Continuous Probability Distributions

7. The Normal Distribution

Random variables can be either **discrete or continuous**. Discrete variables and their distributions were explained in lecture 6. Recall that a discrete variable cannot assume all values between any two given values of the variables. On the other hand, a continuous variable can assume all values between any two given values of the variables. Examples of continuous variables are the heights of adult men, body temperatures , and cholesterol levels of adults. Many continuous variables, such as the examples just mentioned, have distributions that are bell-shaped, and these are called ***approximately normally distributed variables.***

For example, if a researcher selects a random sample of 100 adult women, measures their heights, and constructs a histogram, the researcher gets a graph like the one shown in Figure 1(a). Now, if the researcher increases the sample size and decreases the width of the classes, the histograms will look like the ones shown in Figure 1(b) and (c). Finally, if it were possible to measure exactly the heights of all adult females in the United States and plot them, the histogram would approach what is called a ***normal distribution***, shown in Figure 1(d).

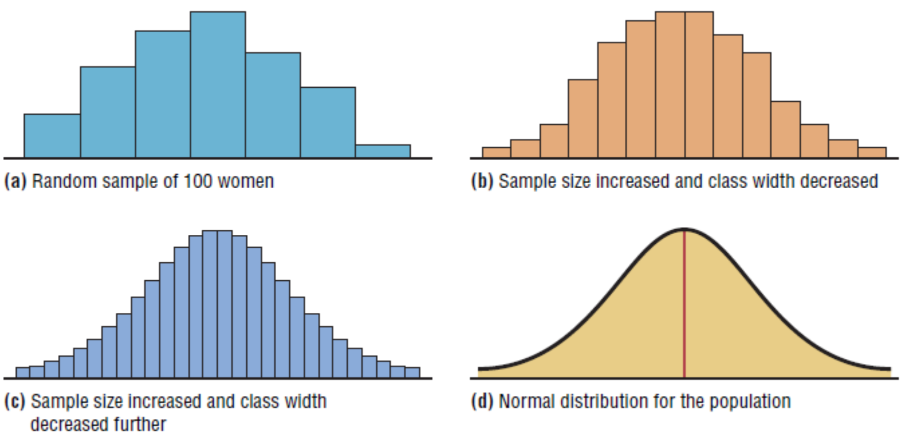


Figure 1. Histograms for the Distribution of Heights of Adult Women

This distribution is also known as **a bell curve or a Gaussian distribution**, named for the German mathematician Carl Friedrich Gauss (1777–1855), who derived its equation.

When the data values are evenly distributed about the mean, a distribution is said to be a symmetric distribution. (A normal distribution is symmetric.) Figure 2(a) shows a symmetric distribution. When the majority of the data values fall to the left or right of the mean, the distribution is said to be skewed. When the majority of the data values fall to the right of the mean, the distribution is said to be a negatively or left-skewed distribution. The mean is to the left of the median, and the mean and the median are to the left of the mode. See Figure 2(b). When the majority of the data values fall to the left of the mean, a distribution is said to be a positively or right-skewed distribution. The mean falls to the right of the median, and both the mean and the median fall to the right of the mode. See Figure 2(c).

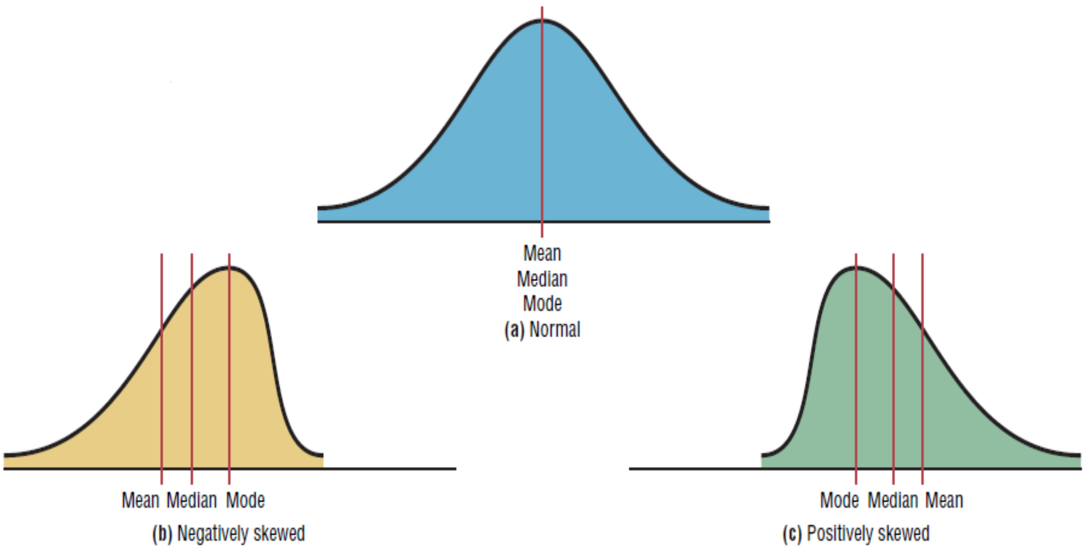


Figure 2. Normal and Skewed Distributions

- **Continuous variable** are variables that can assume to take all values between any two given values of the variables. For examples: the heights of adult men, body temperatures of rats, ground water level, and cholesterol levels of adults.

7.1. Equation of normal distribution

In mathematics, curves can be represented by equations. For example, the equation of the circle is **($x^2 + y^2 = r^2$)** , where r is the radius. A circle can be used to represent many physical objects, such as a wheel or a gear. Even though it is not possible to manufacture a wheel that is perfectly round, the equation and the properties of a circle can be used to study many aspects of the wheel, such as area, velocity, and acceleration. In a similar manner, the theoretical curve, called a normal distribution curve, can be used to study many variables that are not perfectly normally distributed but are nevertheless approximately normal.

The mathematical equation for a normal distribution is

$$\mathbf{y = \frac{e^{-(X-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}}$$

Where:

$e \approx 2.718$ (means ≈ is approximately equal to”)

μ is population mean

σ is population standard deviation

The shape and position of a 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. Figure 3(a) shows two normal distributions with the same mean values but different standard deviations. The larger the standard deviation, the more dispersed, or spread out, the distribution is. Figure 3(b) shows two normal distributions with the same standard deviation but with different means. These curves have the same shapes but are located at different positions on the x axis. Figure 3(c) shows two normal distributions with different means and different standard deviations.

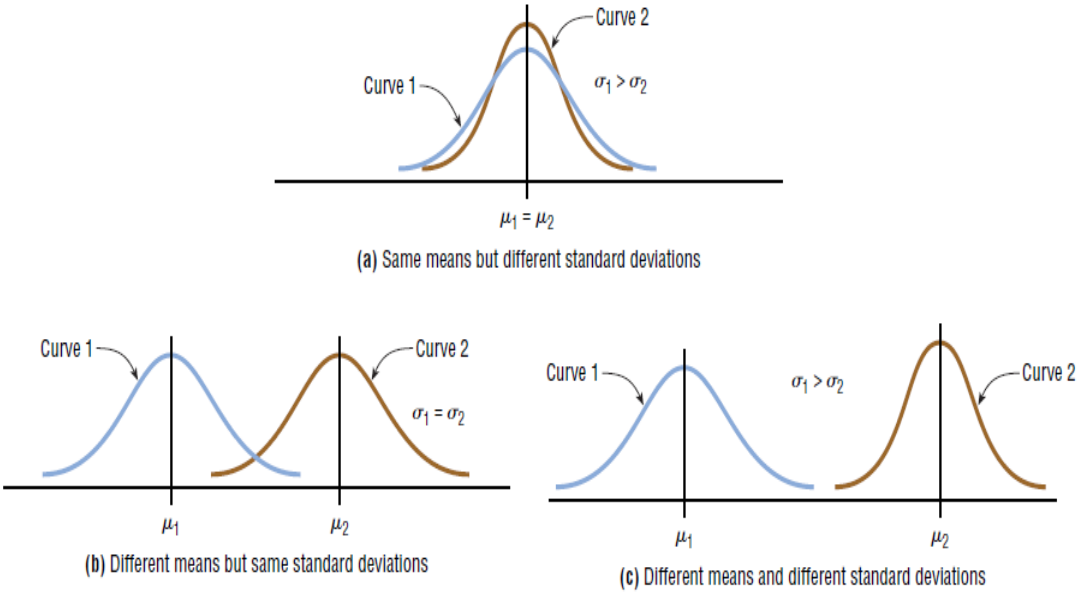


Figure 3. Shapes of Normal Distributions

7.2. Summary of the Properties of the Theoretical Normal Distribution

1. A normal distribution curve is bell-shaped.

2. The mean, median, and mode are equal and are located at the center of the distribution.

3. The curve is symmetric 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.

4. The curve is continuous.

5. The curve never touches the *x* axis. The total area under a normal distribution curve is equal to 1.00, or 100%.

6. The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%; and within 3 standard deviations, about 0.997, or 99.7%.

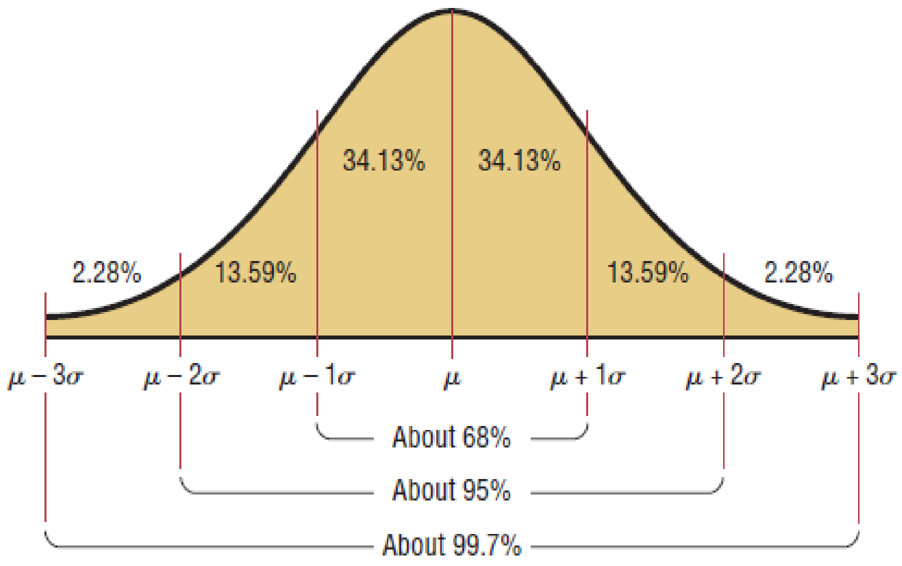


Figure 4. Areas Under a Normal Distribution Curve

7.3. The Standard Normal Distribution

Since each normally distributed variable has its own mean and standard deviation. So, the shape and location of these curves will vary. In practical applications, statisticians used what is called the **standard normal distribution**. Then, a table can be used to determine the area under the curve for each variable. The standard normal distribution is shown in Figure 5 below.

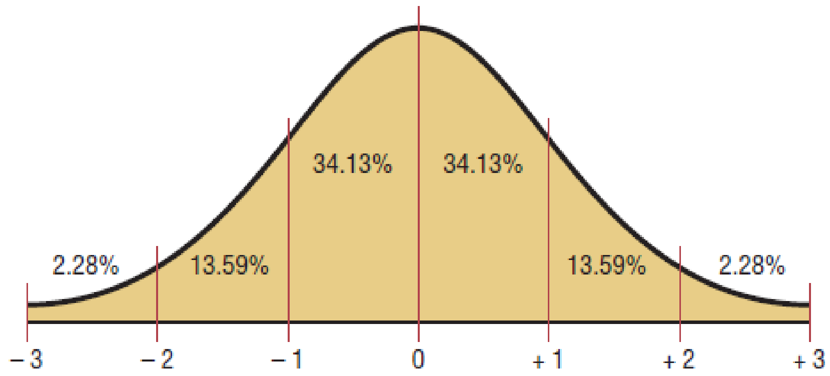


Figure 5. Standard Normal Distribution

The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1. The formula for the standard normal distribution is

$$y = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

All normally distributed variables can be transformed into the standard normally distributed variable using the formula for the standard score:

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \quad \text{or} \quad z = \frac{X - \mu}{\sigma}$$

The values under the curve indicate the proportion of area in each section. For example, the area between the mean and 1 standard deviation above or below the mean is about 0.3413, or 34.13%.

7.4. Finding Areas Under the Standard Normal Distribution Curve

The area under a normal distribution curve is used to finding the probability of the continuous variables for any range would be found. A two-step process is recommended with the use of the Procedure Table shown.

Step 1: Draw the normal distribution curve and shade the area.

Step 2: Find the appropriate figure in the Procedure Table.

Example: find the area under Z = 1.39

Figure 6 . Table of Z for determine the area under the curve

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
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
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

For z values greater than 3.49, use 0.9999.

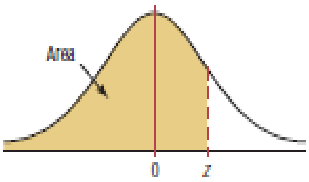
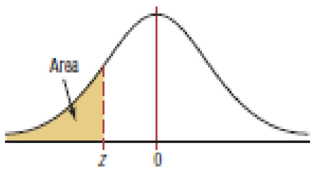


Table E The Standard Normal Distribution

Cumulative Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−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
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

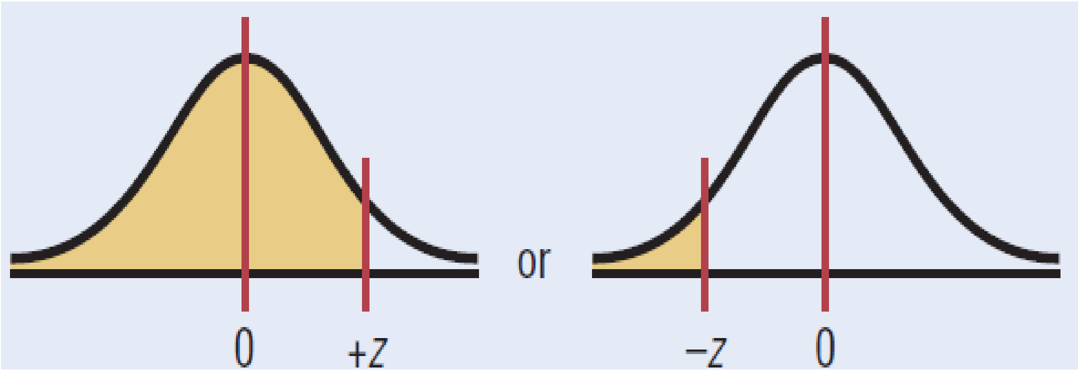
For z values less than −3.49, use 0.0001.



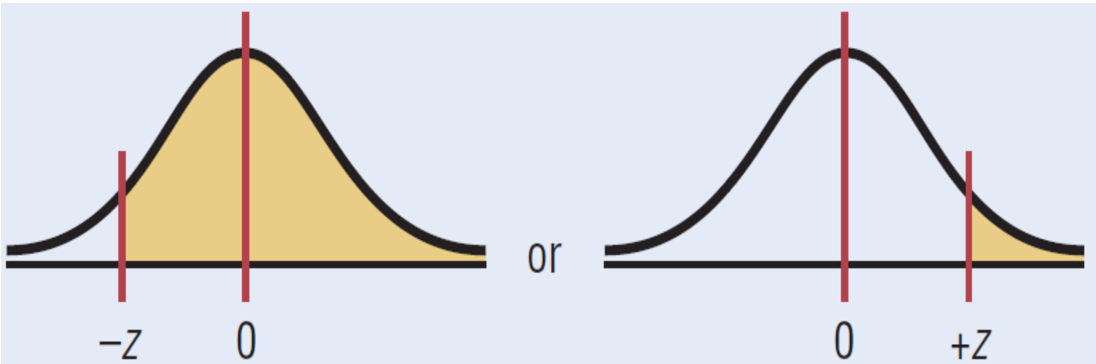
7.4.1. Procedure Table

Finding the Area Under the Standard Normal Distribution Curve

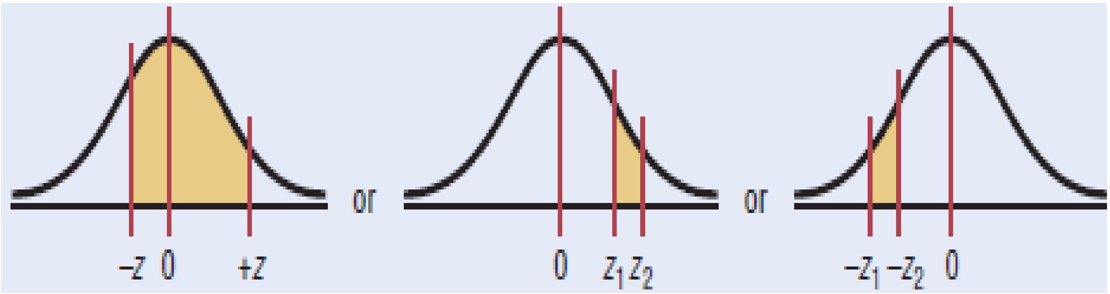
1. To the left of any z value: Look up the z value in the table and use the area given



2. To the right of any z value: Look up the z value and subtract the area from 1. (1- area)



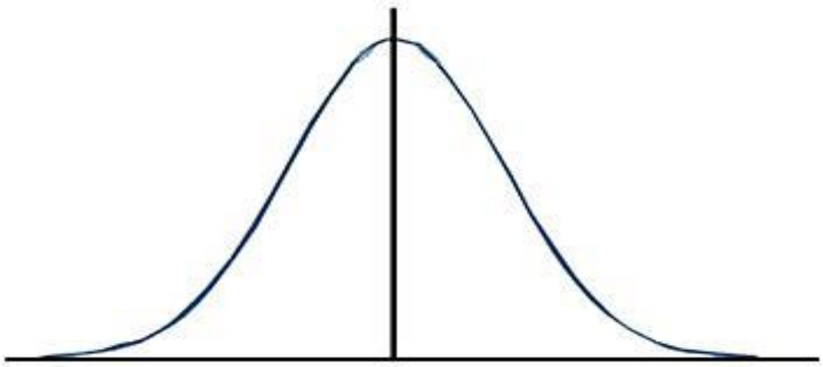
3. Between any two z values: Look up both z values and subtract the corresponding areas. **(Area$_{z2}$ – Area$_{z1}$)**



Example 1: Find the area to the left of $z = 2.06$.

Solution

Step 1 Draw the figure. The desired area is shown in the figure below.

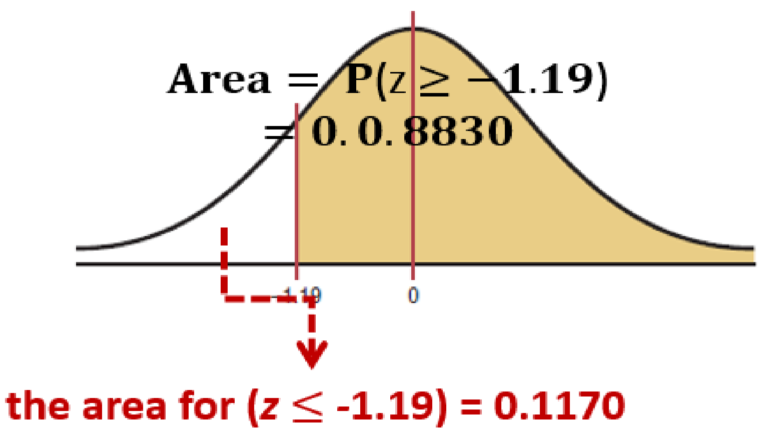


Step 2 We are looking for the area under the standard normal distribution to the left of $z = 2.06$. Since this is an example of **<u>the first case</u>**, look up the area in the table. It is **0.9803.** Hence, **98.03%** of the area is less than **z = 2.06**.

Example 2: Find the area to the right of $z = -1.19$.

Solution

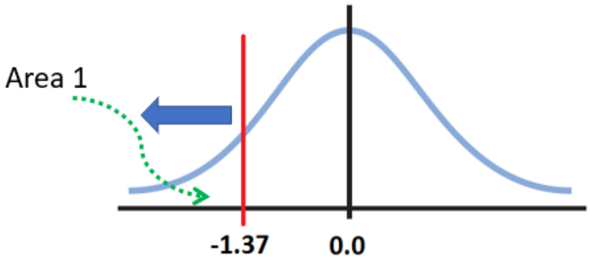
Step 1 Draw the figure. The desired area is shown in the figure.

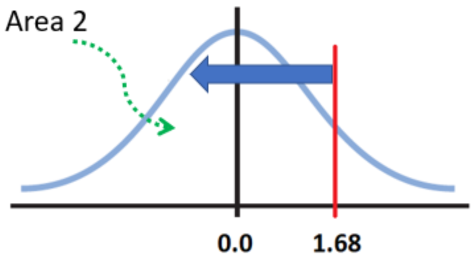


Example 3: Find the area between $z = + 1.68$ and $z = - 1.37$.

<u>Solution</u>

This is case 3. Draw the figure as shown below . The desired area is shown in the figures below





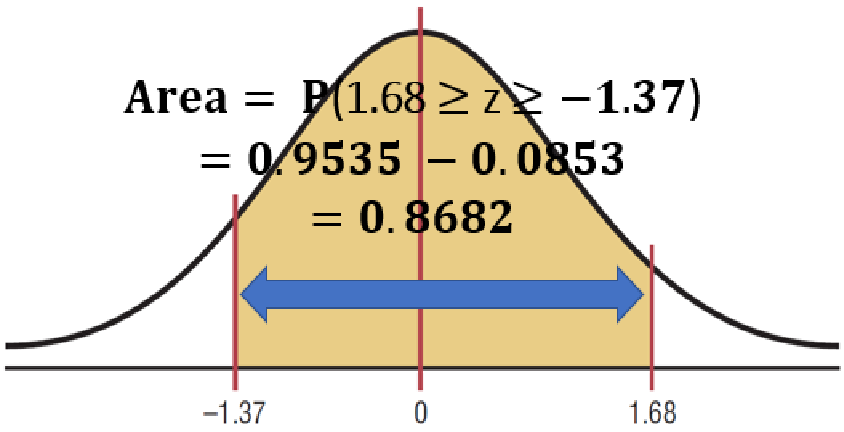
for the small area z = -1.37,
from table area = 0.0853

for the large area z = 1.68,
from table area = 0.9535

The area between the two z values is:

= 0.9535 - 0.0853

= 0.8682 or 86.82%.



7.5. A Normal Distribution Curve as a Probability Distribution Curve

The area under the standard normal distribution curve can be used for calculation the probability for any continuous random variable.

Example 4: Find the probability for each. a) $P(0 < z < 2.32)$; b) $P(z < 1.65)$; c) $P(z > 1.91)$

Solution

a)P(0 < z < 2.32) = P(z< 2.32) – P(z<0) OR

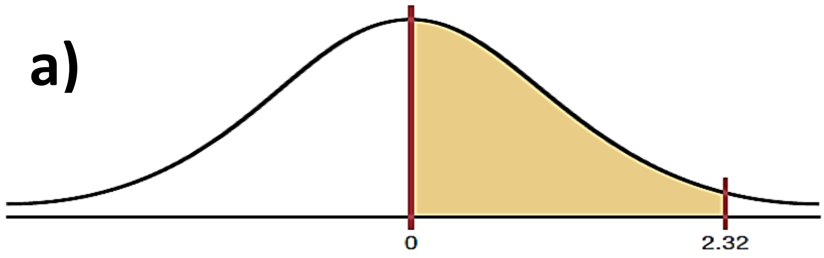
area to the left of(z = 2.32)

– area to the left of (z = 0)

P (z < 2.32) = the area from the table = 0.9898

P (z< 0.0) = the area from the table = 0.5000.

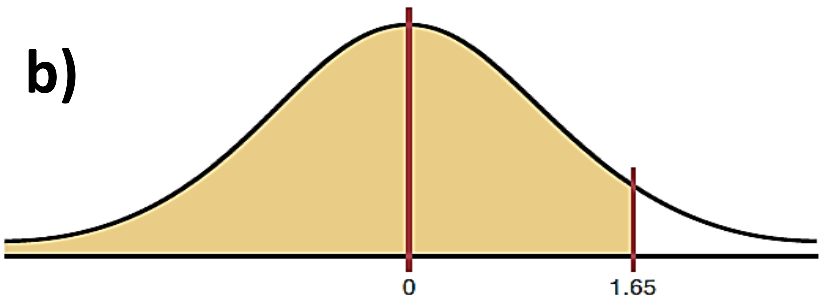
∴ P (0 < z < 2.32) = 0.9898 – 0.500 = 0.4898.



b) P(*z* < 1.65) is the area from the table

to the left of Z = 1.65.

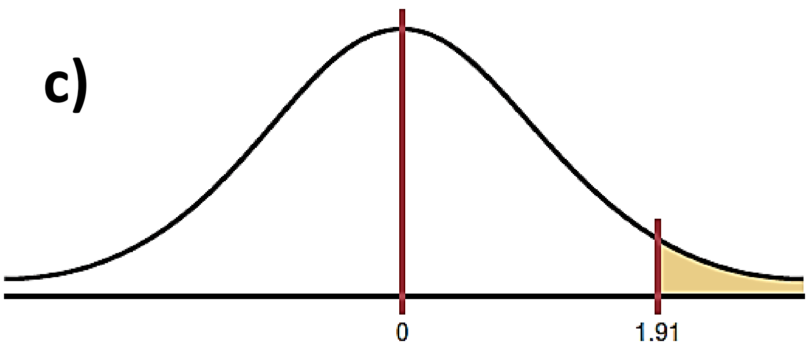
P(z < 1.6) = 0.9505



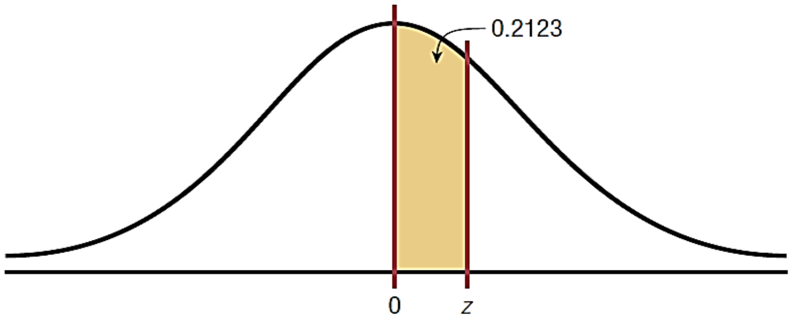
c) P(z > 1.91) = 1 – P (z < 1.91)

= 1 – the area to the left of 1.91

= 1 – 0.9719 = 0.0281, or 2.81%.



Example 5: Find the z value such that the area under the standard normal distribution curve between 0 and the z value is 0.2123.



Solution

Draw the figure. The area is shown in the figure.

The total area = the area of (z = 0) + 0.2123

= 0.5000 + 0.2123 = 0.7123

From the table. The value in the left column is 0.5, and the top value is 0.06. Add these two values to get $\boldsymbol{z = 0.56}$.

7.6. Applications of the Normal Distribution

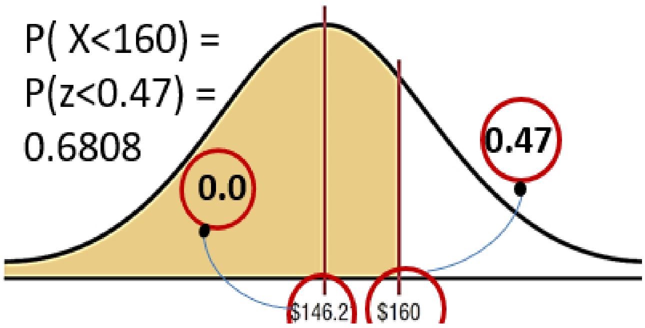
The standard normal distribution curve can be used to solve a wide variety of practical problems. To solve problems by using the standard normal distribution, transform the original variable to a standard normal distribution variable (Z) using the formula:

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

Example 6: A survey found that women spend on average \$146.21 on beauty products during the summer months. Assume the standard deviation is \$29.44. Find the percentage (Probability) of women who spend less than \$160.00. Assume the variable is normally distributed.

Solution

Draw the figure and represent the area as shown in the figure. Then Find the *z* value corresponding to \$160.00.



$$z = \frac{X-\mu}{\sigma} = \frac{160.00-146.21}{29.44} = 0.47$$

From the table: P(X<160) = P(z<0.47) = area to the left of z = 0.6808, or 68.08% (percent of the women spend less than \$160.00 on beauty products).

Example 7: Each month, an American household generates an average of 28 pounds of newspaper for garbage or recycling. Assume the standard deviation is 2 pounds. If a household is selected at random, find the probability of its generating.

a) Between 27 and 31 pounds per month.

b) More than 30.2 pounds per month.

Solution

a) Draw the figure and represent the area. Then find the two *z* values.

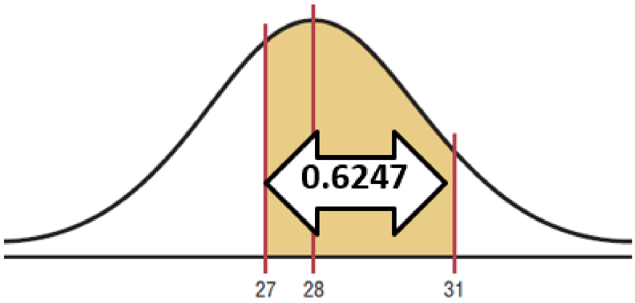
$P(27<X<31) = P(Z_1<z<Z_2)$

$z1 = \frac{X1-\mu}{\sigma} = \frac{27-28}{2} = -0.5$

$z2 = \frac{X2-\mu}{\sigma} = \frac{31-28}{2} = 1.5$

$Area_1 = 0.3085$ and $Area_2 = 0.9332$

$P(27<X<31) = P(-0.5<z<1.5) = 0.9332 - 0.3085 = 0.6247$



b) Draw the figure and represent the area. Then find the two z values.

$P(X>30.2)= P(Z>z_1)= 1-P(X<30.2) = 1 - P(Z<z_1)$

$z1 = \frac{X-\mu}{\sigma} = \frac{30.2-28}{2} = 1.1$ and $Area_1 = 0.8643$

$P(X > 30.2) = 1 - P(X < 30.2) = 1 - P(Z<z_1) = 1 - 0.8643 = 0.1357$ or 13.57%

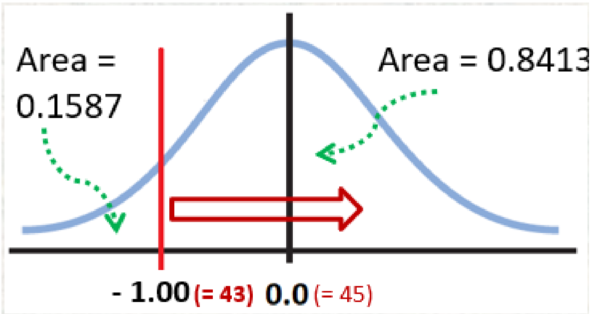
Example 8: A steel factory produces deformed bars with average yield force 45 kN and standard deviation 2 kN, if a bare has been tested, determine the probability of? (a) strength force $\geq$ 43 kN; (b) strength force $\leq$ 47 kN; and (c) strength force between 44 to 46 kN.

Solution

(a) $\mathbf{P(X \geq 43) = P\left((z \geq \frac{X-\mu}{\sigma}\right)} = 1 - P\left((z < \frac{X-\mu}{\sigma}\right)$ = area to the right

$z = \frac{43-45}{2}$ = - 1.00 , then Area to the left = 0.1587

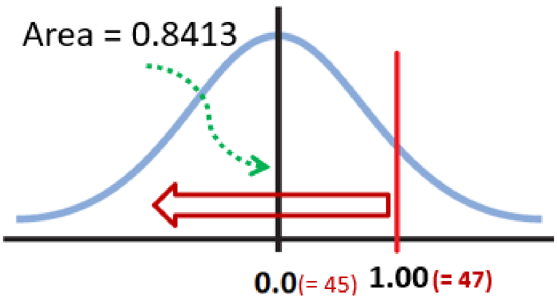
$P(X \geq 43) = P\left((z \geq \frac{X-\mu}{\sigma}\right) = 1 - P\left(((z < \frac{X-\mu}{\sigma}\right) = 1 - 0.1587 = 0.8413$



(b) $\mathbf{P(X \leq 47) = P\left((z \leq \frac{X-\mu}{\sigma}\right)}$

$z = \frac{47-45}{2}$ = 1.00 , then Area to the left= 0.8413

$P(X \leq 47) = P\left((z \leq \frac{X-\mu}{\sigma}\right)$ =0.8413 or 84.13 %



(c) $\mathbf{P(44 \leq X \leq 47) = P((z1 \leq Z \leq z_2)}$

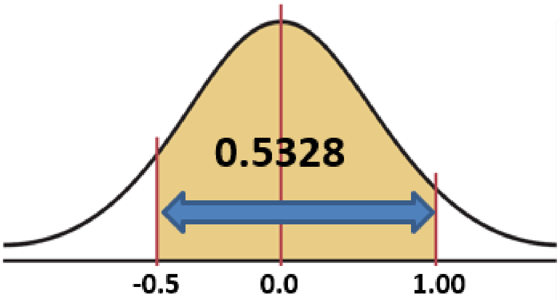
$z1 = \frac{44-45}{2}$ = -0.500, Area to the left = 0.3085

$z2 = \frac{47-45}{2}$ = 1.00 , Area to the left = 0.8413

$P(44 \leq X \leq 47) = P((z1 \leq Z \leq z_2)$

$= 0.8413 – 0.3085$

$= 0.5328$ OR 53.28 %



Example 9: An engineering in PVC pipe factory wishes to select a pipe bearing pressure in the middle 60%. If the mean hydraulic pressure is 120 and the standard deviation is 8, find the upper and lower pressure that meet the requirement.

Solution:

The two values (X1 and X2) must be determined based on the area to the left side of each values

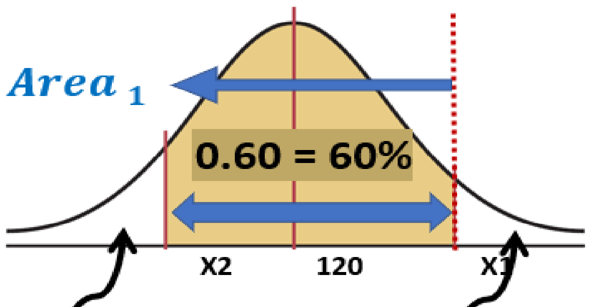
- From Table; $Area_2 = 0.2$, $z_2 = -0.84$
- From Table; $Area_1 = 0.8$, $z_1 = 0.84$

$X1 = z1.\sigma + \mu \Rightarrow \quad X_1 = 0.84*8 + 120 = 126.72$

$X2 = z2.\sigma + \mu \Rightarrow \quad X_2 = -0.84 \times 8 + 120 = 113.28$

Therefore, the middle 60% will have pressure readings of**: $\mathbf{113.28 \leq X \leq 126.72}$.**

Show the following figure.



$$Area_2 = \frac{1-0.6}{2} = 0.2 = 20\%$$

$$Area_1 = 0.6 + 0.2 = 0.8$$

7.7. Determining Normality

The distribution is being normally or approximately normally shaped:

- The easiest way is to draw a histogram for the data and check its shape. If the histogram is not approximately bell shaped, then the data are not normally distributed.
- Skewness coefficient (Pearson's index (PC))

$$PC = \frac{3(\bar{X} - M_e)}{S}$$

The Normality distribution: $-1 \leq PC \leq +1$

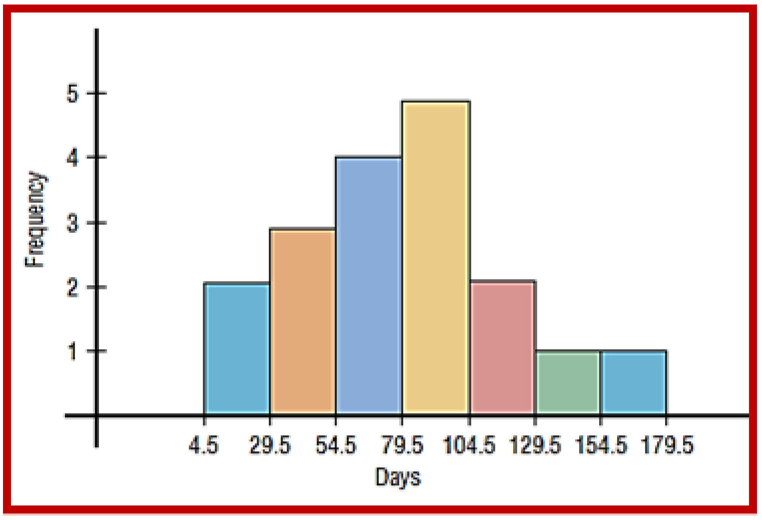
Example 11: A survey of 18 high-technology firms showed the number of days' inventory they had on hand. Determine if the data are approximately normally distributed.

5	29	34	44	45	63	68	74	74
81	88	91	97	98	113	118	151	158

Solution

- Construct a frequency distribution table and draw a histogram for the data.
- The histogram is approximately bell-shaped, we can say that the distribution is approximately normal.

Class	Frequency
5–29	2
30–54	3
55–79	4
80–104	5
105–129	2
130–154	1
155–179	1



Using PC to check the normality

(average = 79.5, median = 77.5, and S = 40.5)

$PC = \frac{3(\bar{X} - M_e)}{S} = \frac{3(79.5 - 77.5)}{40.5}$ = 0.148 within

-1 ≤ PC ≤ +1, it is normal distribution

Homework

To qualify of a steel factory quality, a tensile strength must score in the top 10% on a general test. The tensile mean is 200 and a standard deviation of 20. Find the lowest possible tensile strength to qualify. Assume the test scores are normally distributed.