Topic: Exploratory Data Analysis (EDA)

Presentation of Bivariate Data

Part C: Two quantitative variables

School of Mathematics and Applied Statistics

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Bivariate data: Two Variables

Different tables / plots for different data types . . .

For two qualitative variables:

- two-way tables
- stacked bar graphs
- clustered bar graphs

For one quantitative and one qualitative variable:

- side-by-side box plots
- back-to-back stem & leaf plots

For two quantitative variable/s:

- scatterplots
- line plots (against time)

Two Continuous Variables: Where in the statistical process?

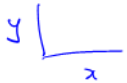
- Ethics
- Nature of the question to be answered
  Is there a linear relationship between two quantitative variables?
- Context/Expertise
- Design:
  - Experiment vs. observational study
  - Sampling
  - Measurement
- **Description and analysis**
  Scatterplots
- Conclusions and decision making

VARIATION

Bivariate Data Analysis

		x	y
1	Mid Session		Exam
2	1	51.7	48.3
3	2	96.7	63.9
4		38.3	48.9
5		78.3	56.1
6		91.7	72.2
7		68.3	53.3
8		83.3	80.0
9		63.3	53.9
10		58.3	33.3
11		60.0	76.1
12		53.3	60.6
13		98.3	86.1
14		73.3	71.1
15		81.7	53.9
16		70.0	56.1
17		80.0	61.7
18		93.3	84.4
19		91.7	71.7
20		83.3	68.9
21		90.0	75.6
22	21	65.0	74.4

Two quantitative variables



- Plot the (x,y) pairs of points on a **scatterplot**
  - one to be a response variable which is on the y-axis

    It is sometimes called the dependent variable
  - and an explatory variable on the x-axis

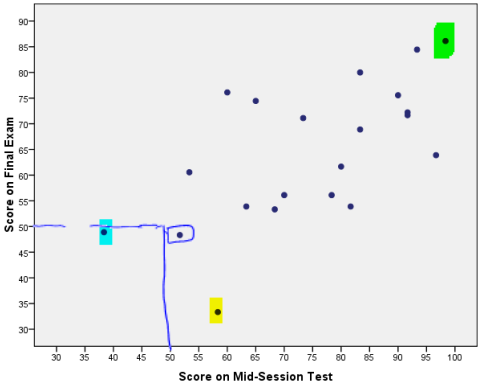
    It is sometimes called called the independent variable

n = 21

Bivariate: Scatterplot

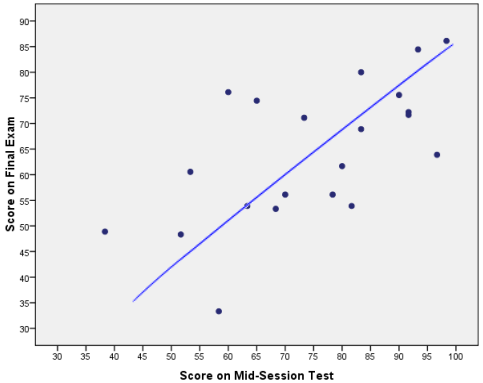
Consider the mark on a mid-session test and the mark on the final exam for 21 students studying a statistics course.

	Mid Session	Exam
1		
2	51.7	48.3
3	96.7	63.9
4	38.3	48.9
5	78.3	56.1
6	91.7	72.2
7	68.3	53.3
8	83.3	80.0
9	63.3	53.9
10	58.3	33.3
11	60.0	76.1
12	53.3	60.6
13	98.3	86.1
14	73.3	71.1
15	81.7	53.9
16	70.0	56.1
17	80.0	61.7
18	93.3	84.4
19	91.7	71.7
20	83.3	68.9
21	90.0	75.6
22	65.0	74.4



Bivariate: Scatterplot

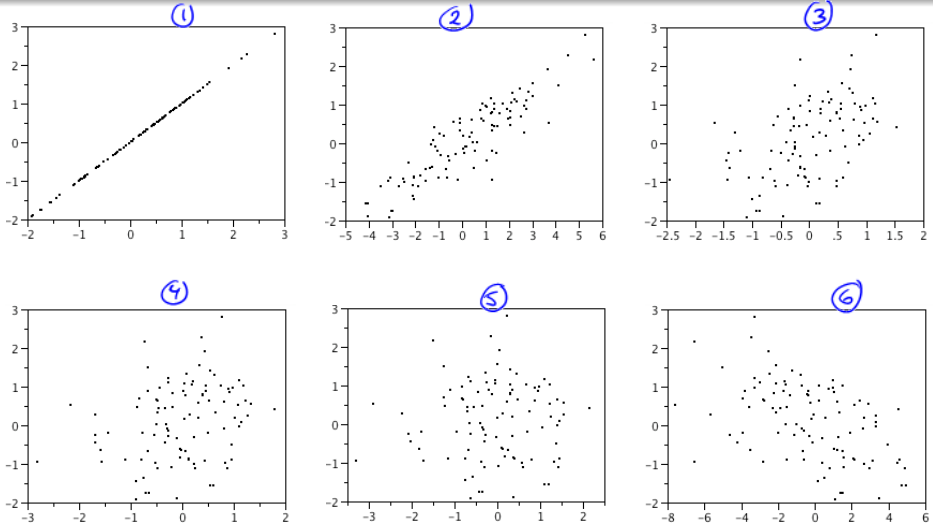
Scatterplot: Examines how two variables vary together.



Consider the mark on a mid-session test and the mark on the final exam for 21 students studying a statistics course.

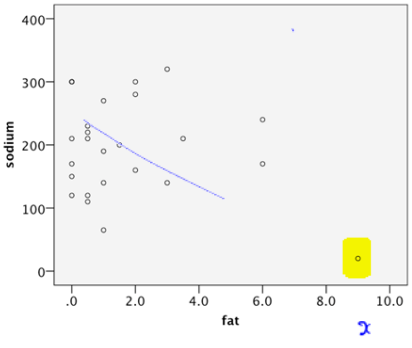
As marks on Mid-session test increase, marks on Final exam tend to _____increase_____.

More Examples



Check Plot - Make Sense of Data in Context

Context is used to make sense of real data: Cans of soup.



What do you see?

Sodium (Y) versus Fat (X)

- There is one outlier for Fat
- Is it also an outlier for Sodium?
- It is a bivariate outlier

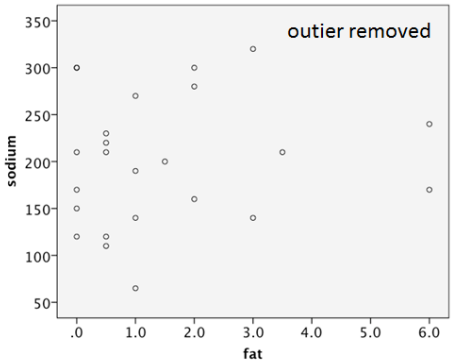
What do we do?

- Check original data for possible recording error since any further analysis will be affected by it.
- Remove the point and redo the plot.

Redo Plot

Context is used to make sense of real data: Cans of soup.

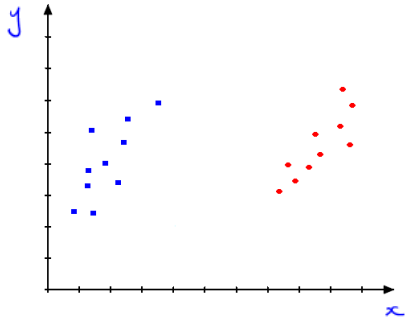
Scatterplot with outlier removed:



Effectively there is _____ relationship now present in the plot.

Plot Reveals Clusters of Points

Sometimes data points separate into two or more different **clusters** which may indicate different groups that should be examined separately:



Ask: Is there another **factor or qualitative variable** which explains the separate clusters?