

Section 4.1

Scatter Diagrams and Correlation

Definitions

- The **Response Variable** is the variable whose value can be explained by the value of the **explanatory** or **predictor variable**.

Scatter Diagram

- A graph that shows the relationship between two quantitative variables measured on the same individual. Each individual in the data set is represented by a point in the scatter diagram. The explanatory variable is plotted on the horizontal axis and the response variable is plotted on the vertical axis.

Finding Scatter Diagram

1. Put x values (explanatory variable) into L1
2. Put y values (response variable) into L2
3. “2nd” button, “y=” button
4. “Enter” on 1: Plot1
5. Choose “On”, “1st type”, L1, L2, “1st mark”
6. “Zoom” -> ZoomStat

1. Find the scatter diagram for the following data

X	Y
1	20
2	50
3	60
4	65

Positively vs. Negatively Associated

- Positively Associated = As the x value increases, the y value increases
- Negatively Associated = As the x value increases, the y value decreases

where x = explanatory variable, y = response variable

Sample Linear Correlation Coefficient (r) or Pearson Product Moment Correlation Coefficient

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{n - 1}$$

2. Find the linear correlation coefficient (by hand)

X	Y
1	10
2	15
8	35
13	44

Sample Linear Correlation Coefficient
(r) or Pearson Product Moment
Correlation Coefficient (shortcut
formula)

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right)} \sqrt{\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n} \right)}}$$

Linear Correlation Coefficient

- Aka (Pearson Product Moment Correlation Coefficient) = a measure of the strength of the linear relation between two variables.
- Represented by r
- Between -1 and 1 (including -1 and 1)
- -1 represents perfect negative correlation, 1 represents perfect positive correlation

Finding r

1. Put x values (explanatory variable) into L1
2. Put y values (response variable) into L2
3. “Stat” button
4. Right arrow to CALC
5. Down arrow to LinReg ($ax + b$)
6. “enter” button
7. “enter” button

* Make sure Diagnostics is On

3. Find the linear correlation coefficient (by TI-83/84)

X	Y
5	50
9	27
11	15
15	3

Testing for a Linear Relation

1. Determine the absolute value of the correlation coefficient $|r|$
2. Find the critical value (CV) in Table II from Appendix A for the given sample size
3. if $|r| > CV$: linear relation exists
if $|r| \leq CV$: no linear relation exists

TABLE II**Critical Values for Correlation Coefficient**

<i>n</i>	
3	0.997
4	0.950
5	0.878
6	0.811
7	0.754
8	0.707
9	0.666
10	0.632
11	0.602
12	0.576
13	0.553
14	0.532
15	0.514
16	0.497
17	0.482

4. Test to see if there is a linear relation between x and y

X	Y
1	33
2	42
3	57
4	62
5	33

Correlation versus Causation

Note: A linear correlation coefficient that implies a strong positive or negative association does not imply causation if it was computed using observational data