Normal Distribution

Normal Distribution

Normal Distribution

A random variable having the normal distribution with mean μ and variance σ^2,
denoted by $X \sim N(\mu, \sigma^2)$,
has the probability density function

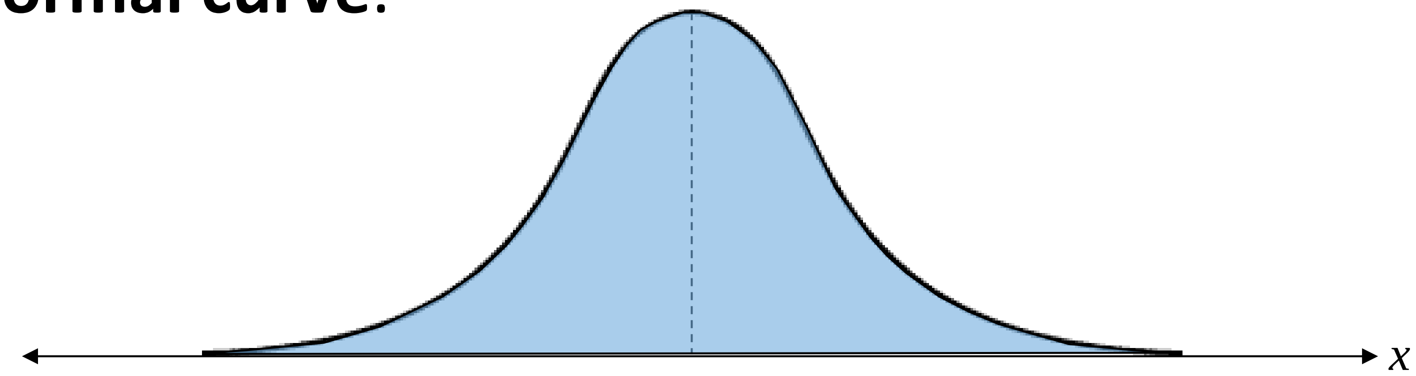
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; -\infty < x < \infty; \; -\infty < \mu < \infty, \; \sigma > 0.$$

The mean μ is called a *location parameter* while the variance σ^2 is called a *scale parameter*.

Properties of Normal Distributions

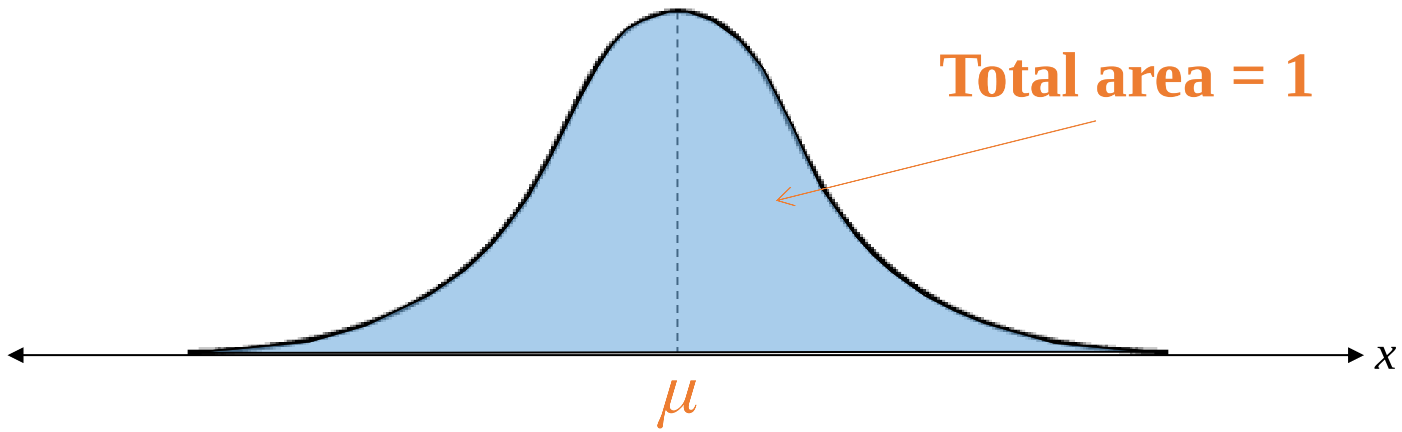
Normal distribution

- A continuous probability distribution for a random variable, *x*.
- The most important continuous probability distribution in statistics.
- The graph of a normal distribution is called the **normal curve**.



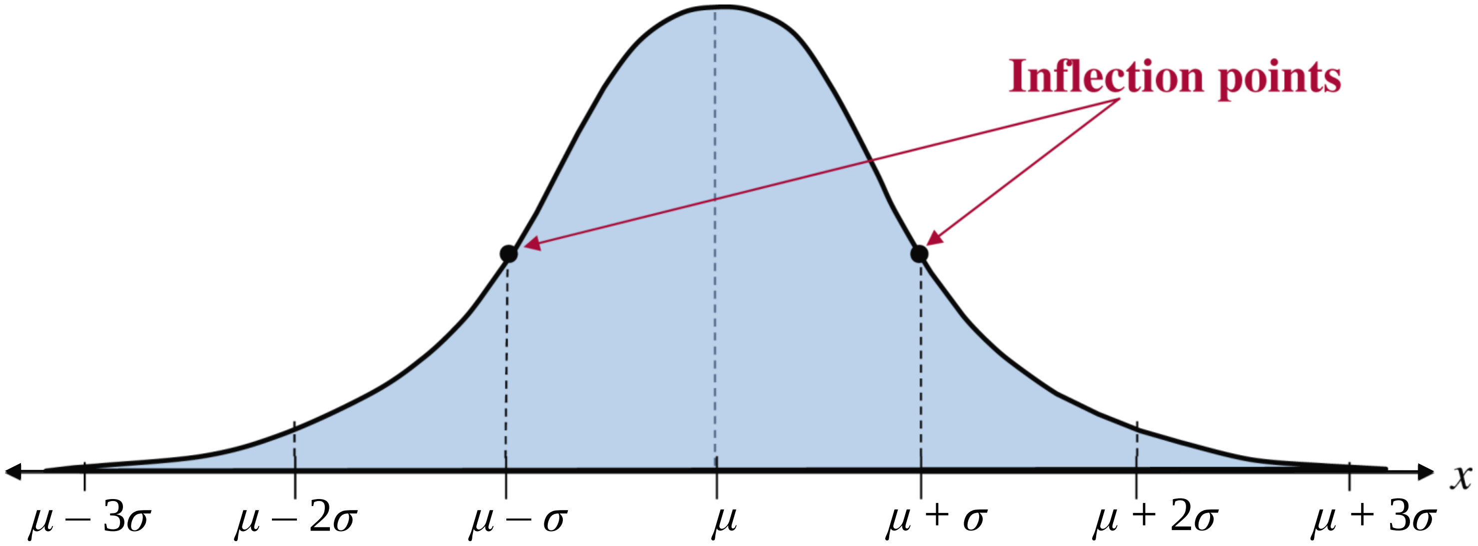
Properties of Normal Distributions

1. The mean, median, and mode are equal.
2. The normal curve is bell-shaped and is symmetric about the mean.
3. The total area under the normal curve is equal to 1.
4. The normal curve approaches, but never touches, the *x*-axis as it extends farther and farther away from the mean.



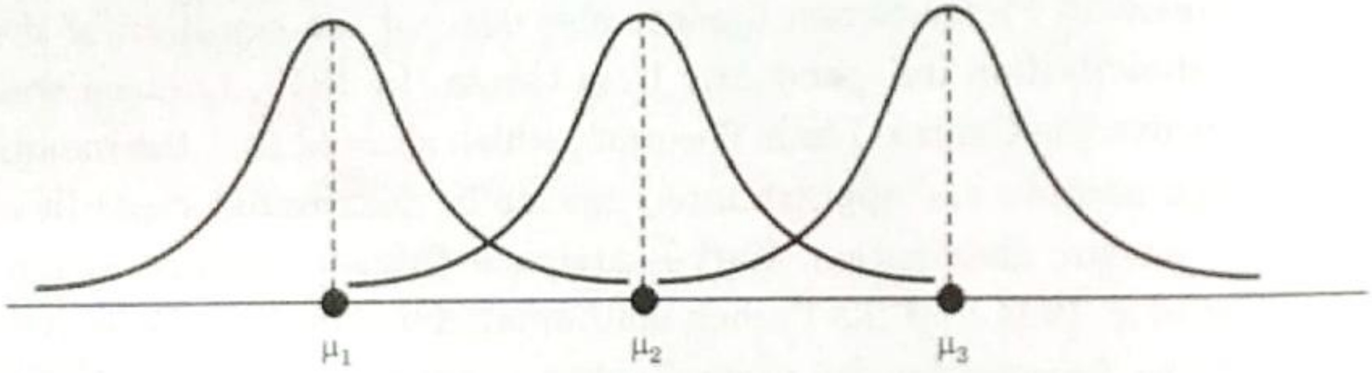
Properties of Normal Distributions

5. Between $\mu - \sigma$ and $\mu + \sigma$ (in the center of the curve), the graph curves downward. The graph curves upward to the left of $\mu - \sigma$ and to the right of $\mu + \sigma$. The points at which the curve changes from curving upward to curving downward are called the **inflection points**.



Mean as location parameter

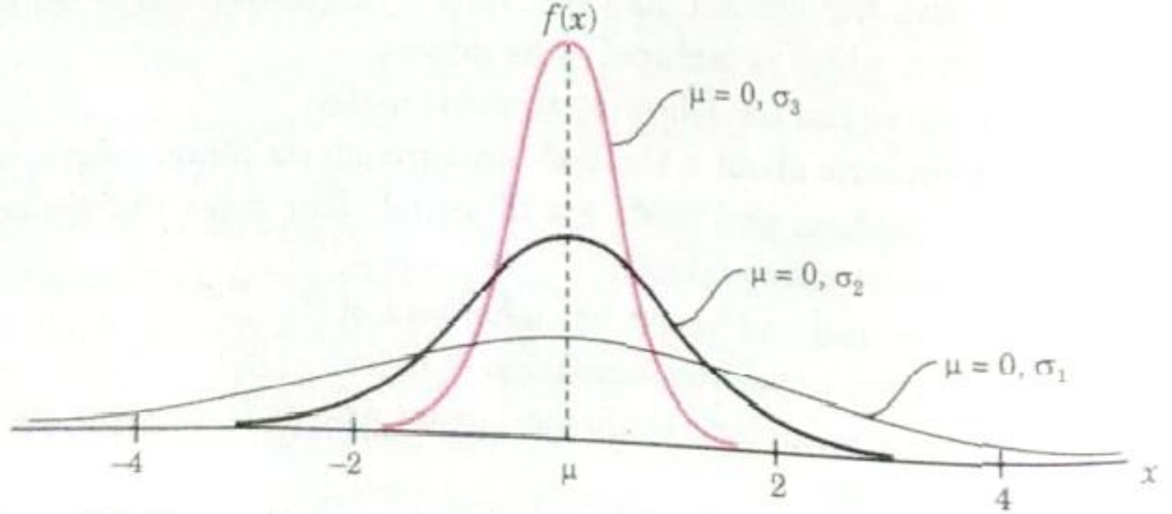
The mean μ, called a *location parameter*, gives the location of the normal distribution.



$\mu_1 < \mu_2 < \mu_3$

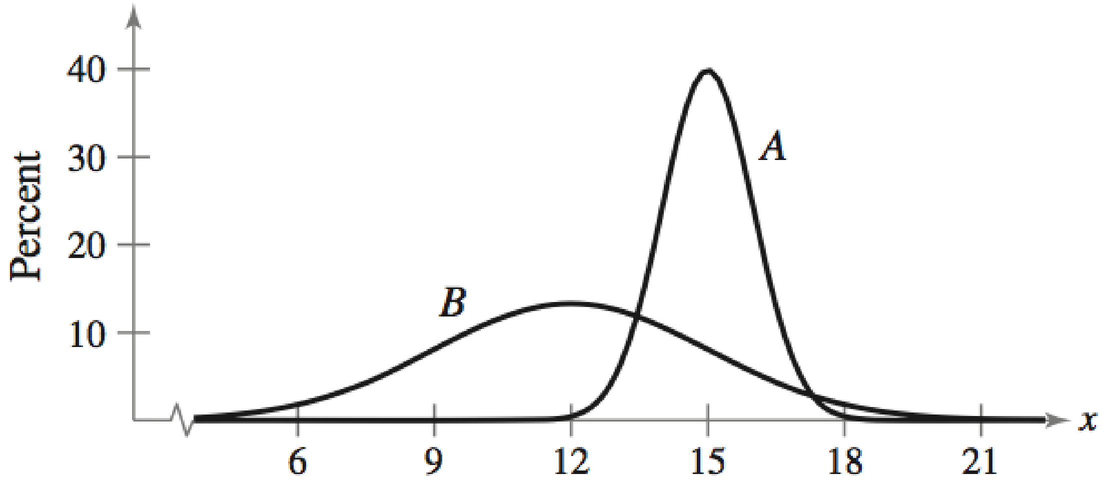
Variance as scale parameter

The variance σ^2, called a *scale parameter*, gives the dispersion of the normal distribution.



Example: Understanding Mean and Standard Deviation

1. Which normal curve has the greater mean?

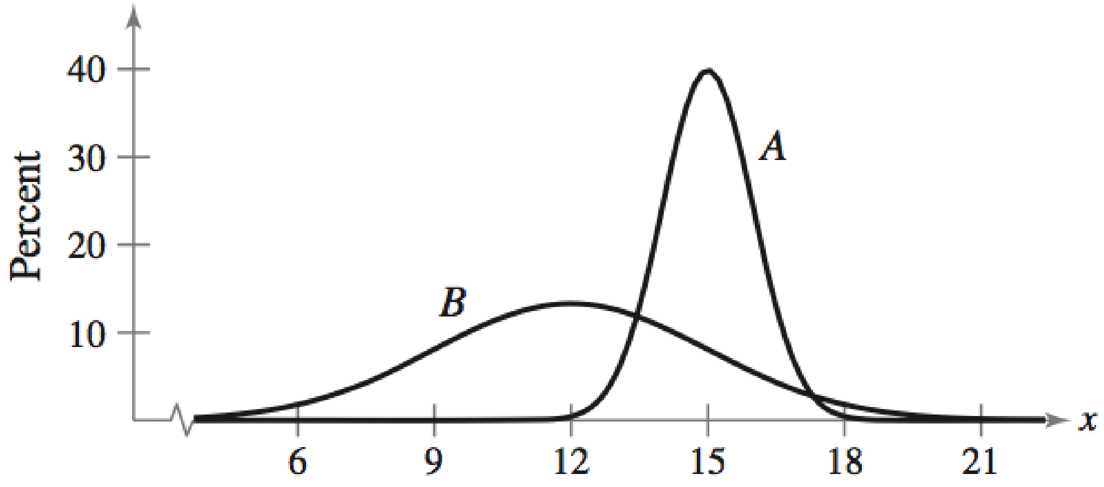


Solution:

Curve *A* has the greater mean (The line of symmetry of curve *A* occurs at $x = 15$. The line of symmetry of curve *B* occurs at $x = 12$.)

Example: Understanding Mean and Standard Deviation

2. Which curve has the greater standard deviation?



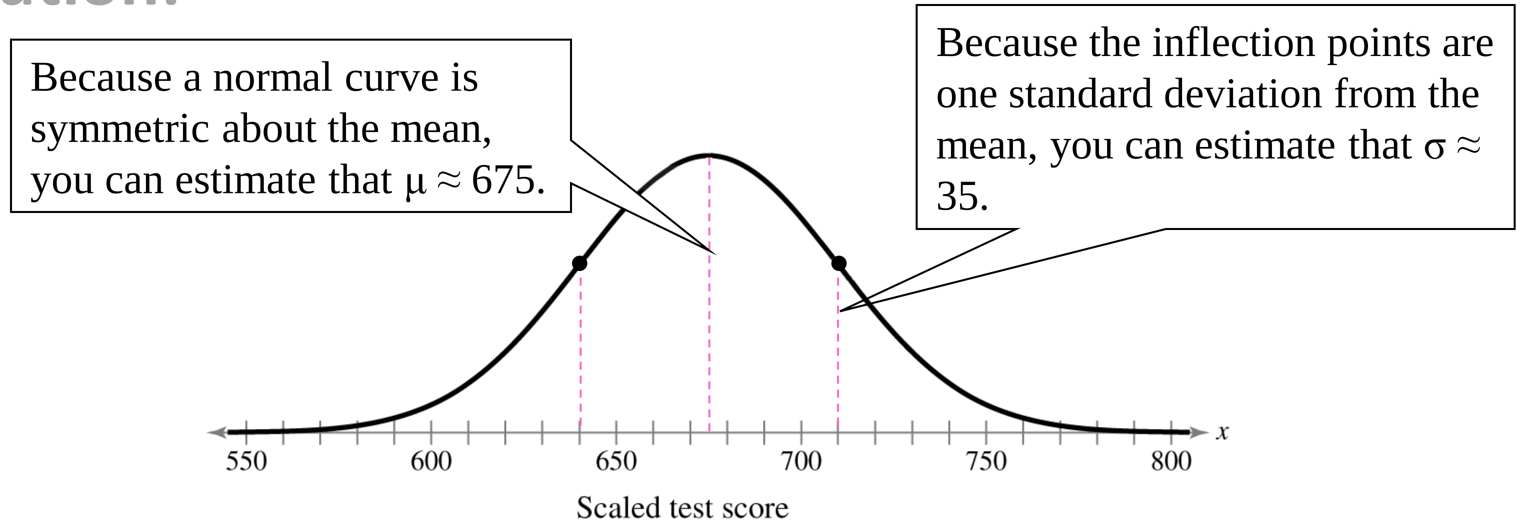
Solution:

Curve *B* has the greater standard deviation (Curve *B* is more spread out than curve *A*.)

Example: Interpreting Graphs

The scaled test scores for the New York State Grade 8 Mathematics Test are normally distributed. The normal curve shown below represents this distribution. What is the mean test score? Estimate the standard deviation.

Solution:



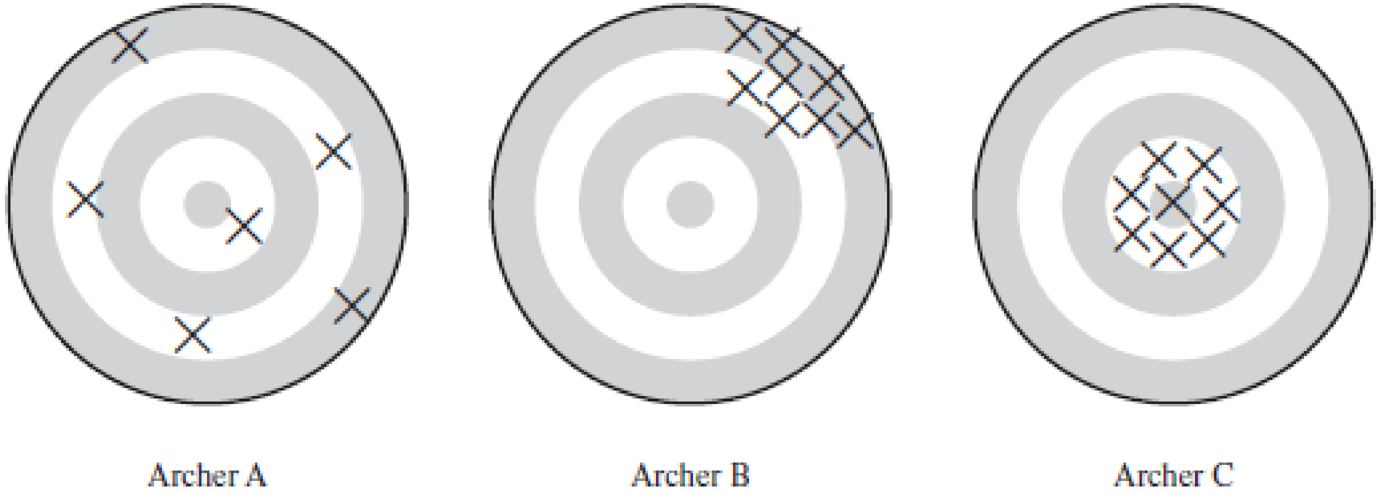
Standard Normal Distribution

A special case of the normal distribution is the standard normal distribution with a mean of 0 and a variance of 1 . The random variable that has this distribution is denoted by $Z \sim N(0,1)$.

“Z follows a normal distribution with mean 0 and variance 1.”

FIGURE 2.3

Unbiased, precise, and accurate archers. Archer A is unbiased—the average position of all arrows is at the bull's-eye. Archer B is precise but not unbiased—all arrows are close together but systematically away from the bull's-eye. Archer C is accurate—all arrows are close together and near the center of the target.



Thus, an estimator $\hat{t}$ of t is **unbiased** if $E(\hat{t}) = t$, **precise** if $V(\hat{t}) = E[(\hat{t} - E[\hat{t}])^2]$ is small, and **accurate** if $\text{MSE}[\hat{t}] = E[(\hat{t} - t)^2]$ is small. A badly biased estimator may be precise but it will not be accurate; accuracy (MSE) is how close the estimate is to the true value, while precision (variance) measures how close estimates from different samples are to each other. Figure 2.3 illustrates these concepts.

Lohr, S. L. (2021). *Sampling: design and analysis*. Chapman and Hall/CRC.

Standard Normal Distribution

A special case of the normal distribution is the standard normal distribution.

The probability density function of a normal random variable Z,
denoted by $Z - N(0,1)$, is

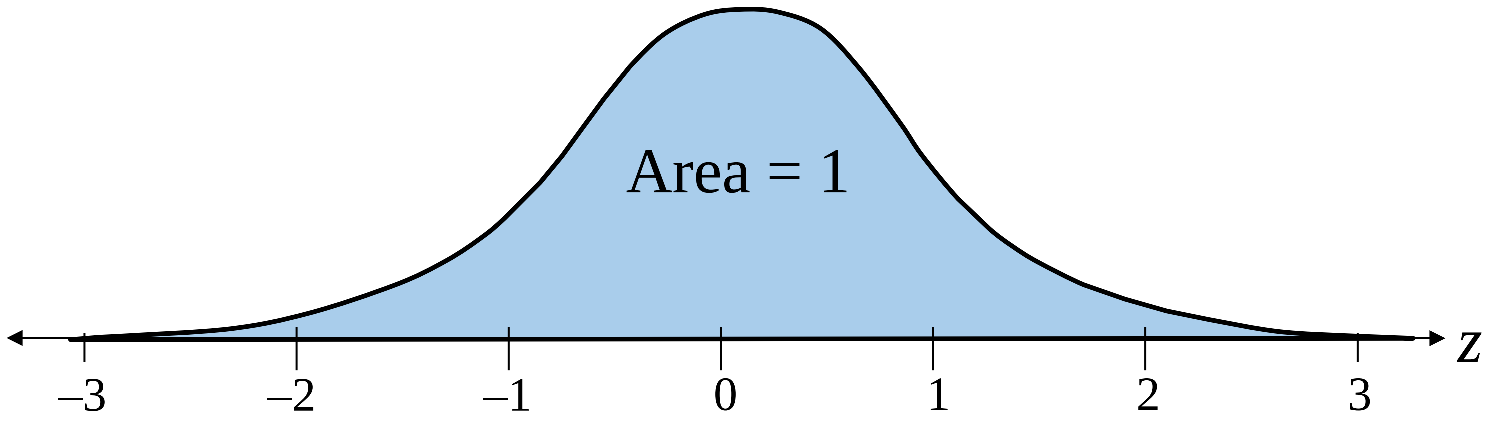
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty$$

with mean $\mu = 0$ and variance $\sigma^2 = 1$

The Standard Normal Distribution

Standard normal distribution

- A normal distribution with a mean of 0 and a standard deviation of 1.

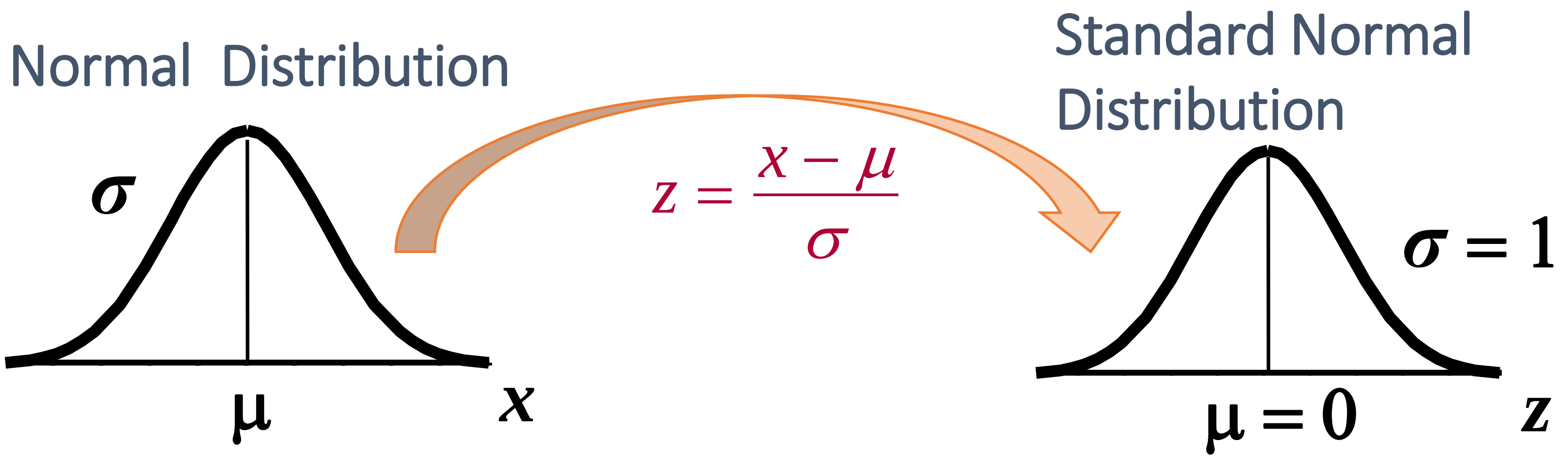


- Any x-value can be transformed into a z-score by using the formula

$$z = \frac{\text{Value} - \text{Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}$$

The Standard Normal Distribution

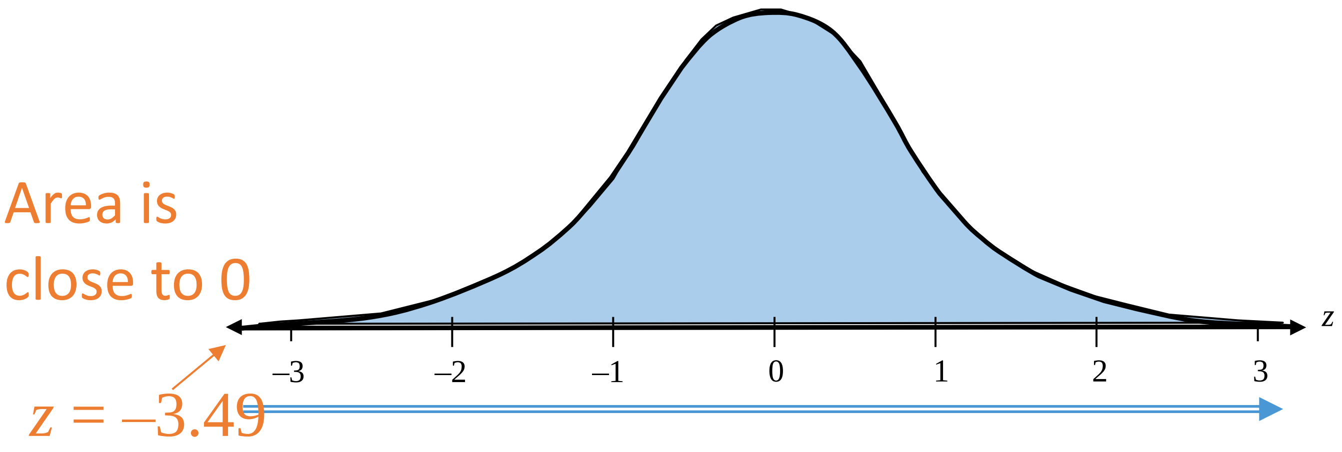
- If each data value of a normally distributed random variable *x* is transformed into a *z*-score, the result will be the standard normal distribution.



- Use the Standard Normal Table to find the cumulative area under the standard normal curve.

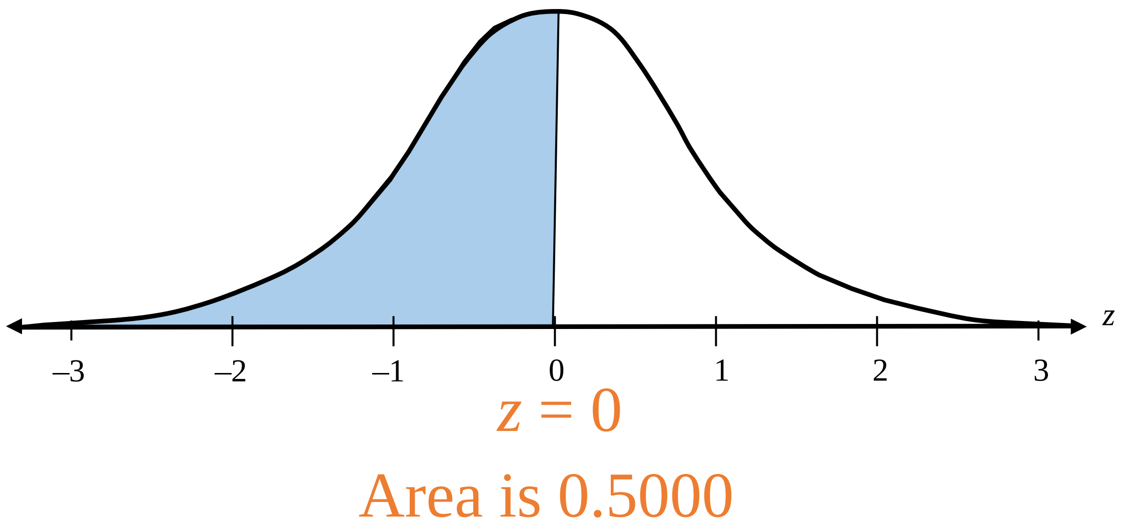
Properties of the Standard Normal Distribution

1. The cumulative area is close to 0 for *z*-scores close to $z = -3.49$.
2. The cumulative area increases as the *z*-scores increase.



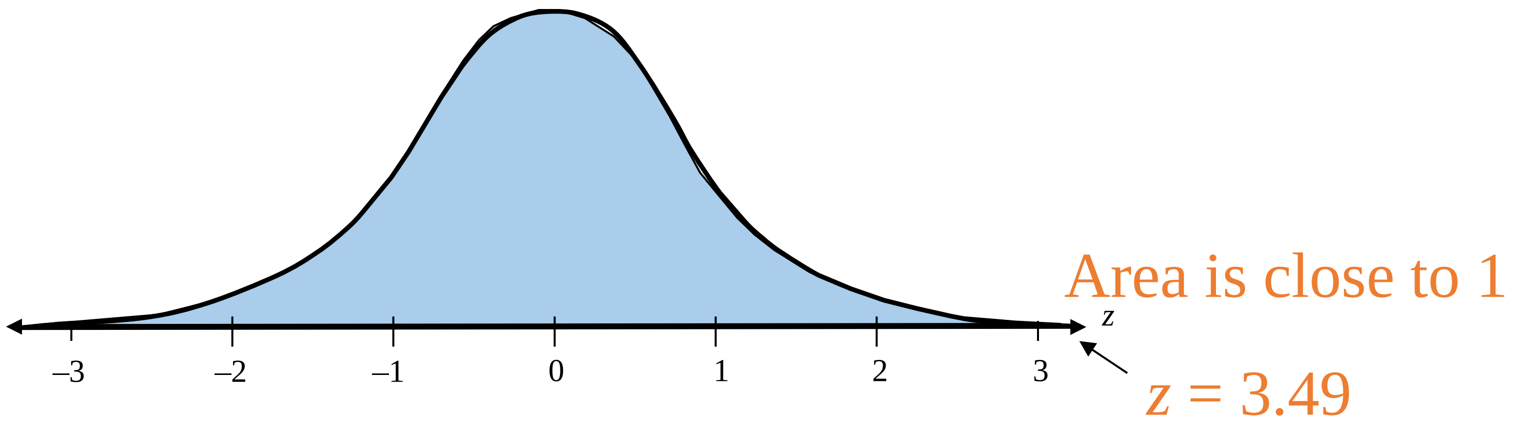
Properties of the Standard Normal Distribution

3. The cumulative area for $z = 0$ is 0.5000.



Properties of the Standard Normal Distribution

4. The cumulative area is close to 1 for z-scores close to $z = 3.49$.

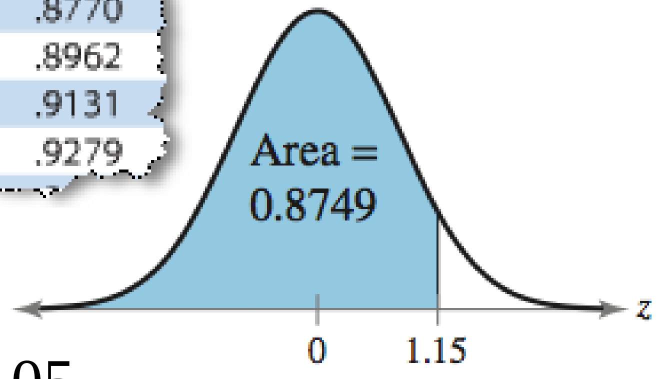


- A z-score table gives the cumulative areas under the standard normal curve.
- It gives the probability value $P(Z < z)$ as an area under the standard normal Curve to the left of $Z = z$.
- The first column of the table are the values of Z correct up to the first decimal
- place and
- the first row of the table are the column headings $0.00°$ through "0.09" corresponding to the
- second decimal place of Z.
- Thus, to find $P(X <*)$ from this table, you need first to compute the corresponding z-score
- of $X = x$ from the formula $z = \frac{x-\mu}{\sigma}$, rounded off to two
- decimal places.
- Then locate the resulting z-score up to its first decimal place in the first column and its second
- decimal place in the column headings of the table.
- The resulting entry in the body of the table is the area under the standard normal curve to the left of $Z =$ z.
- This area is P($Z <$ z), which also gives the desired probability $P(X < x)$.

Example: Using The Standard Normal Table

Find the cumulative area that corresponds to a *z*-score of 1.15.

z	.00	.01	.02	.03	.04	.05	.06
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279



Solution:

Find 1.1 in the left hand column.

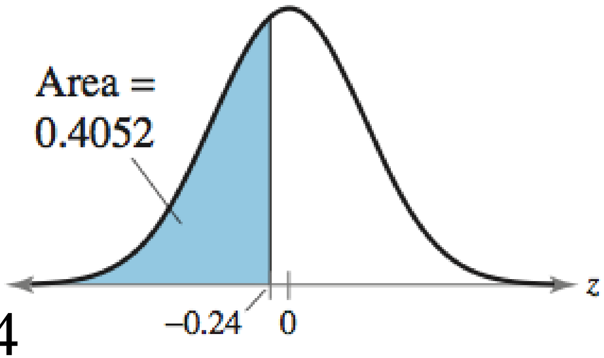
Move across the row to the column under 0.05

The area to the left of *z* = 1.15 is 0.8749.

Example: Using The Standard Normal Table

Find the cumulative area that corresponds to a *z*-score of –0.24.

z	.09	.08	.07	.06	.05	.04	.03
–3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003
–3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004
–3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006
–0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981
–0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336
–0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707
–0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090
–0.1	.4247	.4286	.4325	.4364	.4404	.4443	
–0.0	.4641	.4681	.4721	.4761	.4801	.4840	



Solution:

Find –0.2 in the left hand column.

Move across the row to the column under 0.04

The area to the left of *z* = –0.24 is 0.4052.

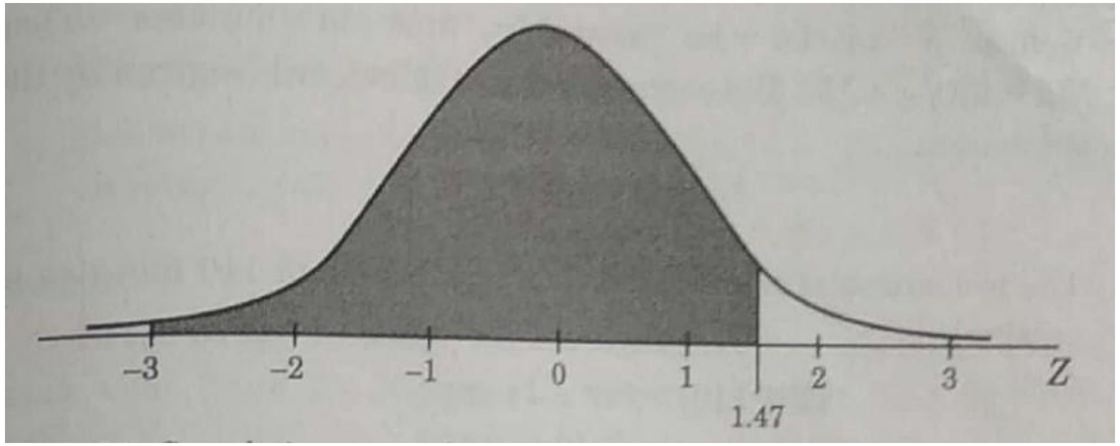
$P(Z < 1.47)$

Example

Find $P(Z < 1.47)$.

The entry in the body of the table at the intersection of the row "1 . 4" and column "0.07" is 0.9292 .

$P(Z < 1.47) = 0.9292$, normal curve to the left of $z = 1.47$.



Cumulative area under the standard normal curve for $z = 1.47$

Example
$P(\mathrm{Z} < -0.75)$

$P(Z > 1.47)$

$P(-0.75 < Z < 1.47)$

Example Answers

$P(\mathrm{Z} < -0.75) = 0.2266.$

$P(Z > 1.47) = 1 - P(Z < 1.47) = 1 - .9292 = .0708$

$$\begin{aligned} P(-0.75 < Z < 1.47) = \quad & P(Z < 1.47) - P(Z < -0.75) \\ & = 0.9292 - 0.2266 \\ & = 0.7026 \end{aligned}$$

Example 1

Arlene, the owner of a medium-scale native handicrafts shop, knows from experience that the time needed to assemble a piece of handicraft is normally distributed with a mean of 130 minutes and a standard deviation of 15 minutes.

a. Find the proportion of handicrafts that took less than 110 minutes to assemble.

b. Find the proportion of handicrafts that took anywhere between 125 and 140 minutes to assemble.

Example 2. Kilio, a running enthusiast, regularly runs around an Olympic size track oval measuring 400 meters. His trainer, Joy, observed that his running times are normally distributed with a standard deviation of ten seconds (s).

a. What is Kiko's average running time around the 400 -meter track oval if 33% of his runs took him less than 120.6 s ?

b. What proportion of Kiko's running times are above 140 s ?

c. Below how many seconds would you find the fastest 10 s of Kiko's running times?

Example 3

MENSA is an international organization requiring a high IQ for membership. It was founded in 1946 by Roland Berrill and Dr. Lancelot Ware as a nonpolitical society of bright people representing different point of view. To qualify for membership, one needs to have an IQ in the top 2% on some approved intelligence tests. One such test is the Stanford-Binet Intelligence Scales, which assumes that IQ scores are normally distributed with a mean of 100 and a standard deviation of 16 .

a. Find the proportion of people in the general population with IQs between 90 and 110 .

b. What is the lowest IQ score in the Stanford-Binet Intelligence Scales that would qualify you for MBNSA membership?

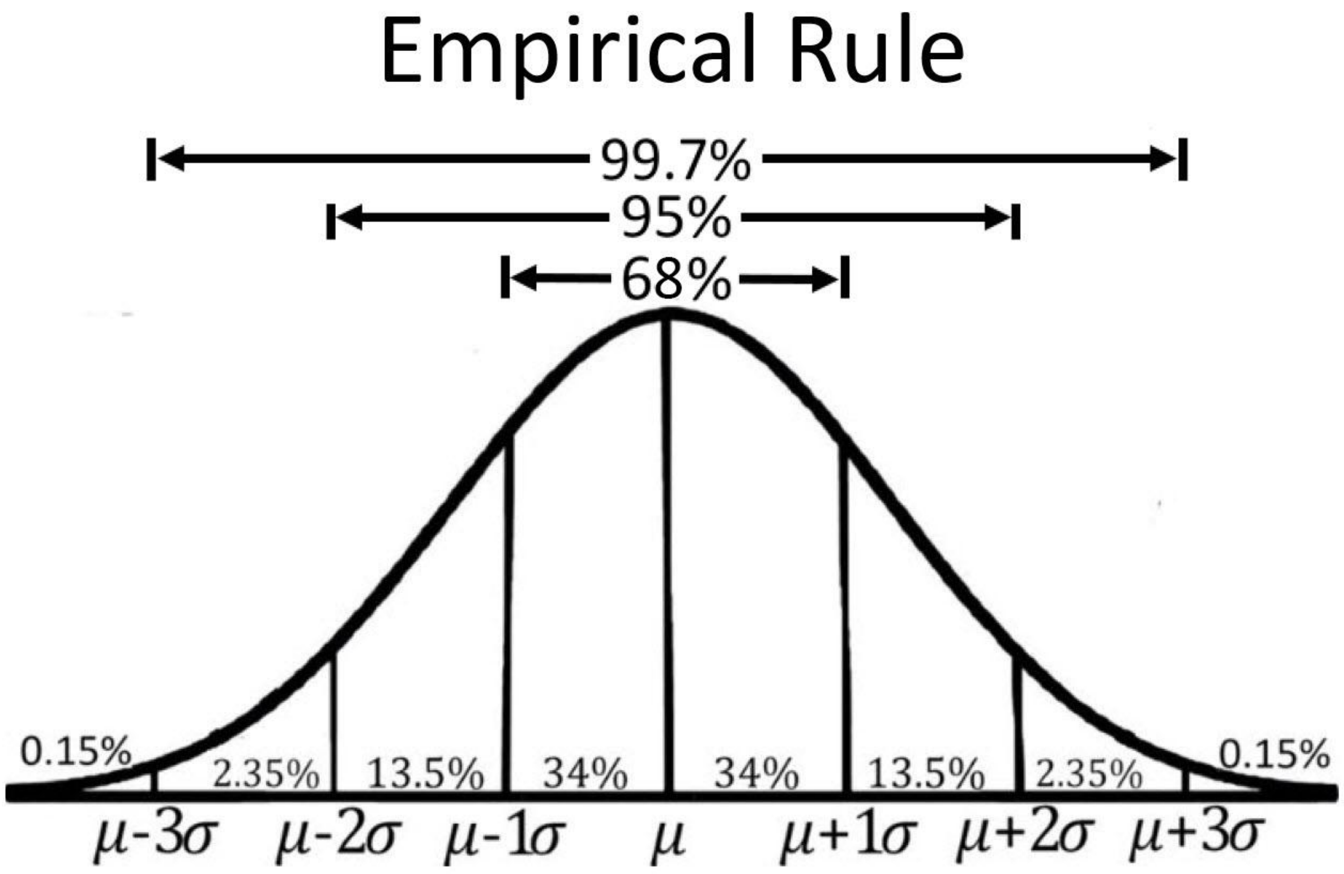
c. Leonor, a very diligent student, has an IQ score of 150 . What proportion of the general population have IQ scores above Leonor's IQ score?

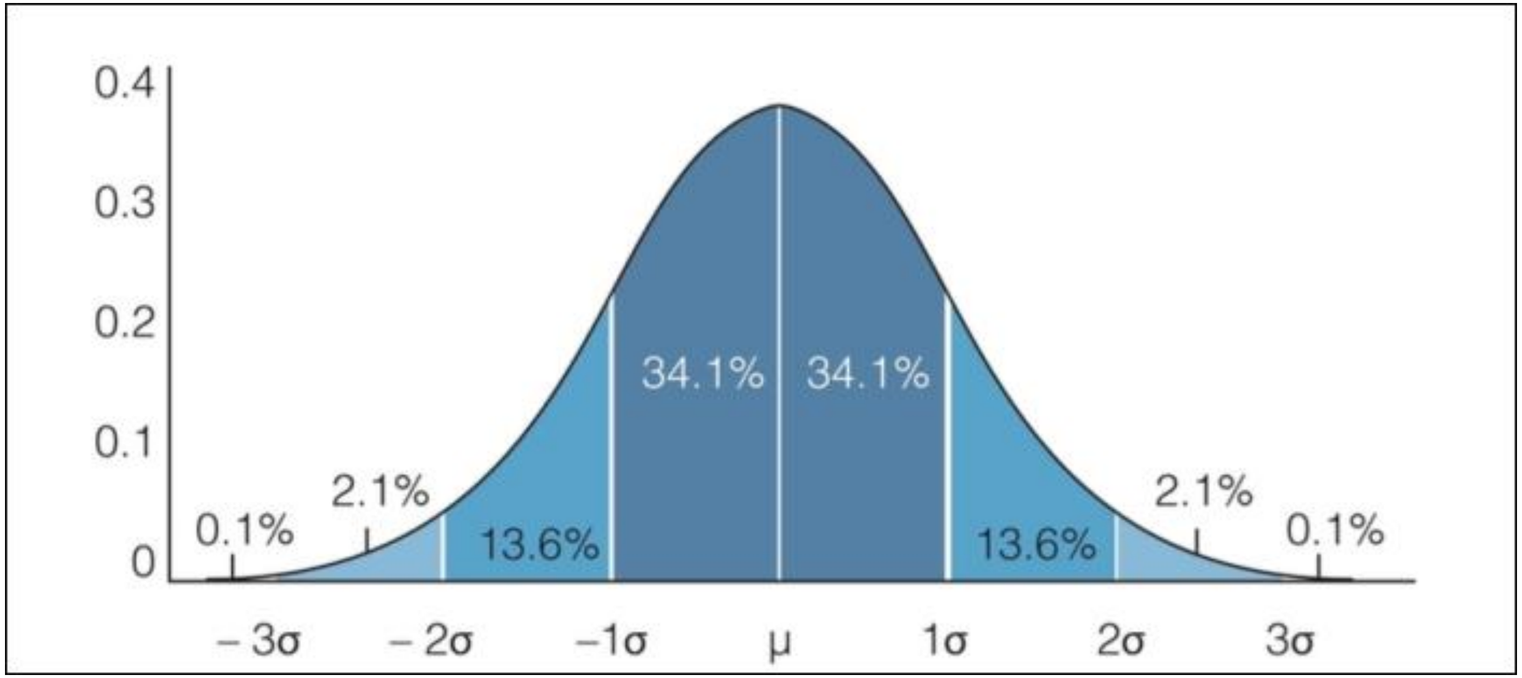
Empirical Rule

The Empirical Rule is a very useful rule of thumb for normally distributed sets of data values.
Empirical Rule
For a bell-shaped distribution of data values,

1. approximately 68% of the observations lie in the interval $[\mu - \sigma, \mu + \sigma]$;
2. approximately 95% of the observations lie in the interval $[\mu - 2\sigma, \mu + 2\sigma]$; and
3. approximately 99.7% of the observations lie in the interval $[\mu - 3\sigma, \mu + 3\sigma]$.

The following figure gives the approximate percentages of cases under the normal curve as stated in the Empirical Rule.




0.4
0.3
0.2
0.1
0
34.1%
34.1%
13.6%
13.6%
2.1%
2.1%
0.1%
0.1%
– 3σ
– 2σ
–1σ
μ
1σ
2σ
3σ

The interval $[\mu - 3\sigma, \mu + 3\sigma]$ is referred to as the 3σ-limits. Observations outside this interval are considered potential outliers or extreme values.

Example 4

If the scores in a statistics test of a large group of students are known to have a normal distribution with $\mu = 86$ and $\sigma = 4$, apply the Empirical Rule.

Solution