NORMAL PLOTS ON THE COMPUTER

The best way to tell whether your data can be modeled well by a Normal model is to make a picture or two. We've already talked about making histograms. Normal probability plots are almost never made by hand because the values of the Normal scores are tricky to find. But most statistics software make Normal plots, though various packages call the same plot by different names and array the information differently.

EXERCISES

1. **Shipments.** A company selling clothing on the Internet reports that the packages it ships have a median weight of 68 ounces and an IQR of 40 ounces.
   a) The company plans to include a sales flyer weighing 4 ounces in each package. What will the new median and IQR be?
   b) If the company recorded the shipping weights of these new packages in pounds instead of ounces, what would the median and IQR be? (1 lb. = 16 oz.)

2. **Hotline.** A company's customer service hotline handles many calls relating to orders, refunds, and other issues. The company's records indicate that the median length of calls to the hotline is 4.4 minutes with an IQR of 2.3 minutes.
   a) If the company were to describe the duration of these calls in seconds instead of minutes, what would the median and IQR be?
   b) In an effort to speed up the customer service process, the company decides to streamline the series of push-button menus customers must navigate, cutting the time by 24 seconds. What will the median and IQR of the length of hotline calls become?

3. **Payroll.** Here are the summary statistics for the weekly payroll of a small company: lowest salary = \$300, mean salary = \$700, median = \$500, range = \$1200, IQR = \$600, first quartile = \$350, standard deviation = \$400.
   a) Do you think the distribution of salaries is symmetric, skewed to the left, or skewed to the right? Explain why.
   b) Between what two values are the middle 50% of the salaries found?
   c) Suppose business has been good and the company gives every employee a \$50 raise. Tell the new value of each of the summary statistics.
   d) Instead, suppose the company gives each employee a 10% raise. Tell the new value of each of the summary statistics.

4. **Hams.** A specialty foods company sells "gourmet hams" by mail order. The hams vary in size from 4.15 to 7.45 pounds, with a mean weight of 6 pounds and standard deviation of 0.65 pounds. The quartiles and median weights are 5.6, 6.2, and 6.55 pounds.
   a) Find the range and the IQR of the weights.
   b) Do you think the distribution of the weights is symmetric or skewed? If skewed, which way? Why?
   c) If these weights were expressed in ounces (1 pound = 16 ounces) what would the mean, standard deviation, quartiles, median, IQR, and range be?
   d) When the company ships these hams, the box and packing materials add 30 ounces. What are the mean, standard deviation, quartiles, median, IQR, and range of weights of boxes shipped (in ounces)?
   e) One customer made a special order of a 10-pound ham. Which of the summary statistics of part d might *not* change if that data value were added to the distribution?

5. **SAT or ACT?** Each year thousands of high school students take either the SAT or the ACT, standardized tests used in the college admissions process. Combined SAT Math and Verbal scores go as high as 1600, while the maximum ACT composite score is 36. Since the two exams use very different scales, comparisons of performance are difficult. A convenient rule of thumb is $SAT = 40 \times ACT + 150$; that is, multiply an ACT score by 40 and add 150 points to estimate the equivalent SAT score. An admissions officer reported the following statistics about the ACT scores of 2355 students who applied to her college one year. Find the summaries of equivalent SAT scores.

   Lowest score = 19 Mean = 27 Standard deviation = 3

   Q3 = 30 Median = 28 IQR = 6

6. **Cold U?** A high school senior uses the Internet to get information on February temperatures in the town where he'll be going to college. He finds a Web site with some statistics, but they are given in degrees Celsius. The conversion formula is $°F = 9/5\ °C + 32$. Determine the Fahrenheit equivalents for the summary information below.

   Maximum temperature = 11°C Range = 33°

   Mean = 1° Standard deviation = 7°

   Median = 2° IQR = 16°

7. **Stats test.** Suppose your Statistics professor reports test grades as z-scores, and you got a score of 2.20 on an exam. Write a sentence explaining what that means.

8. **Checkup.** One of the authors has an adopted grandson whose birth family members are very short. After examining him at his 2-year checkup, the boy's pediatrician said that the z-score for his height relative to American 2-year-olds was −1.88. Write a sentence explaining what that means.

9. **Stats test, part II.** The mean score on the Stats exam was 75 points with a standard deviation of 5 points, and Gregor's z-score was -2. How many points did he score?

10. **Mensa.** People with z-scores above 2.5 on an IQ test are sometimes classified as geniuses. If IQ scores have a mean of 100 and a standard deviation of 16 points, what IQ score do you need to be considered a genius?

11. **Temperatures.** A town's January high temperatures average 36°F with a standard deviation of 10°, while in July the mean high temperature is 74° and the standard deviation is 8°. In which month is it more unusual to have a day with a high temperature of 55°? Explain.

12. **Placement exams.** An incoming freshman took her college's placement exams in French and mathematics. In French, she scored 82 and in math 86. The overall results on the French exam had a mean of 72 and a standard deviation of 8, while the mean math score was 68, with a standard deviation of 12. On which exam did she do better compared with the other freshmen?

13. **Combining test scores.** The first Stats exam had a mean of 65 and a standard deviation of 10 points; the second had a mean of 80 and a standard deviation of 5 points. Derrick scored an 80 on both tests. Julie scored a 70 on the first test and a 90 on the second. They both totaled 160 points on the two exams, but Julie claims that her total is better. Explain.

14. **Combining scores again.** The first Stat exam had a mean of 80 and a standard deviation of 4 points; the second had a mean of 70 and a standard deviation of 15 points. Reginald scored an 80 on the first test and an 85 on the second. Sara scored an 88 on the first but only a 65 on the second. Although Reginald's total score is higher, Sara feels she should get the higher grade. Explain her point of view.

15. **Final exams.** Anna, a language major, took final exams in both French and Spanish and scored 83 on each. Her roommate Megan, also taking both courses, scored 77 on the French exam and 95 on the Spanish exam. Overall, student scores on the French exam had a mean of 81 and a standard deviation of 5, and the Spanish scores had a mean of 74 and a standard deviation of 15.
    a) To qualify for language honors, a major must maintain at least an 85 average for all language courses taken. So far, which student qualifies?
    b) Which student's overall performance was better?

16. **MP3s.** Two companies market new batteries targeted at owners of personal music players. DuraTunes claims a mean battery life of 11 hours, while RockReady advertises 12 hours.
    a) Explain why you would also like to know the standard deviations of the battery lifespans before deciding which brand to buy.
    b) Suppose those standard deviations are 2 hours for DuraTunes and 1.5 hours for RockReady. You are headed for 8 hours at the beach. Which battery is most likely to last all day? Explain.
    c) If your beach trip is all weekend, and you probably will have the music on for 16 hours, which battery is most likely to last? Explain.

17. **Cattle.** The Virginia Cooperative Extension reports that the mean weight of yearling Angus steers is 1152 pounds. Suppose that weights of all such animals can be described by a Normal model with a standard deviation of 84 pounds.
    a) How many standard deviations from the mean would a steer weighing 1000 pounds be?
    b) Which would be more unusual, a steer weighing 1000 pounds or one weighing 1250 pounds?

T 18. **Car speeds.** John Beale of Stanford, CA, recorded the speeds of cars driving past his house, where the speed limit read 20 mph. The mean of 100 readings was 23.84 mph, with a standard deviation of 3.56 mph. (He actually recorded every car for a two-month period. These are 100 representative readings.)
    a) How many standard deviations from the mean would a car going under the speed limit be?
    b) Which would be more unusual, a car traveling 34 mph or one going 10 mph?

19. **More cattle.** Recall that the beef cattle described in Exercise 17 had a mean weight of 1152 pounds, with a standard deviation of 84 pounds.
    a) Cattle buyers hope that yearling Angus steers will weigh at least 1000 pounds. To see how much over (or under) that goal the cattle are, we could subtract 1000 pounds from all the weights. What would the new mean and standard deviation be?
    b) Suppose such cattle sell at auction for 40 cents a pound. Find the mean and standard deviation of the sale prices for all the steers.

T 20. **Car speeds again.** For the car speed data of Exercise 18, recall that the mean speed recorded was 23.84 mph, with a standard deviation of 3.56 mph. To see how many cars are speeding, John subtracts 20 mph from all speeds.
    a) What is the mean speed now? What is the new standard deviation?
    b) His friend in Berlin wants to study the speeds, so John converts all the original miles-per-hour readings to kilometers per hour by multiplying all speeds by 1.609 (km per mile). What is the mean now? What is the new standard deviation?

21. **Cattle, part III.** Suppose the auctioneer in Exercise 19 sold a herd of cattle whose minimum weight was 980 pounds, median was 1140 pounds, standard deviation 84 pounds, and IQR 102 pounds. They sold for 40 cents a pound, and the auctioneer took a \$20 commission on each animal. Then, for example, a steer weighing 1100 pounds would net the owner $0.40(1100) - 20 = \$420$. Find the minimum, median, standard deviation, and IQR of the net sale prices.

22. **Caught speeding.** Suppose police set up radar surveillance on the Stanford street described in Exercise 18. They handed out a large number of tickets to speeders going a mean of 28 mph, with a standard deviation of 2.4 mph, a maximum of 33 mph, and an IQR of 3.2 mph. Local law prescribes fines of \$100, plus \$10 per mile per hour over the 20 mph speed limit. For example, a driver convicted of going 25 mph would be fined $100 + 10(5) = \$150$. Find the mean, maximum, standard deviation, and IQR of all the potential fines.

23. **Professors.** A friend tells you about a recent study dealing with the number of years of teaching experience among current college professors. He remembers the mean but can't recall whether the standard deviation was 6 months, 6 years, or 16 years. Tell him which one it must have been, and why.

24. **Rock concerts.** A popular band on tour played a series of concerts in large venues. They always drew a large crowd, averaging 21,359 fans. While the band did not announce (and probably never calculated) the standard deviation, which of these values do you think is most likely to be correct: 20, 200, 2000, or 20,000 fans? Explain your choice.

25. **Guzzlers?** Environmental Protection Agency (EPA) fuel economy estimates for automobile models tested recently predicted a mean of 24.8 mpg and a standard deviation of 6.2 mpg for highway driving. Assume that a Normal model can be applied.
a) Draw the model for auto fuel economy. Clearly label it, showing what the 68–95–99.7 Rule predicts.
b) In what interval would you expect the central 68% of autos to be found?
c) About what percent of autos should get more than 31 mpg?
d) About what percent of cars should get between 31 and 37.2 mpg?
e) Describe the gas mileage of the worst 2.5% of all cars.

26. **IQ.** Some IQ tests are standardized to a Normal model, with a mean of 100 and a standard deviation of 16.
a) Draw the model for these IQ scores. Clearly label it, showing what the 68–95–99.7 Rule predicts.
b) In what interval would you expect the central 95% of IQ scores to be found?
c) About what percent of people should have IQ scores above 116?
d) About what percent of people should have IQ scores between 68 and 84?
e) About what percent of people should have IQ scores above 132?

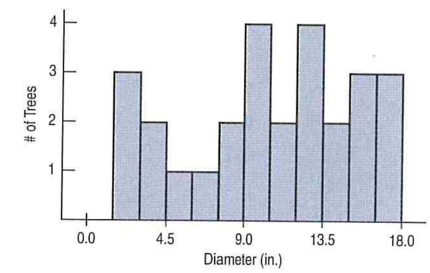
27. **Small steer.** In Exercise 17 we suggested the model $N(1152, 84)$ for weights in pounds of yearling Angus steers. What weight would you consider to be unusually low for such an animal? Explain.

28. **High IQ.** Exercise 26 proposes modeling IQ scores with $N(100, 16)$. What IQ would you consider to be unusually high? Explain.

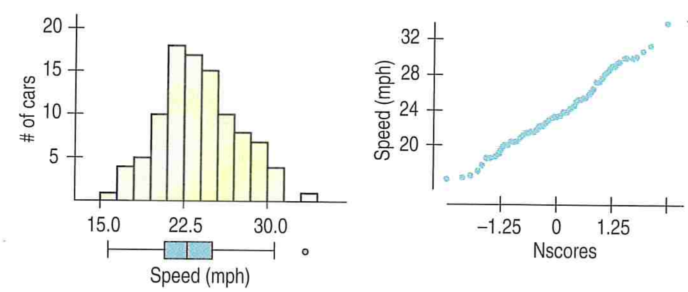
29. **Trees.** A forester measured 27 of the trees in a large woods that is up for sale. He found a mean diameter of 10.4 inches and a standard deviation of 4.7 inches. Suppose that these trees provide an accurate description of the whole forest and that a Normal model applies.
a) Draw the Normal model for tree diameters.
b) What size would you expect the central 95% of all trees to be?
c) About what percent of the trees should be less than an inch in diameter?
d) About what percent of the trees should be between 5.7 and 10.4 inches in diameter?
e) About what percent of the trees should be over 15 inches in diameter?

30. **Rivets.** A company that manufactures rivets believes the shear strength (in pounds) is modeled by $N(800, 50)$.
a) Draw and label the Normal model.
b) Would it be safe to use these rivets in a situation requiring a shear strength of 750 pounds? Explain.
c) About what percent of these rivets would you expect to fall below 900 pounds?
d) Rivets are used in a variety of applications with varying shear strength requirements. What is the maximum shear strength for which you would feel comfortable approving this company's rivets? Explain your reasoning.

31. **Trees, part II.** Later on, the forester in Exercise 29 shows you a histogram of the tree diameters he used in analyzing the woods that was for sale. Do you think he was justified in using a Normal model? Explain, citing some specific concerns.



T 32. **Car speeds, the picture.** For the car speed data of Exercise 18, here is the histogram, boxplot, and Normal probability plot of the 100 readings. Do you think it is appropriate to apply a Normal model here? Explain.



T 33. **Winter Olympics 2006 downhill.** Fifty-three men qualified for the men's alpine downhill race in Torino. The gold medal winner finished in 1 minute, 48.8 seconds. All competitors' times (in seconds) are found in the following list:

108.80	109.52	109.82	109.88	109.93	110.00
110.04	110.12	110.29	110.33	110.35	110.44
110.45	110.64	110.68	110.70	110.72	110.84
110.88	110.88	110.90	110.91	110.98	111.37
111.48	111.51	111.55	111.70	111.72	111.93
112.17	112.55	112.87	112.90	113.34	114.07
114.65	114.70	115.01	115.03	115.73	116.10
116.58	116.81	117.45	117.54	117.56	117.69
118.77	119.24	119.41	119.79	120.93	

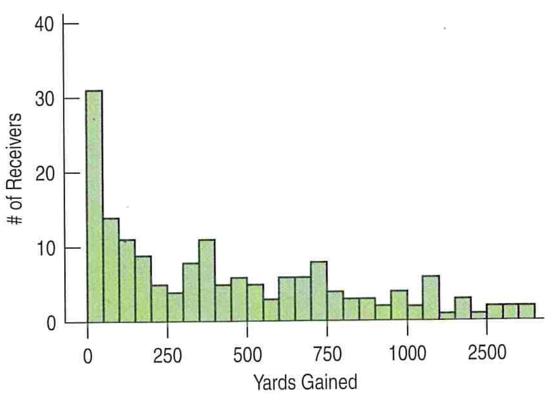
a) The mean time was 113.02 seconds, with a standard deviation of 3.24 seconds. If the Normal model is appropriate, what percent of times will be less than 109.78 seconds?
b) What is the actual percent of times less than 109.78 seconds?
c) Why do you think the two percentages don't agree?
d) Create a histogram of these times. What do you see?

T 34. **Check the model.** The mean of the 100 car speeds in Exercise 20 was 23.84 mph, with a standard deviation of 3.56 mph.
a) Using a Normal model, what values should border the middle 95% of all car speeds?
b) Here are some summary statistics.

Percentile		Speed
100%	Max	34.060
97.5%		30.976
90.0%		28.978
75.0%	Q3	25.785
50.0%	Median	23.525
25.0%	Q1	21.547
10.0%		19.163
2.5%		16.638
0.0%	Min	16.270

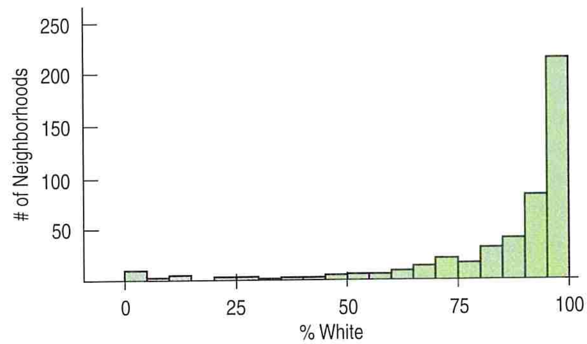
From your answer in part a, how well does the model do in predicting those percentiles? Are you surprised? Explain.

T 35. **Receivers.** NFL data from the 2006 football season reported the number of yards gained by each of the league's 167 wide receivers:



The mean is 435 yards, with a standard deviation of 384 yards.
a) According to the Normal model, what percent of receivers would you expect to gain fewer yards than 2 standard deviations below the mean number of yards?
b) For these data, what does that mean?
c) Explain the problem in using a Normal model here.

36. **Customer database.** A large philanthropic organization keeps records on the people who have contributed to their cause. In addition to keeping records of past giving, the organization buys demographic data on neighborhoods from the U.S. Census Bureau. Eighteen of these variables concern the ethnicity of the neighborhood of the donor. Here are a histogram and summary statistics for the percentage of whites in the neighborhoods of 500 donors:



Count	500
Mean	83.59
Median	93
StdDev	22.26
IQR	17
Q1	80
Q3	97

a) Which is a better summary of the percentage of white residents in the neighborhoods, the mean or the median? Explain.
b) Which is a better summary of the spread, the IQR or the standard deviation? Explain.
c) From a Normal model, about what percentage of neighborhoods should have a percent white within one standard deviation of the mean?
d) What percentage of neighborhoods actually have a percent white within one standard deviation of the mean?
e) Explain the discrepancy between parts c and d.

37. **Normal cattle.** Using $N(1152, 84)$, the Normal model for weights of Angus steers in Exercise 17, what percent of steers weigh
a) over 1250 pounds?
b) under 1200 pounds?
c) between 1000 and 1100 pounds?

38. **IQs revisited.** Based on the Normal model $N(100, 16)$ describing IQ scores, what percent of people's IQs would you expect to be
a) over 80?
b) under 90?
c) between 112 and 132?

39. **More cattle.** Based on the model $N(1152, 84)$ describing Angus steer weights, what are the cutoff values for
a) the highest 10% of the weights?
b) the lowest 20% of the weights?
c) the middle 40% of the weights?

40. **More IQs.** In the Normal model $N(100, 16)$, what cutoff value bounds
a) the highest 5% of all IQs?
b) the lowest 30% of the IQs?
c) the middle 80% of the IQs?

41. Cattle, finis. Consider the Angus weights model $N(1152, 84)$ one last time.

a) What weight represents the 40th percentile?
b) What weight represents the 99th percentile?
c) What's the IQR of the weights of these Angus steers?

42. IQ, finis. Consider the IQ model $N(100, 16)$ one last time.

a) What IQ represents the 15th percentile?
b) What IQ represents the 98th percentile?
c) What's the IQR of the IQs?

43. Cholesterol. Assume the cholesterol levels of adult American women can be described by a Normal model with a mean of 188 mg/dL and a standard deviation of 24.

a) Draw and label the Normal model.
b) What percent of adult women do you expect to have cholesterol levels over 200 mg/dL?
c) What percent of adult women do you expect to have cholesterol levels between 150 and 170 mg/dL?
d) Estimate the IQR of the cholesterol levels.
e) Above what value are the highest 15% of women's cholesterol levels?

44. Tires. A tire manufacturer believes that the treadlife of its snow tires can be described by a Normal model with a mean of 32,000 miles and standard deviation of 2500 miles.

a) If you buy a set of these tires, would it be reasonable for you to hope they'll last 40,000 miles? Explain.
b) Approximately what fraction of these tires can be expected to last less than 30,000 miles?
c) Approximately what fraction of these tires can be expected to last between 30,000 and 35,000 miles?
d) Estimate the IQR of the treadlives.
e) In planning a marketing strategy, a local tire dealer wants to offer a refund to any customer whose tires fail to last a certain number of miles. However, the dealer does not want to take too big a risk. If the dealer is willing to give refunds to no more than 1 of every 25 customers, for what mileage can he guarantee these tires to last?

45. Kindergarten. Companies that design furniture for elementary school classrooms produce a variety of sizes for kids of different ages. Suppose the heights of kindergarten children can be described by a Normal model with a mean of 38.2 inches and standard deviation of 1.8 inches.

a) What fraction of kindergarten kids should the company expect to be less than 3 feet tall?
b) In what height interval should the company expect to find the middle 80% of kindergarteners?
c) At least how tall are the biggest 10% of kindergarteners?

46. Body temperatures. Most people think that the "normal" adult body temperature is 98.6°F. That figure, based on a 19th-century study, has recently been challenged. In a 1992 article in the *Journal of the American Medical Association*, researchers reported that a more accurate figure may be 98.2°F. Furthermore, the standard deviation appeared to be around 0.7°F. Assume that a Normal model is appropriate.

a) In what interval would you expect most people's body temperatures to be? Explain.
b) What fraction of people would be expected to have body temperatures above 98.6°F?
c) Below what body temperature are the coolest 20% of all people?

47. Eggs. Hens usually begin laying eggs when they are about 6 months old. Young hens tend to lay smaller eggs, often weighing less than the desired minimum weight of 54 grams.

a) The average weight of the eggs produced by the young hens is 50.9 grams, and only 28% of their eggs exceed the desired minimum weight. If a Normal model is appropriate, what would the standard deviation of the egg weights be?
b) By the time these hens have reached the age of 1 year, the eggs they produce average 67.1 grams, and 98% of them are above the minimum weight. What is the standard deviation for the appropriate Normal model for these older hens?
c) Are egg sizes more consistent for the younger hens or the older ones? Explain.

48. Tomatoes. Agricultural scientists are working on developing an improved variety of Roma tomatoes. Marketing research indicates that customers are likely to bypass Romas that weigh less than 70 grams. The current variety of Roma plants produces fruit that averages 74 grams, but 11% of the tomatoes are too small. It is reasonable to assume that a Normal model applies.

a) What is the standard deviation of the weights of Romas now being grown?
b) Scientists hope to reduce the frequency of undersized tomatoes to no more than 4%. One way to accomplish this is to raise the average size of the fruit. If the standard deviation remains the same, what target mean should they have as a goal?
c) The researchers produce a new variety with a mean weight of 75 grams, which meets the 4% goal. What is the standard deviation of the weights of these new Romas?
d) Based on their standard deviations, compare the tomatoes produced by the two varieties.

JUST CHECKING
Answers

1. a) On the first test, the mean is 88 and the SD is 4, so $z = (90 - 88)/4 = 0.5$. On the second test, the mean is 75 and the SD is 5, so $z = (80 - 75)/5 = 1.0$. The first test has the lower z-score, so it is the one that will be dropped.
   b) No. The second test is 1 standard deviation above the mean, farther away than the first test, so it's the better score relative to the class.
2. a) The mean would increase to 500.
   b) The standard deviation is still 100 points.
   c) The two boxplots would look nearly identical (the shape of the distribution would remain the same), but the later one would be shifted 50 points higher.
3. The standard deviation is now 2.54 millimeters, which is the same as 0.1 inches. Nothing has changed. The standard deviation has "increased" only because we're reporting it in millimeters now, not inches.
4. The mean is 184 centimeters, with a standard deviation of 8 centimeters. 2 meters is 200 centimeters, which is 2 standard deviations above the mean. We expect 5% of the men to be more than 2 standard deviations below or above the mean, so half of those, 2.5%, are likely to be above 2 meters.
5. a) We know that 68% of the time we'll be within 1 standard deviation (2 min) of 20. So 32% of the time we'll arrive in less than 18 or more than 22 minutes. Half of those times (16%) will be greater than 22 minutes, so 84% will be less than 22 minutes.
   b) 24 minutes is 2 standard deviations above the mean. Because of the 95% rule, we know 2.5% of the times will be more than 24 minutes.
   c) Traffic incidents may occasionally increase the time it takes to get to school, so the driving times may be skewed to the right, and there may be outliers.
   d) If so, the Normal model would not be appropriate and the percentages we predict would not be accurate.