

THE NEW NORMAL

TEACHER VERSION

Subject Level:
High School Math

Grade Level:
11–12

Approx. Time Required:
45–60 minutes

Learning Objectives:

- Students will be able to determine if a distribution is normal based on histograms and on normal probability plots.
- Students will be able to identify which data sets can and cannot be modeled accurately using a normal distribution.

Activity Description

Students will explore distributions of various census data sets to determine whether it can be reasonably assumed that those data follow a normal distribution, based on students' analysis of either a histogram or a normal probability plot for each data set. They will then discuss their findings with a partner who analyzed the other type of graph for each data set.

Suggested Grade Level:
11–12

Approximate Time Required:
45–60 minutes

Learning Objectives:

- Students will be able to determine if a distribution is normal based on histograms and on normal probability plots.
- Students will be able to identify which data sets can and cannot be modeled accurately using a normal distribution.

Topics:

- Histograms
- Normal models
- Normal probability plots

Skills Taught:

- Estimating values based on histograms and normal probability plots
- Identifying normality in a data set

Materials Required

- The student version of this activity, 14 pages

A graphing calculator, graphing software (e.g., Microsoft Excel), or other graphing technology is optional.

Activity Items

There are no separate items for this activity.

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 chart 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.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).	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 determine whether distributions are approximately normal based on two types of graphs of the data.
CCSS.MATH.PRACTICE.MP3. Construct viable arguments and critique the reasoning of others. In pairs, students will discuss their decisions about normality based on two different types of graphs of the same data.

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

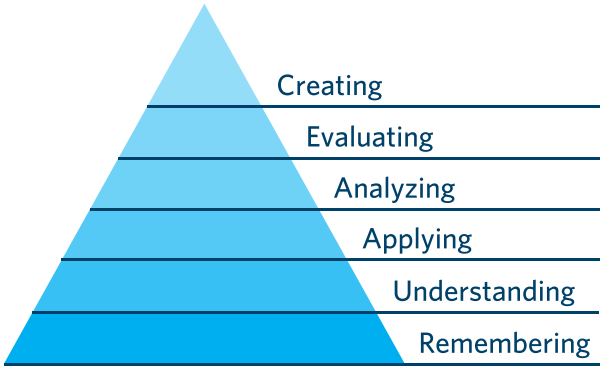
Content Standard	Students should be able to:	Expectation for Grade Band
Data Analysis and Probability	Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.	Understand histograms, parallel box plots, and scatterplots and use them to display data.

Guidelines for Assessment and Instruction in Statistics Education

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

Bloom's Taxonomy

Students will ***analyze*** histograms and normal probability plots of various data distributions and ***evaluate*** whether a normal distribution can accurately be used as a model for those data.



Teacher Notes

Before the Activity

Students must understand the following key terms:

- **Normal probability plot** – a graph that assesses whether a data set has an approximately normal distribution, evidenced by data points that form an approximate straight line
- **Histogram** – a way of displaying numeric data on a graph using horizontal or vertical bars so that the height or the length of the bars indicates frequency
- **Normal distribution** – a data distribution that appears symmetrical and creates a unimodal bell-shaped curve, where the mean and median are of equal value
- **Mean** – a measure of center in a set of numerical data, computed by adding 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)
- **Mode** – the value that appears most frequently in a data set
- **Unimodal** – a distribution that has a single peak or mode
- **Bimodal** – a distribution that has two peaks or modes
- **Bell-shaped curve** – a distribution that is both symmetrical and unimodal, like the outline of a bell
- **Skewness** – a measure of the unevenness of values in a data set in which one "tail" of the distribution has more extreme values than the other "tail," described using terms like "skewed right" and "skewed left"
- **Symmetry** – when the data points on both sides of a distribution appear approximately the same
- **Quantiles** – a value that results from dividing a frequency distribution into equal parts, each the same fraction of the total distribution
- **Gross rent** – the monthly housing expenses for people paying rent, including rent and the cost of utilities

Students should have a basic understanding of how to:

- Analyze a histogram
- Analyze a normal probability plot

Data in this activity come from the 2014 American Community Survey, which is conducted monthly by the Census Bureau and is designed to show how communities are changing. Through asking questions of a sample of the population, it produces national data on more than 35 categories of information, such as education, income, housing, and employment.

Teachers should review with students the two methods used in this activity for determining normality in a distribution: that the histogram of the distribution is unimodal, symmetrical, and bell-shaped, and that the data points of its normal probability plot form an approximate straight line. (Teachers could use TI-83/84 calculators or other graphing software to show students how to create this type of plot.)

Teachers should also review how a normal probability plot can be used to identify skewness: If its smallest and largest data points curve to the left of the normal line, then the original data distribution likely has a long tail to the left (skewed left). If both sets of data points curve to the right, then the original distribution likely has a long tail to the right (skewed right). If the largest and smallest data points curve away from the normal line in opposite directions, then the original distribution may be bimodal or have a different, non-normal shape. Teachers should explain to students that these interpretations of skewness hold true only for plots with normal quantiles on the *y*-axis, as in this activity.

During the Activity

Teachers should have one half of the class complete Version A of the activity and the other half complete Version B. Students should work individually to complete questions 1–6, and then with a classmate who completed the other activity version to discuss their answers and to complete question 7.

Teachers should explain to students that sometimes the number, center, and width of a histogram's bars can affect its shape as either skewed or symmetrical, so normal probability plots can be better a tool for assessing normality.

After the Activity

Teachers should lead a class discussion about what students learned, having them share their responses to question 7.

Extension Ideas

- Teachers could have students use graphing technology to create histograms and normal probability plots to see if one method is easier to use in determining normality.
- If students use TI-83/84 calculators, teachers should have them use *x* values for the Data Axis option, as was done to create the graphs in this activity.
- Teachers could have more advanced students discuss normal quantile values (like *z*-scores) and how the curve of the normal probability plot relates to the normal distribution's shape. For example, teachers could present a distribution where the $z = 2$ and $z = 3$ values are larger than would be expected in a normal distribution.
- Teachers could also have more advanced students predict and examine a normal probability plot for binomial data and then could have them contrast discrete and continuous distributions.

Student Activity

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

Activity Items

There are no separate items for this activity.

Student Learning Objectives

- I will be able to determine if a distribution is normal based on histograms and on normal probability plots.
- I will be able to identify which data sets can and cannot be modeled accurately using a normal distribution.

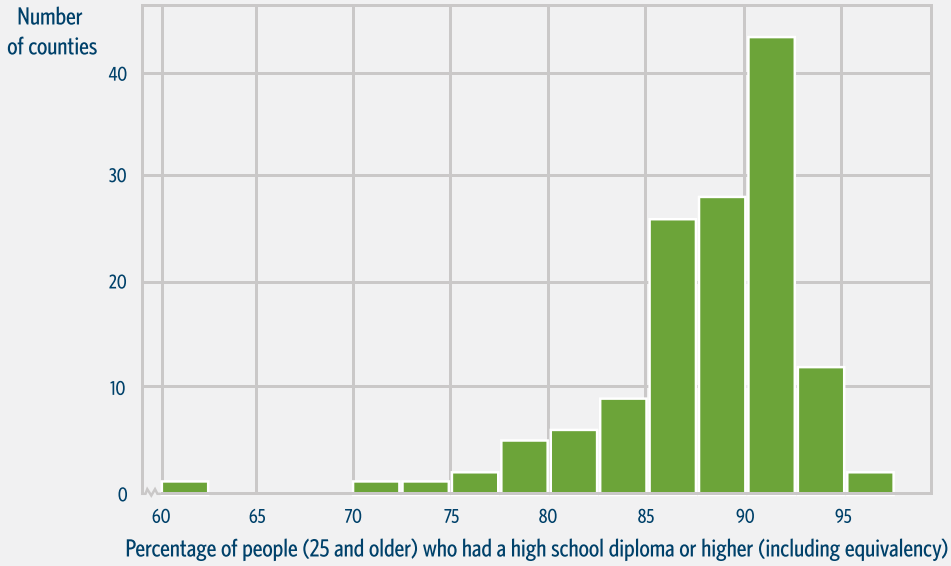
Statistical measures such as median, mean, range, standard deviation (SD), and interquartile range (IQR) provide useful and interesting snapshots of data sets. In this activity, you will analyze county-level data from the U.S. Census Bureau.

Complete either Version A or Version B of this activity according to your teacher's instructions.

Activity Version A

For questions 1–6, you will explore histograms and normal probability plots that use various data from 136 U.S. counties surveyed by the Census Bureau in 2014.

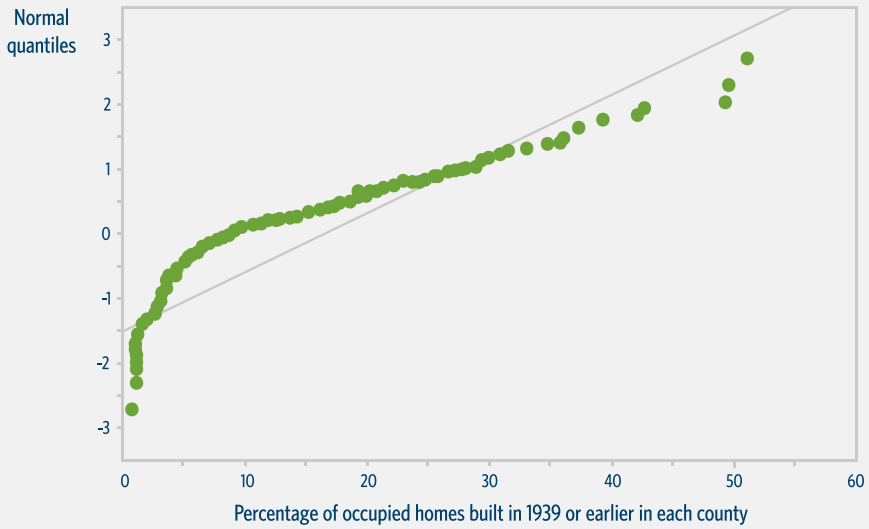
1. Examine the histogram below, which shows the percentage of people (25 and older) in these counties who had the equivalent of a high school diploma or higher, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

a. Based on what you see in this histogram, is the data distribution unimodal? **Yes**
b. Is it approximately symmetrical? **No**
c. Is it approximately bell-shaped? **No**
d. Is it approximately normal? **No**
e. If not normal, is the distribution skewed? If so, in which direction? **Yes, to the left**

2. Examine the normal probability plot below, which shows the percentage of occupied homes built in 1939 or earlier in these counties, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

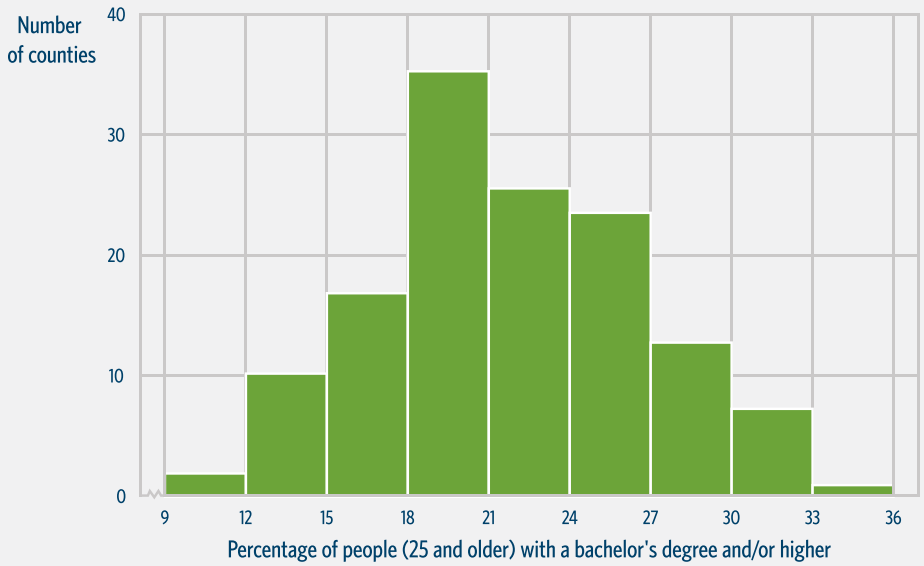
a. Based on what you see in this normal probability plot, is the data distribution approximately normal? Explain how you know.

No, because the pattern of the data points is not linear.

b. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.

Yes, skewed right, because the smallest and largest data points curve to the right of the normal line.

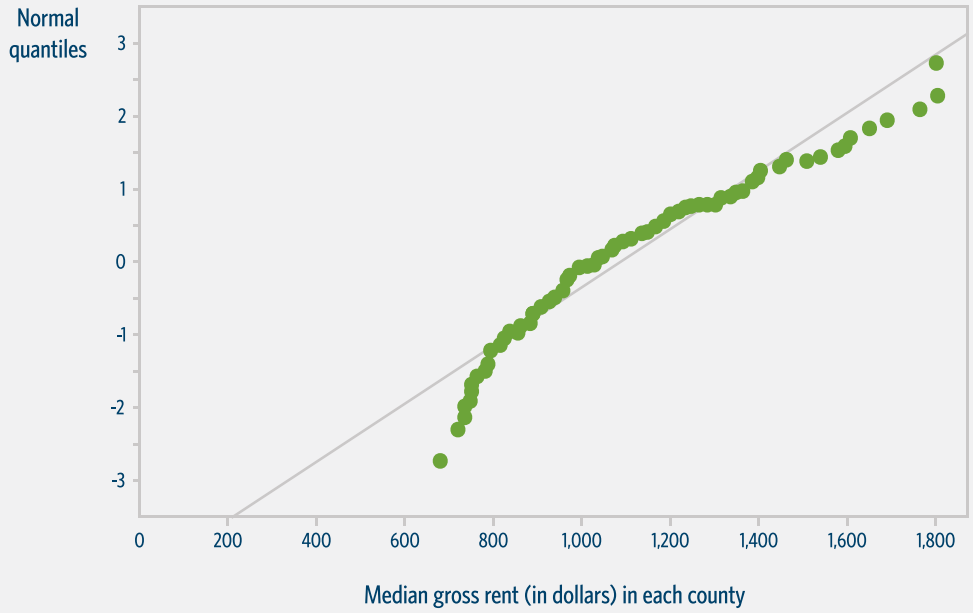
3. Examine the histogram below, which shows the percentage of people (25 and older) in these counties who had bachelor's degrees, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

a. Based on what you see in this histogram, is the data distribution unimodal? **Yes**
b. Is it approximately symmetrical? **Yes**
c. Is it approximately bell-shaped? **Yes**
d. Is it approximately normal? **Yes**
e. If not normal, is the distribution skewed? If so, in which direction? **N/A**

4. Examine the normal probability plot below, which shows the median gross household rent in these counties, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

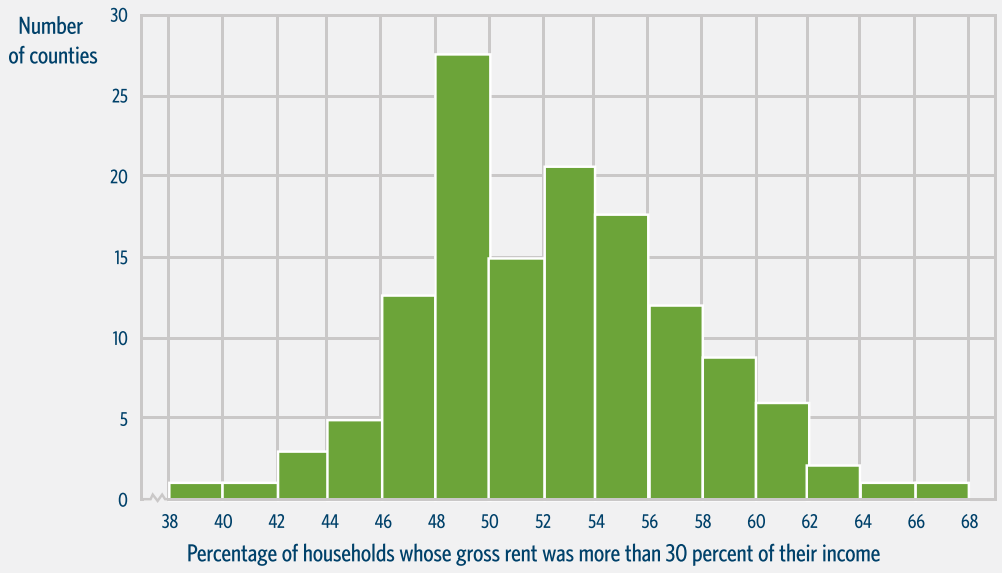
a. Based on what you see in this normal probability plot, is the distribution approximately normal? Explain how you know.

No, because the pattern of the data points is not linear.

b. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.

Yes, skewed right, because the smallest and largest data points curve to the right of the normal line.

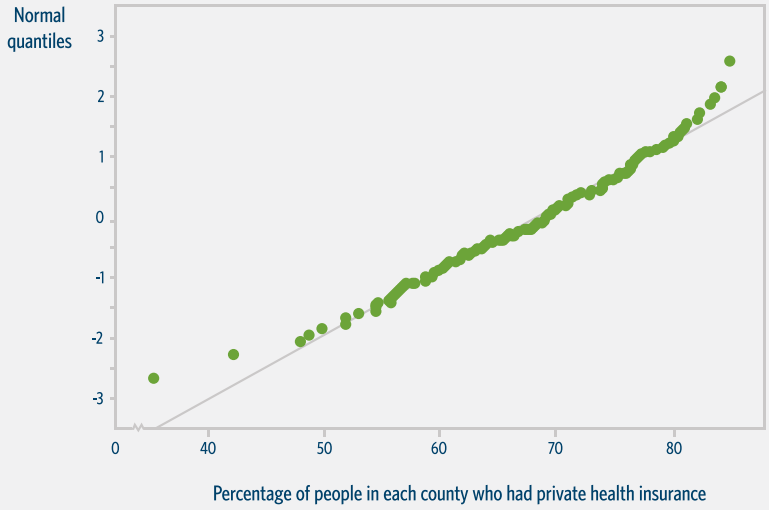
5. Examine the histogram below, which shows the percentage of households in these counties whose gross rent was more than 30 percent of their income, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

a. Based on what you see in this histogram, is the distribution unimodal? **Yes**
b. Is it approximately symmetrical? **Yes**
c. Is it approximately bell-shaped? **Yes**
d. Is it approximately normal? **Yes**
e. If not normal, is the distribution skewed? If so, in which direction? **N/A because the distribution appears normal.**

6. Examine the normal probability plot below, which shows the percentage of people in these counties who had private health insurance, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

a. Based on what you see in this normal probability plot, is the distribution approximately normal? Explain how you know.

No, because while most of the data points fall directly on the line, there are enough that fall on either side of it to indicate skewness.

b. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.

Yes, likely skewed left, because many of the largest data points, and enough of the smallest data points, curve to the left of the normal line.

7. Work with a partner to answer the following questions, using your answers from questions 1–6 and explaining your reasoning.

a. Was it easier to determine whether a distribution was approximately normal based on a data set's histogram or on its normal probability plot?

Student answers will vary.

b. Was it easier to determine whether a distribution was skewed based on a data set's histogram or on its normal probability plot?

Student answers will vary.

c. Was it easier to determine the direction of skewness based on a data set's histogram or on its normal probability plot?

Student answers will vary.

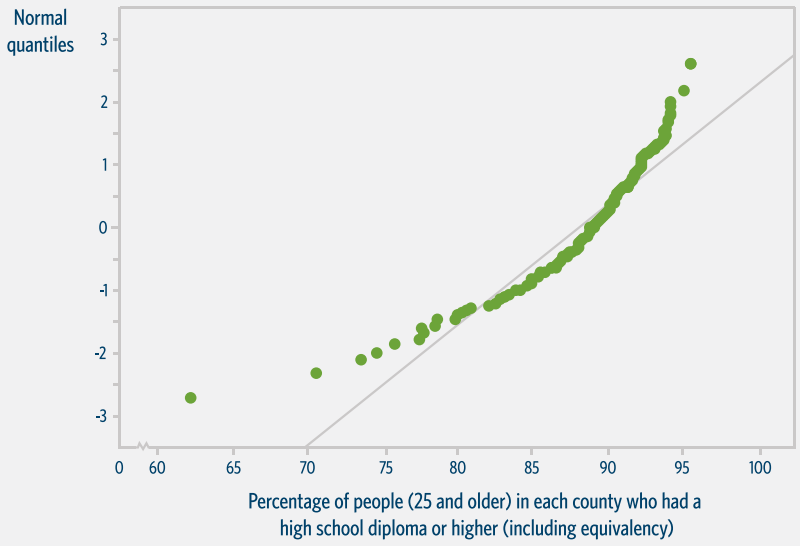
d. Did you and your partner ever make different conclusions about a distribution's shape based on what you saw in the two types of graphs?

Student answers will vary.

Activity Version B

For questions 1–6, you will explore histograms and normal probability plots that use various data from 136 U.S. counties surveyed by the U.S. Census Bureau in 2014.

1. Examine the normal probability plot below, which shows the percentage of people (25 and older) in these counties who had the equivalent of a high school diploma or higher, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

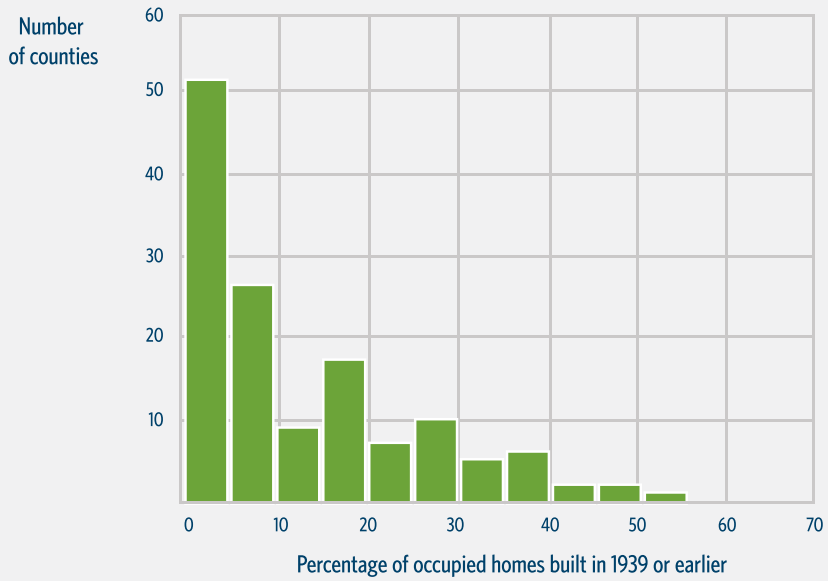
a. Based on what you see in this normal probability plot, is the data distribution approximately normal? Explain how you know.

No, because the data points do not follow a linear pattern.

b. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.

Yes, likely skewed left, because the smallest and largest data points curve to the left of the normal line.

2. Examine the histogram below, which shows the percentage of occupied homes built in 1939 or earlier in these counties, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

a. Based on what you see in this histogram, is the distribution unimodal? **Yes**
b. Is it approximately symmetrical? **No**
c. Is it approximately bell-shaped? **No**
d. Is it approximately normal? **No**
e. If not normal, is the distribution skewed? If so, in which direction? **Yes, skewed right**

3. Examine the normal probability plot below, which shows the percentage of people (25 and older) in each county who had bachelor's degrees, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates.

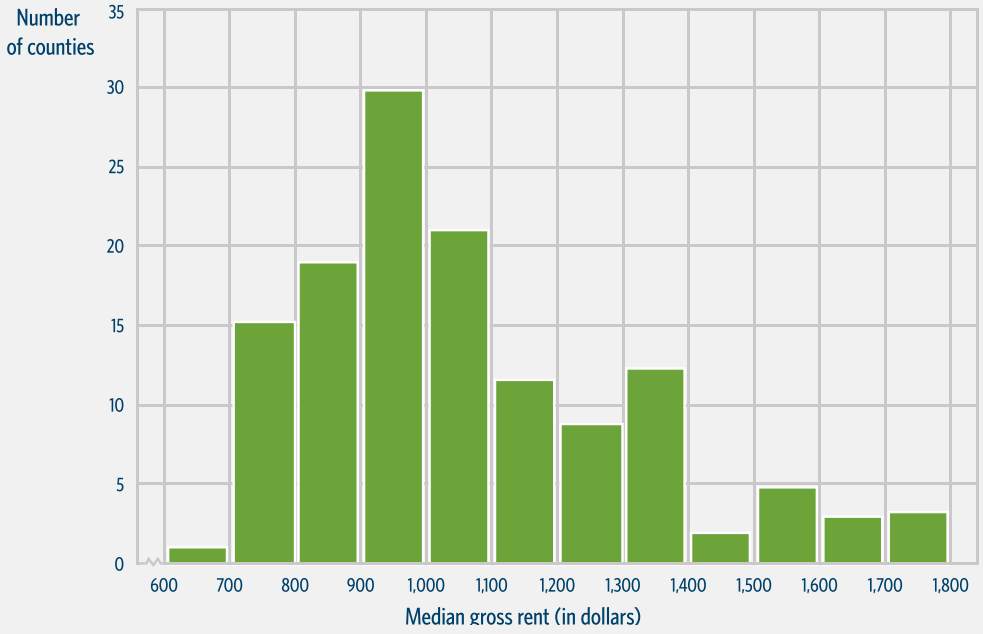
a. Based on what you see in this normal probability plot, is the distribution approximately normal? Explain how you know.

Yes, because almost all of the data points follow a straight line and because not many of the smallest or largest points curve away from the normal line in the same direction, which would indicate skewness.

b. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.

N/A because the distribution appears normal.

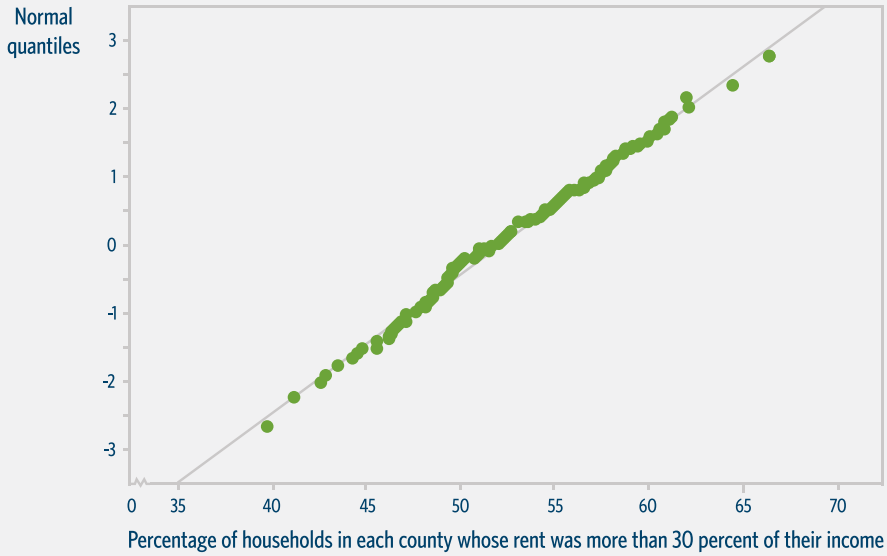
4. Examine the histogram below, which shows the median gross rent in these counties, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

a. Based on what you see in this histogram, is the distribution unimodal? **Yes**
b. Is it approximately symmetrical? **No**
c. Is it approximately bell-shaped? **No**
d. Is it approximately normal? **No**
e. If not normal, is the distribution skewed? If so, in which direction? **Yes, skewed right**

5. Examine the normal probability plot below, which shows the percentage of households in each county whose rent was more than 30 percent of their income, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

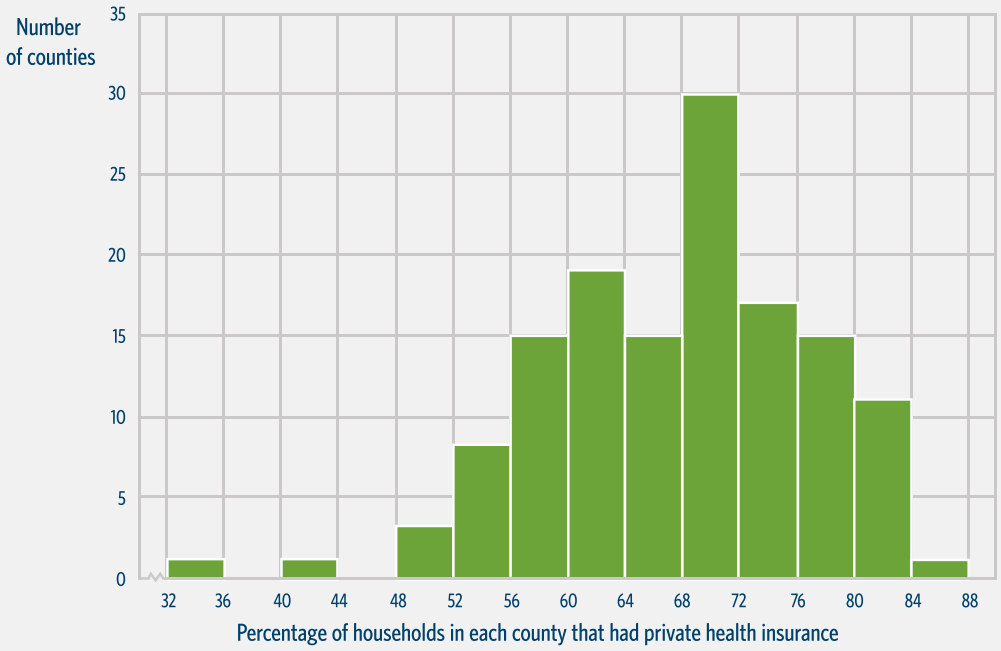
a. Based on what you see in this normal probability plot, is the distribution approximately normal? Explain how you know.

Yes, because the data points follow a very linear pattern.

b. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.

N/A because the distribution appears normal.

6. Examine the histogram below, which shows the percentage of households in these counties that had private health insurance, to answer the questions that follow.



Source: U.S. Census Bureau, 2014 American Community Survey 1-Year Estimates

a. Based on what you see in this histogram, is the distribution unimodal? **Yes**
b. Is it approximately symmetrical? **No**
c. Is it approximately bell-shaped? **No**
d. Is it approximately normal? **No**
e. If not normal, is the distribution skewed? If so, in which direction? **Yes, skewed left**

7. Work with a partner to answer the following questions, using your answers from questions 1–6 and explaining your reasoning.
   a. Was it easier to determine whether a distribution was approximately normal based on a data set's histogram or on its normal probability plot?

      Student answers will vary.

   b. Was it easier to determine whether a distribution was skewed based on a data set's histogram or on its normal probability plot?

      Student answers will vary.

   c. Was it easier to determine the direction of skewness based on a data set's histogram or on its normal probability plot?

      Student answers will vary.

   d. Did you and your partner ever make different conclusions about a distribution's shape based on what you saw in the two types of graphs?

      Student answers will vary.