General Sir John Kotelawala Defense University

Department of Mathematics

MA 3102 APPLIED STATISTICS

Prof. AM Razmy

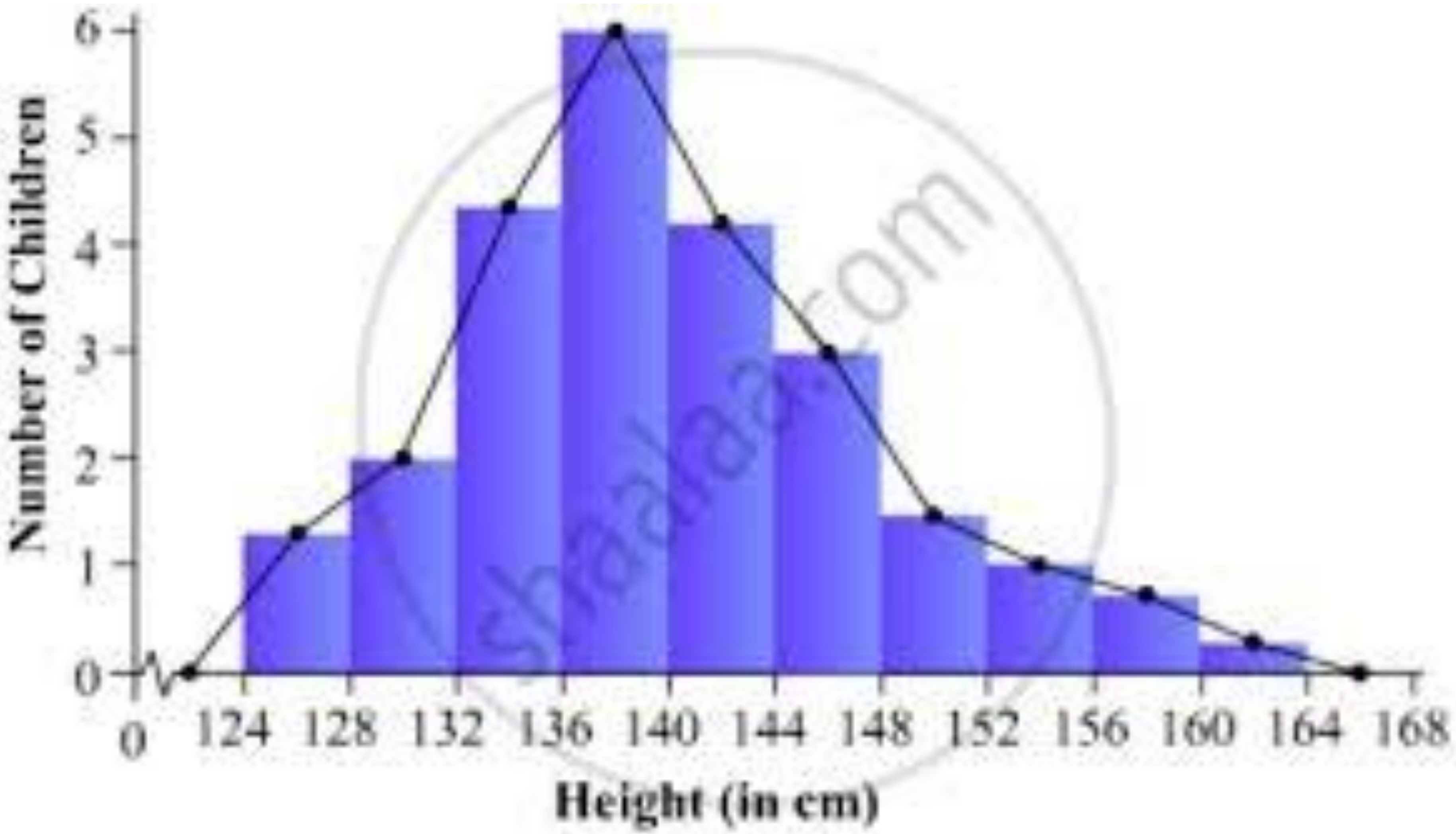
Module 1

Week	Teaching & Learning Activities	Duration Minutes	TLA Type	LO
1	▪ Introduction to Normal distribution	15	Lecture	LO1
	▪ Use Normal Tables	15		
	▪ Applications of the Normal Distribution	30		
	▪ Sampling distributions for population means & the Central Limit Theorem	15		
	▪ Applications of Central Limit Theorem	15		
	▪ Student t distribution	15		
	▪ Application of t distribution and Use of t table	15		

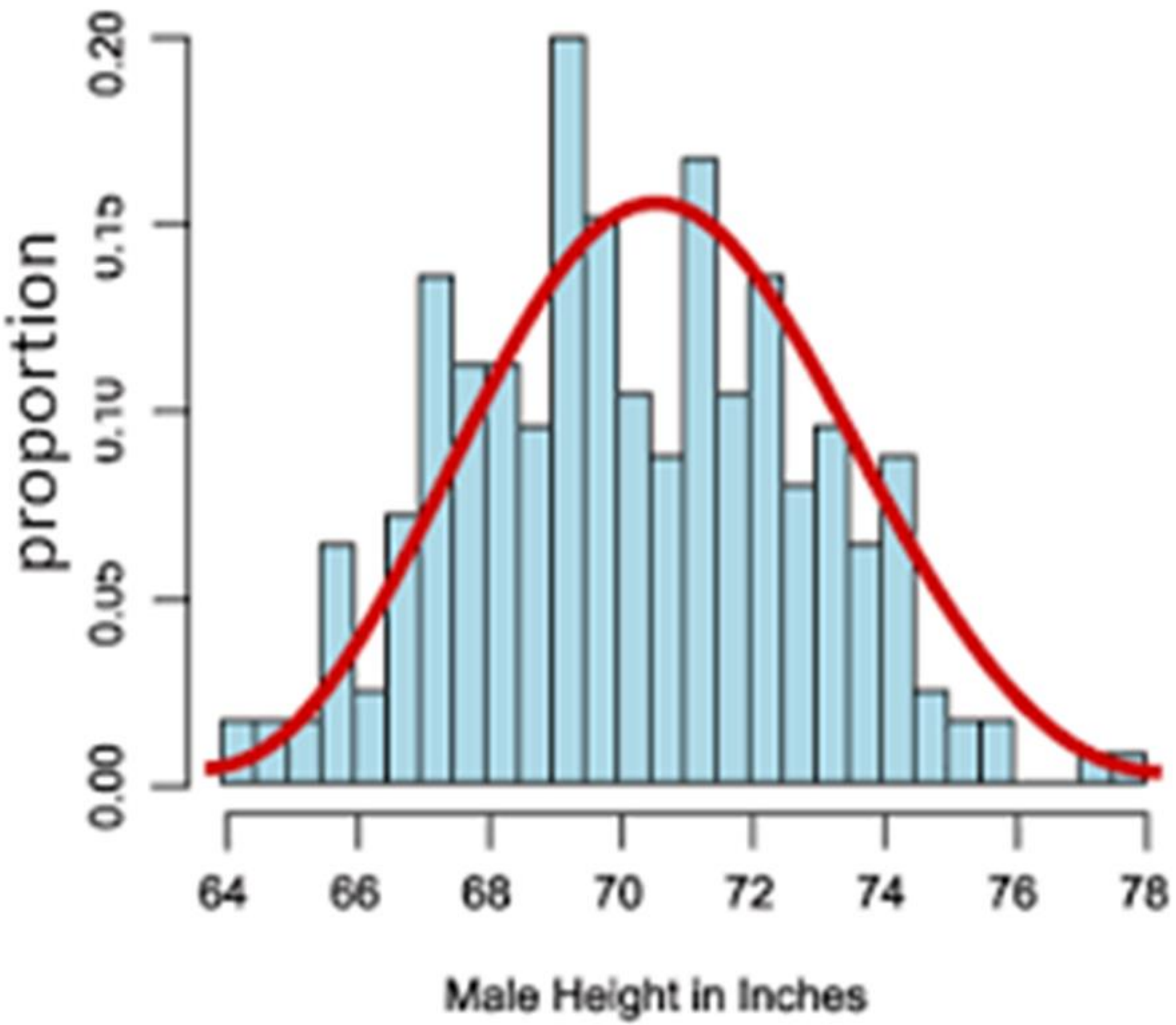
Probability and Statistics for Engineers

Fifth Edition

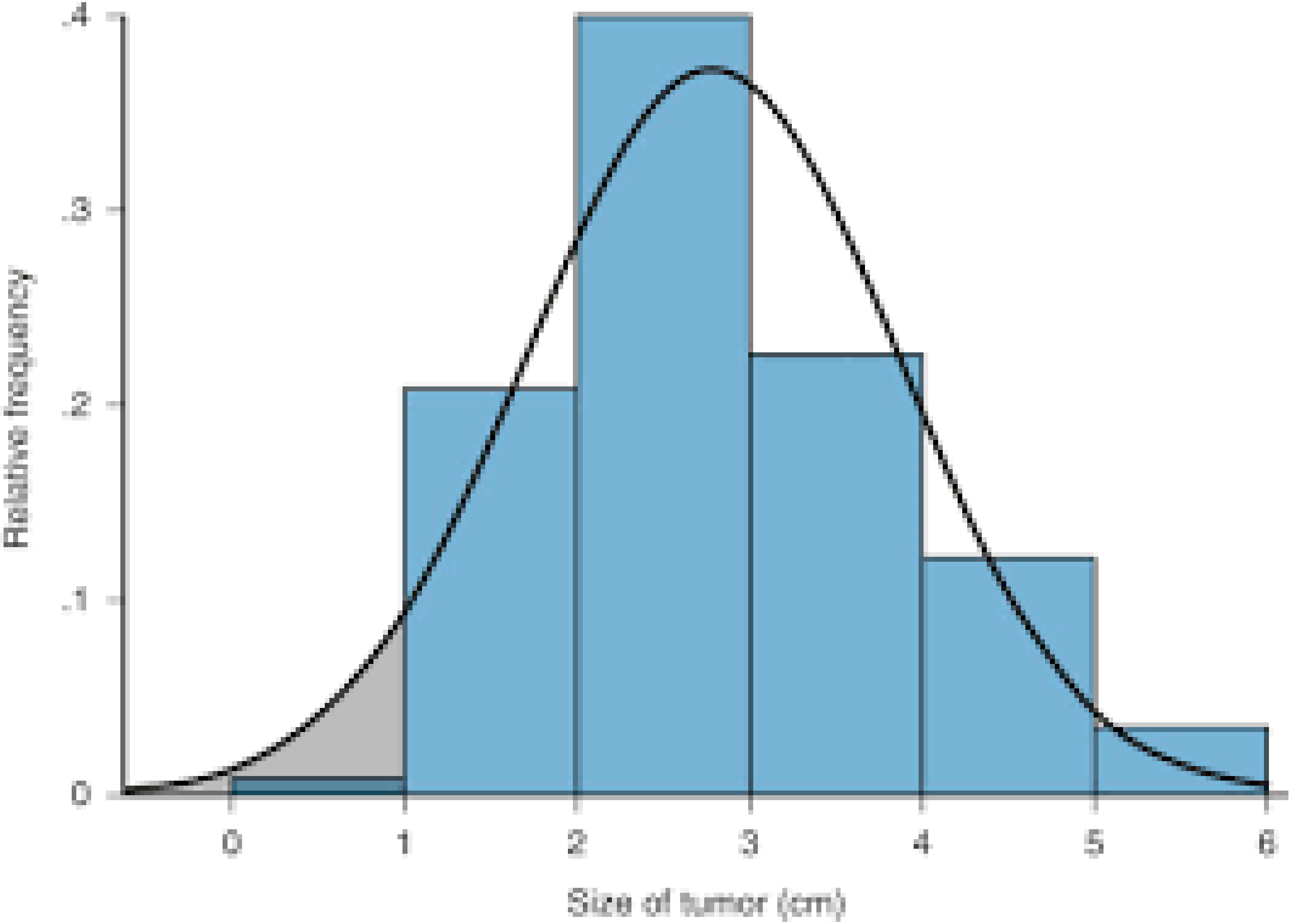
Scheaffer   Mulekar   McClave



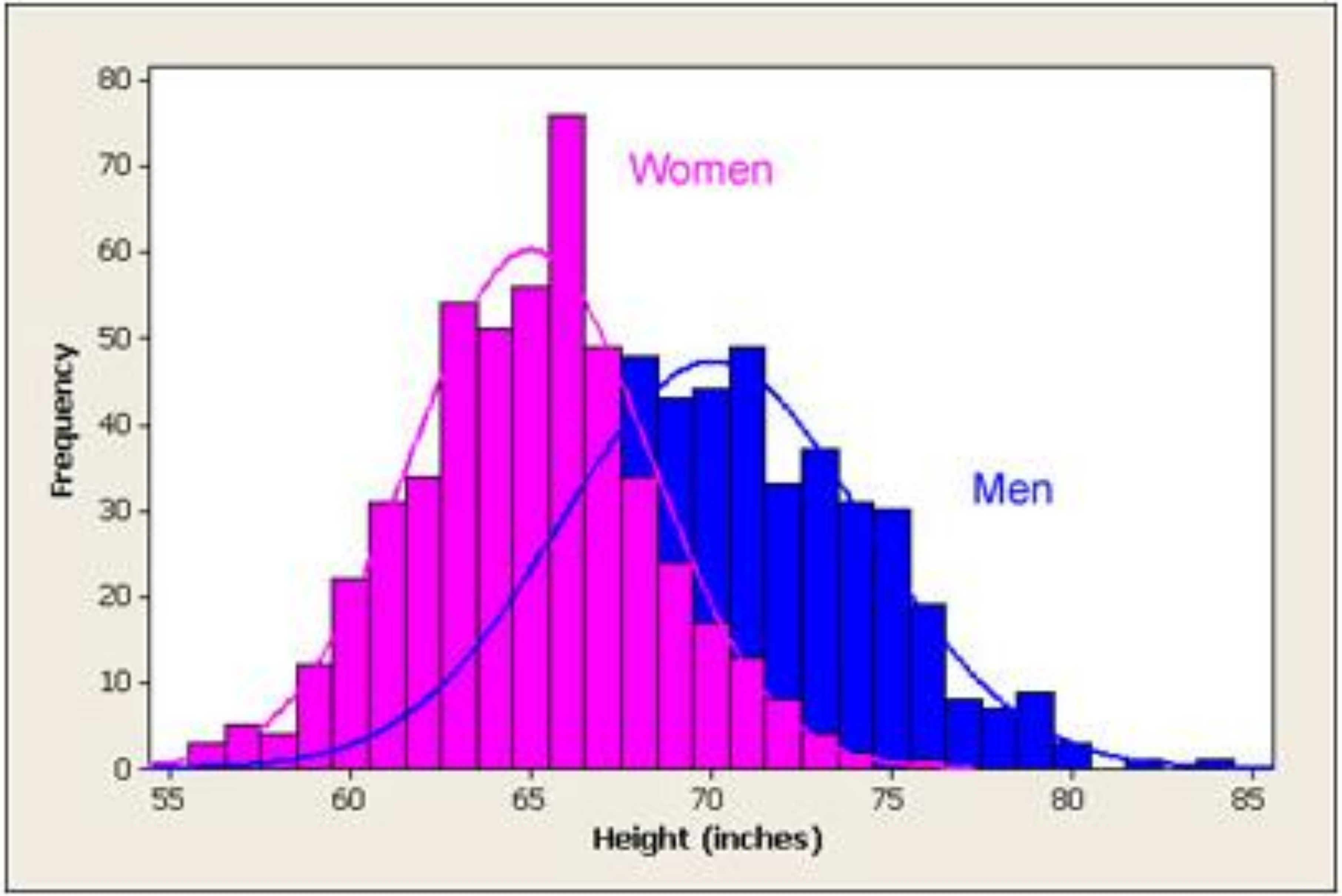
Number of Children
6
5
4
3
2
1
0
0
124
128
132
136
140
144
148
152
156
160
164
168
Height (in cm)



proportion
0.00
0.05
0.10
0.15
0.20
64
66
68
70
72
74
76
78
Male Height in Inches



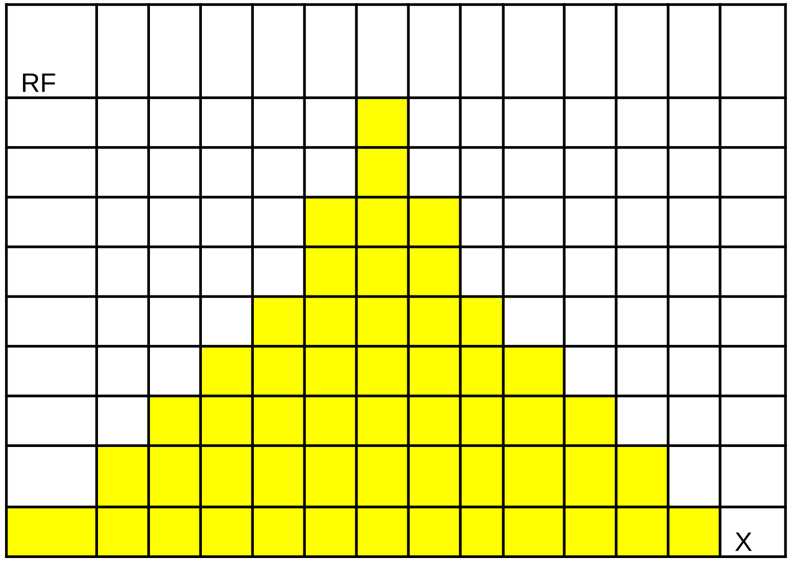
Relative frequency
.4
.3
.2
.1
0
0
1
2
3
4
5
6
Size of tumor (cm)

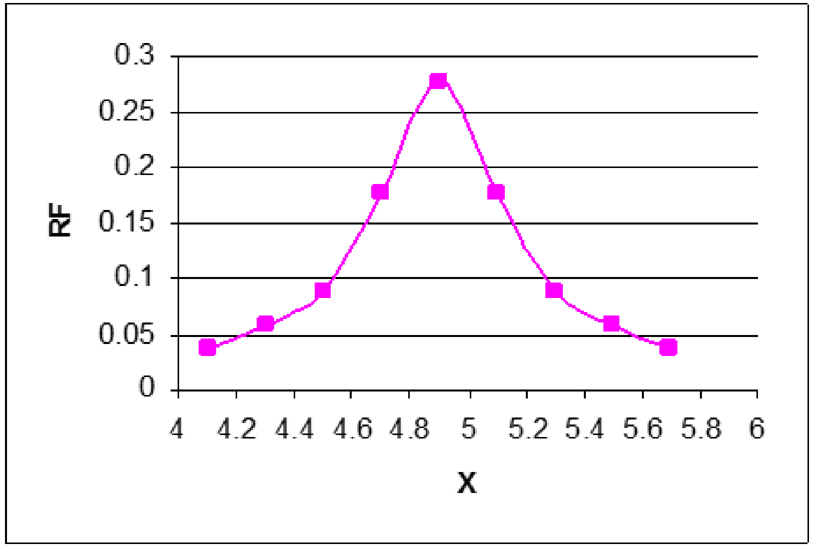


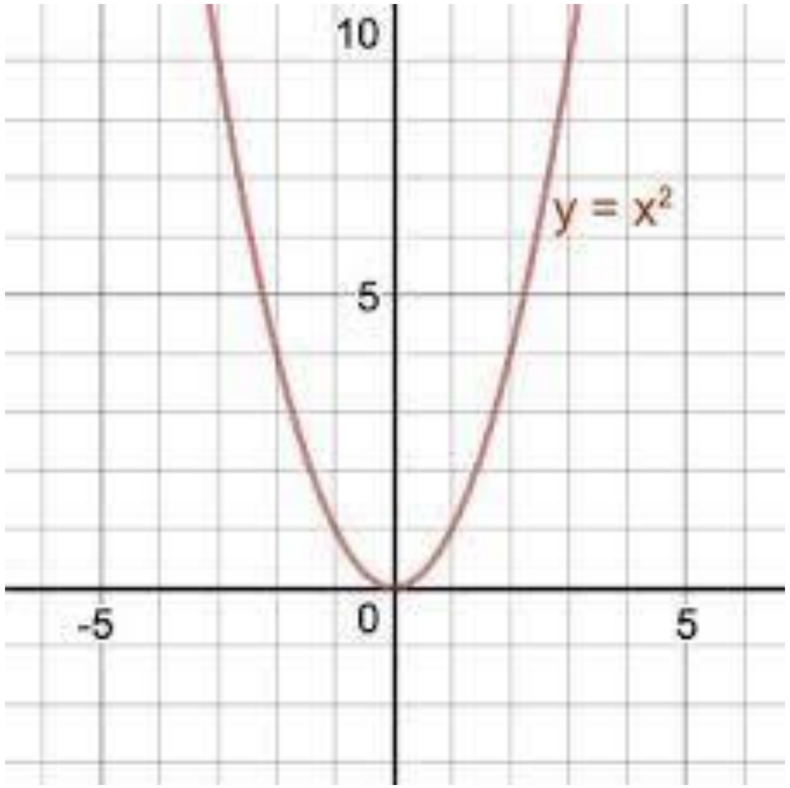
Women
Men
Frequency
Height (inches)
0
10
20
30
40
50
60
70
80
55
60
65
70
75
80
85

The Normal Distribution

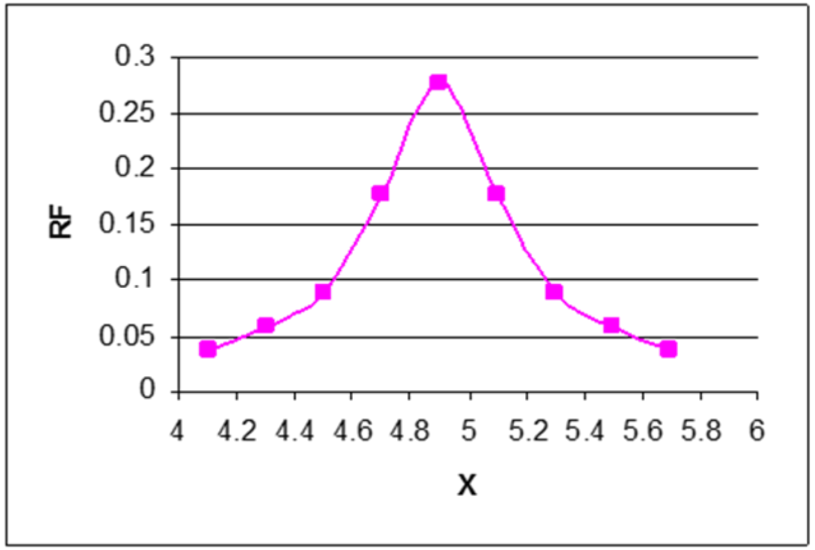
For most of the data found in nature, when the number of observations increases and the class intervals very small, the relative frequency histogram verges towards smooth symmetrical bell shaped curve called normal frequency curve or normal curve or normal distribution density curve. The function (equation for this curve) is called the normal frequency function or more commonly known as the normal probability density function, denoted by f.





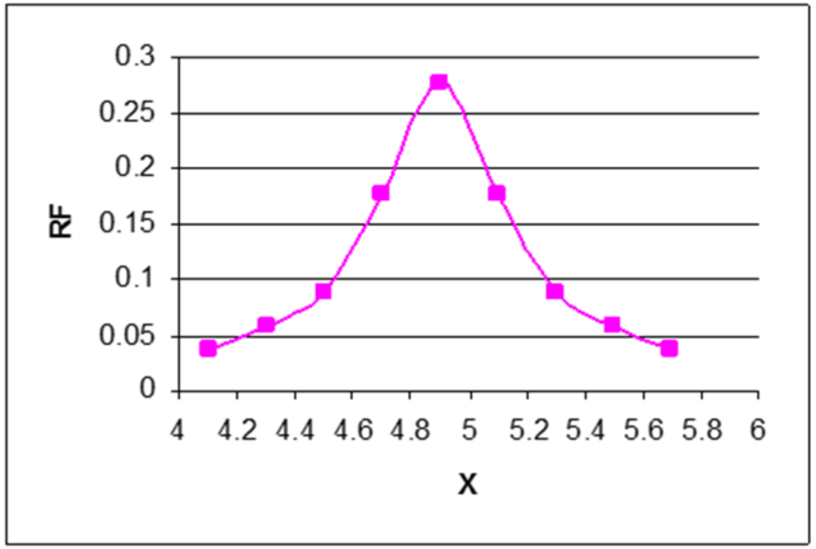


$$f(x) = x^2$$



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

- Many continuous variables have distributions that are bell-shaped and are called **approximately normally distributed variables**.
- The theoretical curve, called the **bell curve** or the **Gaussian distribution**, can be used to study many variables that are not normally distributed but are approximately normal.



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

$$where$$

$$e \approx 2.718$$

$$\pi \approx 3.14$$

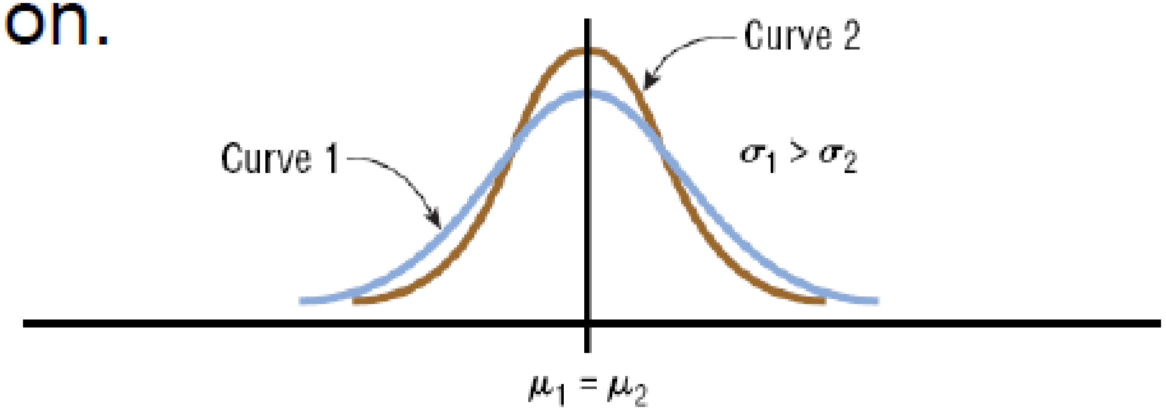
$$\mu = population\ mean$$

$$\sigma = population\ standard\ deviation$$

Normal distribution is determined by the fixed values of the mean μ and standard deviation σ.

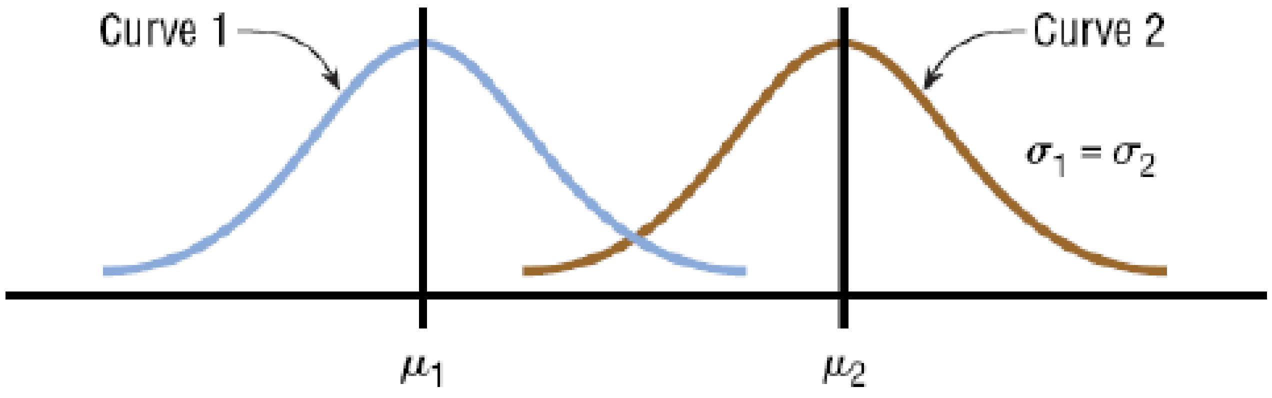
Normal Distributions

- The shape and position of the normal distribution curve depend on two parameters, the **mean** and the **standard deviation**.
- Each normally distributed variable has its own normal distribution curve, which depends on the values of the variable's mean and standard deviation.

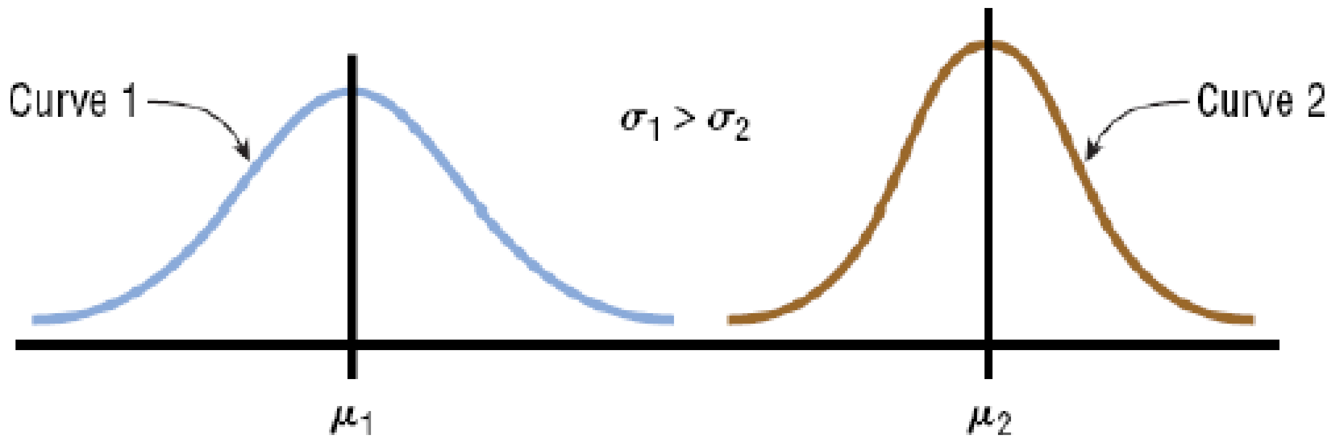


(a) Same means but different standard deviations

Normal Distributions

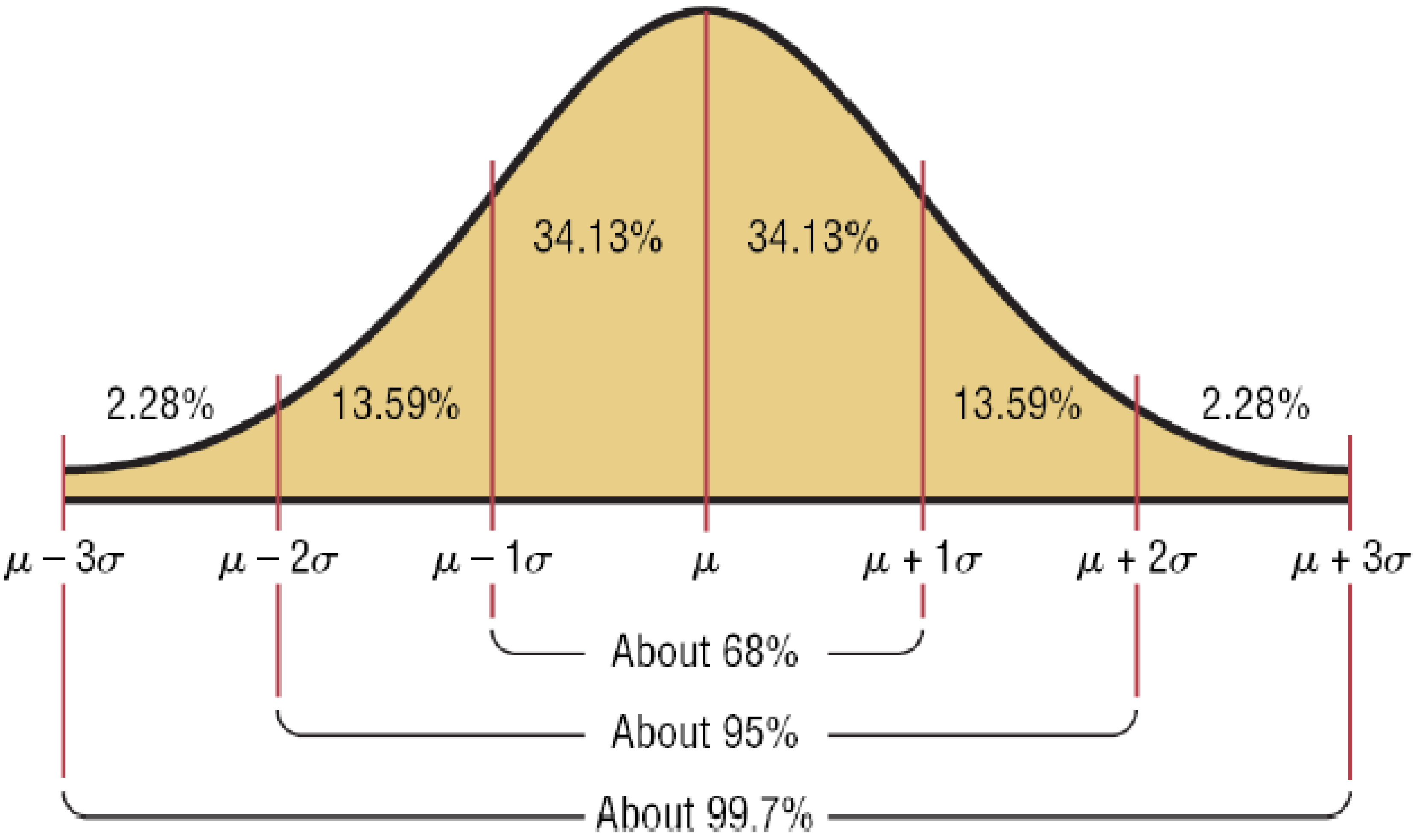


(b) Different means but same standard deviations



(c) Different means and different standard deviations

Normal Distribution Properties



Normal Distribution Properties

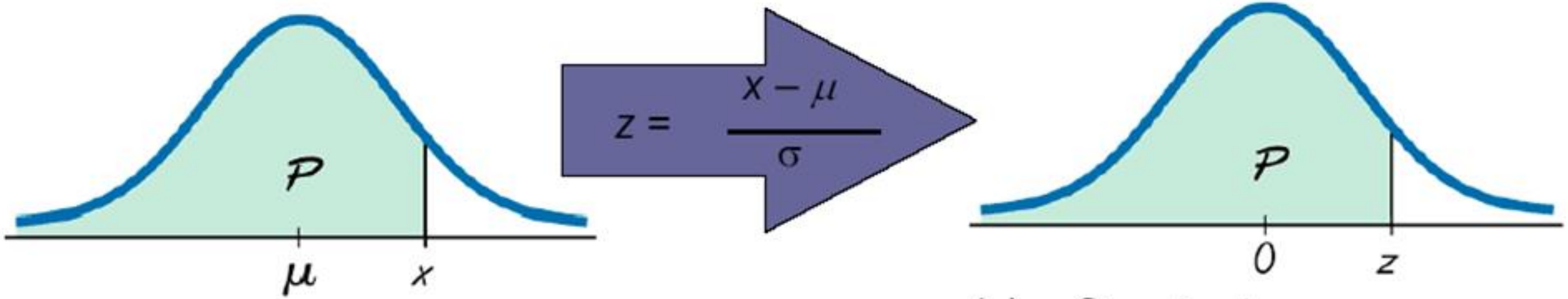
- The normal distribution curve is bell-shaped.
- The mean, median, and mode are equal and located at the center of the distribution.
- The normal distribution curve is **unimodal** (i.e., it has only one mode).
- The curve is symmetrical about the mean, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.

- The curve is continuous—i.e., there are no gaps or holes. For each value of *X*, here is a corresponding value of *Y*.
- The curve never touches the *x* axis. Theoretically, no matter how far in either direction the curve extends, it never meets the *x* axis—but it gets increasingly closer.

Standard Normal Distribution

- Since each normally distributed variable has its own mean and standard deviation, the shape and location of these curves will vary. In practical applications, one would have to have a table of areas under the curve for each variable. To simplify this, statisticians use the standard normal distribution.
- The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1.

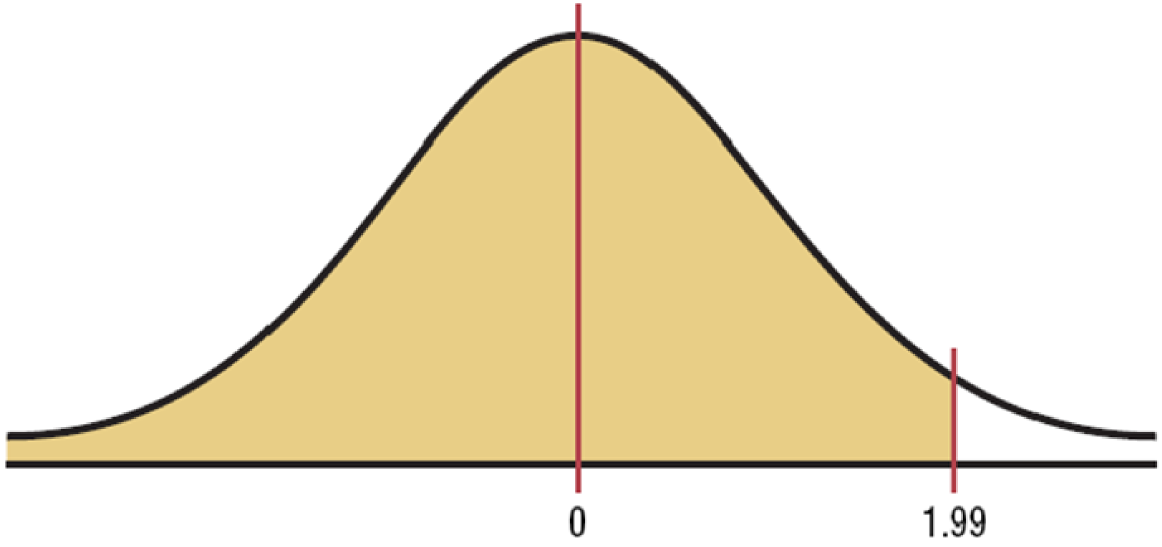
Converting to a Standard Normal Distribution



(a) Nonstandard Normal Distribution

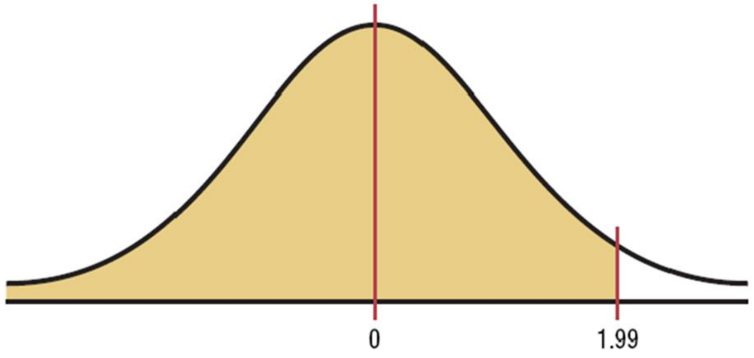
(b) Standard Normal Distribution

Find the area to the left of z = 1.99.



The value in the 1.9 row and the .09 column of Table E is .9767. The area is .9767.

Find the area to the left of z = 1.99.



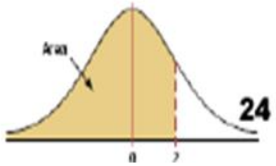
The value in the 1.9 row and the .09 column of Table E is .9767. The area is .9767.

Table E *(continued)*

Cumulative Standard Normal Distribution

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
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952

Gives the cumulative area from the left up to a vertical line above a specific value of z.



Find the area to right of z = -1.16.

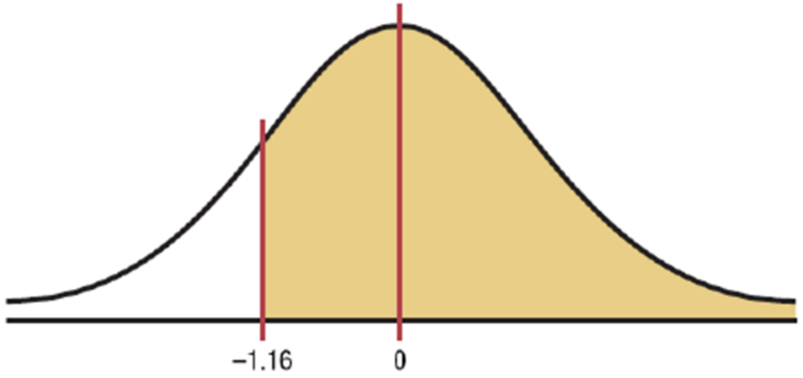
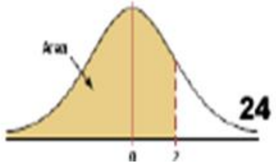


Table E *(continued)*

Cumulative Standard Normal Distribution

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
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952

Gives the cumulative area from the left up to a vertical line above a specific value of z.



Find the probability: $P(0 < z < 2.32)$

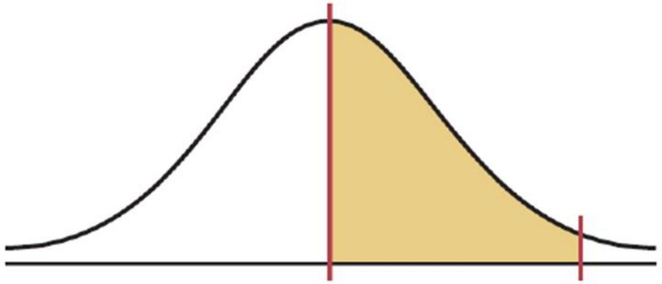
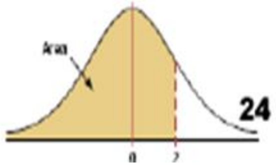


Table E *(continued)*

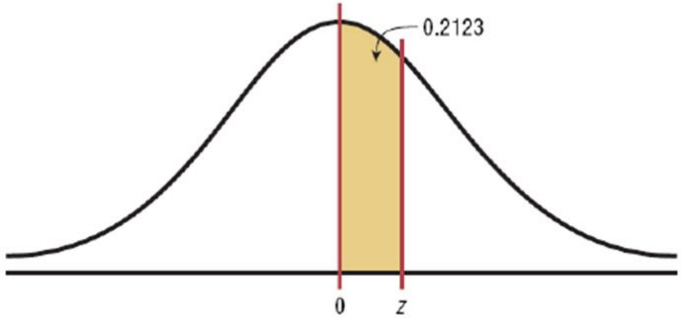
Cumulative Standard Normal Distribution

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
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952

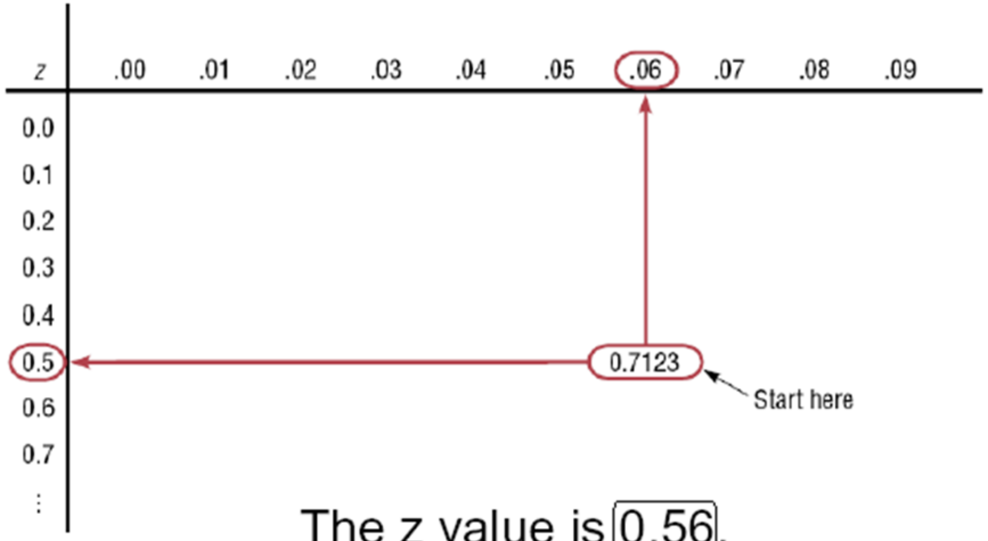
Gives the cumulative area from the left up to a vertical line above a specific value of z.



Find the z value such that the area under the standard normal distribution curve between 0 and the z value is 0.2123.



Add .5000 to .2123 to get the cumulative area of .7123. Then look for that value inside Table E.



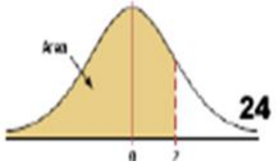
The z value is $\boxed{0.56}$.

Table E (*continued*)

Cumulative Standard Normal Distribution

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
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952

Gives the cumulative area from the left up to a vertical line above a specific value of z.



EX1

A firm that manufactures and bottles apple juice has a machine that automatically fills 16-ounce bottles. There is, however, some variation in the amount of liquid dispensed (in ounces) into each bottle by the machine. Over a long period of time, the average amount dispensed into the bottles was 16 ounces, but there is a standard deviation of 1 ounce in these measurements. If the amount filled per bottle can be assumed to be normally distributed, find the probability that the machine will dispense more than 17 ounces of liquid into any one bottle.

- Average fill μ is 16 oz.
- $\sigma = 1$ oz.
- Amount of fill is normally distributed
- $P(X > 17 \text{ oz}) = ?$

EX 2

Suppose that another machine, similar to the one of Ex 1, operates so that ounces of fill have a mean equal to the dial setting for "amount of liquid" but have a standard deviation of 1.2 ounces. Find the proper setting for the dial so that 17-ounce bottles will overflow only 5% of the time. Assume that the amounts dispensed have a normal distribution.

- $\sigma = 1.2$ oz.
- Amount of fill is normally distributed
- Have to set the μ so that 17 oz. will overflow and it is only 5%
- Find the μ that has to be set.

EX 3

The SAT and ACT college entrance exams are taken by thousands of students each year. The scores on the exam for any one year produce a histogram that looks verymuch like a normal curve. Thus, we can say that the scores are approximately normally distributed. In recent years, the SAT mathematics scores have averaged around 480 with a standard deviation of 100. The ACT mathematics scores have averaged around 18 with a standard deviation of 6.

a An engineering school sets 550 as the minimum SAT math score for new students. What percent of students would score less than 550 in a typical year? (This percentage is called the percentile score equivalent for 550.)

b What would the engineering school set as a comparable standard on the ACT math test?

c What is the probability that a randomly selected student will score over 700 on the SAT math test?

- Exam I – score ND with $\mu = 480$, $\sigma = 100$
- Exam II – score ND with $\mu = 18$, $\sigma = 6$

a. For a school admission, sets min. 550 score in Exam I. What % students would score less than 550 in a year?

b. What would be the cutoff mas for Exam II, if one needs to maintain the same standard for Exam II?

c. P (a randomly selected student scores more than 700 in Exam I)

EX 4

The yield force of a steel-reinforcing bar of a certain type is found to be normally distributed with a mean of 8,500 pounds and a standard deviation of 80 pounds. If three such bars are to be used on a certain project, find the probability that all three will have yield forces in excess of 8,700 pounds.

EX 5

A machine for filling cereal boxes has a standard deviation of 1 ounce of fill per box. What setting of the mean ounces of fill per box will allow 16-ounce boxes to overflow only 1% of the time? Assume that the ounces of fill per box are normally distributed.

The Central Limit Theorem

In addition to knowing how individual data values vary about the mean for a population, statisticians are interested in knowing how the means of samples of the same size taken from the same population vary about the population mean.

Distribution of Sample Means

- A ***sampling distribution of sample means*** is a distribution obtained by using the means computed from random samples of a specific size taken from a population.
- ***Sampling error*** is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

Properties of the Distribution of Sample Means

- The mean of the sample means will be the same as the population mean.
- The standard deviation of the sample means will be smaller than the standard deviation of the population, and will be equal to the population standard deviation divided by the square root of the sample size.

The Central Limit Theorem

- As the sample size *n* increases, the shape of the distribution of the sample means taken with replacement from a population with mean μ and standard deviation σ will approach a normal distribution.
- The mean of the sample means equals the population mean. $\mu_{\overline{X}} = \mu$.
- The standard deviation of the sample means is called the **standard error of the mean** (S.E.)

$$\sigma_{\overline{X}} = \sigma / \sqrt{n}.$$

The Central Limit Theorem

- The central limit theorem can be used to answer questions about sample means in the same manner that the normal distribution can be used to answer questions about individual values.
- A new formula must be used for the z values:

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

Students sometimes have difficulty deciding whether to use

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad \text{or} \qquad z = \frac{X - \mu}{\sigma}$$

The formula

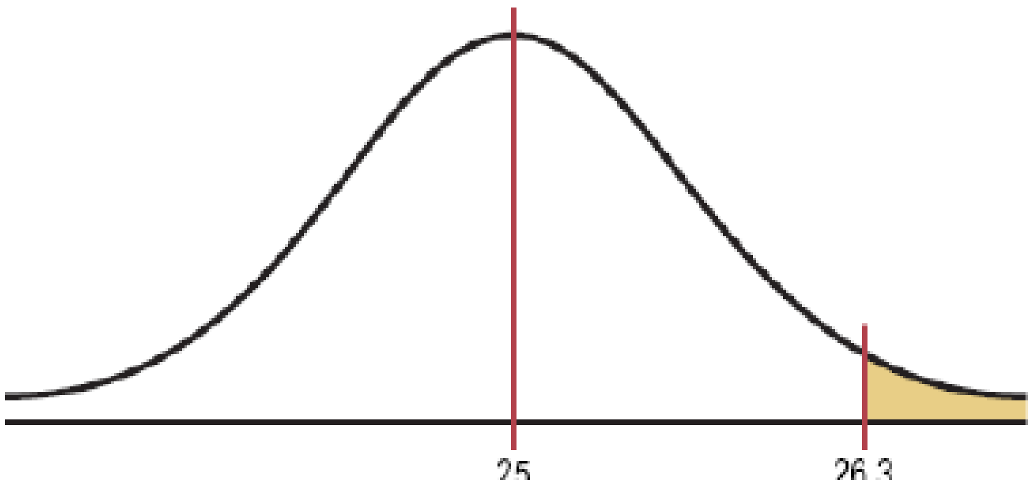
$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

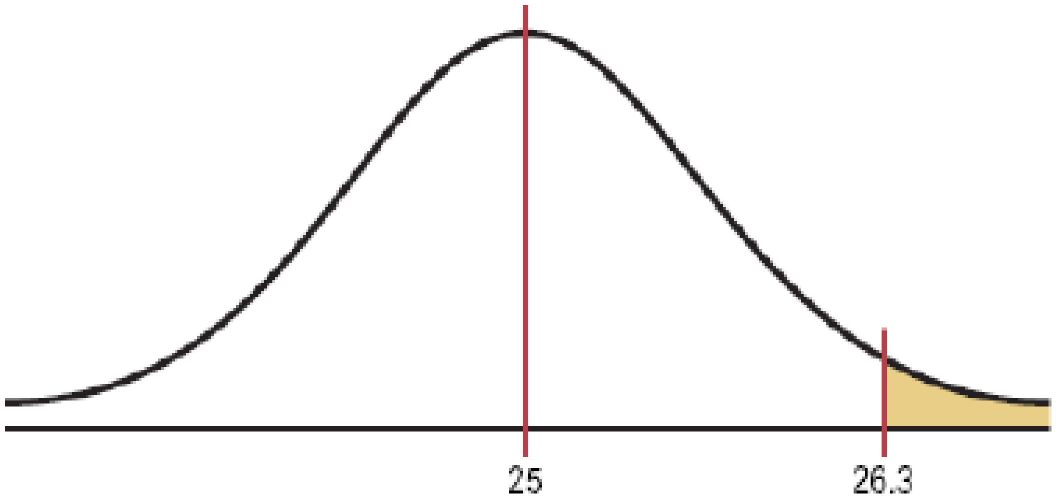
should be used to gain information about a sample mean, as shown in this section. The formula

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

is used to gain information about an individual data value obtained from the population. Notice that the first formula contains $\bar{X}$, the symbol for the sample mean, while the second formula contains X, the symbol for an individual data value.

A. C. Neilsen reported that children between the ages of 2 and 5 watch an <u>average</u> of 25 hours of television per week. Assume the variable is normally distributed and the standard deviation is 3 hours. If <u>20</u> children between the ages of 2 and 5 are randomly selected, find the probability that the <u>mean</u> of the number of hours they watch television will be greater than 26.3 hours.





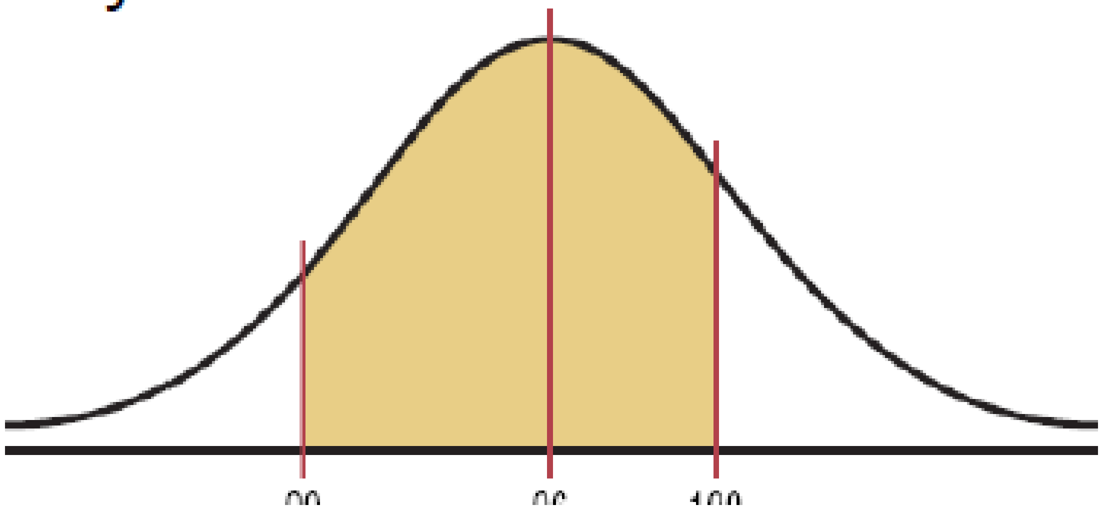
Since we are calculating probability for a sample mean, we need the Central Limit Theorem formula

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{26.3 - 25}{3/\sqrt{20}} = 1.94$$

The area is 1.0000 – 0.9738 = 0.0262. The probability of obtaining a sample mean larger than 26.3 hours is 2.62%.

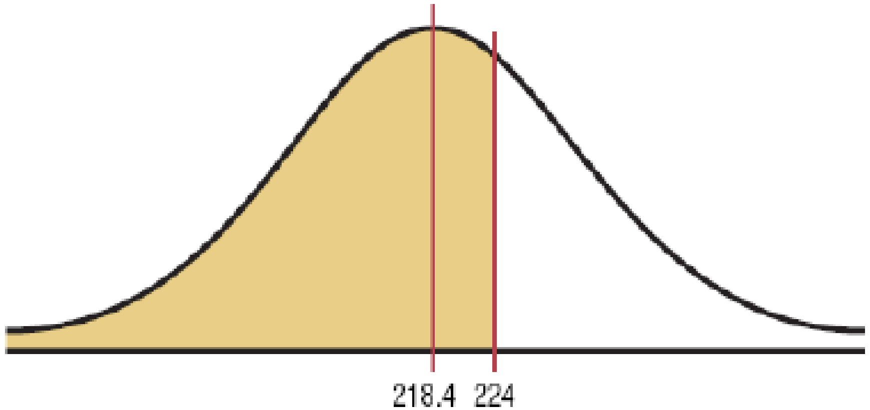
The average age of a vehicle registered in the United States is 8 years, or 96 months. Assume the standard deviation is 16 months. If a random sample of 36 vehicles is selected, find the probability that the mean of their age is between 90 and 100 months.

Since the sample is 30 or larger, the normality assumption is not necessary.



The average number of pounds of meat that a person consumes per year is 218.4 pounds. Assume that the standard deviation is 25 pounds and the distribution is approximately normal.

a. Find the probability that a person selected at random consumes less than 224 pounds per year.



Distribution of individual data values for the population

b. If a sample of 40 individuals is selected, find the probability that mean of the sample will be less than 224 pounds per year.

https://youtu.be/KxE2XBPcgeM