

OVER THE HILL — AGING ON A NORMAL CURVE

TEACHER VERSION

Subject Level:
High School Math

Grade Level:
11–12

Approx. Time Required:
30–45 minutes

Learning Objectives:

- Students will be able to estimate percentages based on a normal distribution using the "empirical rule."
- Students will be able to use a data set's mean and standard deviation to fit it to a normal distribution.

Activity Description

Students will use census data from a sample of 136 U.S. counties and other sample data to make estimates about the U.S. population that is 65 or older in all other counties and about other variables, using normal distribution models.

Suggested Grade Level:
11–12

Approximate Time Required:
30–45 minutes

Learning Objectives:

- Students will be able to estimate percentages based on a normal distribution using the "empirical rule."
- Students will be able to use a data set's mean and standard deviation to fit it to a normal distribution.

Topics:

- Empirical rule
- Histograms
- Normal models
- Proportions

Skills Taught:

- Identifying normality in a data set
- Estimating percentages from a normal distribution
- Making predictions

Materials Required

- The student version of this activity, 12 pages
- Calculators

Activity Items

The following items are part of this activity. The items, their data sources, and instructions for viewing the source data online appear at the end of this teacher version.

- Item 1: Percentage of Population Aged 65 and Older
- Item 2: Percentage of Family Households
- Item 3: Average Hours of Student Sleep on a School Night

For more information to help you introduce your students to the U.S. Census Bureau, read "*Census Bureau 101 for Students*." This information sheet can be printed and passed out to your students as well.

Standards Addressed

See charts below. For more information, read "*Overview of Education Standards and Guidelines Addressed in Statistics in Schools Activities*."

Common Core State Standards for Mathematics

Standard	Domain	Cluster
CCSS.MATH.CONTENT.HSS.ID.A.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. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.	ID – Interpreting Categorical & Quantitative Data	Summarize, represent, and interpret data on a single count or measurement variable.

Common Core State Standards for Mathematical Practice

Standard
CCSS.MATH.PRACTICE.MP2. Reason abstractly and quantitatively. Students will use means and standard deviations to draw conclusions about data distributions, based on a normal distribution model.

National Council of Teachers of Mathematics' Principles and Standards for School Mathematics

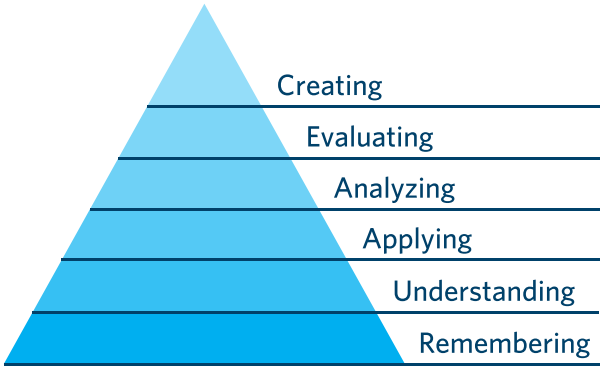
Content Standard	Students should be able to:	Expectation for Grade Band
Data Analysis and Probability	Select and use appropriate statistical methods to analyze data.	For univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics.

Guidelines for Assessment and Instruction in Statistics Education

GAISE	Level A	Level B	Level C
Formulate Questions			
Collect Data			
Analyze Data			X
Interpret Results			X

Bloom's Taxonomy

Students will engage with, ***evaluate***, and ***analyze*** mathematical problems involving normal models.



Teacher Notes

Before the Activity

Students must understand the following key terms:

- **Family household** – a household in which at least one person is related to the householder by birth, marriage, or adoption
- **Mean (μ)** – a measure of center in a set of numerical data, computed by adding together the values in a list and then dividing by the number of values in the list
- **Median** – a measure of center in a set of numerical data, identified as the value appearing at the middle of a sorted version of the list (or the mean of the two middle values if the list contains an even number of values)
- **Shape** – the general form of a data distribution (e.g., bell-shaped, bimodal, irregular, uniform)
- **Normal distribution** – a distribution of a data set that appears symmetrical and creates a bell-shaped curve, with equal values for the mean and median
- **Standard deviation (SD or σ)** – a measure of spread for a set of numerical data, calculated by taking the square root of the variance, that increases in value as the data in the set become more spread out
- **Variance** – a measure of variability in a data set, calculated by first finding the numerical difference between each value and the mean, squaring the differences, and then finding the average of all the squared differences

During the Activity

Teachers should review the different aspects of the histograms in the items to confirm that students understand how to read them. Teachers should also make sure students understand that the mean and the median are equal in normal distributions and that not all data distributions fit under a normal model.

After the Activity

Teachers should ask students to reflect on what they learned.

Extension Idea

- Teachers could ask students to use the Census at School Random Sampler (*amstat.org/censusatschool/randomsampleconditions.cfm*) to first identify a data set that they think will have a normal distribution and another that will not, and then look at the histograms to evaluate their conjectures.

Student Activity

Click ***here*** to download a printable version for students.

Activity Items

The following items are part of this activity and appear at the end of this student version.

- Item 1: Percentage of Population Aged 65 and Older
- Item 2: Percentage of Family Households
- Item 3: Average Hours of Student Sleep on a School Night

Student Learning Objectives

- I will be able to estimate percentages based on a normal distribution using the "empirical rule."
- I will be able to use a data set's mean and standard deviation to fit it to a normal distribution.

Part 1 – Examine a Normal Distribution

1. Examine the histogram in **Item 1: Percentage of Population Aged 65 and Older**, which shows data for all counties that are designated by the U.S. Census Bureau in the United States. What do you notice?

   Student answers will vary but could include: The numbers on the x-axis represent the percentage of people 65 and older in each county. The bars on the left of the graph represent counties with lower percentages of these residents. The tallest bars encompass more than 300 counties.

2. Approximately how many counties have populations in which roughly 24 percent of people are 65 and older?

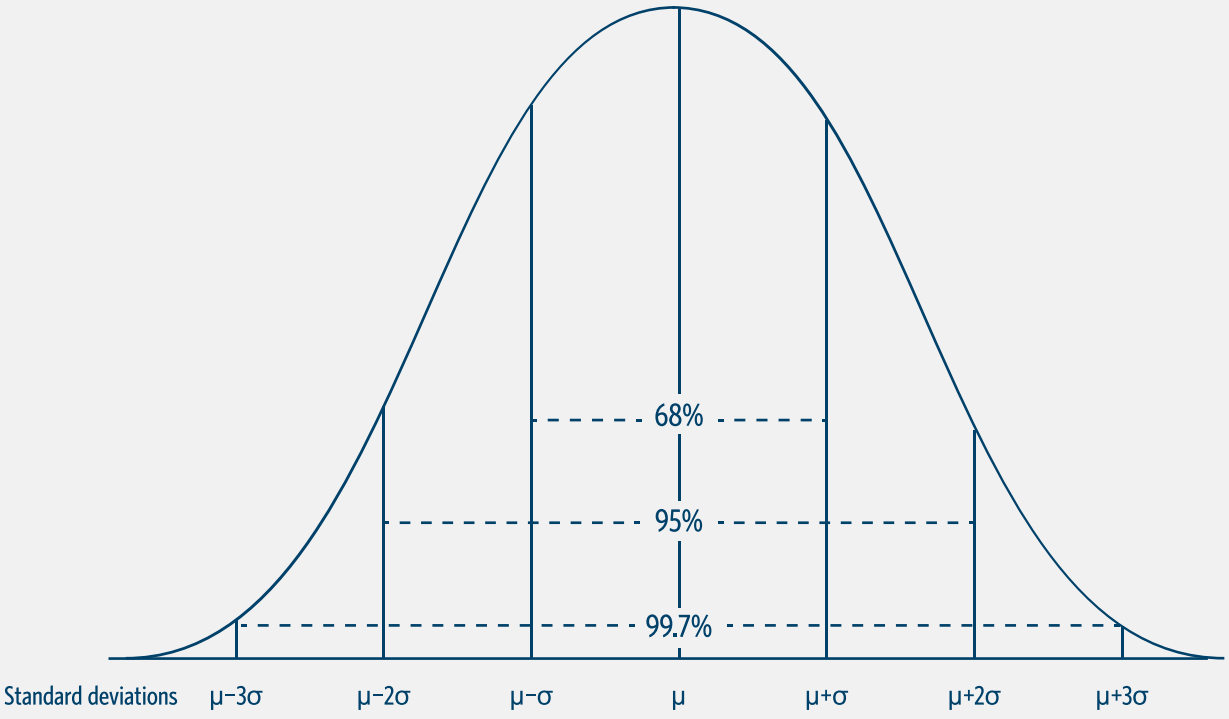
   Student answers should be between 80 and 100 counties.

3. Knowing that **Item 1** represents a normal distribution, what values would you estimate for the median and for the mean? Explain.

   The median appears to be between 16 and 20 percent. The mean should also be between these two values, because the mean and the median are roughly equal in a normal distribution. Because the graph is slightly skewed right, the mean is likely to be slightly greater than the median.

Part 2 – Estimate Values Using the "Empirical Rule"

One reason to use a normal distribution model is to make estimates about data — especially data that are not readily available. The model gives statisticians confidence that the mean is equal to the median and that the empirical rule holds true. This rule, also called the "68-95-99.7 rule," states that for normal distributions:

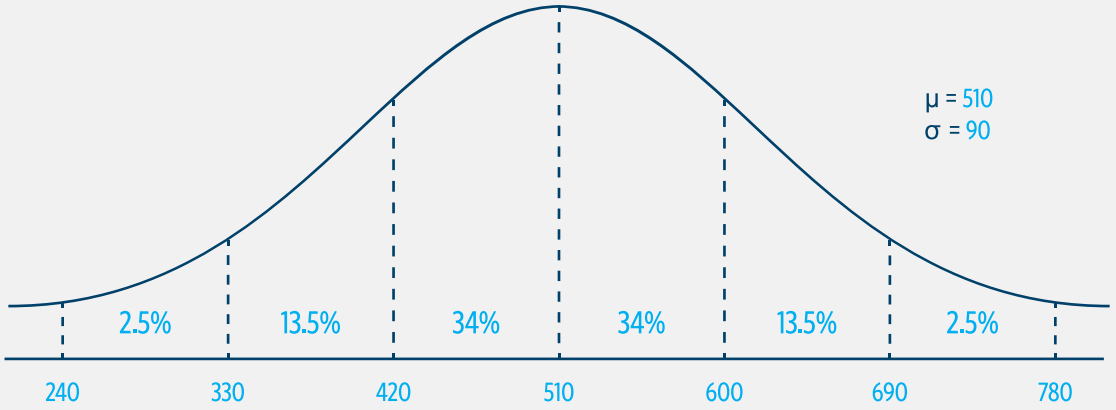


- Approximately 68 percent of the data are within one standard deviation (σ) of the mean (μ).
- Approximately 95 percent of the data are within two standard deviations (σ) of the mean (μ).
- Approximately 99.7 percent of the data are within three standard deviations (σ) of the mean (μ).

In this activity, you will use the normal model and the empirical rule to find estimates for various data.

1. Examine the following hypothetical data distribution. The superintendent of a large county school system believes that the SAT mathematics scores of high school seniors in her county follow an approximately normal distribution. She knows that the mean score in her county is 510 and that the standard deviation (SD) is 90, so she can use the empirical rule to make other estimates.

   a. Fill in the normal curve below with values for μ and σ, and label each interval and the percentage of data each comprises, based on the normal approximation of those values.



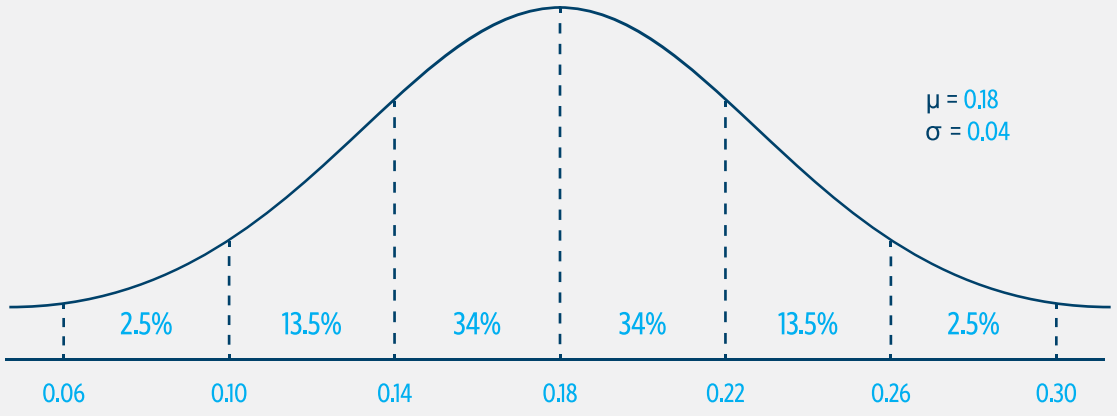
SAT mathematics-section scores of high school seniors

   b. The superintendent wants to know: In this county, what percentage of students would you predict would get an SAT math score between 600 and 690? Explain, and show your work.

13.5 %, because 600 and 690 represent one and two SDs above the mean, respectively, and 95% - 68% = 27%, with half of this percentage (13.5%) falling above the mean.

2. Use **Item 1**, where the mean proportion of people 65 and older in populations across all U.S. counties is 0.18 and the SD is 0.04, to answer the following questions and prompts. Note: Percentages are written as proportions in this case — so, for example, 22 percent would be 0.22.

   a. Fill in the normal curve below with values for μ and σ, and label each interval and the percentage of data each comprises, based on the normal approximation of those values.



   b. Now find the values of μ - 3σ, μ - 2σ, and so on, showing your work below. Write these values in the correct places on the normal curve as well.

μ - 3σ = 0.18 - (3*0.04) = 0.06

μ - 2σ = 0.18 - (2*0.04) = 0.10

μ - σ = 0.18 - 0.04 = 0.14

μ = 0.18

μ + σ = 0.18 + 0.04 = 0.22

μ + 2σ = 0.18 + (2*0.04) = 0.26

μ + 3σ = 0.18 + (3*0.04) = 0.30

   c. If we assume the normal model for these data, in about what percentage of counties would the proportion of people 65 and older be more than 0.22? Show your work and explain your answer in the context of the data.

16 %, because one SD added to the mean is 0.22 (i.e., 0.18 + 0.04 = 0.22) and the percentage of these data beyond one SD is 16 % (i.e., 50% - 34% = 16%). The normal model predicts that about 16 % of U.S. counties have populations in which at least 22 % of their residents are 65 and older.

d. In about what percentage of counties would the proportion of people 65 and older be less than 0.14? Explain, showing your work.

16 percent, because 0.14 corresponds with one SD below the mean, and 50% - 34% = 16%.

e. In about what percentage of counties would the proportion of people 65 and older be less than 0.10? Explain, showing your work.

2.5 percent, because 0.10 corresponds with two SDs below the mean, and 100% - 95% = 5%, with half of this percentage (2.5 percent) falling below the mean.

f. In about what percentage of counties would the proportion of people 65 and older be greater than 0.10? Explain, showing your work.

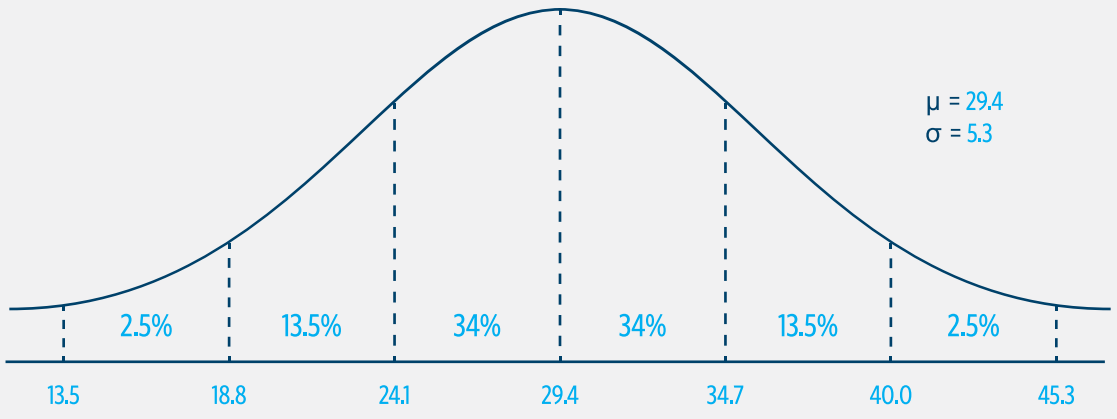
97.5 percent, because 100% - 2.5% = 97.5% (using the answer from question 2e).

g. According to the actual census data, in 12.6 percent of counties the proportion of residents 65 and older was more than 0.22; in 16.5 percent of counties it was less than 0.14; and in 2.2 percent it was less than 0.10. Compared with the predictions you made earlier, do the actual data support, or fail to support, the use of a normal model for this distribution? Explain.

Student answers may vary but could include: They support it. The percentages from the normal distribution are surprisingly close to the actual census values. From looking at just the shape of the distribution in Item 1, it might not be expected that the predictions would be so close, but it appears that the main departure from the normal distribution is in the upper tail (more than one SD above the mean).

3. Item 2: Percentage of Family Households shows the percentages of households with children younger than 18, by county, for a random selection of 136 of the more than 3,000 counties in the United States. The mean of the data is 29.4 percent and the SD is 5.3 percent. Because the distribution is roughly bell-shaped, a normal model can be used to make predictions about all counties in the country.

a. Fill in the normal curve below with values for μ and σ, and label each interval and the percentage of data each comprises, based on the normal approximation of those values.



Percentage of households with children under 18

b. About what percentage of U.S. counties have children in 24.1 to 34.7 percent of their households? Explain how you know.

68 percent, because 24.1 and 34.7 correspond with one SD below and above the mean, respectively, and approximately 68 percent of the data are within this range, according to the empirical rule.

c. About what percentage of U.S. counties have children in 34.7 to 40.0 percent of their households? Explain how you know.

13.5 percent, because 34.7 and 40.0 correspond with one and two SDs above the mean, respectively, and 95% - 68% = 27%, with half of this percentage (13.5 percent) falling above the mean.

d. About what percentage of U.S. counties have children in less than 18.8 percent of their households? Explain how you know.

2.5 percent, because 18.8 corresponds with two SDs below the mean, and 100% - 95% = 5%, with half of this percentage (2.5 percent) falling below the mean.

e. Is your answer to question 3d consistent with the histogram in **Item 2**? Explain.

Yes, there were very few counties with that low of a percentage of family households.

f. According to the census data for just the 136 counties surveyed, 75.7 percent had children in 24.1 to 34.7 percent of their households; 8.9 percent had children in 34.7 to 40.0 percent of them; and 2.9 percent had children in less than 18.8 percent of them. Compared with the predictions you made earlier, do the actual data support, or fail to support, the use of a normal model for this distribution? Explain.

Student answers may vary but could include: The percentages support the use of a normal model for these data. However, the percentages of the counties with between 24.1 and 34.7 percent in the actual distribution differ from the normal distribution by 7.7 percentage points.

g. List two reasons that the percentages you predicted earlier could differ from those actually reported by the Census Bureau.

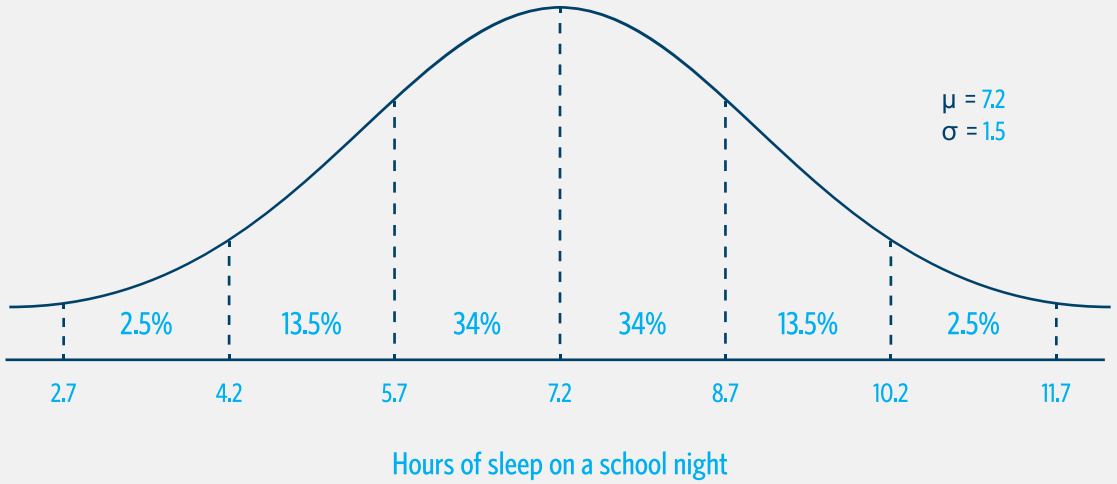
Student answers may vary but could include:

- **The sample of counties may not exactly mirror the distribution for all counties in the United States.**
- **The percentages found here were based on the normal model, which may not fit the total population results exactly.**

Teachers could explain to students that any differences are probably a combination of these two issues.

4. **Item 3: Average Hours of Student Sleep on a School Night** is derived from a random sample of 441 high school students. The histogram is approximately bell-shaped and fairly symmetric. Its mean is 7.2 hours, its SD is 1.5 hours, and it approximately fits a normal model.

   a. Fill in the normal curve below with values for μ and σ, and label each interval and the percentage of data each comprises, based on the normal approximation of those values. Assume that the model applies to all high school students.



   b. About what percentage of all high school students would you predict to sleep between 5.7 and 8.7 hours each night? Explain how you know.

      68 percent, because 5.7 and 8.7 hours correspond with one SD below and above the mean, respectively, and approximately 68 percent of the data are within this range, according to the empirical rule.

   c. About what percentage of all high school students would you predict to sleep more than 8.7 hours each night? Explain how you know.

      16 percent, because 8.7 hours corresponds with one SD above the mean, and 50% - 34% = 16%.

   d. About what percentage of all high school students would you predict to sleep fewer than 8.7 hours each night? Explain how you know.

      84 percent, because 8.7 hours corresponds with one SD above the mean, and 50% + 34% = 84%.

   e. According to **Item 3**, some students reported that they slept very little each night. Based on the normal model, about what percentage of all students would you estimate sleep fewer than 2.7 hours a night? Explain how you know.

      0.15 percent, because 2.7 hours corresponds with three SDs below the mean, and 100% - 99.7% = 0.3%, with half of this percentage (0.15 percent) falling below the mean.

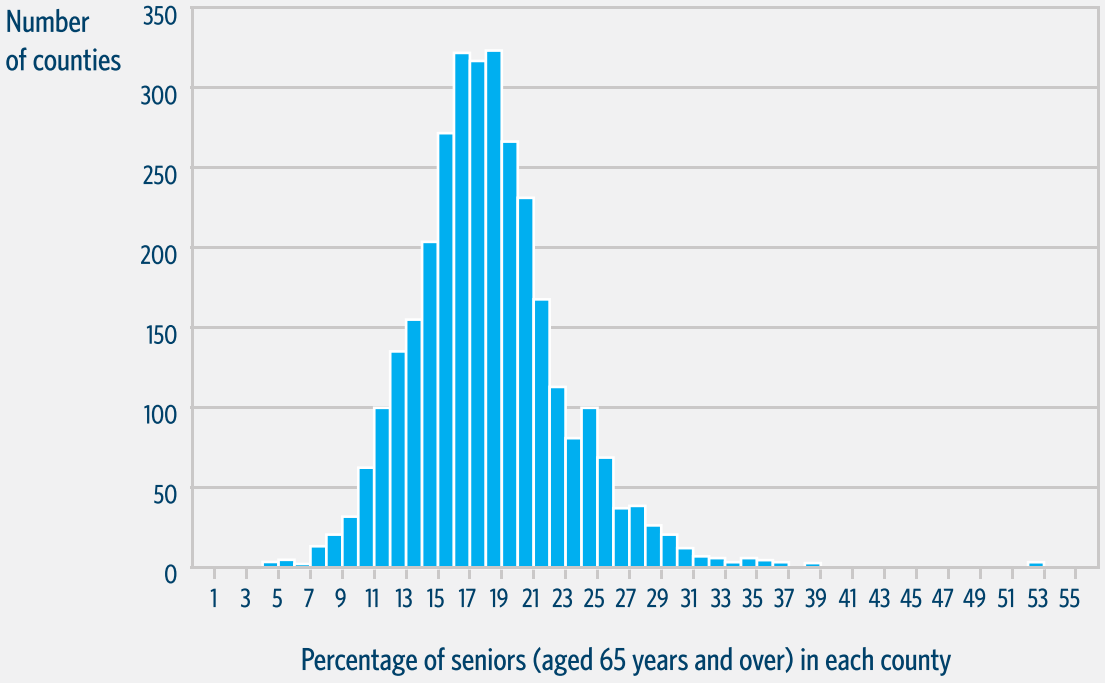
f. According to the actual survey data, 70 percent of the students reported sleeping between 5.7 and 8.7 hours each night; 15.6 percent reported sleeping more than 8.7 hours; and 0.01 percent reported sleeping 3 or fewer hours. Compared with the predictions you made earlier, do the actual data support, or fail to support, the use of a normal model for this distribution? Explain.

Student answers may vary but could include: The percentages support the use of a normal model for these data, with a slight difference in the percentages comparing 3 or fewer hours a night, as the normal model uses 2.7 hours (instead of the actual data's 3 hours) as a data point.

5. Not all data distributions can use the normal model to make estimates. Based on what you know about normality and what you learned in this activity, describe a hypothetical distribution of any variable that cannot be modeled using a normal distribution.

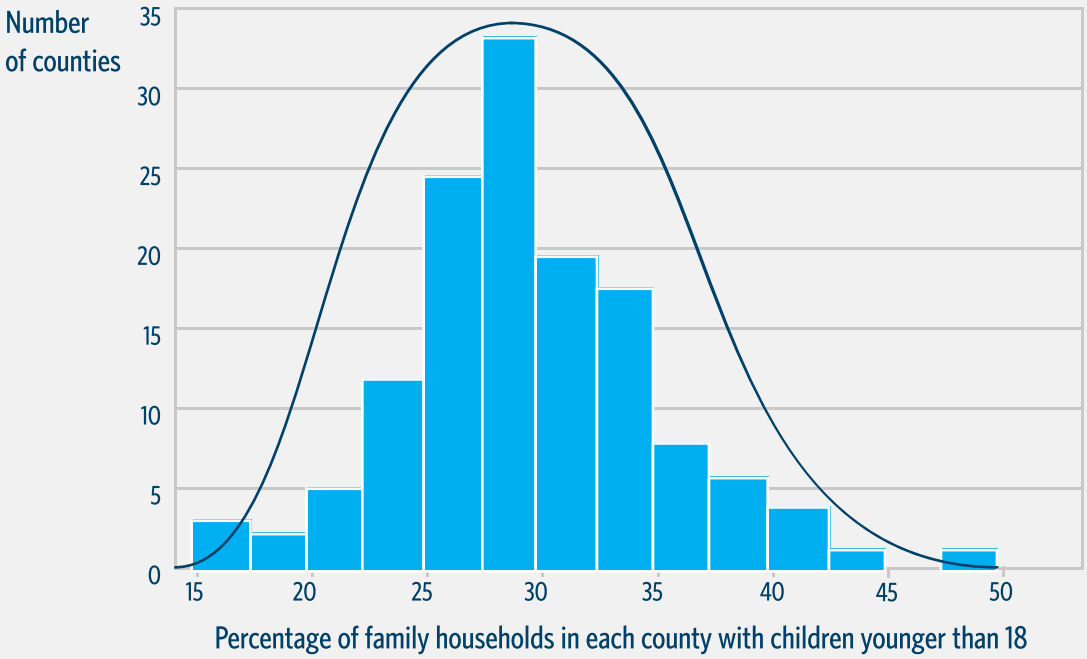
Student answers will vary but could describe a distribution where the mean and the median are very far apart, indicating a skewed distribution that does not fit a normal model. More advanced students could also describe a distribution where the mean is less than two SDs above 0, such as a distribution of weekly hours of television watched whose mean is 14 hours and whose SD is 9 hours. Because the mean in this scenario is less than two SDs above 0, it is not distinct enough from the median, which indicates that the distribution is likely skewed and cannot be accurately modeled by a normal distribution or according to the empirical rule.

Item 1: Percentage of Population Aged 65 and Older



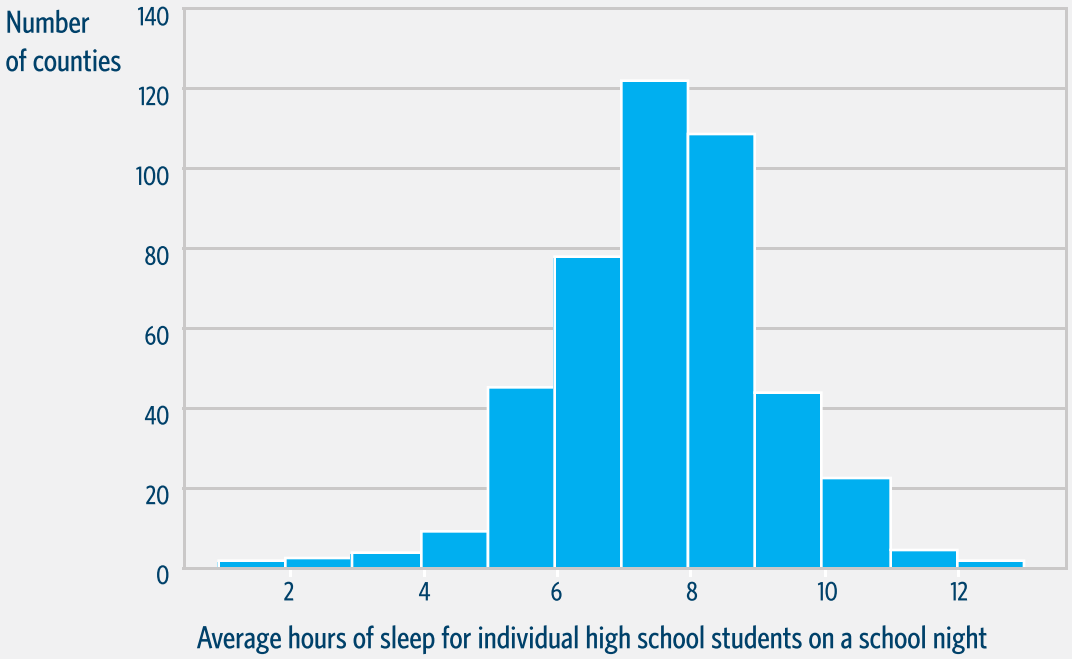
Source for data: U.S. Census Bureau, Population Division, Annual Estimates of the Resident Population for Selected Age Groups by Sex for the United States, States, Counties, and Puerto Rico Commonwealth and Municipios: April 1, 2010 to July 1, 2014.

Item 2: Percentage of Family Households



Source for data: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates, Selected Population Profile in the United States.

Item 3: Average Hours of Student Sleep on a School Night



Note: These data come from a random sample generated by the Census at School Random Sampler data tool:

www.amstat.org/censusatschool/RandomSampleExport.cfm.