Review for Quiz 3

1. a) Find the area between -1 and 1 under the normal curve.

b) Find the area to the left of 1.5 under the normal curve.

c) Find the area between 1.5 and 2.5 under the normal curve.

2. Women age 18-24 have average height of 64.3 inches, with an SD of 2.6 inches, and their heights follow the normal curve.

a) What does it mean to say that their heights follow the normal curve?

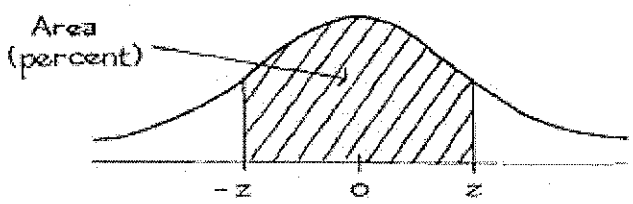
b) Estimate the percentage of these women who are between 62 inches and 66 inches in height.

c) Estimate the percentage of these women who are taller than 70 inches.

d) What height represents the 75-th percentile?

3. Suppose the GPA's for a group of students has an average of 3.5 and SD equal to 0.5. Do these GPA's follow the normal curve?

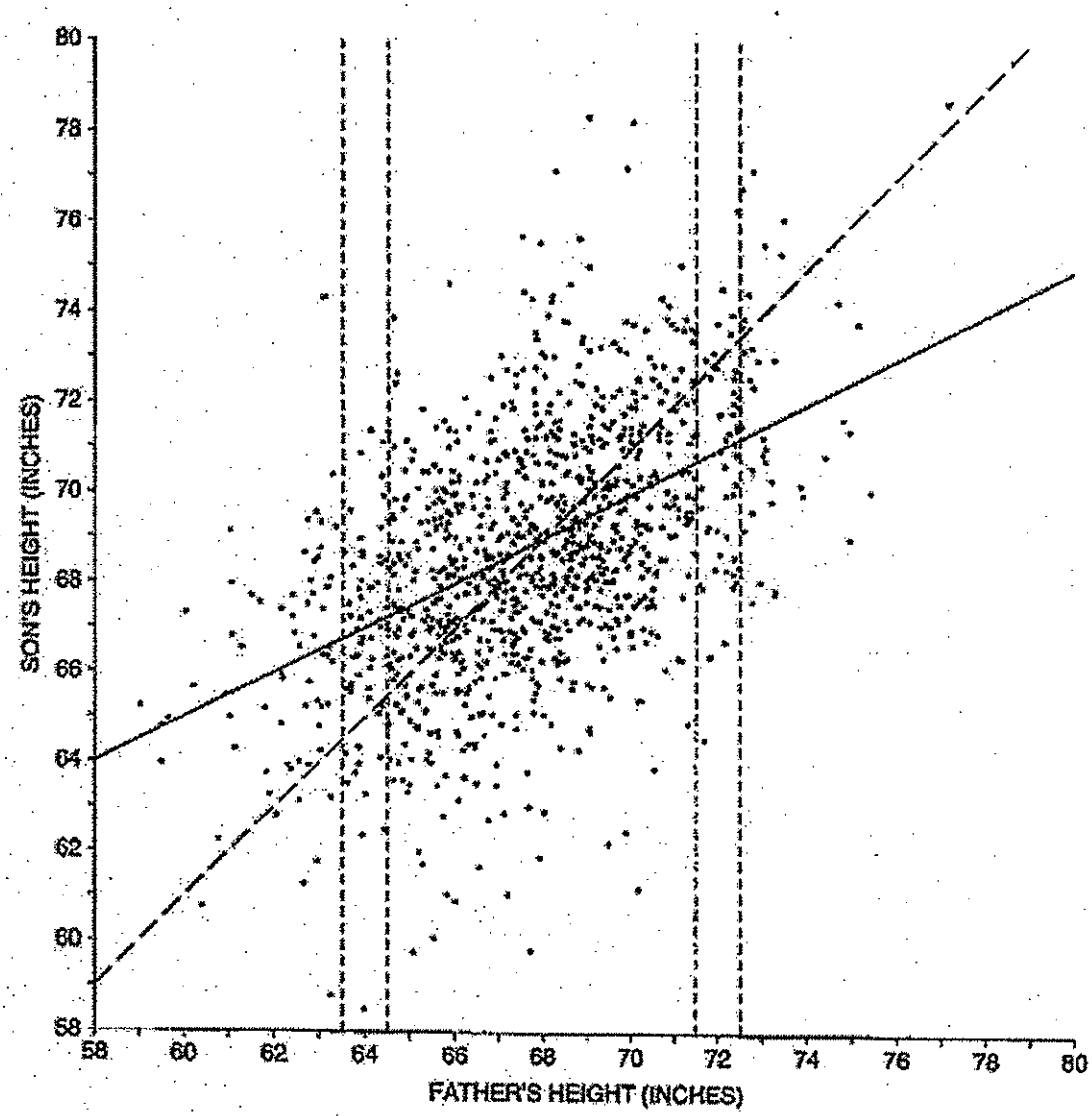
A NORMAL TABLE



z	*Area*	*z*	*Area*	*z*	*Area*
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	99.9991

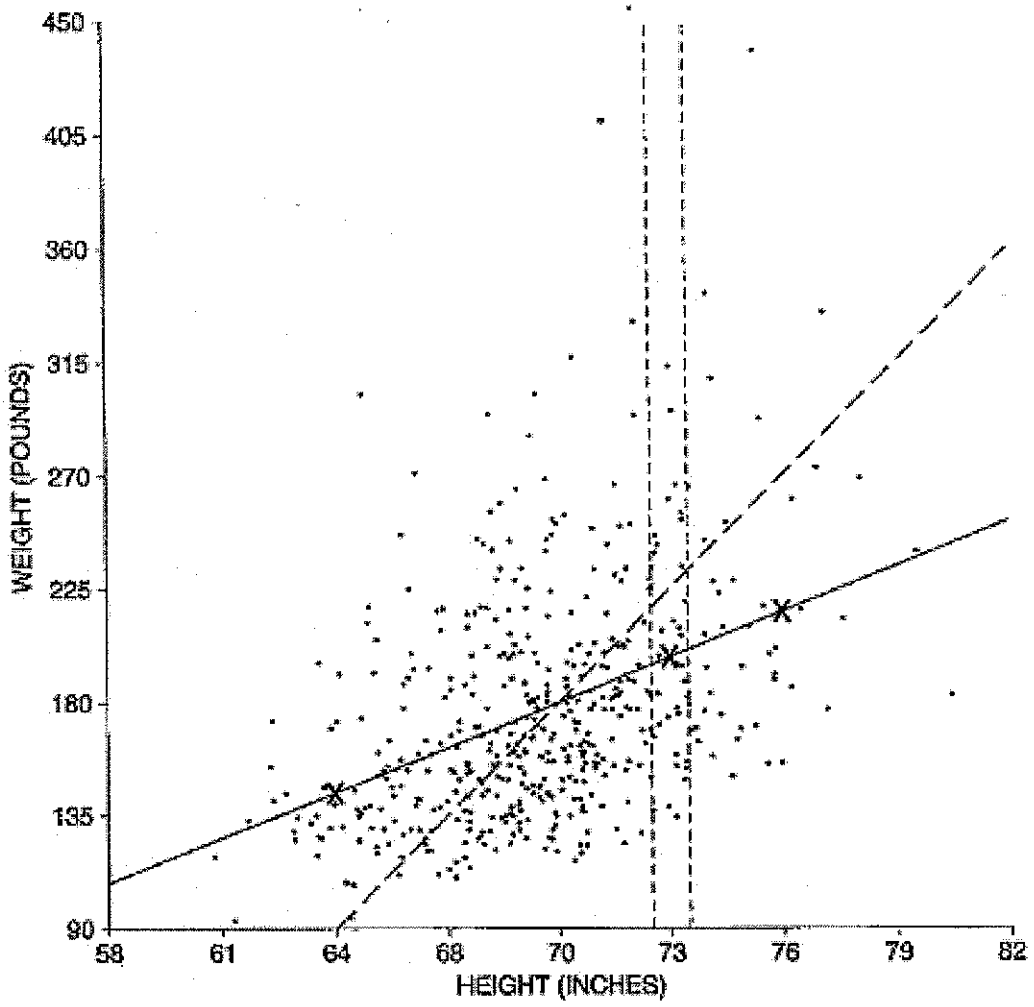
1,078 Pairs of Fathers and Sons

Average height of fathers $\approx$ 68 inches, SD $\approx$ 2.7
Average height of sons $\approx$ 69 inches, SD $\approx$ 2.7
$r \approx 0.5$



Heights and Weights for 471 Men (HANES 5)

Average height $\approx$ 70 inches, SD $\approx$ 3 inches
Average weight $\approx$ 180 pounds, SD $\approx$ 45 pounds
$r \approx 0.40$



Goal: Analyze the association or correlation between two variables
Why? Prediction, Understanding, Control

The correlation coefficient is a measure of linear association between two variables.

<u>Computing the correlation coefficient:</u>

1. **Convert the x - values to standard units.**

2. **Convert the y - values to standard units.**

3. **Multiply each x - value (in standard units) by each y - value (in standard units) .**

4. **The average of these products is equal to r .**

r = average of the products $\left(\frac{x-AV_x}{SD_x}\right) \cdot \left(\frac{y-AV_y}{SD_y}\right)$

$$-1 \leq r \leq 1$$

Chapter 9: More about Correlation

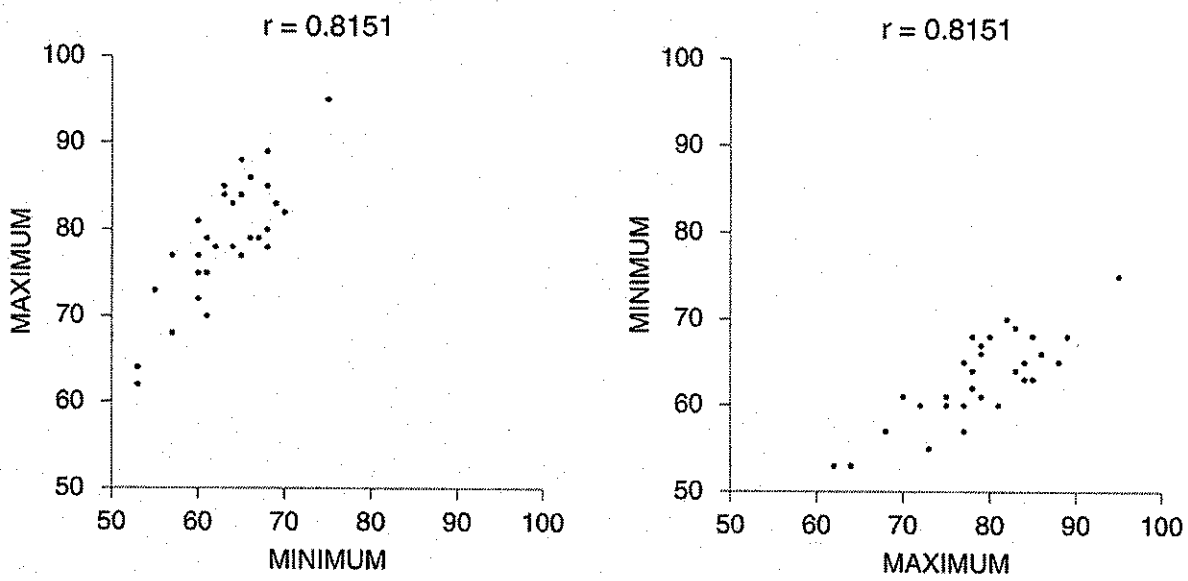
Some facts:

- r is a pure number (no units)

- r does not change if you
  - switch x and y
  - add the same number to each x value
  - add the same number to each y value
  - multiply each x value by a positive number
  - multiply each y value by a positive number

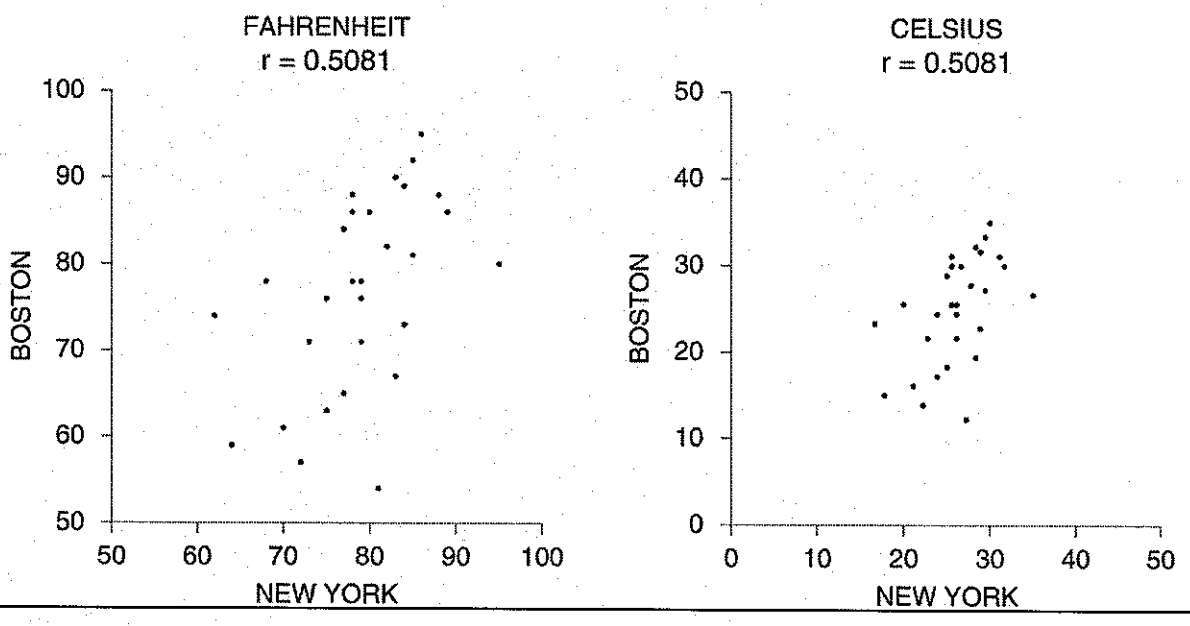
e.g. r is not changed if we switch x and y

Figure 2. Daily temperatures. New York, June 2005.



e.g. the correlation between temperatures in Fahrenheit is the same as it would be if the temperatures were measured in Celsius, because Celsius = (5/9) (Fahrenheit – 32)

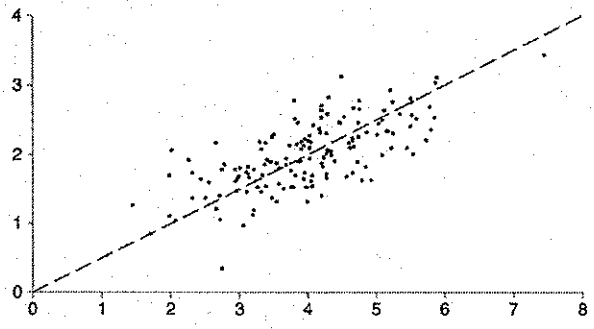
Figure 1. Daily maximum temperatures. New York and Boston, June 2005. The left hand panel plots the data in degrees Fahrenheit; the right hand panel, in degrees Celsius. This does not change *r*.



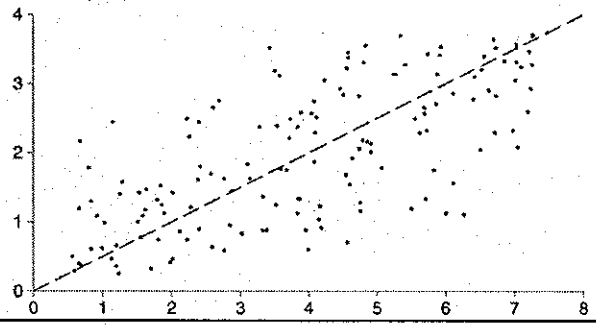
$F = \frac{9}{5} C + 32$ $\quad C = (F - 32)\frac{5}{9}$

The correlation looks stronger if the SDs are smaller:

Figure 3. The effect of changing SDs. The two scatter diagrams have the same correlation coefficient of 0.70. The top diagram looks more tightly clustered around the SD line because its SDs are smaller.



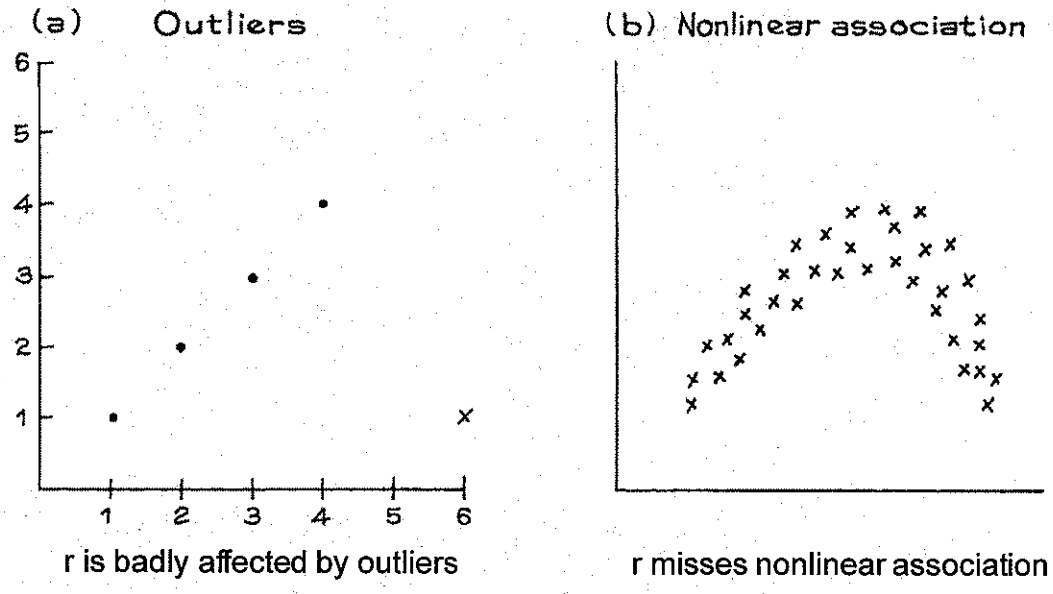
r = 0.7



r = 0.7

r measures association about a LINE:

Figure 5. The correlation coefficient can be misleading in the presence of outliers or non-linear association.



Show on Scatterplot

Ecological Correlations

Sometimes, each point on the plot represents the average or rate for a whole group of individuals.

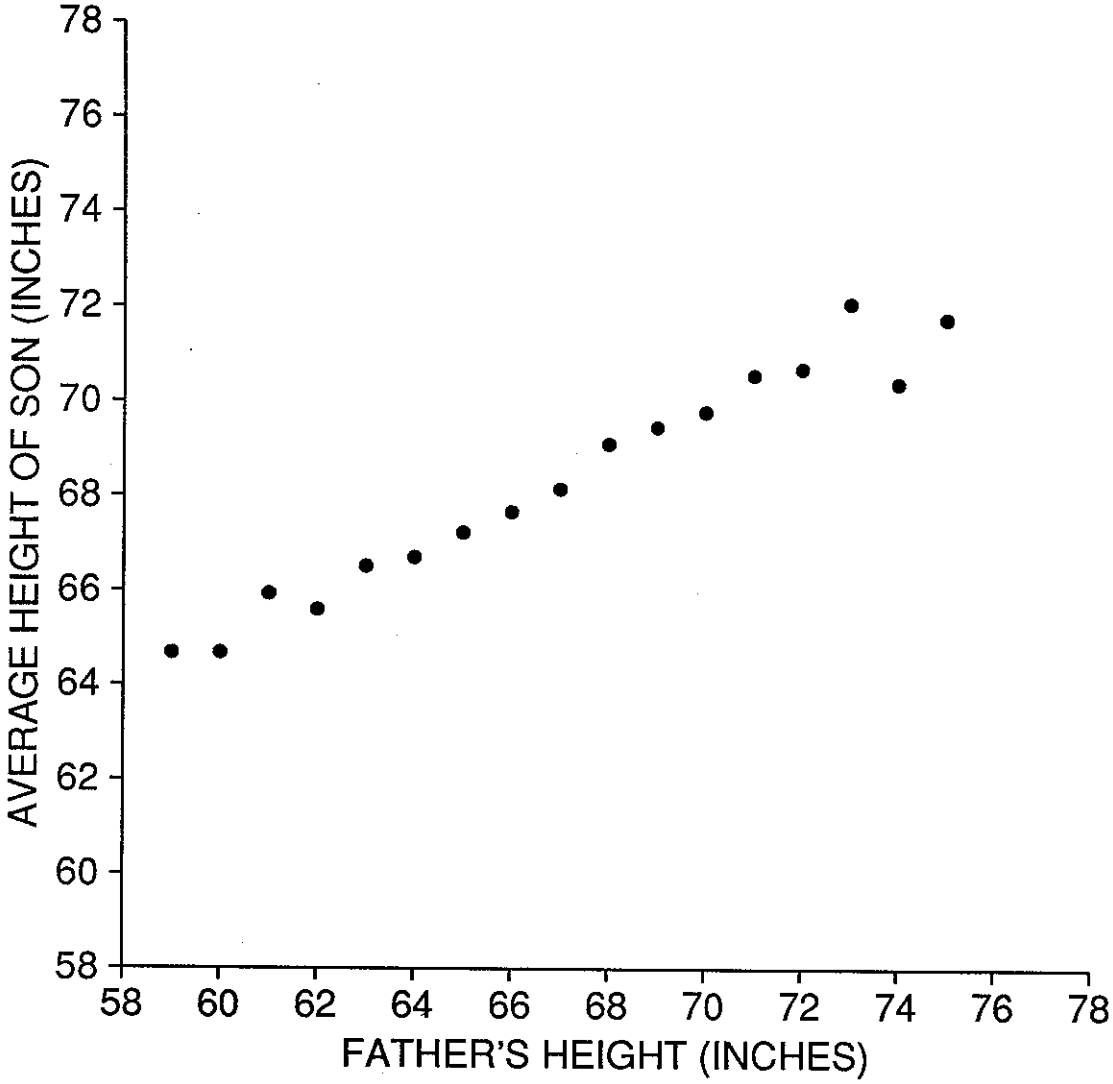
E.g. In various countries, find the death rate from cancer versus rate of cigarette smoking. Plot each country on the scatter diagram.

The correlation from a plot of averages or rates is called an ECOLOGICAL CORRELATION.

Ecological Correlations are artificially strong!

ECOLOGICAL CORRELATIONS

(FATHERS AND SONS HEIGHTS)



Ecological Correlations

Section 1:

Midterm	Final
20	89
77	80
95	47
AV = 64	**AV = 72**

Section 2:

Midterm	Final
26	45
82	90
93	87
AV = 67	**AV = 74**

Section 3:

Midterm	Final
70	96
80	60
66	84
AV = 72	**AV = 80**

The correlation for all nine midterm and final test scores is $r = .05$ **.**

When using just averages, the correlation is $r = .99$

Association is not causation!

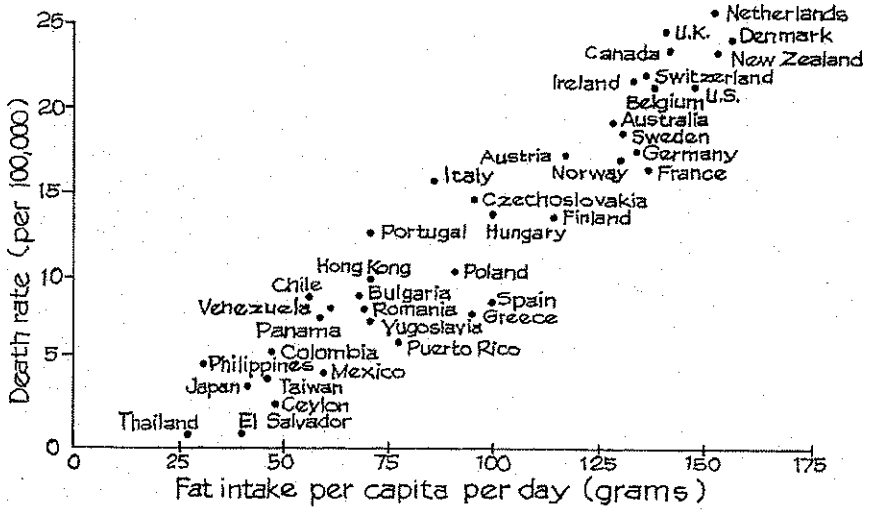
- In children, shoe size is positively associated with reading ability

- Looking weekly, chocolate consumption is positively associated with car accidents

- Number of years of education is positively associated with income

- Scores on tests are negatively correlated with hours spent doing homework

Figure 8. Death rates from breast cancer plotted against fat in the diet, for a sample of countries.

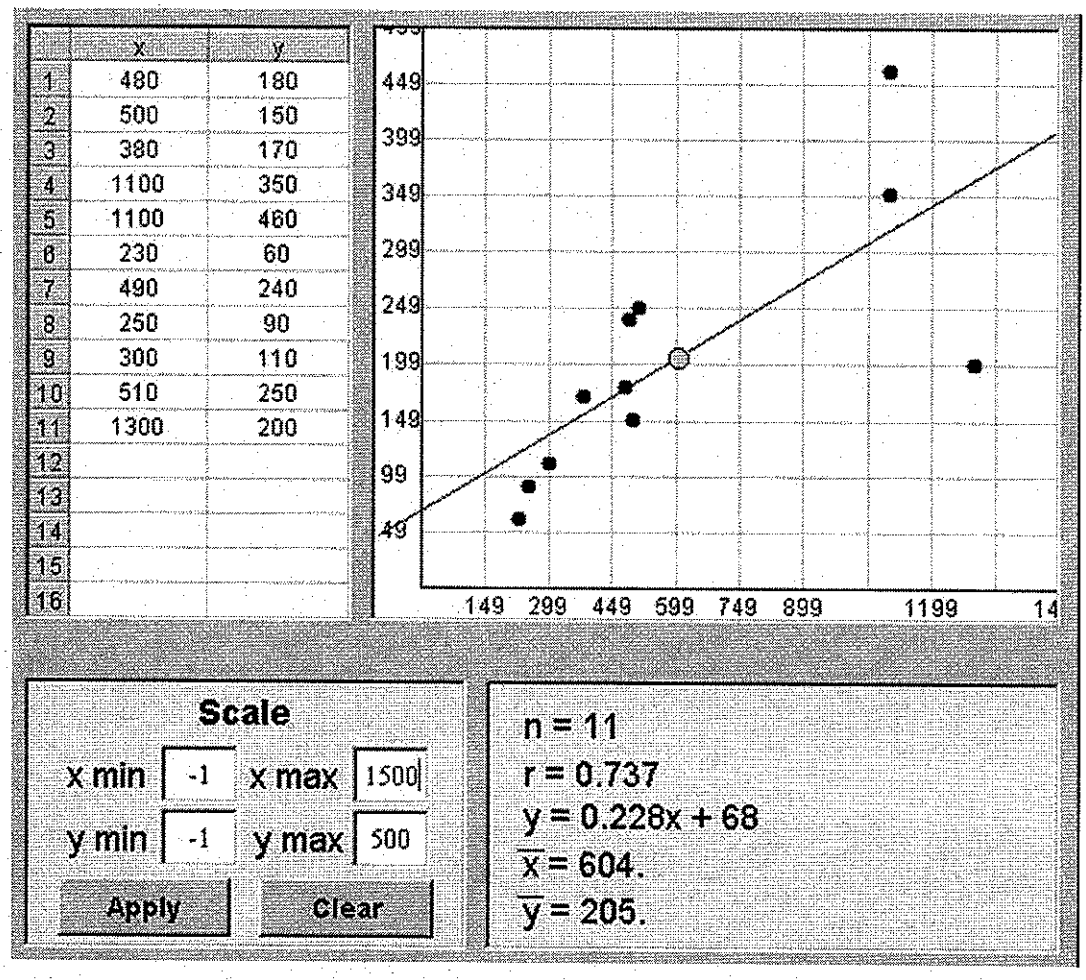


Note: Age standardized.

Source: K. Carroll, "Experimental evidence of dietary factors and hormone-dependent cancers," *Cancer Research* vol. 35 (1975) p. 3379. Copyright by *Cancer Research*. Reproduced by permission.

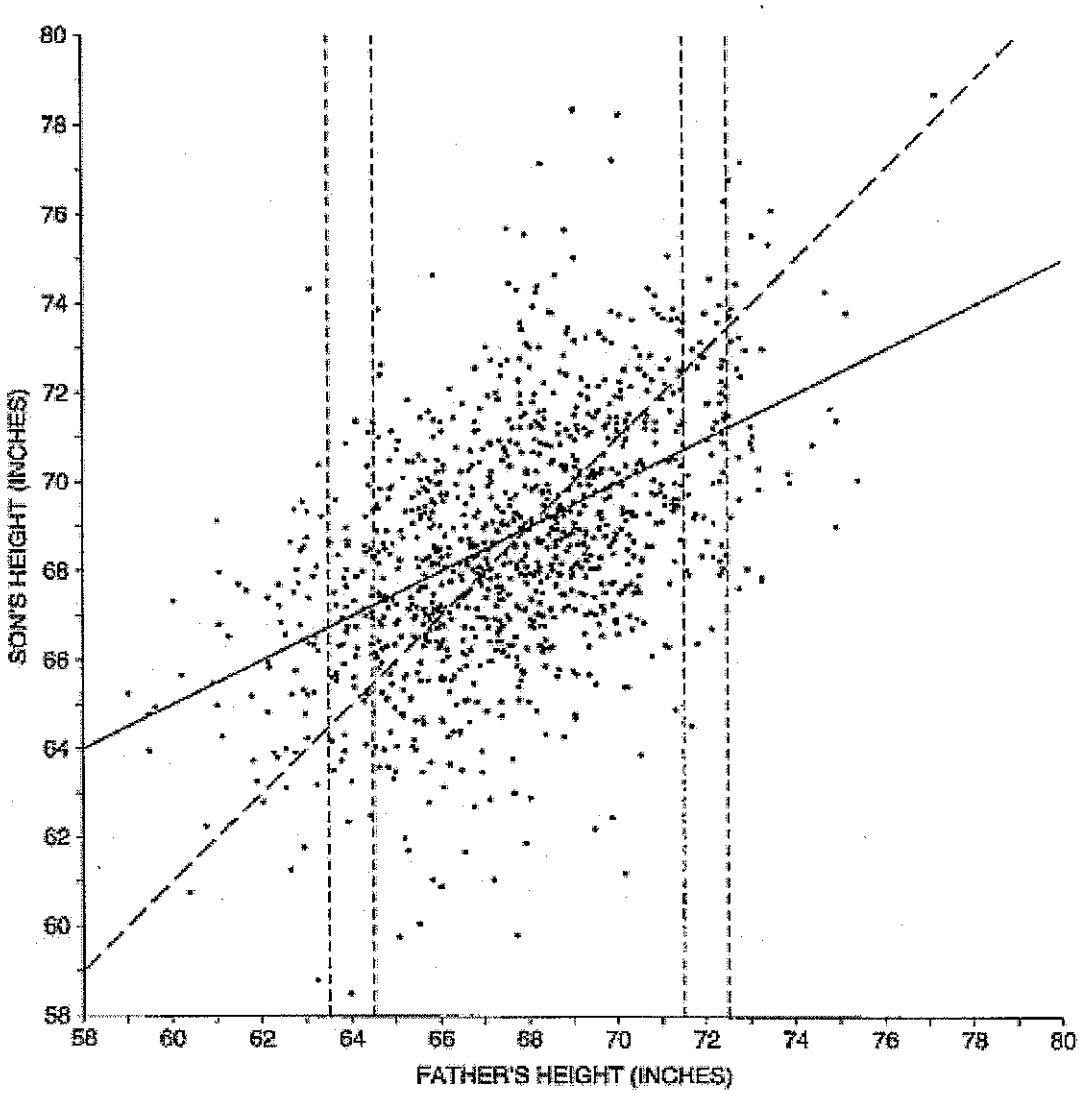
The table and scatter-diagram shows per capita consumption of cigarettes in various countries in 1930, and the death rates from lung cancer for men in 1950.

Country	*Cigarette consumption*	*Deaths per million*
Australia	480	180
Canada	500	150
Denmark	380	170
Finland	1,100	350
Great Britain	1,100	460
Iceland	230	60
Netherlands	490	240
Norway	250	90
Sweden	300	110
Switzerland	510	250
U.S.	1,300	200



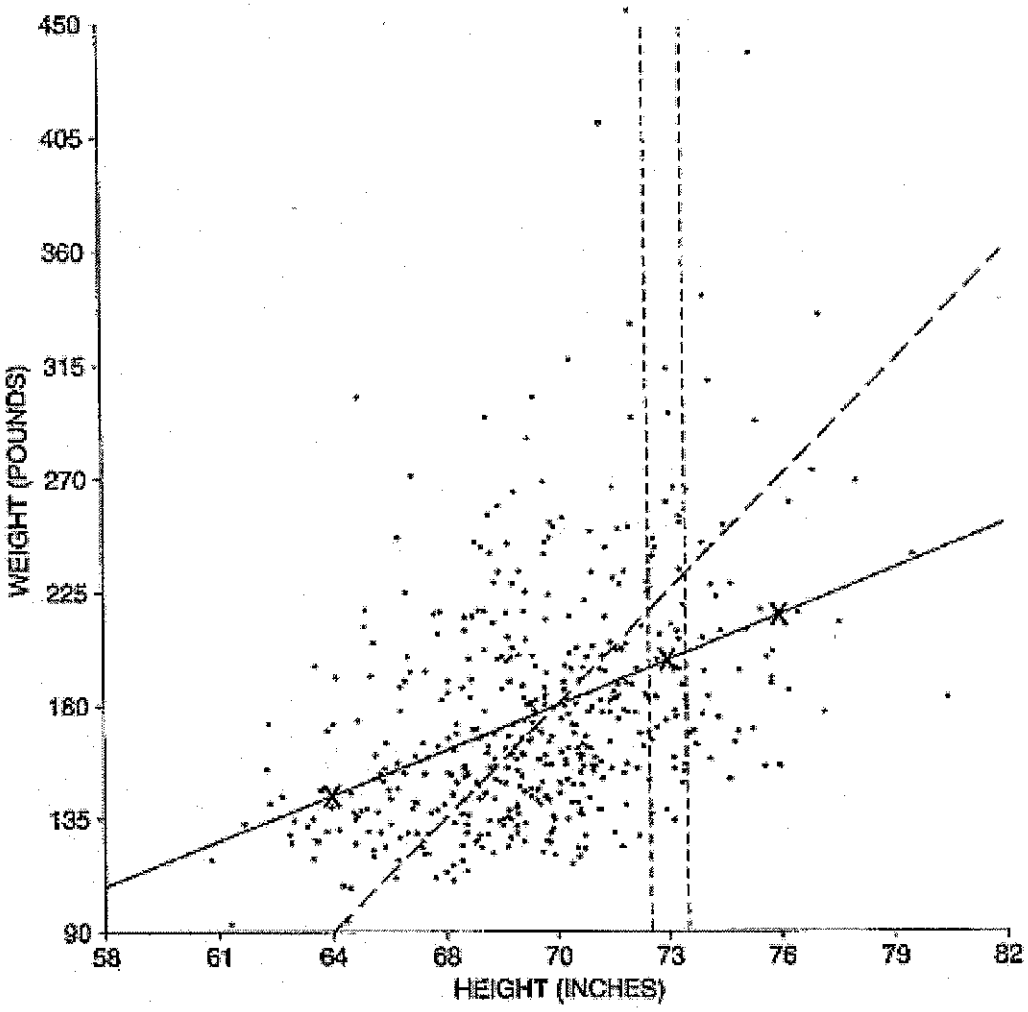
1,078 Pairs of Fathers and Sons

Average height of fathers $\approx$ 68 inches, SD $\approx$ 2.7
Average height of sons $\approx$ 69 inches, SD $\approx$ 2.7
$r \approx 0.5$



Heights and Weights for 471 Men (HANES 5)

Average height $\approx$ 70 inches, SD $\approx$ 3 inches
Average weight $\approx$ 180 pounds, SD $\approx$ 45 pounds
$r \approx 0.40$



The SD line goes through the middle of the football.

The SD line:

- goes through the point of averages (ave_X,ave_Y)
- with

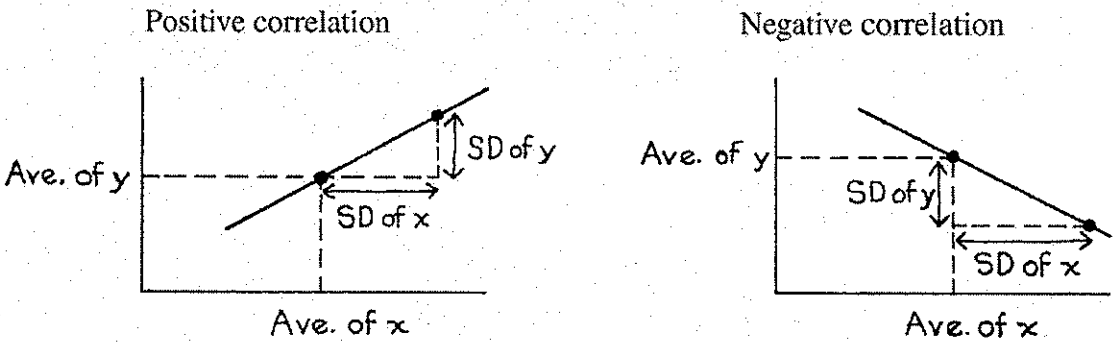
$$\text{slope} = \frac{SD_Y}{SD_X} \quad \text{if r is positive}$$

$$\text{slope} = -\frac{SD_Y}{SD_X} \quad \text{if r is negative}$$

To draw the SD line:

- go to the point of averages and put a dot
- go across SD_X and up SD_Y, put another dot
- join the dots

Figure 8. Plotting the SD line.



Someone who is RIGHT ON the SD line is the same number of SDs above average in y as they are in x.

Example: midterm ave = 75, SD = 10, r = 0.7

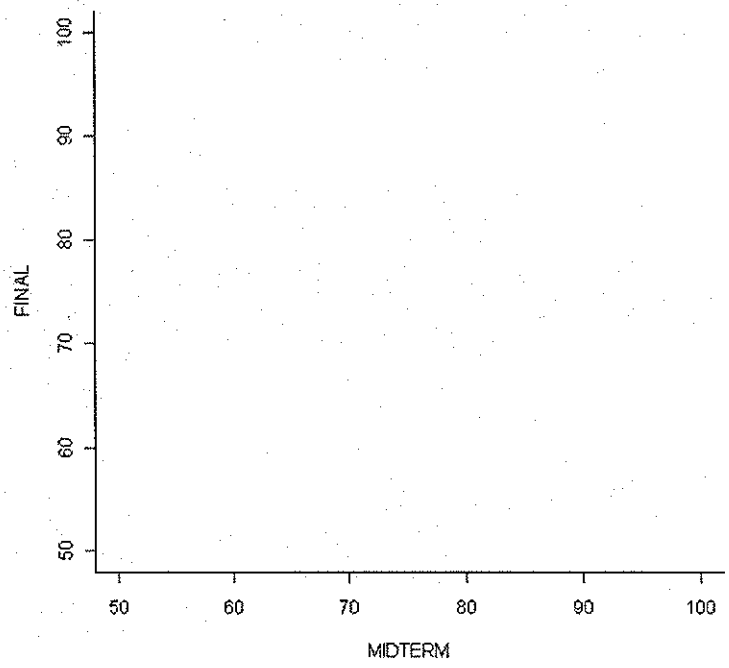
final ave = 70, SD = 12

- Someone who got 85 on the midterm would get 82 on the final if they were on the SD line.
- Someone who got 95 on the midterm would get 94 on the final if they were on the SD line.
- Someone who got 55 on the midterm would get 46 on the final if they were on the SD line.
- Someone who got 60 on the midterm would get 52 on the final if they were on the SD line.

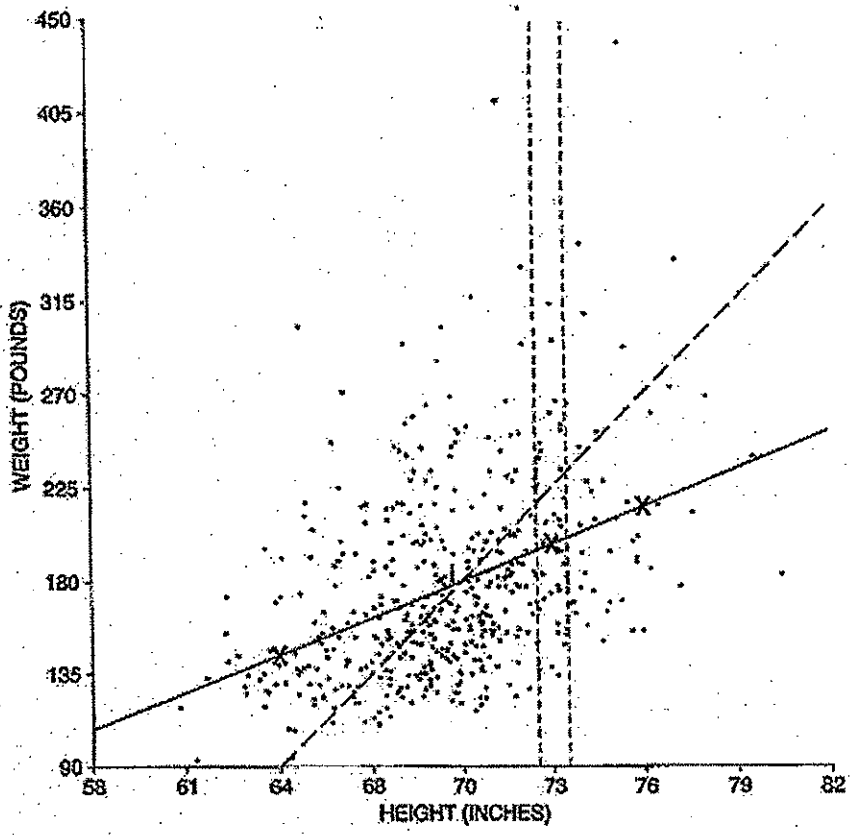
Midterm: ave = 65 SD = 16 r = 0.7

Final: ave = 60 SD=10

Draw the SD line



REGRESSION



The regression method describes how one variable is linearly related to another variable.

The regression line for y on x estimates the average of the y-values for a corresponding x value.