MAT 1372 Statistics with Probability classwk 14 Spring 2012

# 2.5 Sets of paired data and scatter plots

Given a data set of paired values (ordered), we can use a scatter plot to see if there is any relation (correlation, not causation) between the 2 variables. We can fit a line by eye to the data or we can use Excel to do so (trend line). One good exercise is to print out the scatter plot for a data set, put the line in by eye and hand then have Excel do it and compare. With a little practice, you will soon get close to what the software can do. We will learn next week how to mathematically find the line which best fits your data.

2.5.5 The following data relate the attention span (in minutes) to a score on an IQ examination of 18 preschool-age children.



2.5.7 A random group of 12 high school juniors were asked to estimate the average number of hours they study each week. The grade point averages of these students were then determined, with the resulting data being as given in the following. Use it to represent these data in a scatter diagram.



# 3.7 Correlation Coefficient

Given a data set of paired values (ordered), covariance is a measure of how much two variables change together.



In the formula, notice the appearance of the deviations from the mean  that we encountered earlier. The values of the covariance are between  and the.

We normalize or scale the covariance by dividing by  to get the correlation coefficient ***r*** :



The correlation coefficient measures how far the data vary from a line. A value close to 1 or -1 indicates that the data will be close to a line. Anything close to a zero indicates that there is no connection between the 2 variable components; the data will appear randomly scattered.

There are important properties of the coefficient. If we change the units, it has no affect on the coefficient. Algebraically, this means that if substitute axi+b for the xi part of the data, a and b some fixed constants, the coefficient will not change. Converting from Celsius to Fahrenheit, the constants are 9/5 and -32 respectively. Likewise, we can do a similar transformation on the 2nd variable and the coefficient will not change.

3.7.4 The following is a sampling of 10 recently released first-time federal prisoners. The data give their crime, their sentence, and the actual time that they served.



Draw a scatter diagram of the sentence time versus time actually served. Compute the sample correlation coefficient. What does this say about the relationship between the length of a sentence and the time actually served?

3.7.5 Using the data of 3.7.4, determine the sample correlation coefficient of the sentence time and the proportion of that time actually served.

3.7.8 The following table gives yearly per capita soft drink consumption (in litres) and the yearly per capita milk consumption (in kg) for a variety of countries. Use it to find the sample correlation coefficient between soft drink and milk consumption.

