

Chapter 6

The Normal Distribution

6.1 The Normal Distribution¹

6.1.1 Student Learning Objectives

By the end of this chapter, the student should be able to:

- Recognize the normal probability distribution and apply it appropriately.
- Recognize the standard normal probability distribution and apply it appropriately.
- Compare normal probabilities by converting to the standard normal distribution.

6.1.2 Introduction

The normal, a continuous distribution, is the most important of all the distributions. It is widely used and even more widely abused. Its graph is bell-shaped. You see the bell curve in almost all disciplines. Some of these include psychology, business, economics, the sciences, nursing, and, of course, mathematics. Some of your instructors may use the normal distribution to help determine your grade. Most IQ scores are normally distributed. Often real estate prices fit a normal distribution. The normal distribution is extremely important but it cannot be applied to everything in the real world.

In this chapter, you will study the normal distribution, the standard normal, and applications associated with them.

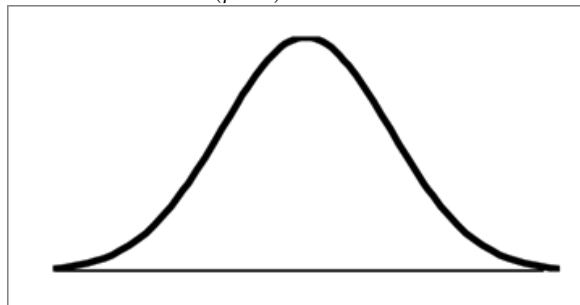
6.1.3 Optional Collaborative Classroom Activity

Your instructor will record the heights of both men and women in your class, separately. Draw histograms of your data. Then draw a smooth curve through each histogram. Is each curve somewhat bell-shaped? Do you think that if you had recorded 200 data values for men and 200 for women that the curves would look bell-shaped? Calculate the mean for each data set. Write the means on the x-axis of the appropriate graph below the peak. Shade the approximate area that represents the probability that one randomly chosen male is taller than 72 inches. Shade the approximate area that represents the probability that one randomly chosen female is shorter than 60 inches. If the total area under each curve is one, does either probability appear to be more than 0.5?

¹This content is available online at <<http://http://cnx.org/content/m16979/1.10/>>.

The normal distribution has two parameters (two numerical descriptive measures), the mean (μ) and the standard deviation (σ). If X is a quantity to be measured that has a normal distribution with mean (μ) and the standard deviation (σ), we designate this by writing

NORMAL: $X \sim N(\mu, \sigma)$



The probability density function is a rather complicated function. **Do not memorize it.** It is not necessary.

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

The cumulative distribution function is $P(X < x)$. It is calculated either by a calculator or a computer or it is looked up in a table.

The curve is symmetrical about a vertical line drawn through the mean, μ . In theory, the mean is the same as the median since the graph is symmetric about μ . As the notation indicates, the normal distribution depends only on the mean and the standard deviation. Since the area under the curve must equal one, a change in the standard deviation, σ , causes a change in the shape of the curve; the curve becomes fatter or skinnier depending on σ . A change in μ causes the graph to shift to the left or right. This means there are an infinite number of normal probability distributions. One of special interest is called the **standard normal distribution**.

6.2 The Standard Normal Distribution²

The **standard normal distribution** is a normal distribution of **standardized values called z-scores**. A **z-score is measured in units of the standard deviation**. For example, if the mean of a normal distribution is 5 and the standard deviation is 2, the value 11 is 3 standard deviations above (or to the right of) the mean. The calculation is:

$$x = \mu + (z)\sigma = 5 + (3)(2) = 11 \quad (6.1)$$

The z-score is 3.

The mean for the standard normal distribution is 0 and the standard deviation is 1. The transformation

$z = \frac{x-\mu}{\sigma}$ produces the distribution $Z \sim N(0, 1)$. The value x comes from a normal distribution with mean μ and standard deviation σ .

²This content is available online at <<http://http://cnx.org/content/m16986/1.7/>>.

6.3 Z-scores³

If X is a normally distributed random variable and $X \sim N(\mu, \sigma)$, then the z-score is:

$$z = \frac{x - \mu}{\sigma} \quad (6.2)$$

The z-score tells you how many standard deviations that the value x is above (to the right of) or below (to the left of) the mean, μ . Values of x that are larger than the mean have positive z-scores and values of x that are smaller than the mean have negative z-scores. If x equals the mean, then x has a z-score of 0.

Example 6.1

Suppose $X \sim N(5, 6)$. This says that X is a normally distributed random variable with mean $\mu = 5$ and standard deviation $\sigma = 6$. Suppose $x = 17$. Then:

$$z = \frac{x - \mu}{\sigma} = \frac{17 - 5}{6} = 2 \quad (6.3)$$

This means that $x = 17$ is **2 standard deviations** (2σ) above or to the right of the mean $\mu = 5$. The standard deviation is $\sigma = 6$.

Notice that:

$$5 + 2 \cdot 6 = 17 \quad (\text{The pattern is } \mu + z\sigma = x.) \quad (6.4)$$

Now suppose $x = 1$. Then:

$$z = \frac{x - \mu}{\sigma} = \frac{1 - 5}{6} = -0.67 \quad (\text{rounded to two decimal places}) \quad (6.5)$$

This means that $x = 1$ is 0.67 standard deviations (-0.67σ) below or to the left of the mean $\mu = 5$. Notice that:

$$5 + (-0.67)(6) \text{ is approximately equal to } 1 \quad (\text{This has the pattern } \mu + (-0.67)\sigma = 1)$$

Summarizing, when z is positive, x is above or to the right of μ and when z is negative, x is to the left of or below μ .

Example 6.2

Some doctors believe that a person can lose 5 pounds, on the average, in a month by reducing his/her fat intake and by exercising consistently. Suppose weight loss has a normal distribution. Let X = the amount of weight lost (in pounds) by a person in a month. Use a standard deviation of 2 pounds. $X \sim N(5, 2)$. Fill in the blanks.

Problem 1

(Solution on p. 278.)

Suppose a person **lost** 10 pounds in a month. The z-score when $x = 10$ pounds is $z = 2.5$ (verify). This z-score tells you that $x = 10$ is _____ standard deviations to the _____ (right or left) of the mean _____. (What is the mean?).

Problem 2

(Solution on p. 278.)

Suppose a person **gained** 3 pounds (a negative weight loss). Then $z =$ _____. This z-score tells you that $x = -3$ is _____ standard deviations to the _____ (right or left) of the mean.

Suppose the random variables X and Y have the following normal distributions: $X \sim N(5, 6)$ and $Y \sim N(2, 1)$. If $x = 17$, then $z = 2$. (This was previously shown.) If $y = 4$, what is z ?

$$z = \frac{y - \mu}{\sigma} = \frac{4 - 2}{1} = 2 \quad \text{where } \mu=2 \text{ and } \sigma=1. \quad (6.6)$$

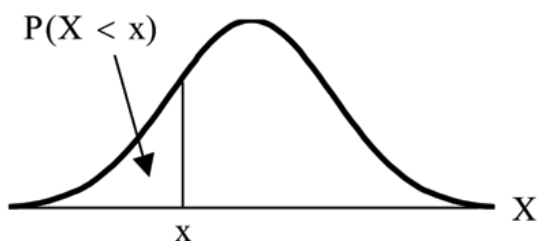
³This content is available online at <<http://cnx.org/content/m16991/1.7/>>.

The z-score for $y = 4$ is $z = 2$. This means that 4 is $z = 2$ standard deviations to the right of the mean. Therefore, $x = 17$ and $y = 4$ are both 2 (of **their**) standard deviations to the right of **their** respective means.

The z-score allows us to compare data that are scaled differently. To understand the concept, suppose $X \sim N(5, 6)$ represents weight gains for one group of people who are trying to gain weight in a 6 week period and $Y \sim N(2, 1)$ measures the same weight gain for a second group of people. A negative weight gain would be a weight loss. Since $x = 17$ and $y = 4$ are each 2 standard deviations to the right of their means, they represent the same weight gain **in relationship to their means**.

6.4 Areas to the Left and Right of x^4

The arrow in the graph below points to the area to the left of x . This area is represented by the probability $P(X < x)$. Normal tables, computers, and calculators provide or calculate the probability $P(X < x)$.



The area to the right is then $P(X > x) = 1 - P(X < x)$.

Remember, $P(X < x) = \text{Area to the left of the vertical line through } x$.

$P(X > x) = 1 - P(X < x) = \text{Area to the right of the vertical line through } x$

$P(X < x)$ is the same as $P(X \leq x)$ and $P(X > x)$ is the same as $P(X \geq x)$ for continuous distributions.

6.5 Calculations of Probabilities⁵

Probabilities are calculated by using technology. There are instructions in the chapter for the TI-83+ and TI-84 calculators.

NOTE: In the Table of Contents for **Collaborative Statistics**, entry **15. Tables** has a link to a table of normal probabilities. Use the probability tables if so desired, instead of a calculator.

Example 6.3

If the area to the left is 0.0228, then the area to the right is $1 - 0.0228 = 0.9772$.

Example 6.4

The final exam scores in a statistics class were normally distributed with a mean of 63 and a standard deviation of 5.

Problem 1

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

⁴This content is available online at <http://cnx.org/content/m16976/1.5/>.

⁵This content is available online at <http://cnx.org/content/m16977/1.10/>.

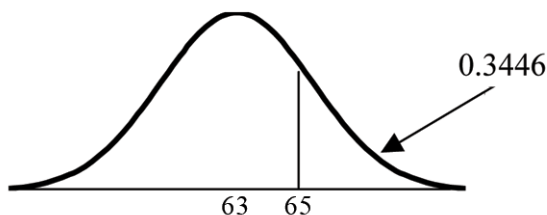
Solution

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

Draw a graph.

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

$P(X > 65) = 0.3446$ (calculator or computer)



The probability that one student scores more than 65 is 0.3446.

Using the TI-83+ or the TI-84 calculators, the calculation is as follows. Go into 2nd DISTR.

After pressing 2nd DISTR, press 2:normalcdf.

The syntax for the instructions are shown below.

normalcdf(lower value, upper value, mean, standard deviation) For this problem: normalcdf(65,1E99,63,5) = 0.3446. You get 1E99 ($= 10^{99}$) by pressing 1, the EE key (a 2nd key) and then 99. Or, you can enter 10^{99} instead. The number 10^{99} is way out in the right tail of the normal curve. We are calculating the area between 65 and 10^{99} . In some instances, the lower number of the area might be -1E99 ($= -10^{99}$). The number -10^{99} is way out in the left tail of the normal curve.

HISTORICAL NOTE: The TI probability program calculates a z-score and then the probability from the z-score. Before technology, the z-score was looked up in a standard normal probability table (because the math involved is too cumbersome) to find the probability. In this example, a standard normal table with area to the left of the z-score was used. You calculate the z-score and look up the area to the left. The probability is the area to the right.

$$z = \frac{65-63}{5} = 0.4 \quad . \text{ Area to the left is } 0.6554. \quad P(X > 65) = P(Z > 0.4) = 1 - 0.6554 = 0.3446$$

Problem 2

Find the probability that a randomly selected student scored less than 85.

Solution

Draw a graph.

Then find $P(X < 85)$. Shade the graph. $P(X < 85) = 1$ (calculator or computer)

The probability that one student scores less than 85 is approximately 1 (or 100%).

The TI-instructions and answer are as follows:

normalcdf(0,85,63,5) = 1 (rounds to 1)

Problem 3

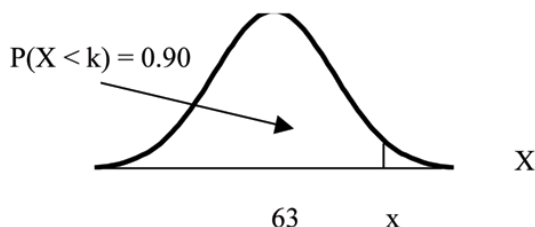
Find the 90th percentile (that is, find the score k that has 90 % of the scores below k and 10% of the scores above k).

Solution

Find the 90th percentile. For each problem or part of a problem, draw a new graph. Draw the x-axis. Shade the area that corresponds to the 90th percentile.

Let k = the 90th percentile. k is located on the x-axis. $P(X < k)$ is the area to the left of k . The 90th percentile k separates the exam scores into those that are the same or lower than k and those that are the same or higher. Ninety percent of the test scores are the same or lower than k and 10% are the same or higher. k is often called a **critical value**.

$k = 69.4$ (calculator or computer)



The 90th percentile is 69.4. This means that 90% of the test scores fall at or below 69.4 and 10% fall at or above. For the TI-83+ or TI-84 calculators, use `invNorm` in 2nd DISTR. `invNorm(area to the left, mean, standard deviation)` For this problem, `invNorm(.90,63,5) = 69.4`

Problem 4

Find the 70th percentile (that is, find the score k such that 70% of scores are below k and 30% of the scores are above k).

Solution

Find the 70th percentile.

Draw a new graph and label it appropriately. $k = 65.6$

The 70th percentile is 65.6. This means that 70% of the test scores fall at or below 65.5 and 30% fall at or above.

`invNorm(.70,63,5) = 65.6`

Example 6.5

More and more households in the United States have at least one computer. The 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 2 hours per day. Assume the times for entertainment are normally distributed and the standard deviation for the times is half an hour.

Problem 1

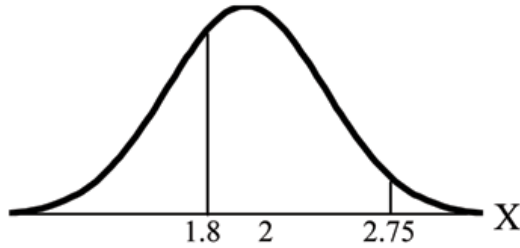
Find the probability that a household personal computer is used between 1.8 and 2.75 hours per day.

Solution

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 $x = 2.75$. $P(1.8 < X < 2.75) = 0.5886$



$$\text{normalcdf}(1.8, 2.75, 2, .5) = 0.5886$$

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

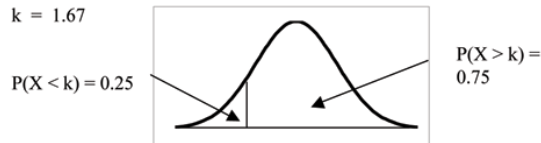
Problem 2

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

Solution

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

$$k = 1.67$$



$$\text{invNorm}(.25, 2, .5) = 1.67$$

The maximum number of hours per day that the bottom quartile of households uses a personal computer for entertainment is 1.67 hours.

6.6 Summary of Formulas⁶

Rule 6.1: Normal Probability Distribution

$$X \sim N(\mu, \sigma)$$

μ = the mean σ = the standard deviation

Rule 6.2: Standard Normal Probability Distribution

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

Z = a standardized value (z-score)

mean = 0 standard deviation = 1

Rule 6.3: Finding the kth Percentile

To find the **kth** percentile when the z-score is known: $k = \mu + (z) \sigma$

Rule 6.4: z-score

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

Rule 6.5: Finding the area to the left

The area to the left: $P(X < x)$

Rule 6.6: Finding the area to the right

The area to the right: $P(X > x) = 1 - P(X < x)$

⁶This content is available online at <<http://cnx.org/content/m16987/1.4/>>.