

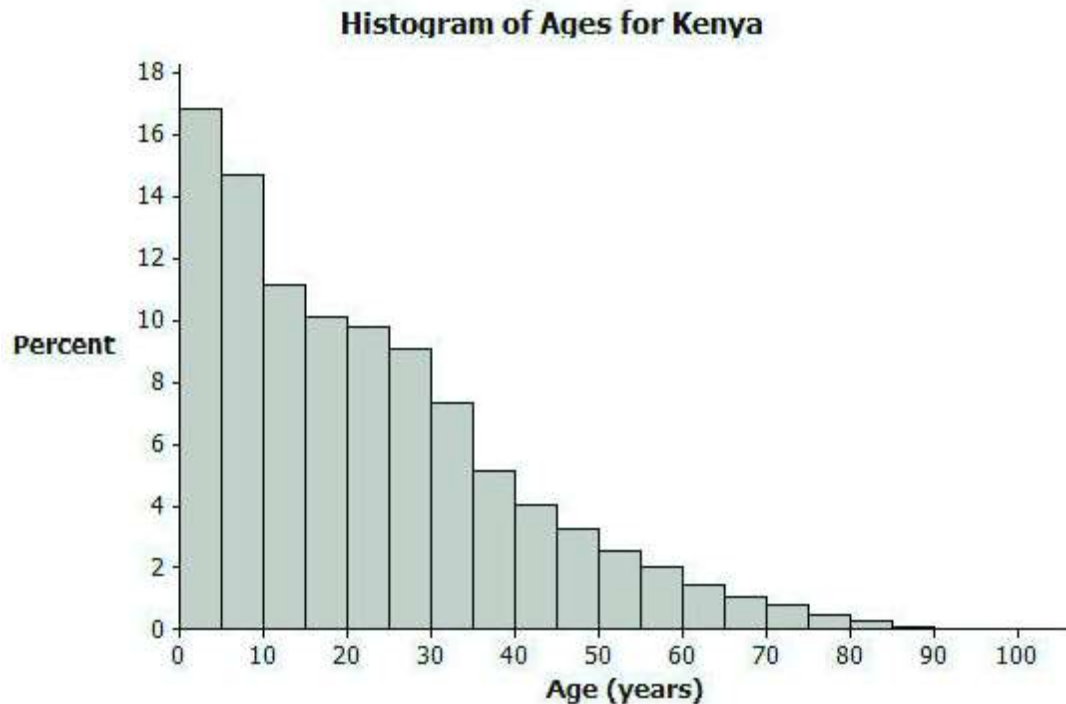
Conditional relative frequency (A conditional relative frequency compares a frequency count to the marginal total that represents the condition of interest.)

Residual (The residual of the data point (x_i, i) is the (actual y_i -value) - (predicted y -value) for the given x_i .)

Residual plot (Given a bivariate data set and linear equation used to model the data set, a residual plot is the graph of all ordered pairs determined as follows: for each data point (x_i, i) in the data set, the first entry of the ordered pair is the x -value of the data point and the second entry is the residual of the data point.)

Correlation coefficient (The correlation coefficient, often denoted by r , is a number between -1 and $+1$ inclusively that measures the strength and direction of a linear relationship between the two types of quantities. If $r=1$ (likewise, $r=-1$), then the graph of data points of the bivariate data set lie on a line of positive slope (negative slope).)

- **Histograms:** A graph of data that groups the data based on intervals and represents the data in each interval by a bar.



Analysis:

1. What do you think this graph is telling us about the population of Kenya?

A large percentage of the people in Kenya are ages 10 or younger.

2. Why might we want to study the data represented by this graph?

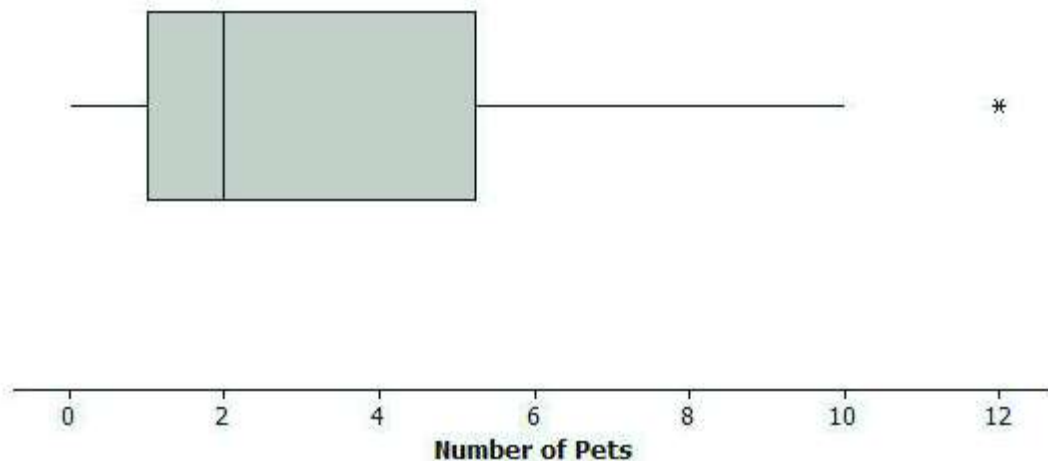
It tells us about Kenya and its challenges based on its population and demographics. Caring for and educating young people are major challenges for this country.

3. Based on your previous work with histograms, would you describe this histogram as representing a symmetrical or a skewed distribution? Explain your answer.

Skewed; it has a tail to the right.

- **Box plots:** A graph that provides a picture of the data ordered and divided into four intervals that each contains approximately 25% of the data.

Boxplot of Number of Pets



Analysis:

1. What does the box plot tell us about the number of pets owned by the thirty students at River City High School?

50% of students own between 1 and 5 pets.

2. Why might understanding the data behind this graph be important?

Understanding the data is important for planning special events involving pets and understanding interests of a group of people.

Two-Way Frequency Table:

	To Fly	Freeze Time	Invisibility	Super Strength	Telepathy	Total
Females	49	60	48	1	70	228
Males	51	71	27	25	48	222
Total	100	131	75	26	118	450

The relative frequency table would be found by dividing each of the above cell values by 450. For example, the relative frequency of females selecting “To Fly” is $49/450$, or approximately 0.109 to the nearest thousandth. A few of the other relative frequencies to the nearest thousandth are shown in the following relative frequency table:

	To Fly	Freeze Time	Invisibility	Super Strength	Telepathy	Total
Females	$49/450$ ≈ 0.109					$228/450$ ≈ 0.507
Males			$27/450$ ≈ 0.06			
Total		$131/450$ ≈ 0.291			$118/450$ ≈ 0.262	

Calculate the remaining relative frequencies in the table above. Write the value in the table as a decimal rounded to the nearest thousandth or as a percent.

Two-Way Frequency Table:

	To Fly	Freeze Time	Invisibility	Super Strength	Telepathy	Total
Females	<i>0.109</i> <i>10.9%</i>	<i>0.133</i> <i>13.3%</i>	<i>0.107</i> <i>10.7%</i>	<i>0.002</i> <i>0.2%</i>	<i>0.156</i> <i>15.6%</i>	<i>0.507</i> <i>50.7%</i>
Males	<i>0.113</i> <i>11.3%</i>	<i>0.158</i> <i>15.8%</i>	<i>0.060</i> <i>6.0%</i>	<i>0.056</i> <i>5.6%</i>	<i>0.107</i> <i>10.7%</i>	<i>0.493</i> <i>49.3%</i>
Total	<i>0.222</i> <i>22.2%</i>	<i>0.291</i> <i>29.1%</i>	<i>0.167</i> <i>16.7%</i>	<i>0.058</i> <i>5.8%</i>	<i>0.262</i> <i>26.2%</i>	<i>1.00</i> <i>100%</i>

1. Based on previous work with frequency tables, which cells in this table would represent the joint relative frequencies? *The joint relative frequency cells are represented by the unshaded cells within the body of the table.*

2. Which cells in the relative frequency table would represent the marginal relative frequencies? *The marginal relative frequency cells are represented by the shaded cells in the total row and total column.*

3. What is the joint relative frequency for females who selected “invisibility” as their favorite superpower? *The joint relative frequency for females who selected “invisibility” as their favorite superpower is 0.107 or approximately 10.7%. This indicates approximately 11% of the students sampled were female who selected “invisibility” as their favorite superpower.*

4. What is the marginal relative frequency for “freeze time?” Interpret the meaning of this value? *29.1% or approximately 29% of the total number of people surveyed selected “freeze time” as their favorite superpower.*

5. What is the difference in the joint relative frequencies for males and for females who selected “to fly” as their favorite superpower? *The difference in the relative frequencies is 0.004 or 0.4%, or $0.113 - 0.109$.*

6. Is there a noticeable difference between the genders and their favorite superpowers? *Yes. The most noticeable differences are in the following superpowers: invisibility, super strength, telepathy.*

Graphing Calculator Help....

Finding the Regression Line (TI-84 Plus)

Step 1: From your home screen, press STAT.

Step 2: From the STAT menu, select the EDIT option. (EDIT enter)

Step 3: Enter the x-values of the data set in L1.

Step 4: Enter the y-values of the data set in L2.

Step 5: Select STAT. Move cursor to the menu item CALC and then move the cursor to option 4: LinReg($ax+b$) or option 8: LinReg($a+bx$). Press enter. (Discuss with students that both options 4 and 8 are representations of a linear equation. It is anticipated that most students are familiar with option 4, or the slope y -intercept form. Option 8 is essentially the same representation using different letters to represent slope and y -intercept. Option 8 is the preferred option in statistical studies.)

Step 6: With option 4 or option 8 on the screen, enter L1, L2, and Y1 as described in the following notes.

LinReg($a+bx$) L1, L2, Y1

Select enter to see results. The least-squares regression will be stored in Y1. Work with students in graphing the scatter plot and Y1.

Note: L1 represents the x -values of the regression function, L2 the y -values, and Y1 represents the least squares regression function.

To obtain Y1, go to VARS, move cursor to Y-VARS, and then Functions (enter). You are now at the screen highlighting the y -variables. Move cursor to Y1 and hit enter.

Y1 is the linear regression line, and will be stored in Y1.

Construction of scatter plot:

1. From the home screen press 2nd, STAT PLOT.
2. Select Plot1 and press ENTER.
3. Select "On", under "Type" choose the first (scatter plot) icon, for Xlist enter L₁, for Ylist enter L₂, and under "Mark" chose the first (square) symbol.
4. Press 2nd, QUIT to return to the home screen.
5. Press Y=.
6. Go to any unwanted graph equations and press CLEAR. Make sure that only Plot1 is selected (not Plot2 or Plot3).
7. Press Zoom, select ZoomStat (option 9), press ENTER.
8. The scatter plot and the least-squares line are displayed.

Construction of residual plot:

1. From the home screen, press 2nd, STATPLOT.
2. Select Plot2 and press ENTER.
3. Select "On", under "Type" choose the first (scatter plot) icon, for Xlist enter L₁, for Ylist enter RESID, and under "Mark" choose the first (square) symbol. ("RESID" is accessed by pressing 2nd, LIST, selecting NAMES, scrolling down to RESID, and pressing ENTER.)
4. Press 2nd, QUIT to return to the home screen.
5. Press Y=.
6. First, deselect the equation of the least-squares line in Y₁ by going to the "=" sign for Y₁ and pressing ENTER. Then deselect Plot1 and make sure that Plot2 is selected.
7. Press Zoom, select ZoomStat (option 9), press ENTER.
8. The residual plot is displayed.

Steps for calculating the correlation coefficient using a TI-84 Plus

Step 1: Determine which variable represents x , and which variable represents y based on x and y variable designations.

Step 2: From the home screen, select STAT.

Step 3: Click ENTER from the Edit option of the menu.

Step 4: Enter the values of x in L₁ and the values of y in L₂.

Step 5: When complete, enter 2nd QUIT.

Step 6: Select STAT.

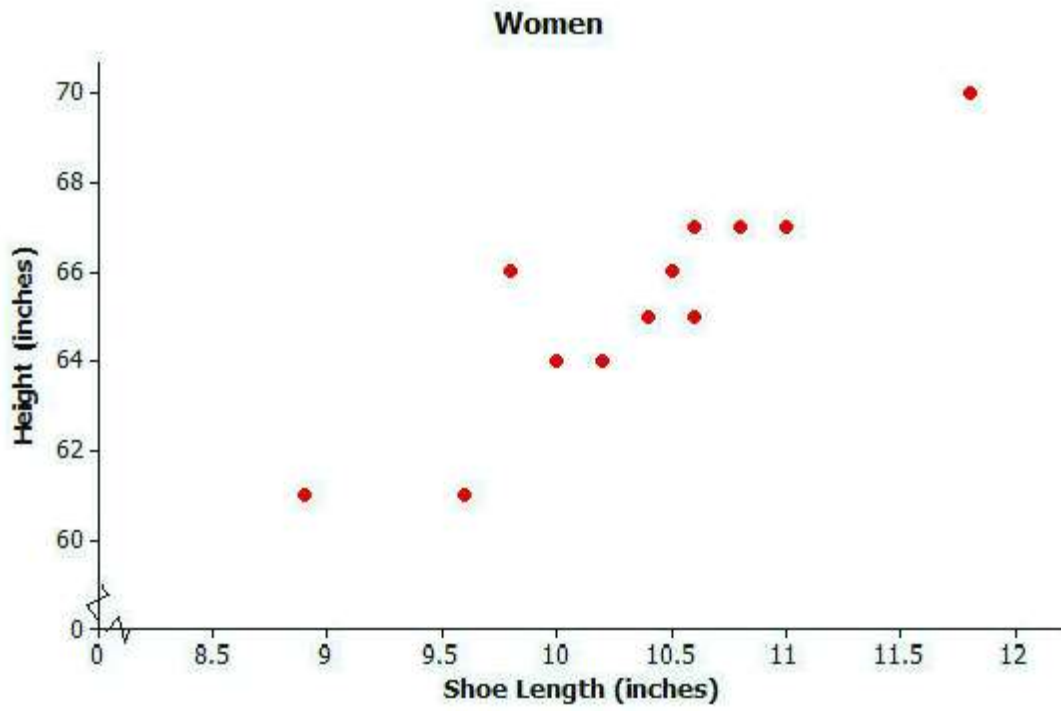
Step 7: With the arrows, move the top cursor over to the option CALC and move the down cursor to 8: LinReg($a+bx$), and then click ENTER.

Step 8: With LinReg($a+bx$) on the screen, enter L₁, L₂ and then ENTER.

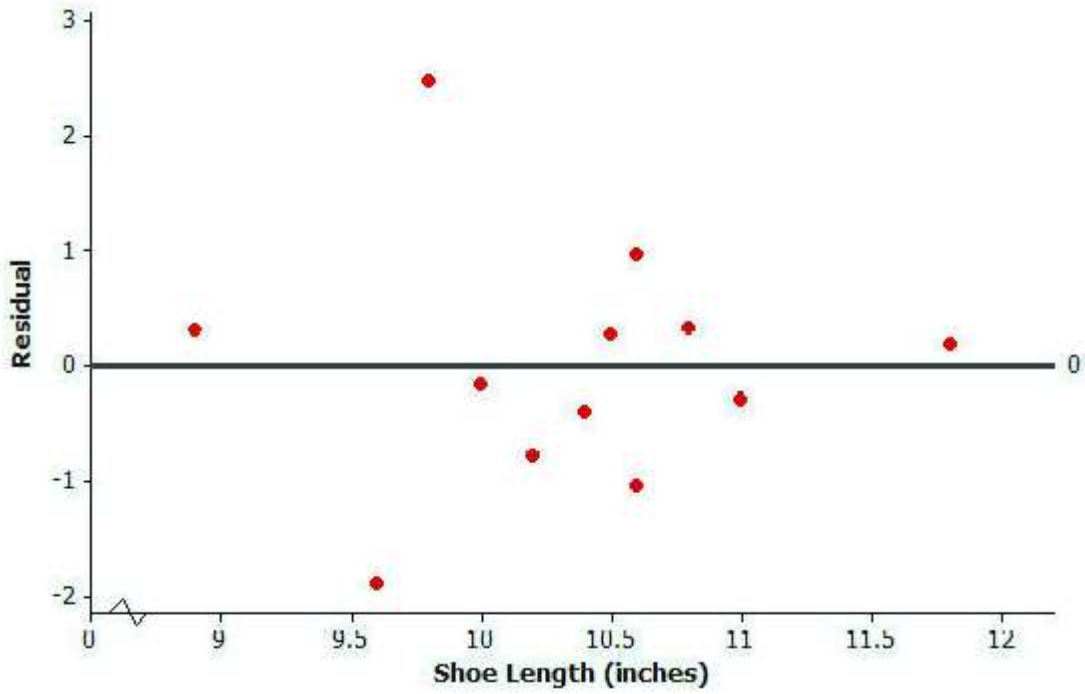
Step 9: The value of r , the correlation coefficient, should appear on the screen.

Scatter Plot:

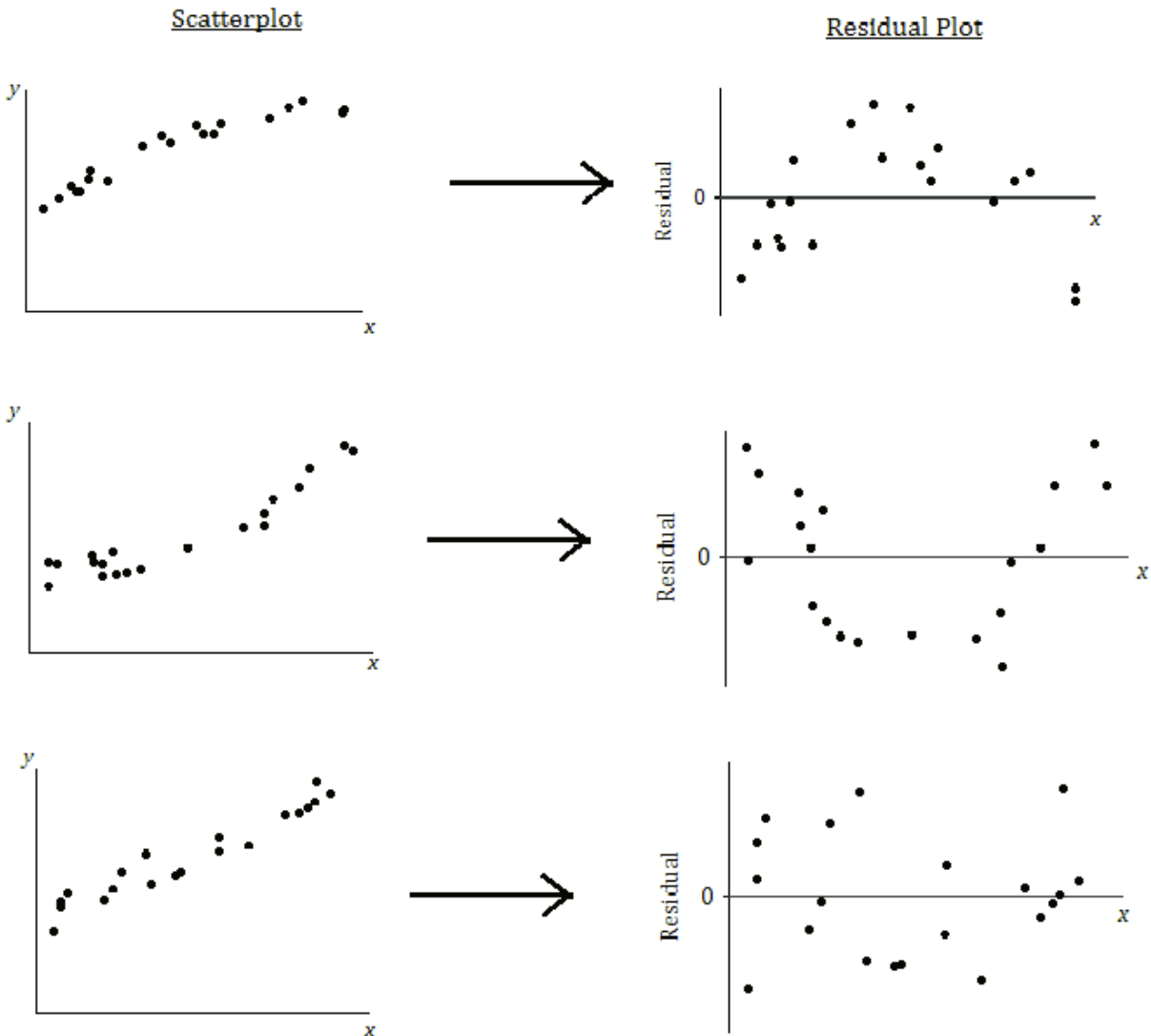
x



Residual Plot:



Analysis:



What does it mean when there is a curved pattern in the residual plot?

Allow students to answer this question. Discuss with students that a curved pattern in the residual plot indicates that the relationship would be better described by a nonlinear function. When you see a curved pattern in the residual plot, you should investigate nonlinear models rather than using a line to describe the relationship. (See the first two scatter plots and residual plots in the figure above, which show a curved pattern in both the scatter plot and the residual plot.)

What does it mean when the points in the residual plot appear to be scattered at random with no visible pattern?

Again, allow students to answer this question. Discuss with students that when the relationship is approximately linear, the points in the residual plot will be scattered at random around the horizontal line at zero. This indicates that a linear model is an appropriate way to describe the relationship. (See the third scatter plot and residual plot in the figure above, which shows a linear pattern in the scatter plot and no pattern in the residual plot.)

