https://www.desmos.com/calculator/lywhybetzt

Goal:

How do we know if a linear model is adequate?

After all, isn't it possible another type of model might be even better?

Warm Up --- Section 3.2_Day 2

The Least Squares Regression Line

And Using Residual Plots to Determine if a Linear Model is appropriate

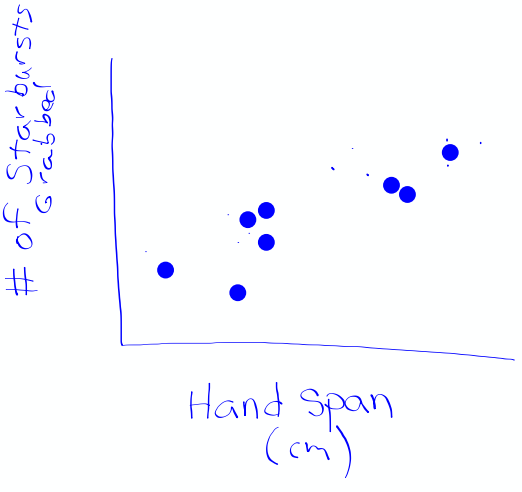
1. Let's return to the Candy Grab data from an earlier class.

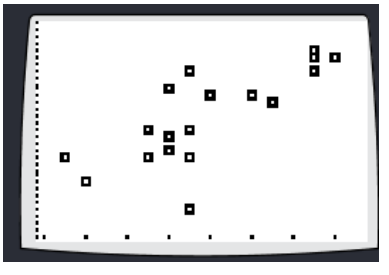
   Enter x = hand span in L_3 and y = number starbursts in L_4.

2. Use your graphing calculator to create a scatter plot. Sketch it below.

Hand Span (cm)	Number of Candies
19.5	15
19	33
17	19
21.5	31
19.3	36
18.5	27
20	32
21	32
19.5	23
23	38
18.5	23
19	26
19	24
22½	39
19	24
19.5	27
19	26
16.5	23
22.5	36
22.5	38

2. Use your graphing calculator to create a scatter plot. Sketch it below.





2. Calculate the equation of the least-squares regression line. Make sure to define variables by writing your equation in context as well. Add the line to your sketch above (you can graph in Y_1)

$\hat{y} = -28.54 + 2.88x$

$\widehat{\#\text{pieces}} = -28.54 + 2.88(\text{handspan})$

Interpret the slope and *y* intercept in context.

Slope For each additional cm of handspan, the number of starbursts grabbed rises by 2.88 starbursts

y-intercept - Nothing meaningful

4. Calculate and interpret the residual for the 2nd student in the list.

(19, 33)
↑
x-value

$\widehat{\#\text{Starbursts}} = -28.5 + 2.88(\text{hanspan})$

$= -28.5 + 2.88(19)$

$= 26.2$ pieces

4. Calculate and interpret the residual for the 2nd student in the list.

(19, 33)
↑
x-value

$\widehat{\#Starbursts} = -28.5 + 2.88(\text{hanspan})$

$= -28.5 + 2.88(19)$

$= 26.2$ pieces

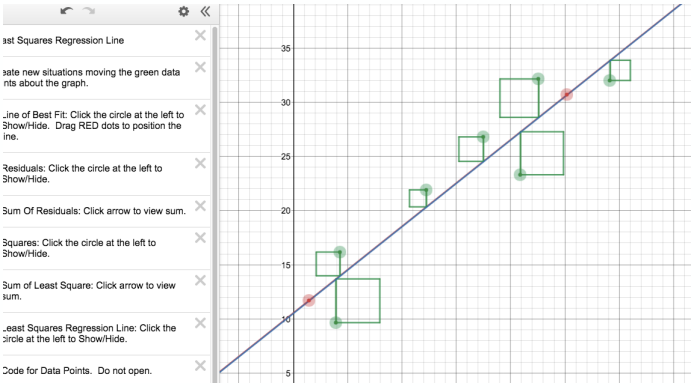
Residual = Actual − Predicted

$= 33 - 26.2 = 6.8$

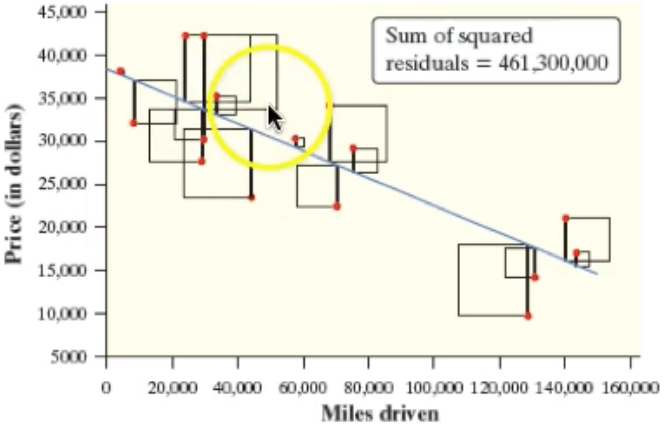
The #pieces of candy grabbed for a 19 cm handspan is 6.8 cm more than predicted by the LSRL.

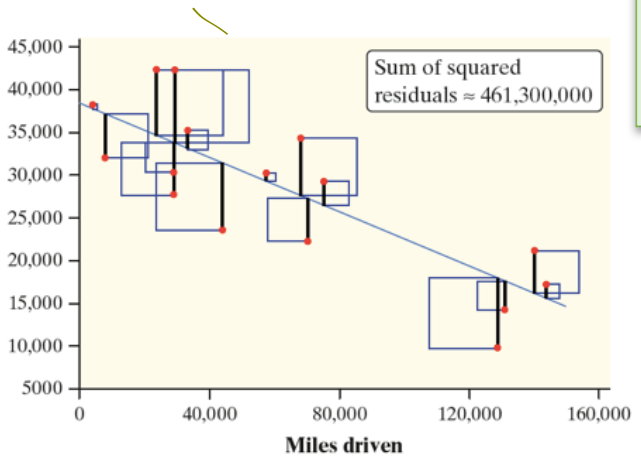
5. The least-squares regression line is the line that makes .....

The Least-Squares Regression Line
(pages 183-184)



- The LSRL minimizes the sum of squared residuals





The **least-squares regression line** is the line that makes the sum of the squared residuals as small as possible.

6. Is the LSRL resistant to outliers?

Residual Plots

(pages 185-188)

What is a residual plot?

A tool to address FORM

To determine if the model we are using has the right form !

One of the first principles of data analysis is to look for an overall pattern and for striking departures from the pattern.

A regression line describes the overall pattern of a linear relationship between an explanatory variable and a response variable.

We see departures from this pattern by looking at a residual plot.

A **residual plot** is a scatterplot that displays the residuals on the vertical axis and the explanatory variable on the horizontal axis.

A **residual plot** is a scatterplot that displays the residuals on the vertical axis and the explanatory variable on the horizontal axis.

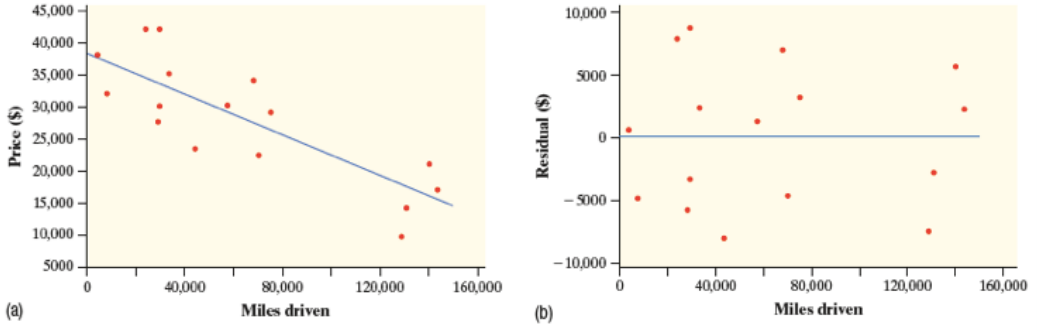


FIGURE 3.11 The (a) scatterplot and (b) residual plot for the relationship between price and miles driven for Ford F-150s.

Determining if a Linear Model Is Appropriate: Residual Plots

A residual plot magnifies the deviations of the points from the line, making it easier to see unusual observations and patterns. If a regression model is appropriate:

- The residual plot should show no obvious patterns.
- The residuals should be relatively small in size.

How to Interpret a Residual Plot

To determine whether the regression model is appropriate, look at the residual plot.

- If there is no leftover curved pattern in the residual plot, the regression model is appropriate.
- If there is a leftover curved pattern in the residual plot, consider using a regression model with a different form.

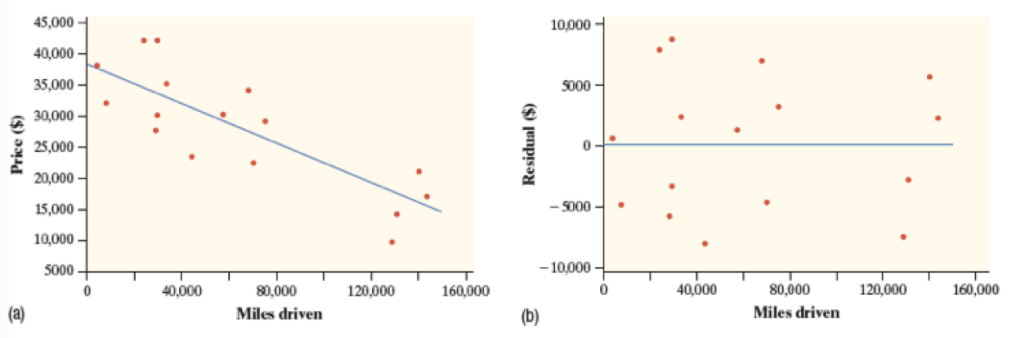
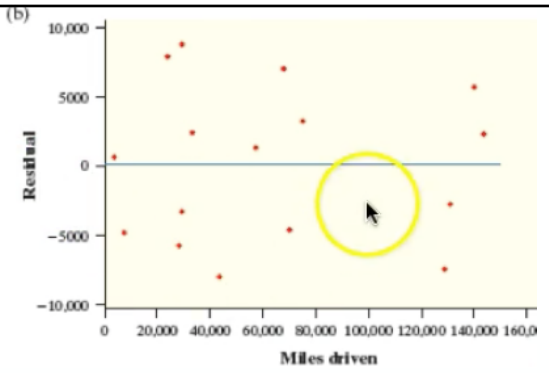
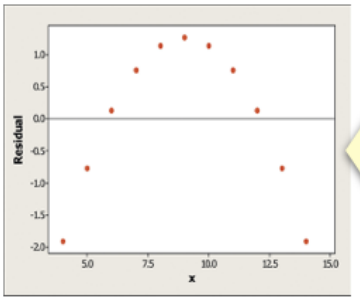


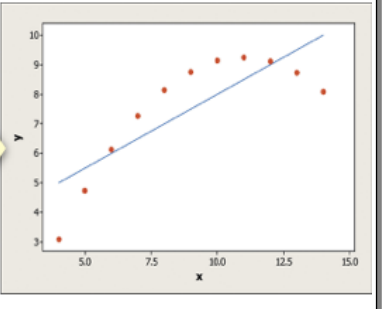
FIGURE 3.11 The (a) scatterplot and (b) residual plot for the relationship between price and miles driven for Ford F-150s.

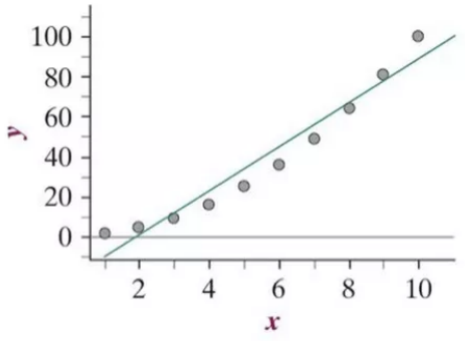


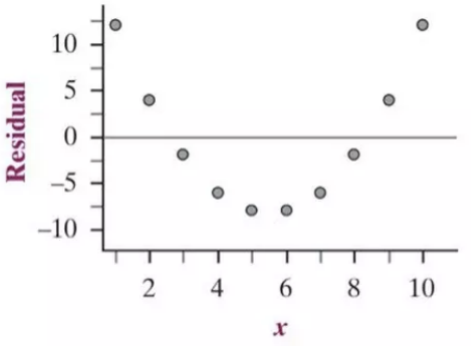
- No leftover pattern → model we are using matches the form of the association and is appropriate.
- Leftover pattern → the model we are using doesn't match the form of the association.



Pattern in residuals
Linear model not appropriate



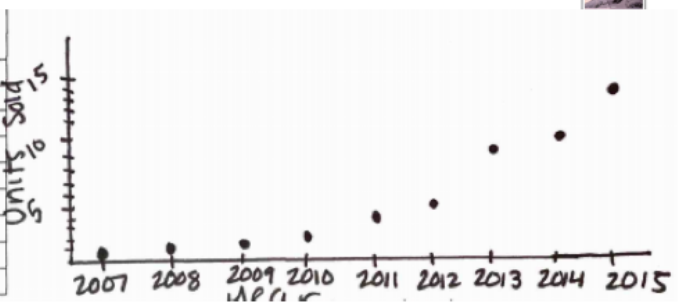




New
Classwork

Lesson 3.2: Day 2: How many iPhones will be sold?

iPhone	Year	Units Sold (millions)
Original	2007	0.5
3G	2008	1
3Gs	2009	1
4	2010	1.7
4S	2011	4
5	2012	5
5C, 5S	2013	9
6, 6 Plus	2014	10
6S, 6S Plus	2015	13

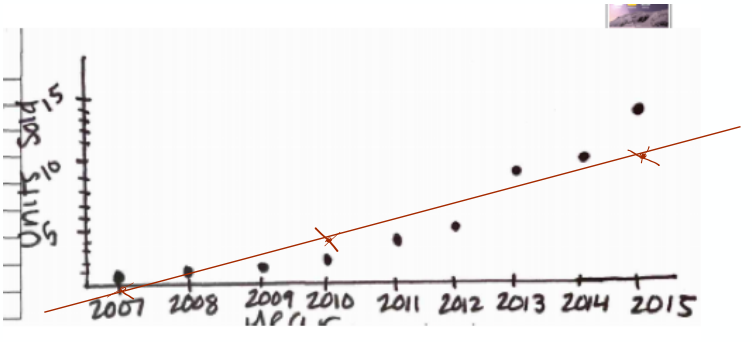


1. Above is the data of all iPhone sales during their opening weekend. From the scatter plot, describe the **form** of the distribution.

The distribution appears non-linear

2. Using your textbook applet gives the least squares regression line of:

$\hat{y} = -3222.633 + 1.605x$ Graph this line on to your scatterplot above as best you can. Also write the LSRL in context.



3. Calculate the residual for 2007. Interpret the residual.

Units Sold $= -3222.633 + 1.665(2007)$

$= -1.40$ (predicted)

Residual $= A - P$

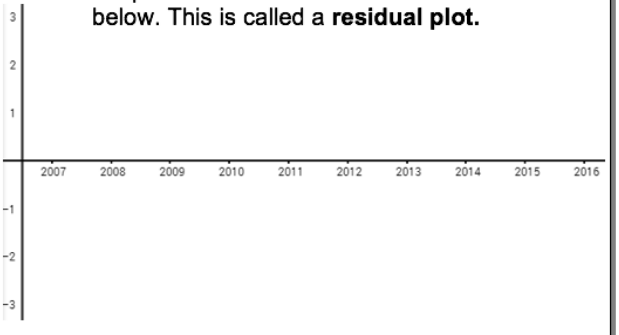
$= 0.5 - (-1.40) = 1.9$ million

The actual number of units sold in 2007 was/~~were~~ 1.9 million greater than expected

4. Complete the table below.

Year	Actual Units Sold (millions)	Predicted Units Sold (millions)	Residual A−P
2007	0.5		
2008	1		
2009	1		
2010	1.7		
2011	4		
2012	5		
2013	9		
2014	10		
2015	13		

5. Graph the residuals on the axes below. This is called a **residual plot.**



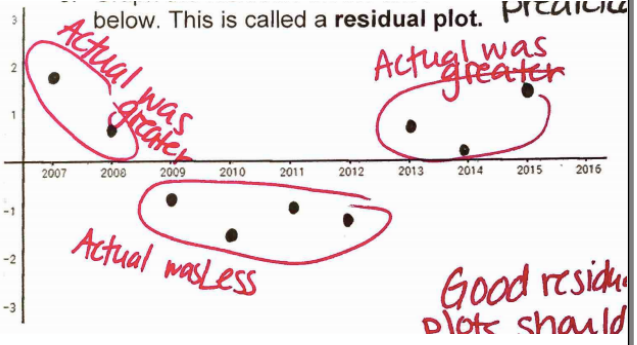
6. For which points was the actual greater than the predicted? Which were less than predicted? Identify these on the graph.

7. Do you think the regression line is a good fit for the data? Why or why not? Explain using the residual plot.

5. Complete the table below.

Year	Actual Units Sold (millions)	Predicted Units Sold (millions)	Residual
2007	0.5	-1.40	1.90
2008	1	0.21	0.79
2009	1	1.81	-.81
2010	1.7	3.42	-1.72
2011	4	5.02	-1.02
2012	5	6.63	-1.63
2013	9	8.23	0.77
2014	10	9.84	0.16
2015	13	11.44	1.56

below. This is called a **residual plot.** Predicte



Good residu
plot should

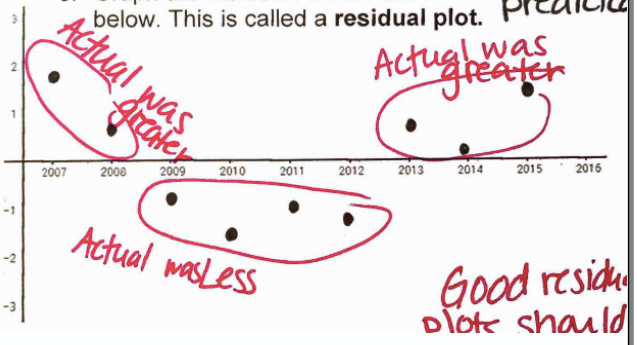
6. For which points was the actual greater than the predicted? Which were less than predicted? Identify these on the graph.

7. Do you think the regression line is a good fit for the data? Why or why not? Explain using the residual plot.

5. Complete the table below.

Year	Actual Units Sold (millions)	Predicted Units Sold (millions)	Residual
2007	0.5	-1.40	1.90
2008	1	0.21	0.79
2009	1	1.81	-.81
2010	1.7	3.42	-1.72
2011	4	5.02	-1.02
2012	5	6.63	-1.63
2013	9	8.23	0.77
2014	10	9.84	0.16
2015	13	11.44	1.56

below. This is called a **residual plot.** Predicte



Good residu
plot should

6. For which points was the actual greater than the predicted? Which were less than predicted? Identify these on the graph.

7. Do you think the regression line is a good fit for the data? Why or why not? Explain using the residual plot.

NO. The residual plot does not show a random scatter. It looks curved so the form is non-linear.

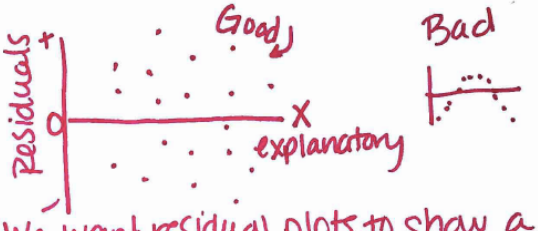
Lesson 3.2 – LSRL and Residual Plots

Big Ideas:

Big Ideas:

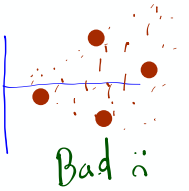
LT #1: Least squares regression line (LSRL)
- The line with the smallest sum of (residuals)^2.

LT #2 Residual Plot



We want residual plots to show a random scatter with no leftover pattern.

Check Your Understanding:



Bad :(

Check Your Understanding:

Ninth-grade students at the Webb Schools go on a backpacking trip each fall. Students are divided into hiking groups of size 8 by selecting names from a hat. Before leaving, students and their backpacks are weighed. The data here are from one hiking group.

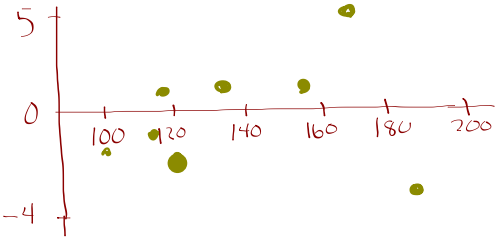
Body weight (lb)	120	187	109	103	131	165	158	116
Backpack weight (lb)	26	30	26	24	29	35	31	28

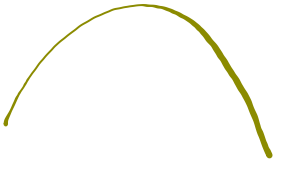
1. Use your textbook 2-variable quantitative Applet to calculate the equation of the least-squares regression line.

$\hat{y} = 16.2649 + 0.0908x$

- $\widehat{\text{backpack weight}} = 16.2649 + 0.0908(\text{Body Weight})$

2. Make a residual plot for the linear model in Question 1 using this same applet. Sketch and label it below.



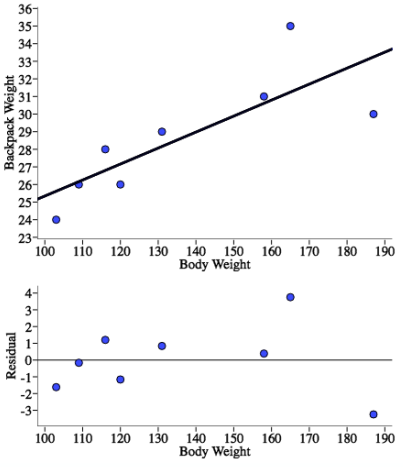


3. What does the residual plot indicate about the appropriateness of the linear model? Explain your answer.

Because there is a negative/positive/negative pattern, a linear model is not appropriate.

8. Construct and interpret a Residual plot for the Candy Grab.

What does it tell you about the appropriateness of using a linear model for this dat?



← Not a random scatter
- + -
Linear model not appropriate

Assignment

3.2.....47, 49, 51, 53

Read/Study pp.183-187