

Lab 1

Topic: Summations

The attached table gives data for some popular fast foods.

Do the following:

(a) Enter each column of data as a list. Be sure to keep the order of top to bottom in each case. For example let us call the "Calories" data, "cal":

>cal:=`[660,538,591,413,356,445,351,370,483,439,383,584,405,409,266,506,429]`;

`cal := [660, 538, 591, 413, 356, 445, 351, 370, 483, 439,
383, 584, 405, 409, 266, 506, 429]`

(b) Let's agree to call the remaining data, fat, carb and na. Now enter each of these data as lists.

In statistics there is a quantity we can compute which gives us a sense of how closely sets of data are linked, or, more formally a measure of the correlation between data, namely the Pearson correlation coefficient. Suppose we have two sets of data

$X = [x_1, x_2, x_3, \dots, x_n]$ and $Y = [y_1, y_2, y_3, \dots, y_n]$ and further suppose we have calculated the mean of each set by the formulas,

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} \quad \text{and} \quad \bar{y} = \sum_{i=1}^n \frac{y_i}{n},$$

then the Pearson correlation coefficient is given by

$$\rho(X, Y) = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

where $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ and $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$

(c) Now calculate the following: $\rho(\text{cal}, \text{fat})$, $\rho(\text{cal}, \text{carb})$, $\rho(\text{cal}, \text{na})$, $\rho(\text{fat}, \text{carb})$, $\rho(\text{fat}, \text{na})$ and $\rho(\text{carb}, \text{na})$. For each of these quantities you should get a number between -1 and 1.

For example you should get $\rho(\text{cal}, \text{fat}) = .8542285013$. A "1" indicates perfect correlation. A "0" indicates no correlation while a "-1" indicates perfect anticorrelation. Thus in the case of calories and fat there is a strong linkage or correlation. That is to say if a food is high (low) in calories it will tend to be high (low) in fat.

Which two sets of data are the most closely linked?

Which two sets are the least close?

Explain your answers in a couple of sentences which would be suitable for an article in the News and Observer.

(d) Plot the graphs of the two pairs of data you chose at the end of part (c). Do the graphs reinforce your selections?