Section 3.2 Notes

Getting a Line on the Pattern

Two reasons to fit a line to a set of data:

1. To find a ______________________, or model, that describes the relationship between the two variables.
2. To use the line to predict _________________________ when you know ________________________
   - ________________________________: variables on the x-axis.
   - ________________________________: variable on the y-axis.

Two types of predictions:

1. ___________________________: making a prediction when the value of x falls outside the range of the actual data.
2. ___________________________: making a prediction when the x falls inside the range of the actual data.

______________________: the error in the predicted values and actual values.

Residual = observed vale of y - predicted value of y.

$$y - \hat{y}$$

_________________________________: the line for which the sum of the squared errors (residuals), or SSE, is as small as possible.

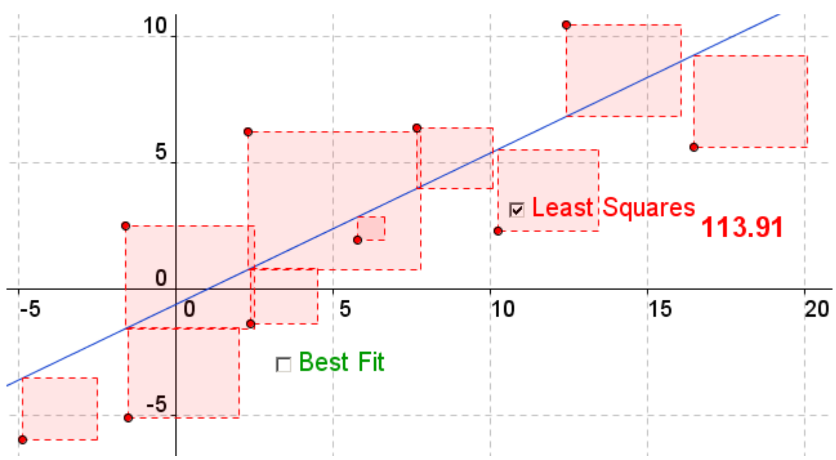
$$SSE = \sum (residuals)^2$$

$$SSE = \sum (y - \hat{y})^2$$

Properties of Least Squares Regression Line

1. The sum (and mean) of the residuals is _________.
2. The line contains the point of the _________________________.
3. The standard deviation of the residuals is _________________ than for any other line that goes through the point ____________.
4. The line has a slope, b_1, where

$$b_1 = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$



Section 3.3 Notes

Correlation

____________________________: measures the strength of a linear relationship numerically.

Formula for the Correlation:

$$r = \frac{1}{n-1} \sum z_x \cdot z_y$$

McDonald's Revisited:

Sandwich	Fat, x	Calories, y	z_x	z_y	$z_x \cdot z_y$
Hamburger	9	250			
Double Cheeseburger	23	440			
Quarter Pounder with Cheese	26	510			
Double Quarter Pounder with Cheese	42	740			
Big Mac	29	540			
Angus Deluxe	39	750			
Filet-O-Fish	18	380			
McChicken	16	360			
McRib	26	500			
Big N' Tasty	24	460			
Sum					
Mean					
Standard Deviation					

Correlation: ________________

Relationship between slope and correlation:

$$b_1 = r \cdot \frac{s_y}{s_x}$$

BEWARE: Correlation does not ALWAYS imply causation.

A ______________________________ is a variable that you didn't include in your analysis but that might explain the relationship between the variables you did include. That is, when variables x and y are correlated, it might be because both are consequences of a third variable, z, that is lurking in the background.

____________________________________ (r^2): the proportion of the total variation in the y's that is explained by the relationship with x.

$$r^2 = \frac{SST - SSE}{SST}$$

Finding the Correlation of determination, r^2:

Sandwich	Fat, x	Calories, y	Predicted $\bar{y}$	$(y-\bar{y})^2$	Predicted $\hat{y}$	$(y-\hat{y})^2$
Hamburger	9	250				
Double Cheeseburger	23	440				
Quarter Pounder with Cheese	26	510				
Double Quarter Pounder with Cheese	42	740				
Big Mac	29	540				
Angus Deluxe	39	750				
Filet-O-Fish	18	380				
McChicken	16	360				
McRib	26	500				
Big N' Tasty	24	460				

SST: ________________ SSE: ________________

__________________________: the tendency of y-values to be closer to their mean.

Section 3.4 Notes
Diagnostics

Judging a Point's Influence

Points are separated from the bulk of the data by white spaces are ________________ and are ____________ ____________________. To judge a point's influence, compare the regression equation and correlation computed first with and then without the point in question.

Best way to do this: Fit the line ___________________________ the questionable point and see what happens. Then report all results, with appropriate explanation.

McDonald's Data Revisted (AGAIN!)

Sandwich	Fat, x	Calories, y
Hamburger	9	250
Double Cheeseburger	23	440
Quarter Pounder with Cheese	26	510
Double Quarter Pounder with Cheese	42	740
Big Mac	29	540
Angus Deluxe	39	750
Filet-O-Fish	18	380
McChicken	16	360
McRib	26	500
Big N' Tasty	24	460
Grilled Chicken Sandwich	10	420

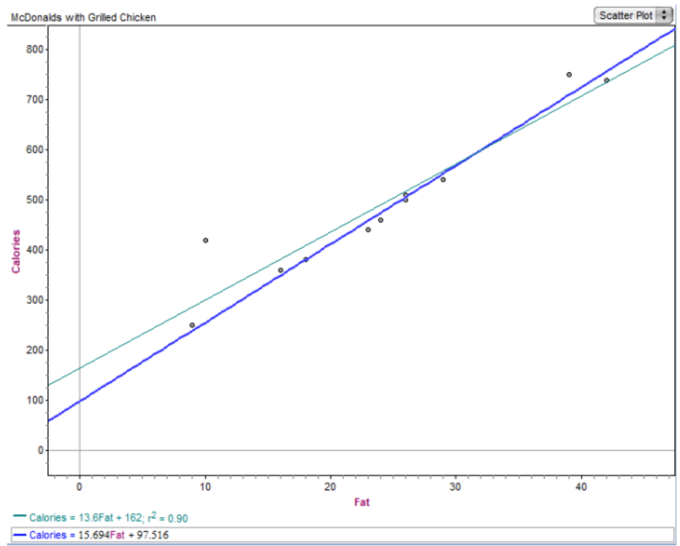
Without Grilled Chicken:

$$\hat{y} = 15.694x + 97.516$$
$$r = 0.993$$

With Grilled Chicken:

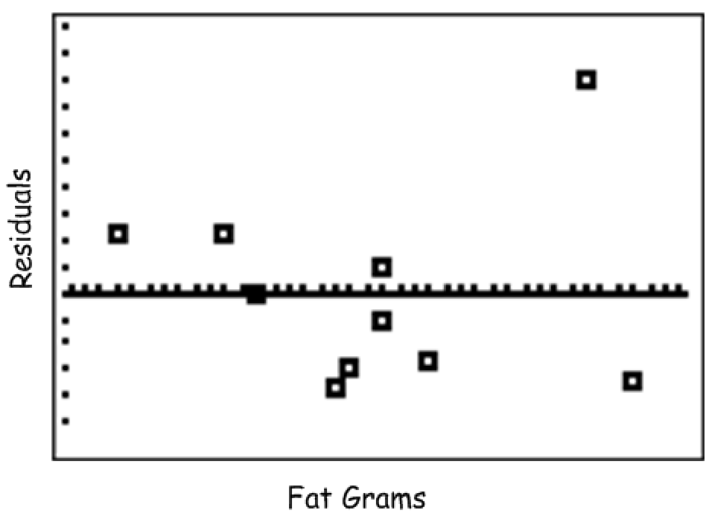
$\hat{y} =$ ____________________

$r =$ __________



A ______________________ is a scatterplot of residuals $y - \hat{y}$, versus predictor values x.

Residual Plot of McDonald's Data



Note: Only when the clouds of points is ________________ does the least squares line, together with the correlation, give a good summary of the relationship described by the plot.

The residual plot will look like: ____

Section 3.5 Notes

Shape-Changing Transformations

Basic rules of logarithms:

Logarithms	Powers
$\log_b(mn) = \log_b m + \log_b n$	$a^{m+n} = a^m \cdot a^n$
$\log_b m^n = n \log_b m$	$(a^m)^n = a^{mn}$
$\log_b b^m = m$	$a^{\log_a m} = m$

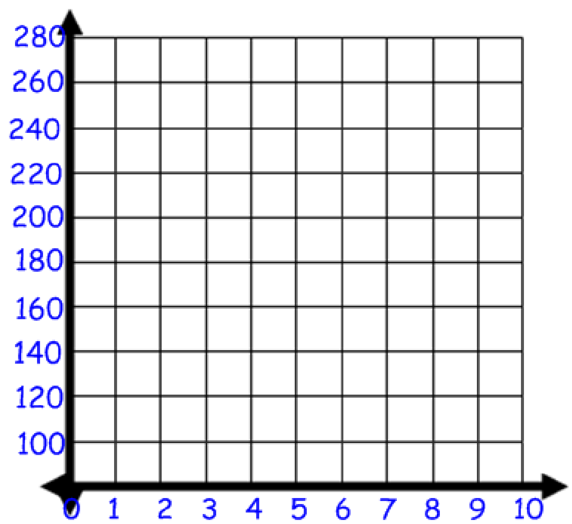
Exponential Growth and Decay: ___________________

The points can be "linearized" by taking the log (base 10 or base e) of each value of y: __________________________

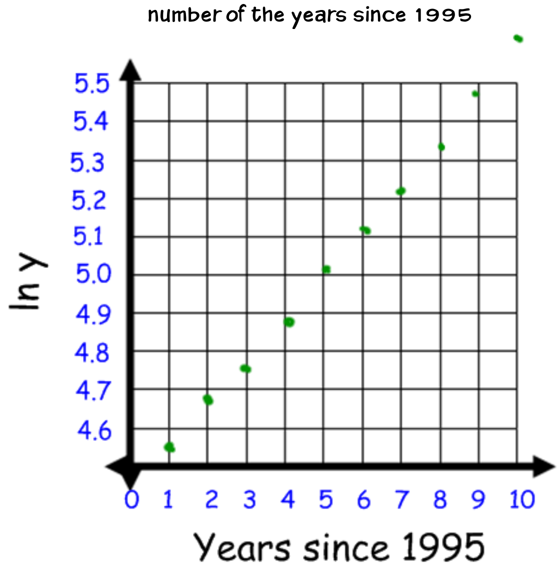
Example: # of rabbits in Central Park (Let x = years since 1995)

Year	1 1996	2 1997	3 1998	4 1999	5 2000	6 2001	7 2002	8 2003	9 2004	10 2005
# of rabbits (round to nearest rabbit)	95	107	119	134	150	168	188	210	236	264

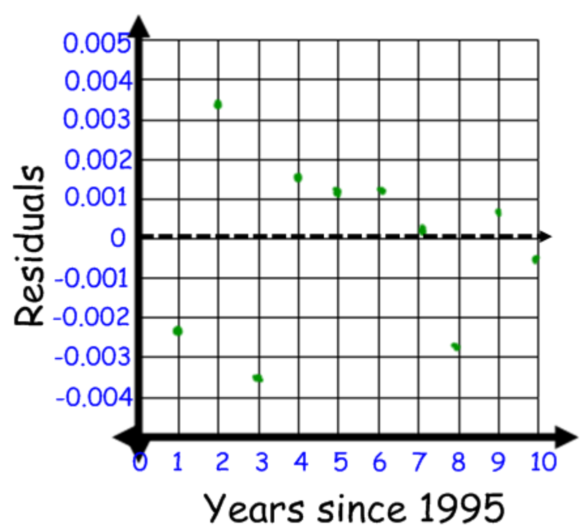
Graph the function:



A plot of ln(rabbits) versus the number of the years since 1995



Graph the residuals for the ln(rabbits) regression



Linear regression of ln y vs. x: ___________________________________

Write a function to represent the equation: ________________________________

Log Log Transformation of Powers

____________________________: have an equation of the form _________________ as the underlying model.

The point can be "linearized" by taking the log of both the values of x and the values of y. _______________________

Power Transformations:
Sometimes your knowledge of the situation allow you to go directly to an appropriate power transformation without using logs.

Common Power Transformations: ____

Example:
Volume of Ball vs. Radius

Radius (m)	Volume
1	4
1.2	7
1.5	15
1.9	29
2	33
2	35
2.4	56
2.9	102
3	113
3.3	151
3.4	163
3.7	210
3.7	215
4	260
4	268
4	275
4.3	300
4.4	350
4.4	400
4.5	400