4.3: Using the Normal Distribution

The shaded area in the following graph indicates the area to the right of x. This area is represented by the probability $P(X > x)$. Normal tables - see Appendix A in Chapter 11.1 - provide the probability above a specific value such as x_1. This is the shaded part of the graph below.

This is a normal distribution curve. A value, x, is labeled on the horizontal axis, X. A vertical line extends from point x to the curve, and the area under the curve to the left of x is shaded. The area of this shaded section represents the probability that a value of the variable is less than x.

Figure $4.3.1$

Because the normal distribution is symmetrical, if x_1 were the same distance to the left of the mean, the probability in the left tail (*below* that value), would be the same as the shaded area in the right tail as shown in the figure. In general, bear in mind that because of the symmetry of this distribution, one-half of the probability is to the right of the mean and one-half is to the left of the mean.

Calculations of Probabilities

Let's discuss how to find the probability of a specified region in a standard normal distribution. The shaded region in the figure below shows that the area between x_1 and x_2 is the probability as stated in the formula: $P(X_1 \leq X \leq X_2)$. In this case, suppose we have μ = 5 and σ = 2. Suppose that x_1 = 7 and x_2 = 9. We want to find the probability of a score falling between 7 and 9 on this distribution.

Figure $4.3.2$

The solution is to convert the distribution we have with its mean and standard deviation to the Standard Normal Distribution. The Standard Normal has a random variable called Z. To compute probabilities, areas, for any normal distribution, we need only to convert the particular normal distribution to the standard normal distribution and look up the answer in the tables. As review, here again is the **standardizing formula**:

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

where z is the value on the standard normal distribution, x is the value from a normal distribution one wishes to convert to the standard normal, μ and σ are, respectively, the mean and standard deviation of that population. Note that the equation uses μ and σ which denotes population parameters. This is still dealing with probability so we always are dealing with the population, with **known** parameter values and a **known** distribution. It is also important to note that because the normal distribution is symmetrical it does not matter if the *z*-score is positive or negative when calculating a probability. One standard deviation to the left (negative *z*-score) covers the same area as one standard deviation to the right (positive *z*-score). This fact is why the Standard Normal tables do not provide areas for the left side of the distribution. Because of this symmetry, the *z*-score formula can also be written as:

$$Z=\frac{|x-\mu|}{\sigma}$$

where the vertical lines in the equation means the absolute value of the number.

Using this formula, we can determine that x_1 = 7 converts to a *z*-score of +1, and x_2 = 9 converts to a *z*-score of 2.

Then, using the z table, to find the probability of z = 1, go to the z column, reading down to 1.0 and then read at column 0. That number, $0.1587$is the probability of a score falling at or above z = 1. We repeat this process - using the z table - to find the probability of a score falling at or above z = 2, which is 0.0228. To obtain the shaded area that we wish to know about (as shown in Figure 4.3.2 above), we will need to subtract $0.1587-0.0228$to obtain our final probability, $.1359$.

To compute probabilities, areas, for any normal distribution, we need only to convert the particular normal distribution to the standard normal distribution using the *z*-formula and look up the answer in the tables.

What the standardizing formula is really doing is computing the number of standard deviations x is from the mean of its own distribution. The standardizing formula and the concept of counting standard deviations from the mean is the secret of all that we will do in this statistics class. The reason this is true is that **all** of statistics boils down to variation, and the counting of standard deviations is a measure of variation.

This formula, in many disguises, will reappear over and over throughout this course.

Example $4.3.1$

The final exam scores in a statistics class were normally distributed with a mean of 63 and a standard deviation of five.

a. Find the probability that a randomly selected student scored more than 65 on the exam.
b. Find the probability that a randomly selected student scored less than 85.

Answer

a. Let x = a score on the final exam. $X \sim N(63, 5)$, where $\mu = 63$ and $\sigma = 5$.

Draw a graph.

Then, find $P(x > 65)$.

$P(x>65)=0.3446$

This is a normal distribution curve. The peak of the curve coincides with the point 63 on the horizontal axis. The point 65 is also labeled. A vertical line extends from point 65 to the curve. The probability area to the right of 65 is shaded; it is equal to 0.3446.

Figure 4.3.3

$$z_1=\frac{x_1-\mu}{\sigma}=\frac{65-63}{5}=0.4$$

Looking up this value of *z* in our *z*-table, $P\left(x \geq x_1\right)=P\left(z \geq z_1\right)=0.3446$

The probability that any student selected at random scores more than 65 is 0.3446.

Answer

b.

$z=\frac{x-\mu}{\sigma}=\frac{85-63}{5}=4.4$. The closest value in our *z*-table is 4.5, which has a probability of .00000340.

Therefore, the probability that one student scores *less* than 85 is approximately one or 100% (i.e., 1 - the probability of scoring *above* that value).

Exercise 4.3.1

The golf scores for a school team were normally distributed with a mean of 68 and a standard deviation of three.

Find the probability that a randomly selected golfer scored less than 65.

Example 4.3.2

A personal computer is used for office work at home, research, communication, personal finances, education, entertainment, social networking, and a myriad of other things. Suppose that the average number of hours a household personal computer is used for entertainment is two hours per day. Assume the times for entertainment are normally distributed and the standard deviation for the times is half an hour.

a. Find the probability that a household personal computer is used for entertainment between 1.8 and 2.75 hours per day.

Answer

a. Let x = the amount of time (in hours) a household personal computer is used for entertainment. $X \sim N(2,0.5)$ where $\mu=2$ and $\sigma=0.5$.

Find $P(1.8<x<2.75)$.

The probability for which you are looking is the area **between** $x=1.8$ and $x=2.75$. $P(1.8<x<2.75)=0.5886$

This is a normal distribution curve. The peak of the curve coincides with the point 2 on the horizontal axis. The values 1.8 and 2.75 are also labeled on the x-axis. Vertical lines extend from 1.8 and 2.75 to the curve. The area between the lines is shaded.

Figure 4.3.4

$P(1.8 \leq x \leq 2.75)=P(z_1 \leq z \leq z_2)$

The probability that a household personal computer is used between 1.8 and 2.75 hours per day for entertainment is 0.5886.

b. Find the maximum number of hours per day that the bottom quartile of households uses a personal computer for entertainment.

Answer

b. To find the maximum number of hours per day that the bottom quartile of households uses a personal computer for entertainment, first find the 25^{th} percentile, k, where $P(x<k)=0.25$.

This is a normal distribution curve. The area under the left tail of the curve is shaded. The shaded area shows that the probability that x is less than k is 0.25. It follows that k = 1.67.

Figure 4.3.5

$f(Z)=0.5-0.25=0.25,$ therefore $z \approx -0.675($ or just 0.68 using the table$)$ $z=\frac{x-\mu}{\sigma}=\frac{x-2}{0.5}=-0.675,$ therefore $x=-0.675*0.5+2=1.66$

The maximum number of hours per day that the bottom quartile of households uses a personal computer for entertainment is 1.66 hours.

Exercise 4.3.2

The golf scores for a school team were normally distributed with a mean of 68 and a standard deviation of three. Find the probability that a golfer scored between 66 and 70.

Example 4.3.3

In the United States the ages 13 to 55+ of smartphone users approximately follow a normal distribution with approximate mean and standard deviation of 36.9 years and 13.9 years, respectively.

a. Determine the probability that a random smartphone user in the age range 13 to 55+ is between 23 and 64.7 years old.

Answer

a. 0.8186

b. Determine the probability that a randomly selected smartphone user in the age range 13 to 55+ is at most 50.8 years old.

Answer

b. 0.8413

Example 4.3.4

A citrus farmer who grows mandarin oranges finds that the diameters of mandarin oranges harvested on his farm follow a normal distribution with a mean diameter of 5.85 cm and a standard deviation of 0.24 cm.

a. Find the probability that a randomly selected mandarin orange from this farm has a diameter larger than 6.0 cm. Sketch the graph.

Answer

This is a normal distribution curve. The peak of the curve coincides with the point 2 on the horizontal axis. The values 1.8 and 2.75 are also labeled on the x-axis. Vertical lines extend from 1.8 and 2.75 to the curve. The area between the lines is shaded.

Figure 4.3.6

$$z_1 = \frac{6 - 5.85}{.24} = .625$$

$P(x \geq 6) = P(z \geq 0.625) = 0.2670$

b. The middle 20% of mandarin oranges from this farm have diameters between ______ and ______.

Answer

$f(z) = \frac{0.20}{2} = 0.10, \text{ therefore } z \approx \pm 0.25$

$z = \frac{x-\mu}{\sigma} = \frac{x-5.85}{0.24} = \pm 0.25 \to \pm 0.25 \cdot 0.24 + 5.85 = (5.79, 5.91)$