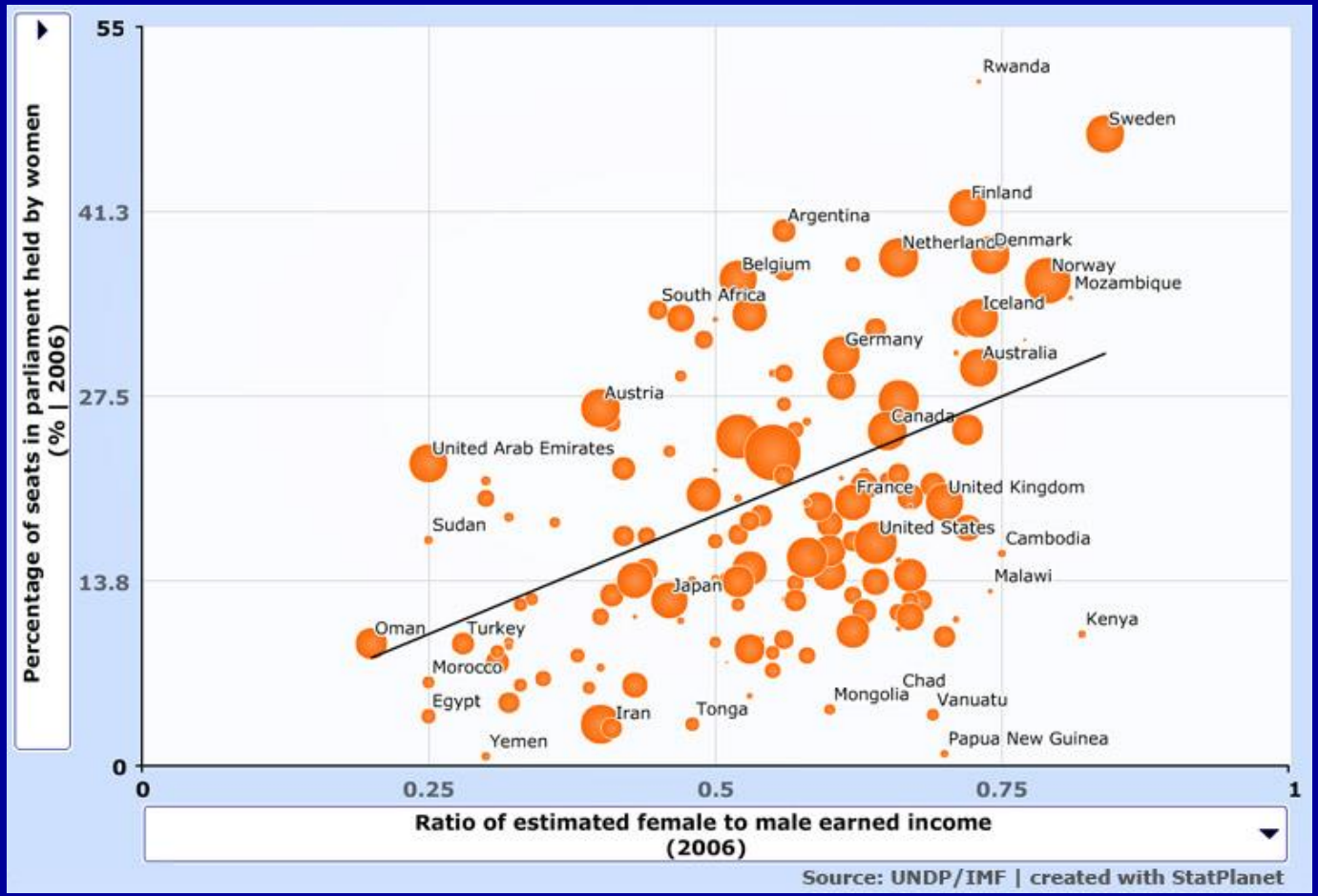
Correlation

Dr. Hawkar Qasim Birdawod

Defining Correlation

- Co-variation or co-relation between two variables
- These variables change together
- Usually scale (interval or ratio) variables

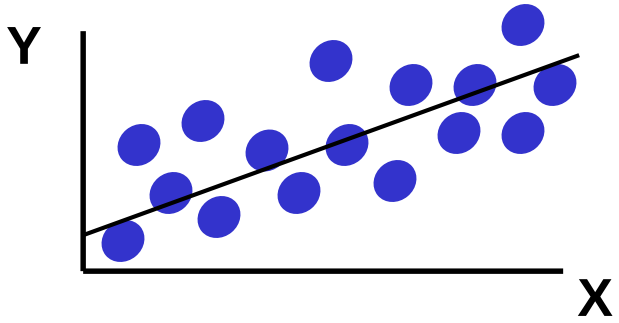

Percentage of seats in parliament held by women (% | 2006)
55
41.3
27.5
13.8
0
0
0.25
0.5
0.75
1
Ratio of estimated female to male earned income (2006)
Source: UNDP/IMF | created with StatPlanet
Rwanda
Sweden
Finland
Argentina
Netherland
Denmark
Belgium
Norway
Mozambique
South Africa
Iceland
Germany
Australia
Austria
Canada
United Arab Emirates
France
United Kingdom
Sudan
United States
Cambodia
Malawi
Japan
Kenya
Oman
Turkey
Morocco
Chad
Egypt
Mongolia
Vanuatu
Iran
Tonga
Yemen
Papua New Guinea

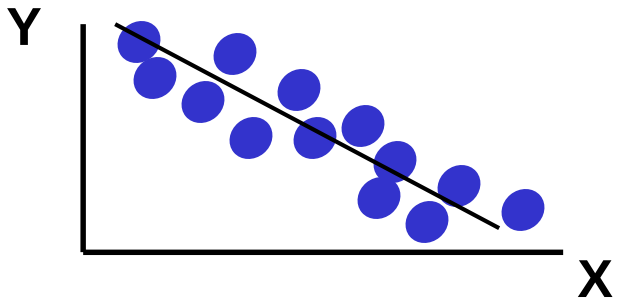
Correlation Coefficient

- A statistic that quantifies a relation between two variables
- Can be either positive or negative
- Falls between -1.00 and 1.00
- The value of the number (not the sign) indicates the strength of the relation

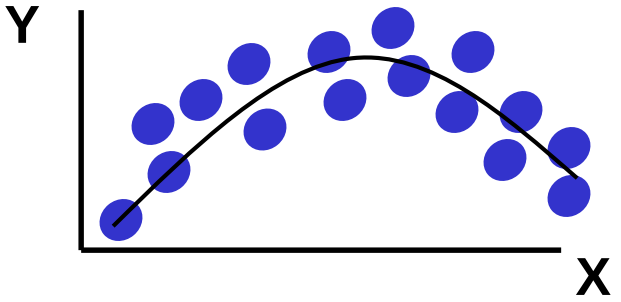
Linear Correlation

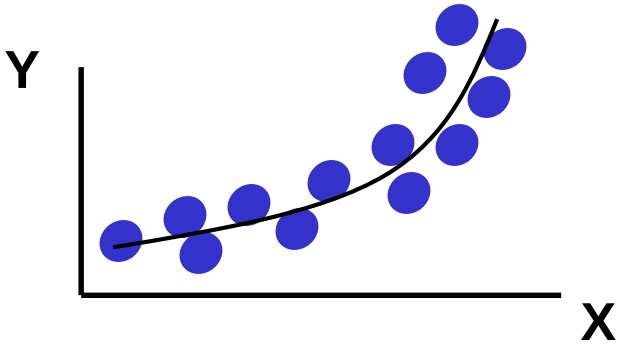
Linear relationships





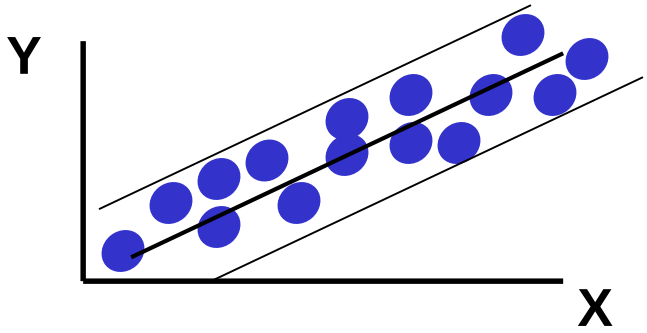
Curvilinear relationships

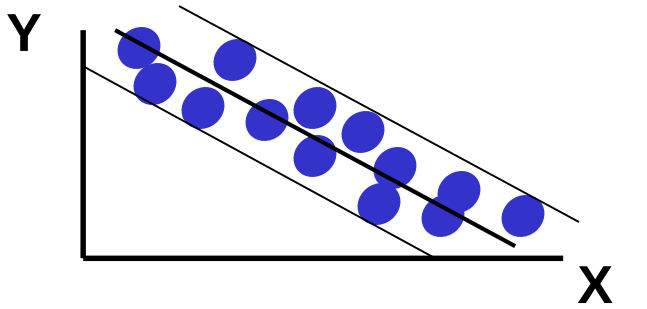




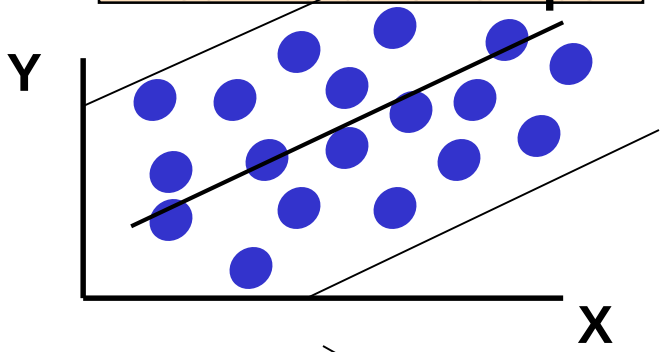
Linear Correlation

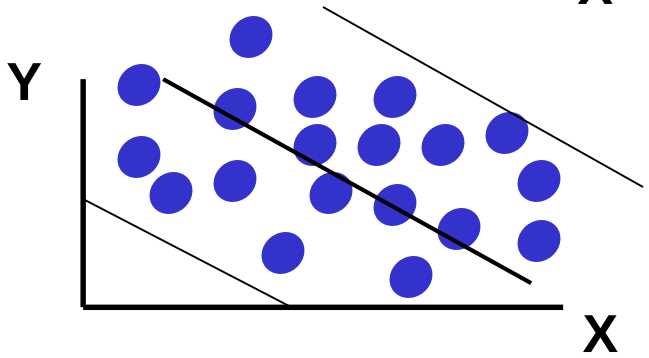
Strong relationships





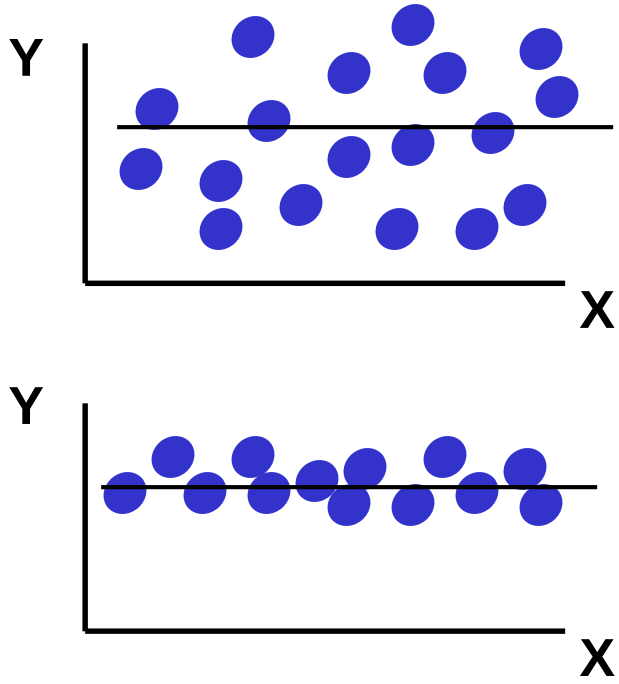
Weak relationships



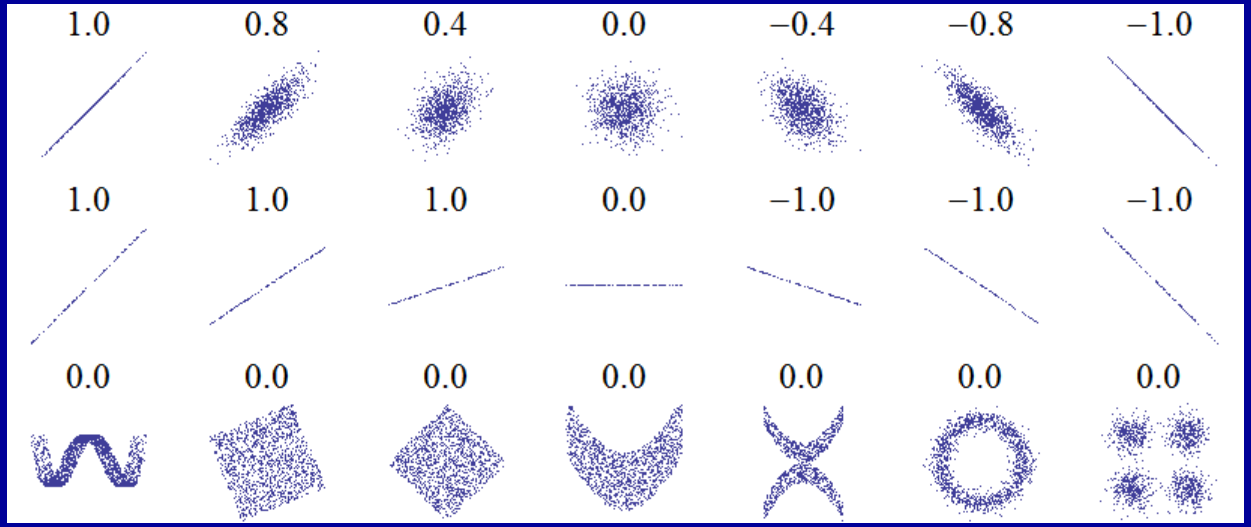


Linear Correlation

No relationship



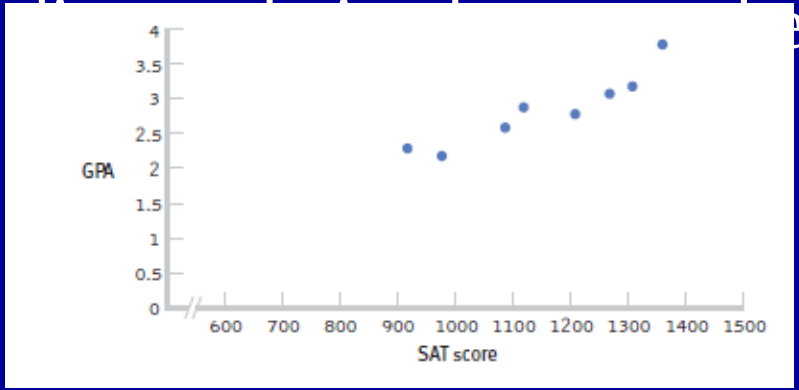
Correlation



Positive Correlation

Association between variables such that high scores on one variable tend to have high scores on the other variable

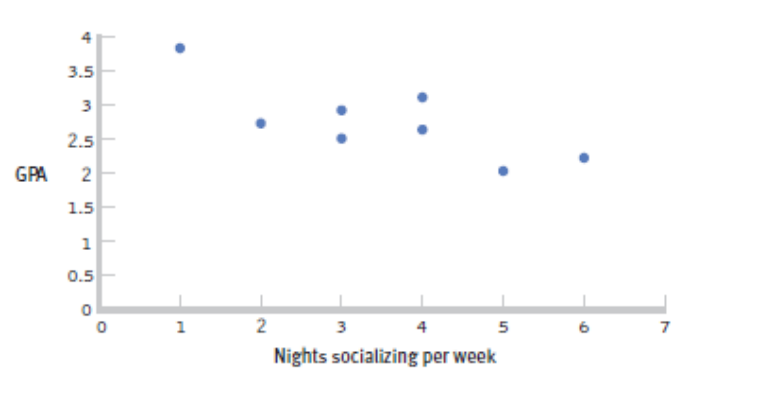
A [illegible] variables



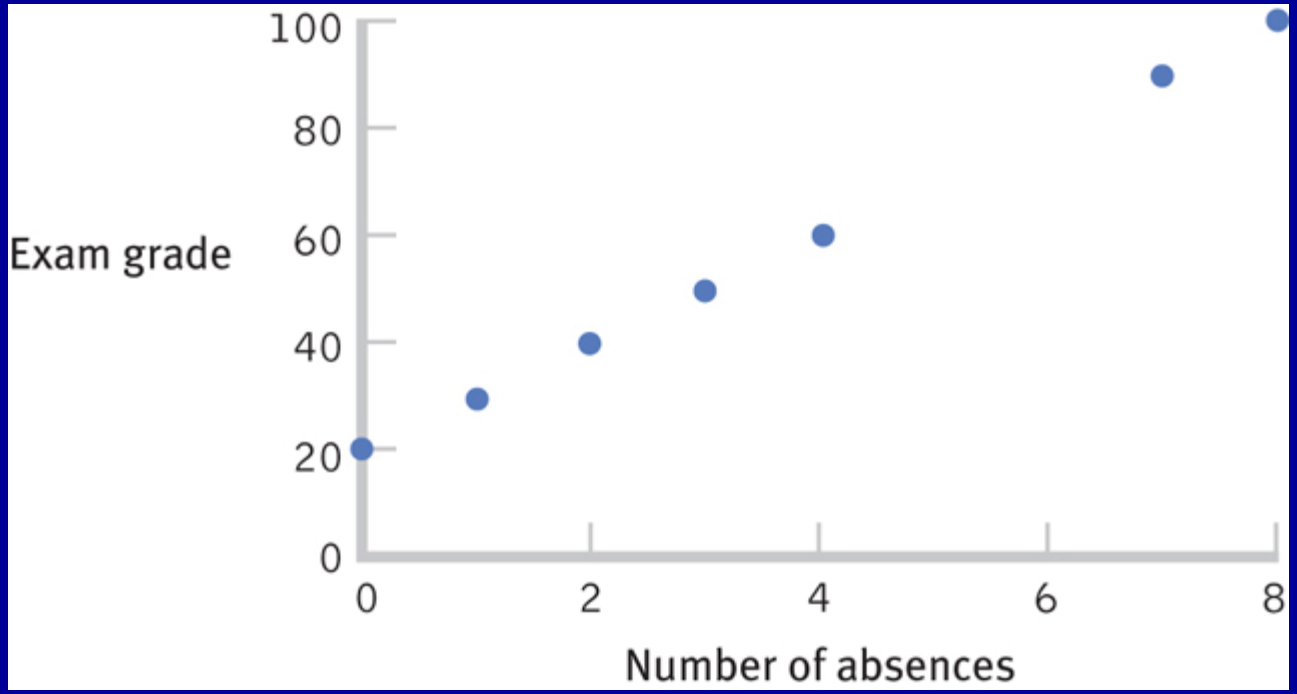
Negative Correlation

Association between variables such that high scores on one variable tend to have low scores on the other variable

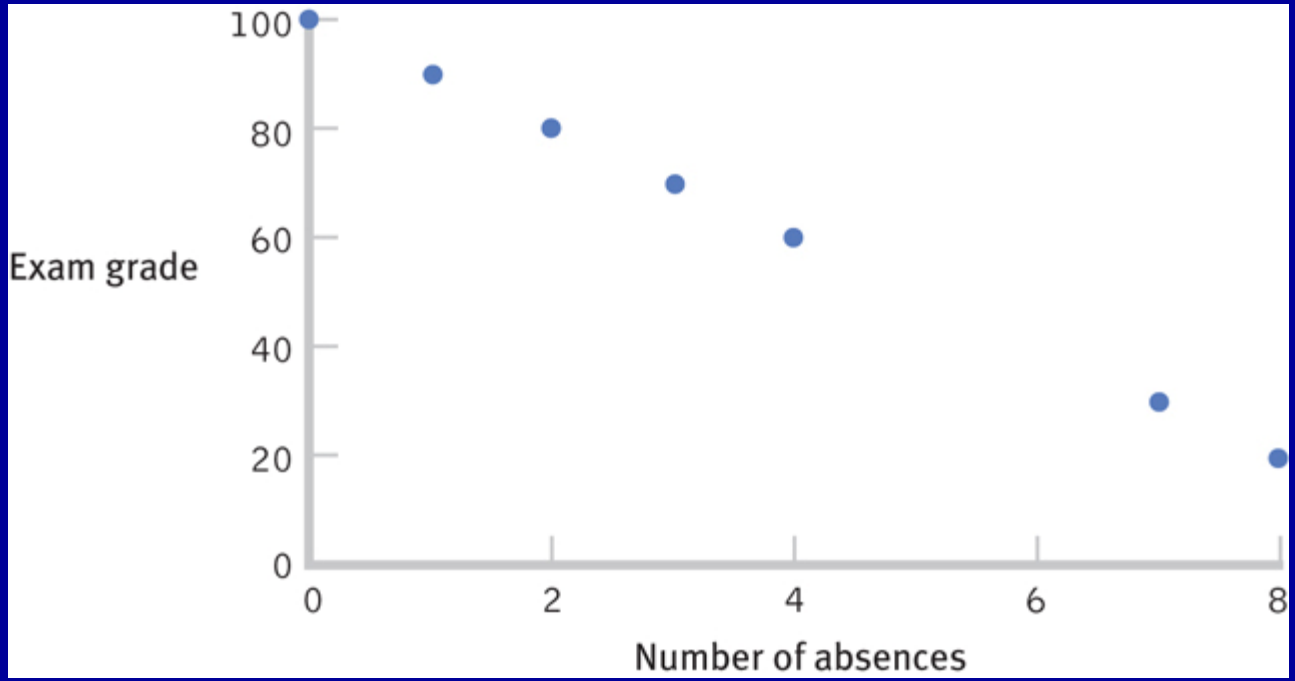
An in[illegible]e variables



A Perfect Positive Correlation

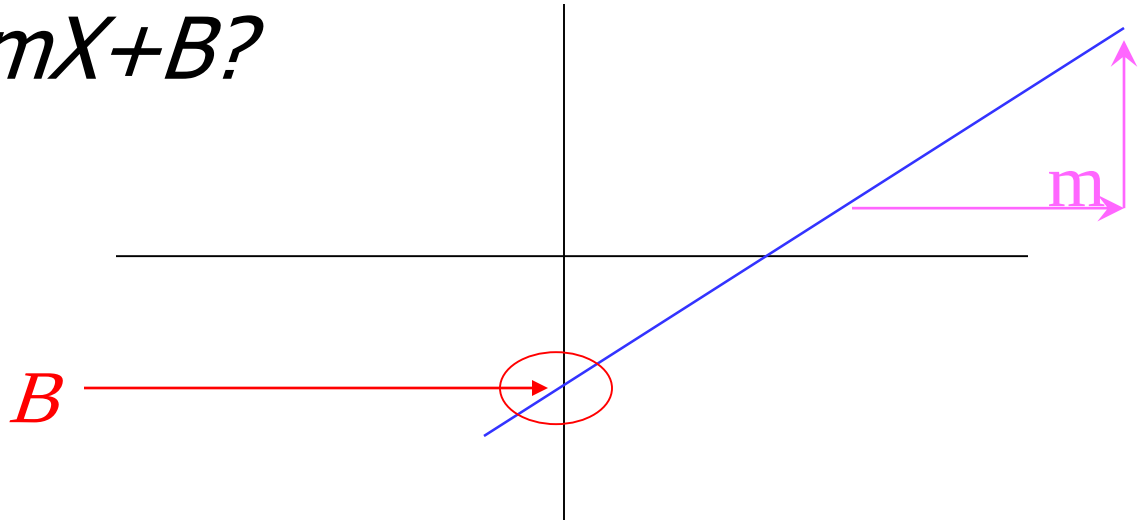


A Perfect Negative Correlation



What is "Linear"?

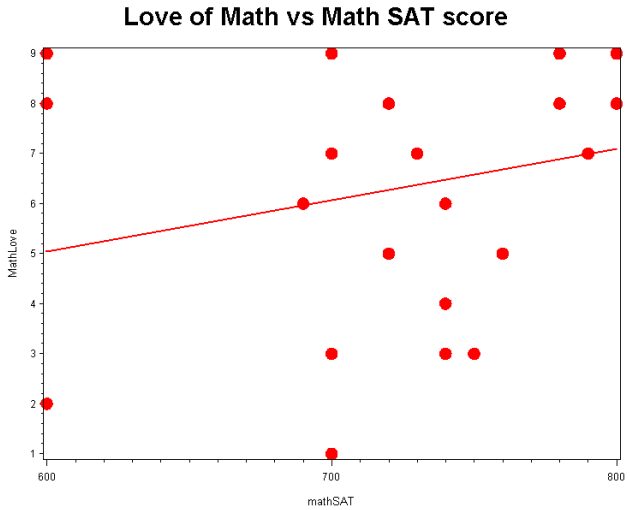
- Remember this:
- *Y=mX+B?*



What's Slope?

A slope of 2 means that every 1-unit change in X yields a 2-unit change in Y.

Simple linear regression



P=.22; not significant

The linear regression model:

Love of Math = 5 + .01*math SAT score

intercept

slope

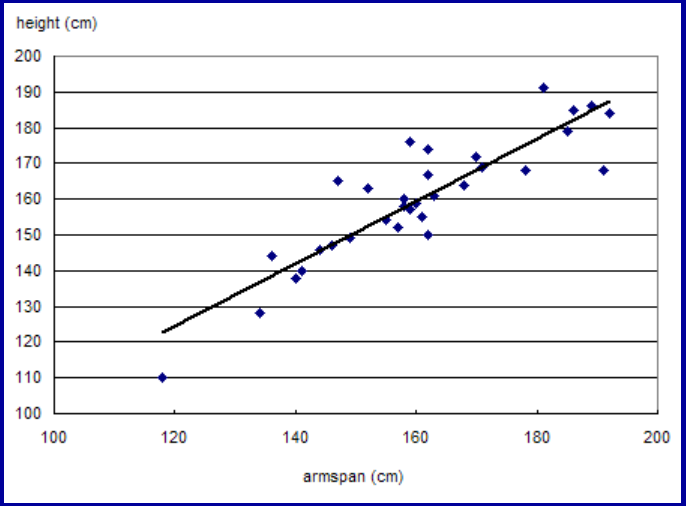
Misleading Correlations

- Something to think about
  - There is a 0.91 correlation between ice cream consumption and drowning deaths.
    - Does eating ice cream cause drowning?
    - Does grief cause us to eat more ice cream?

Correlation

Correlation is NOT causation

-e.g., armspan and height



The Pearson Correlation Coefficient

- A statistic that quantifies a linear relation between two scale variables.
- Symbolized by the italic letter *r* when it is a statistic based on sample data.
- Symbolized by the italic letter *p* "rho" when it is a population parameter.

- Pearson correlation coefficient
  - r
  - Linear relationship

$$r = \frac{\sum[(X - M_X)(Y - M_Y)]}{\sqrt{(SS_X)(SS_Y)}}$$

Correlation Hypothesis Testing

- Step 1. Identify the population, distribution, and assumptions
- Step 2. State the null and research hypotheses.
- Step 3. Determine the characteristics of the comparison distribution.
- Step 4. Determine the critical values.
- Step 5. Calculate the test statistic

TABLE 15-2. Is Skipping Class Related to Statistics Exam Grades?

Here are the scores for 10 students on two scale variables: number of absences from class in one semester and exam grade.

Student	Absences	Exam Grade
1	4	82
2	2	98
3	2	76
4	3	68
5	1	84
6	0	99
7	4	67
8	8	58
9	7	50
10	3	78

Always Start with a Scatterplot

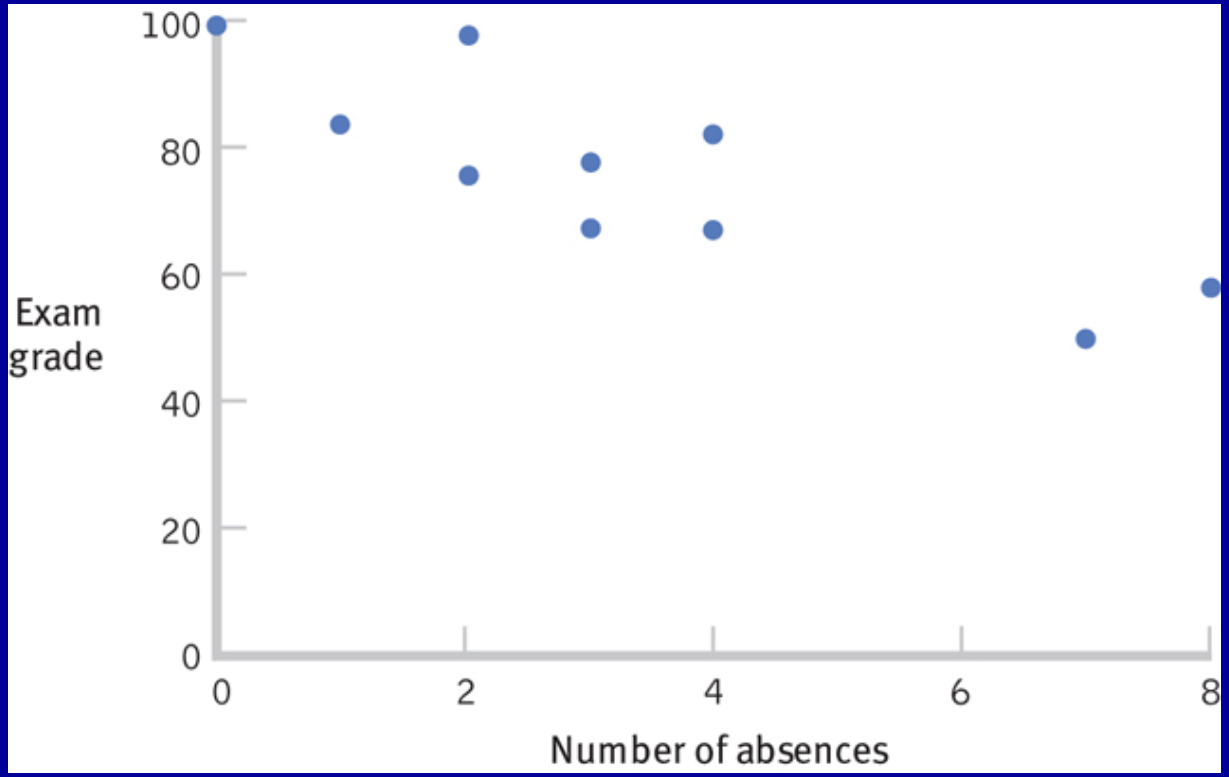


TABLE 15-3. Calculating the Numerator of the Correlation Coefficient

Absences (X)	$(X - M_X)$	Exam Grade (Y)	$(Y - M_Y)$	$(X - M_X)(Y - M_Y)$
4	0.6	82	6	3.6
2	−1.4	98	22	−30.8
2	−1.4	76	0	0.0
3	−0.4	68	−8	3.2
1	−2.4	84	8	−19.2
0	−3.4	99	23	−78.2
4	0.6	67	−9	−5.4
8	4.6	58	−18	−82.8
7	3.6	50	−26	−93.6
3	−0.4	78	2	−0.8
$M_X = 3.400$		$M_Y = 76.000$		$\Sigma[(X - M_X)(Y - M_Y)] = -304.0$

TABLE 15-4. Calculating the Denominator of the Correlation Coefficient

Absences (X)	$(X - M_X)$	$(X - M_X)^2$	Exam Grade (Y)	$(Y - M_Y)$	$(Y - M_Y)^2$
4	0.6	0.36	82	6	36
2	−1.4	1.96	98	22	484
2	−1.4	1.96	76	0	0
3	−0.4	0.16	68	−8	64
1	−2.4	5.76	84	8	64
0	−3.4	11.56	99	23	529
4	0.6	0.36	67	−9	81
8	4.6	21.16	58	−18	324
7	3.6	12.96	50	−26	676
3	−0.4	0.16	78	2	4
		$\Sigma(X - M_X)^2 = 56.4$			$\Sigma(Y - M_Y)^2 = 2262$

Thank You