9-1 Paired Data

- is there a relationship
- if so, what is the equation
- use the equation for prediction

Apr 26-8:27 AM

9-2 Correlation

- **Correlation**

exists between two variables when one of them is related to the other in some way

Apr 26-8:29 AM

Positive Linear Correlation

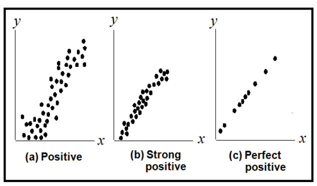
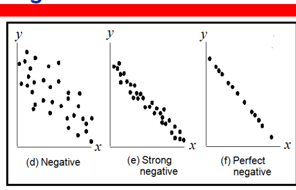


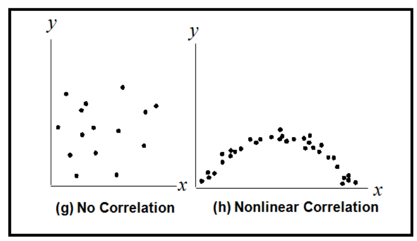
Figure 9-1 Scatter Plots

Negative Linear Correlation



Apr 26-8:30 AM

No Linear Correlation



Apr 26-8:32 AM

Ex1) Is there a correlation between the bill amount and the tip you leave? Make a scatterplot for the following data on page 505.

L1	L2
33.46	5.5
50.68	5
87.92	8.08
98.84	17
63.6	12
107.34	16

- enter data in L_1 and L_2
- 2nd y =
- on twice
- select scatterplot for L_1 and L_2
- Zoom
- #9
- Enter

Apr 30-1:37 PM

Linear Correlation Coefficient r

measures **strength** of the linear relationship between paired x and y values in a **sample**

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

Rounding the Linear Correlation Coefficient r

- Round to **three** decimal places so that it can be compared to critical values in Table A-6
- Use calculator or computer if possible

Apr 26-8:32 AM

Interpreting the Linear Correlation Coefficient

- If the absolute value of r exceeds the value in Table A - 6, conclude that there is a significant linear correlation.
- Otherwise, there is not sufficient evidence to support the conclusion of significant linear correlation.

TABLE A-6 Critical Values of the Pearson Correlation Coefficient r

n	$\alpha = .05$	$\alpha = .01$
4	.950	.999
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
25	.396	.505
30	.361	.463
35	.335	.430
40	.312	.402
45	.294	.378
50	.279	.361
60	.254	.330
70	.236	.305
80	.220	.286
90	.207	.269
100	.196	.256

Apr 26-8:35 AM

Properties of the Linear Correlation Coefficient r

1. $-1 \le r \le 1$
2. Value of r does not change if all values of either variable are converted to a different scale.
3. The r is not affected by the choice of x and y. Interchange x and y and the value of r will not change.
4. r measures strength of a linear relationship.

Apr 26-8:36 AM

Common Errors Involving Correlation

1. Causation: It is wrong to conclude that correlation implies causality.
2. Averages: Averages suppress individual variation and may inflate the correlation coefficient.
3. Linearity: There may be some relationship between x and y even when there is no significant linear correlation.

Apr 26-8:37 AM

Formal Hypothesis Test

- To determine whether there is a significant linear correlation between two variables
- Two methods
- Both methods let $H_0: \rho = 0$ (no significant linear correlation)
  $H_1: \rho \ne 0$ (significant linear correlation)

Method 1: Test Statistic is t

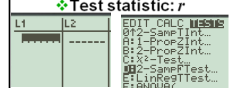
$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

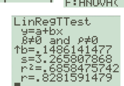
Critical values:

use Table A-3 with degrees of freedom = n - 2

Method 2: Test Statistic is r (uses fewer calculations)

- Test statistic: r





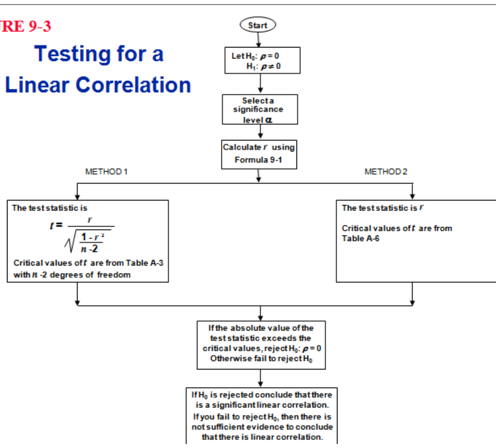
- Critical values:

Refer to Table A-6 (no degrees of freedom)

Apr 26-8:37 AM

FIGURE 9-3

Testing for a Linear Correlation

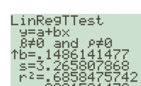


Apr 26-8:43 AM

Ex1b)Is there a correlation between the bill amount and the tip you leave? Use $\alpha = .05$

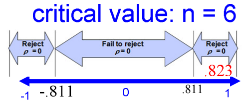
Hypothesis: $H_0: \rho = 0$ (no significant linear correlation)
$H_1: \rho \neq 0$ (significant linear correlation)

test statistic:



critical value: n = 6

n	$\alpha = .05$	$\alpha = .01$
4	.950	.999
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798



Conclusion: reject H_0. There is a linear relationship.

Apr 27-2:08 PM

Is there a significant linear correlation?

Data from the Garbage Project

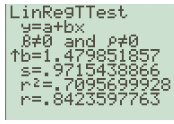
x Plastic (lb)	0.27	1.41	2.19	2.83	2.19	1.81	0.85	3.05
y Household	2	3	3	6	4	2	1	5

$n = 8$ $\alpha = 0.05$ $H_0: \rho = 0$

$H_1: \rho \neq 0$

use calculator - linear regression t test

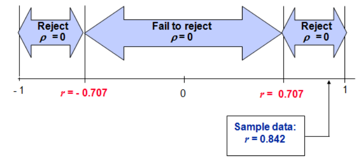
Test statistic is $r = 0.842$



Apr 26-8:44 AM

n	$\alpha = .05$	$\alpha = .01$
4	.950	.999
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765

Critical values are r = - 0.707 and 0.707
(Table A-6 with n = 8 and α = 0.05)



0.842 > 0.707, That is the test statistic does fall within the critical region.

Therefore, we REJECT H_0: $\rho = 0$ (no correlation) and conclude there is a significant linear correlation between the weights of discarded plastic and household size.

Apr 27-3:31 PM