AP STAT Handout: Influential Points and Lurking Variables Name _____
Section 3.3: Correlation and Regression Wisdom

EQ:

Recall:

- Outliers--- points that are well _____ from the _____ that the other points seem to _____.

Outliers in Univariate Data Set

- value in a set of data that _____ with the rest of the _____
- more than _____ from the _____
- lies outside the _____ wall

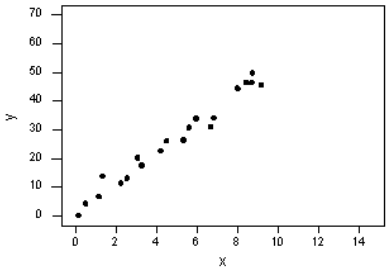
Outliers in Bivariate Data

- _____ with respect to other _____
- in _____, a point that has an _____

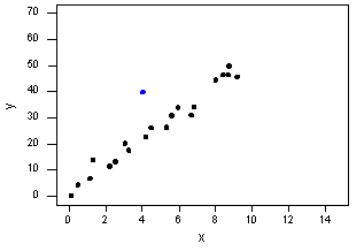
Influential Point in Bivariate Data

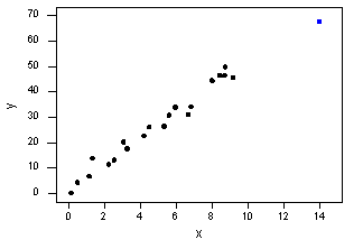
- when _____the _____ changes
- leverage on the _____(aka known as _____)
- normally outliers in _____, but are not always _____in terms of _____ (i.e. _____not large)

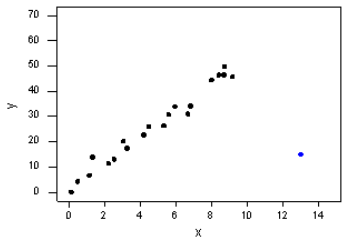
❖ The original data set is graphed at the right.



Classify the new point as a possible outlier and/or an influential point. State whether its **presence** increases or decreases the strength of the association of the variables.

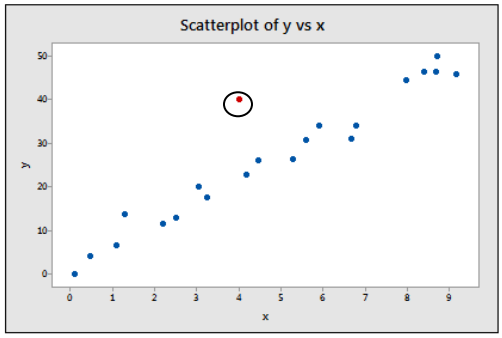


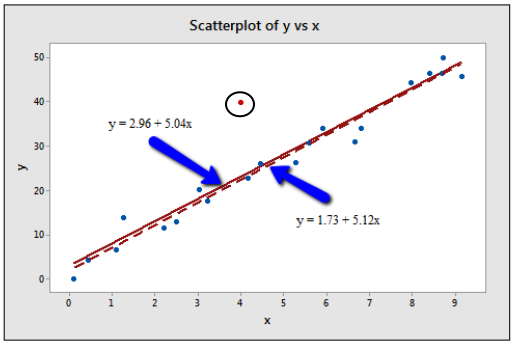




- Go over graphs on pp 234 – 236.

Possible Outlier in Bivariate Data





LSRL including point

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.71075	91.01%	90.53%	89.61%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.96	2.01	1.47	0.157	
x	5.037	0.363	13.86	0.000	1.00

Regression Equation

$y = 2.96 + 5.037\ x$

LSRL excluding point

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.59199	97.32%	97.17%	96.63%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.73	1.12	1.55	0.140	
x	5.117	0.200	25.55	0.000	1.00

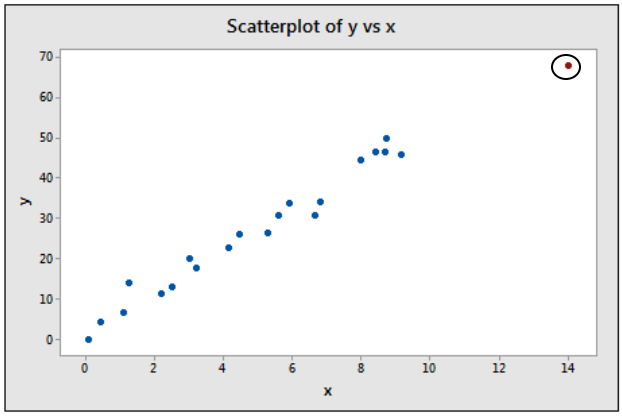
Regression Equation

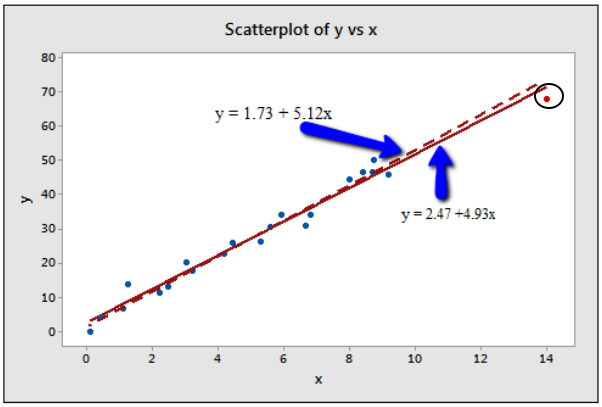
$y = 1.73 + 5.117\ x$

- Presence of point **does not** impact _____ or _____very much, therefore _____ _____.

- Still an _____ in terms of y-values. See S, the standard deviation of the _____.

Possible Influential Point in Bivariate Data





LSRL including point

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.70911	97.74%	97.62%	97.04%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.47	1.08	2.29	0.033	
x	4.927	0.172	28.66	0.000	1.00

Regression Equation

y = 2.47 + 4.927 x

LSRL excluding point

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.59199	97.32%	97.17%	96.63%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.73	1.12	1.55	0.140	
x	5.117	0.200	25.55	0.000	1.00

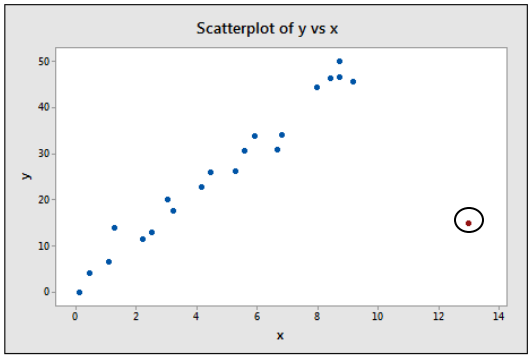
Regression Equation

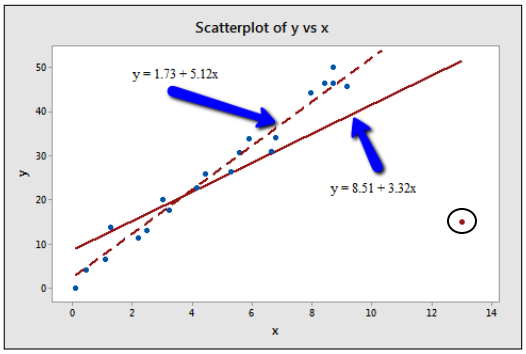
y = 1.73 + 5.117 x

- Presence of point **does not** impact _____ or _____ very much, therefore _____ _____.

- _____ in terms of y-values. See *S*, standard deviation of the _____.

Possible Outlier and Influential Point in Bivariate Data





LSRL including point

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
10.4459	55.19%	52.84%	19.11%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	8.50	4.22	2.01	0.058	
x	3.320	0.686	4.84	0.000	1.00

Regression Equation

y = 8.50 + 3.320 x

LSRL excluding point

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.59199	97.32%	97.17%	96.63%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.73	1.12	1.55	0.140	
x	5.117	0.200	25.55	0.000	1.00

Regression Equation

y = 1.73 + 5.117 x

- Presence of point **does** impacted _____ and _____, therefore _____.

- Also _____ in terms of y-values. See *S*, standard deviation of the _____.

❖ Important Notes:

- _____ points are almost always _____, but not vice-versa.
- _____ in ___ direction may influence _____ but not _____ of the regression line.
- _____ for _____ is "_____" when point is removed
- No _____ for determining _____ and _____. Be able to explain what happens to _____, _____, _____, and _____ when these points are _____ or _____ a scatterplot.

➢ **Assignment p. 238 #59 - 62**

During the months of March and April of a certain year, the weekly weight increases of a puppy in New York were collected. For the same time frame, the retail price increases of snowshoes in Alaska were collected.

Create a scatterplot for this data. Analyze both your graph and the summary statistic in a few sentences below.

The weight of a growing puppy in New York (in pounds)	The retail price of snowshoes in Alaska (in dollars)
8	32.45
8.5	32.95
9	33.45
9.6	34.00
10.1	34.50
10.7	35.10
11.5	35.63

_____

_____

_____

Can we draw this conclusion? " The weight increase of a puppy in New York is _____ the price of snowshoes in Alaska to increase or vice-versa." _____

RULE: _____ and _____ do not demonstrate _____

_____ does not imply _____!!!

_____ follows from _____ only.

Lurking Variables ---has an _____ _____ and yet is not included among the _____ _____ under consideration. Perhaps its _____ is unknown or its _____ _____.

- What could be a lurking variable in these examples? Relate it back to both the explanatory and the response variables.

a. There is a strong positive correlation between the foot length of K-12 students and reading scores.

b. Students who have lower test scores tend to use tutors more often than students who don't.

c. A survey shows a strong positive correlation between the percentage of a country's inhabitants that use cell phones and the life expectancy in that country.

Ex. A group of college students believes that herbal tea has remarkable powers. To test this belief, they make weekly visits to a local nursing home, where they visit with the residents and serve them herbal tea. The nursing home staff reports that after several months many of the residents are more cheerful and healthy.

A skeptical sociologist commends the students for their good deeds but scoffs at the idea that herbal tea helped the residents. Identify the explanatory and response variables in this informal study. Then explain why lurking variables account for the observed association.

Explanatory Variable: _____

Response Variable: _____

_____

_____

_____

_____

_____

KEY IDEAS TO FOCUS ON:

	UNIVARIATE DATA	BIVARIATE DATA
KEY IDEA		
PLOTS		
SHAPE		
IDEAL SHAPE		
MEASURE OF CENTER		
MEASURE OF SPREAD FROM CENTER		

- **Assignment: p. 242 – 243 #63, 64, 66, 67 p. 244 – 247 #69, 70, 73**