	PURPOSE	SYNTAX	EXAMPLE
norm.rvs	Generates random numbers from normal distribution	norm.rvs(loc=0, scale=1, size=1, random_state=None)	norm.rvs(1000, 3, .25) Generates 1000 numbers from a normal with mean 3 and sd=.25
norm.pdf	Probability Density Function (PDF)	norm.pdf(x, loc = 0, scale = 1)	norm.pdf(0, 0, .5) Gives the density (height of the PDF) of the normal with mean=0 and sd=.5.
norm.cdf	Cumulative Distribution Function (CDF)	norm.cdf(x, loc = 0, scale = 1)	norm.cdf(1.96, 0, 1) Gives the area under the standard normal curve to the left of 1.96, i.e. ~0.975
norm.ppf	Quantile Function – inverse of pnorm	norm.ppf(x, loc = 0, scale = 1)	norm.ppf(0.975, 0, 1) Gives the value at which the CDF of the standard normal is .975, i.e. ~1.96

Forward School

Exercise 1

- With a normal random variable with mean 22 and variance 25, find put the following probabilities:
  - Lies between 16.2 and 27.5
  - Is greater than 29
  - Is less than 17
  - Is less than 15 or greater than 25

Exercise 2

- Assume that the test scores of a college entrance exam fits a normal distribution. Furthermore, the mean test score is 72, and the standard deviation is 15.2 What is the percentage of students scoring 84 or more in the exam?