Mathematics for Analytics and Finance

Sami Najafi

MSIS2402/2405

Module 7

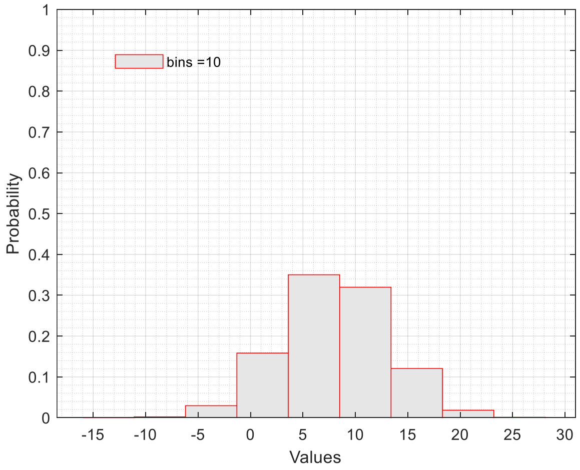
Continuous Random Variables

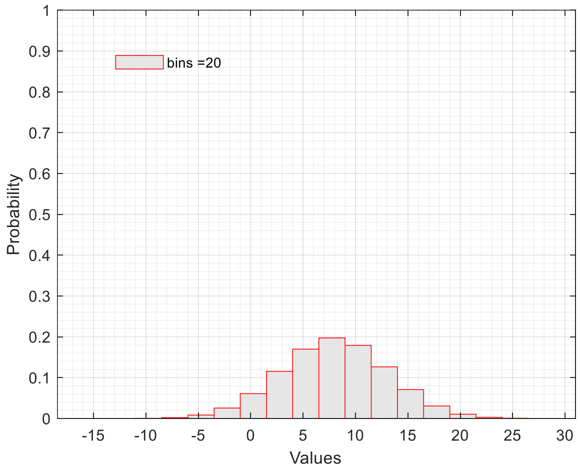
Continuous Random Variables

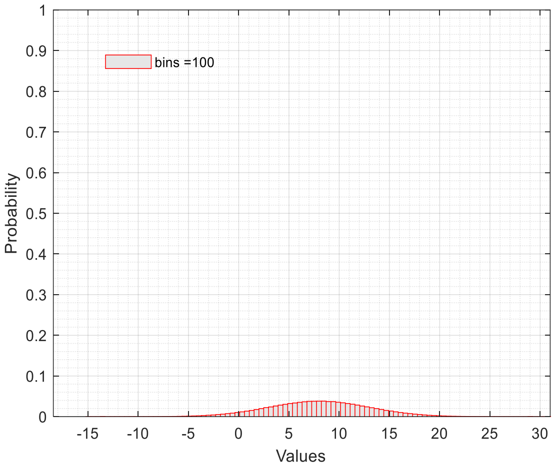
- A continuous random variable is a variable that can assume any value on a continuum (that is, the set of possible values is continuous or "connected with no gaps", i.e., each point is equal to the left and right limits to that point)
  - thickness of an item
  - time required to complete a task
  - temperature of a solution
  - height, in inches

Increasing The Number of Bins

Remark: As the number of bins increases, the probability associated with each bin decreases. Eventually, as the bins become extremely narrow, the probability of each individual bin approaches zero.



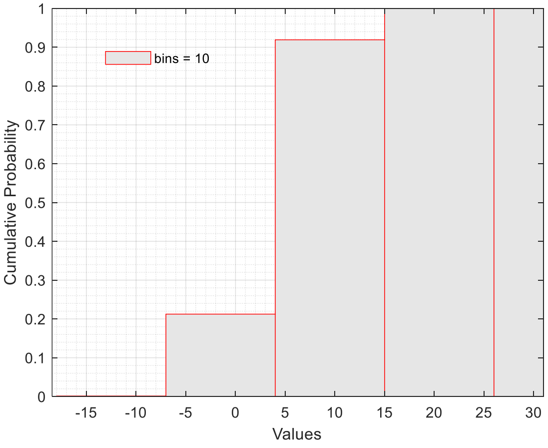


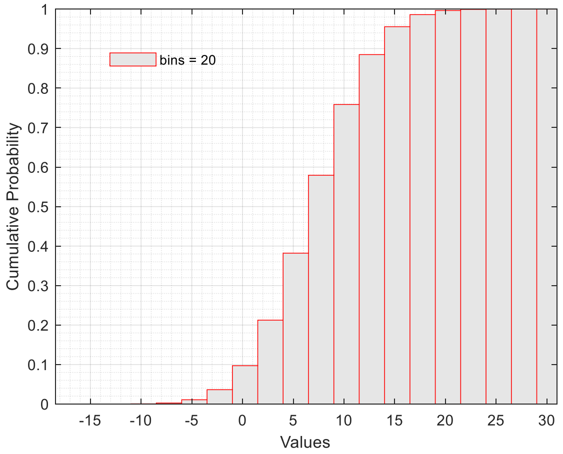


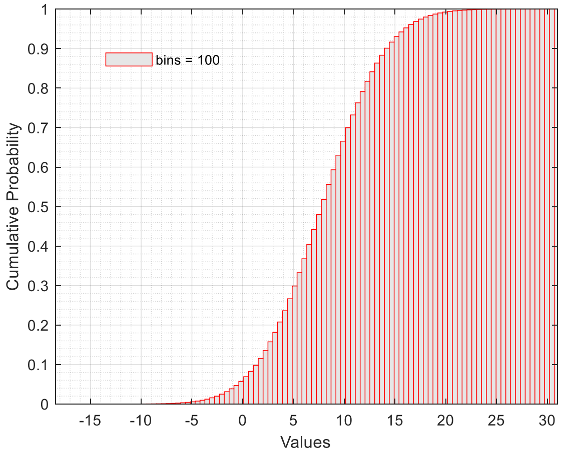
Increasing The Number of Bins

- As the number of bins increases, the cumulative probability (CDF) becomes increasingly continuous and smooth eliminating all jumps. So, it becomes differentiable, meaning that $f(x) = \frac{dF(x)}{dx}$ is defined for all points.

- Given this, by the Fundamental Theorem of Calculus: $F(x) = \int_{-\infty}^{x} f(u)du\,,\ x \in \mathbb{R}$.







Continuous Random Variables

Definition: The random variable X is called continuous if its distribution function can be expressed as

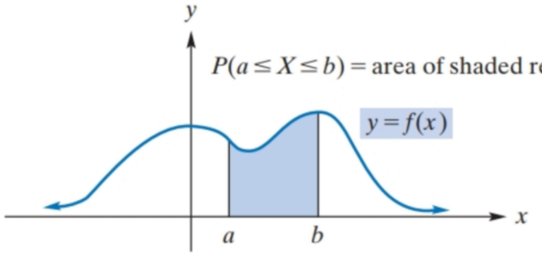
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du\,,\ \ x \in \mathbb{R}.$$

For some integrable (continuous) function $f\colon \mathbb{R} \to [0, \infty)$, called the probability density function of X. $f(x)$ is called a (probability) density function, and has the following properties:

- $f(x) \geq 0$ (because $F(x)$ is increasing in x)
- $\int_{-\infty}^{\infty} f(x)dx = 1$ (because $P(X \leq \infty) = 1$)
- $P(a \leq X \leq b) = \int_a^b f(x)dx$ (because of the Fundamental Theorem of Calculus)

Remark: The probability of any individual value is zero (i.e., $P(X = a) = 0$ for any value a). Hence:

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



Probability density function.

Continuous Random Variables

Remark: When we discretize x, we mean is an acceptable range for $x \leq X < x + \Delta x$. By Taylor Expansion (or the notion of differential we studied in Module 2):

$$F(x+\Delta x) = F(x) + f(x)\Delta x + o(\Delta x) \Rightarrow P(x < X \leq x + \Delta x) = F(x+\Delta x) - F(x) = \Delta F(x) = f(x)\Delta x + o(\Delta x)$$

$$P(X = x) \approx \Delta F(x) \approx f(x)\Delta x$$

Similarly (see Module 2):

$$\mu = E(X) = \lim_{n\to\infty} \sum_{k=1}^{n} x_k p(x_k) = \lim_{n\to\infty} \sum_{k=1}^{n} x_k f(x_k)\Delta x = \int_a^b x f(x) dx$$

$$\Delta x = \frac{b-a}{n}, \qquad x_k = k\Delta x, \qquad f(x_k) = f(k\Delta x)$$

All results in the continuous setting can be obtained through the limit of discretized values $x_k = k\Delta x$ with probabilities $p(x_k) \approx f(x_k)\Delta x$. These results will remain the same as discrete RV case, with the only difference that integration replaces summation and $f(x)dx$ replaces $p(x)$.

Summary of Some Properties

For a random variable X with density function f:

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$\sigma^2 = var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

For a random variable X with density function f and a real function $g(X)$:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

For random variables X and Y with the joint density function $f(X,Y)$:

$$cov(X,Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(x - E(X)\big)\big(y - E(Y)\big) f(x,y)\, dxdy = E(XY) - E(X)E(Y)$$

Two random variables X and Y with the joint density function $f(X,Y)$ are independent if and only if

$f(a,b) = f_X(a)f_Y(b)$ for all $a, b \in \mathbb{R}$

Example

For a random variable X with density function:

$$f(x) = \begin{cases} ax & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a) Find the value of a

b) Find the mean and variance of X

Solution

a)

$$\int_0^2 f(x)dx = 1 \Rightarrow \int_0^2 axdx = 2a = 1 \Rightarrow a = \frac{1}{2}$$

b)

$$\mu = \int_0^2 xf(x)dx = \frac{1}{2}\int_0^2 x^2dx = \frac{8}{6} = \frac{4}{3}$$

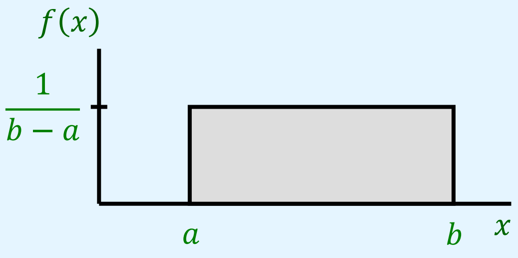
$$E(X^2) = \int_0^2 x^2 f(x)dx = \frac{1}{2}\int_0^2 x^3dx = \frac{2^4}{8} - 0 = 2$$

$$var(X) = E(X^2) - \mu^2 = 2 - \frac{16}{9} = \frac{2}{9}$$

The Uniform Distribution

We say that X is a uniform random variable on the interval (a, b) if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & , \text{if } a < x < b \\ 0 & , \text{otherwise} \end{cases}$$



Remark: all the possibilities have a uniform density

If X is a uniform random variable on the interval (a, b) then

$$E(X) = \frac{a+b}{2}, \qquad var(X) = \frac{(b-a)^2}{12}$$

The Uniform Mean and Variance

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \left. \frac{x^2}{2(b-a)} \right|_a^b = \frac{a+b}{2}$$

$$E(X^2) = \int_a^b \frac{x^2}{b-a} dx = \left. \frac{x^3}{3(b-a)} \right|_a^b = \frac{a^2+b^2+ab}{3}$$

$$var(X) = E(X^2) - E(X)^2 = \frac{a^2+b^2+ab}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{4(a^2+b^2+ab) - 3(a^2+b^2+2ab)}{12} = \frac{(b-a)^2}{12}$$

Example

Let X be a random variable from a uniform distribution over the range $2 \leq X \leq 6$.

(i) Write the probability distribution.

(ii) Find the mean and standard deviation of X.

(iii) What is the probability that X takes a value between 3 and 5?

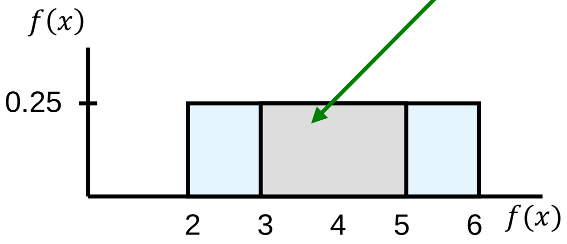
Solution

$$f(x) = \frac{1}{6-2} = \frac{1}{4}, \qquad 2 \le x \le 6$$

$$\mu = \frac{a+b}{2} = \frac{2+6}{2} = 4$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(6-2)^2}{12}} = 1.1547$$

$$P(3 \le X \le 5) = \frac{5-3}{6-2} = 0.5$$



The Exponential Distribution

Often used to model the length of time between the occurrence of two successive events (the interarrival times)

Examples:

- Time between two successive trucks arriving at an unloading dock
- Time between two successive transactions at an ATM Machine
- Time between two successive phone calls to the main operator

Note: in many cases, service times and waiting times are exponentially distributed.

Exponential Distribution

A random variable T is an exponential random variable with parameter λ, for some $\lambda > 0$:

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & , if\ t \geq 0 \\ 0 & , if\ t < 0 \end{cases}$$

where:

- λ: the average number of events in a unit of the interval (the rate of occurrence of the events)
- t: the interval of time or space over which events occur (e.g., a specific time duration, area, or volume)

The Exponential Distribution

Remark: The CDF of an exponential random variable is:

$$F_T(t) = P(T \leq t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}$$

In R: $F_T(t) = \text{pexp}(t, \lambda)$

Remark: If X is an exponential random variable with parameter λ then its mean and variance are:

$$E(T) = \frac{1}{\lambda}, \qquad var(T) = \frac{1}{\lambda^2}$$

Example

Customers arrive at the service counter at the rate of 15 per hour. What is the probability that the inter-arrival time between two customers is less than three minutes?

Solution

$$\left.\begin{aligned} \lambda &= 15/\mathrm{hr} \\ t &= \frac{1}{20}\,\mathrm{hr} \end{aligned}\right\} \Rightarrow \lambda t = \frac{15}{20} = 0.75$$

$P\left(T < \dfrac{1}{20}\right) = 1 - e^{-\lambda t} = 1 - e^{-0.75} = 0.5276$ = `pexp(0.05,15)`

Example

A company that receives most of its orders by telephone conducted a study to determine how long customers were willing to wait on hold before ordering a product. The length of waiting time was found to be a variable best approximated by an exponential distribution with a mean waiting time of 3 minutes. What proportion of customers having to wait more than 4.5 minutes before placing an order?

Solution

$$\lambda = \frac{1}{E(T)} = \frac{1}{3} / \text{ min}$$

$$t = 4.5 \text{ min}$$

$$P(T \geq t) = e^{-\lambda t} = e^{-\frac{4.5}{3}} = e^{-1.5} = 0.2231$$

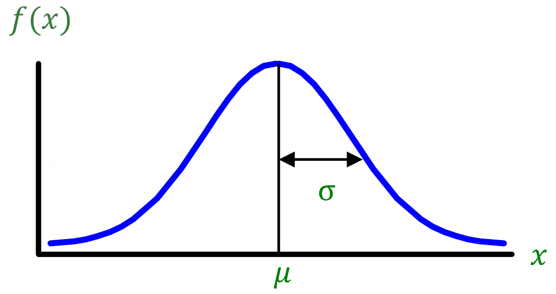
The Normal Distribution

Most important distribution in statistics (also called "Gaussian")

- Often occurs, e.g., heights, IQ, returns, errors
- Used to approximate others (even discrete), e.g., Binomial
- Useful in problems involving large amount of data

- Bell Shaped
- Symmetric
- Mean = Median = Mode

- Location is determined by the mean, μ
- Spread is determined by the standard deviation, σ
- The random variable has an infinite range: $-\infty$ to $+\infty$
- We use the notation: $N(\mu, \sigma)$

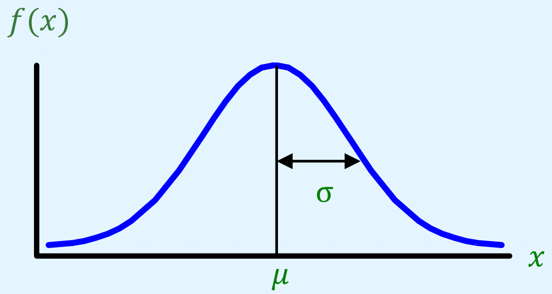


The Normal Distribution

We say that X is a normal random variable (or simply that X is normally distributed)

with parameters μ and σ if the density of X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < +\infty$$



where:

$e =$ Euler's value $\approx 2.718,$ $\qquad \pi =$ Pi value ≈ 3.14159

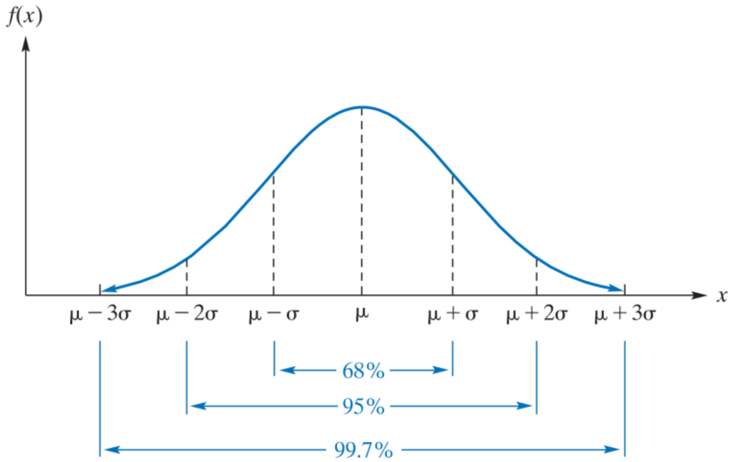
Remark: If X is a normal random variable with parameters μ and σ then its mean and variance are:

$$E(X) = \mu, \qquad var(X) = \sigma^2$$

Properties of the Normal Distribution

In a normal distribution:

- Approximately 68% of all possible values lie within one standard deviation (σ) from the mean (μ).
- Roughly 95% of all possible values fall within two standard deviations (2σ) from the mean.
- Almost 99.7% of all possible values are encompassed within three standard deviations (3σ) from the mean.



Areas under normal curve

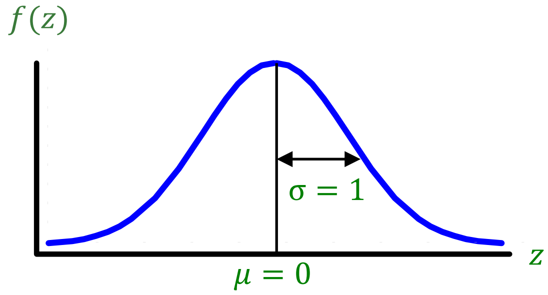
The Standard Normal Distribution

Standard Normal Distribution: Also known as: "Z" distribution

- Mean: $\mu = 0$
- Standard Deviation: $\sigma = 1$

Key observations:

- Values above the mean → Positive Z-values
- Values below the mean have → Negative Z-values



Standardizing Process for a Normal Distribution

- Any normal distribution (with any mean and standard deviation) can be transformed into the standard normal distribution (Z) (Standardization)
- Standardization is often used an alternative approach for computing normal probabilities.
- Translate from any point x on the original normal distribution to the corresponding point z (or z_x) on the standard normal distribution by subtracting the mean of X and dividing by its standard deviation:

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

What does Z-value tell us?

Example

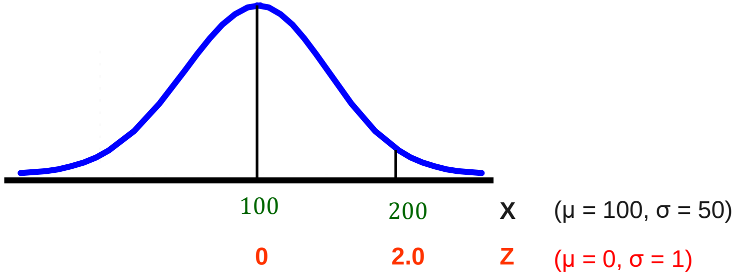
If X is distributed normally with mean of 100 and standard deviation of 50, find the z value for $x = 200$.

Solution

If X is distributed normally with mean of 100 and standard deviation of 50, the z value for $x = 200$ is

$$z_{200} = \frac{x - \mu}{\sigma} = \frac{200 - 100}{50} = 2$$

This says that $= 200$ is two standard deviations (2 increments of 50 units) above the mean of 100.

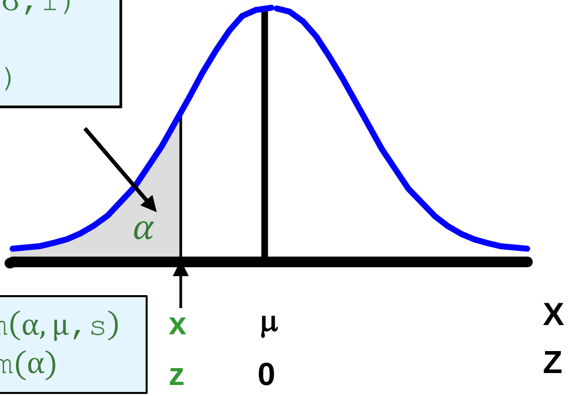


Remark: Converting to z does not change the probabilities.

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = P(Z \leq z_x)$$

Normal Probabilities in R

$\alpha = P(X \leq x) =$ `pnorm(x,μ,σ,1)`
Or:
$\alpha = P(Z \leq z) =$ `pnorm(z)`



$x = \mathrm{qnorm}(\alpha, \mu, s)$
$z_x = \mathrm{qnorm}(\alpha)$

Example: $P(Z \leq 0.12) = \mathrm{pnorm}(0.12) = 0.5478$

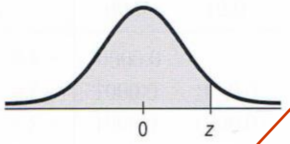
Normal Distribution in Table

z-table: The value in the table indicates the probability $P(Z \leq z)$ for a specified value of z, like $z = 0.12$.

The column gives the second decimal value of z

Table Z (cont.)

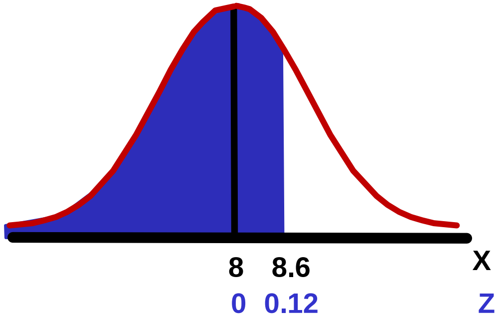
Areas under the standard Normal curve



	Second decimal place in z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879

The row gives the value of z up to the first decimal point

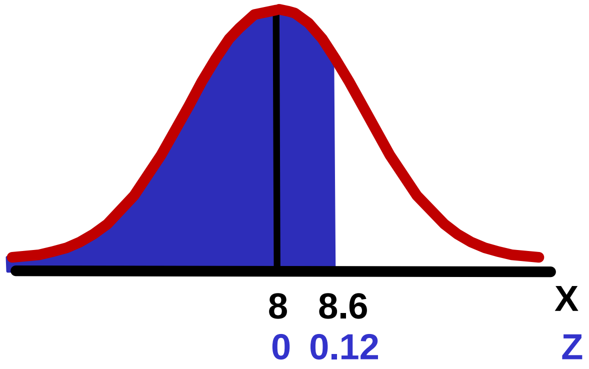
Example: $P(Z \leq 0.12) = 0.5478$



Example

Let X be the time to download a movie from the internet. Suppose X is normal with mean 8 minutes and standard deviation 5 minutes. Find $P(X < 8.6)$.

Solution



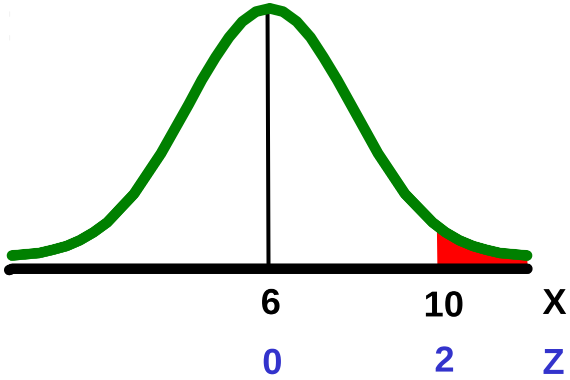
$$z_{8.6} = \frac{x-\mu}{\sigma} = \frac{8.6-8}{5} = 0.12$$

$$P(X < 8.6) = P(Z < 0.12) = pnorm(0.12) = 0.5478$$

Example

At a bank, not all customers experience the same waiting time in line. Assuming the waiting times are normally distributed with an average of 6 minutes and a standard deviation of 2 minutes, what is the probability that a customer will wait more than 10 minutes?

Solution



$$z_{10} = \frac{10 - 6}{2} = 2$$

$$P(X > 10) = P(Z > 2) = 1 - pnorm(2) = 0.023$$

Example

Last year's customer satisfaction scores for a product followed a normal distribution with an average score of 72 and a standard deviation of 12. Out of 500 customers surveyed, how many, on average, gave satisfaction scores higher than 78?

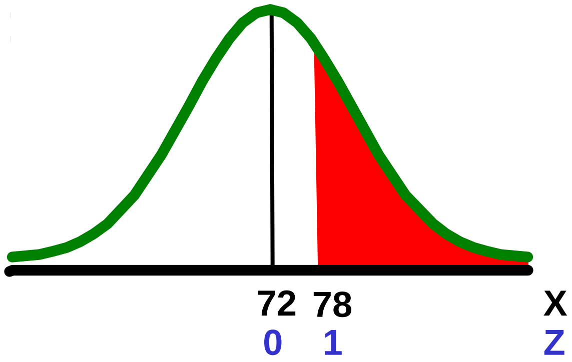
Solution

$$z_{78} = \frac{78 - 72}{12} = 0.5$$

$$P(X > 78) = P(Z > 0.5) = 1 - pnorm(0.5) = 0.31$$

Number of customers: $Y \sim Binomial(500, 0.31)$

$$E(Y) = 500 \times 0.31 = 155$$



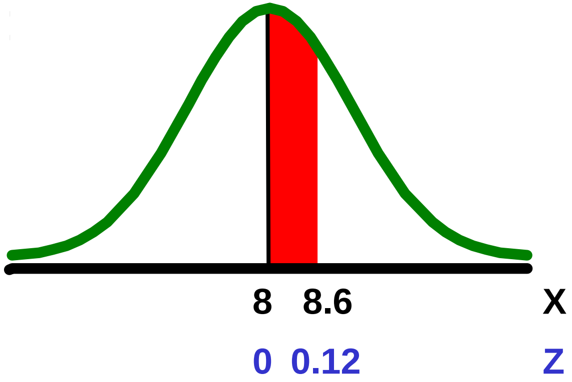
Example:

Let X be the time to download a movie from the internet. Suppose X is normal with mean 8 minutes and standard deviation 5 minutes. Find $P(8 < X < 8.6)$.

Example:

$$z_{8.6} = \frac{8.6 - 8}{5} = 0.12$$

$$z_8 = \frac{8 - 8}{5} = 0$$



$$P(8 < X < 8.6) = P(0 < Z < 0.12) = P(Z \leq 0.12) - P(Z < 0)$$

$$= pnorm(0.12) - pnorm(0) = 0.047$$

Given a Normal Probability Find the x Value

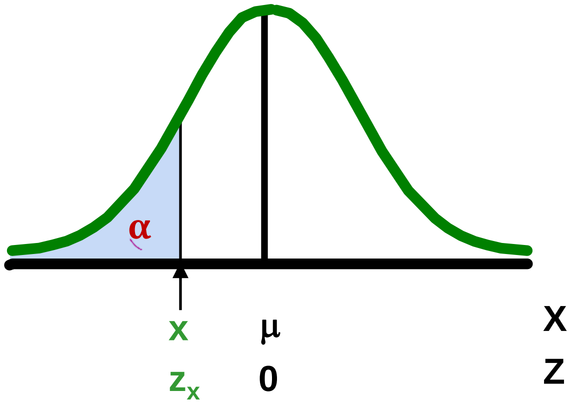
Steps to find the x value for a known probability α :

1. Find z_x for the known cumulative probability α:

$$z_x = \mathrm{qnorm}(\alpha)$$

2. Convert to x by using the formula:

$$x = \mu + z_x\sigma$$



$$z_x = qnorm(\alpha)$$

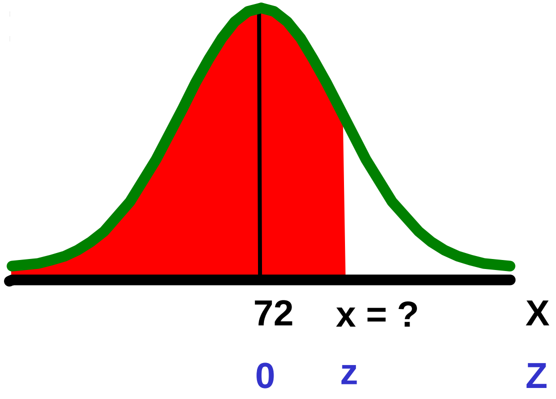
Example

Last year's customer satisfaction scores for a product followed a normal distribution with an average score of 72 and a standard deviation of 12. What is the 90^{th} percentile of our customer satisfaction scores?

Solution

$$z_x = qnorm(0.9) \ = 1.28$$

$$x = \mu + z_x \sigma = 72 + 1.28 \times 12 = 87.36$$

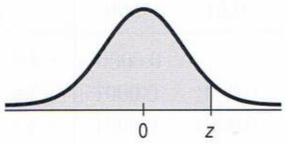


Solution by Using Table

Second decimal: 0.08

Table Z (cont.)

Areas under the standard Normal curve



First decimal: 1.20

	Second decimal place in z									
z	*0.00*	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*	*0.06*	*0.07*	*0.08*	*0.09*
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000†									

† For $z \geq 3.90$, the areas are 1.0000 to four decimal places.