

8.7 Applications of the Normal Distribution

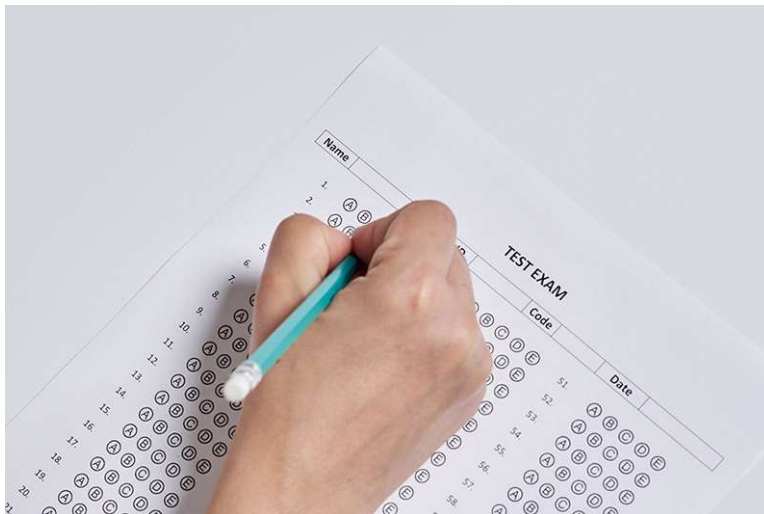


Figure 8.63 Standardized test results generally adhere to the normal distribution. (credit: “Taking a Test” by Marco Verch Professional Photographer/Flickr, CC BY 2.0)

Learning Objectives

After completing this section, you should be able to:

1. Apply the normal distribution to real-world scenarios.

As we saw in [The Normal Distribution](#), the word “standardized” is closely associated with the normal distribution. This is why tests like college entrance exams, state achievement tests for K–12 students, and Advanced Placement tests are often called “standardized tests”: scores are assigned in a way that forces them to follow a normal distribution, with a mean and standard deviation that are consistent from year to year. Standardization also allows people like college admissions officers to directly compare an applicant who took the ACT (a college entrance exam) to an applicant who instead chose to take the SAT (a different college entrance exam). Standardization allows us to compare individuals from different groups; this is among the most important applications of the normal distribution. We’ll explore this and other real-world uses of the normal distribution in this section.

College Entrance Exams

There are two good ways to compare two data values from different groups: using z-scores and using percentiles. The two methods will always give consistent results (meaning that we won’t find, for example, that the first value is better using z-scores but the second value is better using percentiles), so use whichever method is more comfortable for you.

EXAMPLE 8.41

Evaluating College Entrance Exam Scores

According to the [Digest of Education Statistics \(https://openstax.org/r/nces_ed_gov\)](https://openstax.org/r/nces_ed_gov), composite scores on the SAT have mean 1060 and standard deviation 195, while composite scores on the ACT have mean 21 and standard deviation 5.

1. At what percentile would an SAT score of 990 fall?
2. What is the z-score of an ACT score of 27?
3. Which is better: a score of 1450 on the SAT or 29 on the ACT?

Solution

1. Using Google Sheets, we can answer this question with the formula “=NORM.DIST(990, 1060, 195, TRUE)”. A score of 990 would fall at the 36th percentile.
2. Using the formula $z = \frac{x-\mu}{s}$, we get $z = \frac{27-21}{5} = 1.2$.
3. Let’s compare the values using both percentiles and z-values:

Percentiles: Using “=NORM.DIST(1450, 1060, 195, TRUE)” we find that an SAT score of 1450 is at the 98th percentile. Meanwhile, by entering “=NORM.DIST(29, 21, 5, TRUE)” we see that an ACT score of 29 is around the 95th percentile. Since it’s at a higher percentile, we can conclude that an SAT score of 1450 is better than an ACT score of 29.

z-scores: Using the formula, we see that the z-score for an SAT score of 1450 is $z = \frac{1450-1060}{195} = 2$, while the z-score for an ACT score of 29 is $z = \frac{29-21}{5} = 1.6$. Since it has a higher z-score, an SAT score of 1450 is better than an ACT score of 29.

> YOUR TURN 8.41

According to [the Graduate Management Admission Council \(GMAC\) \(https://openstax.org/r/mba_com\)](https://openstax.org/r/mba_com), the mean score on the GMAT (an entrance exam for graduate schools in business management) is 565, with standard deviation 116. For the LSAT (an entrance exam for law schools), [Kaplan \(https://openstax.org/r/kaptest_com\)](https://openstax.org/r/kaptest_com) informs us that the mean score is 150 with standard deviation 10.

1. What is the z-score for a GMAT score of 715?
2. At what percentile is an LSAT score of 166?
3. Which is better: a GMAT score of 650 or an LSAT score of 161?

Coin flipping

In the opening of [The Normal Distribution](#), we saw that the number of heads we get when we flip a coin 100 times is distributed normally. It can be shown that if n is the number of flips, then the mean of that distribution is $\frac{n}{2}$ and the standard deviation is $\frac{\sqrt{n}}{2}$ (as long as $n \geq 20$). So, for 100 flips, the mean of the distribution is 50 and the standard deviation is 5. In that opening example, one of our early runs gave us 70 heads in 100 flips, which we noted seemed unusual. Using the normal distribution, we can identify exactly how unusual that really is. Using Google Sheets, the formula “=NORM.DIST(70, 50, 5, TRUE)” gives us 0.999968, which is the 99.997th percentile! How is that useful? Suppose you need to test whether a coin is fair, and so you flip it 100 times. While we might be suspicious if we get 70 heads out of the 100 flips, we now have a numerical measure for how unusual that is: If the coin were fair, we would expect to see 70 heads (or more) only $100 - 99.9968 = 0.0032\%$ of the time. That’s really unlikely! Analysis like this is related to *hypothesis testing*, an important application of statistics in the sciences and social sciences.

EXAMPLE 8.42

Flipping a Coin

Let’s say we flip a coin 64 times and count the number of heads.

1. What would be the mean of the corresponding distribution?
2. What would be the standard deviation of the corresponding distribution?
3. Suppose we got 25 heads, which seems a little low. At what percentile would 25 heads fall?

✓ Solution

1. Since $n = 64$, the mean is $\frac{64}{2} = 32$.
2. Again using $n = 64$, we get a standard deviation of $\frac{\sqrt{64}}{2} = 4$.
3. Using “=NORM.DIST(25, 32, 4, TRUE)”, we see that 25 heads is at the 4th percentile. Reading and Interpreting Scatter Plots

> YOUR TURN 8.42

You flip a coin 144 times and count the number of heads.

1. What is the mean of the corresponding distribution?
2. What is the standard deviation of the corresponding distribution?
3. You flipped 81 heads. At what percentile does that fall?

Analyzing Data That Are Normally Distributed

Whenever we’re working with a dataset that has a distribution that looks symmetric and bell-shaped, we can use

techniques associated with the normal distribution to analyze the data.

EXAMPLE 8.43

Using Normal Techniques to Analyze Data

The data in “AvgSAT” (https://openstax.org/r/Chapter8_Data-Sets) contains the average SAT score for students attending every institution of higher learning in the United States for which data is available. In [Example 8.12](#), we created a histogram for these data:

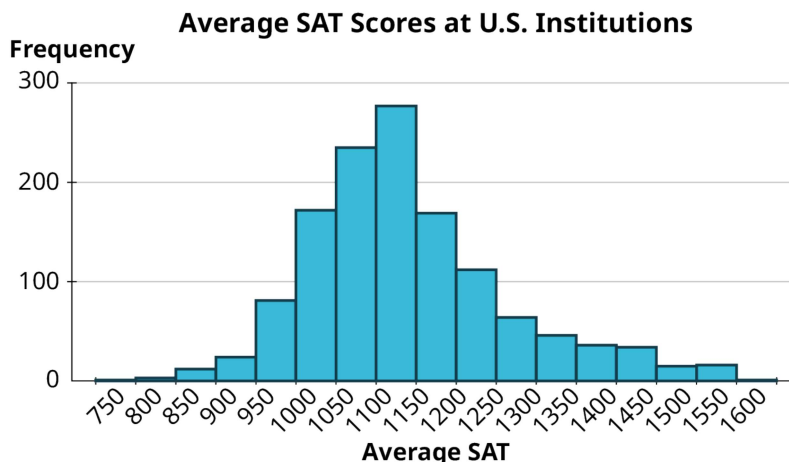


Figure 8.64 (data source: <https://data.ed.gov>)

This distribution is fairly symmetric (it’s just a little right-skewed) and bell-shaped, so we can use normal distribution techniques to analyze the data.

1. What is the mean of these average SAT scores?
2. What is the standard deviation of these SAT scores?
3. Using the answers to the previous two questions, use NORM.DIST in Google Sheets to estimate at what percentile the University at Buffalo in New York (average SAT: 1250) falls.
4. Use PERCENTRANK to find the actual percentile of the University at Buffalo, and see how close the estimate in the previous question came.

✓ Solution

1. Using the AVERAGE function in Google Sheets, we find that the mean is 1141.174.
2. Using the STDEV function, we get that the standard deviation is 125.517.
3. Entering “=NORM.DIST(1250, 1141, 125.517, TRUE)” into Google Sheets, we estimate that the University at Buffalo is at the 81st percentile.
4. Using PERCENTRANK, we find that the actual percentile is the 84th. These are close!

> YOUR TURN 8.43

1. Again using the data in “AvgSAT” (https://openstax.org/r/Chapter8_Data-Sets), find the average SAT score of a school at the 35th percentile in two ways: using NORM.INV and using PERCENTILE.

? WHO KNEW?

Political Meddling Exposed

The normal distribution pops up in some unusual places. Recently, a team at Duke University has been using statistics to help identify partisan gerrymandering, where electoral districts have been carefully drawn in a way that benefits one political party over another. In their analysis, they found that hypothetical election results in randomly drawn districts are normally distributed. By using techniques similar to the ones we used above, they can quantify precisely

how biased a particular electoral map is by finding the percentile rank of the actual election result on the normal distribution of the hypothetical results. You can find out more about their work at the "[Quantifying Gerrymandering](https://openstax.org/r/quantifying-gerrymandering)" [site here.](https://openstax.org/r/quantifying-gerrymandering) (<https://openstax.org/r/quantifying-gerrymandering>)

Check Your Understanding

For the following problems, recall that the SAT exam has mean 1060 and standard deviation 195, and that composite scores on the ACT have mean 21 and standard deviation 5.

49. At what percentile would an SAT score of 940 fall? Round to the nearest whole number.
50. What score would be at the 67th percentile on the ACT?
51. Which is a better score: 1300 on the SAT or 27 on the ACT?

For the following problems, recall that if we flip a coin at least 20 times, the distribution of the number of heads is approximately normal with mean equal to half the number of flips and standard deviation equal to half of the square root of the number of flips.

52. Suppose we flip a coin 120 times. What are the mean and standard deviation of the corresponding distribution of heads?
53. Let's say you flip 70 heads in 120 flips. At what percentile would that fall?
54. How many heads would be at the 30th percentile? Round to the nearest whole number.

For the following problems, use the data in "[World Tax,](https://openstax.org/r/Chapter8_Data-Sets)" (https://openstax.org/r/Chapter8_Data-Sets) which gives the tax revenue of many countries of the world in 2017, expressed as a percentage of their gross domestic products. You can assume that the distribution is approximately normal (but you can make a histogram to check, if you want).

55. What tax revenue percentage falls at the third quartile? Answer this question using Google Sheets in two ways: using PERCENTILE and using NORM.INV.
56. What tax revenue percentage falls at the 20th percentile? Answer this question using Google Sheets in two ways: using PERCENTILE and using NORM.INV.
57. At what percentile does the United Kingdom (25.62%, found on row 46 in the spreadsheet) fall? Answer this question using Google Sheets in two ways: using PERCENTRANK and using NORM.DIST.
58. At what percentile does Kiribati (21.97%, found on row 62 in the spreadsheet) fall? Answer this question using Google Sheets in two ways: using PERCENTRANK and using NORM.DIST.



SECTION 8.7 EXERCISES

For the following exercises, assume we're looking at results from two different standardized tests. The first, called the ABC, has mean 250 and standard deviation 50. The second, called the XYZ, has mean 80 and standard deviation 10.

1. What would be the z -score of a result of 322 on the ABC?
2. What would be the z -score of a result of 57 on the XYZ?
3. At what percentile would a result of 211 on the ABC fall? Round your answer to the nearest percentile.
4. At what percentile would a result of 94 on the XYZ fall? Round your answer to the nearest percentile.
5. What score would fall at the first quartile on the ABC? Round your answer to the nearest whole number.
6. What score would fall at the 60th percentile on the XYZ? Round your answer to the nearest whole number.
7. Which score is better: 202 on the ABC or 72 on the XYZ?
8. Which score is better: 324 on the ABC or 94 on the XYZ?

For the following exercises, recall that if we flip a coin at least 20 times, the distribution of the number of heads is approximately normal with mean equal to half the number of flips and standard deviation equal to half of the square root of the number of flips.

9. What would be the mean and standard deviation for the number of heads in 80 coin flips?
10. What would be the mean and standard deviation for the number of heads in 144 coin flips?
11. How many heads would be at the 80th percentile for 80 coin flips? Round your answer to the nearest whole number.
12. How many heads would be at the 20th percentile for 144 coin flips? Round your answer to the nearest whole number.
13. What is the z -score for 51 heads in 80 coin flips? Round your answer to the nearest hundredth.
14. What is the z -score for 87 heads in 144 coin flips? Round your answer to the nearest hundredth.
15. Which would be more unusual: flipping 51 heads in 80 tries or flipping 87 heads in 144 tries? How do you know?

For the following exercises, use the data in "[Wheat,](https://openstax.org/r/Chapter8_Data-Sets)" (https://openstax.org/r/Chapter8_Data-Sets) which gives the yield

of wheat (in bushels) produced over several equally sized plots. (These data were collected as part of an experiment reported in ["The experimental error of field trials"](https://openstax.org/r/sets-059) (<https://openstax.org/r/sets-059>) by W.B. Mercer and A.D. Hall).

16. Create a histogram of Yield to check that it's approximately normally distributed.
17. Find the mean and standard deviation of Yield. Round answers to the nearest thousandth.
18. What proportion of yields does the normal distribution predict should fall below 4.56? Round to the nearest thousandth.
19. At what percentile does the normal distribution predict a yield of 3.6 would fall? Round to the nearest whole number.
20. What yield does the normal distribution estimate would fall at the 80th percentile? Round to the nearest hundredth.
21. What yield does the normal distribution estimate would fall at the 95th percentile? Round to the nearest hundredth.
22. If one of these plots yields 2.8 bushels, would that be a good or bad result? How do you know?

8.8 Scatter Plots, Correlation, and Regression Lines

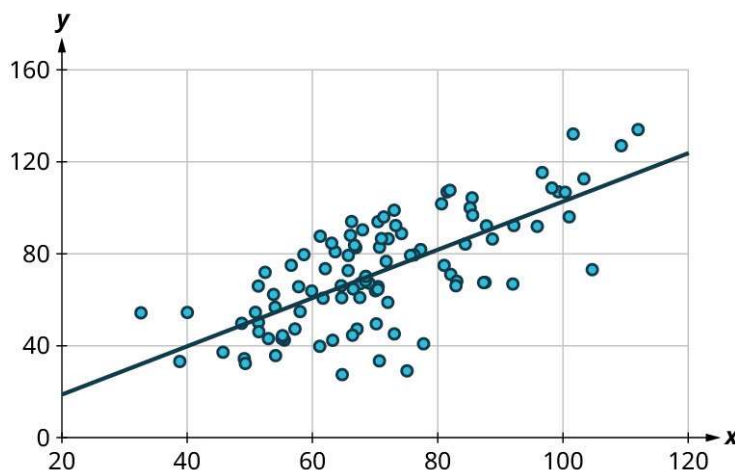


Figure 8.65 A scatter plot is a visualization of the relationship between quantitative dataset.

Learning Objectives

After completing this section, you should be able to:

1. Construct a scatter plot for a dataset.
2. Interpret a scatter plot.
3. Distinguish among positive, negative and no correlation.
4. Compute the correlation coefficient.
5. Estimate and interpret regression lines.

One of the most powerful tools statistics gives us is the ability to explore relationships between two datasets containing quantitative values, and then use that relationship to make predictions. For example, a student who wants to know how well they can expect to score on an upcoming final exam may consider reviewing the data on midterm and final exam scores for students who have previously taken the class. It seems reasonable to expect that there is a relationship between those two datasets: If a student did well on the midterm, they were probably more likely to do well on the final than the average student. Similarly, if a student did poorly on the midterm, they probably also did poorly on the final exam.

Of course, that relationship isn't set in stone; a student's performance on a midterm exam doesn't cement their performance on the final! A student might use a poor result on the midterm as motivation to study more for the final. A student with a really good grade on the midterm might be overconfident going into the final, and as a result doesn't prepare adequately.

The statistical method of *regression* can find a formula that does the best job of predicting a score on the final exam based on the student's score on the midterm, as well as give a measure of the confidence of that prediction! In this section, we'll discover how to use regression to make these predictions. First, though, we need to lay some graphical groundwork.