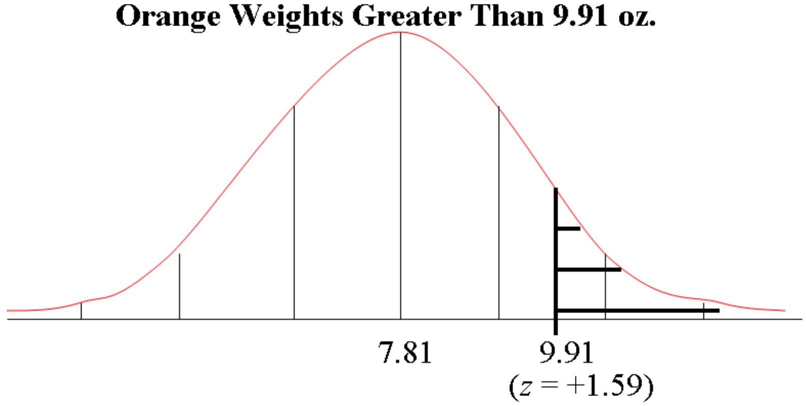
Unit 16
Normal Distributions

Objectives:

- To obtain relative frequencies (probabilities) and percentiles with a population having a normal distribution

While there are many different types of distributions that a population may have (see Figures 15-1a to 15-1f), normal distributions are of particular importance in statistics. One of the primary reasons normal distributions are so important is because, as we have seen previously, the sampling distribution of $\bar{x}$ with simple random sampling takes on the properties of a normal distribution, with a sufficiently large sample size *n*. Figure 15-2 gave us a somewhat detailed description of a normal distribution in terms of one, two, and three standard deviations away from the mean. We have seen that a normal distribution is symmetric and bell-shaped, and that practically all of the items in a normally distributed population are within three standard deviations of the mean.

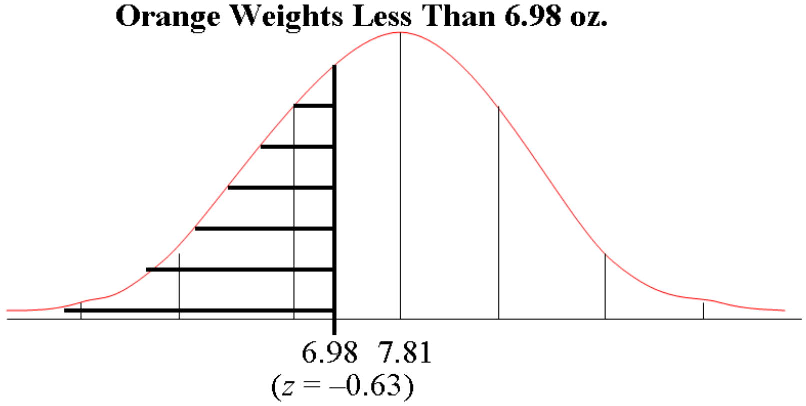
Figure 16-1
Orange Weights Greater Than 9.91 oz.



We now want to describe normal distributions in a much more detailed manner than merely in terms of one, two, and three standard deviations away from the mean. Table A.2 in the appendix provides areas under a normal density curve in terms of *z*-scores. A normal curve described entirely in terms of *z*-scores is called a *standard normal curve*. By converting raw scores to *z*-scores, we can use Table A.2 to find many different areas under any normal density curve. These areas under a normal density curve can be interpreted as relative frequencies or as probabilities. In addition to using Table A.2, statistical software packages, spreadsheets, programmable calculators, etc. can also be used to find areas under a normal density curve.

The figure at the top of the Table A.2 indicates that you can use this table directly to find the area under a normal curve above any value greater than the mean μ. Such an area is found by obtaining the *z*-score of a value greater than the mean and reading the corresponding area in the body of the table. The *z*-scores in the table are all to two decimal place accuracy, with the *z*-scores to the first decimal place displayed as the row labels, and the column labels providing the second decimal place. The areas in the body of the table are all given to four decimal place accuracy.

Figure 16-2
Orange Weights Less Than 6.98 oz.

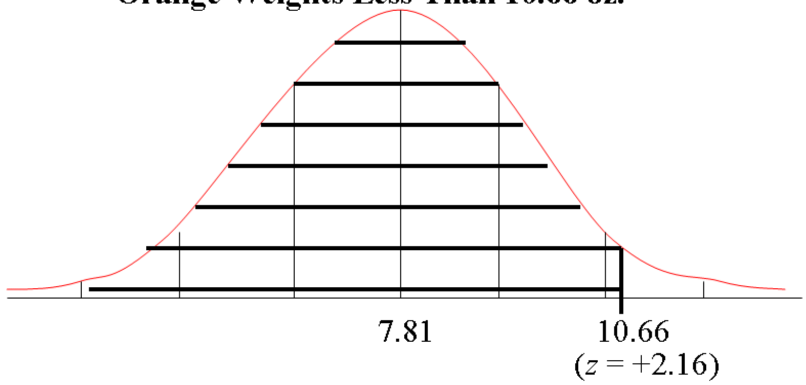


To illustrate the use of Table A.2, we shall consider several examples involving the population of weights of oranges from a particular grove. Let us suppose that the weights of oranges from the grove have a normal distribution with mean μ = 7.81 oz. and standard deviation σ = 1.32 oz. We shall consider finding the percentage of orange weights that lie in a given range, or, in other words, the probability that one randomly selected orange has a weight in the given range. These relative frequencies (or probabilities) will of course correspond to an appropriate area under a normal density curve.

To begin, we shall find the percentage, or relative frequency, of oranges that weigh more than 9.91 oz., which can also be interpreted as the probability that one randomly selected orange has a weight more than 9.91 oz. To find this probability, we first use $\mu = 7.81$ and $\sigma = 1.32$ to find the z-score of 9.91 oz. as follows:

$$\frac{9.91 - 7.81}{1.32} = +1.59 \ .$$

Figure 16-3
Orange Weights Less Than 10.66 oz.



The proportion of shaded area in Figure 16-1 is the desired probability. This shaded area corresponds exactly to the shaded area in the figure at the top of Table A.2. The desired area is found directly from Table A.2 in the row labeled 1.5 and the column labeled 0.09. The probability that one randomly selected orange weighs more than 9.91 oz. is 0.0559 (or 5.59%).

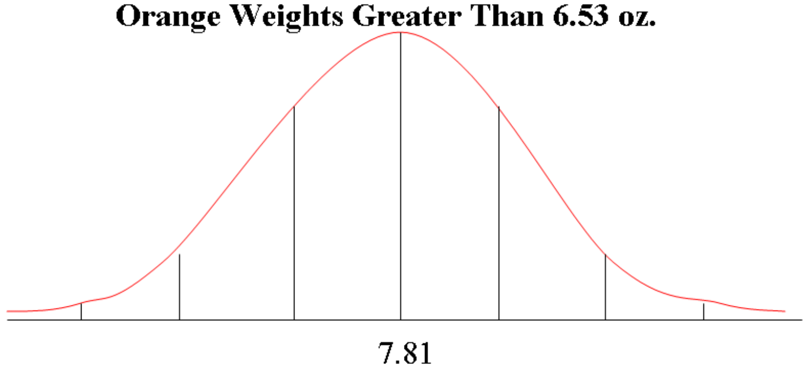
Next, we shall find the percentage, or relative frequency, of oranges that weigh less than 6.98 oz., which can also be interpreted as the probability that one randomly selected orange has a weight less than 6.98 oz. To find this probability, we first use $\mu = 7.81$ and $\sigma = 1.32$ to find the z-score of 6.98 oz. as follows:

$$\frac{6.98 - 7.81}{1.32} = -0.63 \ .$$

The proportion of shaded area in Figure 16-2 is the desired probability. This shaded area is a mirror image of the shaded area in the figure at the top of Table A.2. Since a normal curve is symmetric, the desired area is found directly from Table A.2 in the row labeled 0.6 and the column labeled 0.03. The probability that one randomly selected orange weighs less than 6.98 oz. is 0.2643 (or 26.43%).

We cannot always obtain the desired area under the standard normal curve directly from Table A.2. It is sometimes necessary to use the fact that the total area under the standard normal curve is one (or 100%) to find desired areas under a standard normal curve.

Figure 16-4
Orange Weights Greater Than 6.53 oz.



To illustrate, we shall find the percentage, or relative frequency, of oranges that weigh less than 10.66 oz., which can also be interpreted as the probability that one randomly selected orange has a weight less than 10.66 oz. To find this probability, we first use $\mu = 7.81$ and $\sigma = 1.32$ to find the z-score of 10.66 oz. as follows:

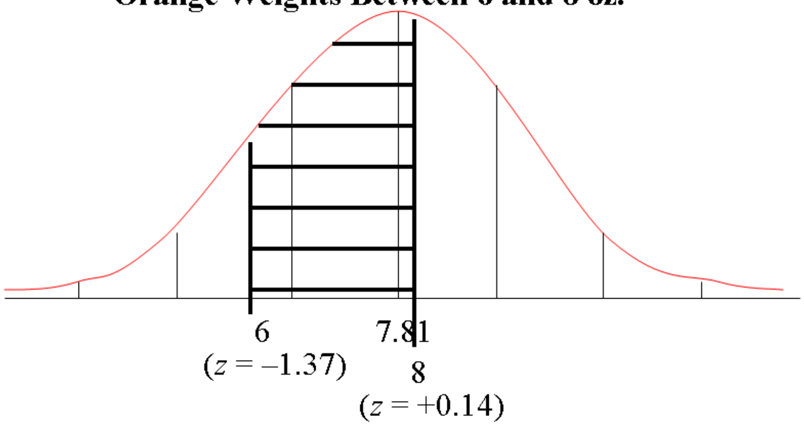
$$\frac{10.66 - 7.81}{1.32} = +2.16 \ .$$

The proportion of shaded area in Figure 16-3 is the desired probability, but this shaded area does not correspond to the shaded area in the figure at the top of Table A.2; however, the unshaded area in Figure 16-3 does correspond exactly to the shaded area in the figure at the top of Table A.2. This unshaded area is then found directly from Table A.2 in the row labeled 2.1 and the column labeled 0.06; since the total area under a normal curve is equal to 1, the desired shaded area in Figure 16-3 is found by subtracting this entry of Table B.2 from 1.

The probability that one randomly selected orange weighs less than 10.66 oz. is 1 – 0.0154 = 0.9846 (or 98.46%).

As another illustration, we shall have you find the percentage, or relative frequency, of oranges that weigh more than 6.53 oz., which can also be interpreted as the probability that one randomly selected orange has a weight more than 6.53 oz. Find the *z*-score for 6.53 oz., and label the value 6.53 oz. on the horizontal axis in Figure 16-4. Then, shade the desired area under the normal curve, and use Table A.2 to obtain the desired probability. (You should find that the *z*-score is –0.97, and that the probability that one randomly selected orange weighs more than 6.53 oz. is 0.8340 (or 83.40%).)

Figure 16-5
Orange Weights Between 6 and 8 oz.



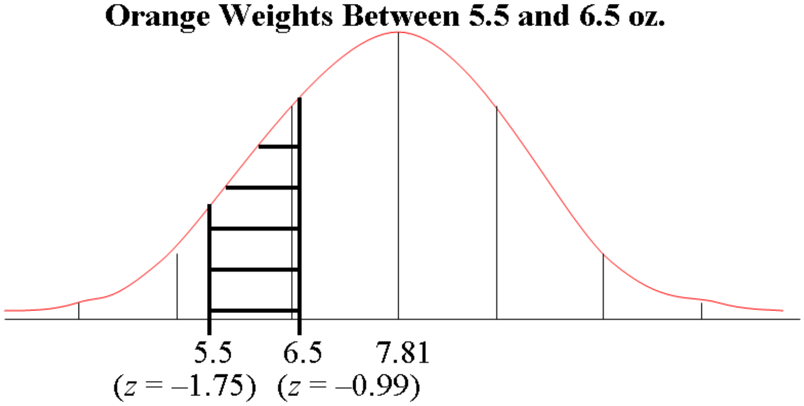
If the desired area under the standard normal curve is in between two given values, then we need to read Table A.2 twice. Let us obtain the percentage, or relative frequency, of oranges that weigh between 6 and 8 oz., which can also be interpreted as the probability that one randomly selected orange has a weight between 6 and 8 oz. To find this probability, we first use $\mu = 7.81$ and $\sigma = 1.32$ to find the *z*-score of 6 oz. to be

$$\frac{6-7.81}{1.32} = -1.37 \ ,$$

and to find the *z*-score of 8 oz. to be

$$\frac{8-7.81}{1.32} = +0.14 \ .$$

Figure 16-6
Orange Weights Between 5.5 and 6.5 oz.



The proportion of shaded area in Figure 16-5 is the desired probability. The unshaded area below 6 oz. in Figure 16-5 is a mirror image of the shaded area in the figure at the top of Table A.2; also, the unshaded area above 8 oz. in Figure 16-5 corresponds exactly to the shaded area in the figure at the top of Table A.2. We find the total unshaded area by adding the entry of Table A.2 in the row labeled 1.3 and the column labeled 0.07 to the entry of Table A.2 in the row labeled 0.1 and the column labeled 0.04; we then obtain the desired area by subtracting this unshaded area from 1. The probability that one randomly selected orange weighs between 6 and 8 oz. is 1 – (0.4443 + 0.0853) = 0.4704 (or 47.04%).

In the illustration just completed, the desired area under the standard normal curve was in between two values, where one was below μ, and the other was above μ. We shall now consider instances where the desired area under the standard normal curve is in between two values, where either both values are greater than μ, or both values are less than μ.

First, we obtain the percentage, or relative frequency, of oranges that weigh between 5.5 and 6.5 oz., which can also be interpreted as the probability that one randomly selected orange has a weight between 5.5 and 6.5 oz. To find this probability, we first use $\mu = 7.81$ and $\sigma = 1.32$ to find the *z*-score of 5.5 oz. to be

$$\frac{5.5-7.81}{1.32} = -1.75 \ ,$$

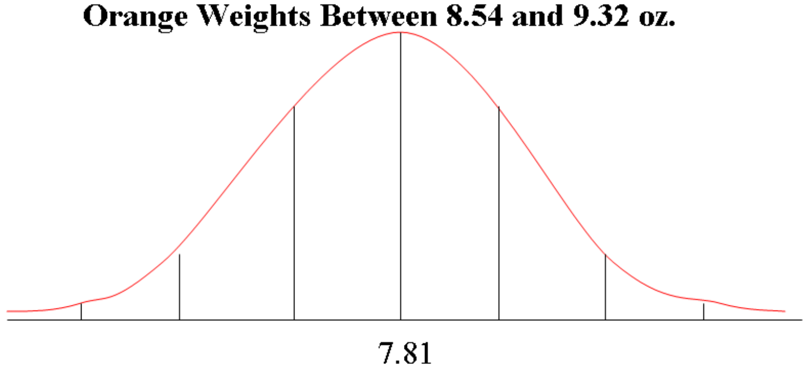
and to find the *z*-score of 6.5 oz. to be

$$\frac{6.5 - 7.81}{1.32} = -0.99 \ .$$

The proportion of shaded area in Figure 16-6 is the desired probability. The unshaded area below 5.5 oz. in Figure 16-6 is a mirror image of the shaded area in the figure at the top of Table A.2; also, if the shaded and unshaded areas below 6.5 oz. are combined together in Figure 16-6, this combined area is a mirror image of the shaded area in the figure at the top of Table A.2. We find the desired area by subtracting the entry of Table A.2 in the row labeled 1.7 and the column labeled 0.05 from the entry of Table A.2 in the row labeled 0.9 and the column labeled 0.09. The probability that one randomly selected orange weighs between 5.5 and 6.5 oz. is 0.1611 – 0.0401 = 0.1210 (or 12.10%).

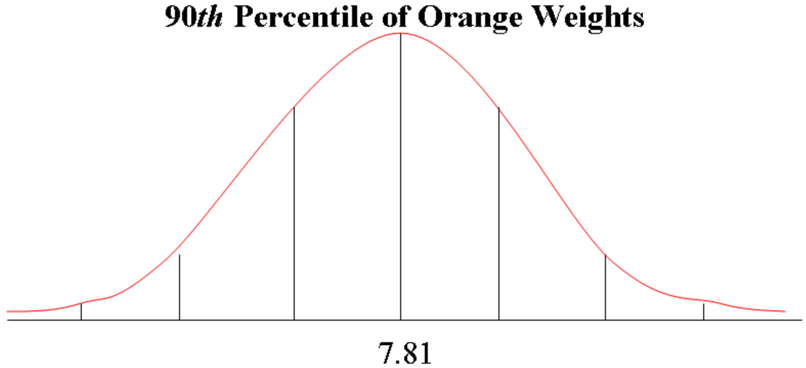
Now, we shall let you find the percentage, or relative frequency, of oranges that weigh between 8.54 and 9.32 oz., which can also be interpreted as the probability that one randomly selected orange has a weight between 8.54 and 9.32 oz. Find the *z*-score for each of 8.54 and 9.32 oz., and label the values for 8.54 and 9.32 oz. on the horizontal axis in Figure 16-7. Then, shade the desired area under the normal curve, and use Table A.2 to obtain the desired probability. (You should find that the *z*-scores are +0.55 and +1.14, and that the probability that one randomly selected orange weighs between 8.54 and 9.32 oz. is 0.1641 (or 16.41%).)

Figure 16-7
Orange Weights Between 8.54 and 9.32 oz.



Since practically all of the area under a normal curve is within three standard deviations of the mean, there are no *z*-scores in Table A.2 below –3.09 or above +3.09. When one orange weight is randomly selected from the population with mean μ = 7.81 oz. and standard deviation σ = 1.32 oz., we would consider the probability of observing a weight greater than 20 oz. to be practically zero (0), because the *z*-score of 20 oz. is +9.23. Of course, common sense might suggest to us that the likelihood of finding an orange weighing more than 20 oz. is extremely small! In a similar fashion, we would conclude that practically all oranges weigh more than 2 oz. (for which the z-score is – 4.40), or in other words, the probability of selecting an orange weighing more than 2 oz. is one (1).

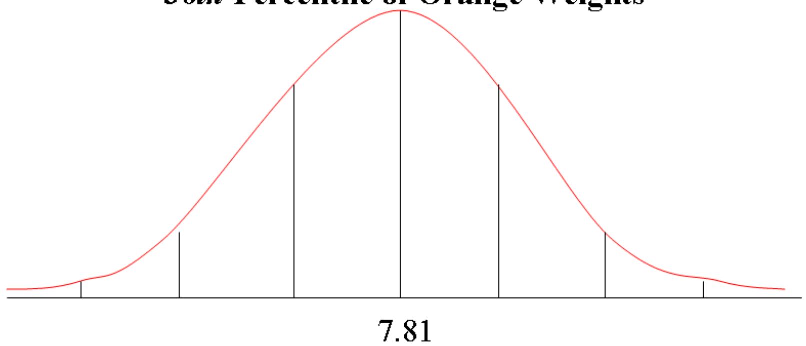
Figure 16-8
90*th* Percentile of Orange Weights



It is sometimes desirable to obtain percentiles and percentile ranks from a density curve. We define the *p*th percentile to be the value with *p*% of the total area below it and (100 – *p*)% of the total area above it; also, the percentage of area below a given value *x* is defined to be the percentile rank of *x*. To illustrate the use of Table B.2 in obtaining percentiles and percentile ranks with a normally distributed population, we shall return to the population of orange weights having a normal distribution with mean μ = 7.81 oz. and standard deviation σ = 1.32 oz.

Finding the percentile rank for a value *x* from a normally distributed population simply amounts to calculating the area under the normal density curve which lies below *x*. For instance, earlier we found the percentage of oranges weighing less than 6.98 oz. to be 26.43%; we can then say that the percentile rank of an orange weighing 6.98 oz. is 26.43 or 26. Often, the percent sign (%) is omitted when stating a percentile rank, since a percentile rank is understood to be a percentage. As another illustration, recall that we have previously found that the percentage of oranges weighing less than 10.66 oz. is 98.46%; we can then say that the percentile rank of an orange weighing 10.66 oz. must be 98.46 or 98.

Figure 16-9
30*th* Percentile of Orange Weights



Finding the *p*th percentile of a normally distributed population amounts to finding the value below which lies *p* percent of the area under the normal density curve. Since a normal curve is symmetric, the 50th percentile (i.e., the median) is equal to the mean μ. We determine other percentiles by using Table A.2 to find the *z*-score of the desired percentile and converting this *z*-score to a raw score. To illustrate, we shall obtain the 90th percentile of the orange weights and the 30th percentile of the orange weights.

To find the 90th percentile, we first realize that the 90th percentile must be greater than the median (which is the mean μ), implying that the *z*-score of the 90th percentile must be positive. We then use Table A.2 to determine the desired *z*-score, that is, the *z*-score of the orange weight below which lie 90% of all the orange weights and above which lie 10% of all the orange weights. Since the areas listed in Table A.2 are the areas under the normal curve above a positive *z*-score, by searching for the area closest to 0.10 (which is 0.1003), we find that the desired *z*-score is +1.28. In Figure 16-8, label the *z*-score +1.28 on the horizontal axis; if you draw a vertical line through the graph at the location of the *z*-score +1.28, you should be able to see that roughly 90% of the area under the normal curve lies below this *z*-score, and 10% of the area under the normal curve lies above this *z*-score. Recall that we use $x = \mu + z\sigma$ to convert a *z*-score to a raw score. Convert the *z*-score +1.28 to a raw score to obtain the 90th percentile of the orange weights. (You should find that the 90th percentile is approximately 9.50 oz.)

To find the 30th percentile of the orange weights, we must first realize that the 30th percentile must be smaller than the median (which is the mean μ), implying that the *z*-score of the 30th percentile must be negative. We then use Table A.2 to determine the desired *z*-score, that is, the *z*-score of the orange weight below which lie 30% of all the orange weights and above which lie 70% of all the orange weights. The areas listed in Table A.2 are the areas under the normal curve above a positive *z*-score, but since normal distributions are symmetric, the areas listed can also be treated as the areas under the normal curve below a negative *z*-score. By searching for the area closest to 0.30 (which is 0.3015), we find that the desired *z*-score is –0.52. In Figure 16-9, label the *z*-score –0.52 on the horizontal axis; if you draw a vertical line through the graph at the location of the *z*-score –0.52, you should be able to see that roughly 30% of the area under the normal curve lies below this *z*-score, and 70% of the area under the normal curve lies above this *z*-score. Use $x = \mu + z\sigma$ to convert the *z*-score –0.52 to a raw score to obtain the 30th percentile of the orange weights. (You should find that the 30th percentile is approximately 7.12 oz.)

We have now illustrated the use of Table A.2 in obtaining desired areas under the standard normal curve and in obtaining percentiles. Now that we have thoroughly discussed normal distributions, we once again recall how we have previously observed that the sampling distribution of $\bar{x}$ with simple random samples of sufficiently large size *n*, takes on the properties of a normal distribution; consequently, the sampling distribution of $\bar{x}$ can be approximated using a normal distribution, and we shall very shortly begin making heavy use of this fact.

Self-Test Problem 16-1. Suppose the right-hand grip strength for men between the ages of 20 and 40 is normally distributed with mean 86.3 lbs. and standard deviation 7.8 lbs. Draw a sketch illustrating the probability that one randomly selected male between the ages of 20 and 40 will have a right-hand grip strength

(a) over 90 lbs., and find this probability;
(b) under 96 lbs., and find this probability;
(c) over 70 lbs., and find this probability;
(d) under 75 lbs., and find this probability;
(e) between 85 and 100 lbs., and find this probability;
(f) between 88 and 95 lbs., and find this probability;
(g) between 75 and 82 lbs, and find this probability.
(h) Draw a sketch illustrating the probability that the right-hand grip strength for one randomly selected male is within 4 lbs. of the population mean, and find this probability.
(i) Find the percentile rank for a male whose right-hand grip strength is 90 lbs.
(j) Find the quartiles for the distribution of right-hand grip strengths.

Answers to Self-Test Problems

16-1 (a) 0.3192 or 31.92% (b) 0.8925 or 89.25% (c) 0.9817 or 98.17% (d) 0.0735 or 7.35% (e) 0.5283 or 52.83% (f) 0.2815 or 28.15% (g) 0.2177 or 21.77% (h) 0.3900 or 39.00% (i) 68.08 or 68 (j) The quartiles are approximately 81.1, 86.3, and 91.5 lbs.

Summary

A normal curve described entirely in terms of *z*-scores is called a *standard normal curve*. Tables of standard normal probabilities (e.g., Table A.2) provide a detailed description of a normal density curve in terms of *z*-scores. Using such a table together with the fact that a normal distribution is symmetric, we can obtain areas under a normal density curve, which can be interpreted as relative frequencies or as probabilities, and we can obtain percentile ranks and percentiles. Statistical software packages, spreadsheets, programmable calculators, etc. are often able to supply the same information as a table of standard normal probabilities. Since the sampling distribution of $\bar{x}$ with simple random samples of sufficiently large size *n*, takes on the properties of a normal distribution, the sampling distribution of $\bar{x}$ can be approximated using a normal distribution.