Stats Review
Chapter 4

Note:

This review is composed of questions similar to those found in the chapter review and/or chapter test. This review is meant to highlight basic concepts from the course. It does not cover all concepts presented by your instructor. Refer back to your notes, unit objectives, handouts, etc. to further prepare for your exam.

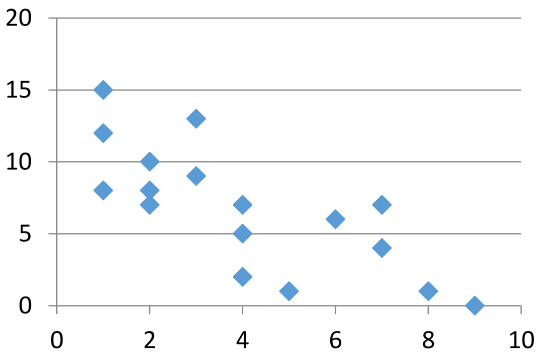
The questions are displayed on one slide followed by the answers are displayed in red on the next.

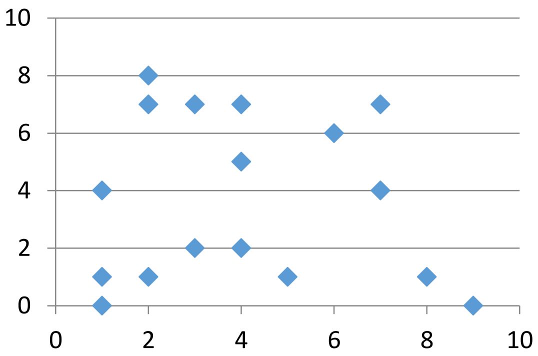
This review is available in alternate formats upon request.

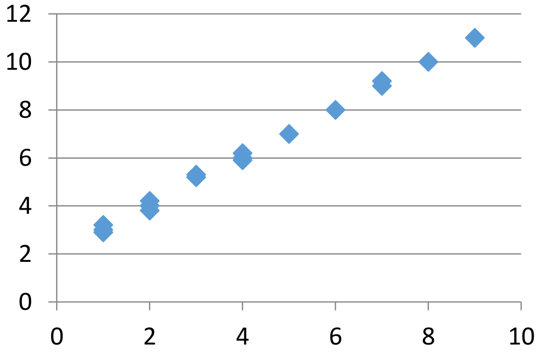
Correlation Coefficient

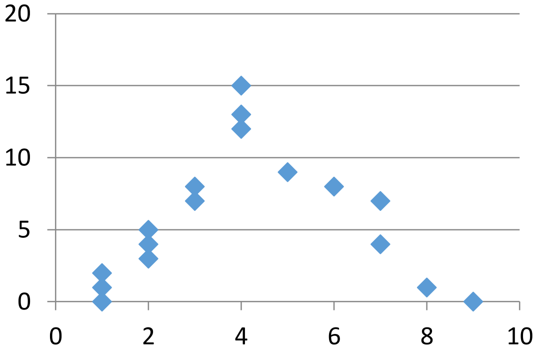
Match the Correlation with the Graph

r=-.6 r=0 r=.99 should not use correlation







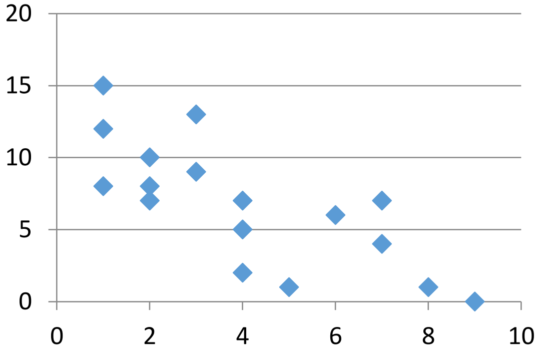


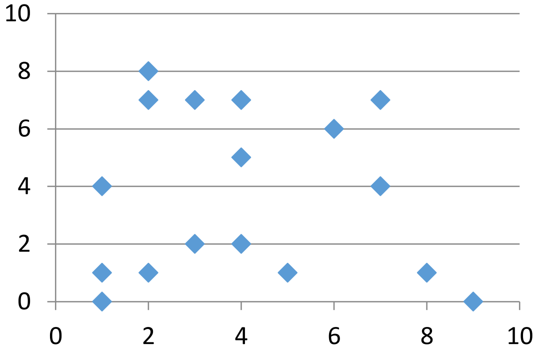
Correlation Coefficient

Match the Correlation with the Graph

r=-.6 r=0 r=.99 should not use correlation

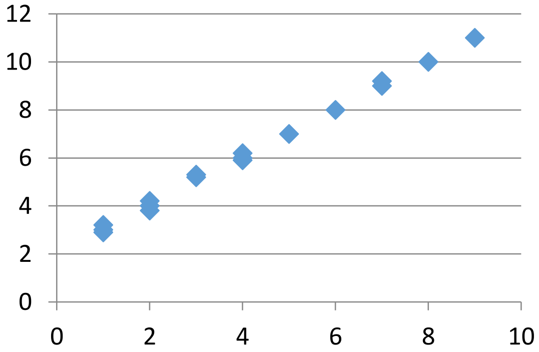
r=-.6

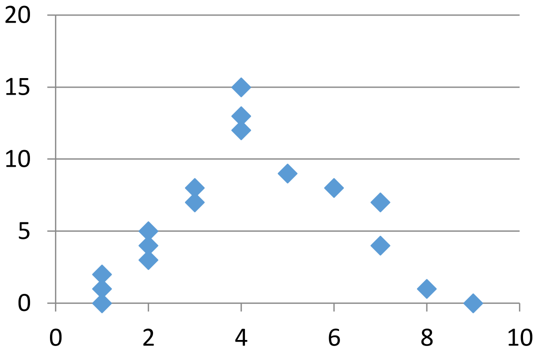




r=0

r=.999





Should not use correlation

Calculate Correlation Coefficient

For the following data

x	41	30	35	45	33
y	94	80	82	102	86

a) determine the correlation coefficient by hand when $\bar{x} = 36.8\ \ s_x = 6.099\ \ \bar{y} = 88.8\ \ s_y = 9.121$

b) Determine the critical value for the correlation coefficient

c) Is there a linear relationship between x and y?

Calculate Correlation Coefficient

a) determine the correlation coefficient

		Step 1 ↓	Step 2 ↓	Step 3 ↓
x	y	$\frac{x_i-\bar{x}}{s_x}$	$\frac{y_i-\bar{y}}{s_y}$	$\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$
41	94	$\frac{41-36.8}{6.099}=$ $.6886$	$\frac{94-88.8}{9.121}$ $=.5701$	$.6886\cdot.5701$ $=.3926$
30	80	-1.1149	-.9648	1.0757
35	82	-.2951	-.7455	.2200
45	102	1.3444	1.4472	1.9456
33	86	-.6231	-.3070	.1913

Step 4: Add the numbers from step 3
=3.8252

Step 5: Divide the sum by n-1

$$=\frac{3.8252}{4}=.9563$$

r=.9563

b) Determine the critical value for the correlation coefficient

Using table II (page A-2), when n=5, the critical value is .878.

c) Is there a linear relationship between x and y?

Yes because the correlation coefficient is greater than the critical value.

Least-Squares Regression Line

The data are the average one-way commute times (in minutes) for selected students and the number of absences for those students during the term.

a) Find the equation of the regression line for the given data. Given $\bar{x}$=86.556, $\bar{y}$=8.556, s_x=9.593, s_y=4.39, r=.98. Round the regression line values to the nearest hundredth.

b) What would be the predicted number of absences if the commute time was 40 minutes? Is this a reasonable question?

c) Interpret the Slope

d) Is it appropriate to determine the y-intercept.

Commute time (x)	Number of absences (y)
72	3
85	7
91	10
90	10
88	8
98	15
75	4
100	15
80	5

Least-Squares Regression Line

a) Find the least-squares regression Line Given $\overline{x}$=86.556, $\overline{y}$=8.556, s_x=9.593, s_y=4.39, r=.98.

Slope (b_1)=r$\left(\frac{s_y}{s_x}\right)$=.98$\left(\frac{4.39}{9.593}\right)$=.45

Y-Intercept (b_0)= $\overline{y}$-b1 $\bar{x}$ =8.556-.45(86.556)=-30.3

The regression Line is $\hat{y} = b_1\text{x} + b_o$ so ours is $\boldsymbol{\hat{y} = .45\text{x} - 30.3}$

b) What would be the predicted number of absences if the commute time was 40 minutes? Is this a reasonable question

Time is 40 minutes or x=40. Put this value into our least-squares regression line $\hat{y} = .45(40) - 30.3$=-12.3

This means that when the commute time is 40 minutes, the number of absences is -12.3. This is not a reasonable question since 40 is outside the scope (i.e. 40 is not within the given range of x values).

c) Interpret the slope

The slope is .45. **This means that for every minute we increase our commute the number of absences increases by .45.**

d) Is it appropriate to determine the y-intercept.

No, because the y-intercept is outside the scope and it does not make sense to have a negative amount absences when the commute time is 0 minutes.

Residuals

Remember from the previous problem that the least squares regression line is $\hat{y} = .45\text{x} - 30.3.$

The time number of absences is 11 when the commute time is 95. Is the number of absences above or below average at this temperature?

Residuals

Remember from the previous problem that the least squares regression line is $\hat{y} = .45\text{x} - 30.3$.

The time number of absences is 11 when the commute time is 95. Is the number of absences above or below average at this temperature?

To answer the question find the residual.
residual = observed y – predicted y (or $y - \hat{y}$)
The observed y (number of absences) is 11 when x (commute time) is 95.
Find the predicted y by substituting 95 into the least squares regression line.

$$\hat{y} = .45(95) - 30.3 = 12.45$$

The residual is then
11-12.45=-1.45

Since the residual is negative, **the observed 11 absences is below average at 95 minutes**.

Find the Sum of Residuals

Remember:

$$\hat{y} = .45\text{x} - 30.3$$

Commute time (x)	Number of absences (y)
72	3
85	7
91	10
90	10
88	8
98	15
75	4
100	15
80	5

Find the Sum of Residuals

Remember:

$$\hat{y} = .45\text{x} - 30.3$$

		Step 1↓	Step 2↓	Step 3↓
Commute time (x)	Number of absences (y)	Predicted $\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
72	3	.45(72)-30.3 =2.1	3-2.1=.9	$(.9)^2$=.81
85	7	7.95	-.95	.9025
91	10	10.65	-.65	.4225
90	10	10.2	-.2	.04
88	8	9.3	-1.3	1.69
98	15	13.8	1.2	1.44
75	4	3.45	.55	.3025
100	15	14.7	.3	.09
80	5	5.7	-.7	.49

Step 1: Using the least squares regression line, find the predicted y values ($\hat{y}$) for each x

Step 2: Calculate the residuals: observed –predicted or y-$\hat{y}$

Step 3: Calculate the residuals squared: $(\text{observed} - \text{predicted})^2$ or $(y-\hat{y})^2$

Step 4: Find the sum of the numbers in the column from step 3. This is the sum of residuals which equals **6.1875**.

Coefficient of Determination (R^2)

a) If the coefficient of determination (R^2) is 86.44% and the data shows a negative association, what is the linear correlation coefficient (r)?

b) Interpret $R^2 = 86.44\%$

Coefficient of Determination (R^2)

a) If the coefficient of determination (R^2) is 86.44% and the data shows a negative association, what is the linear correlation coefficient (r)?

$$r = \sqrt{r^2} = \sqrt{.8644} = .9297$$

Since it has a negative association, **r =-.9297.**

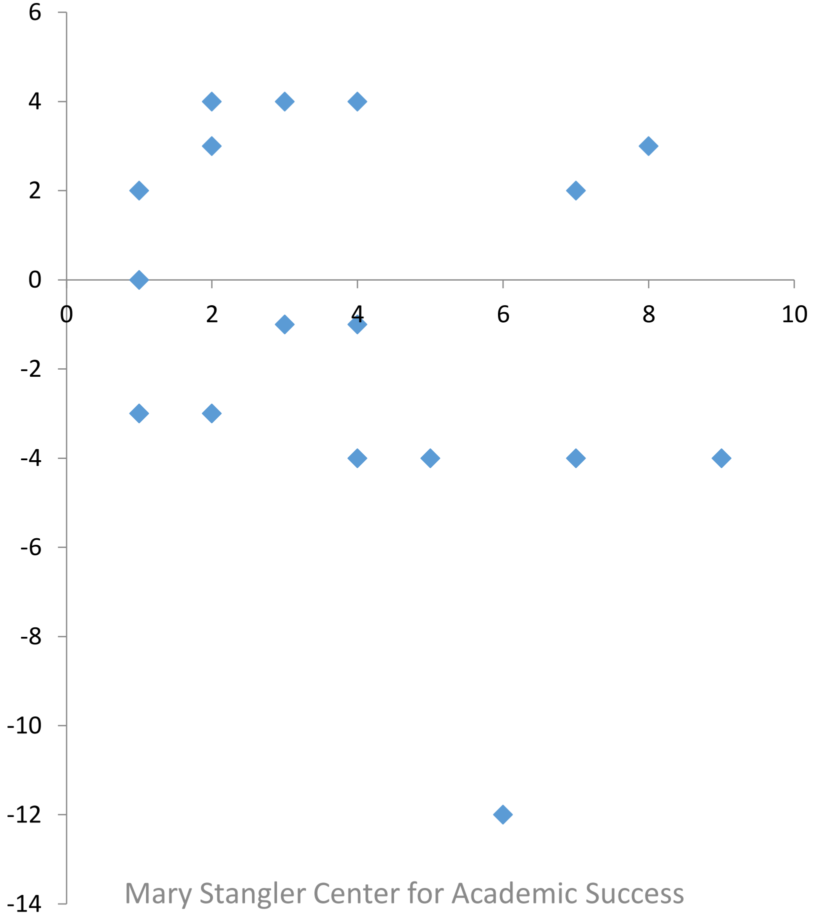
b) Interpret R^2 = 86.44%

86.44% of the variability in y (the response variable) is explained by the least-squares regression line.

Residual Plots

a) What does the residual plot to the right suggest?

b) Removing the outlier, what does the residual plot suggest?



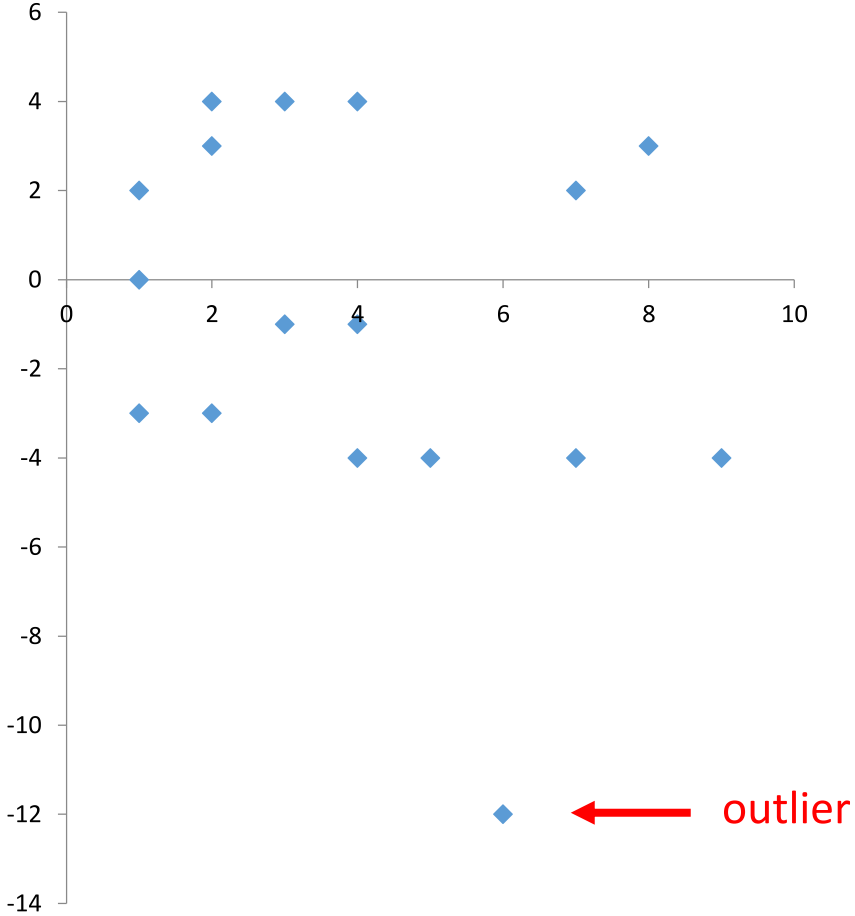
Residual Plots

a) What does the residual plot to the right suggest?

There is an outlier

b) Removing the outlier, what does the residual plot suggest?

No pattern, linear model is appropriate



Contingency Tables

Is there an association between party affliction and gender? The following represents the gender and party affliction of registered voters based on random sample 802 adults.

	Female	**Male**
Republican	105	115
Democrat	150	103
Independent	150	179

a) Construct a frequency marginal distribution
b) Construct a relative frequency marginal distribution
c) Construct a conditional distribution of party affiliation by gender
d) Is gender associated with party affiliation? If so, how?

Contingency Tables parts a) and b)

a) Construct a frequency marginal distribution

	Gender		**frequency marginal distribution**
Party	**Female**	**Male**	
Republican	105	115	=105+115=220
Democratic	150	103	253
Independent	150	179	329
frequency marginal distribution	=105+150+150=405	397	802

To do: Find the total for each row and column

b) Construct a relative frequency marginal distribution

	Gender		**relative frequency marginal distribution**
Party	**Female**	**Male**	
Republican	105	115	.274
Democratic	150	103	.315
Independent	150	179	.410
relative frequency marginal distribution	=405/802=.505	.495	1

To do: Divide the row/column total by the sample size

Contingency Tables parts c) and d)

c) Construct a conditional distribution of party affiliation by gender

	Gender	
Party	**Female**	**Male**
Republican	=105/405=.259	=115/397=.290
Democratic	.370	.259
Independent	.370	.451
Total	1	1

To do: Divide the each cell by its column total

d) Is gender associated with party affiliation? If so, how?

Yes; males are more likely to be Independents and less likely to be democrats.