

Lecture 05

Scatterplots:

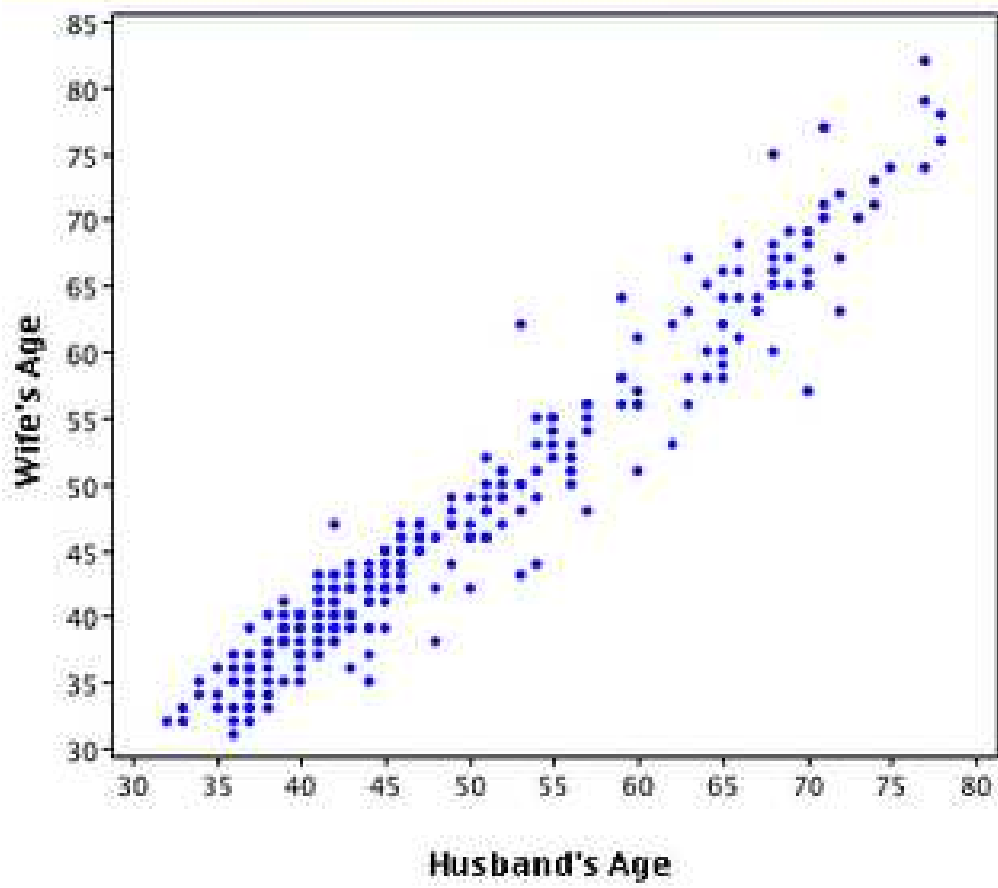
- a graphical descriptive statistic
- for paired quantitative data $(x_1, y_1), \dots, (x_n, y_n)$
- always label axes and provide a title
- focus is on the relationship between x and y
- scatterplots aid in prediction
- interpolation versus extrapolation

Examples: data appropriate for a scatterplot?

(a) Consider 20 patients who take drug 1 and we record their blood pressure (x 's). There are 20 other patients who take drug 2 and we record their blood pressure (y 's).

(b) Consider the monthly immigration rates (x 's) into British Columbia and the monthly emigration rates from British Columbia (y 's).

(c) We consider 10 different colours. In a neighbourhood, we count the number of houses of each colour.



Sample correlation coefficient r :

- a numerical descriptive statistic
- for paired quantitative data $(x_1, y_1) \dots, (x_n, y_n)$
- $r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$
- r describes linearity between x and y

Association versus cause-effect:

- correlation does not imply causation
- the role of lurking variables in causation
- observational studies
- randomized experiments

Example for discussion: “Prayer can Lower Blood Pressure”, USA Today, August 11, 1998.

People who attended a religious service once a week and prayed or studied the Bible were 40% less likely to have high blood pressure.