Linear Regression

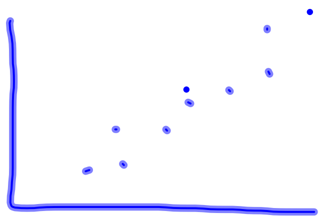
Scatter plot: make sure you have title, labels, and scales
Used for finding a relationship between two variables

Independent variable: the variable we control (x)

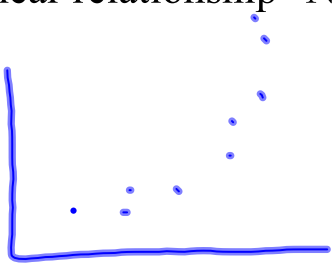
Dependent variable: the variable we measure (y)

Examples:

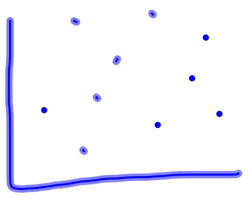
Linear relationship



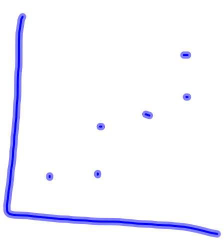
Non-linear relationship



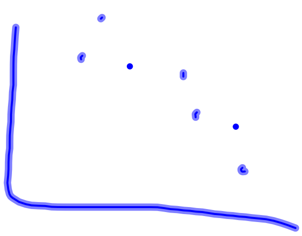
No relationship



Positive relationship



Negative relationship



The correlation coefficient, r, measures the strength of the linear relationship between x and y. It is always between -1 and +1.

$r = +1$ is a perfect positive linear relationship

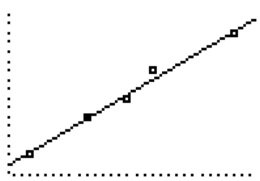
$r = 0$ is absolutely no relationship

$r = -1$ is a perfect negative linear relationship

To find the regression equation for a set of data, fill in a chart similar to the one below.
Write the summations (add all the x's, y's, etc.) in the last row.

# of students	seconds			
x	y	xy	x^2	y^2
5	3	15	25	9
12	7	84	144	49
17	9	153	289	81
20	12	240	400	144
30	16	480	900	256
Σ 84	47	972	1758	539

```
LinReg
 y=ax+b
 a=.5259515571
 b=.5640138408
 r²=.9869708232
 r=.9934640523
```



The regression equation is always written as $y = mx + b$
where m = slope and b = y-intercept.

$$m = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{5(972) - (84)(47)}{5(1758) - (84)^2}$$

$$b = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{\quad}{1734}$$

$$\frac{912}{1734} = .52$$

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{\left(n(\Sigma x^2) - (\Sigma x)^2\right)\left(n(\Sigma y^2) - (\Sigma y)^2\right)}} = \frac{912}{\sqrt{(1734)(\quad)}}$$

Residuals are the distance the actual points are from the regression line.

If the data is linear, the residuals should be random.

If the data is non-linear, the residuals should show a pattern.

Homework
p.125
#1-29
Due Tuesday, March 19