11-10 Normal Distributions

Content Standard

S.ID.4 Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate . . .

Objective To use a normal distribution

SOLVE IT! Getting Ready!

Even and odd functions are defined as follows.

Even function: $f(x) = f(-x)$

Odd function: $-f(x) = f(-x)$

Which is the graph of an even function? Of an odd function? Justify your answers.

Try it out. Suppose (2, 2) is a point on $f(x)$. If $f(x)$ is even, what other point is on the graph of $f(x)$?

MATHEMATICAL PRACTICES A **discrete probability distribution** has a finite number of possible events, or values. The binomial probability distribution you studied in the preceding lesson is a discrete probability distribution.

The events for a **continuous probability distribution** can be any value in an interval of real numbers. If a data set is large, the distribution of its discrete values approximates a continuous distribution.

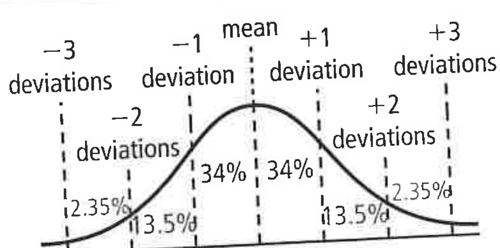
Lesson Vocabulary
- discrete probability distribution
- continuous probability distribution
- normal distribution

Essential Understanding Many common statistics (such as human height, weight, or blood pressure) gathered from samples in the natural world tend to have a *normal distribution* about their mean.

A **normal distribution** has data that vary randomly from the mean. The graph of a normal distribution is a normal curve.

take note

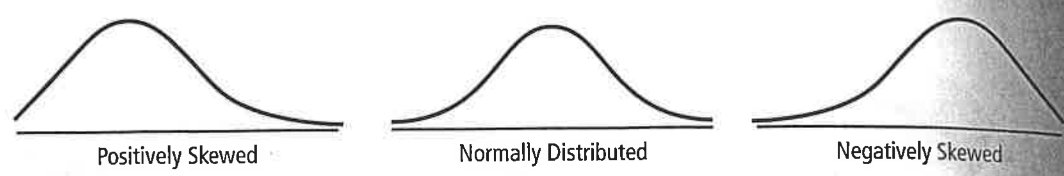
Key Concept Normal Distribution



In a normal distribution,
- 68% of data fall within one standard deviation of the mean
- 95% of data fall within two standard deviations of the mean
- 99.7% of data fall within three standard deviations of the mean

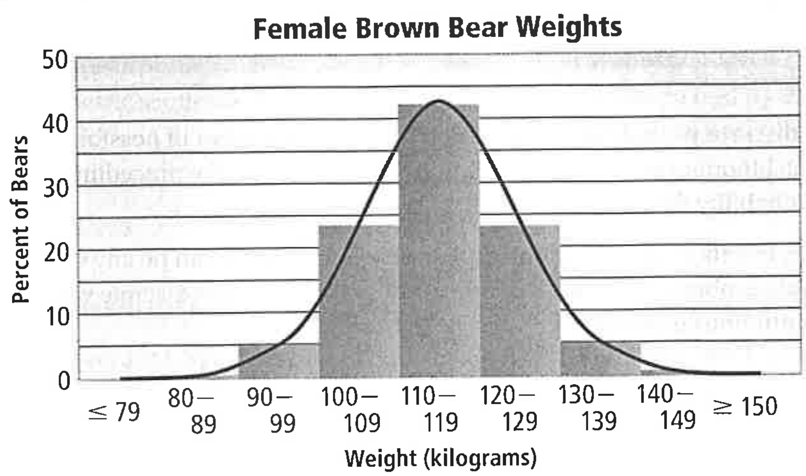
A normal distribution has a symmetric bell shape centered on the mean.

Sometimes data are not normally distributed. A data set could have a distribution that is *skewed*, an asymmetric curve where one end stretches out further than the other end. When a data set is skewed, the data do not vary predictably from the mean. This means that the data do not fall within the standard deviations of the mean like normally distributed data, and so it is inappropriate to use mean and standard deviation to estimate percentages for skewed data.



Problem 1 Analyzing Normally Distributed Data STEM

Zoology The bar graph gives the weights of a population of female brown bears. The red curve shows how the weights are normally distributed about the mean, 115 kg. Approximately what percent of female brown bears weigh between 100 and 129 kg?



Plan

How do you find this percent?
The percents for each bar are based on the same sample population of bears. You can add the percents.

Estimate and add the percents for the intervals 100–109, 110–119, and 120–129.

$23 + 42 + 23 = 88$

About 88% of female brown bears weigh between 100 and 129 kg.

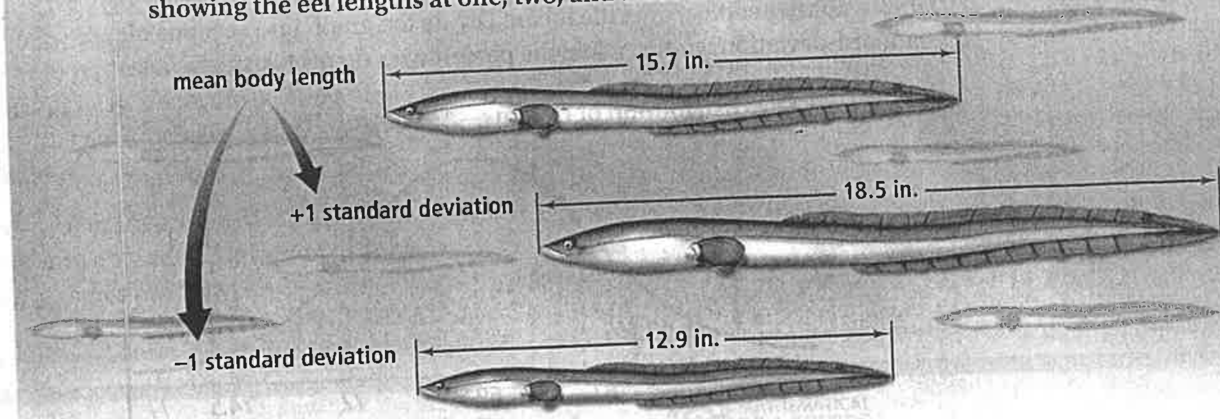
Got It? 1. a. Approximately what percent of female brown bears in Problem 1 weigh less than 120 kg?

b. The standard deviation in the weights of female brown bears is about 10 kg. Approximately what percent of female brown bears have weights that are within 1.5 standard deviations of the mean?

When data are normally distributed, you can sketch the graph of the distribution using the fact that a normal curve has a symmetric bell shape.

Problem 2 Sketching a Normal Curve STEM

Zoology For a population of male European eels, the mean body length and one positive and negative standard deviation is shown below. Sketch a normal curve showing the eel lengths at one, two, and three standard deviations from the mean.



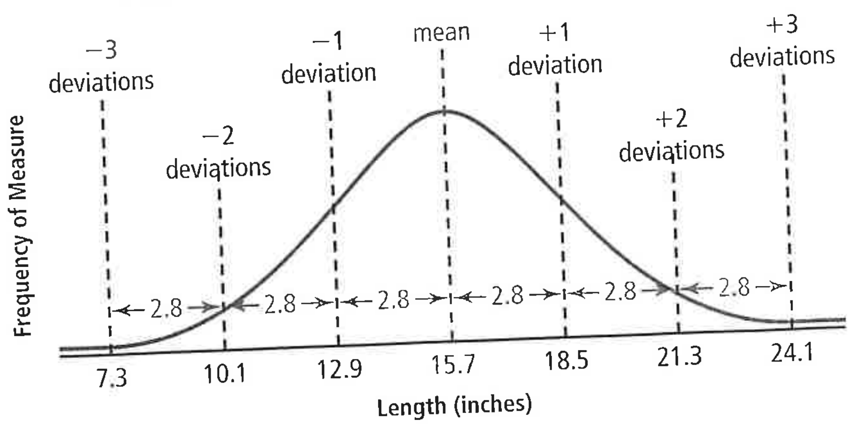
Know
The mean and the standard deviation of the population

Need
Lengths that are one, two, and three standard deviations from the mean

Plan
- Multiply the standard deviation by 1, 2, and 3.
- Draw vertical lines at the mean ± these values.
- Sketch the normal curve.

Think
How high do you draw the curve?
Unless you actually label the vertical scale, it doesn't matter.

Distribution of Body Lengths for Male European Eels



Got It? **2.** For a population of female European eels, the mean body length is 21.1 in. The standard deviation is 4.7 in. Sketch a normal curve showing eel lengths at one, two, and three standard deviations from the mean.

When you show a probability distribution as a bar graph, the height of the bar indicates probability. For a normal distribution, however, the area between the curve and an interval on the x-axis represents probability.

Problem 3 Analyzing a Normal Distribution

Plan

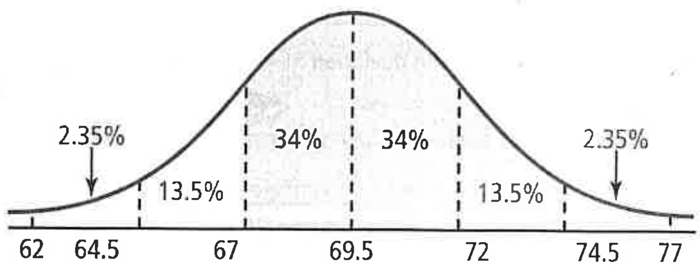
How do you divide the graph of the distribution?
Draw vertical lines at intervals that are one standard deviation wide, on both sides of the mean.

The heights of adult American males are approximately normally distributed with mean 69.5 in. and standard deviation 2.5 in.

A What percent of adult American males are between 67 in. and 74.5 in. tall?

Draw a normal curve. Label the mean. Divide the graph into sections of standard-deviation widths. Label the percentages for each section.

Distribution of Heights—Adult American Males



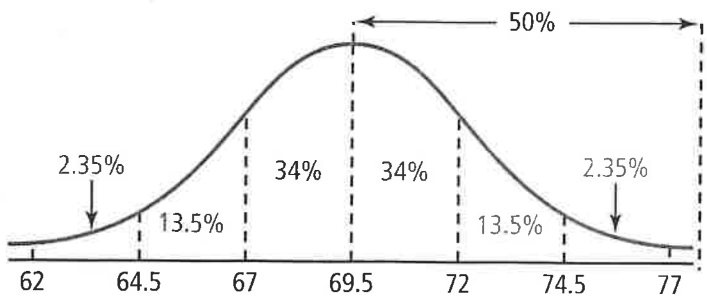
$P(67 < \text{height} < 74.5) = 0.34 + 0.34 + 0.135 = 0.815$

About 82% of adult American males are between 67 in. and 74.5 in. tall.

B In a group of 2000 adult American males, about how many would you expect to be taller than 6 ft (or 72 in.)?

Because the graph is symmetric about the mean, the right half of the distribution contains 50% of the data. If you subtract everything between 69.5 in. and 72 in. from the right half, only the part of the distribution that is greater than 72 in. remains.

Distribution of Heights—Adult American Males



$P(\text{height} > 72) = 0.50 - 0.34 = 0.16$

You would expect about 16% of the 2000 adult American males to be taller than 72 in. You would expect about $0.16 \cdot 2000 = 320$ to be over 6 ft tall.

Got It? 3. The scores on the Algebra 2 final are approximately normally distributed with a mean of 150 and a standard deviation of 15.

a. What percentage of the students who took the test scored above 180?

b. If 250 students took the final exam, approximately how many scored above 135?

c. **Reasoning** If 13.6% of the students received a B on the final, how can you describe their scores? Explain.

Lesson Check

Do you know HOW?

1. Use the graph from Problem 1. What is the approximate percent of female brown bears weighing at least 100 kg?
2. Draw a curve to represent a normally distributed experiment that has a mean of 180 and a standard deviation of 15. Label the x-axis and indicate the probabilities.
3. The scores on an exam are normally distributed, with a mean of 85 and a standard deviation of 5. What percent of the scores are between 85 and 95?

Do you UNDERSTAND? MATHEMATICAL PRACTICES

4. **Vocabulary** Why is a normal distribution "normal"?
5. **Compare and Contrast** How do the mean and median compare in a normal distribution?
6. **Reasoning** What is the effect on a normal distribution if each data value increases by 10? Justify your answer.

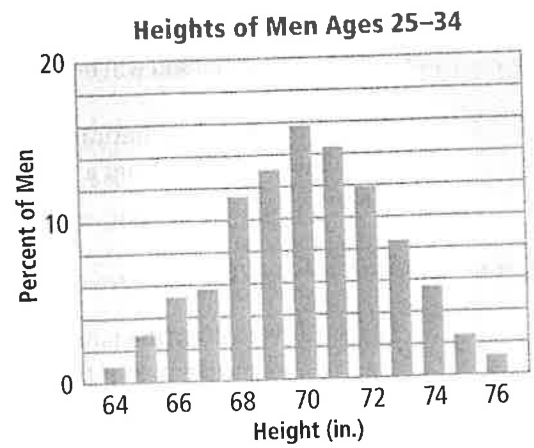
Practice and Problem-Solving Exercises MATHEMATICAL PRACTICES

A Practice

See Problem 1.

Biology The heights of men in a survey are normally distributed about the mean. Use the graph for Exercises 7–10.

7. About what percent of men aged 25 to 34 are 69–71 in. tall?
8. About what percent of men aged 25 to 34 are less than 70 in. tall?
9. Suppose the survey included data on 100 men. About how many would you expect to be 69–71 in. tall?
10. The mean of the data is 70, and the standard deviation is 2.5. Approximately what percent of men are within one standard deviation of the mean height?



See Problem 2.

Sketch a normal curve for each distribution. Label the x-axis values at one, two, and three standard deviations from the mean.

11. mean = 45, standard deviation = 5
12. mean = 45, standard deviation = 10
13. mean = 45, standard deviation = 2
14. mean = 45, standard deviation = 3.5

See Problem 3.

A set of data has a normal distribution with a mean of 50 and a standard deviation of 8. Find the percent of data within each interval.

15. from 42 to 58
16. greater than 34
17. less than 50

B Apply

18. Think About a Plan The numbers of paper clips per box in a truckload of boxes are normally distributed, with a mean of 100 and a standard deviation of 5. Find the probability that a box will *not* contain between 95 and 105 clips.

- How should you label the vertical lines on the graph of the normal distribution?
- Which parts of the graph are *not* between 95 and 105 clips?

19. a. From the table at the right, select the set of values that appears to be distributed normally.

b. Using the set you chose in part (a), make a histogram of the values.

c. Sketch a normal curve over your graph.

Set 1	Set 2	Set 3
1	5	5
10	7	6
5	7	9
19	7	1
2	4	1
7	11	5
1	7	11
7	7	1
2	7	10
10	9	4
6	7	2
9	7	8

20. Writing In a class of 25, one student receives a score of 100 on a test. The grades are distributed normally, with a mean of 78 and a standard deviation of 5. Do you think the student's score is an outlier? Explain.

21. Sports To qualify as a contestant in a race, a runner has to be in the fastest 16% of all applicants. The running times are normally distributed, with a mean of 63 min and a standard deviation of 4 min. To the nearest minute, what is the qualifying time for the race?

22. Agriculture To win a prize at the county fair, the diameter of a tomato must be greater than 4 in. The diameters of a crop of tomatoes grown in a special soil are normally distributed, with a mean of 3.2 in. and a standard deviation of 0.4 in. What is the probability that a tomato grown in the special soil will be a winner?

A normal distribution has a mean of 100 and a standard deviation of 10. Find the probability that a value selected at random is in the given interval.

23. from 80 to 100

24. from 70 to 130

25. from 90 to 120

26. at least 100

27. at most 110

28. at least 80

STEM **29. Weather** The table at the right shows the number of tornadoes that were recorded in the U.S. in 2008.

a. Draw a histogram to represent the data.

b. Does the histogram approximate a normal curve? Explain.

c. Is it appropriate to use a normal curve to estimate the percent of tornados that occur during certain months of the year? Explain.

Month	Tornadoes
1	84
2	147
3	129
4	189
5	461
6	294
7	93
8	101
9	111
10	21
11	20
12	40

30. Error Analysis In a set of data, the value 332 is 3 standard deviations from the mean and the value 248 is 1 standard deviation from the mean. A classmate claims that there is only one possible mean and standard deviation for this data set. Do you agree? Explain.

31. Reasoning Jake and Elena took the same standardized test, but are in different classes. They both received a score of 87. In Jake's group, the mean was 80 and the standard deviation was 6. In Elena's group, the mean was 76 and the standard deviation was 4. Did either student score in the top 10% of his or her group? Explain.

Challenge

32. **Manufacturing** Tubs of yogurt weigh 1.0 lb each, with a standard deviation of 0.06 lb. At a quality control checkpoint, 12 of the tubs taken as samples weighed less than 0.88 lb. Assume that the weights of the samples were normally distributed. How many tubs of yogurt were taken as samples?

33. **Reasoning** Describe how you can use a normal distribution to approximate a binomial distribution. Draw a binomial histogram and a normal curve to help with your explanation.

Standardized Test Prep

SAT/ACT

34. For a daily airline flight between two cities, the number of pieces of checked luggage has a mean of 380 and a standard deviation of 20. On what percent of the flights would you expect from 340 to 420 pieces of checked luggage?

(A) 34% (B) 47.5% (C) 68% (D) 95%

35. A jar contains 37 pennies, 53 nickels, 29 dimes, and 21 quarters. A coin is drawn at random from the jar. What is the probability that the coin drawn is NOT a quarter?

(F) $\frac{56,869}{2,744,000}$ (G) $\frac{3}{20}$ (H) $\frac{3}{17}$ (I) $\frac{17}{20}$

36. A multiple-choice quiz contains five questions, each with three answer choices. You select all five answer choices at random. What is the best estimate of the probability that you will get at least four answers correct?

(A) 4.1% (B) 4.5% (C) 13.2% (D) 46.1%

Short Response

37. Distribution A has 50 data values with mean 40 and standard deviation 2.4. Distribution B has 30 data values with mean 40 and standard deviation 2.8. Which distribution has more data values at or below 40? Show your work.

Mixed Review

Find the probability of x successes in n trials for the given probability of success p on each trial. See Lesson 11-9.

38. $x = 4, n = 7, p = 0.2$ 39. $x = 2, n = 9, p = 0.4$ 40. $x = 6, n = 10, p = 0.3$

Graph each equation. Identify the conic section and describe the graph and its lines of symmetry. Then find the domain and range. See Lesson 10-1.

41. $x^2 + y^2 = 64$ 42. $x^2 - y^2 = 9$ 43. $9x^2 + 25y^2 = 225$

Get Ready! To prepare for Lesson 12-1, do Exercises 44–46.

Write an equation for each horizontal translation of $y = x - 2$. Then graph each translation. See Lesson 2-6.

44. 1 unit right 45. 2 units left 46. $\frac{3}{4}$ unit left

Concept Byte

For Use With Lesson 11-10

ACTIVITY

Margin of Error

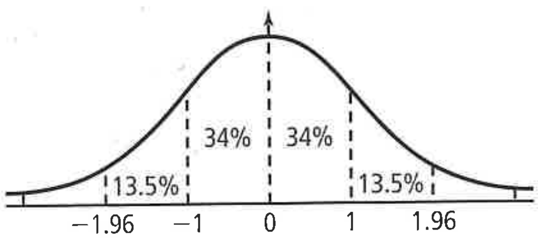
Content Standard

S.IC.4 Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

The mean of a sample may or may not be the mean of the population the sample was drawn from. The **margin of error** helps you find the interval in which the mean of the population is likely to be. The margin of error is based on the size of the sample and the *confidence level* desired.

A 95% confidence level means that the probability is 95% that the true population mean is within a range of values called a **confidence interval**. It also means that when you select many different large samples from the same population, 95% of the confidence intervals will actually contain the population mean.

The means of all the samples follow a normal distribution.



The normal distribution above shows that 95% of values are between −1.96 and 1.96 standard deviations from the mean. To find the margin of error based on the mean of a large set of data at a 95% confidence level, you use the formula $ME = 1.96 \cdot \frac{s}{\sqrt{n}}$, where ME is the margin of error, s is the standard deviation of the sample data, and n is the number of values in the sample. The confidence interval for the population mean μ (pronounced *myoo*) is $\overline{x} - ME \leq \mu \leq \overline{x} + ME$, where $\overline{x}$ is the sample mean.

Activity 1

A grocery store manager wanted to determine the wait times for customers in the express lines. He timed customers chosen at random.

Waiting Time (minutes)				
3.3	5.1	5.2	6.7	7.3
7.5	4.6	6.2	5.5	3.6
3.4	3.5	8.2	4.2	3.8
4.7	4.6	4.7	4.5	9.7
5.4	5.9	6.7	6.5	8.2
3.1	3.2	8.2	2.5	4.8

1. What is the mean and stardard deviation of the sample? Round to the nearest tenth of a minute.
2. At a 95% confidence level, what is the approximate margin of error? Round to the nearest tenth of a minute.
3. What is the confidence interval for a 95% confidence level?
4. What is the meaning of the interval in terms of wait times for customers?

You can also find the margin of error and the confidence interval for a sample proportion. A **sample proporton** $\hat{p}$ is the ratio $\frac{x}{n}$, where x is the number of times an event occurs in a sample of size n.

Activity 2

What is the sample proportion for each situation? Write the ratios as percents rounded to the nearest tenth of a percent.

5. In a poll of 1085 voters selected randomly, 564 favor Candidate A.

6. A coin is tossed 40 times, and it comes up heads 25 times.

To find the margin of error for a sample proportion at a 95% confidence level, use the formula $ME = 1.96 \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$, where ME is the margin of error, $\hat{p}$ is the sample proportion, and n is the sample size. The confidence interval for the population proportion p is $\hat{p} - ME \leq p \leq \hat{p} + ME$.

Activity 3

Find (a) the sample proportion, (b) the margin of error, and (c) the 95% confidence interval for the population proportion.

7. In a survey of 530 randomly selected high school students, 280 preferred watching football to watching basketball.

8. In a simple random sample of 500 people, 342 reported using social networking sites on the Internet.

Exercises

For Exercises 9–10, find the 95% confidence interval for the population mean or population proportion, and interpret the confidence interval in context.

9. A consumer research group tested the battery life of 36 randomly chosen batteries to establish the likely battery life for the population of same type of battery.

Battery Life (In Hours)			
63.2	84.6	78.4	85.8
62.1	81.8	63.6	64.2
79.4	75.2	54.1	73.4
66.3	74.5	71.6	60.1
61.2	74.5	72.4	81.3
61.4	83.6	75.6	74.1
68.3	82.2	59.3	47.6
86.2	64.3	72.7	71.8
71.4	63.6	59.6	68.1

10. In a poll of 720 likely voters, 358 indicate they plan to vote for Candidate A.

11. **Data Collection** Roll a number cube 30 times. Record the results from each roll. In parts (a) and (b), find the sample proportion, the margin of error for a 95% confidence level, and the 95% confidence interval for the population proportion.
    - **a.** rolling a 2
    - **b.** rolling a 3
    - **c.** Is the 95% confidence intervals for the population proportion about the same for rolling a 2 and for rolling a 3?
    - **d.** Compare your sample proportions to the theoretical proportions for parts (a) and (b). Would you expect the theoretical proportion to be within the confidence intervals you found? Explain.

Concept Byte

For Use With Lesson 11-10

ACTIVITY

Drawing Conclusions from Samples

Content Standard

S.IC.5 Using data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

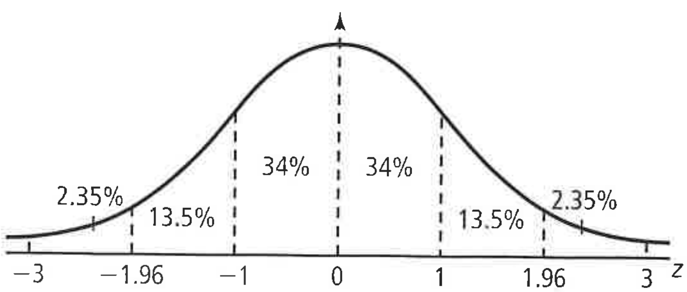
You can compare samples to determine if the difference in mean or proportion for a large population, based on a given confidence level, is significant. If a population is large and there are at least 30 data points in a sample, then the means and proportions can be compared using a normal distribution.

An important measure for normally distributed data is the **z-score,** which indicates the number of standard deviations a value lies above or below the mean of a population. When finding the z-score of a data point of a population, the formula is $z = \frac{x - \mu}{\sigma}$, where x is a data point, μ is the mean of a population, and σ is the standard deviation of the population.

Activity 1

In a given population, the weights of newborns are normally distributed about the mean 3250 g. The standard deviation of the population is 500 g.

1. What is the z-score of a newborn weighing 2500 g?
2. What is the z-score of a newborn weighing 4500 g?
3. What is the probability that a newborn weighs between 2270 g and 4230 g? Use z-scores of the weights and the normal curve.



To compare the mean of a sample with the mean of a population, you use the formula $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$, where $\bar{x}$ is the mean of the sample, μ is the mean of the population, σ is the standard deviation of the population, and n is the sample size. A z-score that is between -1.96 and 1.96 means that at a 95% confidence level, there is not a significant difference between the sample mean and the population mean. A z-score of less than -1.96 or greater than 1.96 indicates that at a 95% confidence level the differences between the sample mean and the population mean is significant and the differences are not simply due to chance.

Activity 2

A company that develops fertilizers wants to know whether either of the two new fertilizers they have in development shows a significant difference in the growth of plants based on a 95% confidence level. The company has data on the growth of bean plants without fertilizers. For a growth period of one month, the population of the beans grown without fertilizers have a mean of 20 cm with a standard deviation of 1 cm.

4. A company researcher chooses 30 plants at random and uses fertilizer A for one month. The researcher finds that after using fertilizer A, the mean of the bean plants' growth is 20.4 cm. What is the z-score for the mean of the sample treated with fertilizer A compared to the population of bean plants without fertilizer?

5. Does fertilizer A meet a 95% confidence level for having growth that is significantly different from growth without fertilizer?

6. To test fertilizer B, the researcher chooses 35 plants at random. The bean plants' growth averages 20.3 cm. What is the z-score for the mean of the sample treated with fertilizer B compared to the population of bean plants without fertilizer?

7. Does fertilizer B meet a 95% confidence level for having growth that is significantly different from growth without fertilizer?

8. Based on these data, would you advise the company to market fertilizer A or fertilizer B? Explain.

To compare the proportion of a sample with the proportion of a population, you use the formula for the z-score for a proportion, $z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$, where $\hat{p}$ is the sample proportion, p is the population proportion, and n is the sample size. For a 95% confidence level, a z-score between -1.96 and 1.96 means that difference between the sample and the population is not significant. A z-score of less than -1.96 or greater than 1.96 means that the difference in the proportions is significant, and not simply due to chance.

Activity 3

Suppose your teacher accidentally gives you a multiple-choice calculus test, instead of an Algebra 2 test. There are 40 true-false questions, and you and your classmates answer the questions randomly. The probability of getting an answer correct is $\frac{1}{2}$. You can use $\frac{1}{2}$ as the population proportion.

9. Develop a simulation for answering the 40 questions. Record your results.

10. What is the z-score for the sample based on your trials?

11. Compare your results with the results obtained by the rest of the class. At a 95% confidence level, what percentage of the results in the class showed a statistically significant difference from a score you would expect to get by guessing?