

So the probability is .95 that a child's score (X) chosen at random would be between 30.4 and 69.6. We may not be very interested in low scores, because they don't represent problems. But anyone with a score of 69.6 or higher is a problem to someone. Only 2.5% of children score at least that high.

What we have just discussed is closely related to, but not quite the same as, what we will later consider under the heading of confidence limits. The major difference is that here we knew the population mean and were trying to predict where a single observation (X) would fall. We will later call something like this a **prediction interval**. When we discuss confidence limits, we will have a sample mean (or some other statistic) and will want to set limits that have a probability of .95 of bracketing the population mean (or some other relevant parameter). You do not need to know anything at all about confidence limits at this point. I simply mention the issue to forestall any confusion in the future.

prediction interval

3.5 Assessing Whether Data Are Normally Distributed

There will be many occasions in this book where we will assume that data are normally distributed, but it is difficult to look at a distribution of sample data and assess the reasonableness of such an assumption. Statistics texts are filled with examples of distributions that look normal but aren't, and these are often followed by statements of how distorted the results of some procedure are because the data were nonnormal. As I said earlier, we can superimpose a true normal distribution on top of a histogram and have some idea of how well we are doing, but that is often a misleading approach. A far better approach is to use what are called **Q-Q plots (quantile-quantile plots)**.

Q-Q plots (quantile-quantile plots)

Q-Q plots

The idea behind quantile-quantile (Q-Q) plots is basically quite simple. Suppose that we have a sample of 100 observations that is perfectly normally distributed with mean = 0 and standard deviation = 1. (The mean and standard deviation could be any values, but 0 and 1 just make the discussion simpler.) With that distribution we can easily calculate what value would cut off, for example, the lowest 1% of the distribution. From Appendix z this would be a value of -2.33 . We would also know that a cutoff of -2.054 cuts off the lowest 2%. We could make this calculation for every value of $0.00 < p < 1.00$, and we could name the results the *expected quantiles* of a normal distribution.

Now we go to the data we actually have. Because I have specified that they are perfectly normally distributed and that there are $n = 100$ observations, the lowest score will be the lowest 1% and it will be -2.33 . Similarly the second lowest value would cut off 2% of the distribution and would be -2.054 . We will call these the *obtained quantiles* because they were calculated directly from the data. For a perfectly normal distribution the two sets of quantiles should agree exactly. The value that forms the 15th percentile of the obtained distribution should be exactly that value that the normal distribution would have given the population mean and standard deviation.

But suppose that our sample data were not normally distributed. Then we might find that the score cutting off the lowest 1% of our sample was -2.8 instead of -2.33 . The same could happen for other quantiles. Here the expected quantiles from a normal distribution and the obtained quantiles from our sample would not agree.

But how do we measure agreement? The easiest way is to plot the two sets of quantiles against each other, putting the expected quantiles on the **Y axis** and the obtained quantiles on the **X axis**. If the distribution is normal the plot should form a straight line running at

a 45-degree angle. These plots are illustrated in Figure 3.10 for a set of data drawn from a normal distribution and a set drawn from a decidedly nonnormal distribution.

In Figure 3.10 you can see that for normal data the Q-Q plot shows that most of the points fall nicely on a straight line. They depart from the line a bit at each end, but that commonly happens unless you have very large sample sizes. For the nonnormal data, however, the plotted points depart drastically from a straight line. At the lower end where we would expect quantiles of around -1 , the lowest obtained quantile was actually about -2 . In other words the distribution was truncated on the left. At the upper right of the Q-Q plot where we obtained quantiles of around 2.0 , the expected value was at least 3.0 . In other words the obtained data did not depart enough from the mean at the lower end and departed too much from the mean at the upper end. A program to plot Q-Q plots in R is available at the book's Web site.

We have been looking at Achenbach's Total Behavior Problem scores and I have suggested that they are very normally distributed. Figure 3.11 presents a Q-Q plot for those scores. From this plot it is apparent that Behavior Problem scores are normally distributed, which is, in part, a function of the fact that Achenbach worked very hard to develop that scale and give it desirable properties.

However, let's look at the reaction-time data that began this book. In discussing Figure 2.4, I said that the data are roughly normally distributed though truncated on the

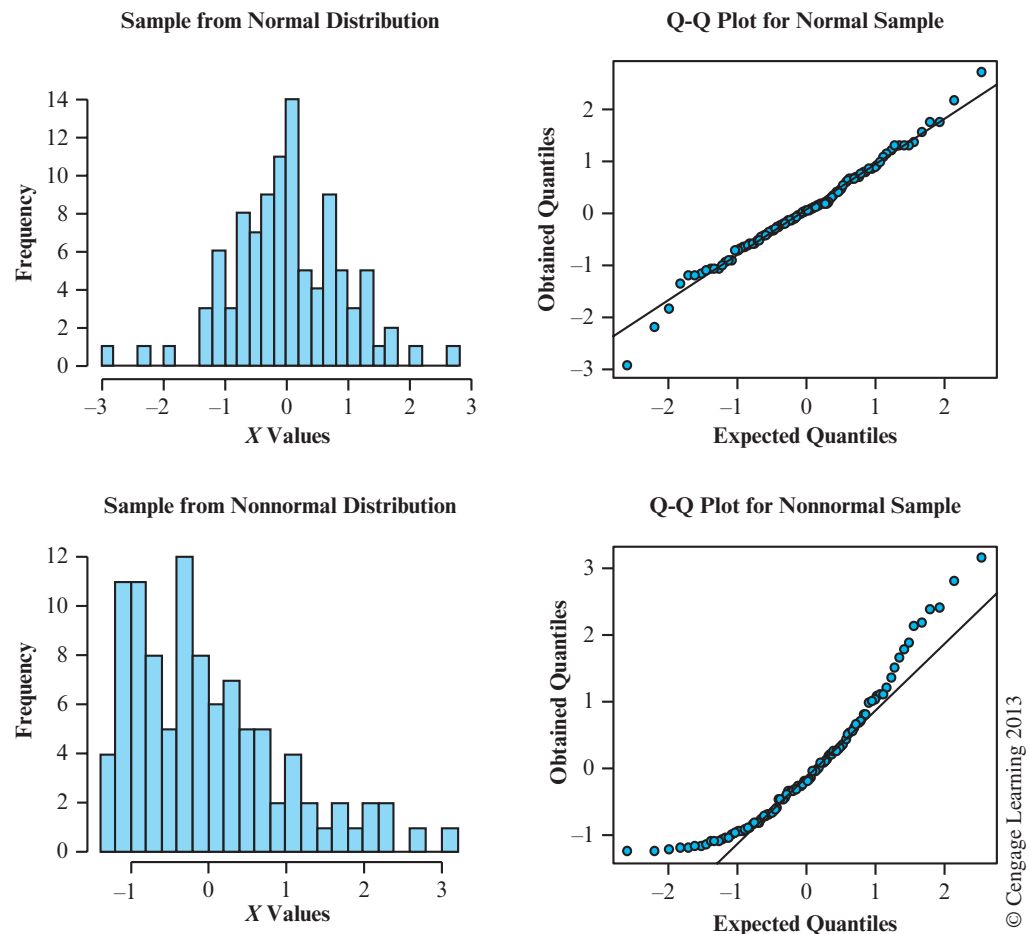


Figure 3.10 Histograms and Q-Q plots for normal and nonnormal data

left. And the figure did, indeed, pretty much show that. But Figure 3.12 shows the Q-Q plot for RxTime, and you can see that it is far from a straight line. I did this to illustrate that it is more difficult than you might think to look at a simple histogram, with or without a superimposed normal distribution, and decide whether or not a distribution is normal.

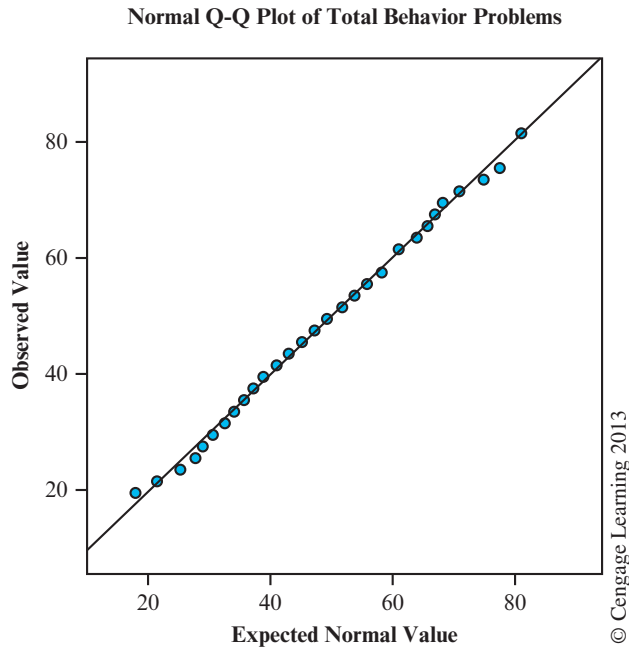


Figure 3.11 Q-Q plot of Total Behavior Problem scores

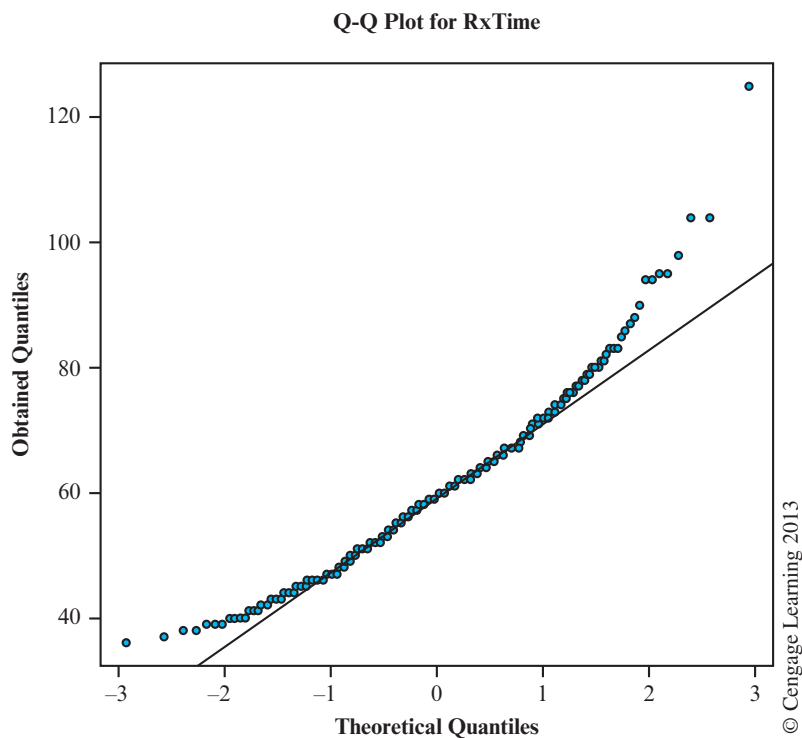


Figure 3.12 Q-Q plot of reaction time data

The Axes for a Q-Q Plot

In presenting the logic behind a Q-Q plot I spoke as if the variables in question were standardized, although I did mention that it was not a requirement. I did so because it was easier to send you to tables of normal distribution. However, you will often come across Q-Q plots where one or both axes are in different units, which is not a problem. The important consideration is the distribution of points within the plot and not the scale of either axis. In fact, different statistical packages not only use different scaling, but they also differ on which variable is plotted on which axis. If you see a plot that looks like a mirror image (vertically) of one of my plots that simply means they have plotted the observed values on the X axis instead of the expected ones.

The Kolmogorov-Smirnov Test

Kolmogorov-Smirnov test

The best-known statistical test for normality is the **Kolmogorov-Smirnov test**, which is available within SPSS under the nonparametric tests. Although you should know that the test exists, *most people do not recommend its use*. In the first place most small samples will pass the test even when they are decidedly nonnormal. On the other hand, when you have very large samples the test is very likely to reject the hypothesis of normality even though minor deviations from normality will not be a problem. D'Agostino and Stephens (1986) put it even more strongly when they wrote, "The Kolmogorov-Smirnov test is only a historical curiosity. It should never be used." I mention the test here only because you will come across references to it and SPSS will offer to calculate it for you. You should know its weaknesses.

3.6 Measures Related to z

standard scores

We already have seen that the z formula given earlier can be used to convert a distribution with any mean and variance to a distribution with a mean of 0 and a standard deviation (and variance) of 1. We frequently refer to such transformed scores as **standard scores**. There are also other transformational scoring systems with particular properties, some of which people use every day without realizing what they are.

A good example of such a scoring system is the common IQ. The raw scores from an IQ test are routinely transformed to a distribution with a mean of 100 and a standard deviation of 15 (or 16 in the case of the Binet). Knowing this, you can readily convert an individual's IQ (e.g., 120) to his or her position in terms of standard deviations above or below the mean (i.e., you can calculate the z score). Because IQ scores are more or less normally distributed, you can then convert z into a percentage measure by use of Appendix z . (In this example, a score of 120 has approximately 91% of the scores below it. This is known as the 91st **percentile**.)

percentile

Other common examples are standard diagnostic tests that are converted to a fixed mean and standard deviation. (Achenbach's test is an example.) The raw scores are transformed by the producer of the test and reported as coming from a distribution with a mean of 50 and a standard deviation of 10 (for example). Such a scoring system is easy to devise. We start by converting raw scores to z scores (using the obtained raw score mean and standard deviation). We then convert the z scores to the particular scoring system we have in mind. Thus

$$\text{New score} = \text{New SD} * (z) + \text{New mean},$$

where z represents the z score corresponding to the individual's raw score. Scoring systems such as the one used on Achenbach's Youth Self-Report checklist, which have a mean set