

1.3 The Normal Distributions

Density Curves

Definition. A (probability) *density curve* is a curve that

- is always on or above the horizontal axis, and
- has area exactly 1 underneath it (that is, the area bounded by the curve and the x -axis).

A density curve describes the overall pattern of a distribution. The area under the curve and above any range of values is the proportion of all observations that fall in that range.

The Median and Mean of a Density Curve

Note. The median of a density curve is the “equal-areas” point. See Figure 1.15 (and TM-19).

Definition. The *mean* of a density curve is the equal-areas point, the point that divides the area under the curve in half. The *mean* of a density curve is the balance point, at which the curve would balance if made of solid material. The median and mean are the same for a symmetric density curve. See Figure 1.16 (and TM-20).

Note. The usual notation for the mean of an idealized distribution is μ

(mu). The standard deviation of a density curve is denoted σ (sigma).

Normal Distributions

Note. A VERY common class of density curves is the *normal distributions*. These curves are symmetric, single-peaked, and bell-shaped. All normal distributions have the same shape and are determined solely by their mean μ and standard deviation σ . Figure 1.19 (see TM-21) gives two examples of normal distributions. The points at which the curves change concavity are located a distance σ on either side of μ . We will use the area under these curves to represent a percentage of observations. (These areas correspond to integrals, for those of you with some experience with calculus.)

Note. In the normal distribution with mean μ and standard deviation σ :

- 68% of the observations fall within σ of the mean μ .
- 95% of the observations fall within 2σ of μ .
- 99.7% of the observations fall within 3σ of μ .

This is called the “68-95-99.7 Rule.” See Figure 1.20 (and TM-22).

Notation. We abbreviate the normal distribution with mean μ and standard deviation σ as $N(\mu, \sigma)$.

Note. Some reasons we are interested in normal distributions are:

- Normal distributions are good descriptions for some distributions of *real data*.
- Normal distributions are good approximations to the results of many kinds of *chance outcomes*.
- Many *statistical inference* procedures based on normal distributions work well for other roughly symmetric distributions.

The Standard Normal Distribution

Definition. If x is an observation from a distribution that has mean μ and standard deviation σ , the *standard value* of x is

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

This value is sometimes called the *z-score* for x .

Definition. The *standard normal distribution* is the normal distribution $N(0, 1)$ with mean 0 and standard deviation 1. If a variable x has any normal distribution $N(\mu, \sigma)$ with mean μ and standard deviation σ , then the standardized variable $z = \frac{x - \mu}{\sigma}$ has the standard normal distribution.

Normal Distribution Calculations

Note. An area under a density curve is a proportion of the observations in a distribution. Because all normal distributions are the same when we standardize, we can find area under any normal curve from a single table, a table that gives areas under the curve for the standard normal distribution.

Example 1.15. What proportion of all young women are less than 68 inches tall? Assume that the relevant distribution is $N(64.5, 2.5)$ (see Example 1.14, page 65).

Solution. The z -score for $x = 68$ inches is

$$z = \frac{x - \mu}{\sigma} = \frac{68 \text{ in} - 64.5 \text{ in}}{2.5 \text{ in}} = 1.4.$$

So we want to find the area to the LEFT of 1.4 in the standard normal distribution (the question says “less than”). See Figure 1.22 (and TM-24). We’ll find this area after one more comment.

Note. *Table A* is a table of areas under the standard normal curve. The table entry for each value z is the area under the curve to the left of z . Table A is reproduced also on TM-139 and TM-140.

Solution to Example 1.15 (continued). We now see that we want the entry in Table A that corresponds to $z = 1.4$. This entry is 0.9192. Therefore 91.92% of the population of young women are less than 68 inches tall.

Note. Fortunately, Table A is built into the Sharp EL-546G. To find the area under the normal distribution to the LEFT of a z -score, do the following:

- Put the calculator in “statistics mode” by pressing $\boxed{\text{MODE}}$ and $\boxed{3}$.
- Press $\boxed{0}$ to put the calculator in single-variable statistics mode (STO appears in the display).
- Press the $\boxed{2\text{ndF}}$ key, then the $\boxed{P(t)}$ key (the $\boxed{1}$ key... “ P ” appears), type in the z value, and hit $\boxed{=}$.

See page 43 of the calculator owner’s manual for more details.

Note. The protocol for finding normal proportions (i.e. areas under $N(0, 1)$ for a given x value) is:

- State the problem in terms of the observed variable x .
- Standardize x to restate the problem in terms of a standard normal variable z . Draw a picture to show the area under the standard normal curve.
- Find the required area under the standard normal curve, using Table A or the calculator and the fact that the total area under the curve is 1.

Example 1.17. The distribution of blood cholesterol levels in a large population of people of the same age and sex is roughly normal. For

14-year-old boys, the mean is $\mu = 170$ milligrams of cholesterol per deciliter of blood (mg/dl) and the standard deviation is $\sigma = 30$ mg/dl. levels above 240 mg/dl may require medical attention. What percent of 14-year-old boys have more than 240 mg/dl of cholesterol?

Solution. The z -score for $x = 240$ is $z = \frac{x - \mu}{\sigma} = \frac{240 - 170}{30} = 2.33$. We want the area to the RIGHT of $z = 2.33$ in $N(0, 1)$ (the question says “more than”). Well, the area to the left of $z = 2.33$ is (Table A or the calculator) .9901. Since the total area under a normal distribution is 1, the desired area is $1 - .9901 = .0099$. So .99% of such boys have more than 240 md/dl of cholesterol. See Figure 1.23 (and TM-25).

Note. We can also calculate area to the RIGHT of a z -score using the calculator:

- Put the calculator in “statistics mode” by pressing $\boxed{\text{MODE}}$ and $\boxed{3}$.
- Press $\boxed{0}$ to put the calculator in single-variable statistics mode (**STO** appears in the display).
- Press the $\boxed{2\text{ndF}}$ key, then the $\boxed{R(t)}$ key (the $\boxed{3}$ key... “ R ” appears), type in the z value, and hit $\boxed{=}$.

See page 44 of the calculator owner’s manual for more details.

Example 1.18. In the above example, what percent of 14-year-old boys have blood cholesterol between 170 and 240 mg/dl?

Solution. We are interested in what proportion of x values satisfy $170 \leq x \leq 240$. The z -score for $x = 170$ is $\frac{170 - 170}{30} = 0$ and the z -score for $x = 240$ is $\frac{240 - 170}{30} = 2.33$. Therefore we want the area under $N(0, 1)$ for $0 \leq z \leq 2.33$ (see Figure 1.24 and TM-26). Well, the area to the LEFT of $z = 0$ is 0.5 (since 0 is the mean), and the area to the LEFT of $z = 2.33$ is .9901 (Table A or the calculator). Therefore, the desired area is $.9901 - .5 = .4901$. So 49.01% of boys fall in this category.

Note. The area bounded under $N(0, 1)$ between 0 and z is also a built in function for the Sharp EL-546G. It is the $Q(t)$ function and is accessed in the same way as the $P(t)$ and $R(t)$ functions. See page 44 of the calculator owner's manual for more details.

Note. If we have to deal with a z value outside the range of Table A, we do so as follows: if the value is less than -3.49 , assume the entry to be 0, and if the value is greater than $z = 3.49$ assume the entry to be 1. When dealing with the calculator, this is not a problem (and you get 2 more decimals of accuracy than given in Table A).

Finding a Value Given a Proportion

Note. Instead of calculating proportions from Table A, we might be given the proportion of a population below a certain unknown value, and asked to find that value. To carry this out, we must use Table 2 backwards (unfortunately, this is not built into your calculator).

Example 1.19. Scores on the SAT for verbal ability follow the $N(430, 100)$ distribution. How high must a student score in order to place in the top 10% of all students taking the SAT?

Solution. We want the area to the LEFT of our z value to be $1 - .1 = .9$ (we are interested in the complement of this area... the problem says “top 10%”). From Table A, we have $z = 1.28$. Now converting this back to a SAT score we solve $\frac{x - 430}{100} = 1.28$ and get $x = 558$.