UCLA College | Physical Sciences
Statistics

# STATS 10: Introduction to Statistical Reasoning
# Chapter 6

Bingling Wang
Department of Statistics

# Probability Distribution

# Random Variable

**Map outcomes of random processes to numbers**

When the value of a variable is determined by a chance event,
that variable is called a random variable.

- **Discrete random variable**: numerical values that can be listed or counted
  - The number of phone calls you will receive during the day
  - The number of burgers sold by In-N-Out daily
  - The number of bus lines that go to UCLA

- **Continuous random variable**: take any values in an interval and are often measurements.
  - The exact length of time your next phone call will last
  - The exact weight of a burger you ordered from In-N-Out
  - The exact time it takes to commute to school by bus

# Probability Distribution

Probability distribution describes a random variable (r.v.).

It tells us:

- **The possible outcomes of a random variable**
- **The probability that each outcome will occur**

Depending on the variable type, we take different approaches to display or analyze the probability distribution of the variable.

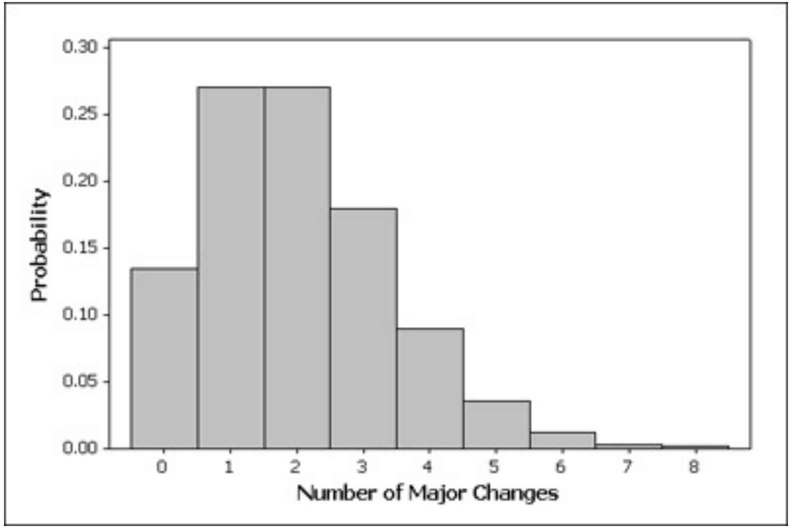
**Tables** **Graphs** **Equations**

- All $P(x)$ values must be between 0 and 1, i.e., $0 \leq P(x) \leq 1$
- The sum of all $P(x)$ values must equal 1

# Probability Distribution

Consider the random variable “The number of times a student changes major at a college.”

| X=# changes in major | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Probability | 0.135 | 0.271 | 0.271 | 0.180 | 0.090 | 0.036 | 0.012 | 0.003 | 0.002 |



# Exercise

Suppose you roll a fair 6-sided die:

- You win 5 points if you roll a 5 or a 6.
- You lose 1 point if you roll a 1.
- For any other outcome you win or lose nothing.

Create a table showing the probability distribution for the points you will win by playing this game.

Let X be the random variable that represents the number of points you win.

| X | P(x) |
|---|---|
| -1 | |
| 0 | |
| 5 | |

# Discrete Probability Distribution

## Probability Distributions Can Also Be Equations

**Example**: suppose we roll a fair die until we see a 6.

X = number of trials

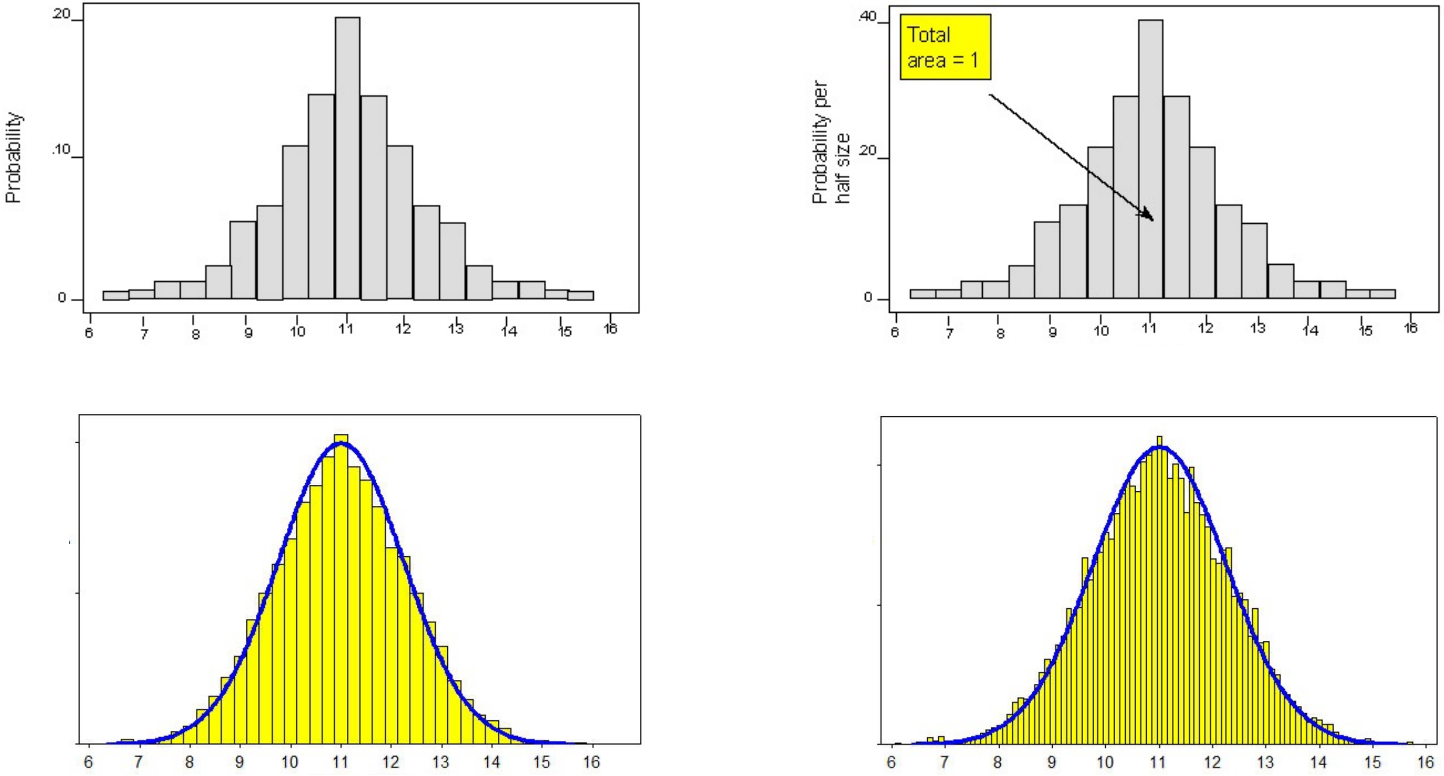
| X | 1 | 2 | 3 | …… |
|---|---|---|---|---|
| Probability | $\frac{1}{6}$ | $\frac{5}{6}\times\frac{1}{6}$ | $\frac{5}{6}\times\frac{5}{6}\times\frac{1}{6}$ | |

$$P(x) = \left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right)$$

$$P(x = 10) = \left(\frac{5}{6}\right)^{10-1}\left(\frac{1}{6}\right) = 0.0323$$

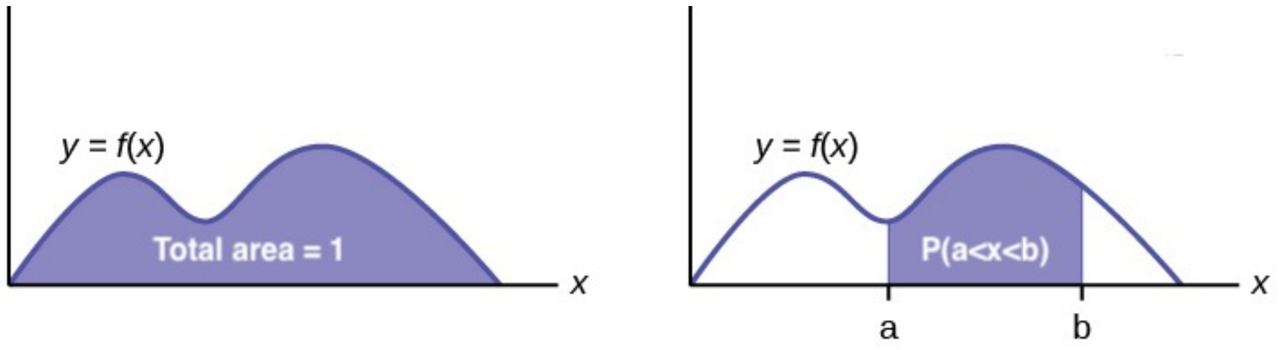
$$\mathrm{P}(\mathrm{x} \leq 3) = \frac{1}{6} + \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^{2}\left(\frac{1}{6}\right) = 0.4213$$

# Transition to Continuous Random Variables



# Continuous Probability Distribution

- The continuous probability distribution is usually visualized by a curve
  -- **Probability density curve**
  - The density curve cannot lie below the x-axis
- The probability for a continuous random variable is represented as
  -- **Area under the curve**
  - The total area under the curve must equal 1
  - The probability of any single value occurring is zero

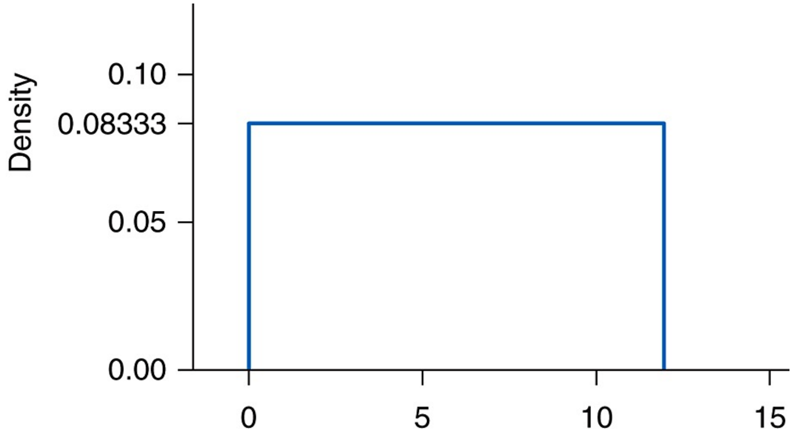


# Finding Probabilities

**Example:**

A bus arrives at a certain stop every 12 minutes. The graph shows the probability distribution for wait times before the bus arrives.

Use the distribution to find the probability of a wait time between 0 and 10 minutes.



# The Normal Model

# The Normal Distribution

**The most widely used probability model for continuous numerical variables.**

Also called Gaussian distribution / Laplace-Gauss distribution

Defined by the **mean** and the **standard deviation**

**Notation:** $N(\mu, \sigma)$

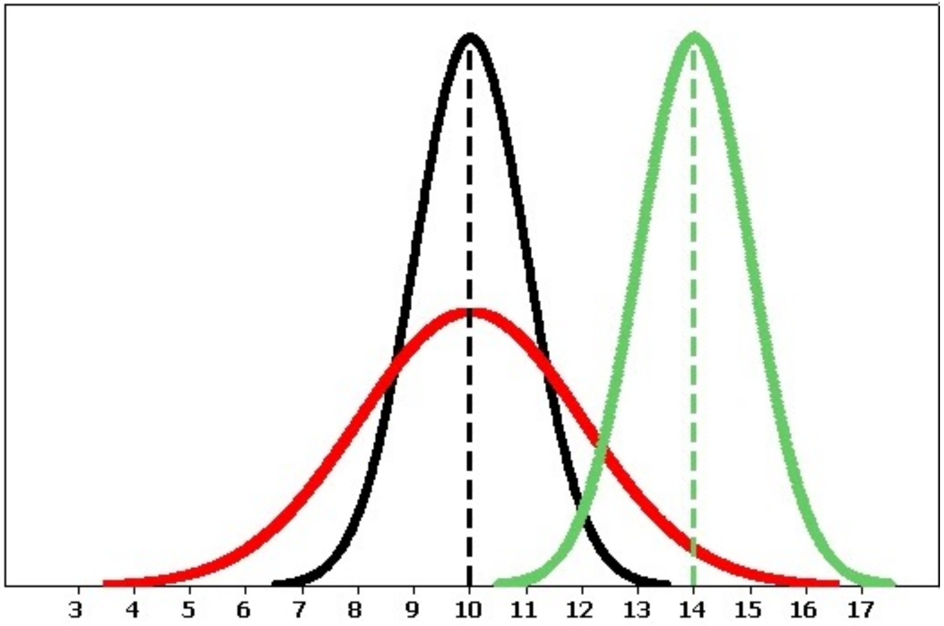
-- a Normal distribution with mean $\mu$, and standard deviation $\sigma$

**Characteristics of the distribution**:

- Unimodal
- Symmetric
- Bell shaped

# Different Mean and Standard Deviation

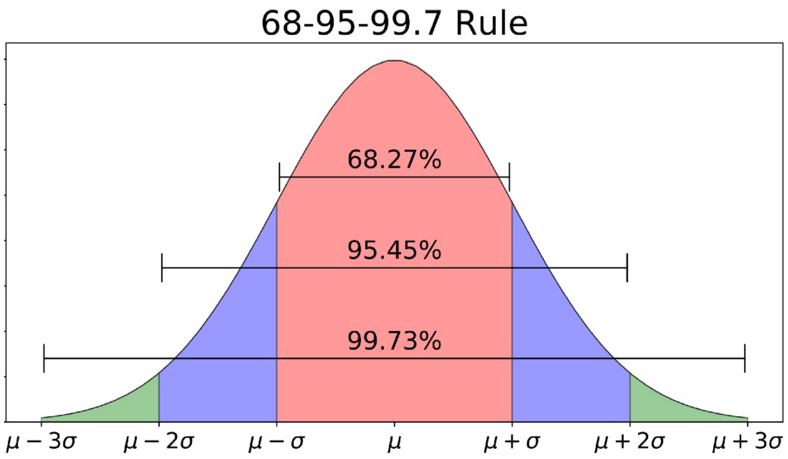
The shape of the normal distribution is completely determined by the values of the mean and standard deviation.



# The Empirical Rule

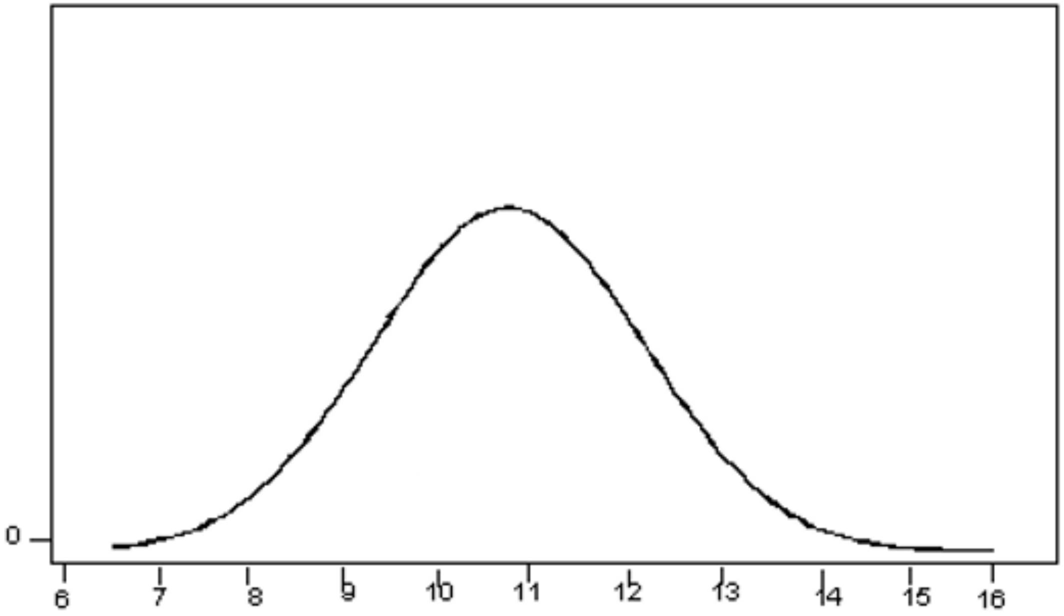
If X is a normal random variable with N($\mu$,$\sigma$), then the probability is

- $P(\mu - \sigma < X < \mu + \sigma) \approx 0.68$
- $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95$
- $P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997$



# Example: Foot Length

Suppose that foot length of a randomly chosen adult male is a normal random variable with mean μ=11 and standard deviation σ=1.5.



# Finding Normal Probabilities

The first and helpful step is usually to:
sketch the curve, label it appropriately, and shade the area of interest.

## 1. Using the exact probability density function (pdf)

pdf for normal distribution:

$$P(a < x < b) = \int_a^b f(x)\,dx \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

## 2. With computer technology

In R: function **pnorm(q, mean, sd)** computes probabilities $P(x \leq q)$ from a normal distribution with specified mean and sd.

## 3. With standard normal table -- Z-table

**Standard Normal:** the normal model with mean 0 and standard deviation 1, $N(0, 1)$.

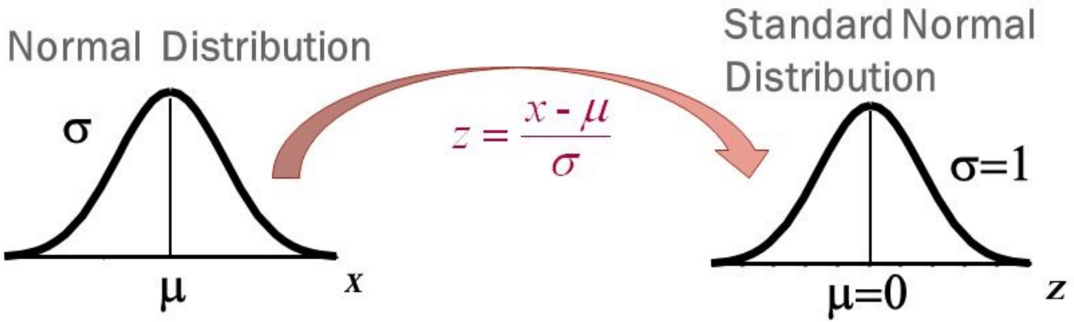
# Finding Normal Probabilities with Z-table

This approach requires us to convert the normal model to a standard normal model by converting the values to standard units — **standardization / z-transformation.**

**Steps**:

1. Convert the values to standard units, $z = \frac{x-\mu}{\sigma}$.
2. Find the probability using the Z-table.

Standardizing data into z-scores does not change the shape of the distribution.



# Z Table

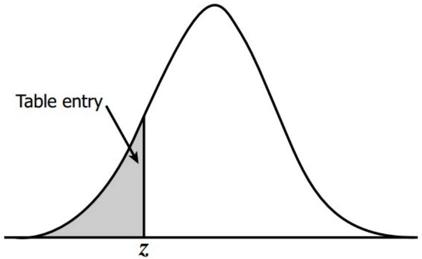


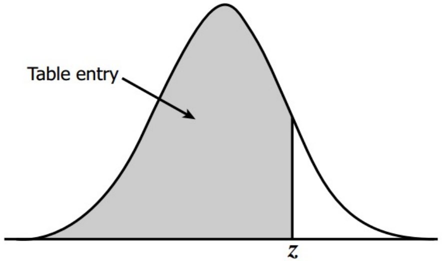

Table entries for z represent the area under the curve to the left of z.

$$\boldsymbol{Z \sim N(0,1)}$$

Example: $P(z < -2.72)$

| Z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| -3.9 | .00005 | .00005 | .00004 | .00004 | .00004 | .00004 | .00004 | .00004 | .00003 | .00003 |
| -3.8 | .00007 | .00007 | .00007 | .00006 | .00006 | .00006 | .00006 | .00005 | .00005 | .00005 |
| -3.7 | .00011 | .00010 | .00010 | .00010 | .00009 | .00009 | .00008 | .00008 | .00008 | .00008 |
| -3.6 | .00016 | .00015 | .00015 | .00014 | .00014 | .00013 | .00013 | .00012 | .00012 | .00011 |
| -3.5 | .00023 | .00022 | .00022 | .00021 | .00020 | .00019 | .00019 | .00018 | .00017 | .00017 |
| -3.4 | .00034 | .00032 | .00031 | .00030 | .00029 | .00028 | .00027 | .00026 | .00025 | .00024 |
| -3.3 | .00048 | .00047 | .00045 | .00043 | .00042 | .00040 | .00039 | .00038 | .00036 | .00035 |
| -3.2 | .00069 | .00066 | .00064 | .00062 | .00060 | .00058 | .00056 | .00054 | .00052 | .00050 |
| -3.1 | .00097 | .00094 | .00090 | .00087 | .00084 | .00082 | .00079 | .00076 | .00074 | .00071 |
| -3.0 | .00135 | .00131 | .00126 | .00122 | .00118 | .00114 | .00111 | .00107 | .00104 | .00100 |
| -2.9 | .00187 | .00181 | .00175 | .00169 | .00164 | .00159 | .00154 | .00149 | .00144 | .00139 |
| -2.8 | .00256 | .00248 | .00240 | .00233 | .00226 | .00219 | .00212 | .00205 | .00199 | .00193 |
| -2.7 | .00347 | .00336 | .00326 | .00317 | .00307 | .00298 | .00289 | .00280 | .00272 | .00264 |
| -2.6 | .00466 | .00453 | .00440 | .00427 | .00415 | .00402 | .00391 | .00379 | .00368 | .00357 |
| -2.5 | .00621 | .00604 | .00587 | .00570 | .00554 | .00539 | .00523 | .00508 | .00494 | .00480 |
| -2.4 | .00820 | .00798 | .00776 | .00755 | .00734 | .00714 | .00695 | .00676 | .00657 | .00639 |
| -2.3 | .01072 | .01044 | .01017 | .00990 | .00964 | .00939 | .00914 | .00889 | .00866 | .00842 |
| -2.2 | .01390 | .01355 | .01321 | .01287 | .01255 | .01222 | .01191 | .01160 | .01130 | .01101 |
| -2.1 | .01786 | .01743 | .01700 | .01659 | .01618 | .01578 | .01539 | .01500 | .01463 | .01426 |
| -2.0 | .02275 | .02222 | .02169 | .02118 | .02068 | .02018 | .01970 | .01923 | .01876 | .01831 |
| -1.9 | .02872 | .02807 | .02743 | .02680 | .02619 | .02559 | .02500 | .02442 | .02385 | .02330 |
| -1.8 | .03593 | .03515 | .03438 | .03362 | .03288 | .03216 | .03144 | .03074 | .03005 | .02938 |
| -1.7 | .04457 | .04363 | .04272 | .04182 | .04093 | .04006 | .03920 | .03836 | .03754 | .03673 |
| -1.6 | .05480 | .05370 | .05262 | .05155 | .05050 | .04947 | .04846 | .04746 | .04648 | .04551 |
| -1.5 | .06681 | .06552 | .06426 | .06301 | .06178 | .06057 | .05938 | .05821 | .05705 | .05592 |

# Z Table Continued




Let X be a normal random variable.

What is the probability of X taking a value less than 0.75 standard deviations above the mean (i.e. having a z-score of 0.75)?

| Z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .50000 | .50399 | .50798 | .51197 | .51595 | .51994 | .52392 | .52790 | .53188 | .53586 |
| 0.1 | .53983 | .54380 | .54776 | .55172 | .55567 | .55962 | .56356 | .56749 | .57142 | .57535 |
| 0.2 | .57926 | .58317 | .58706 | .59095 | .59483 | .59871 | .60257 | .60642 | .61026 | .61409 |
| 0.3 | .61791 | .62172 | .62552 | .62930 | .63307 | .63683 | .64058 | .64431 | .64803 | .65173 |
| 0.4 | .65542 | .65910 | .66276 | .66640 | .67003 | .67364 | .67724 | .68082 | .68439 | .68793 |
| 0.5 | .69146 | .69497 | .69847 | .70194 | .70540 | .70884 | .71226 | .71566 | .71904 | .72240 |
| 0.6 | .72575 | .72907 | .73237 | .73565 | .73891 | .74215 | .74537 | .74857 | .75175 | .75490 |
| 0.7 | .75804 | .76115 | .76424 | .76730 | .77035 | .77337 | .77637 | .77935 | .78230 | .78524 |
| 0.8 | .78814 | .79103 | .79389 | .79673 | .79955 | .80234 | .80511 | .80785 | .81057 | .81327 |
| 0.9 | .81594 | .81859 | .82121 | .82381 | .82639 | .82894 | .83147 | .83398 | .83646 | .83891 |
| 1.0 | .84134 | .84375 | .84614 | .84849 | .85083 | .85314 | .85543 | .85769 | .85993 | .86214 |
| 1.1 | .86433 | .86650 | .86864 | .87076 | .87286 | .87493 | .87698 | .87900 | .88100 | .88298 |
| 1.2 | .88493 | .88686 | .88877 | .89065 | .89251 | .89435 | .89617 | .89796 | .89973 | .90147 |
| 1.3 | .90320 | .90490 | .90658 | .90824 | .90988 | .91149 | .91309 | .91466 | .91621 | .91774 |
| 1.4 | .91924 | .92073 | .92220 | .92364 | .92507 | .92647 | .92785 | .92922 | .93056 | .93189 |
| 1.5 | .93319 | .93448 | .93574 | .93699 | .93822 | .93943 | .94062 | .94179 | .94295 | .94408 |
| 1.6 | .94520 | .94630 | .94738 | .94845 | .94950 | .95053 | .95154 | .95254 | .95352 | .95449 |
| 1.7 | .95543 | .95637 | .95728 | .95818 | .95907 | .95994 | .96080 | .96164 | .96246 | .96327 |
| 1.8 | .96407 | .96485 | .96562 | .96638 | .96712 | .96784 | .96856 | .96926 | .96995 | .97062 |
| 1.9 | .97128 | .97193 | .97257 | .97320 | .97381 | .97441 | .97500 | .97558 | .97615 | .97670 |
| 2.0 | .97725 | .97778 | .97831 | .97882 | .97932 | .97982 | .98030 | .98077 | .98124 | .98169 |
| 2.1 | .98214 | .98257 | .98300 | .98341 | .98382 | .98422 | .98461 | .98500 | .98537 | .98574 |
| 2.2 | .98610 | .98645 | .98679 | .98713 | .98745 | .98778 | .98809 | .98840 | .98870 | .98899 |
| 2.3 | .98928 | .98956 | .98983 | .99010 | .99036 | .99061 | .99086 | .99111 | .99134 | .99158 |
| 2.4 | .99180 | .99202 | .99224 | .99245 | .99266 | .99286 | .99305 | .99324 | .99343 | .99361 |

# Z Table Exercises

**Suppose X is a normal random variable, $X \sim N(\mu, \sigma)$.**

1. What is the probability of X taking a value more than 1.57 standard deviations below its mean?

2. What is the probability of X taking a value within 1.57 standard deviations from its mean?

3. What is the probability of X taking a value between 1 standard deviation below and 1.57 standard deviation above its mean?

| $z$ | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| −1.5 | .0668 | .0655 | .0643 | .0630 | .0618 | .0606 | .0594 | .0582 | .0571 | .0559 |
| −1.0 | .1587 | .1562 | .1539 | .1515 | .1492 | .1469 | .1446 | .1423 | .1401 | .1379 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |

# Exercise

Suppose the distribution of teenagers' weights follows a normal distribution $N(125, 10)$.

What is the probability that a randomly selected teenager weighs between 135.5 and 140.6 pounds?

1. Convert the values to standard units;

2. Use the Z-table to find the area under the curve between the two values.

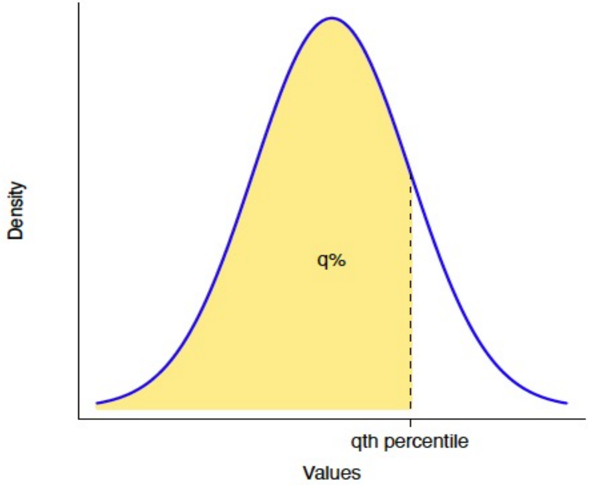
| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |

# Finding Percentiles

**Given a probability, what is the associated value of the normal random variable?**

**Percentile**

The $q$**th percentile** of a probability distribution is the value where $q\%$ of the distribution lies to the left and $(100 - q)\%$ lies to the right.



# Finding Percentiles

To find the $q$th percentile of a $N(\mu, \sigma)$ distribution:

**1. With computer technology**

R function: $\mathrm{qnorm(p, mean, sd)}$ computes value for that probability value $p$ from a normal distribution with specified mean and sd.

$\mathrm{qnorm}(0.5, 0, 1)$ $\mathrm{qnorm}(0.5, 1, 2)$ $\mathrm{qnorm}(0.84, 1, 2)$

**2. With z table**

I. Use the z-table to locate the given probability.

II. Find the corresponding z-score.

III. Convert the z-score to original units.

$$q\text{th percentile} = \mu + \sigma z$$

# Example

Suppose the heights of kindergarten children follow a $N(39, 2)$ normal distribution, how tall (in inches) does a child need to be in order to be in the 90$^{th}$ percentile?

1. Locate the given probability in the table

| $z$ | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|
| 0.8 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |

2. Find the z-score using standard normal table

3. Convert the z-score into the original scale