Day 18

Review A17

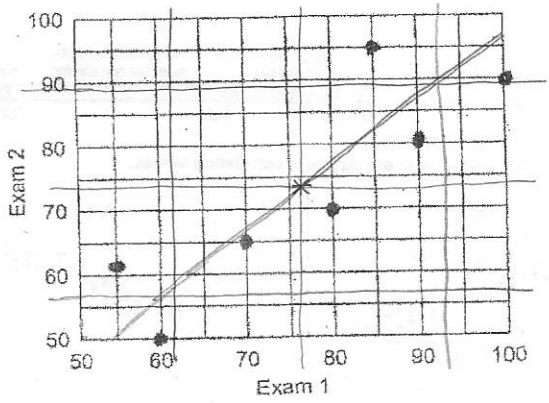
For quantitative data, we make a picture (usually a histogram) we describe a distribution by telling about its: Shape, center, spread, and unusual features

Now we are concerned with describing the association between two quantitative variables. Once we make a picture (a scatterplot) of paired x- and y-values we describe the association by telling about the: form, direction, strength, and unusual features

Rather than use generic terms such as a little, some, or a lot to describe the spread of a distribution we developed numerical summaries IQR or standard deviation. Well now we need to develop a numerical summary for the strength of the association between two quantitative variables.

Example 1: Construct a scatter plot for the Exam 1 and Exam 2 scores of 7 students:

Exam 1	Exam 2
55	62
60	50
70	65
80	70
85	95
90	80
100	90
$\bar{x}$ 77.14	73.14
s 16.29	16.05



b) describe the association

Form: straight

Direction: +

strength: somewhat

unusual: none

, Calculate the mean and standard deviations for Exam 1 and 2. Label (mean(x), mean(y)) on the graph and use SD_x and SD_y to create a "box".

The above statistics tell us the location and size of the "cloud" of data but nothing about the relationship between x and y.

Correlation describes the strength and direction of the linear relationship between two quantitative variables, without significant outliers	
3 conditions needed for Correlation:	1. quantitative variables 2. straight enough - check for linearity with scatterplot 3. outlier

Correlation measures the strength of association between x and y by measuring how closely the points of a scatterplot hug a line that goes through all the points that are an equal number of standard deviations away from the mean for both variables. This line is known as the standard deviation line and goes through (mean(x), mean(y)) and has slope SD_y/SD_x if + correlation and $-SD_y/SD_x$ if -- correlation.

Wait just a minute; we have heard this story before. It's just that now we are comparing many pairs of apples and oranges. This sounds like a job for Z-scores!