Class Worksheet

February 14 and 19 Math 125 Kovitz 2025

Chapter 8

Correlation and Regression

Correlation: a measure of the relationship between two variables.

The Scatter Diagram

If there is a strong association between two variables, then knowing one helps a lot in predicting the other. But when there is a weak association, information about one variable does not help much in guessing the other.

The Correlation Coefficient

The correlation coefficient is a measure of linear association or clustering around a line. The relationship between two variables can be summarized by:

- The average of the *x*-values, the SD of the *x*-values
- the average of the *y*-values, the SD of the *y*-values
- the correlation coefficient r.

Correlations are always between -1 and 1, but can take any value in between. A positive correlation means that the cloud slopes up; as one variable increases, so does the other. A negative correlation means that the cloud slopes down; as one variable increases, the other decreases.

The SD Line: the line around which the points cluster if r is close to ± 1 .

Computing the Correlation Coefficient

- Convert the x- and y-values to standard units by finding the average and SD for both the x-values and the y-values and then subtracting the average from each value and dividing by the SD. Put the results in two new columns of the table of data.
- For each row multiply $(x \text{ in standard units}) \times (y \text{ in standard units})$. Put the resulting products in the last column of the table.
- Take the average of the products. That average is r.

Chapter 9

The Correlation Coefficient (continued)

The correlation coefficient is a pure number, without units. It is not affected by:

- interchanging the two variables.
- adding the same number to all the values of one variable.
- multiplying all the values of one variable by the same positive number.

As r gets close to 1, the distance of a typical point above or below the SD line becomes a small fraction of the vertical SD. That means that r measures clustering relative to the SDs.

Whenever possible, look at the scatter diagram to check for outliers and nonlinear association: r measures linear association, not association in general.

Ecological correlations are based on rates or averages. They are often used in political science and sociology. And they tend to overstate the strength of an association. So watch out.

Correlation measures association. But association is not the same as causation.

Formulas

The slope of the SD line:

When r is positive When r is positive

$$(SD \text{ of } y)/(SD \text{ of } x)$$

When r is negative

$$-(\text{SD of } y)/(\text{SD of } x)$$

To calculate the correlation coefficient, convert each variable to standard units, and then take the average product.

Alternative formula for the correlation coefficient:

$$r = \frac{\operatorname{cov}(x, y)}{(\operatorname{SD of} x) \times (\operatorname{SD of} y)}$$

where

$$cov(x, y) = (average of products xy) - (ave of x) \times (ave of y).$$

Problems to think about for Chapters 8 and 9

Find the correlation coefficient for the following data set. (First draw the scatter diagram and estimate r.)

x	y
2	26
13	24
14	22
16	23
20	20

A survey of weights and heights of male college students included five men of Albanian origin. Find the correlation coefficient for that data set, listed below.

{60 in., 107 lb.; 66 in., 137 lb.; 68 in., 152 lb.; 69 in., 197 lb.; 72 in., 167 lb.}

Which tells you more about the relationship between two variables: the scatter diagram or the correlation coefficient? Explain briefly.

Suppose that the correlation between weight (in pounds) and years of schooling completed is about -0.10. True or false and explain:

- (a) Heavier persons tend to be more educated.
- (b) Persons with more education tend to weigh less.
- (c) The correlation between years of schooling completed and weight (in pounds) is about -0.10.
- (d) The correlation between wt. (in kilograms) and years of schooling completed will not be about -0.10.
- (e) Lighter persons tend to be more educated.
- (f) If you eat and put on 25 pounds, you will become less educated.