UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

DIPARTIMENTO DI MEDICINA E CHIURGIA

Continuous random variables & the Gaussian distribution

Continuous random variables

It can take on an infinite number of values included in an interval of finite or infinite amplitude.

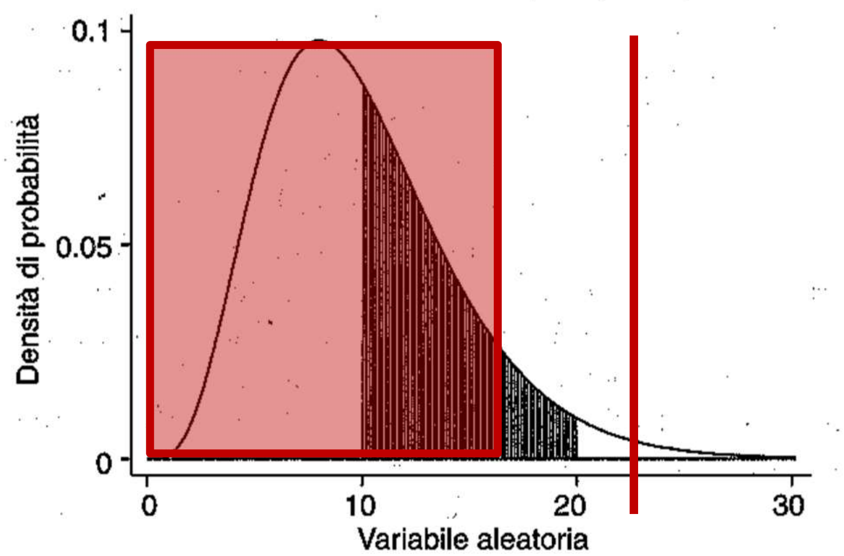
->The probability for any single value is 0 $P(X=x)=0$

->A probability is assigned for a range of values $P(a<X<b) \geq 0$

Example

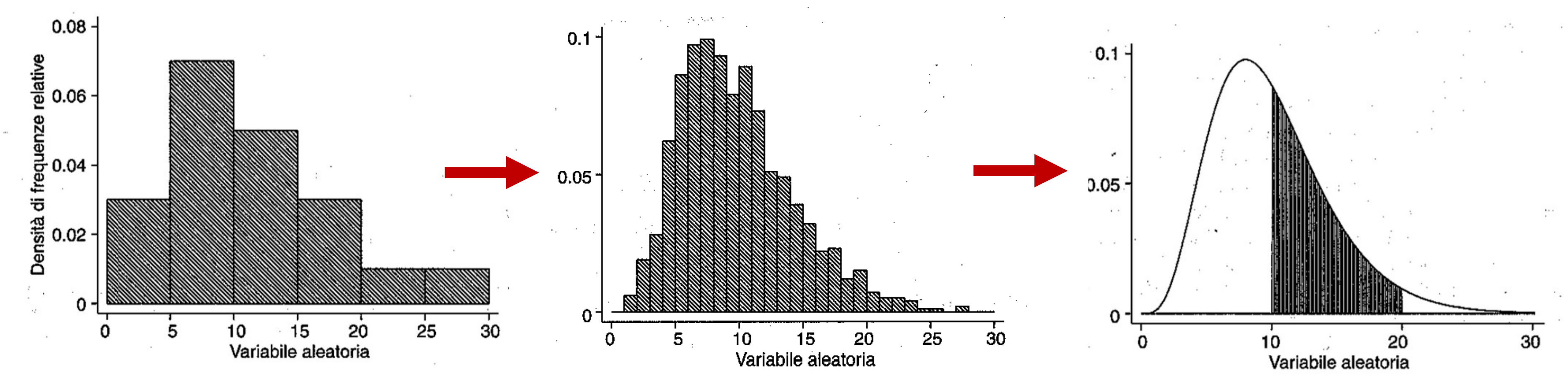
What is the probability of having a BMI of 23 kg/m^2?

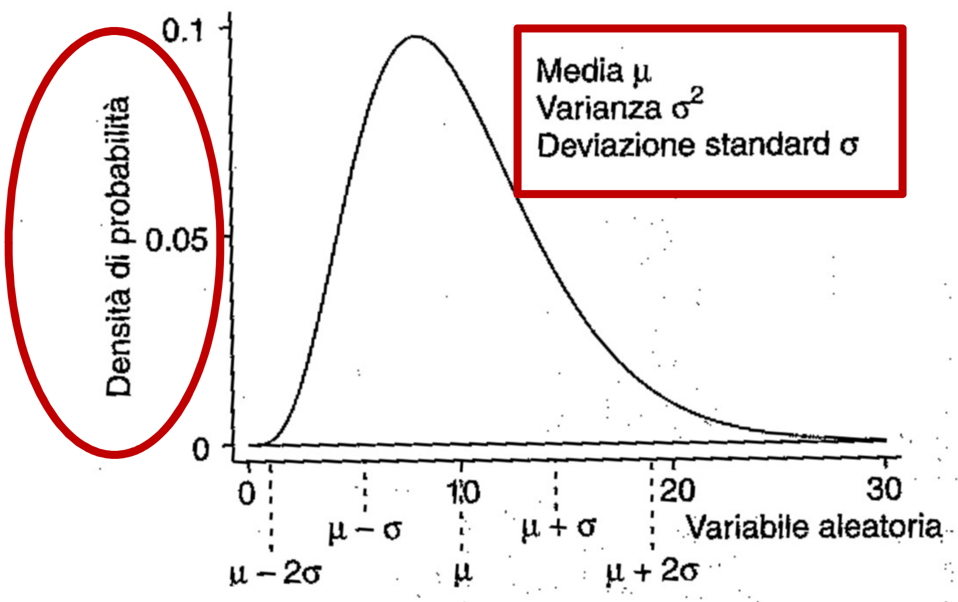
What is the probability of having a BMI < 18 kg/m^2?



From discrete to continuous…

… bins smaller and smaller, n ->∞





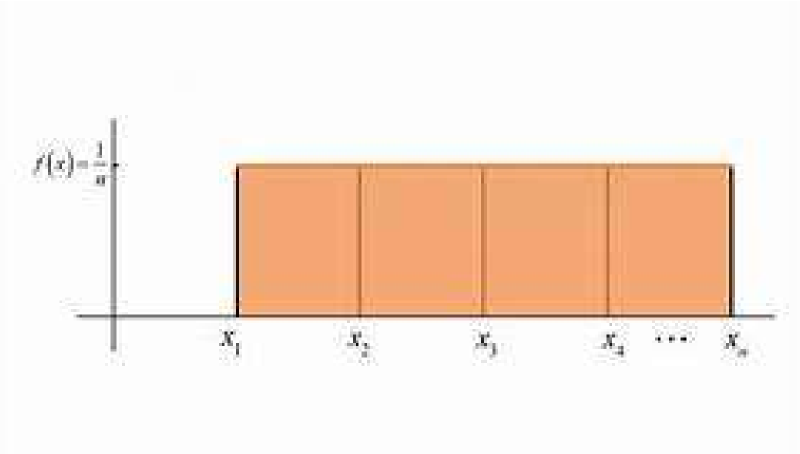
Expected value and variance for continuous random variables

$$E(X) = \int_{\Omega} xf(x)dx = \mu$$

$$Var(X) = \int_{\Omega} [x - E(X)]^2 f(x)dx = \sigma^2$$

Uniform (rectangular) Distribution

A continuous random variable has a uniform distribution if its values are spread evenly over the range. The graph of a uniform distribution results in a rectangular shape.



Density Curve

A density curve is the graph of a continuous probability distribution. It must satisfy the following properties:

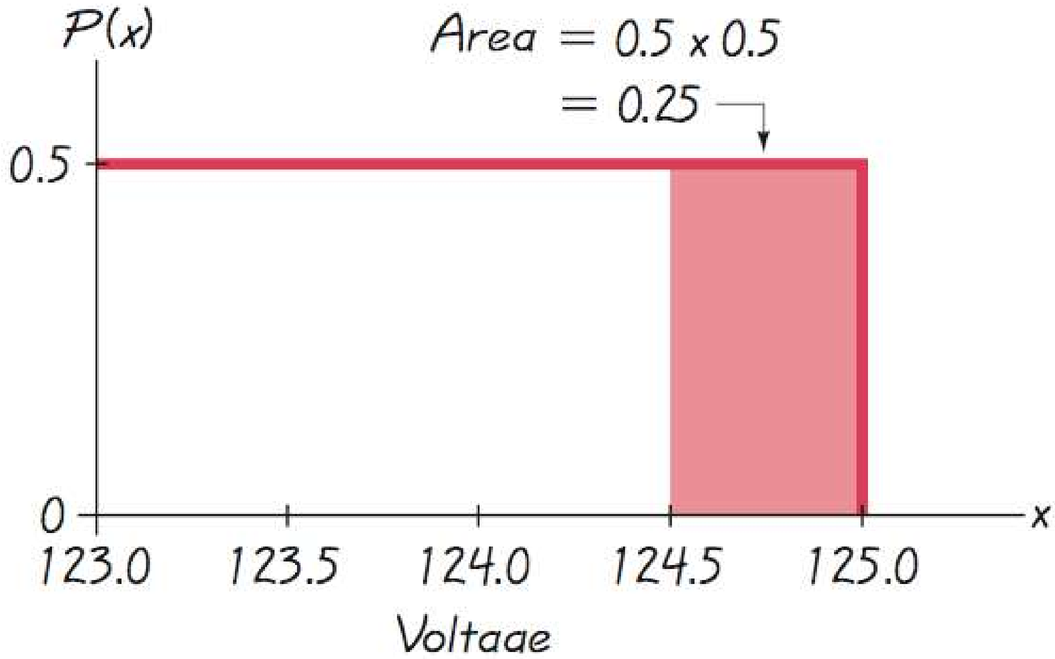
1. The total area under the curve must equal 1.
2. Every point on the curve must have a vertical height that is 0 or greater. (That is, the curve cannot fall below the *x*-axis.)

Area and Probability

Because the total area under the density curve is equal to 1, there is a correspondence between *area* and *probability*.

Using Area to Find Probability

Given the uniform distribution illustrated, find the probability that a randomly selected voltage level is greater than 124.5 volts.



Shaded area represents voltage levels greater than 124.5 volts. Correspondence between area and probability: 0.25.

Gaussian distribution

Gaussian (or normal) distribution

Karl Friedrich Gauss (1777-1855).

The random variable Gaussian plays a fundamental role because:

- describes well the manifestation of many phenomena, for example:

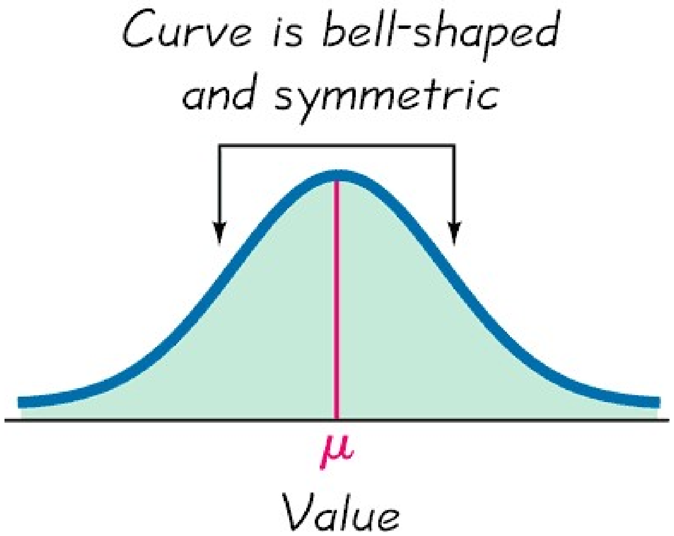
✓ Measurement errors (Gaussian genesis)
✓ Morphological characteristics (height, length)

- enjoys important properties (relevant technical aspect)

Gaussian (or normal) distribution

If a continuous random variable has a symmetric and bell-shaped distribution and it can be described by the following equation we say that it has a normal distribution.

Karl Friedrich Gauss (1777-1855).



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

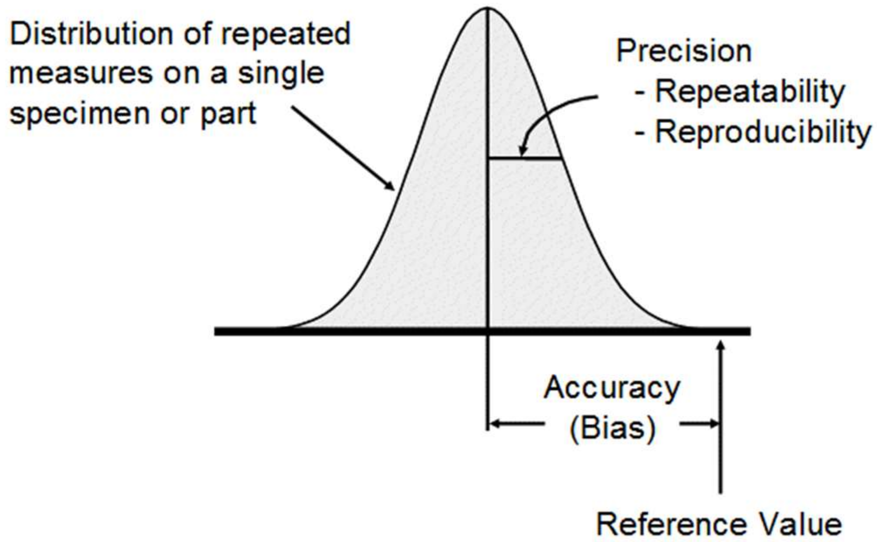
e: Euler's number, - mathematical constant approximately equal to 2.71828

π: mathematical constant, approximately equal to 3.14159

Distribution determined by fixed values of mean and standard deviation

Gaussian distribution & measurment errors

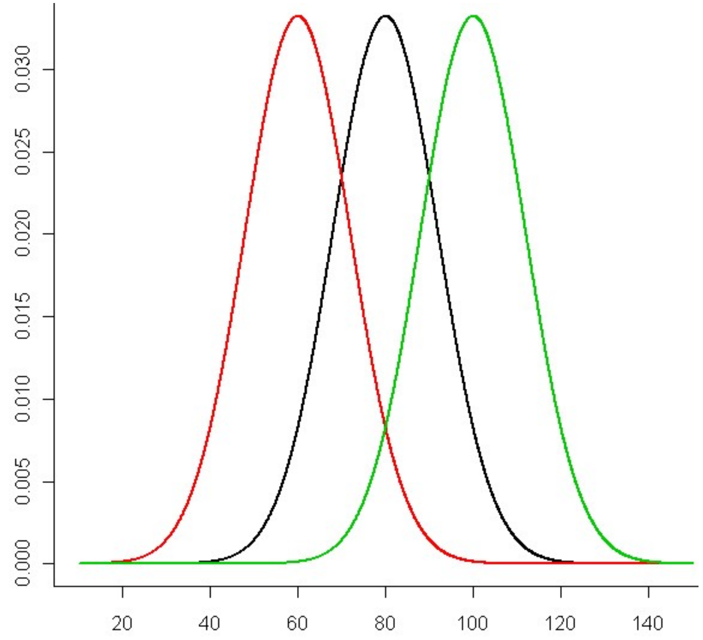
The random measurement errors ($\varepsilon = x - \mu$), taken as a whole, show a typical behavior that can be described as follows:



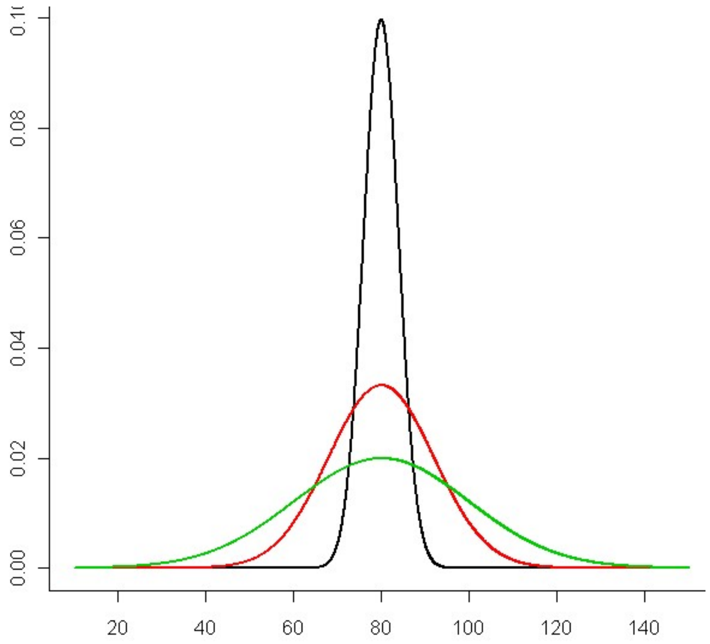
1. small errors are more frequent than large ones;
2. errors of negative sign tend to occur with the same frequency as those with a positive sign;
3. as the number of measures increases, 2/3 of the values tend to be included in the interval mean ± 1 standard deviation & 95% of the values tend to be included in the interval average ± 2 standard deviations

Gaussian parameters: μ and σ

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



N(μ=60, σ=12)
N(μ=80, σ=12)
N(μ=100, σ=12)



N(μ=80, σ=4)
N(μ=80, σ=12)
N(μ=80, σ=24)

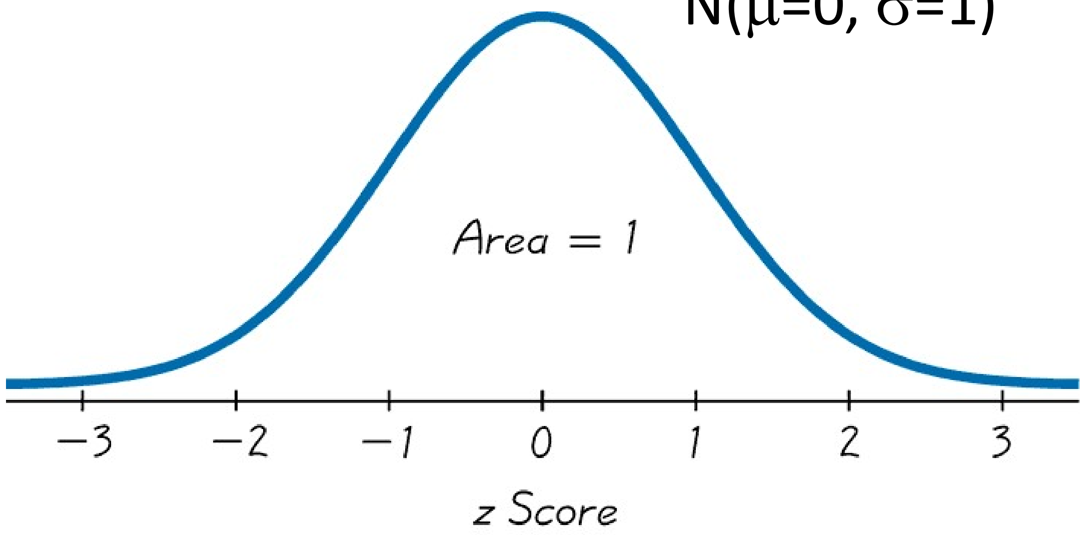
The standard Normal distribution: z score

The standard normal distribution is a specific normal distribution having the following three properties:

1. Bell-shaped (gaussian)
2. μ=0 - null mean
3. σ=1 - standard deviation equal to 1

The total area under its density curve is equal to 1 (corresponding to a probability of 100%)

$N(\mu=0, \sigma=1)$

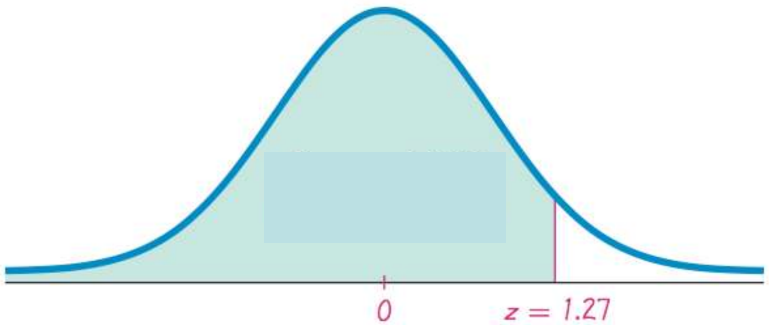


$$y = f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \cdot x^2\right]$$

Example: Bone Density Test

A bone mineral density test can be helpful in identifying the presence or likelihood of osteoporosis, a disease causing bones to become more fragile and more likely to break. The result of a bone density test is commonly measured as a z score. The population of z scores is normally distributed with a mean of 0 and a standard deviation of 1, so test results meet the requirements of a standard normal distribution.

1) A randomly selected adult undergoes a bone density test. Find the probability that this person has a bone density test score less than 1.27.



The Gaussian functions are not integrable and should be tabulated.

From Triola & Triola book

The z score –tabulated values of areas (probabilities)

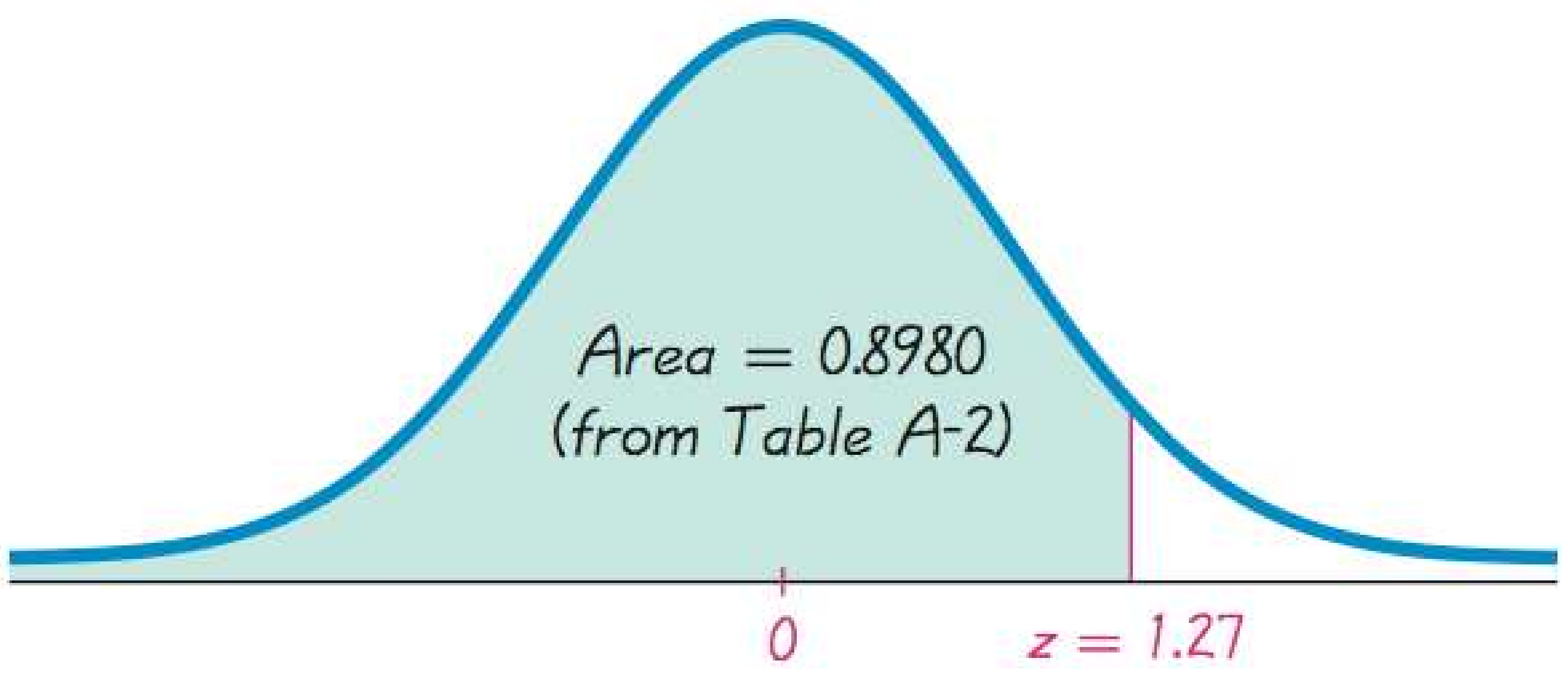
TABLE A-2 Standard Normal (*z*) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	* .0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	* .0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

Example : Bone Density Test

1) A randomly selected adult undergoes a bone density test. Find the probability that this person has a bone density test score less than 1.27.

$P(z < 1.27) =$



Look at Table A-2

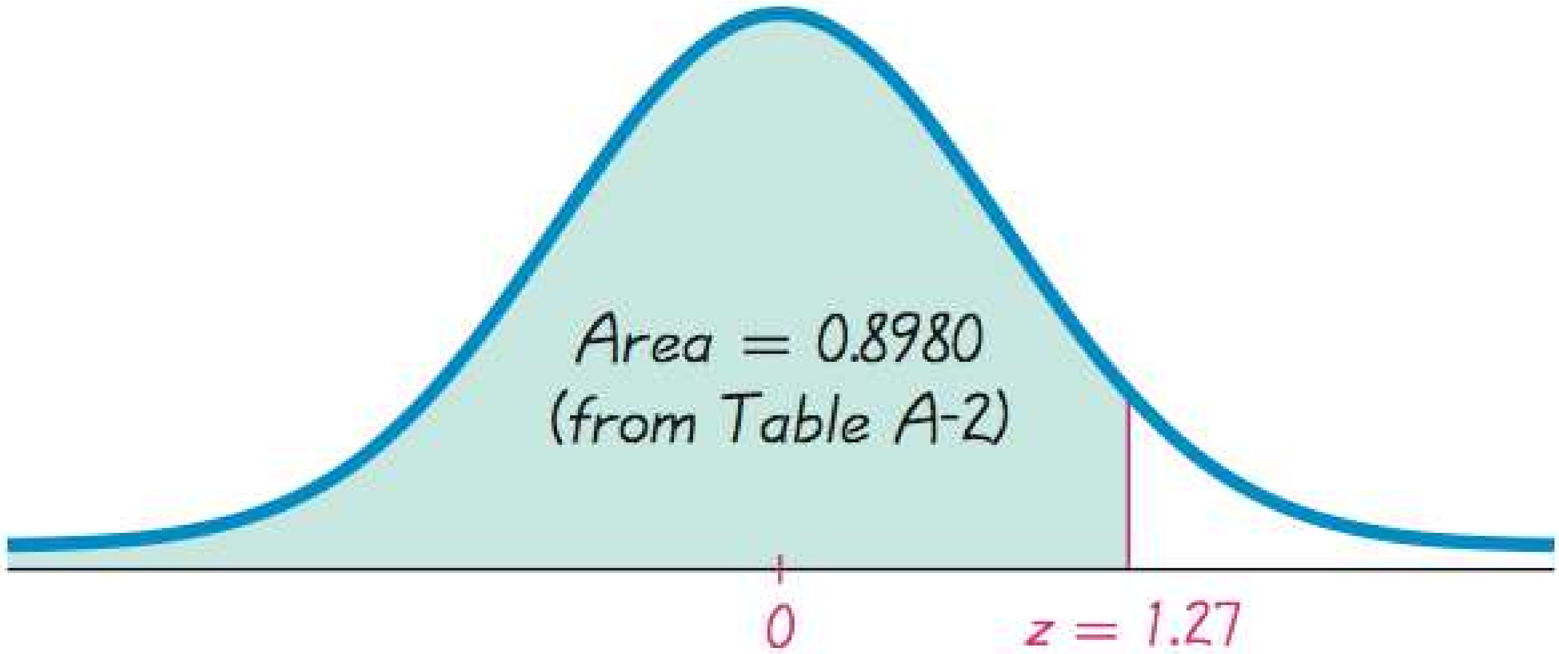
TABLE A-2 (*continued*) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292

Example : Bone Density Test

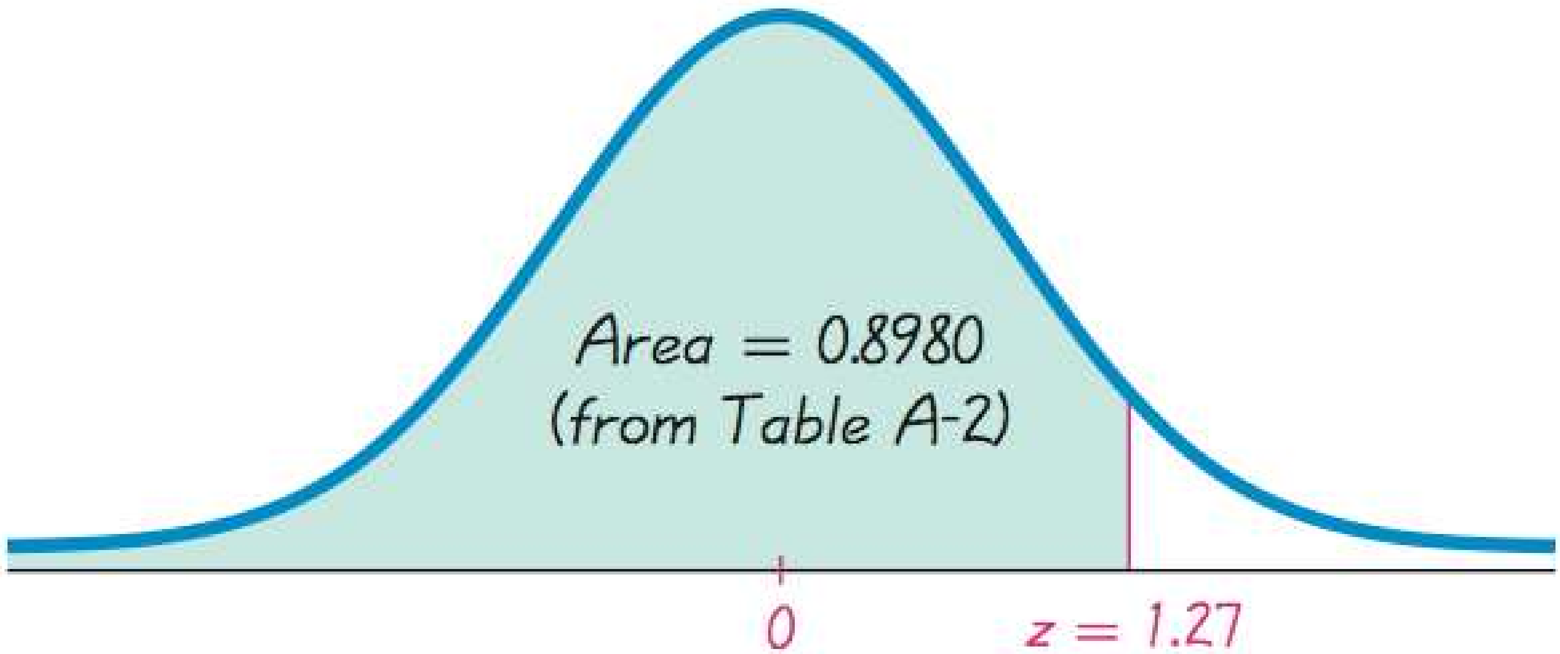
1) A randomly selected adult undergoes a bone density test. Find the probability that this person has a bone density test score less than 1.27.

$$P\ (z < 1.27) = 0.8980$$



Example : Bone Density Test

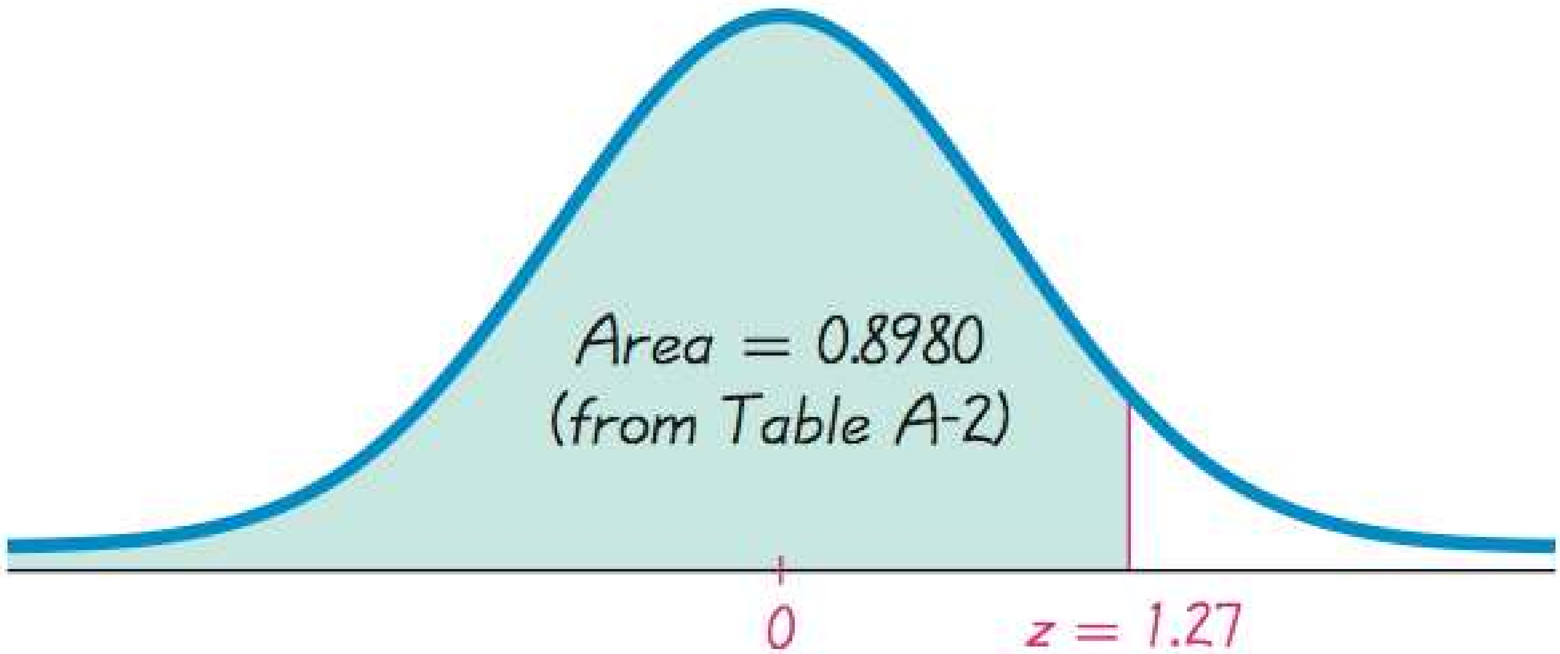
P $(z < 1.27) = 0.8980$



The *probability* of randomly selecting an adult with bone density test score less than 1.27 is 0.8980.

Example : Bone Density Test

$P(z < 1.27) = 0.8980$



Or 89.80% of adults will have bone density test score below 1.27.

Example : Bone Density Test

A bone mineral density test can be helpful in identifying the presence or likelihood of osteoporosis, a disease causing bones to become more fragile and more likely to break. The result of a bone density test is commonly measured as a z score. The population of z scores is normally distributed with a mean of 0 and a standard deviation of 1, so these test results meet the requirements of a standard normal distribution.

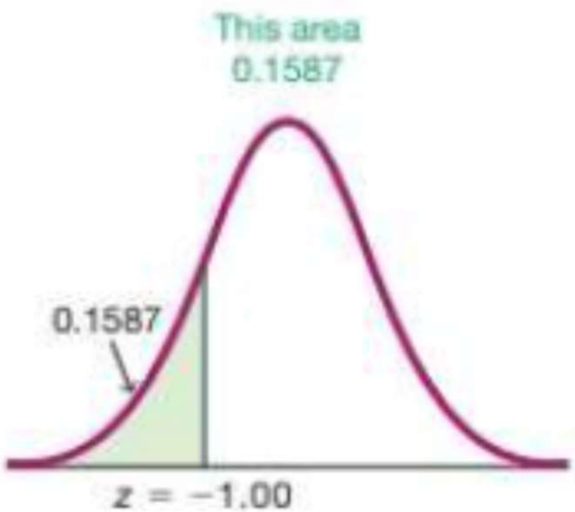
1) A randomly selected adult undergoes a bone density test. Find the probability that this person has a bone density test score less than 1.27.
2) A bone density test reading between -1.00 and -2.50 indicates that the subject has osteopenia, which is some bone loss. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.
3) Find the bone density score corresponding to P_{95}, the 95th percentile. That is, find the bone density score that separates the bottom 95% from the top 5%.

From Triola & Triola book

Exercise: Bone Density Test

2) A bone density test reading between -1.00 and -2.50 indicates that the subject has osteopenia, which is some bone loss. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.

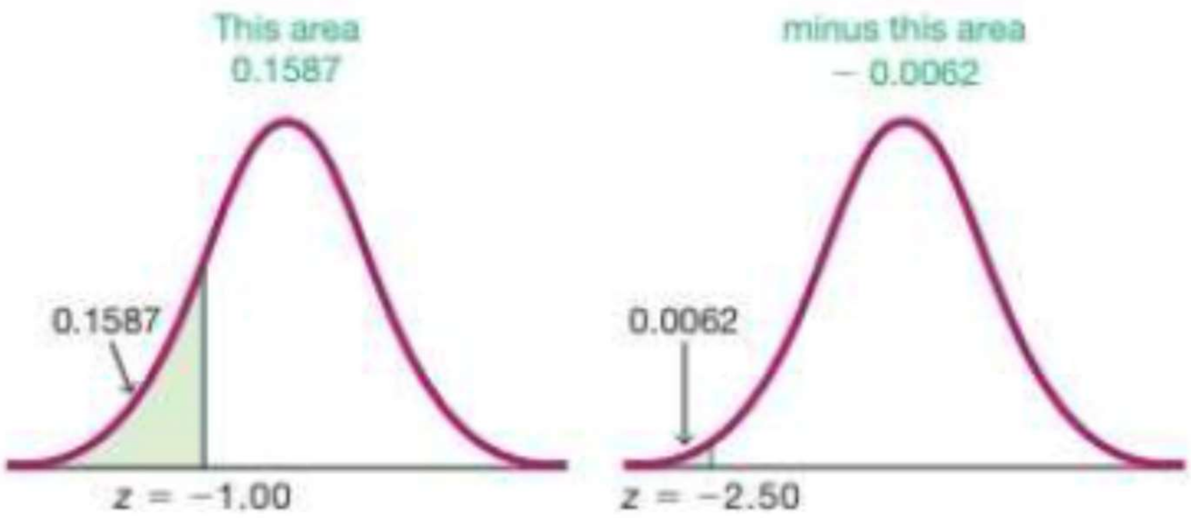
1. The area to the left of z = -1.00 is 0.1587 $\rightarrow$P(z<-1)=0.1587.



Exercise: Bone Density Test

2) A bone density test reading between -1.00 and -2.50 indicates that the subject has osteopenia, which is some bone loss. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.

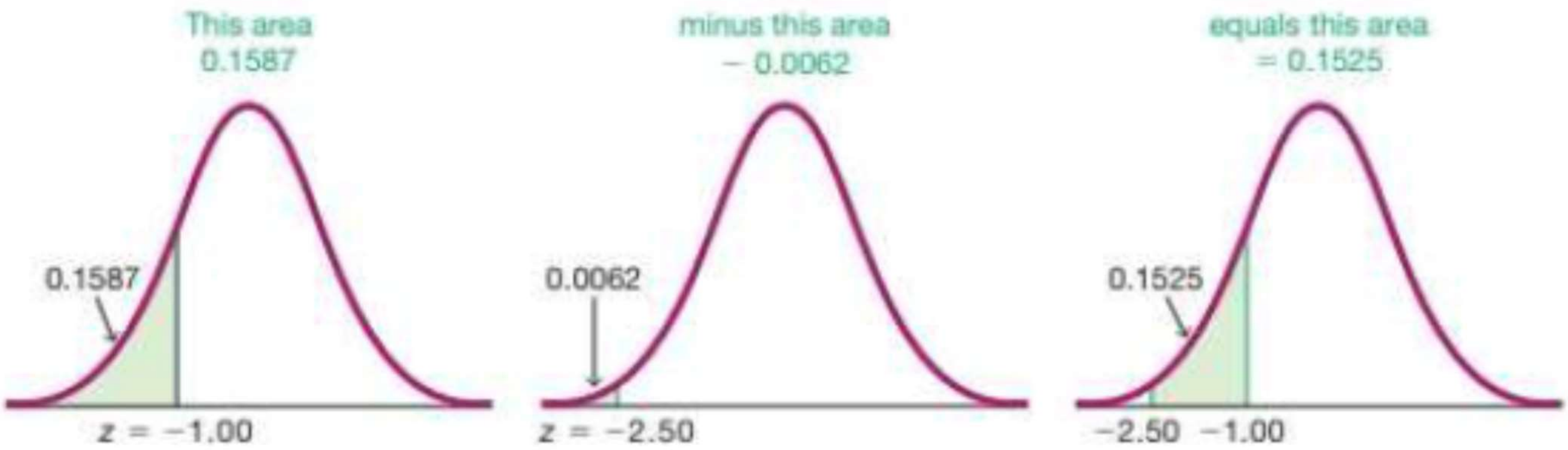
1. The area to the left of z = -1.00 is 0.1587 $\rightarrow$P(z<-1)=0.1587.
2. The area to the left of z = -2.50 is 0.0062 $\rightarrow$P(z<-2.50)= 0.0062.



Exercise: Bone Density Test

2) A bone density test reading between -1.00 and -2.50 indicates that the subject has osteopenia, which is some bone loss. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.

1. The area to the left of z = -1.00 is 0.1587 →P(z<-1)=0.1587.
2. The area to the left of z = -2.50 is 0.0062 →P(z<-2.50)= 0.0062.
3. The area between z = - 2.50 and z = - 1.00 is the difference between the areas found in the preceding two steps:



Notation

$$P(a < z < b)$$

denotes the probability that the *z* score is between *a* and *b*.

$$P(z > a)$$

denotes the probability that the *z* score is greater than *a*.

$$P(z < a)$$

denotes the probability that the *z* score is less than *a*.

Exercise: Bone Density Test

3) Find the bone density score corresponding to P_{95}, the 95th percentile. That is, find the bone density score that separates the bottom 95% from the top 5%.

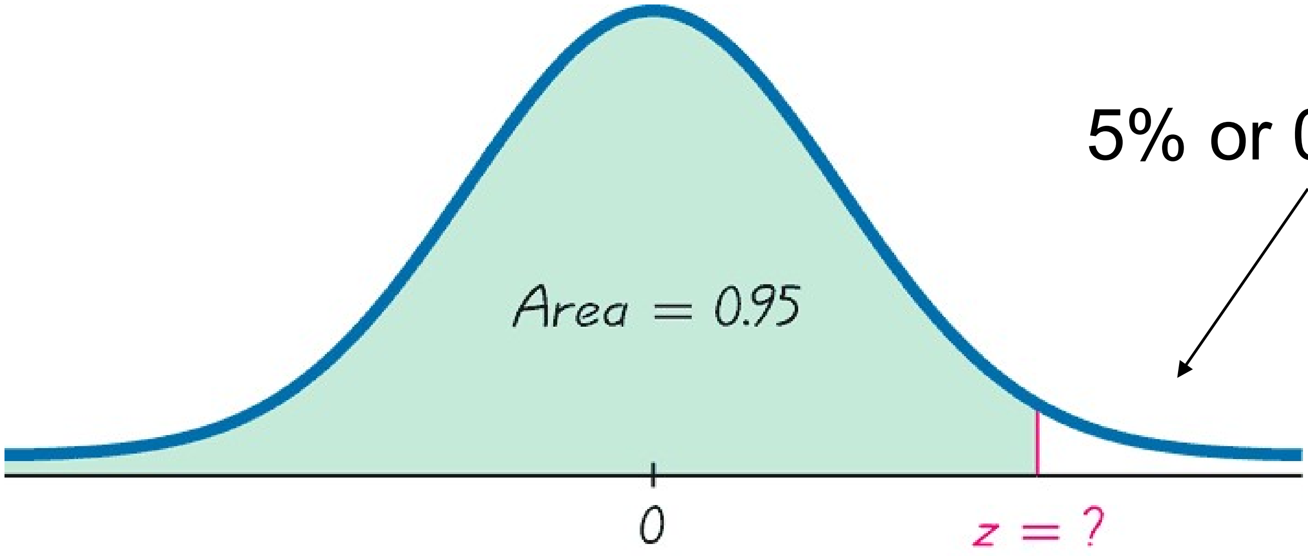
Finding a z Score When Given a Probability Using Table A-2

1. Draw a bell-shaped curve and identify the region under the curve that corresponds to the given probability. If that region is not a cumulative region from the left, work instead with a known region that is a cumulative region from the left.

2. Using the cumulative area from the left, locate the closest probability in the body of Table A-2 and identify the corresponding z score.

Exercise: Bone Density Test

3) Find the bone density score corresponding to P_{95}, the 95th percentile. That is, find the bone density score that separates the bottom 95% from the top 5%.



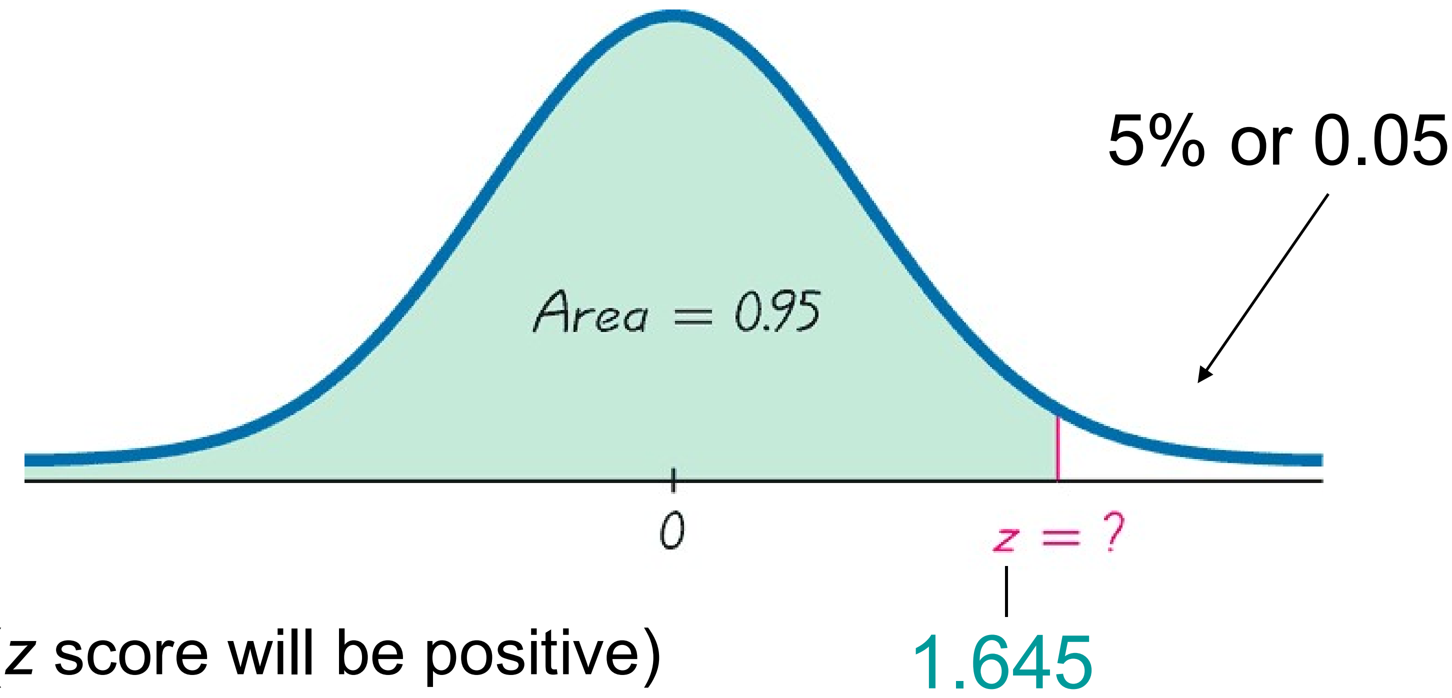
5% or 0.05

(z score will be positive)

Finding the 95th Percentile

Exercise: Bone Density Test

3) Find the bone density score corresponding to P_{95}, the 95th percentile. That is, find the bone density score that separates the bottom 95% from the top 5%.

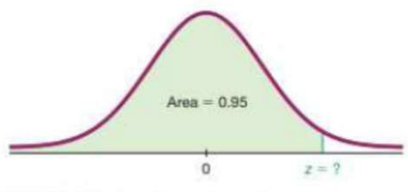


(z score will be positive)

Finding the 95^{th} Percentile

Exercise: Bone Density Test

3) Find the bone density score corresponding to P_{95}, the 95th percentile. That is, find the bone density score that separates the bottom 95% from the top 5%.



INTERPRETATION
For bone density test scores, 95% of the scores are less than or equal to 1.645, and 5% of them are greater than or equal to 1.645.

Example:

Adults have pulse rates with a mean of 72 bpm (beats per minute), a standard deviation of 13 bpm, and a distribution that is approximately normal (data from GISSI-HF prevention trial).

The normal range is generally considered to be between 60 bpm and 100 bpm.

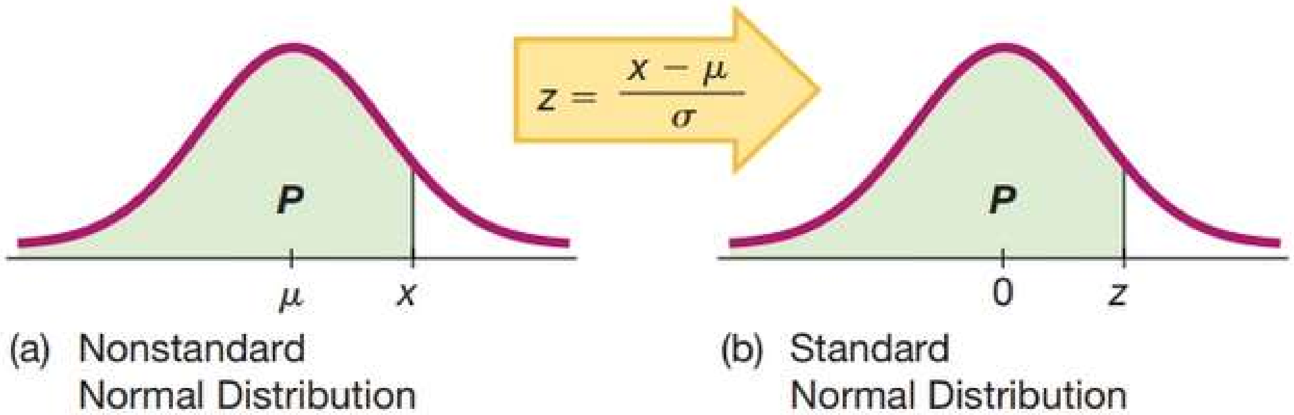
What is the proportion of adults who are expected to have Tachycardia (pulse rates greater than 100 bpm)?

$N(\mu=72, \quad \sigma=13)$

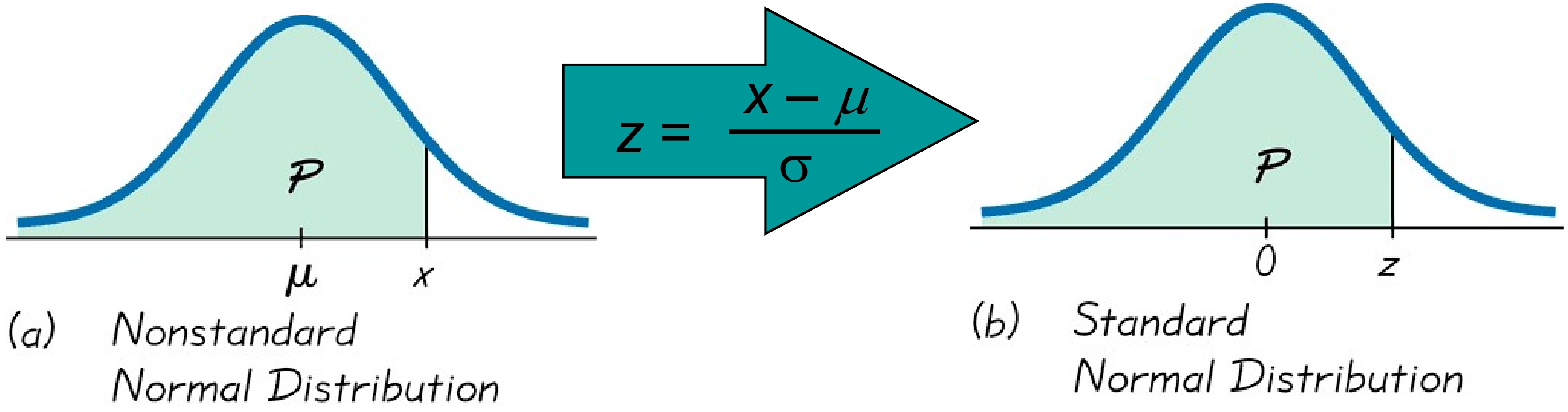
How can we do with PULSE variable?

pulse~N(72,13)

In order to work with any nonstandard normal distribution (with a mean different from 0 and/or a standard deviation different from 1)
the key is a simple conversion that allows us to "standardize" any normal distribution so that x values can be transformed to z scores; then the methods of the preceding section can be used.



Converting to a Standard Normal Distribution



Round *z* scores to 2 decimal places

Exercise

In GISSI-prevention trial we noted that pulse rates of adult are normally distributed with a mean of 72 bpm and a standard deviation of 13 bpm.

1)Find the proportion of adults with a pulse rate greater than 100 bpm. These are considered to be at a high risk of stroke, heart disease, or cardiac death.

2) Normal pulse rates are generally considered to be between 60 bpm and 100 bpm, find the percentage of subjects with normal pulse rates.

3) Find the pulse rate that separates the highest 1% from the lowest 99%. That is, find P_{99}.

pulse~N(72,13)

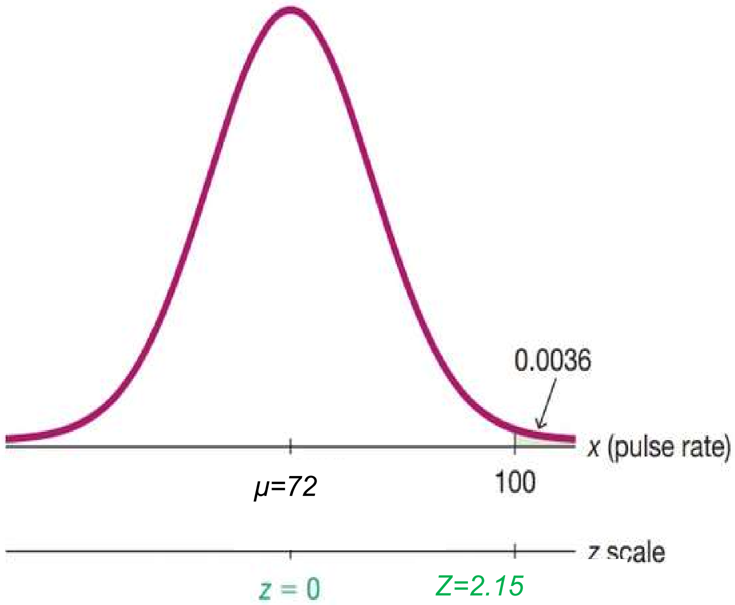
Exercise

In GISSI-prevention trial we noted that pulse rates of adult are normally distributed with a mean of 72 bpm and a standard deviation of 13 bpm.

1)Find the proportion of adults with a pulse rate greater than 100 bpm. These are considered to be at a high risk of stroke, heart disease, or cardiac death.

pulse~N(72,13)

Z=(100-72)/13=2.15



Exercise

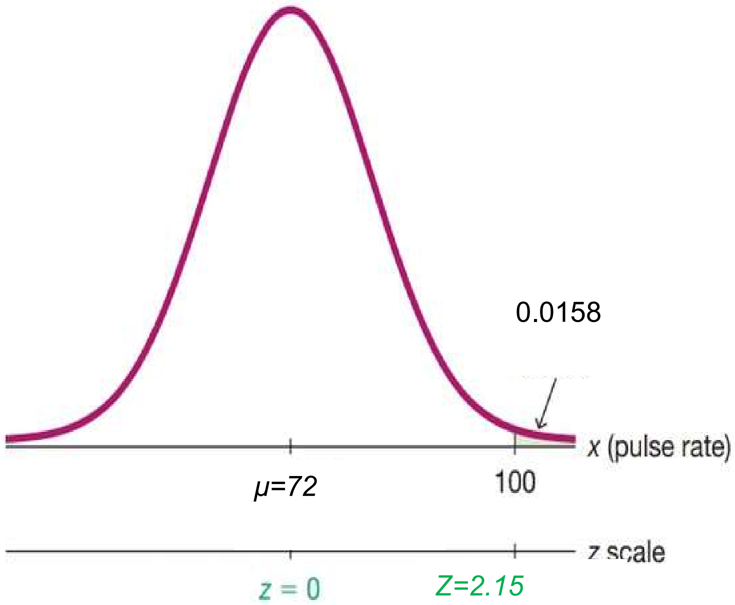
In GISSI-prevention trial we noted that pulse rates of adult are normally distributed with a mean of 72 bpm and a standard deviation of 13 bpm.

1)Find the proportion of adults with a pulse rate greater than 100 bpm. These are considered to be at a high risk of stroke, heart disease, or cardiac death.

pulse~N(72,13)

Z=(100-72)/13=2.15

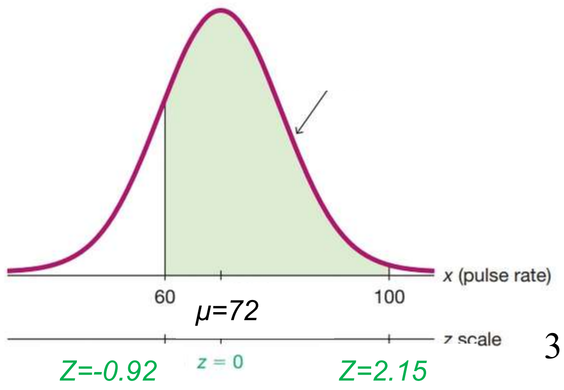
P(X>100)=P(Z>2.15)=1-P(Z<2.15)=
1-0.9842=0.0158



Solution

2) Normal pulse rates are generally considered to be between 60 bpm and 100 bpm, find the percentage of subjects with normal pulse rates.

Z=(60-72)/13=-**0.92**

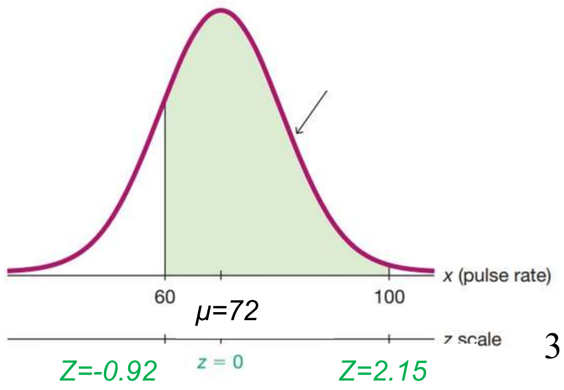


Solution

2) Normal pulse rates are generally considered to be between 60 bpm and 100 bpm, find the percentage of subjects with normal pulse rates.

Z=(60-72)/13=-**0.92**

P(X<60)=P(Z<-0.92)=0.1788 *(from the table)*



Solution

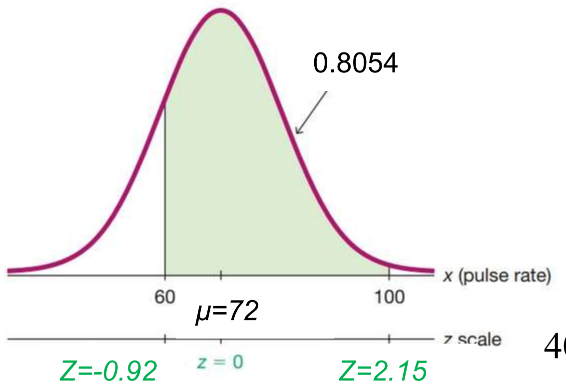
2) Normal pulse rates are generally considered to be between 60 bpm and 100 bpm, find the percentage of subjects with normal pulse rates.

Z=(60-72)/13=-**0.92**

P(X<60)=P(Z<-0.92)=0.1788 *(from the table)*

P(60<X<100)=P(-0.92<Z<2.15)=

=1-0.0158-0.1788=0.8054



Solution

3) Find the pulse rate that separates the highest 1% from the lowest 99%. That is, find P_{99}.

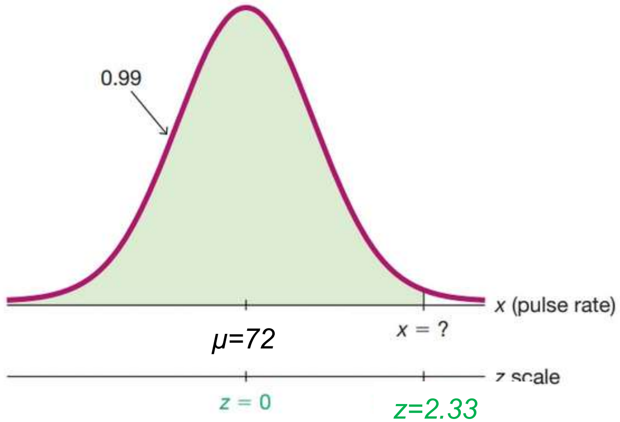


FIGURE 6-14 Finding the 99th Percentile

$P(X>x_{0.99})=P(Z>z_{0.99})=0.01$

$z_{0.99}=2.33$ (from the table)

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

$$2.33 = \frac{x-70}{13}$$

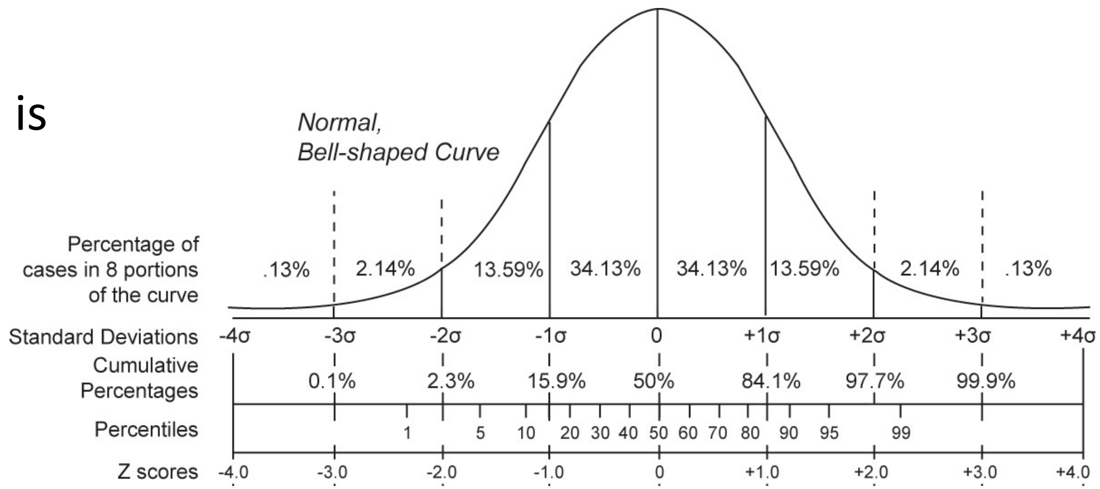
$$x = 70 + (2.33 * 13) = 100.29$$

Conversion of percentiles into z-scores

If a distribution of data is approximately symmetric and bell-shaped, about 95% of the data should fall within two standard deviations of the mean.

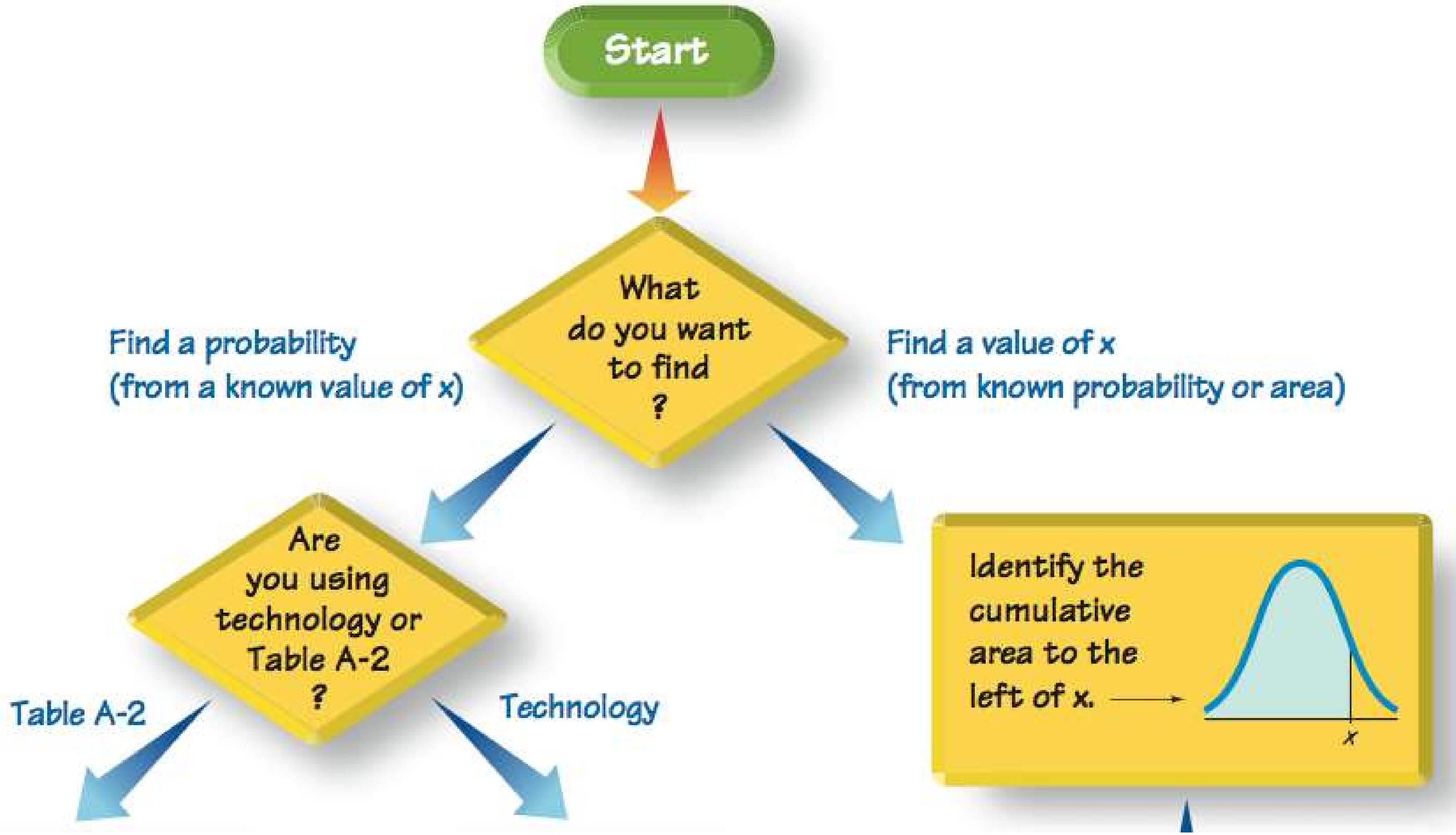
The ***z-score*** for a data value, *x*, is

$$z = \frac{x - \bar{x}}{s}$$

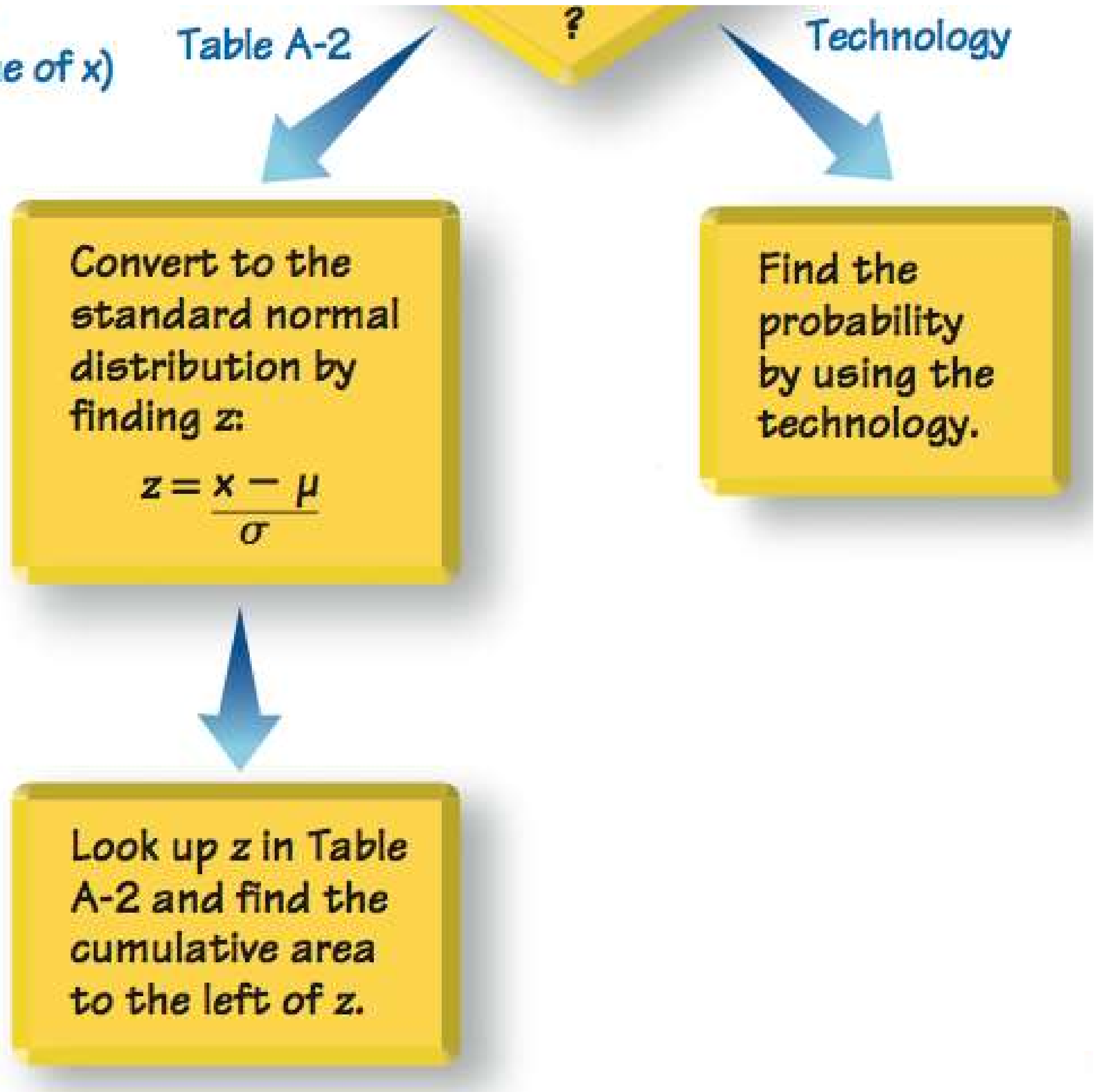


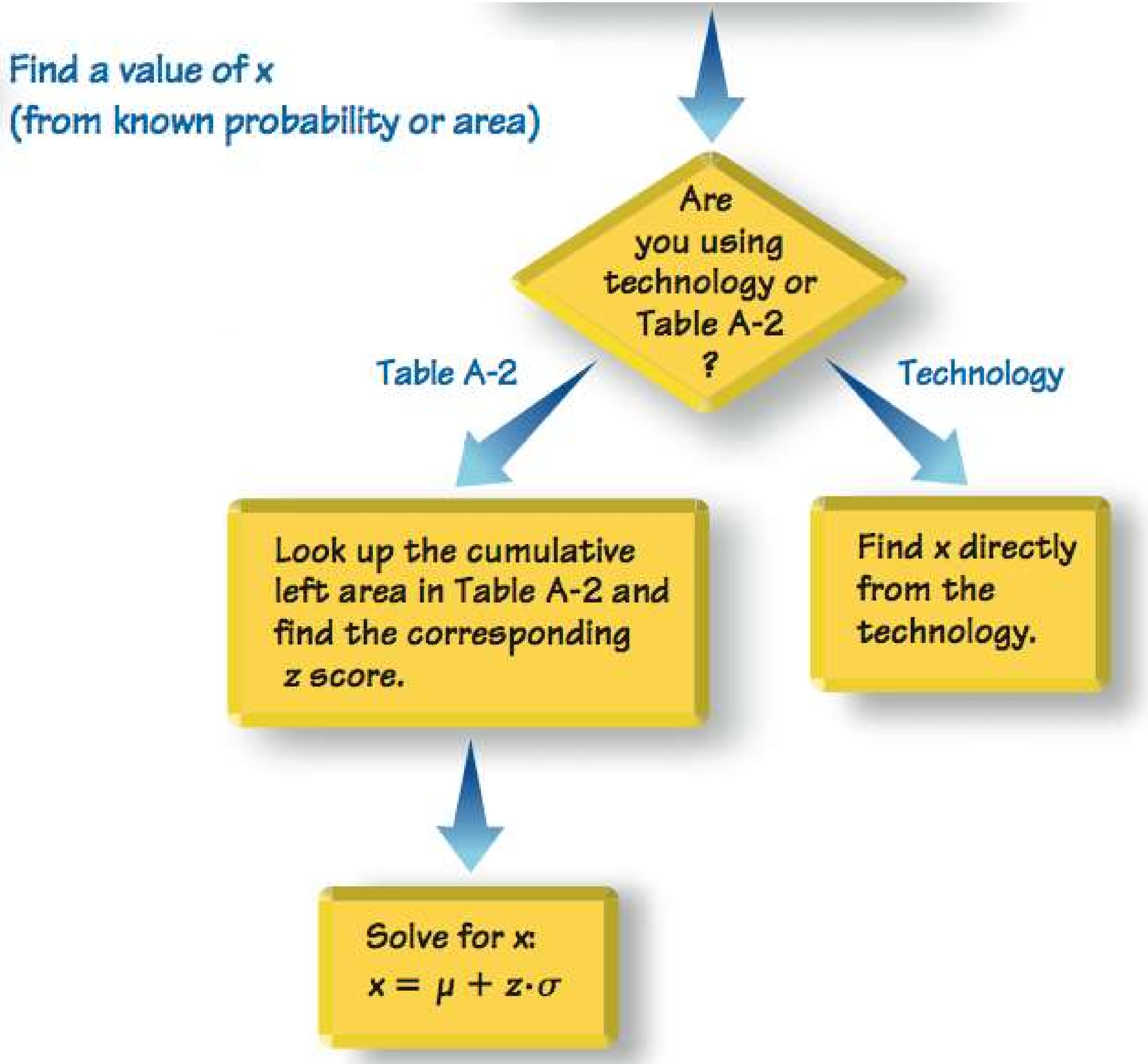
Note 1 - z-score puts values on a common scale
Note 2- z-score is the number of standard deviations a value falls from the mean

Applications with Normal Distributions

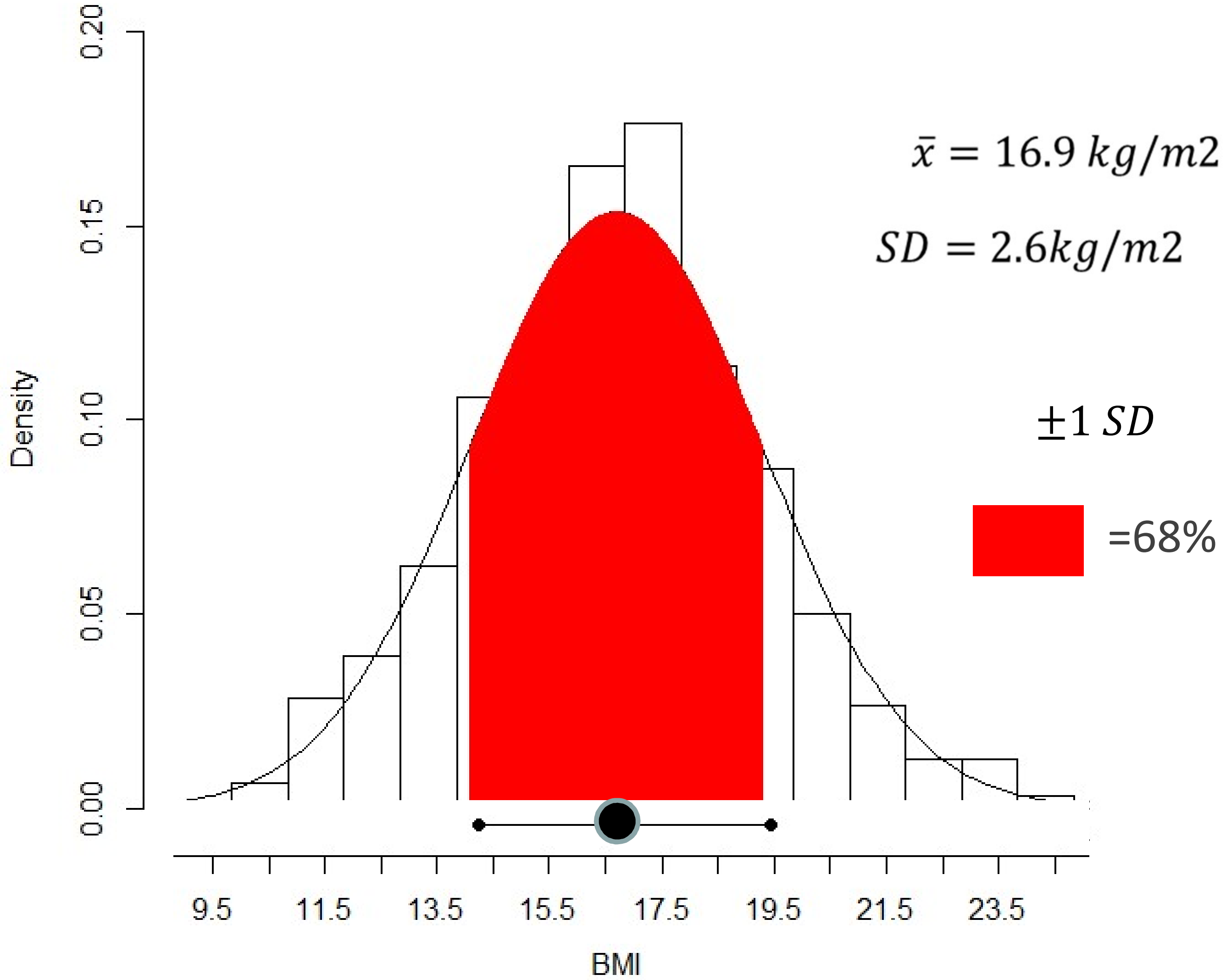


Find a probability
(from a known value of x)

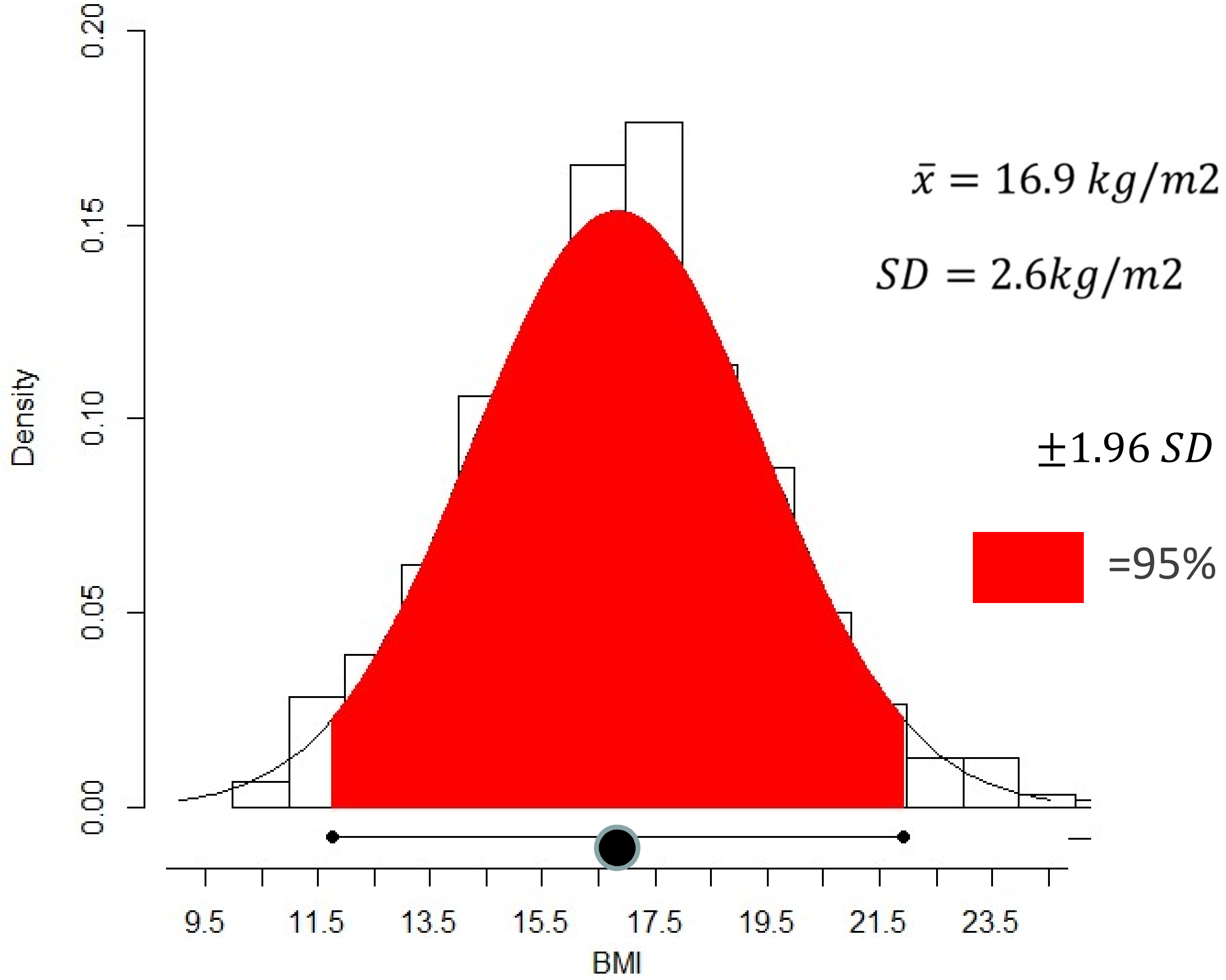



Find a value of x
(from known probability or area)
Are you using technology or Table A-2 ?
Table A-2
Technology
Look up the cumulative left area in Table A-2 and find the corresponding z score.
Find x directly from the technology.
Solve for x:
$x = \mu + z \cdot \sigma$

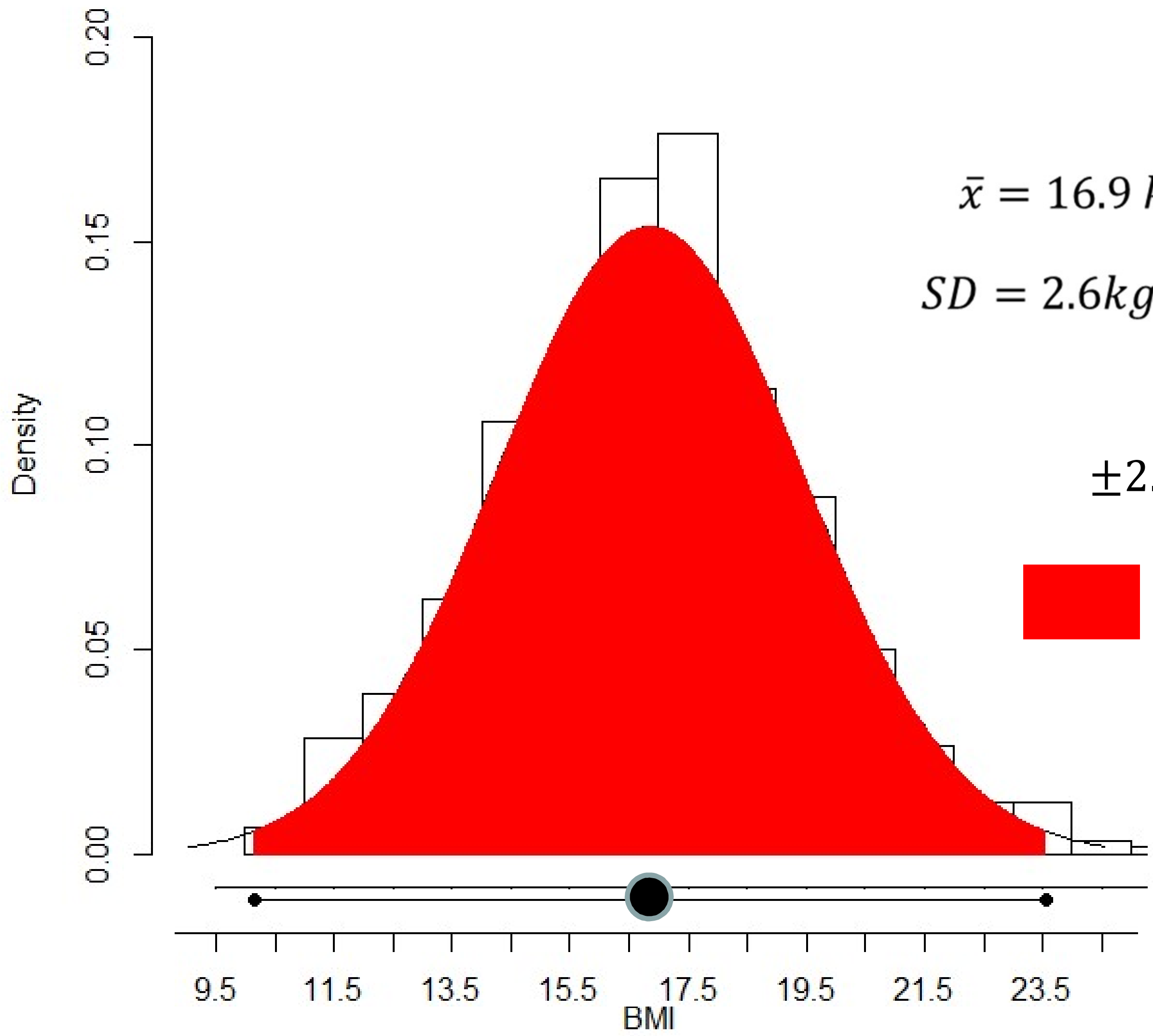
Histogram with Gaussian approximation – intervals around μ



Histogram with Gaussian approximation – intervals around μ



Histogram with Gaussian approximation – intervals around μ



$\bar{x} = 16.9\ kg/m2$

$SD = 2.6 kg/m2$

$\pm 2.54\ SD$

=99%