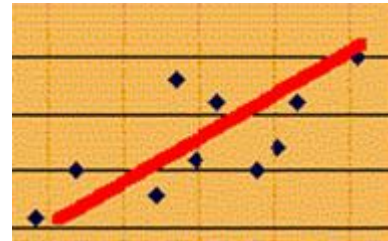
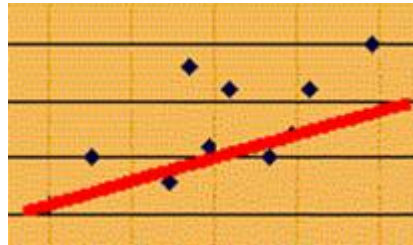
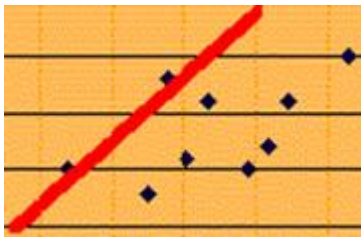


Sample Questions on Linear Regression

1. Please state, in your own words, what the following values mean, and in which situation they apply:
 - a) $r = 0.02$
 - b) $r = -0.8$
 - c) $r = 0.95$
 - d) Suppose you compute the equation of a least-square regression line as $y = -2x + 3$ and the correlation coefficient $r = 0.8$. Can that be right?

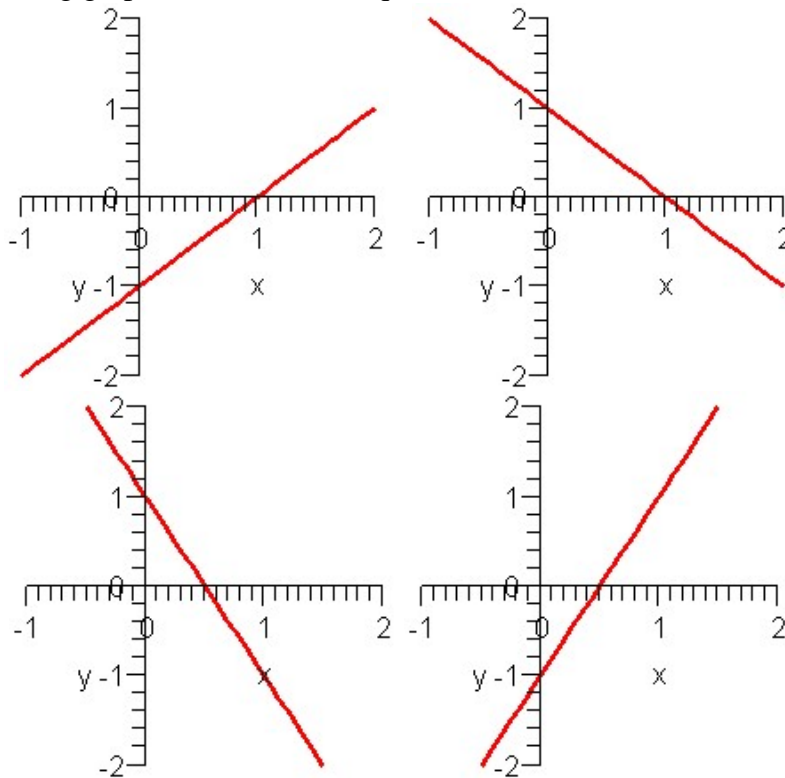
2. Which of the following line would be best for a “least square regression line” through the given data:



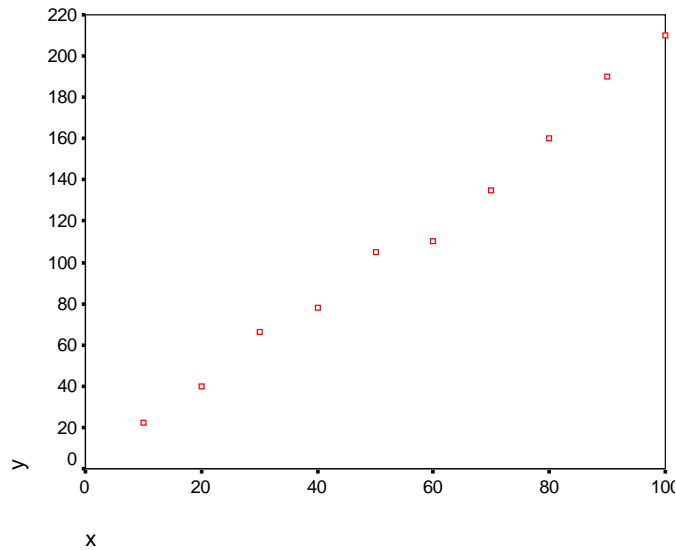
3. Suppose we have four equations of lines as follows:

(a) $y = x - 1$ (b) $y = 2x - 1$ (c) $y = -x + 1$ (d) $y = -2x + 1$

Which of the following graphs matches what equation?



4. When using the computer to draw a “scatter plot, it comes up with the following picture:



- Draw a “best-fit” line through this data.
- Use the line to *estimate* the y-intercept and slope of the equation of the least-square regression line (a very rough estimate of the values is okay)
- Look at the data and your line and estimate whether r would be close to -1, close to 0, or close to 1

5. We want to see if there is a relation between engine size and horse power of cars. Approximately 400 cars were randomly selected for the study, and we used the computer to compute a “linear regression”.

We used as X-variable (independent) the engine size and as Y-variable (dependent) the horse powers. StatCrunch came up with the following output:

Simple linear regression results:

Dependent Variable: Horse powers

Independent Variable: Engine Size (cube in)

Horse powers = 41.00038 + 0.32643723 Engine Size (cube in)

Sample size: 392

R (correlation coefficient) = 0.8959

R-sq = 0.80271465

Estimate of error standard deviation: 17.00383

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	41.00038	1.8043838	≠ 0	390	22.72265	<0.0001
Slope	0.32643723	0.008194721	≠ 0	390	39.835064	<0.0001

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	458801.2	458801.2	1586.8322	<0.0001
Error	390	112760.805	289.13025		
Total	391	571562			

- a) Find the exact equation of the least-square regression line
- b) What is the correlation coefficient r and what does it mean
- c) Predict the horse power of an engine with engine size 500 cubic inches.
- d) Do you think your prediction is accurate? Why?

Please note that the above data is somewhat old. Today's engines are somewhat more efficient (which might warrant another study -☺)

6. Consider the following data, listing years of schooling for a respondent of a survey and his or her father. We want to determine if there is a linear relation between the variables and use the least-square regression line to make predictions.

Highest year of school completed, father	Highest year of school completed			
12	12			
15	16			
5	5			
16	19			

- a) Compute the correlation coefficient r and the equation of the least-square regression line $y = mx + b$. Recall the corresponding formulas:

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}, \quad S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}, \quad S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}, \quad \text{slope: } m = \frac{S_{xy}}{S_{xx}}, \quad \text{y-intercept: } b = \bar{y} - m\bar{x}$$

- b) Draw a scatter plot for the above data and draw a regression line “by hand”. Compare your line with the equation you obtained earlier. Is there a match? Does the value of the correlation coefficient make sense? Explain.
- e) Predict the years of school for a person whose father has completed 18 years of schooling? Do you think your prediction is reasonably accurate? Why?
7. Use the following data file to find out whether there is a linear relationship between the salary and the years of education of an employee. Which variable is the independent, which the dependent one? Make sure to include a scatter plot. Also, predict the salary of an employee with 16 years of education.

To load the data, start StatCrunch, open the StatCrunch spreadsheet, and click (on the left side) on “Web Address”. Copy-and-paste (or type carefully without mistakes) the following address into the “WWW Address” field:

<http://www.mathcs.org/statistics/course/00-data/employeenumeric.xls>

then press RETURN