Probability and Statistics (IT302) Class No. 27
19th October 2020 Monday 09:45 AM - 10:15 AM

Probability and Statistics (IT302) Class No. 28
20th October 2020 Tuesday 10:30 AM - 11:00 AM

Continuous Uniform Distribution

- One of the simplest continuous distributions in all of statistics is the **Continuous Uniform Distribution**.

- This distribution is characterized by a density function that is “flat,” and thus the probability is uniform in a closed interval, say [A, B].

- Although applications of the **Continuous Uniform Distribution** are not as abundant as those for other distributions, it is appropriate for the novice to begin this introduction to **Continuous Distributions with the Uniform Distribution.**

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Uniform Distribution

The **density function** of the Continuous Uniform random variable *X* on the interval [*A*, *B*] is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B, \\ 0, & \text{elsewhere.} \end{cases}$$

The **density function** forms a rectangle with base B−A and constant height 1/B−A. As a result, the uniform distribution is often called **the rectangular distribution**.

Note, however, that the interval **may not always be closed: [A,B]. It can be (A,B) as well**.

The density function for a uniform random variable on the interval [1, 3] is shown in Figure.

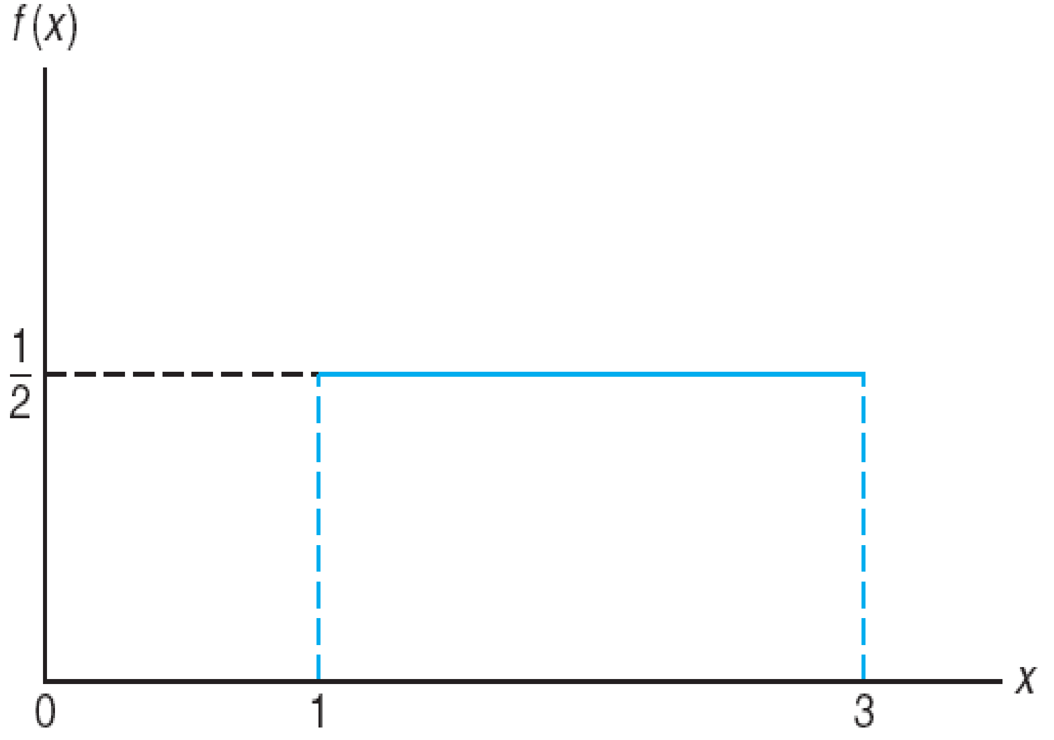


Figure: The density function for a random variable on the interval [1, 3].

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Uniform Distribution Contd.

Probabilities are simple to calculate for the uniform distribution because of the simple nature of the density function. However, **note that the application of this distribution is based on the assumption that the probability of falling in an interval of fixed length within [A, B] is constant.**

Example 6.1: Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conferences occur quite often. In fact, it can be assumed that the length *X* of a conference has a uniform distribution on the interval [0, 4].

(a) What is the probability density function?

(b) What is the probability that any given conference lasts at least 3 hours?

Solution: (a) The appropriate density function for the uniformly distributed random variable X in this situation is

$$f(x) = \begin{cases} \frac{1}{4}, & 0 \le x \le 4, \\ 0, & \text{elsewhere.} \end{cases}$$

(b) $P[X \ge 3] = \int_3^4 \frac{1}{4}\, dx = \frac{1}{4}.$

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

The Mean and Variance of the Uniform Distribution

Theorem 6.1: The mean and variance of the uniform distribution are

$$\mu = \frac{A+B}{2} \text{ and } \sigma^2 = \frac{(B-A)^2}{12}.$$

Normal Distribution

The most important continuous probability distribution in the entire field of statistics is the normal distribution. **Its graph, called the normal curve, is the bell-shaped curve of Figure, which approximately describes many phenomena that occur in nature, industry, and research.**

Example: physical measurements in areas such as meteorological experiments, rainfall studies, and measurements of manufactured parts are often more than adequately explained with a normal distribution.

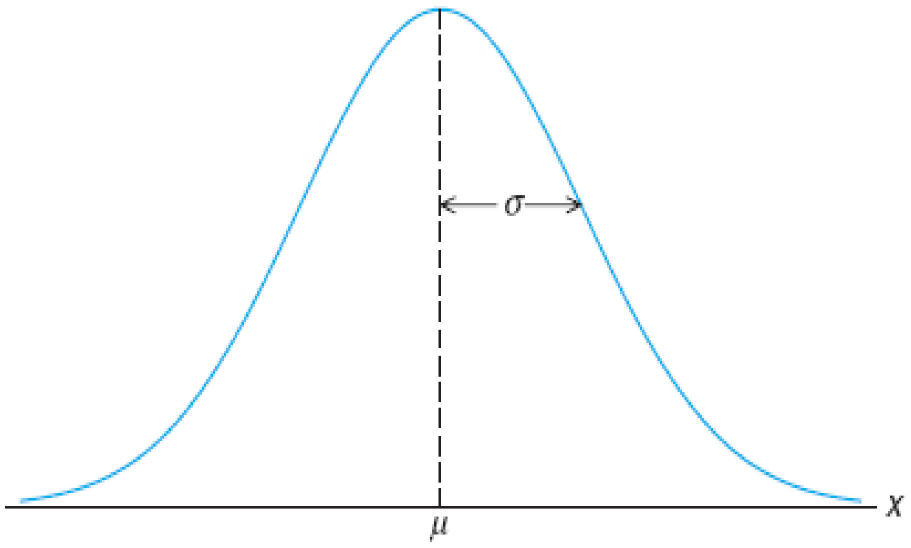


Figure: The normal curve.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Normal Distribution Contd.

- In 1733, Abraham DeMoivre developed the mathematical equation of the normal curve. It provided a basis from which much of the theory of inductive statistics is founded.

- **The normal distribution is often referred to as the Gaussian distribution**, in honor of Karl Friedrich Gauss (1777–1855), who also derived its equation from a study of errors in repeated measurements of the same quantity.

- **A continuous random variable X having the bell-shaped distribution of Figure is called a normal random variable**.

- The mathematical equation for the probability distribution of the normal variable depends on the two parameters μ and σ, its mean and standard deviation, respectively. Hence, values of the density of X denote by $n(x; \mu, \sigma)$.

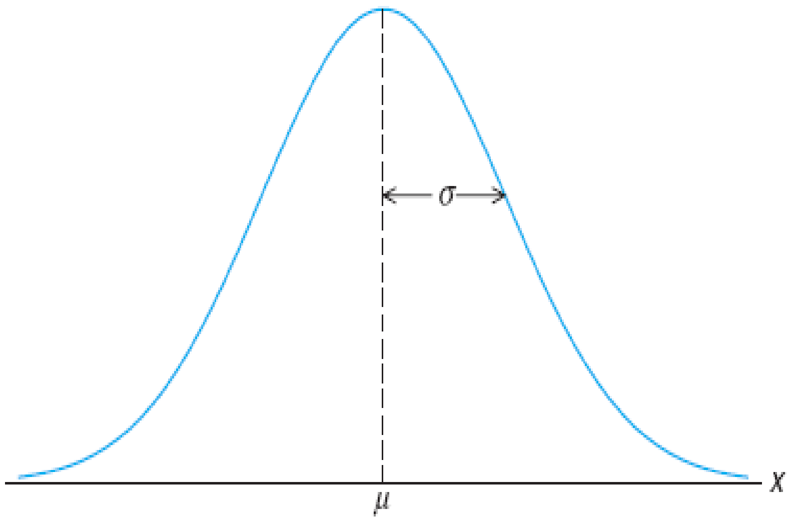


Figure: The normal curve.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Normal Distribution Contd.

The density of the normal random variable X, with mean μ and variance σ^2, is

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < \infty,$$

where $\pi = 3.14159\ldots$ and $e = 2.71828\ldots$.
Once μ and σ are specified, the normal curve is completely determined.

Example: μ = 50 and σ = 5, then the ordinates n(x; 50, 5) can be computed for various values of x and the curve drawn.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Normal Distribution Contd.

Below Figure shows two normal curves having the **same standard deviation but different means.** The two curves are identical in form but are centered at different positions along the horizontal axis.

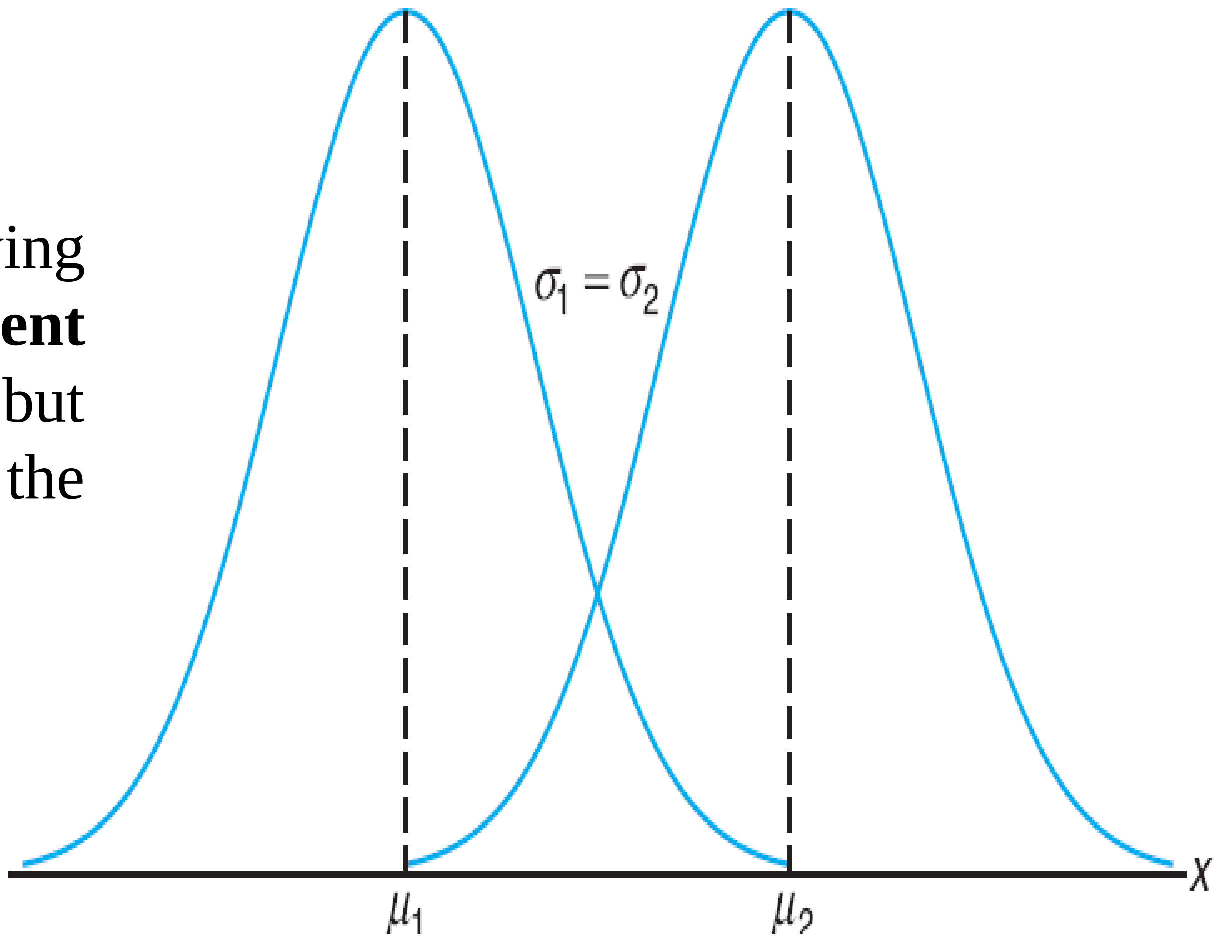


Figure : Normal curves with μ1 < μ2 and σ1 = σ2.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Normal Distribution Contd.

- Below Figure shows two normal curves with the **same mean but different standard deviations.**

- In this two curves are centered at exactly the same position on the horizontal axis, but the curve with the larger standard deviation is lower and spreads out farther.

- Remember that the area under a probability curve must be equal to 1, and therefore the more variable the set of observations, the lower and wider the corresponding curve will be

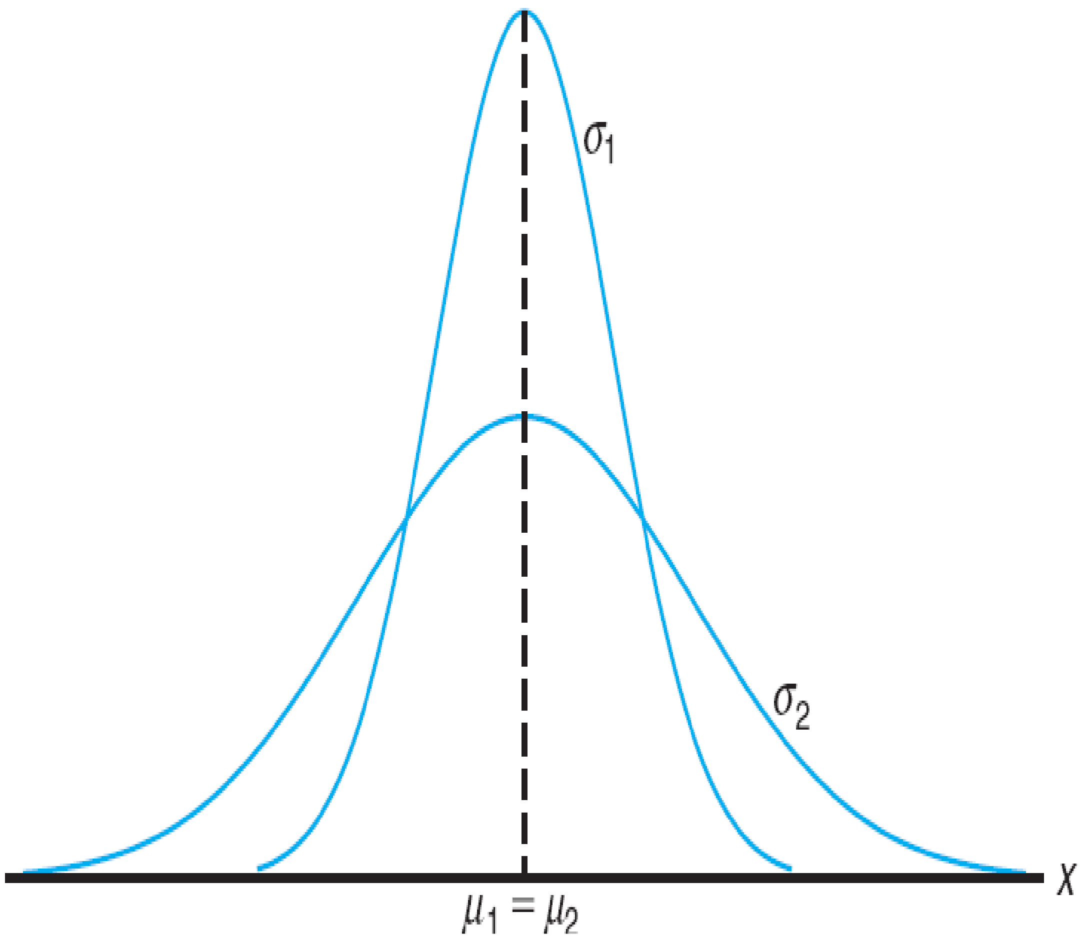


Figure: Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Normal Distribution Contd.

Below Figure shows **two normal curves having different means and different standard deviations**. Clearly, they are centered at different positions on the horizontal axis and their shapes reflect the two different values of σ.

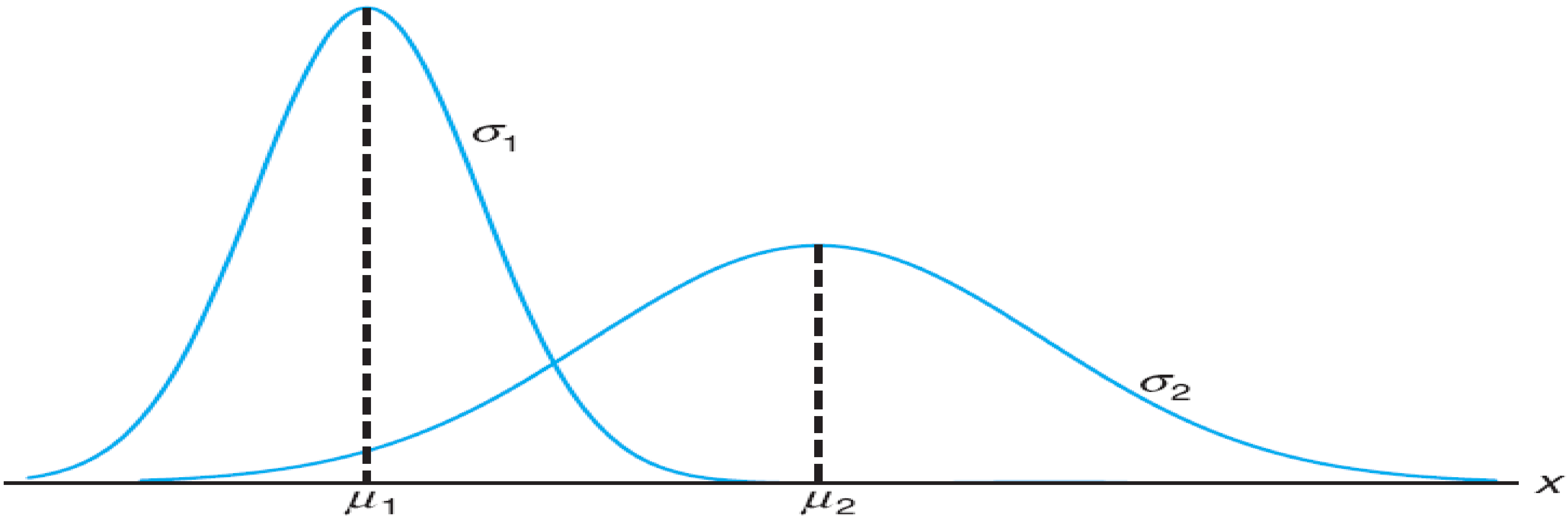


Figure : Normal curves with μ1 < μ2 and σ1 < σ2.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Normal Distribution Contd.

Based on inspection of Figures show in previous slides and examination of the first and second derivatives of $n(x; \mu, \sigma)$, properties of the normal curve are:

1. The mode, which is the point on the horizontal axis where the curve is a maximum, occurs at $x = \mu$.
2. The curve is symmetric about a vertical axis through the mean μ.
3. The curve has its points of inflection at $x = \mu \pm \sigma$; it is concave downward if $\mu - \sigma < X < \mu + \sigma$ and is concave upward otherwise.
4. The normal curve approaches the horizontal axis asymptotically as proceed in either direction away from the mean.
5. The total area under the curve and above the horizontal axis is equal to 1.

Theorem 6.2: **The mean and variance of $n(x; \mu, \sigma)$ are μ and σ^2, respectively. Hence, the standard deviation is σ.**

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Areas under the Normal Curve

The curve of any continuous probability distribution or density function is constructed so that the area under the curve bounded by the two ordinates $x = x_1$ and $x = x_2$ equals the probability that the random variable X assumes a value between $x = x_1$ and $x = x_2$. Thus, for the normal curve in below Figure is represented by the area of the shaded region.

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x;\mu,\sigma)\,dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx$$

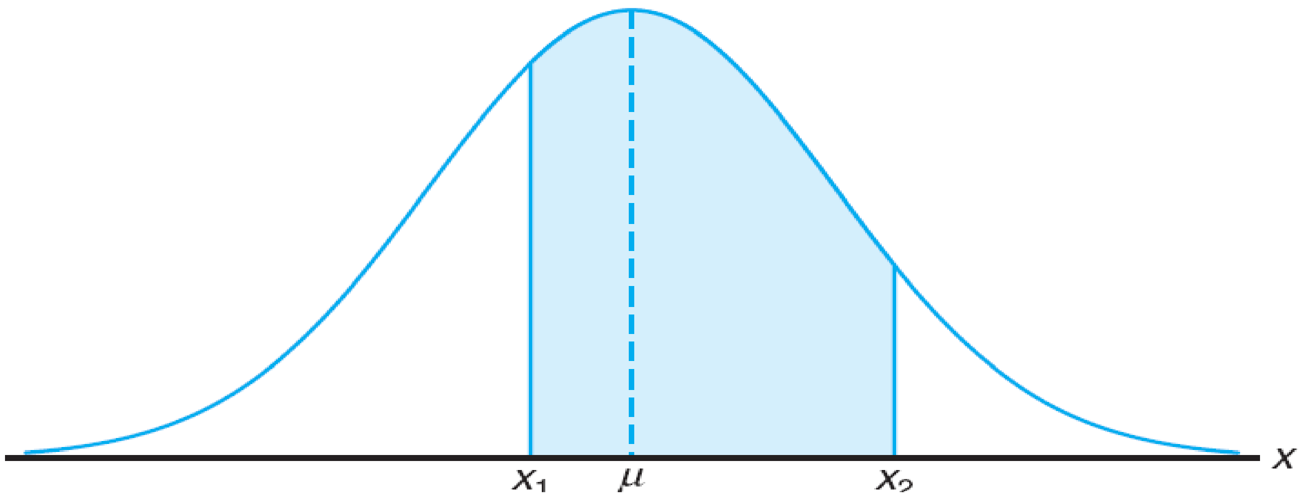


Figure: $P(x_1 < X < x_2)$ = area of the shaded region.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Areas under the Normal Curve Contd.

- The area under the curve between any two ordinates must then also depend on the values μ and σ.

- This is evident in below Figure, where shaded regions corresponding to $P(x_1 < X < x_2)$ for two curves with different means and variances.

- $P(x_1 < X < x_2)$, where X is the random variable describing distribution A, is indicated by the shaded area below the curve of A. If X is the random variable describing distribution B, then $P(x_1 < X < x_2)$ is given by the entire shaded region.

- Obviously, the two shaded regions are different in size; therefore, the probability associated with each distribution will be different for the two given values of X.

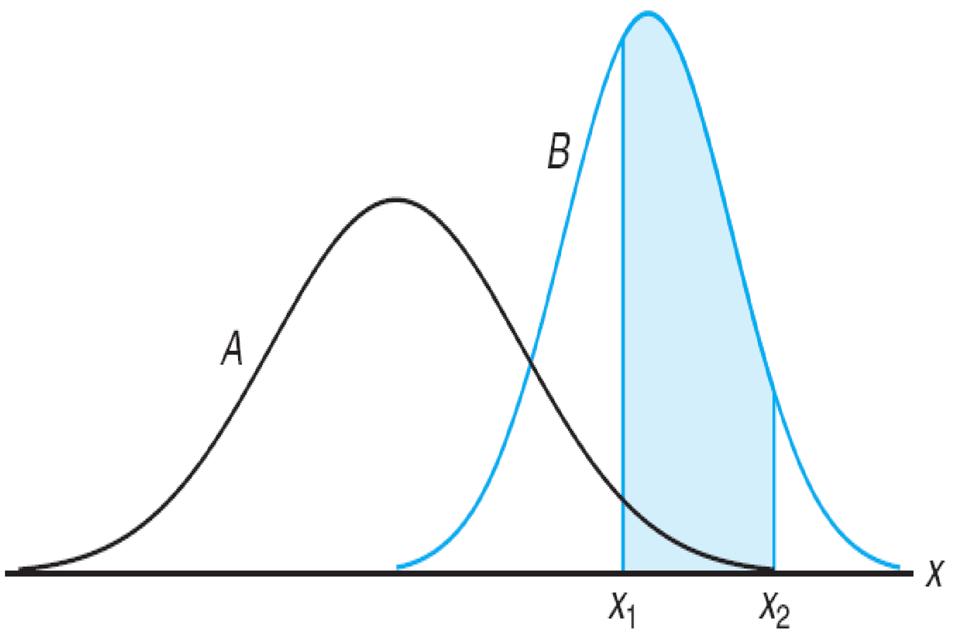


Figure : $P(x_1 < X < x_2)$ for different normal curves.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Areas under the Normal Curve Contd.

Transform all the observations of any normal random variable X into a new set of observations of a normal random variable Z with mean 0 and variance 1. This can be done by means of the transformation

$Z = X - \mu/\sigma$

.

Whenever X assumes a value x, the corresponding value of Z is given by $z = (x - \mu)/\sigma$. Therefore, if X falls between the values $x = x_1$ and $x = x_2$, the random variable Z will fall between the corresponding values $z_1 = (x_1 - \mu)/\sigma$ and $z_2 = (x_2 - \mu)/\sigma$. Consequently, we may write

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}\sigma}\int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx = \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-\frac{1}{2}z^2}\,dz$$

$$= \int_{z_1}^{z_2} n(z;0,1)\,dz = P(z_1 < Z < z_2),$$

where *Z* is seen to be a normal random variable with mean 0 and variance 1.

Standard Normal Distribution

Definition 6.1: The distribution of a normal random variable with **mean 0 and variance 1 is called a standard normal distribution.**

The original and transformed distributions are illustrated in Figure. Since all the values of X falling between x_1 and x_2 have corresponding z values between z_1 and z_2, the area under the X-curve between the ordinates $x = x_1$ and $x = x_2$ in Figure equals the area under the Z-curve between the transformed ordinates $z = z_1$ and $z = z_2$.

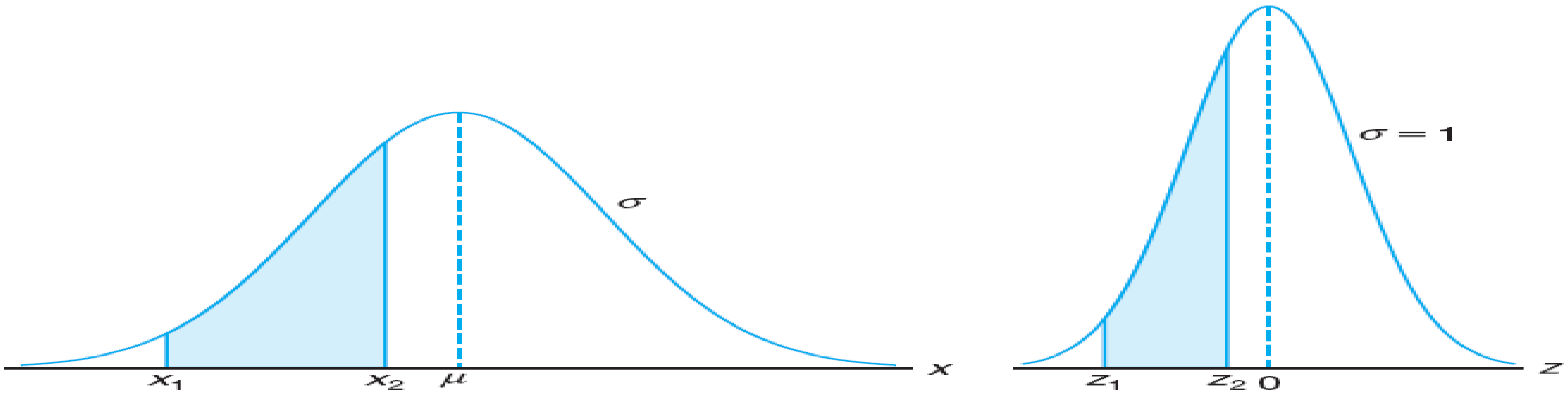
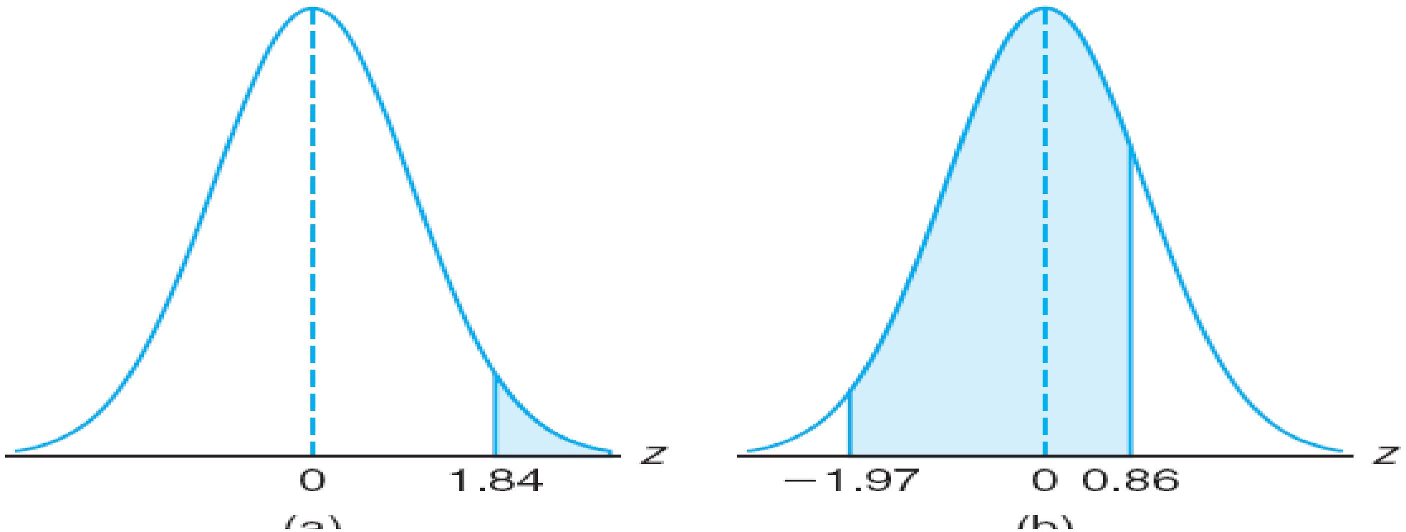


Figure : The original and transformed normal distributions

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Example 6.2

Given a standard normal distribution, find the area under the curve that lies
(a) to the right of $z = 1.84$ and
(b) between $z = -1.97$ and $z = 0.86$.



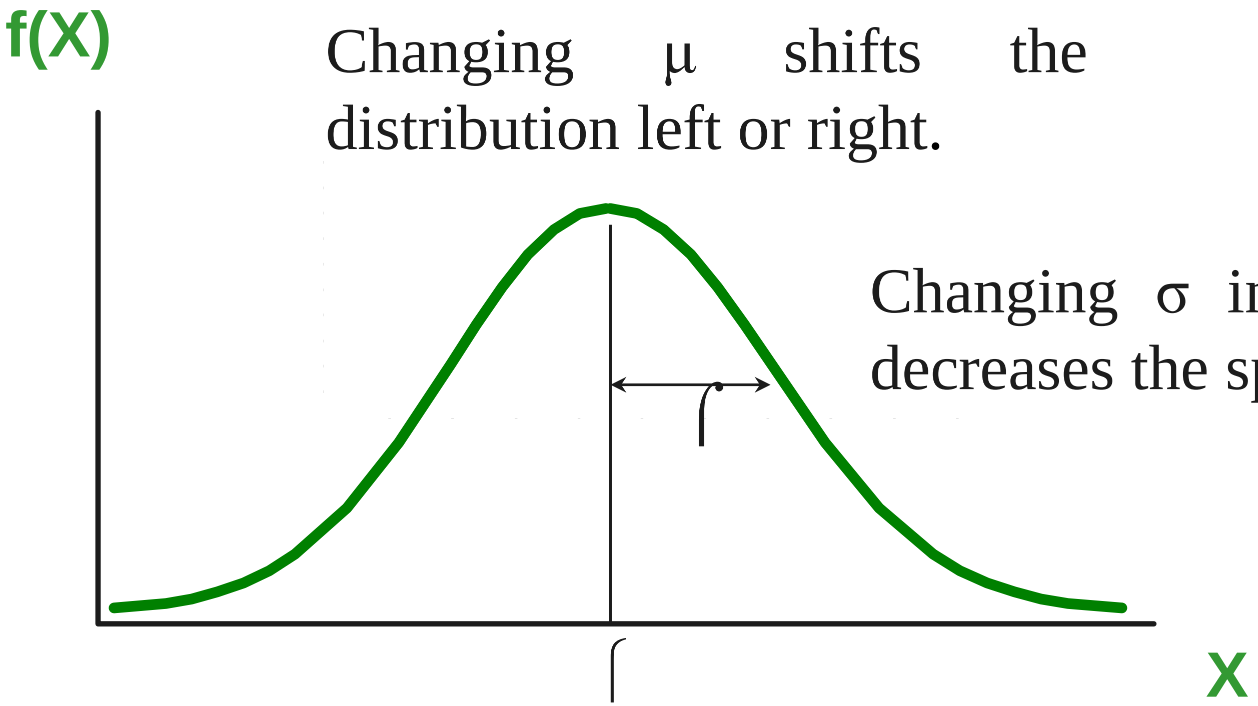
Solution : See above Figure for the specific areas.

a) The area in Figure to the right of $z = 1.84$ is equal to 1 minus the area in Table A.3 to the left of $z = 1.84$, namely, $1 - 0.9671 = 0.0329$.
b) The area in Figure between $z = -1.97$ and $z = 0.86$ is equal to the area to the left of $z = 0.86$ minus the area to the left of $z = -1.97$. From Table A.3 we find the desired area to be $0.8051 - 0.0244 = 0.7807$.

Source : Probability & Statistics for Engineers & Scientists by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, 9th Edition.

Additional Material

The Normal Distribution



Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

The Normal Distribution: as mathematical function (pdf)

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

<u>Note constants:</u>

π=3.14159

e=2.71828

This is a bell shaped curve with different centers and spreads depending on μ and σ

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

The Normal PDF

It's a probability function, so no matter what the values of μ and σ, must integrate to 1!

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Normal distribution is defined by its mean and standard dev.

E(X)=μ = $\int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$

Var(X)=σ^2 = $\int_{-\infty}^{+\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx) - \mu^2$

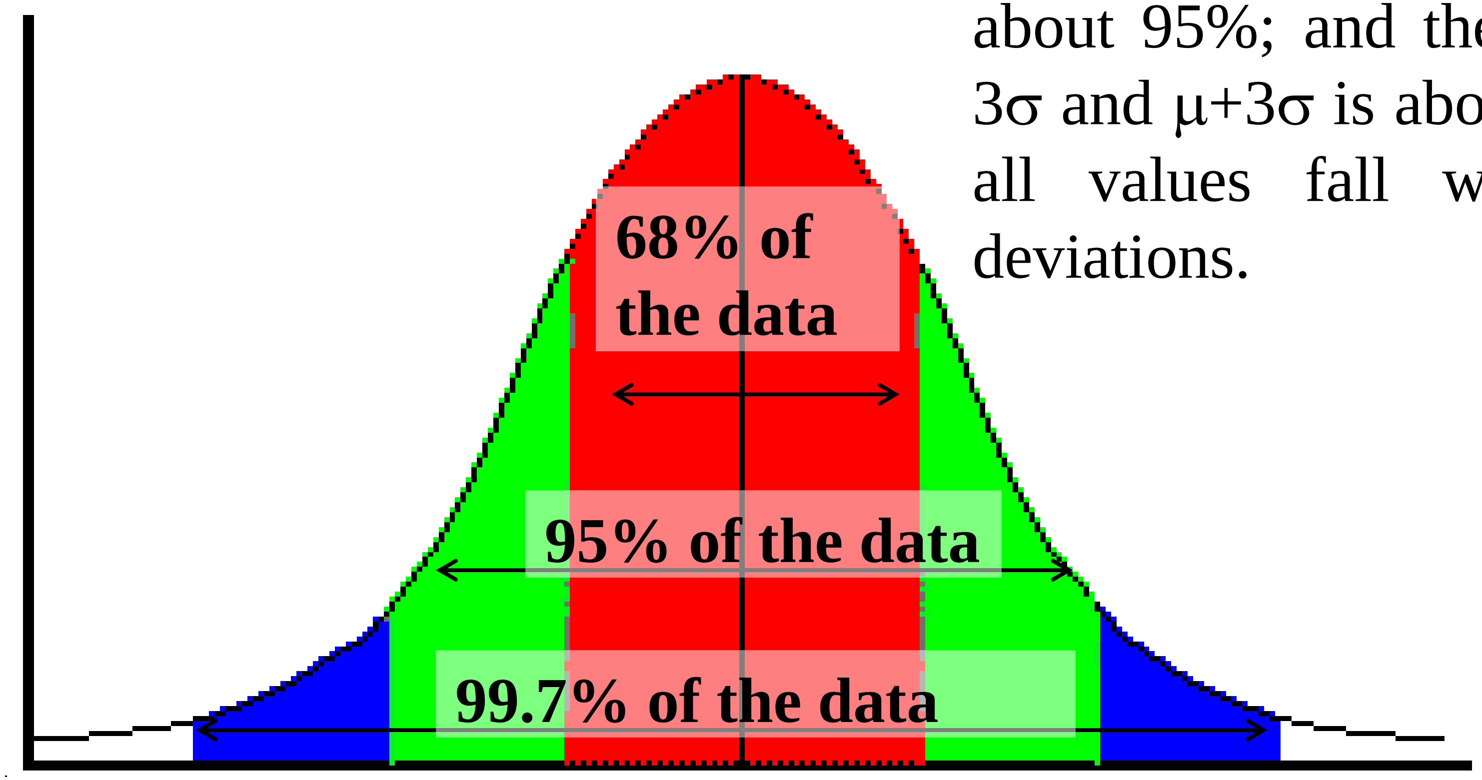
Standard Deviation(X)=σ

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

**The beauty of the normal curve:

No matter what μ and σ are, the area between $\mu-\sigma$ and $\mu+\sigma$ is about 68%; the area between $\mu-2\sigma$ and $\mu+2\sigma$ is about 95%; and the area between $\mu-3\sigma$ and $\mu+3\sigma$ is about 99.7%. Almost all values fall within 3 standard deviations.

68-95-99.7 Rule



Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

68-95-99.7 Rule

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .68$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .95$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .997$$

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

How Good is Rule for Real Data?

Check some example data:

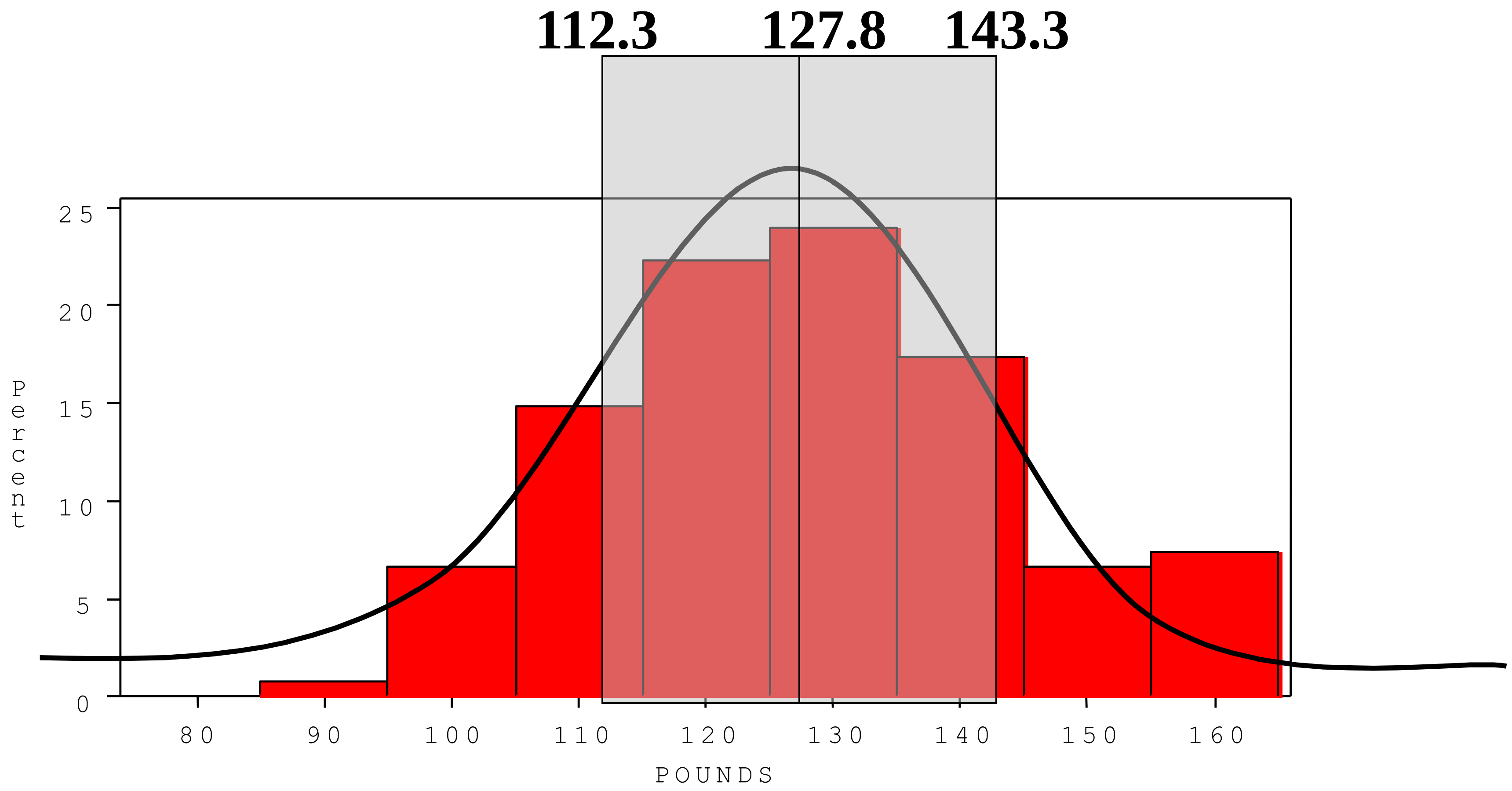
The mean of the weight of the women = 127.8

The standard deviation (SD) = 15.5

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

68% of 120 = .68x120 = ~ 82 runners

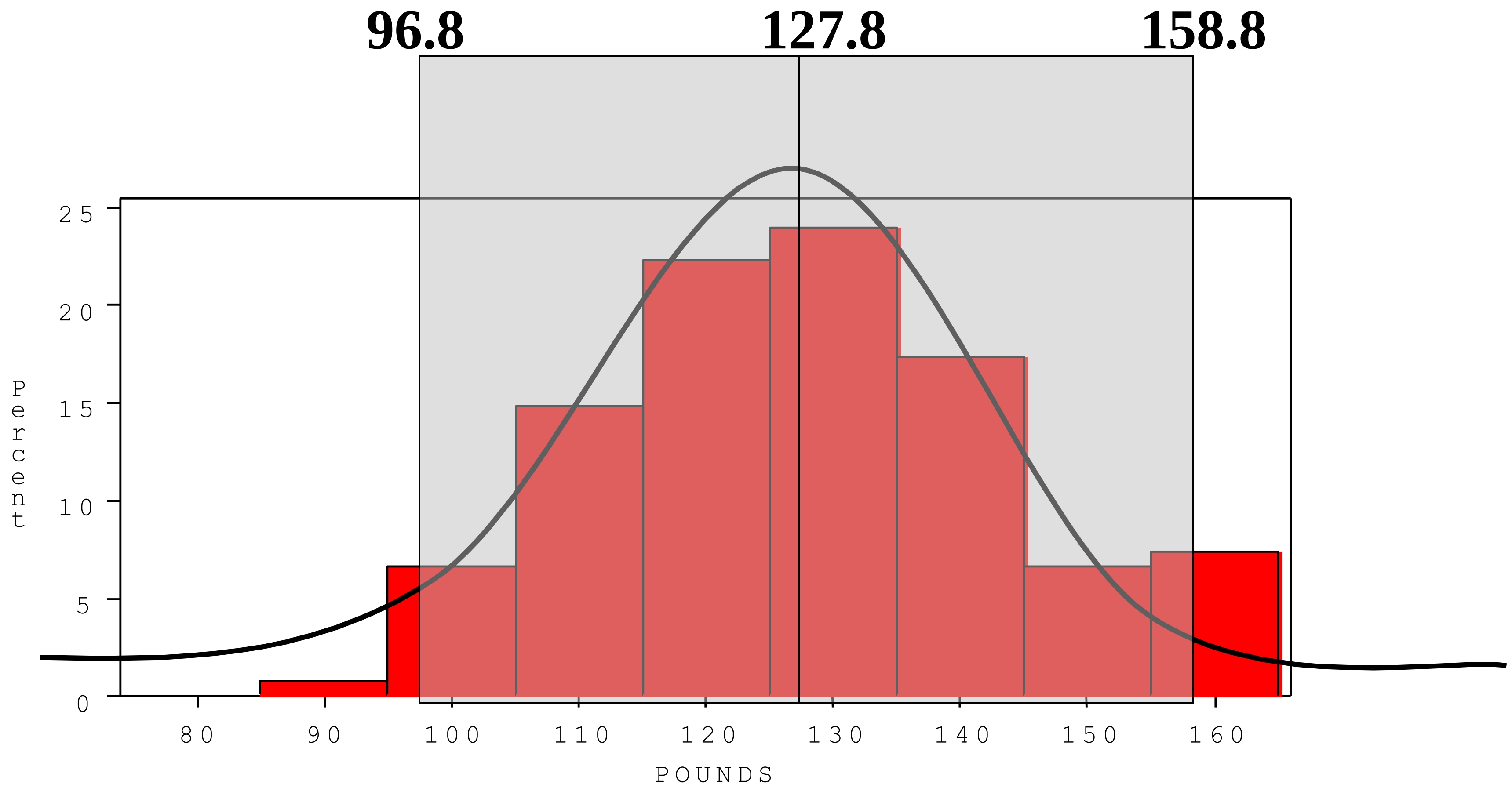
In fact, 79 runners fall within 1-SD (15.5 lbs) of the mean.



Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

95% of 120 = .95 x 120 = ~ 114 runners

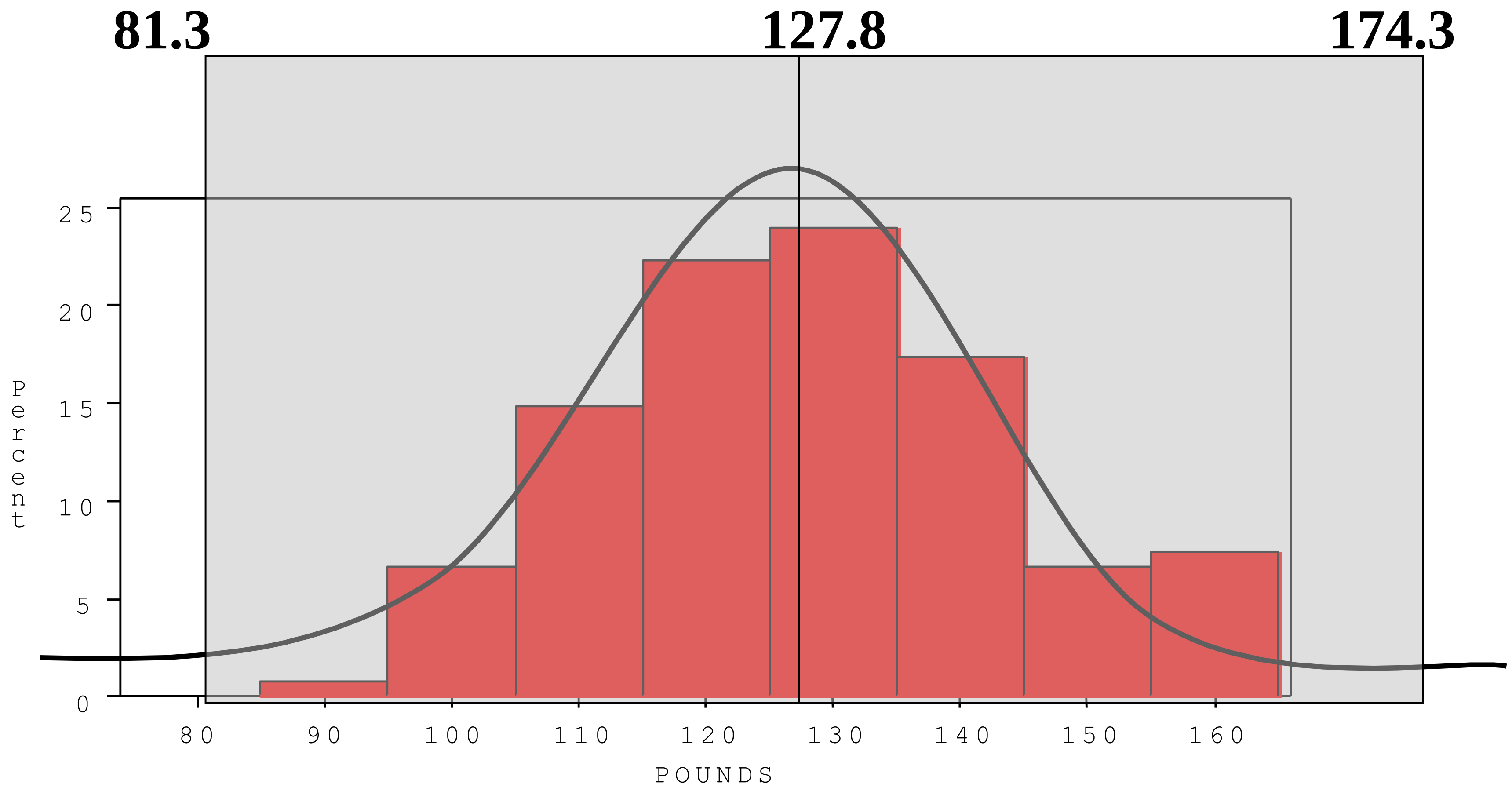
In fact, 115 runners fall within 2-SD's of the mean.



Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

99.7% of 120 = .997 x 120 = 119.6 runners

In fact, all 120 runners fall within 3-SD's of the mean.



Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Example

- Suppose SAT scores roughly follows a normal distribution in the U.S. population of college-bound students (with range restricted to 200-800), and the average math SAT is 500 with a standard deviation of 50, then:
  - 68% of students will have scores between 450 and 550
  - 95% will be between 400 and 600
  - 99.7% will be between 350 and 650

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Example

- BUT…
- What if you wanted to know the math SAT score corresponding to the 90th percentile (=90% of students are lower)?

P(X≤Q) = .90 →

$$\int_{200}^{Q} \frac{1}{(50)\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-500}{50})^2} dx = .90$$

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

The Standard Normal (Z): "Universal Currency"

The formula for the standardized normal probability density function is

$$p(Z) = \frac{1}{(1)\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{Z-0}{1})^2} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(Z)^2}$$

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

The Standard Normal Distribution (Z)

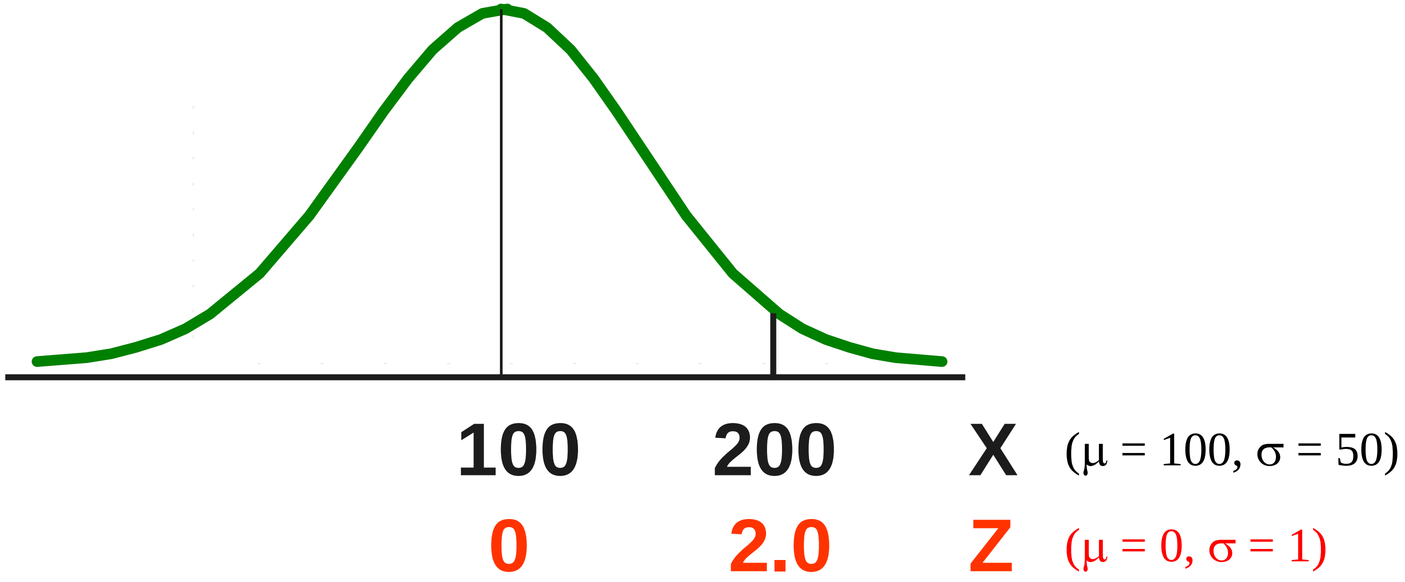
All normal distributions can be converted into the standard normal curve by subtracting the mean and dividing by the standard deviation:

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

Somebody calculated all the integrals for the standard normal and put them in a Table! So we never have to integrate! Even better, computers now do all the integration.

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Comparing X and Z units



Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Example

- For example: What's the probability of getting a math SAT score of 575 or less, μ=500 and σ=50?

$$Z = \frac{575 - 500}{50} = 1.5$$

- i.e., A score of 575 is 1.5 standard deviations above the mean

$$\therefore P(X \le 575) = \int_{200}^{575} \frac{1}{(50)\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-500}{50})^2} dx \longrightarrow \int_{-\infty}^{1.5} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}Z^2} dz$$

But to look up Z= 1.5 in standard normal chart (or enter into SAS)→ no problem! = .9332

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Practice Problem

If birth weights in a population are normally distributed with a mean of 109 oz and a standard deviation of 13 oz,

a) What is the chance of obtaining a birth weight of 141 oz *or heavier* when sampling birth records at random?

b) What is the chance of obtaining a birth weight of 120 *or lighter*?

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Answer

a. What is the chance of obtaining a birth weight of 141 oz *or heavier* when sampling birth records at random?

$$\boxed{Z = \frac{141-109}{13} = 2.46}$$

b. From the chart or SAS □ Z of 2.46 corresponds to a right tail (greater than) area of: $P(Z \geq 2.46) = 1-(.9931) = .0069$ or .69 %

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Answer

b. What is the chance of obtaining a birth weight of 120 *or lighter*?

$$\boxed{Z = \frac{120-109}{13} = .85}$$

From the chart or SAS → Z of .85 corresponds to a left tail area of:
P(Z≤.85) = .8023= 80.23%

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Looking up Probabilities in the Standard Normal Table

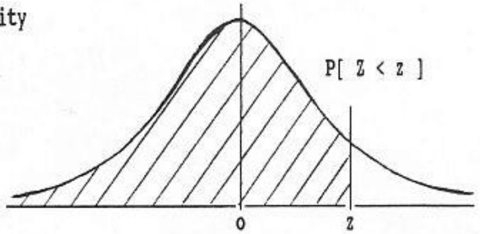
STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z

i.e.

$$P[\,Z < z\,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



What is the area to the left of Z=1.51 in a standard normal curve?

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	0.5159	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
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
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	0.8804	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
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

Z=1.51

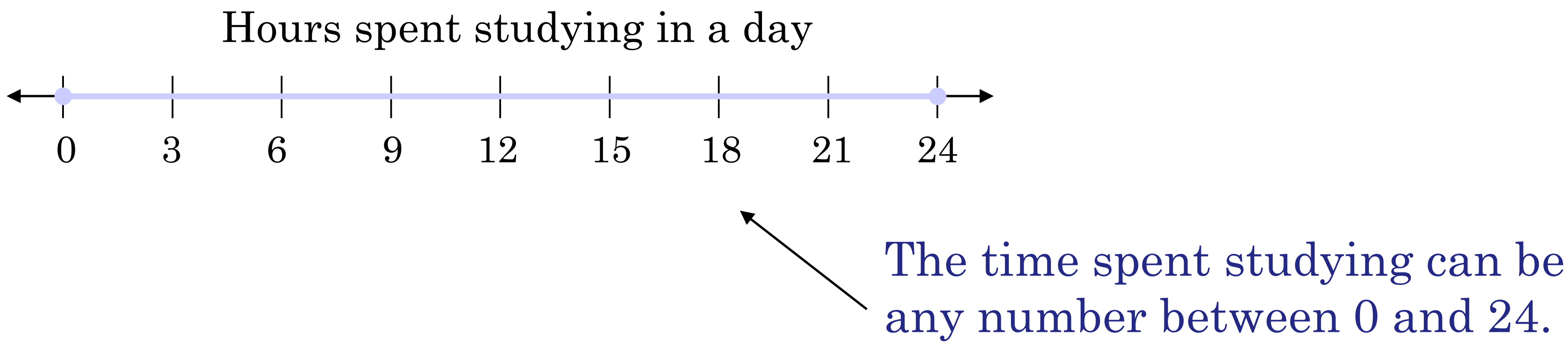
Z=1.51

Area is 93.45%

Source: http://web.stanford.edu/~kcobb/hrp259/lecture6

Properties of Normal Distributions

A **Continuous Random Variable** has an infinite number of possible values that can be represented by an interval on the number line.

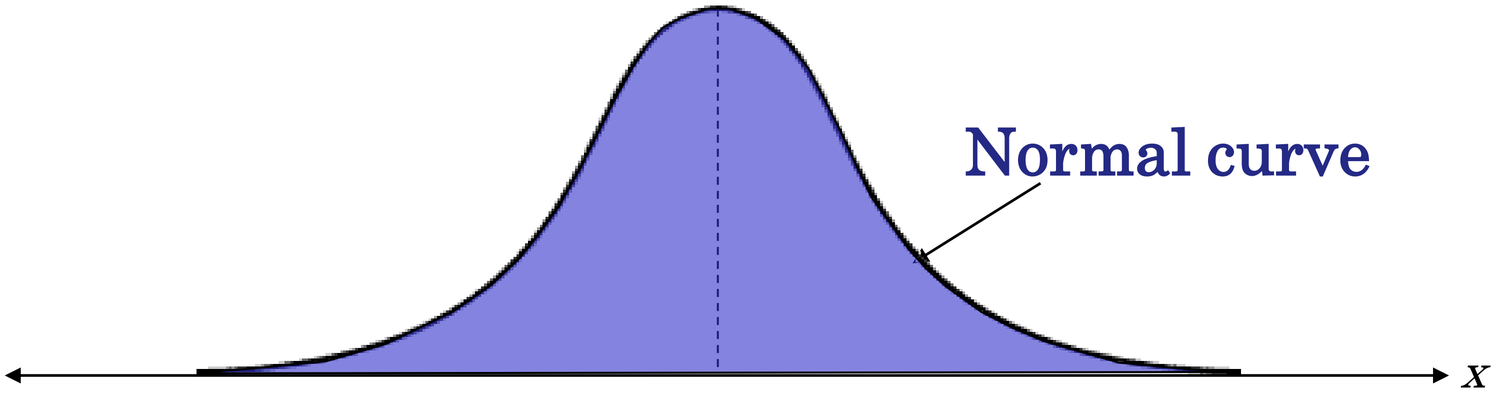


The probability distribution of a continuous random variable is called a **continuous probability distribution.**

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Properties of Normal Distributions

The most important probability distribution in statistics is the **normal distribution**.



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

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

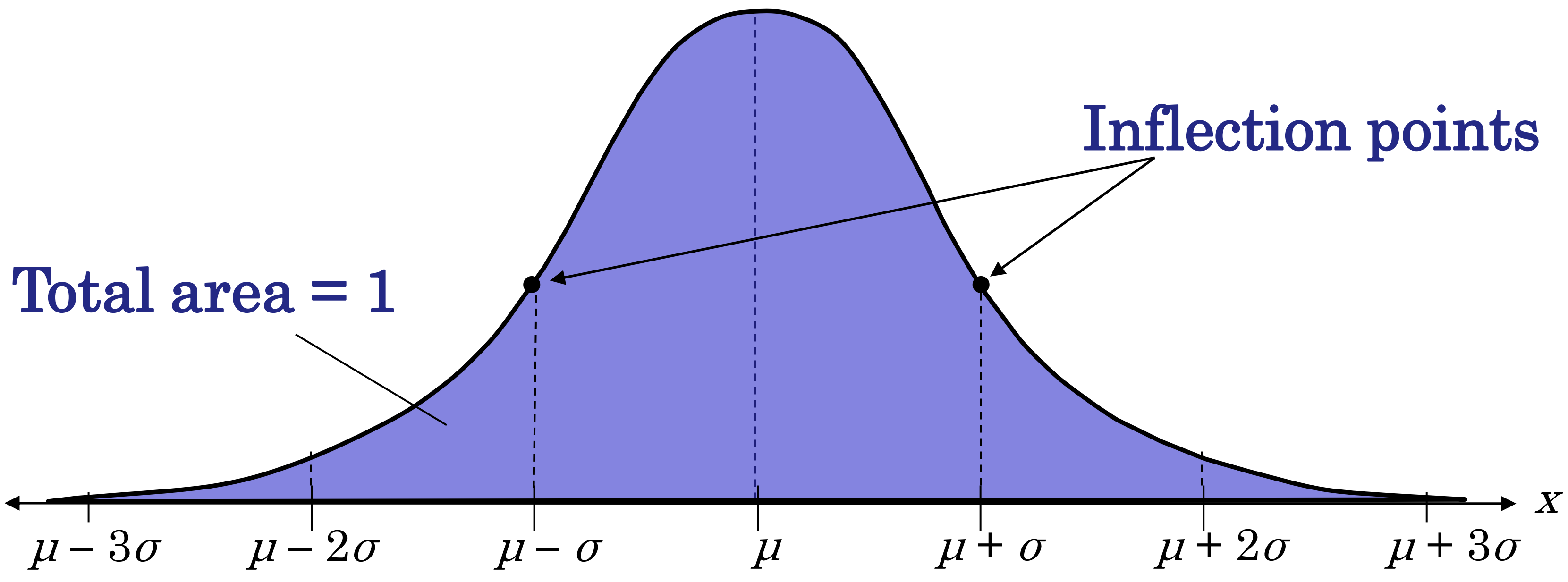
Properties of Normal Distributions

Properties of a Normal Distribution

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

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Properties of Normal Distributions



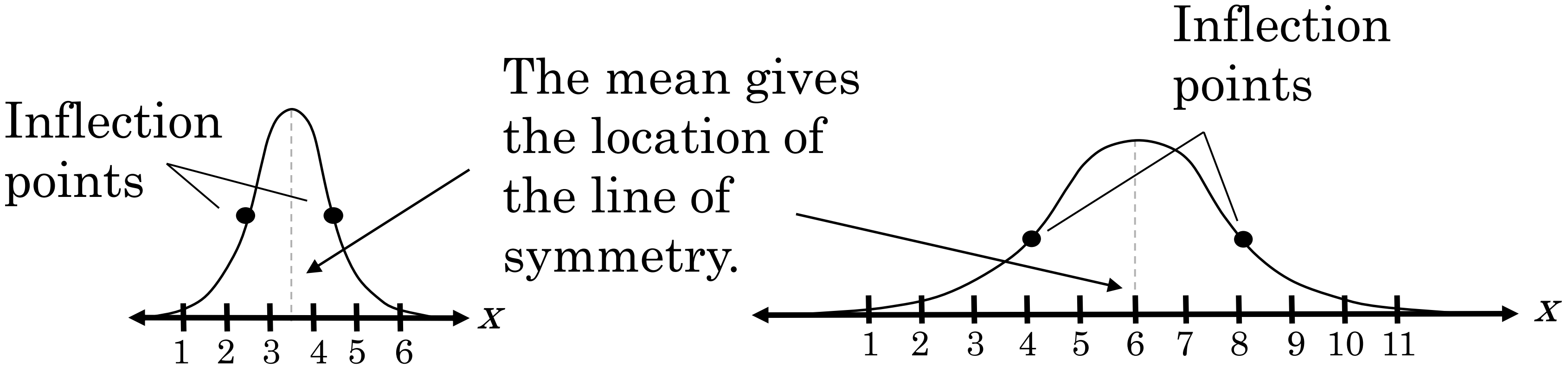
If x is a continuous random variable having a normal distribution with mean μ and standard deviation σ, you can graph a normal curve with the equation

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}. \qquad e = 2.178 \quad \pi = 3.14$$

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Means and Standard Deviations

A normal distribution can have any mean and any positive standard deviation.



Mean: $\mu = 3.5$

Standard deviation: $\sigma \approx 1.3$

Mean: $\mu = 6$

Standard deviation: $\sigma \approx 1.9$

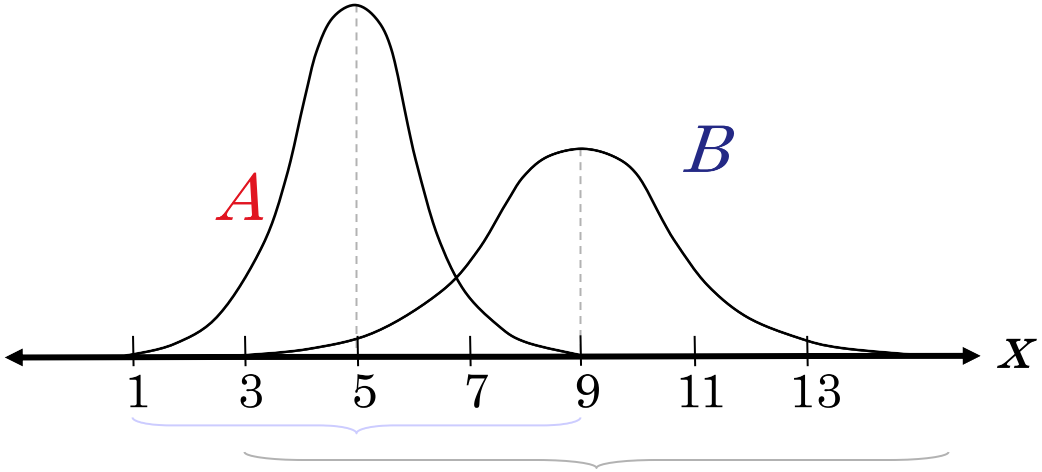
The standard deviation describes the spread of the data.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Means and Standard Deviations

Example:

1. Which curve has the greater mean?
2. Which curve has the greater standard deviation?



The line of symmetry of curve A occurs at $x = 5$. The line of symmetry of curve B occurs at $x = 9$. Curve B has the greater mean.

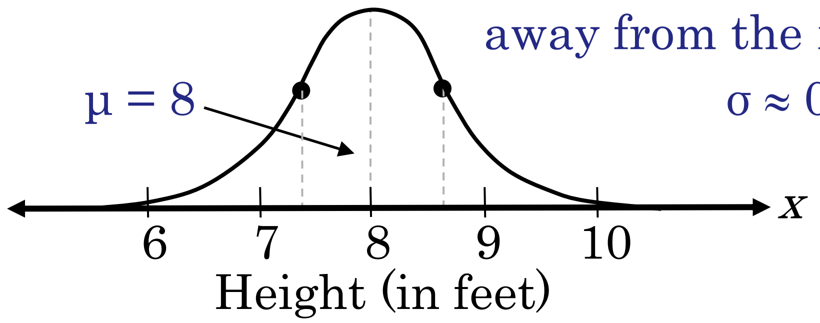
Curve B is more spread out than curve A, so curve B has the greater standard deviation.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Interpreting Graphs

Example

The heights of fully grown magnolia bushes are normally distributed. The curve represents the distribution. What is the mean height of a fully grown magnolia bush? Estimate the standard deviation.



The inflection points are one standard deviation away from the mean.

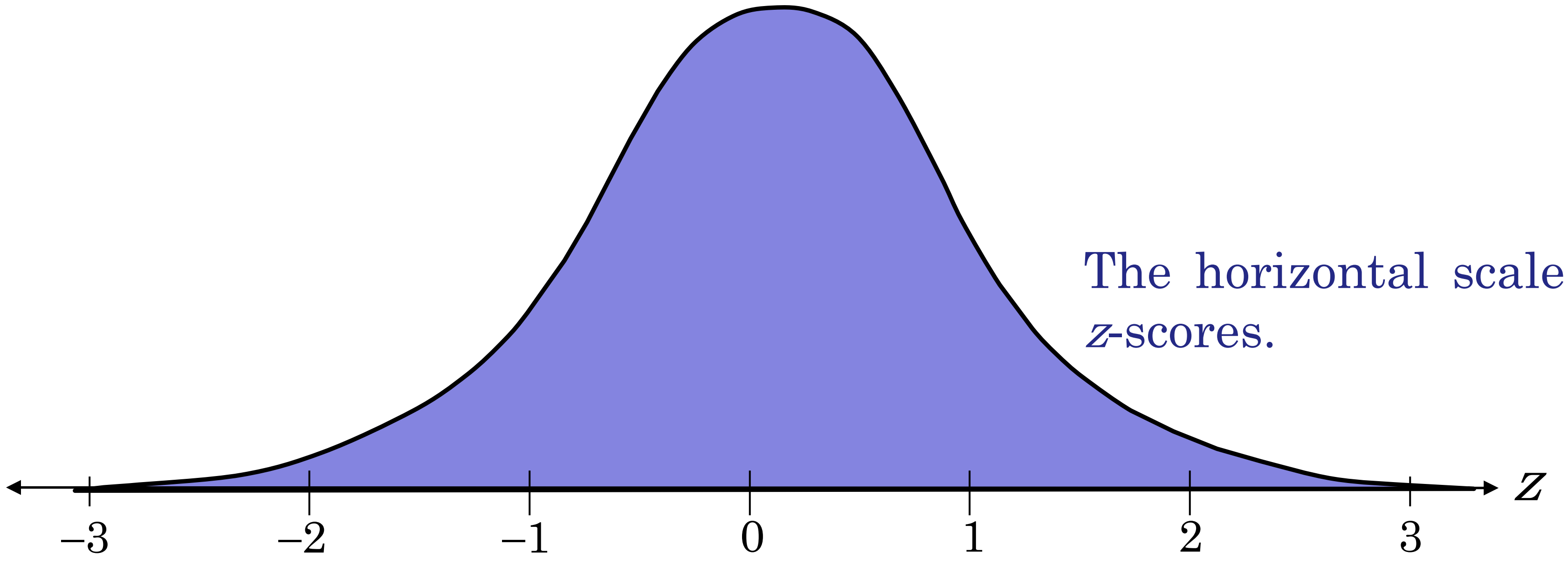
$\sigma \approx 0.7$

The heights of the magnolia bushes are normally distributed with a mean height of about 8 feet and a standard deviation of about 0.7 feet.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

The Standard Normal Distribution

The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1.



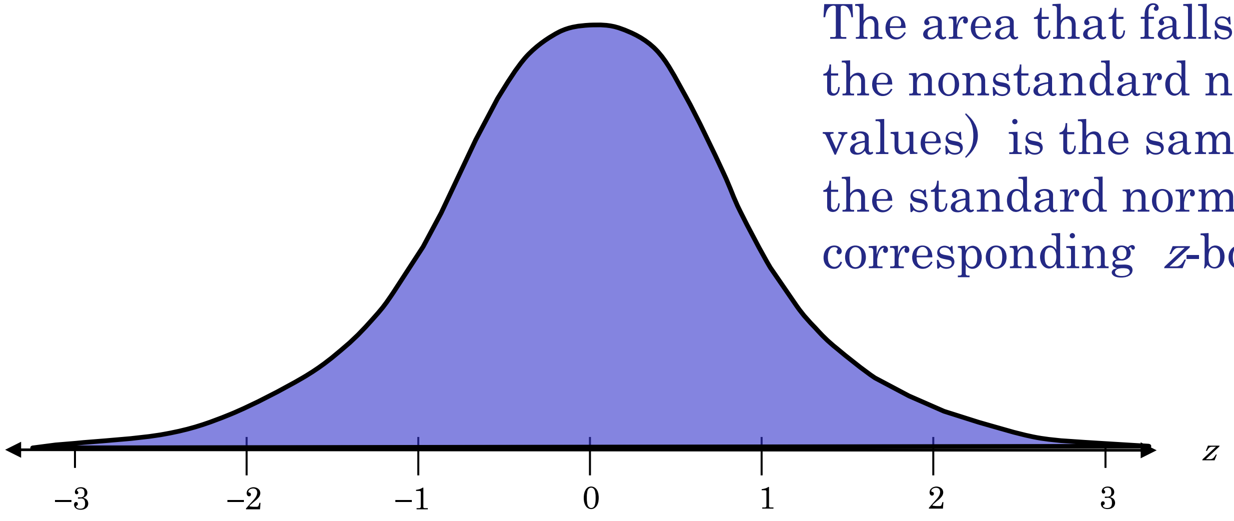
The horizontal scale corresponds to *z*-scores.

Any value can be transformed into a *z*-score by using the formula $z = \frac{\text{Value - Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}$.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

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.



The area that falls in the interval under the nonstandard normal curve (the *x*-values) is the same as the area under the standard normal curve (within the corresponding *z*-boundaries).

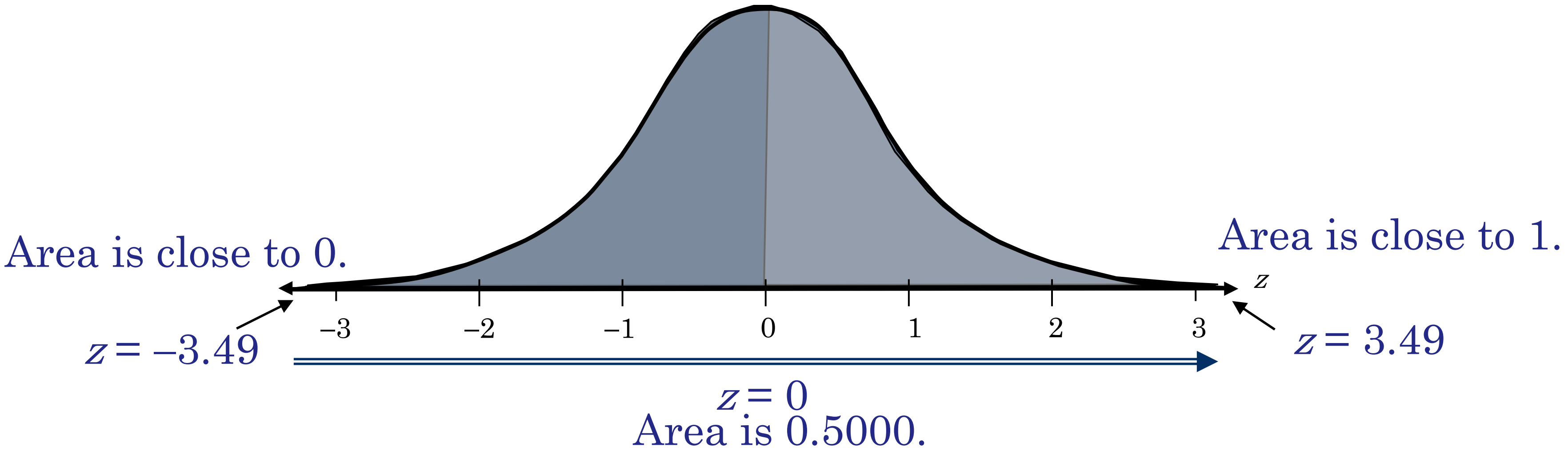
After the formula is used to transform an *x*-value into a *z*-score, the Standard Normal Table in Appendix B is used to find the cumulative area under the curve.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

The Standard Normal Table

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.
3. The cumulative area for $z = 0$ is 0.5000.
4. The cumulative area is close to 1 for *z*-scores close to $z = 3.49$



Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

The Standard Normal Table

Example:
Find the cumulative area that corresponds to a z-score of 2.71.

Appendix B: Standard Normal Table

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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981

Find the area by finding 2.7 in the left hand column, and then moving across the row to the column under 0.01. The area to the left of $z = 2.71$ is 0.9966.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

The Standard Normal Table

Example:
Find the cumulative area that corresponds to a *z*-score of –0.25.

Appendix B: Standard Normal Table

z	.09	.08	.07	.06	.05	.04	.03	.02	.01	.00
–3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
–3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
–0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821
–0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207
–0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602
–0.0	.4641	.4681	.4724	.4761	.4801	.4840	.4880	.4920	.4960	.5000

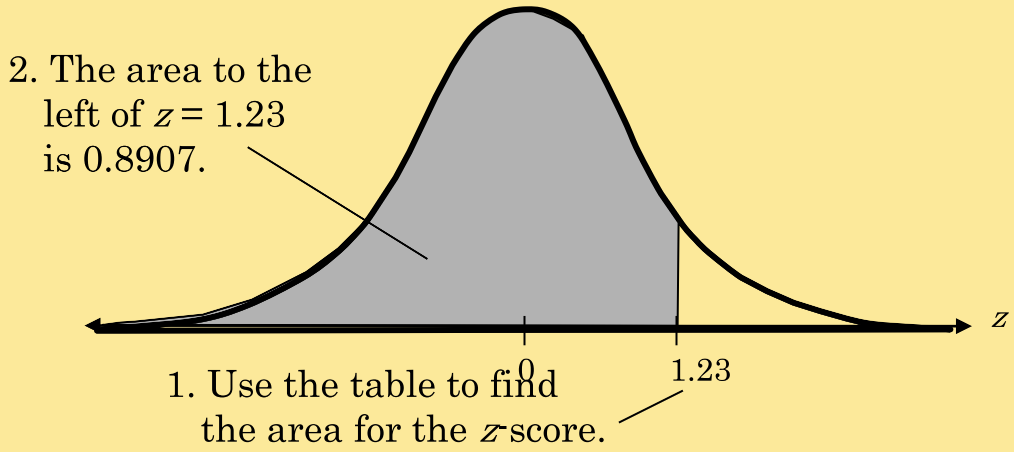
Find the area by finding –0.2 in the left hand column, and then moving across the row to the column under 0.05. The area to the left of $z = –0.25$ is 0.4013

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Guidelines for Finding Areas

Finding Areas Under the Standard Normal Curve

1. Sketch the standard normal curve and shade the appropriate area under the curve.
2. Find the area by following the directions for each case shown.
   a. To find the area to the *left* of z, find the area that corresponds to z in the Standard Normal Table.

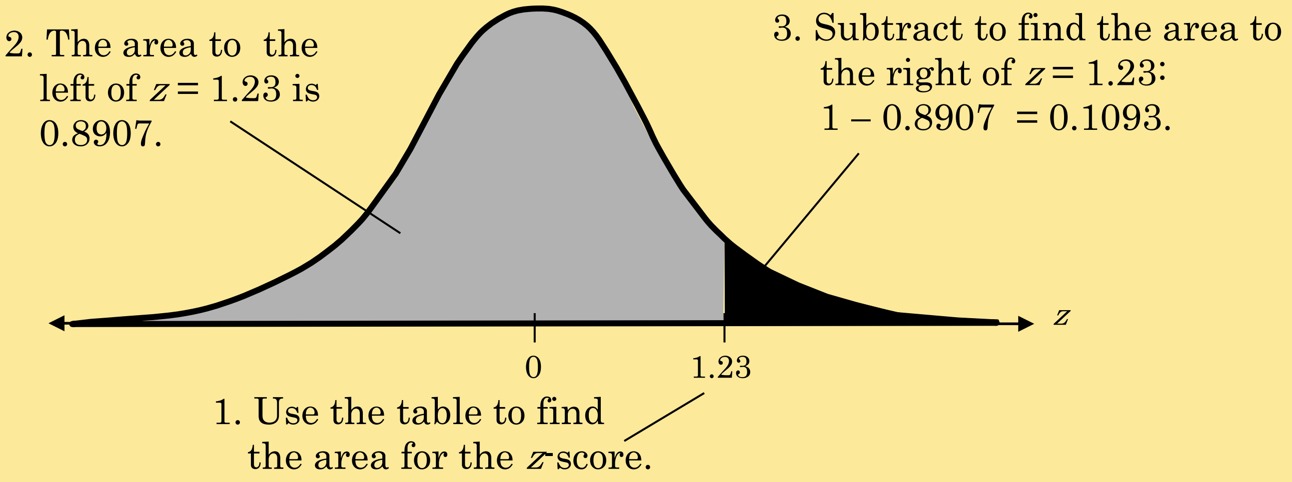


Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Guidelines for Finding Areas

Finding Areas Under the Standard Normal Curve

b. To find the area to the *right* of *z*, use the Standard Normal Table to find the area that corresponds to *z*. Then subtract the area from 1.

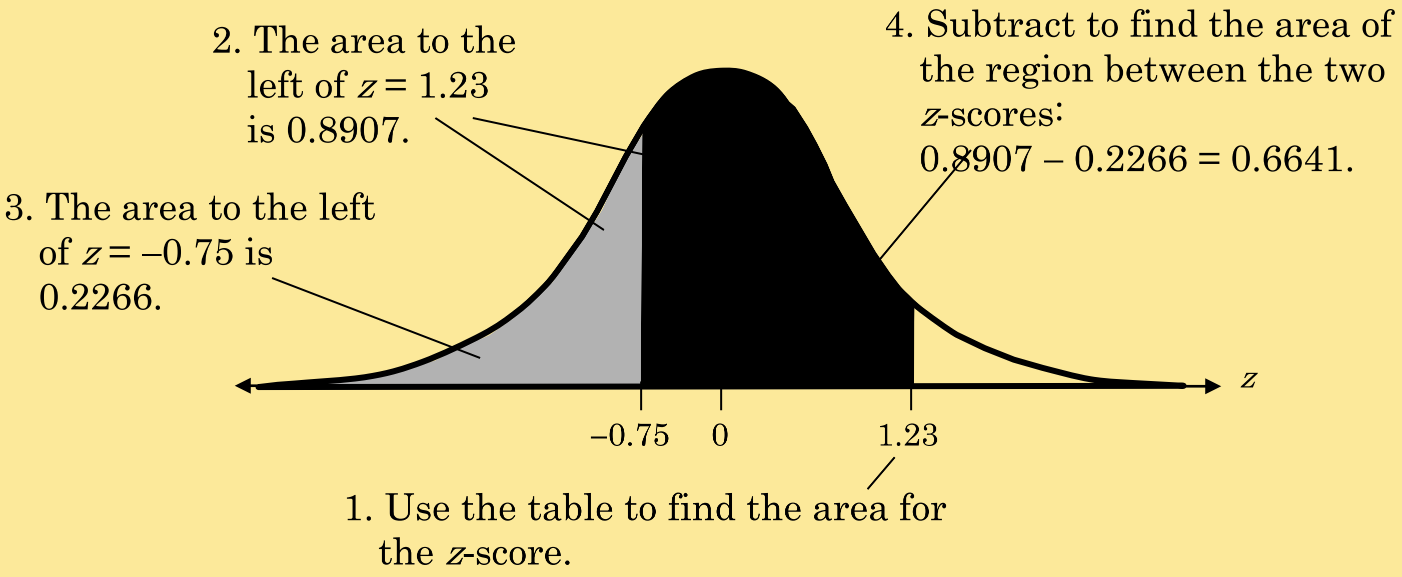


Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Guidelines for Finding Areas

Finding Areas Under the Standard Normal Curve

c. To find the area *between* two *z*-scores, find the area corresponding to each *z*-score in the Standard Normal Table. Then subtract the smaller area from the larger area.

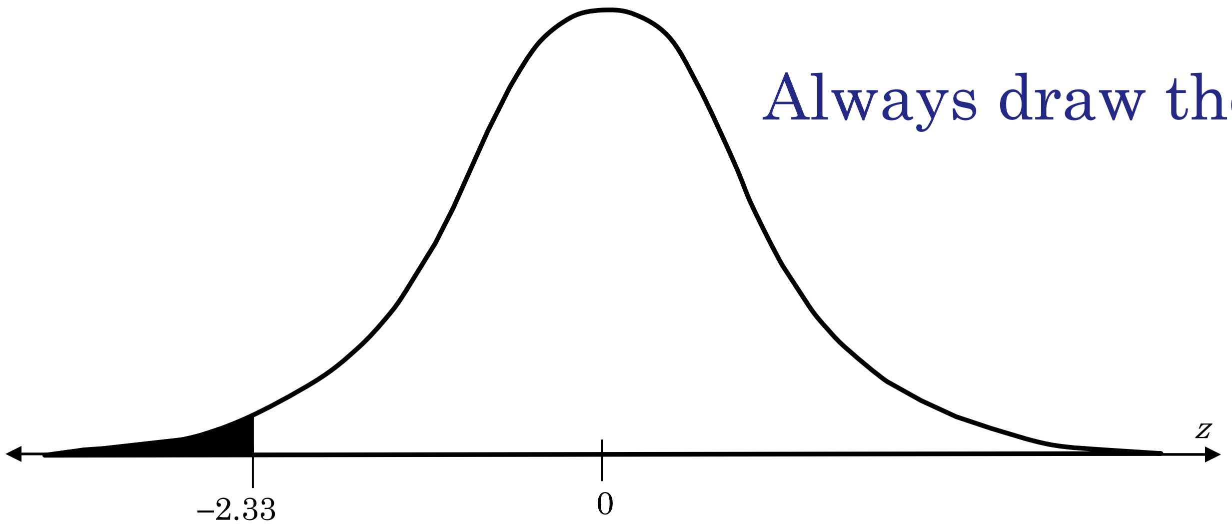


Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Guidelines for Finding Areas

Example:

Find the area under the standard normal curve to the left of $z = -2.33$.



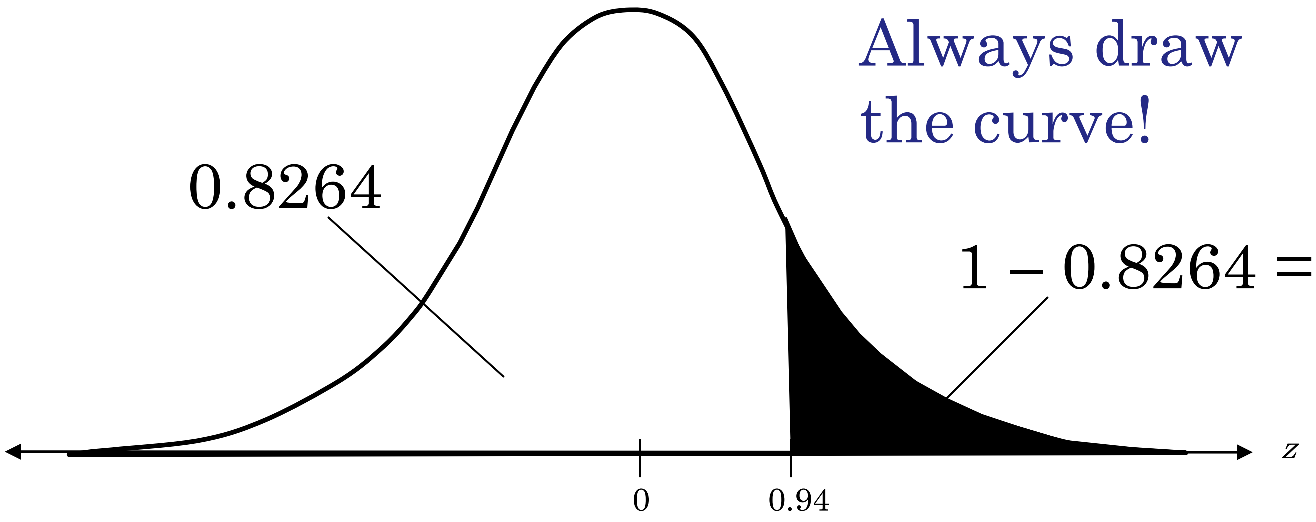
Always draw the curve!

From the Standard Normal Table, the area is equal to 0.0099.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Guidelines for Finding Areas

Example:
Find the area under the standard normal curve to the right of z = 0.94.



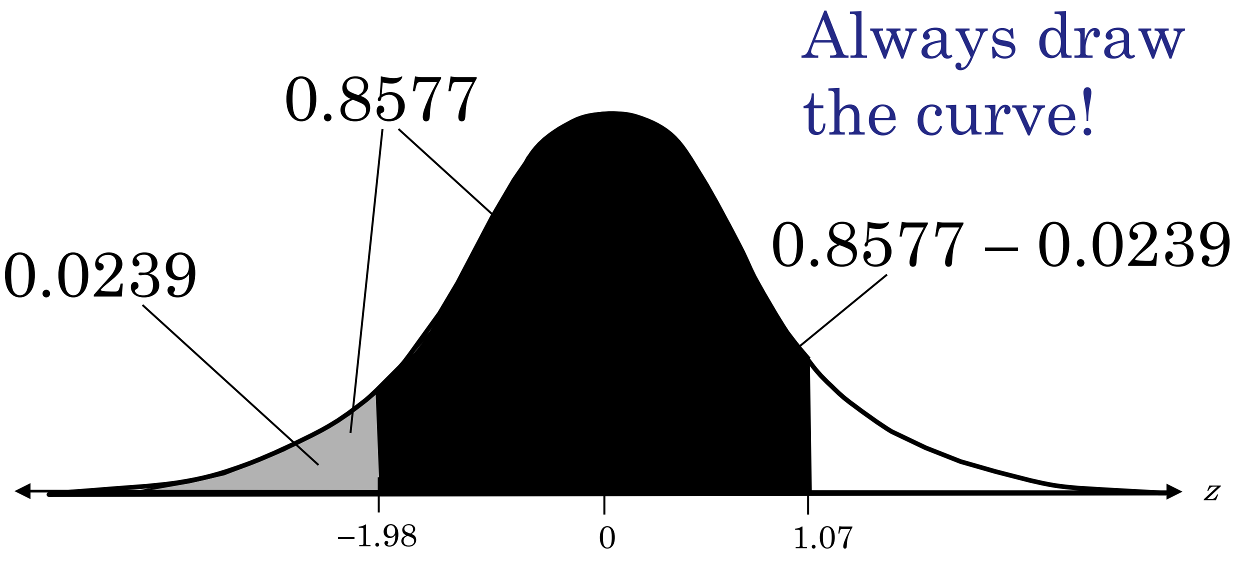
From the Standard Normal Table, the area is equal to 0.1736.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Guidelines for Finding Areas

Example:

Find the area under the standard normal curve between $z = -1.98$ and $z = 1.07$.



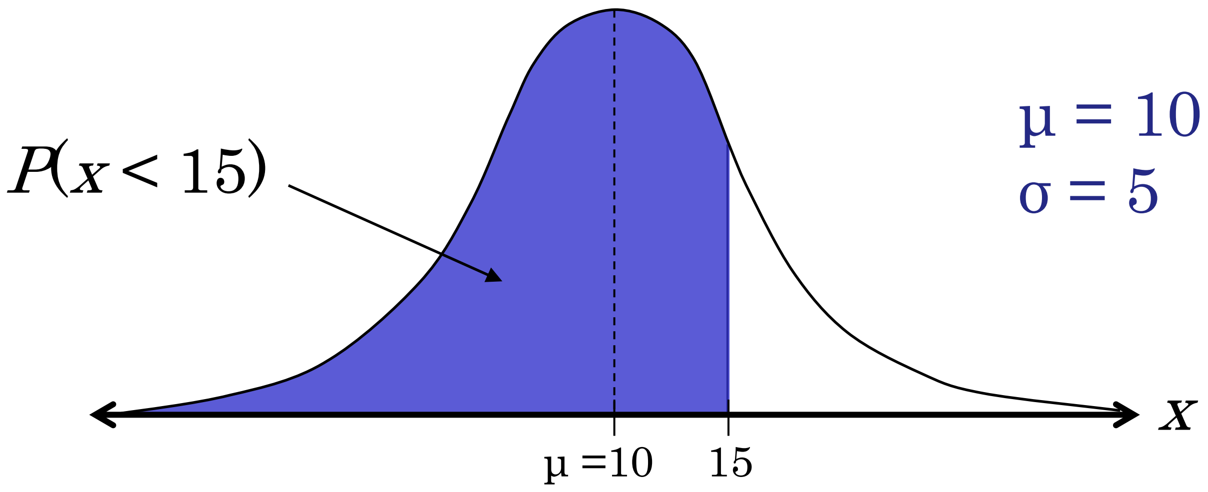
From the Standard Normal Table, the area is equal to 0.8338.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Normal Distributions: Finding Probabilities

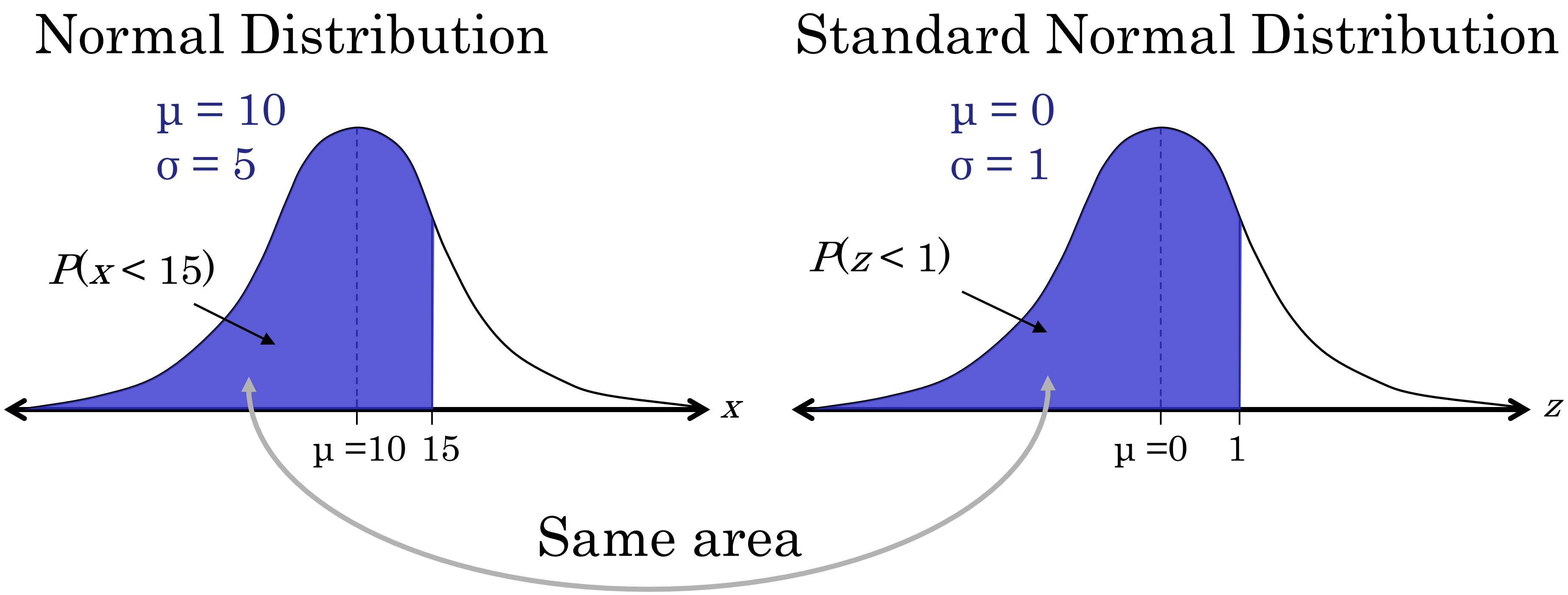
Probability and Normal Distributions

If a random variable, *x*, is normally distributed, you can find the probability that *x* will fall in a given interval by calculating the area under the normal curve for that interval.



Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Probability and Normal Distributions



$$P(x < 15) = P(z < 1) = \text{Shaded area under the curve}$$

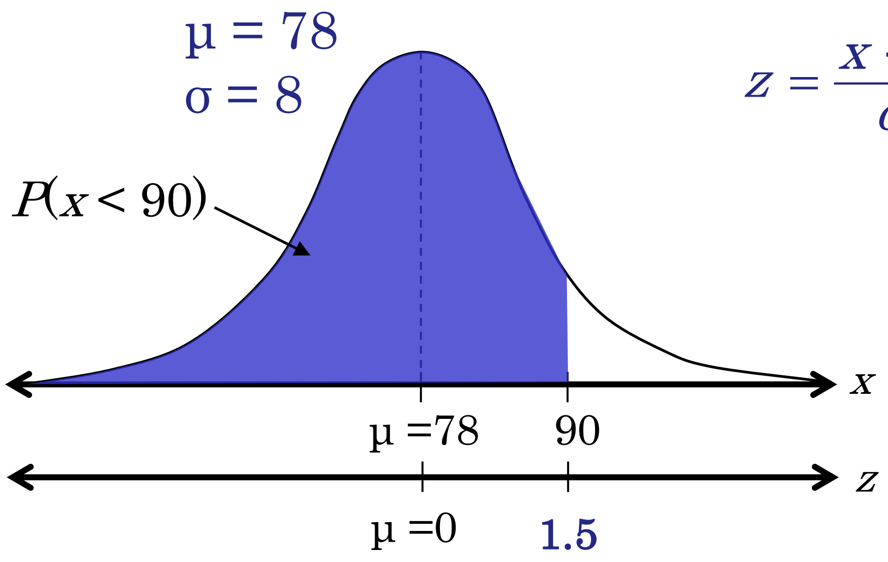
$$= 0.8413$$

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Probability and Normal Distributions

Example:

The average on a statistics test was 78 with a standard deviation of 8. If the test scores are normally distributed, find the probability that a student receives a test score less than 90.



$$z = \frac{x - \mu}{\sigma} = \frac{90 - 78}{8}$$

$$= 1.5$$

The probability that a student receives a test score less than 90 is 0.9332.

$$P(x < 90) = P(z < 1.5) = 0.9332$$

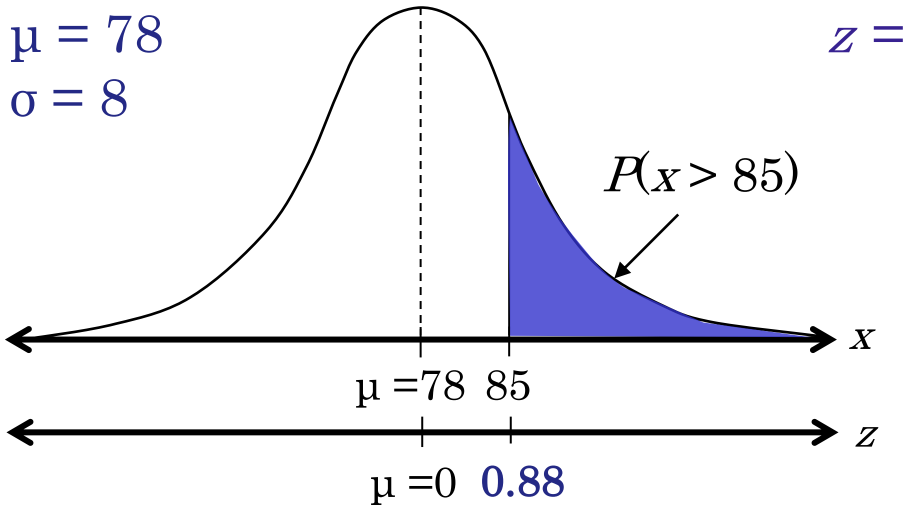
Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Probability and Normal Distributions

Example:

The average on a statistics test was 78 with a standard deviation of 8. If the test scores are normally distributed, find the probability that a student receives a test score greater than than 85.

$\mu = 78$
$\sigma = 8$



$$z = \frac{x - \mu}{\sigma} = \frac{85 - 78}{8}$$
$$= 0.875 \approx 0.88$$

The probability that a student receives a test score greater than 85 is 0.1894.

$$P(x > 85) = P(z > 0.88) = 1 - P(z < 0.88) = 1 - 0.8106 = 0.1894$$

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

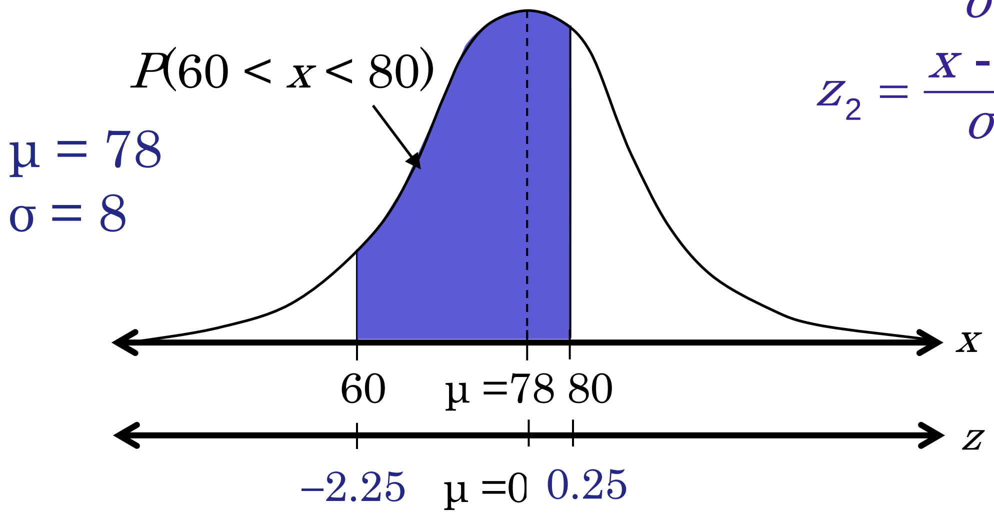
Probability and Normal Distributions

Example:

The average on a statistics test was 78 with a standard deviation of 8. If the test scores are normally distributed, find the probability that a student receives a test score between 60 and 80.

$$z_1 = \frac{x - \mu}{\sigma} = \frac{60 - 78}{8} = -2.25$$

$$z_2 = \frac{x - \mu}{\sigma} = \frac{80 - 78}{8} = 0.25$$



The probability that a student receives a test score between 60 and 80 is 0.5865.

$$P(60 < x < 80) = P(-2.25 < z < 0.25) = P(z < 0.25) - P(z < -2.25)$$

$$= 0.5987 - 0.0122 = 0.5865$$

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Normal Distributions: Finding Values

Finding z-Scores

Example: Find the z-score that corresponds to a cumulative area of 0.9973.

Appendix B: Standard Normal Table

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
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981

Find the z-score by locating 0.9973 in the body of the Standard Normal Table. The values at the beginning of the corresponding row and at the top of the column give the z-score.

The z-score is 2.78.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Finding z-Scores

Example: Find the z-score that corresponds to a cumulative area of 0.4170.

Appendix B: Standard Normal Table

z	.09	.08	.07	.06	.05	.04	.03	.02	.01	.00
–3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
–0.2	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
–0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821
–0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207
–0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602
–0.0	.4641	.4681	.4724	.4761	.4801	.4840	.4880	.4920	.4960	.5000

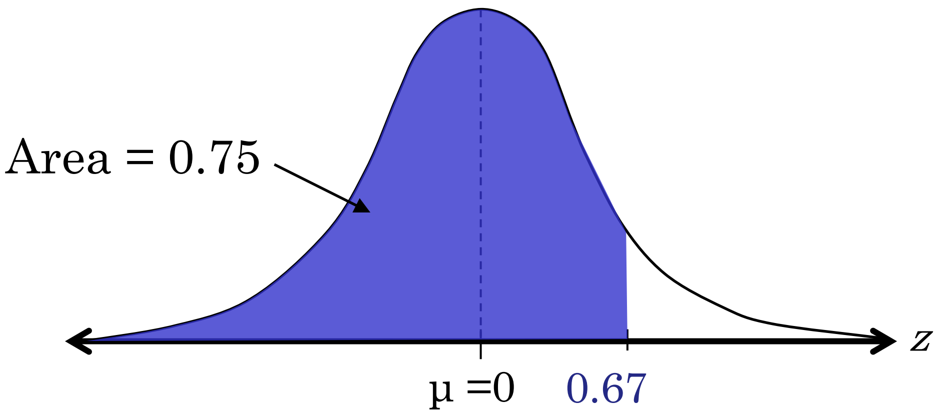
Use the closest area.

Find the z-score by locating 0.4170 in the body of the Standard Normal Table. Use the value closest to 0.4170. The z-score is –0.21.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Finding a z-Score Given a Percentile

Example: Find the z-score that corresponds to P_{75}.



The z-score that corresponds to P_{75} is the same z-score that corresponds to an area of 0.75.

The z-score is 0.67.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Transforming a z-Score to an x-Score

To transform a standard z-score to a data value, x, in a given population, use the formula

$$x = \mu + z\sigma.$$

Example:

The monthly electric bills in a city are normally distributed with a mean of \$120 and a standard deviation of \$16. Find the *x*-value corresponding to a *z*-score of 1.60.

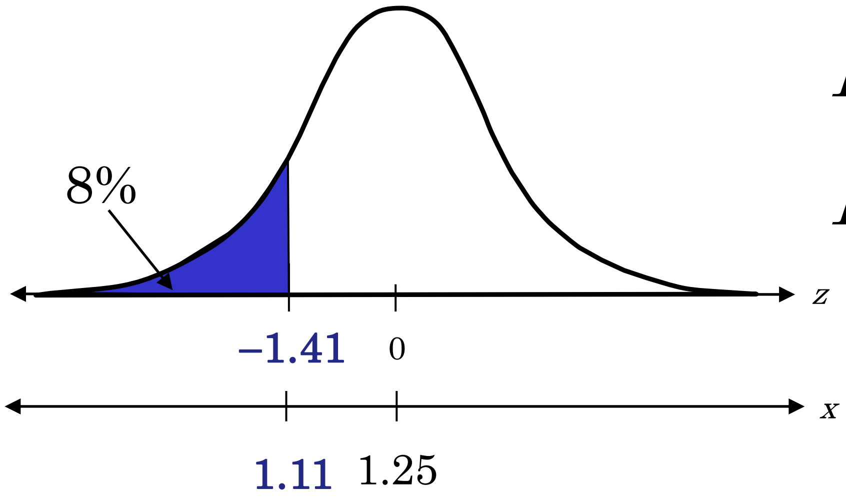
$$\begin{aligned} x &= \mu + z\sigma \\ &= 120 + 1.60(16) \\ &= 145.6 \end{aligned}$$

We can conclude that an electric bill of \$145.60 is 1.6 standard deviations above the mean.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Finding a Specific Data Value

Example:
The weights of bags of chips for a vending machine are normally distributed with a mean of 1.25 ounces and a standard deviation of 0.1 ounce. Bags that have weights in the lower 8% are too light and will not work in the machine. What is the least a bag of chips can weigh and still work in the machine?



$$P(z < ?) = 0.08$$

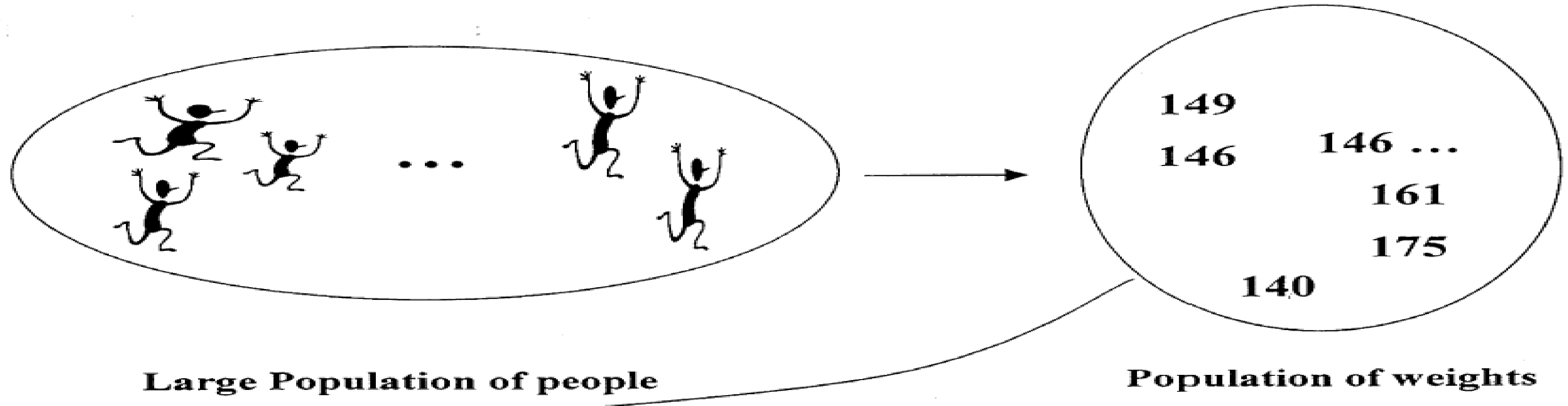
$$P(z < -1.41) = 0.08$$

$$x = \mu + z\sigma$$
$$= 1.25 + (-1.41)0.1$$
$$= 1.11$$

The least a bag can weigh and still work in the machine is 1.11 ounces.

Source: http://www3.govst.edu/kriordan/files/mvcc/math139/ppt/lfstat3e_ppt_05_rev.ppt

Sampling Distributions



Individual observations	Means for n = 5	Means for n = 20
149	153.0	151.6
146	146.4	151.3
.	.	.
.	.	.
.	.	.
μ = 150 lbs	μ = 150 lbs	μ = 150 lbs
$\sigma^2 = 100\text{lbs}$	$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = 20\ \text{lbs}^2$	$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = 5\ \text{lbs}^2$
$\sigma = 10$ lbs	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 4.47$ lbs	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 2.23$ lbs

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Normal Distribution Density Function

The normal distribution is defined by the density function:

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

This function happens to be

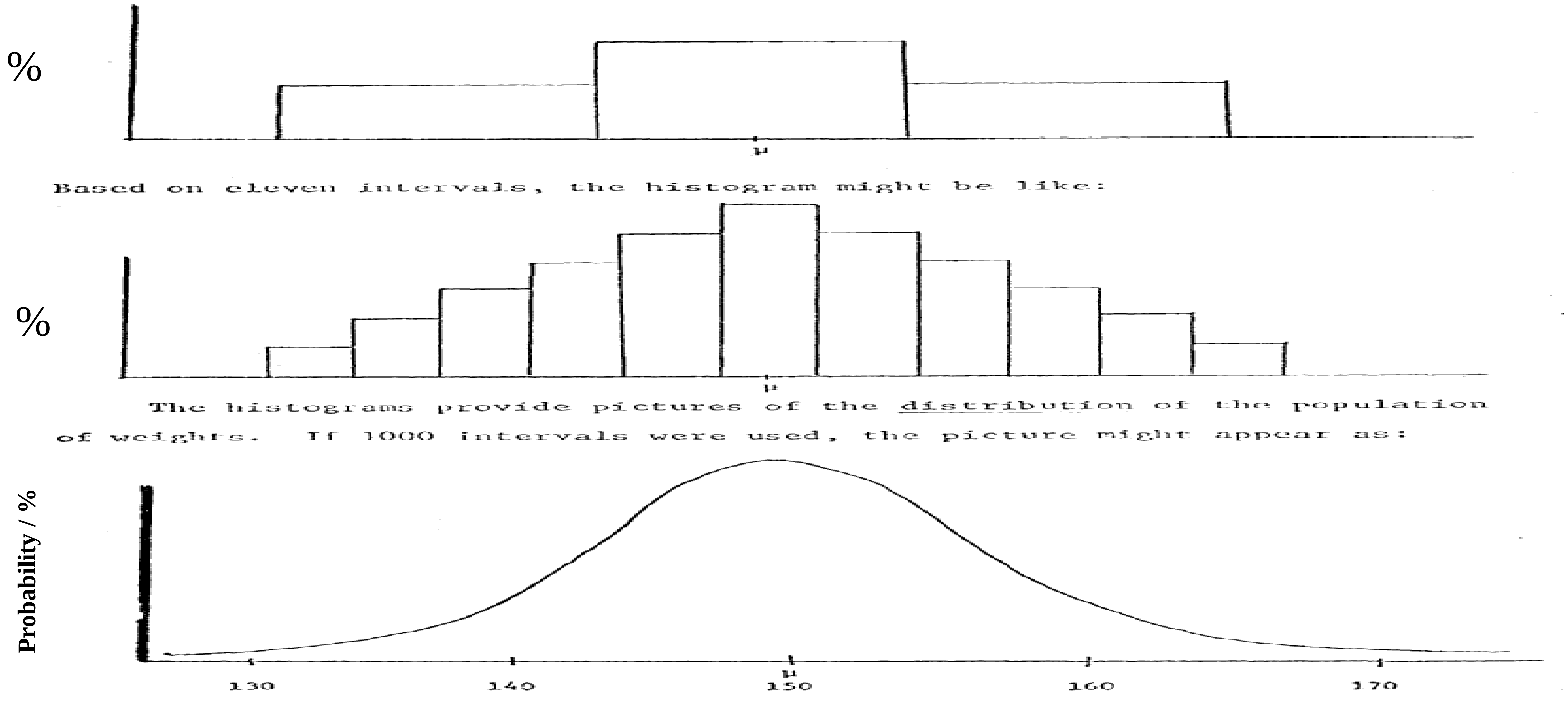
Symmetrical,

Bell-shaped,

and easy to use tables are available.

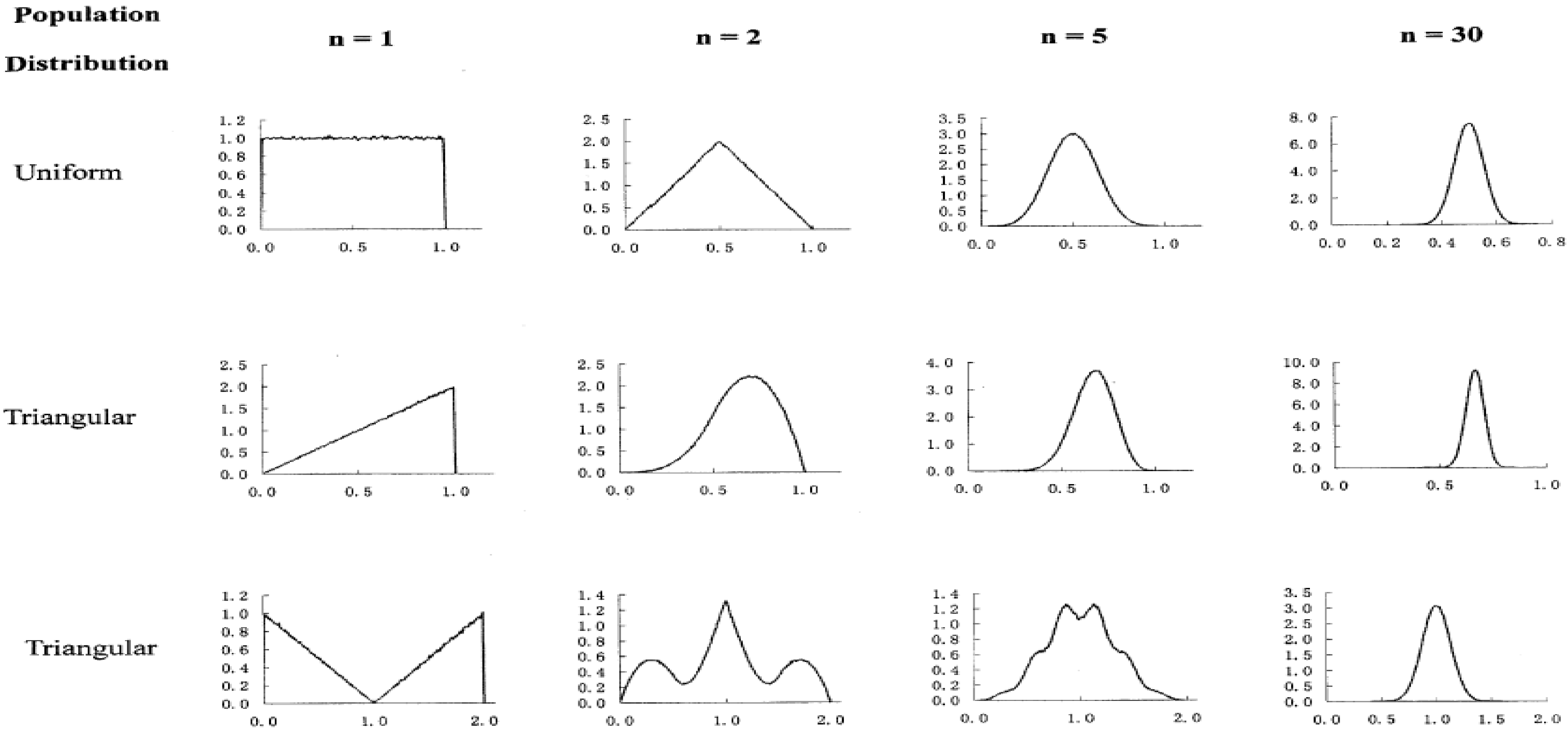
Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Normal Distribution



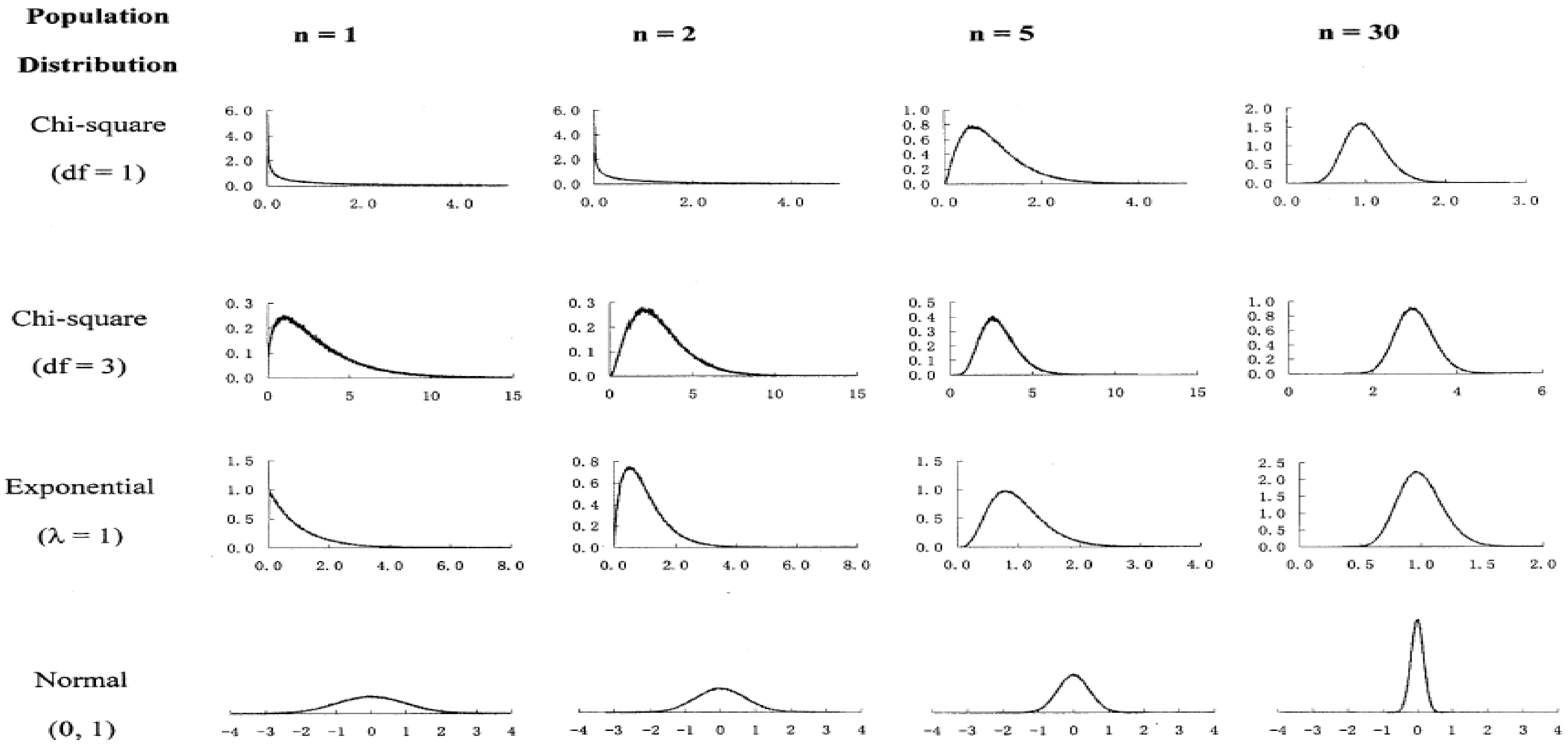
Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Population Distributions



Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Population Distributions



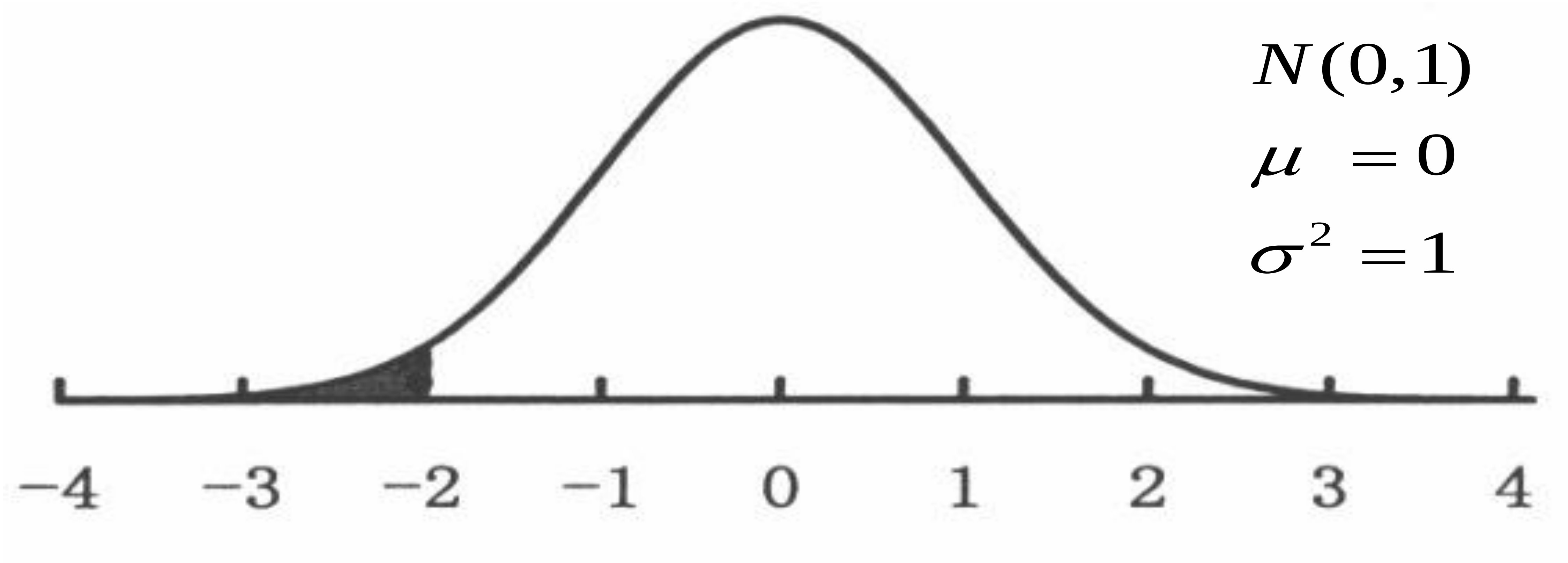
Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Using the Normal Tables

- We can use the normal tables to obtain probabilities for measurements for which this frequency distribution is appropriate. For a reasonably complete set of probabilities, see TABLE MODULE 1: NORMAL TABLE.
- This module provides most of the z-values and associated probabilities you are likely to use; however, it also provides instructions demonstrating how to calculate those not included directly in the table.
- The table is a series of columns containing numbers for z and for P(z). The z represents the z-value for a normal distribution and P(z) represents the area under the normal curve to the left of that z-value for a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Using the Normal Tables

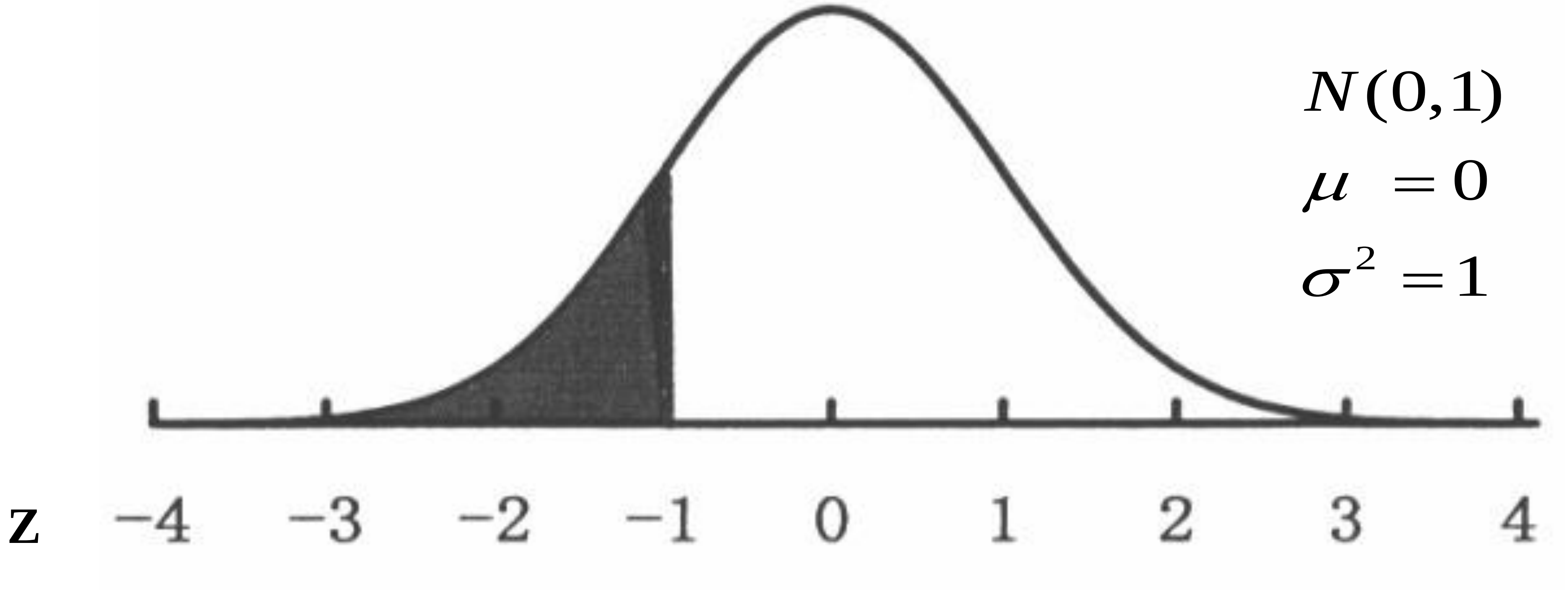


z

(1) Area Below z = -2; P(z < -2) = 0.0228

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

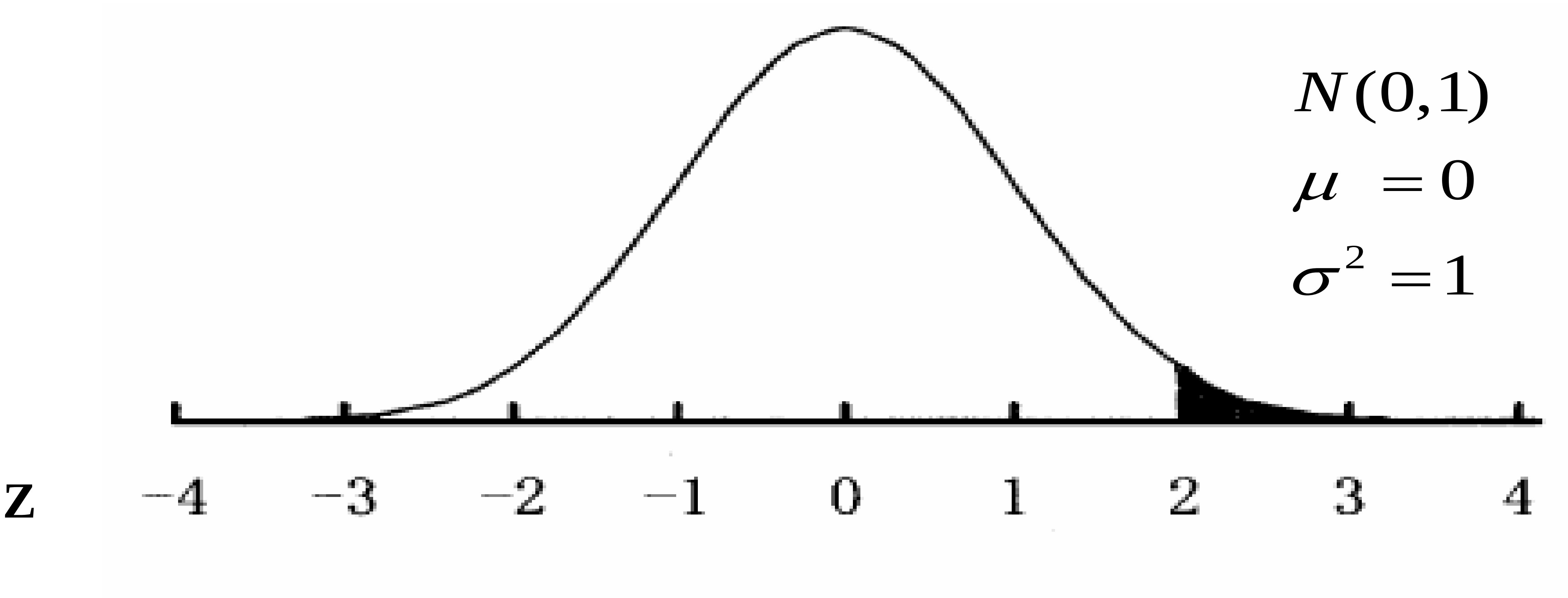
Using the Normal Tables



(2) Area Below z = -1; P(z < -1) = 0.1587

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

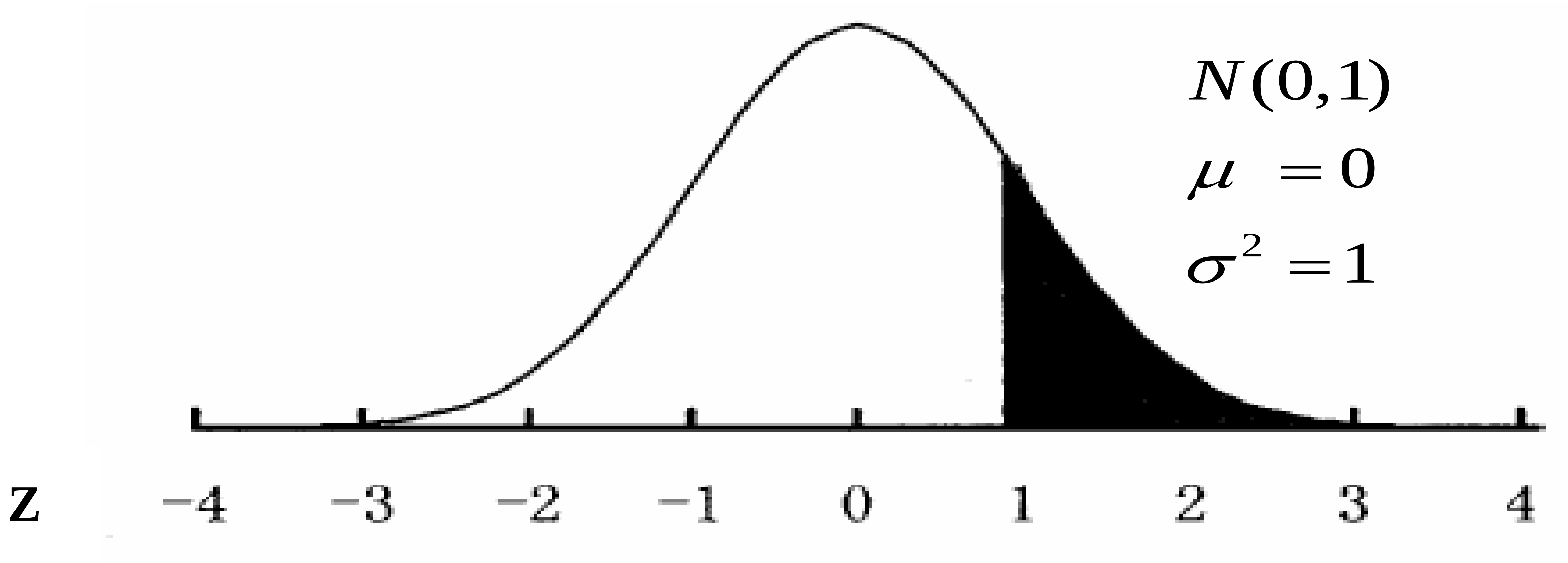
Using the Normal Tables



(1) Area Below z = +2; P(z > +2) = 0.0228

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

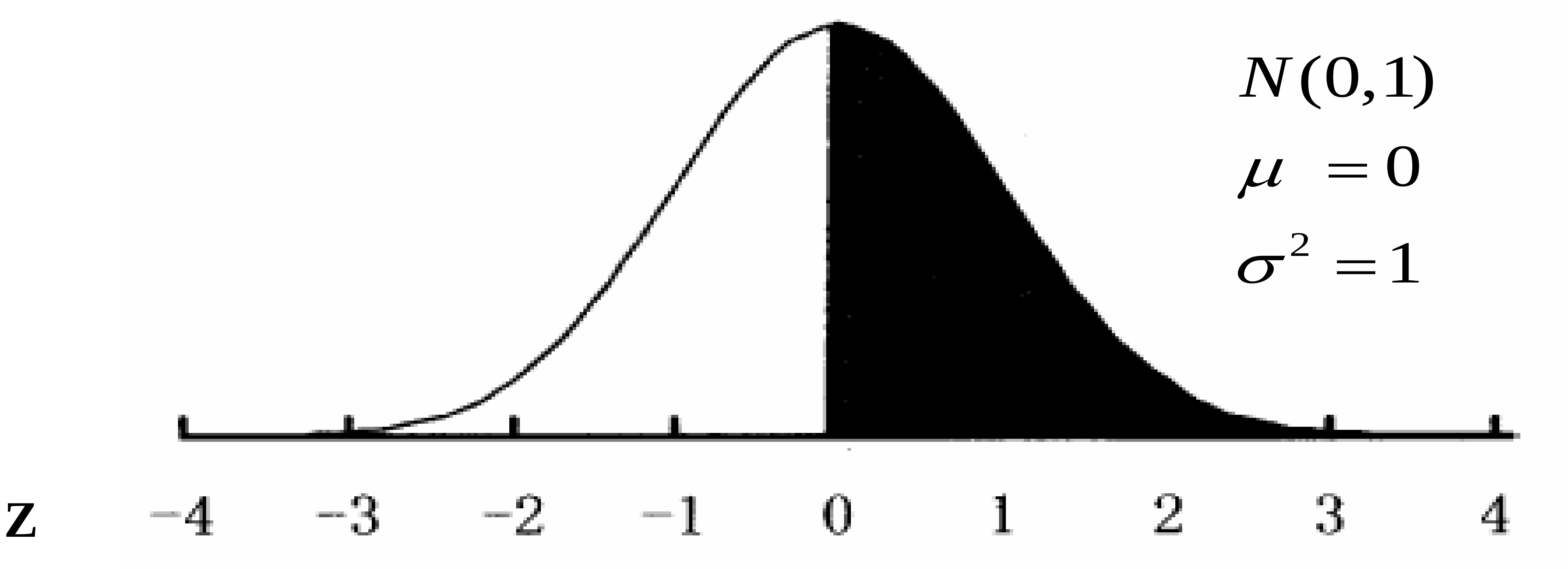
Using the Normal Tables



(2) Area Below z = +1; P(z > +1) = 0.1587

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

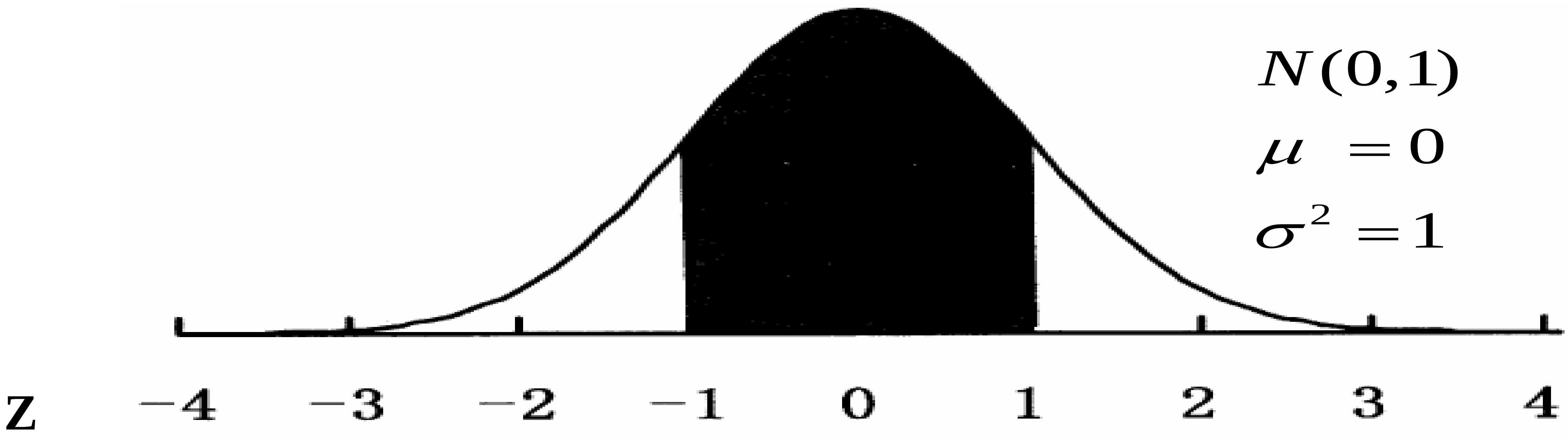
Using the Normal Tables



(3) Area Below z = 0; P(z > 0) = 0.5000

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Calculating the Area Under the Normal Curve



(1) Area between -1, +1; P(-1 < z < +1)

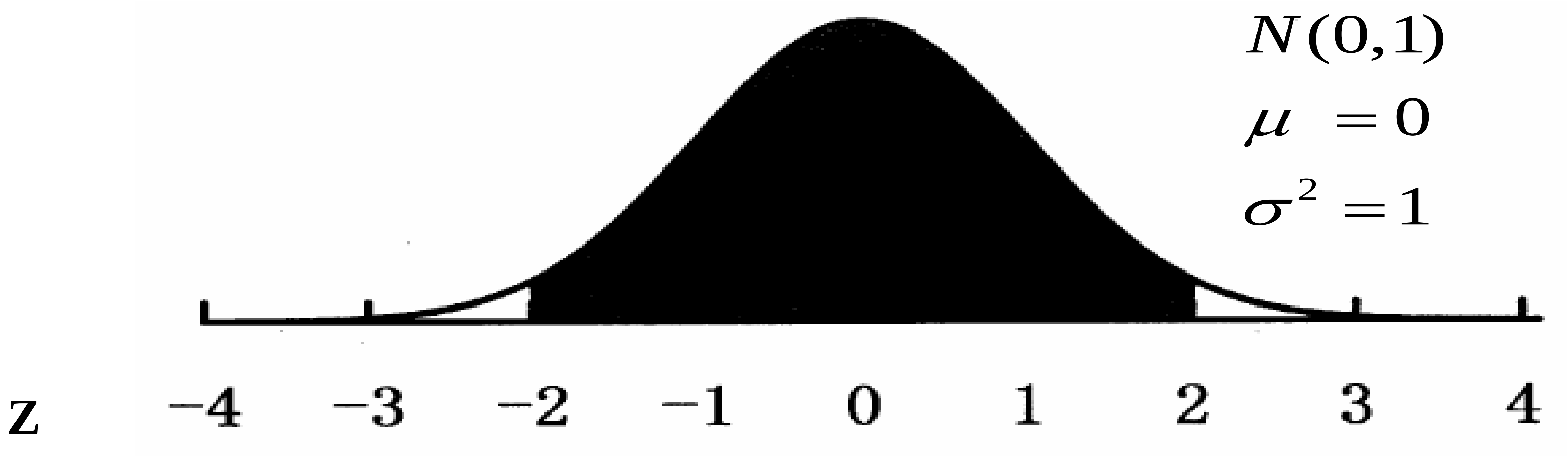
up to z = +1: .8413

up to z = -1 : .1587

.6826

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Calculating the Area Under the Normal Curve



(2) Area between -2, +2; P(-2 < z < +2)

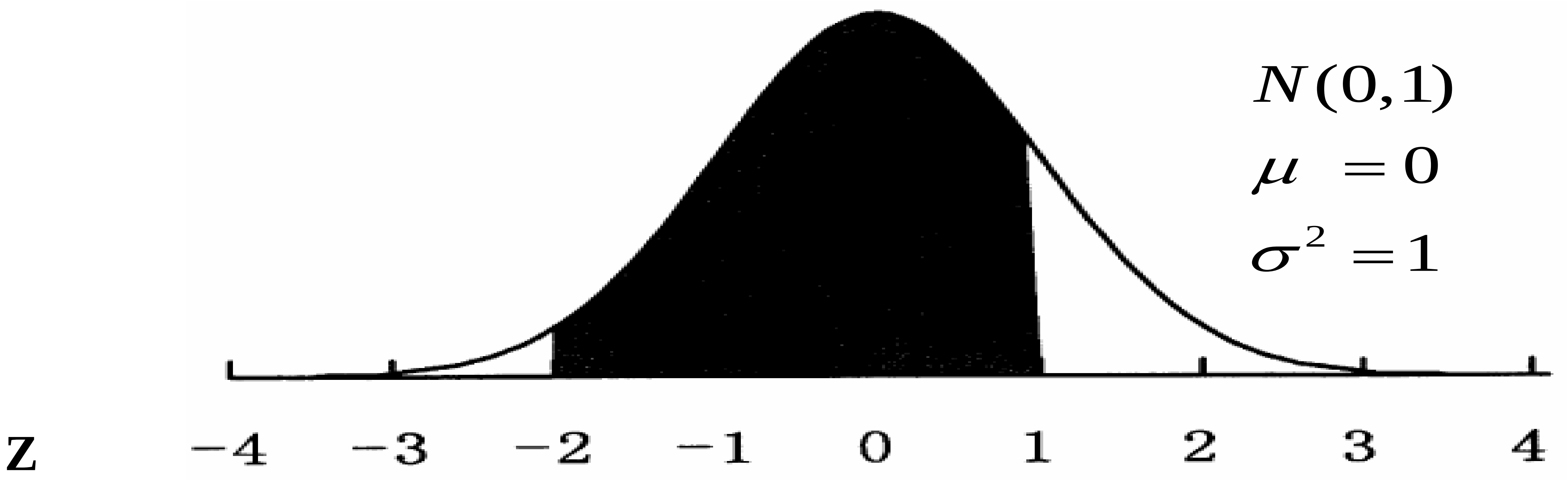
up to z = +2: .9772

up to z = -2 : .0228

.9544

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Calculating the Area Under the Normal Curve



(3) Area between -2, +1; P(-2 < z < +1)

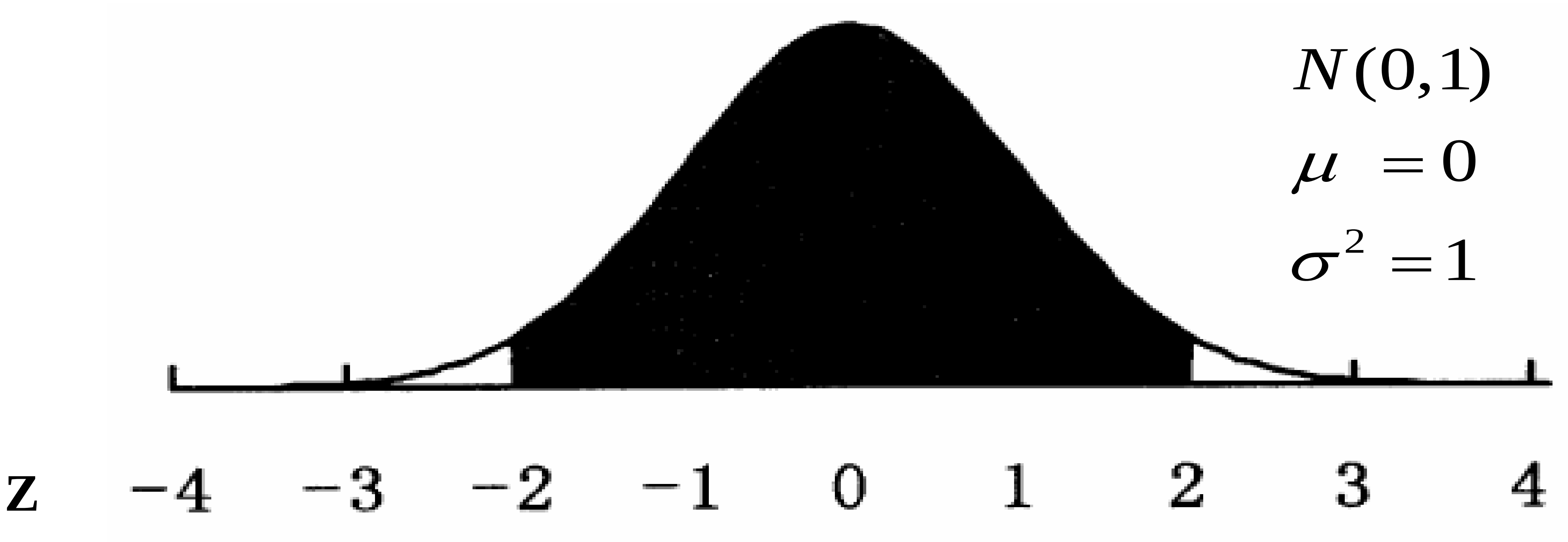
up to z = +1: .8413

up to z = -2 : .0228

.8185

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Standard Normal Distribution

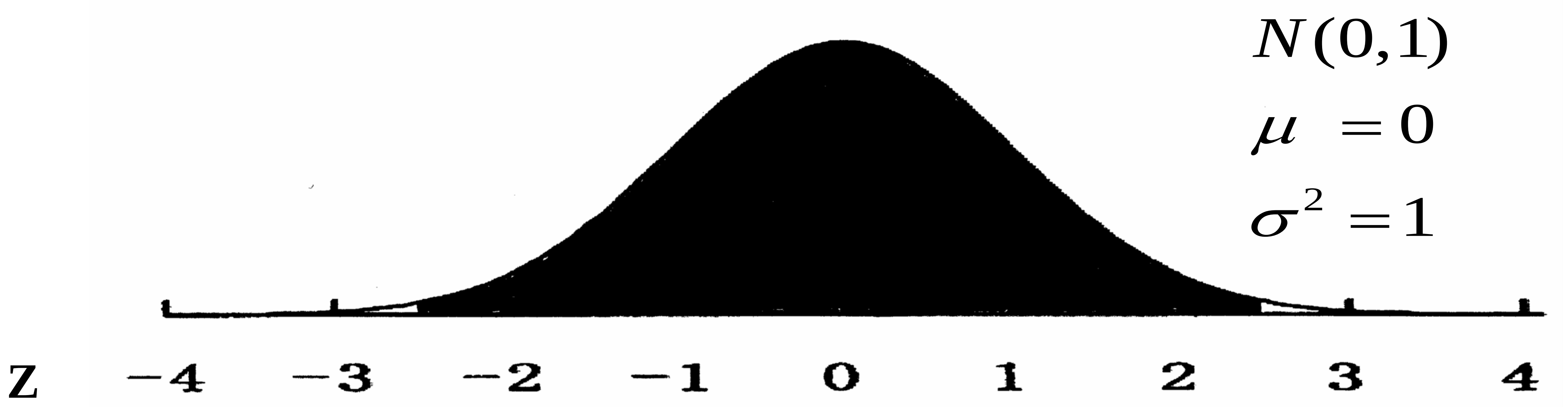


(1) Values of z that bracket middle 95%

-1.96 to +1.96

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Standard Normal Distribution



(1) Values of z that bracket middle 99%

-2.576 to +2.576

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Calculating z-values

If $X \sim N(\mu_x, \sigma_x)$

and $Z \sim N(0,1)$

i.e. $\mu_z = 0$ and $\sigma_z^2 = 1$

then the corresponding z value for x is given as

$$z = \frac{x - \mu_x}{\sigma_x}$$

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Calculating z-values

$Z \sim N(0,1)$

$\mu_z = 0$

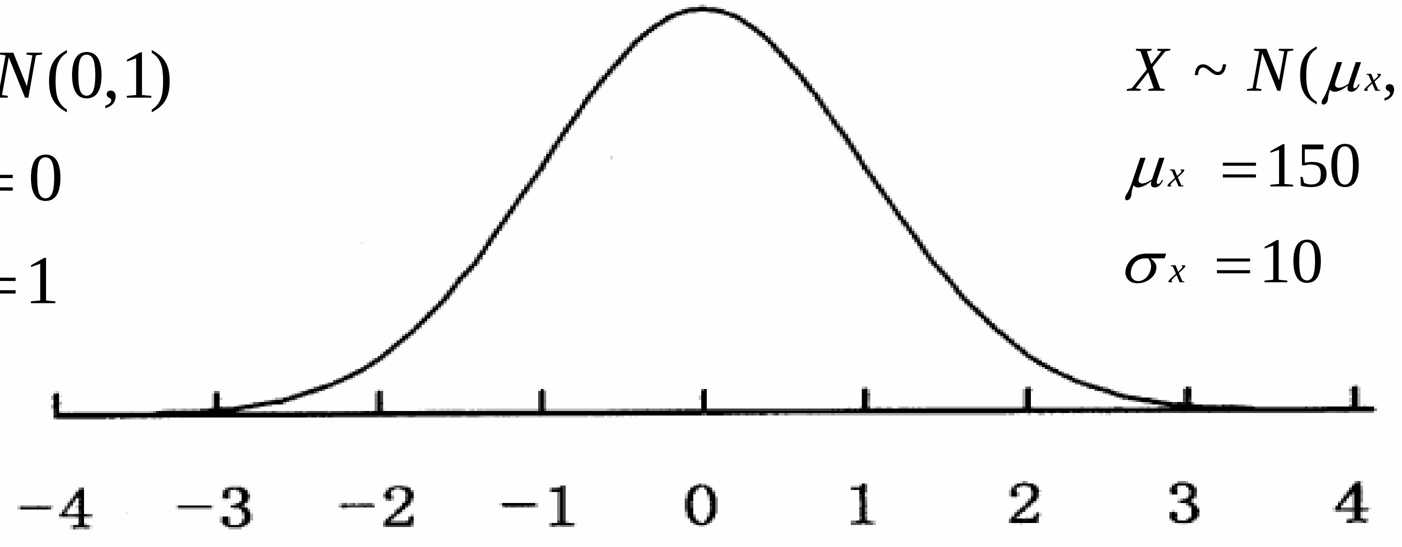
$\sigma_z^2 = 1$

$X \sim N(\mu_x, \sigma_x)$

$\mu_x = 150$

$\sigma_x = 10$

Z



Weight	110	120	130	140	150	160	170	180	190
		$\mu_x - 3\sigma_x$	$\mu_x - 2\sigma_x$	$\mu_x - 1\sigma_x$	μ_x	$\mu_x + 1\sigma_x$	$\mu_x + 2\sigma_x$	$\mu_x + 3\sigma_x$	

$$z = \frac{x - \mu_x}{\sigma_x} \;; \quad \text{if } X \sim \mathrm{N}(150, 10) \; i.e. \; \mu_x = 150, \sigma_x = 10$$

$$\text{when } x = 150; \quad z = \frac{150 - 150}{10} = 0$$

$$\text{when } x = 170; \quad z = \frac{170 - 150}{10} = \frac{20}{10} = 2$$

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Some Questions

The following questions reference a normal distribution with a mean μ = 150 lbs, a variance σ^2 = 100 lbs^2, and a standard deviation σ = 10 lbs. Such a distribution is often indicated by the symbols $N(\mu,\sigma) = N(150, 10)$.

1. What is the likelihood that a randomly selected individual observation is within 5 lbs of the population mean μ = 150lbs?

2. What is the likelihood that a mean from a random sample of size n = 5 is within 5 lbs of μ = 150 lbs?

3. What is the likelihood that a mean from a random sample of size n = 20 is within 5 lbs of μ = 150 lbs?

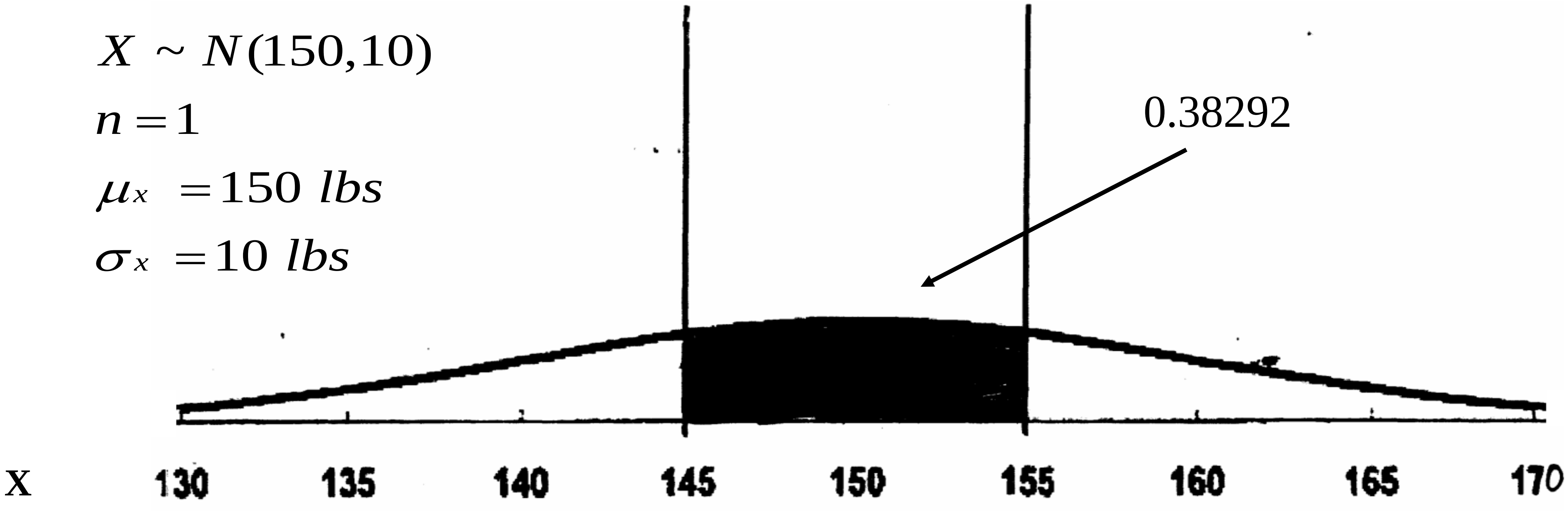
Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

$$X \sim N(150,10)$$

$$n = 1$$

$$\mu_x = 150\ lbs$$

$$\sigma_x = 10\ lbs$$



$$z_{Upper} = \frac{x_{Upper} - \mu_x}{\sigma_x} = \frac{155-150}{10} = 0.5$$, Area up to z $_{\text{upper}}$ = 0.69146

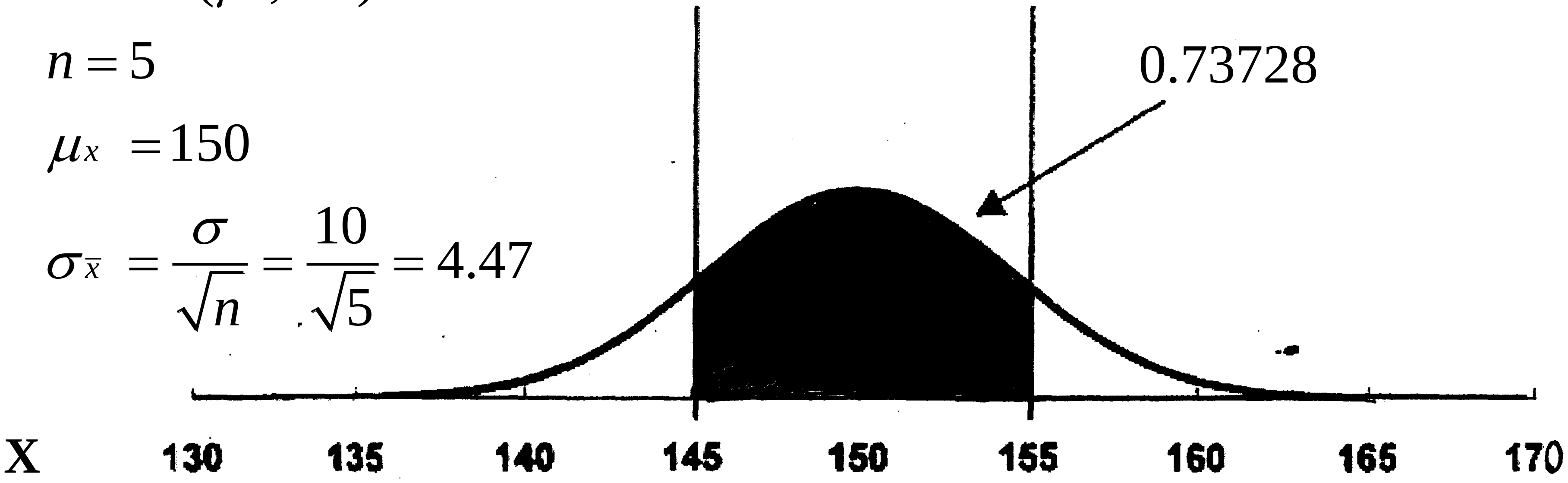
$$z_{Lower} = \frac{x_{Lower} - \mu_x}{\sigma_x} = \frac{145-150}{10} = -0.5$$, Area up to z $_{\text{lower}}$ = 0.30854

Area between z $_{\text{upper}}$ and z $_{\text{lower}}$ = 0.38292

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

Solution to Question 2

$$X \sim N(\mu_x, \sigma_x)$$
$$n = 5$$
$$\mu_x = 150$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{5}} = 4.47$$



$$z_{Upper} = \frac{x_{Upper} - \mu_x}{\sigma_x} = \frac{155-150}{4.47} = 1.12$$, Area up to z $_{upper}$ = 0.86864

$$z_{Lower} = \frac{x_{Lower} - \mu_x}{\sigma_x} = \frac{145-150}{4.47} = -1.12$$, Area up to z $_{lower}$ = 0.13136

Area between z $_{upper}$ and z $_{lower}$ = 0.73728

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt

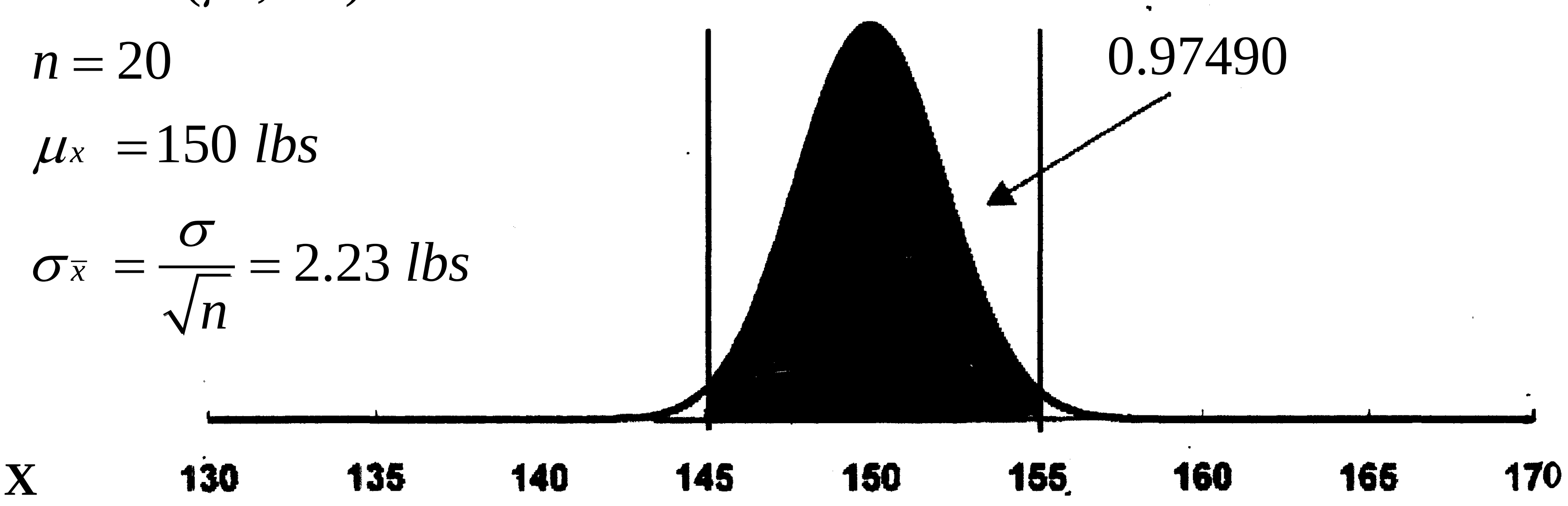
Solution to Question 3

$$X \sim N(\mu_x, \sigma_x)$$

$$n = 20$$

$$\mu_x = 150\ lbs$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 2.23\ lbs$$

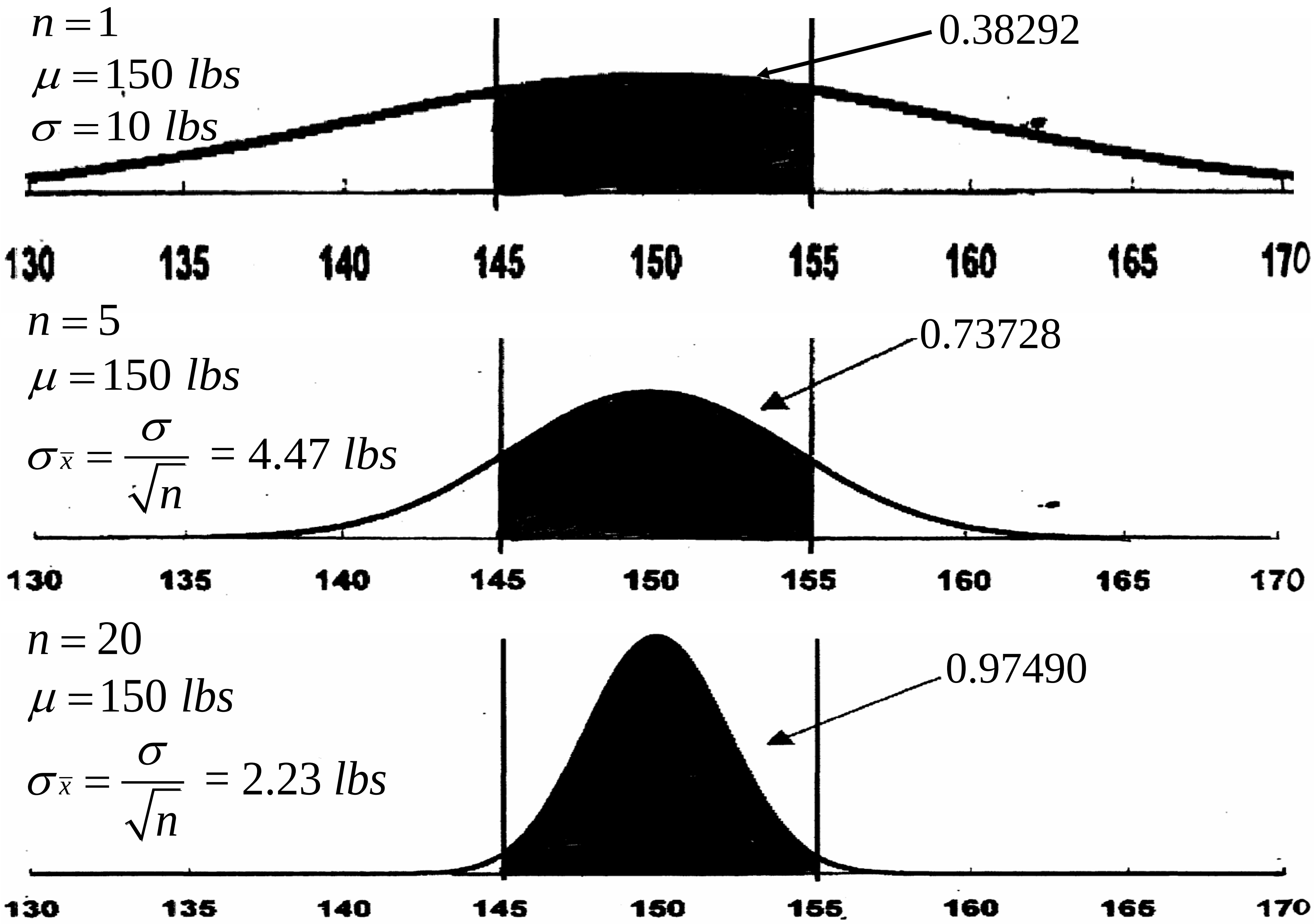


$$z_{Upper} = \frac{x_{Upper} - \mu_x}{\sigma_x} = \frac{155 - 150}{2.23} = 2.24$$, Area up to z upper = 0.98745

$$z_{Lower} = \frac{x_{Lower} - \mu_x}{\sigma_x} = \frac{145 - 150}{2.23} = -2.24$$, Area up to z lower = 0.01255

Area between z upper and z lower = 0.97490

Source : http://biostatcourse.fiu.edu/PPT/MODULE%2013%20Normal%20Distribution.ppt



$n = 1$
$\mu = 150\ lbs$
$\sigma = 10\ lbs$
0.38292
130 135 140 145 150 155 160 165 170
$n = 5$
$\mu = 150\ lbs$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 4.47\ lbs$
0.73728
130 135 140 145 150 155 160 165 170
$n = 20$
$\mu = 150\ lbs$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 2.23\ lbs$
0.97490
130 135 140 145 150 155 160 165 170