Pairs of Data
Linear Correlation

Math 122

Scatter Plots

TI

- Enter data values in lists L1 and L2

- Press 2ND -> Stat Plot

- Select a plot

- Press ZOOM -> ZooomStat

Online Calculator

- Enter data values in List 0 and 1

- Press StatPlot

- Select scatter plot

- Press calculate

Fall Track Testing

400m	30m	Fly 30m	Vert	St LJ	St TJ	Push ups	Sit ups	Triple Bound	Total Pts	Weight	Height
78.30	4.75	4.45	16	77	203	43	48	233	321	145.0	63.25
65.90	4.45	4.06	18	81.5	235.5	38	41	230	393	133.4	68
66.50	4.59	4.17	18.5	84	242	46	58	243	414	140.4	64.25
69.50	4.43	4.04	20	88	240	40	42	246	409	132.2	66.75
64.60	4.25	4.02	20.5	92	242	42	56	262	451	132.2	65.5
68.90	4.36	4.00	18	79	236	50	60	242	425	137.6	62.5
65.40	4.28	3.92	22	91.5	243.5	42	44	261	461	151.0	71.5
72.80	4.58	4.23	17	75	220.5	38	41	233	338	123.0	68.25
61.20	4.24	3.81	18	73	217	42	42	222	405	139.2	65.5
67.80	4.26	3.89	21.5	85	252.5	54	55	266	463	140.2	67
78.30	4.59	4.20	17.5	71.75	227	40	67	234	356	140.2	64.25
66.70	4.39	4.02	20	83	244	40	46	253	414	127.0	66.25
65.70	4.26	3.95	21	78.5	246	42	50	236	423	135.8	66.5
67.80	4.48	4.11	17.5	77	221	54	46	248	400	141.6	65.75
64.40	4.36	3.93	23	81	240	50	60	255	466	164.6	67
64.90	4.37	3.88	21	94.5	264.75	49	61	254	477	111.8	63.5
66.90	4.27	3.92	18.5	83	250	36	33	245	397	114.8	63.5
73.40	4.69	4.04	17	76	213.5	54	60	236	357	113.2	57.75
67.20	4.35	4.04	20	84.5	243	38	51	256	420	141.8	67
62.80	4.15	3.76	23.5	92	252.5	48	57	284	496	118.2	66.25

A scatter plot of matched pairs of data may show a correlation or relationship between the data values

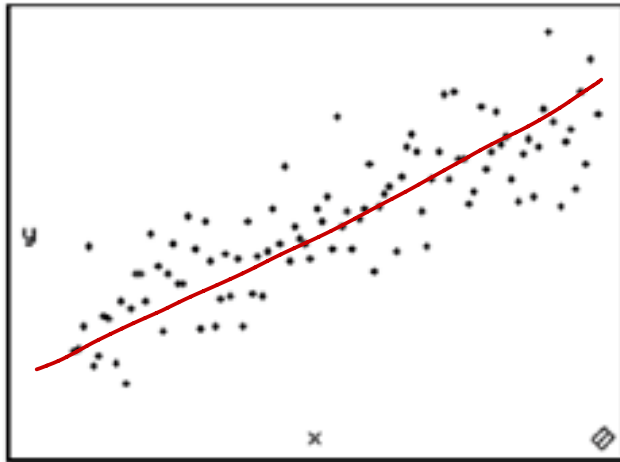# Correlation

A correlation exists between two variables when the values of one are somehow associated with the values of the other in some way.

A linear correlation exists between two variables when there is a correlation and the plotted points of paired data result in a pattern that can be approximated by a straight line.
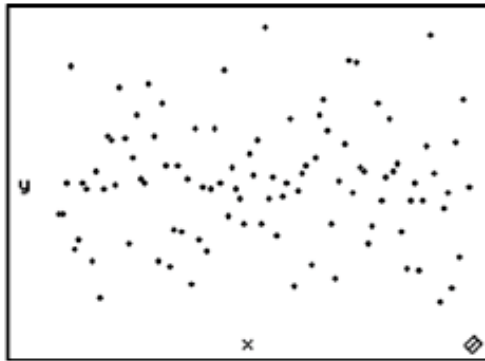
Correlation Coefficient

For any sample of paired data, we can calculate a variable r called the linear correlation coefficient that tells us how much linear correlation there is between the variables involved.
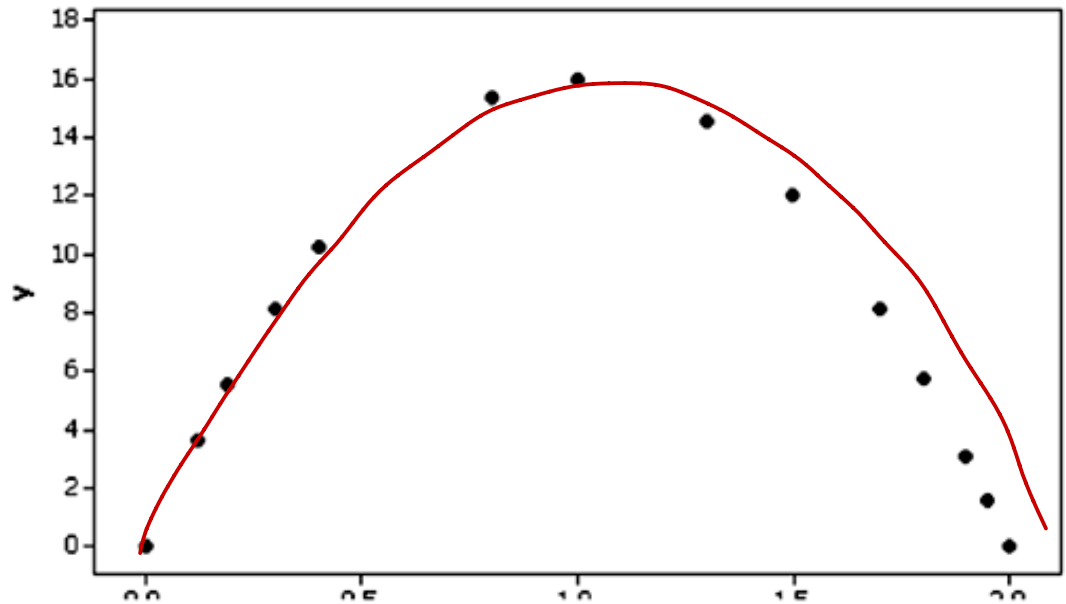
(a) Positive correlation:

$r = 0.851$
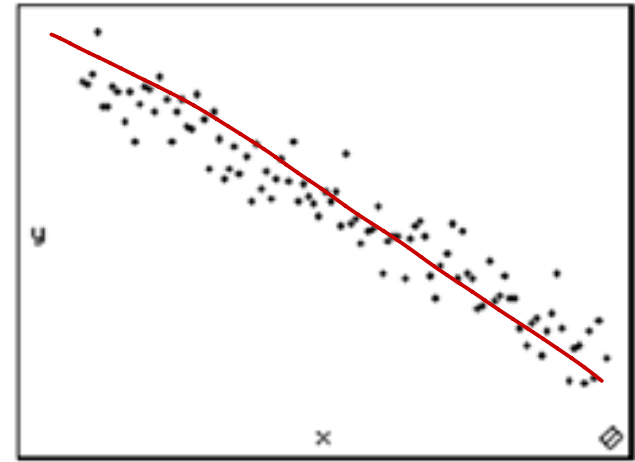
(b) Negative correlation:

$r = -0.965$

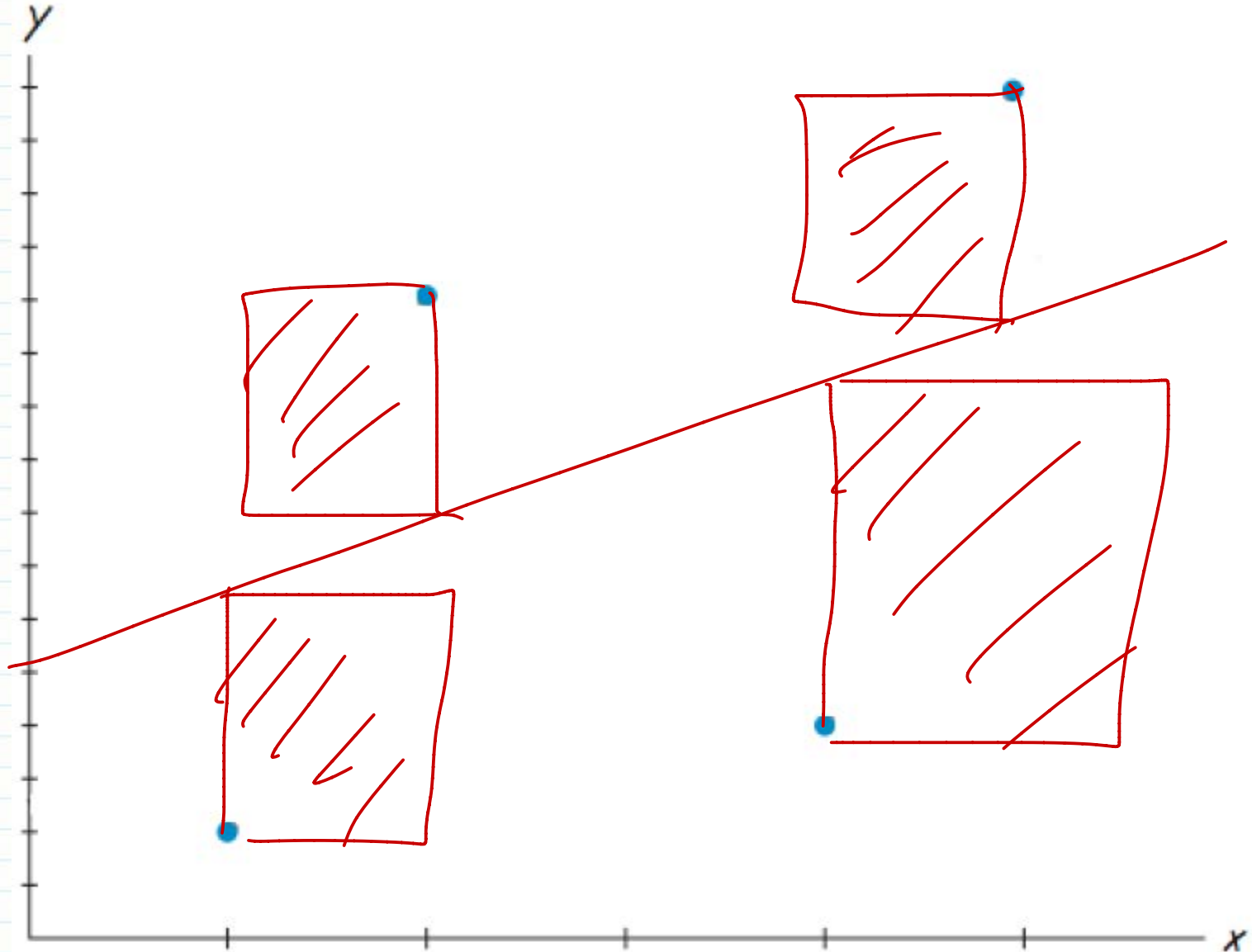**(c) No correlation:** $r = 0$

(d) Nonlinear relationship: $r = -0.087$

Terminology

- Line of best fit

- Regression line

- Least Squares line

- Regression equation

Goal: Find the line which minimizes area

Linear Correlation Coefficient -- r

- If r is close to 1, then there is a positive linear correlation.
- If r is close to -1, there is a negative linear correlation.
- If r is close to 0, there is no linear correlation.

- r^2 is the proportion of variation in the variables which can be explained by the correlation.

P-Values

- H_0 is that r=0 (there is no linear correlation)
- H_1 is that r≠0 (there is linear correlation)

- If P<0.05, reject H_0 – sample data indicates linear correlation

- If P>0.05, do not reject H_0 – sample data does not indicate linear correlation

Fall Track Testing

400m	30m	Fly 30m	Vert	St LJ	St TJ	Push ups	Sit ups	Triple Bound	Total Pts	Weight	Height
78.30	4.75	4.45	16	77	203	43	48	233	321	145.0	63.25
65.90	4.45	4.06	18	81.5	235.5	38	41	230	393	133.4	68
66.50	4.59	4.17	18.5	84	242	46	58	243	414	140.4	64.25
69.50	4.43	4.04	20	88	240	40	42	246	409	132.2	66.75
64.60	4.25	4.02	20.5	92	242	42	56	262	451	132.2	65.5
68.90	4.36	4.00	18	79	236	50	60	242	425	137.6	62.5
65.40	4.28	3.92	22	91.5	243.5	42	44	261	461	151.0	71.5
72.80	4.58	4.23	17	75	220.5	38	41	233	338	123.0	68.25
61.20	4.24	3.81	18	73	217	42	42	222	405	139.2	65.5
67.80	4.26	3.89	21.5	85	252.5	54	55	266	463	140.2	67
78.30	4.59	4.20	17.5	71.75	227	40	67	234	356	140.2	64.25
66.70	4.39	4.02	20	83	244	40	46	253	414	127.0	66.25
65.70	4.26	3.95	21	78.5	246	42	50	236	423	135.8	66.5
67.80	4.48	4.11	17.5	77	221	54	46	248	400	141.6	65.75
64.40	4.36	3.93	23	81	240	50	60	255	466	164.6	67
64.90	4.37	3.88	21	94.5	264.75	49	61	254	477	111.8	63.5
66.90	4.27	3.92	18.5	83	250	36	33	245	397	114.8	63.5
73.40	4.69	4.04	17	76	213.5	54	60	236	357	113.2	57.75
67.20	4.35	4.04	20	84.5	243	38	51	256	420	141.8	67
62.80	4.15	3.76	23.5	92	252.5	48	57	284	496	118.2	66.25

Is there a linear correlation between Vertical Jump and Standing Long Jump?

Vert	St LJ
16	77
18	81.5
18.5	84
20	88
20.5	92
18	79
22	91.5
17	75
18	73
21.5	85
17.5	71.75
20	83
21	78.5
17.5	77
23	81
21	94.5
18.5	83
17	76
20	84.5
23.5	92

H_0: $r=0$ No correlatio

H_1: $r \neq 0$ Correlation

TI: LinReg TTEST

Online: correlation

$P = .0004$

$P < .05$

Reject H_0 / Support H_1

The sample evidence indicates a linear correlation.

Using the Regression Equation

- Use the linear correlation between Vertical jump and Standing Long Jump to estimate the Standing Long Jump of an athlete with a vertical jump of 24 inches.

$X = $ vertical leap

$Y = $ St. LJ

$Y = 2.22x + 39.22$

$R^2 = .5029$

$Y = 2.22 \times 24 + 39.22$

$= 92.5$ inches

$y = mx + b$

$m = 2.220795$

$b = 39.223552$

$r = 0.709182$

$t = 4.267648$

$P = 0.000463$

Which is a better predictor of an athlete's 400m time, height or the number of sit-ups she can do in one minute?

Correlation b/w 400 & Situps

$r = .179$

Correlation b/w 400 & height

$r =$

400m	Sit ups	Height
78.30	48	63.25
65.90	41	68
66.50	58	64.25
69.50	42	66.75
64.60	56	65.5
68.90	60	62.5
65.40	44	71.5
72.80	41	68.25
61.20	42	65.5
67.80	55	67
78.30	67	64.25
66.70	46	66.25
65.70	50	66.5
67.80	46	65.75
64.40	60	67
64.90	61	63.5
66.90	33	63.5
73.40	60	57.75
67.20	51	67
62.80	57	66.25

Correlation with 400m time

	R	R^2
30m	0.82	0.67
Fly 30m	0.81	0.66
Vert	-0.65	0.42
St TJ	-0.59	0.35
St LJ	-0.52	0.27
Triple Bound	-0.41	0.16
Height	-0.38	0.14
Sit ups	0.18	0.03
Push ups	-0.06	0.00
Weight	-0.01	0.00

Where does r come from?