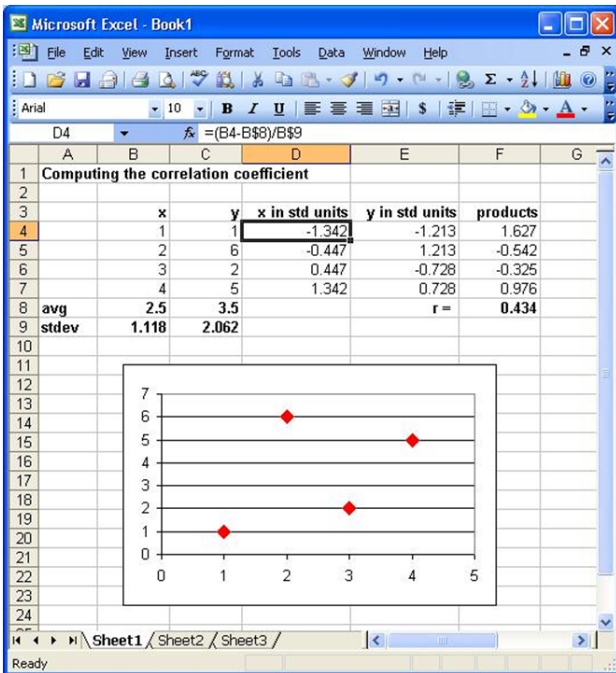


Unit 4: Correlation and Causation



“Covariance formula” for r (FPP p.134)

$$r = \frac{1}{\sigma_x \sigma_y} ((\text{avg of } x_i y_i) - \overline{xy})$$

Numerator of right side is the “covariance of x and y ”

$$\begin{aligned} r &= \text{avg of products } x_i y_i \text{ in std units} = \left(\sum_i \frac{x_i - \bar{x}}{\sigma_x} \frac{y_i - \bar{y}}{\sigma_y} \right) / n \\ &= \frac{1}{\sigma_x \sigma_y} \sum_i (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) / n \\ &= \frac{1}{\sigma_x \sigma_y} \left(\frac{\sum_i x_i y_i}{n} - \frac{\sum_i x_i}{n} \bar{y} - \bar{x} \frac{\sum_i y_i}{n} + \frac{n \bar{x} \bar{y}}{n} \right) \\ &= \frac{1}{\sigma_x \sigma_y} ((\text{avg of } x_i y_i) - \bar{x} \bar{y} - \bar{x} \bar{y} + \bar{x} \bar{y}) \\ &= \frac{1}{\sigma_x \sigma_y} ((\text{avg of } x_i y_i) - \overline{xy}) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \end{aligned}$$

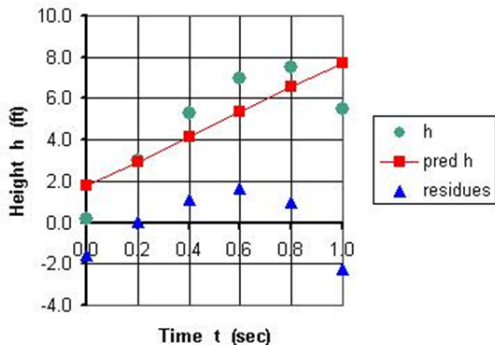
D4 fx =ROUND((B4-B\$12)/B\$13,1)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Computing the correlation coefficient in two ways												
2	by standard formula							by covariance formula					
3		x-value	y-value	x in std units	y in std units	products			x-value	y-value	products		
4		1	2	-1.5	-1.3	2			1	2	2		
5		2	4	-1.1	0.1	-0.1			2	4	8		
6		3	3	-0.7	-0.6	0.4			3	3	9		
7		4	4	-0.2	0.1	0			4	4	16		
8		5	5	0.2	0.7	0.1			5	5	25		
9		6	3	0.7	-0.6	-0.4			6	3	18		
10		7	3	1.1	-0.6	-0.7			7	3	21		
11		8	7	1.5	2.1	3.2			8	7	56		
12	Avg	4.5	3.9			0.6		Avg	4.5	3.9	19.4		
13	SD	2.3	1.5					SD	2.3	1.5	0.5	Real r =	0.582217
14													
15		1001	2.1	-1.5	-1.2	1.8			1001	2.1	2102.1		
16		1002	4.1	-1.1	0.1	-0.1			1002	4.1	4108.2		
17		1003	2.9	-0.7	-0.7	0.5			1003	2.9	2908.7		
18		1004	3.9	-0.2	0	0			1004	3.9	3915.6		
19		1005	5.1	0.2	0.8	0.2			1005	5.1	5125.5		
20		1006	2.9	0.7	-0.7	-0.5			1006	2.9	2917.4		
21		1007	2.9	1.1	-0.7	-0.8			1007	2.9	2920.3		
22		1008	7.1	1.5	2.1	3.2			1008	7.1	7156.8		
23	Avg	1004.5	3.9			0.5		Avg	1005	3.9	3894.3		
24	SD	2.3	1.5					SD	2.3	1.5	6.7	Real r =	0.549869

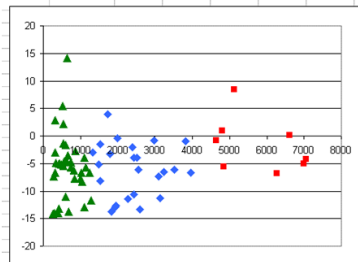
D3

=B3-C3

	A	B	C	D	E	F
1	A ball is thrown upwards. The height h in ft after t sec:					
2	t	h	pred h	residues		
3	0.0	0.2	1.771	-1.571		
4	0.2	3.0	2.963	0.037		
5	0.4	5.3	4.154	1.146		
6	0.6	7.0	5.346	1.654		
7	0.8	7.5	6.537	0.963		
8	1.0	5.5	7.729	-2.229		
9						



Village populations in CNY vs. % population increase 1990 to 1998



r's:

small -0.086383

medium -0.030814

large -0.402356

overall 0.137419

SAT scores Average scores from school districts in Cayuga, Madison, and Oswego counties for the 1998-99 school year

District	C	Overall	Verbal	Math
Auburn	C	1068	525	543
Cato-Meridian	C	1057	528	529
Moravia	C	1064	523	541
Otselic Valley	C	1010	506.4	503.6
Port Byron	C	999	498	501
Southern Cayuga	C	1111	541	570
Brookfield	M	1002	509	493
Canastota	M	1001	494	507
Cazenovia	M	1046	518	528
Chittenango	M	998	491	507
DeRuyter	M	987.7	491.1	496.6
Hamilton	M	1082	534	548
Madison	M	978	496	482
Morrisville-Eaton	M	1035	503	532

Oneida	M	1034	508	526
Stockbridge Valley	M	1000	490	510
A-P-W	O	1039	512.3	526.3
Central Square	O	1023	507	518
Fulton	O	1044	522	522
Hannibal	O	1017	504	513
Mexico	O	1039	519	520
Oswego	O	1059	522	537
Phoenix	O	998	494	504
Pulaski	O	1028	522	502
Sandy Creek	O	1000	490	510

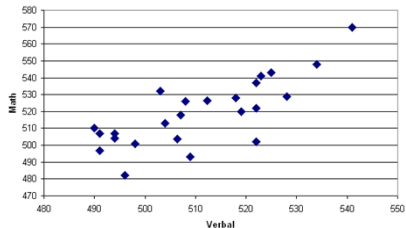
Verbal vs. Math $r = 0.770$

	Overall	Verbal	Math
Cayuga County	1063.9	524.3	539.6
Madison County	1015.8	503.7	512.1
Oswego County	1027.4	510.2	516.9

Ecol $r = 0.989$

Scatterplots of the last two slides

District SATs, 1998-99



County SATs, 1998-99

