Figure 5.2. PROBABILITY MODELLING: Normal Distributions

Program 4 in: *Against All Odds: Inside Statistics*

Histograms and stemplots give a quite detailed picture of a distribution. In this program you will see that the overall pattern of a distribution can often be described compactly by a smooth curve, called a *probability density function* (*p.d.f.*) or *density curve*. Think of a p.d.f. as a smooth curve drawn through the tops of the bars of a relative frequency histogram. A density curve is usually an *idealized* description of a distribution that gives the overall pattern but ignores the irregularities that are present in actual data. In the video, p.d.f.s are used to compare the distributions of the ages of Americans in 1930 and 2075. Of course, the second distribution is an estimate made by the U.S. Census Bureau. This is an effective use of density curves because we are interested in the big picture of changes in the overall distributions of age, *not* in small details.

The median of a distribution can be located on a density curve as the point that divides the curve into two equal areas. The mean of the distribution is the point at which the curve would balance if it were made out of solid material. Because a p.d.f. represents an ideal distribution, we use a special symbol for its mean, the Greek letter μ; the standard deviation is similarly represented by the Greek letter σ. The standard deviation σ cannot be located by eye on most p.d.f.s. The mean and median are equal for symmetric density curves, but the mean of a skewed curve is located farther toward the long tail than the median.

The *normal distributions* are specified by bell-shaped symmetric density curves. Normal distributions are used very widely in statistics, so we will study their properties in detail. Unlike most distributions, a normal distribution is completely described by its mean μ and standard deviation σ. Both μ and σ can be found from the shape of a normal curve. Changing μ moves a normal curve along the axis with*out* changing its shape. Changing σ changes the shape; curves with larger standard deviations are more spread out and less sharply peaked. As the video notes, normal distributions give an approximate description of the overall pattern of many common types of data. But do remember that *no* distribution for actual data will be *exactly* normal. The smooth normal curve is an *ideal* distribution that is easy to work with, not an exact description.

All normal distributions are the *same* when measurements are made in units of σ about the mean. These are called *standardized measurements*. In particular, all normal distributions satisfy the *68–95–99.7 rule*. This rule says that in any normal distribution, about 68% of the observations fall within σ of the mean μ, about 95% fall within two standard deviations of the mean, and about 99.7% fall within 3σ on either side of μ. Study Figure 1.24 on page 72 of the Text and Figure 5.3 of the Course Materials to get a clear picture of the situation. The 68–95–99.7 rule allows us to think about normal distributions without doing detailed calculcations. In the video, we can quickly see how few young women are eligible for membership in a Beanstalk Club that is open only to women at least 5 feet 10 inches (about 178 cm) tall [and to men at least 6 feet 2 inches (about 188 cm) tall].

When we have several observations, each from a different normal distribution, we often use standardized measurements to compare the observations. As the video notes, the distribution of major league batting averages is roughly normal, but the standard deviation has changed over time. To compare how far outstanding hitters of different eras stood above their contemporaries, we standardized their batting averages.

Blank page