Practice Data Analysis by Data Mining

Fumio Shibui

In this paper, the tentative plan which understands the view of "data mining" which attracts attention now through training not only at the industrial world but at a university is tried next, and is stated in detail.

1 Data Analysis Editing by Spreadsheet

There is no telling how it should utilize for there being much data of a total, sale record, etc. of a questionnaire, well. —

I will use the "data analysis function" of "Excel" learned until now at such time.

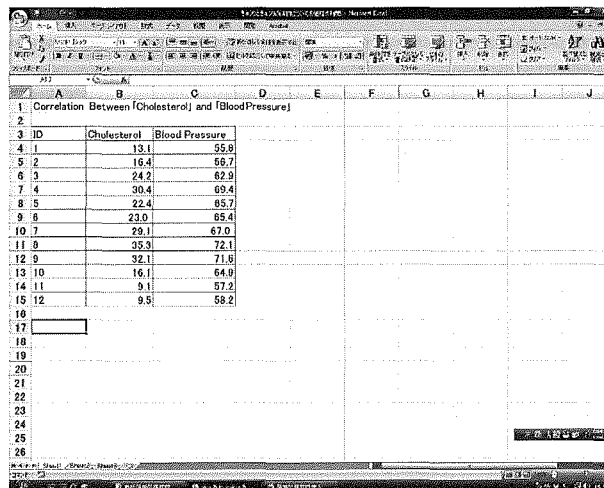For example, the secret of the popularity of a hot-selling product is known by very easy operation.

1. "Tendency of the near future" – the prediction from a data degree of association – here, he studies the procedure of analyzing the relation which is between two data first.

It actually has training in an easy example using PC.

A lower table is a table which summarized cholesterol, and correlation of "Blood Pressure" to monthly.

Since the amount of money of salt or broth has required the way of pickles, it is high.

Data is inputted into a cell as follows.

Next, a "scatter diagram" is made.

The relation of two or more data is called "correlation."

A scatter diagram uses, when investigating a relation about the price about the pickles cholesterol and Blood Pressure, and it can analyze what kind of relation the data assigned to the horizontal axis (X-axis) and the vertical axis (Y-axis) according to the variation condition of the "point" which each data expresses has.

For example, supposing a point is located in a line with an "upward rise", the relation [the value of Y will also increase if the value of X increases] will be realized.

This is told to the wind "positive correlation."

It is called "negative correlation" by the relation of [the value of Y decreasing if the value of X increases] supposing it is reverse and a point is located in a line with "the bottom of a right shoulder is make"

I think when a point appears in a graph scatteringly, there [there is no correlation conspicuous in the value of X-Y].
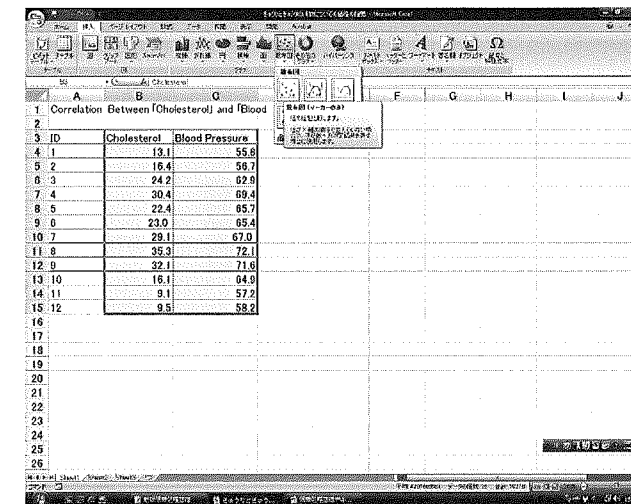
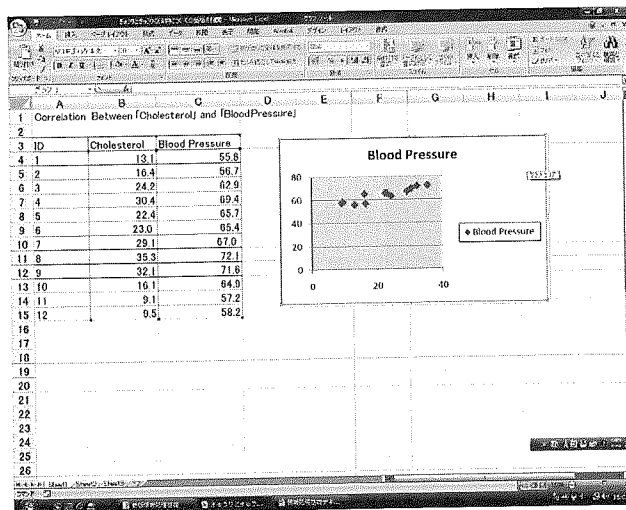A scatter diagram is actually made.
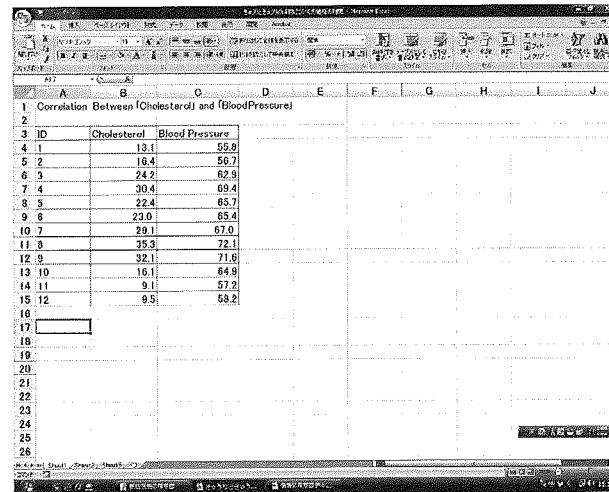
First, C15 is dragged from the cell B3, and a [graph wizard] is opened.
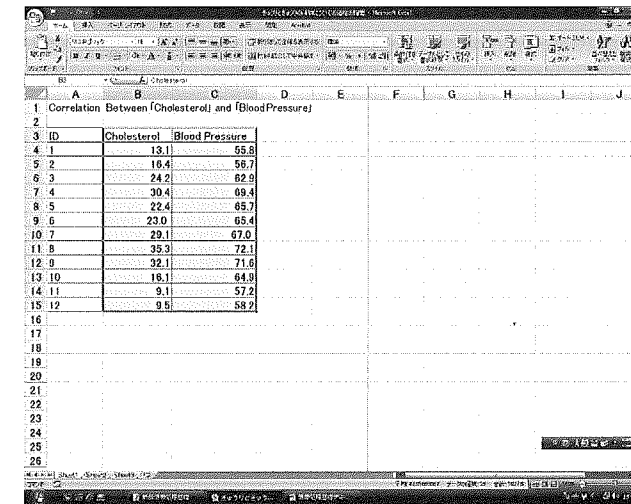
And since a "graph wizard" dialog opens, a "scatter diagram" is clicked from "Graf's kind."

Click the [next] And check whether a radio button called a series is a [sequence], and click the [next].

And a graduation line is set up with a [graduation line] tab so that the position of a mutually related point may become legible previously.

A graduation line is clicked by both of check boxes of the numerical axis of X-Y.
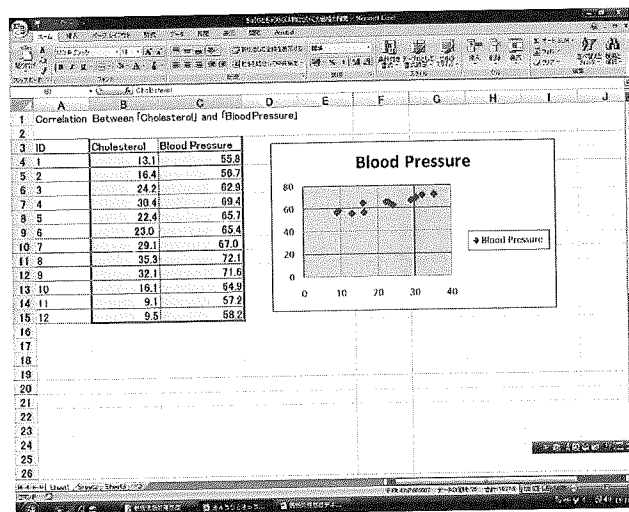
Next, an item name is inputted into a graph title or a numerical axis with a [title and label] tab.

I will input the following character strings into a text box.

Graph title: Correlation of the price about the pickles of Cholesterol and Blood Pressure

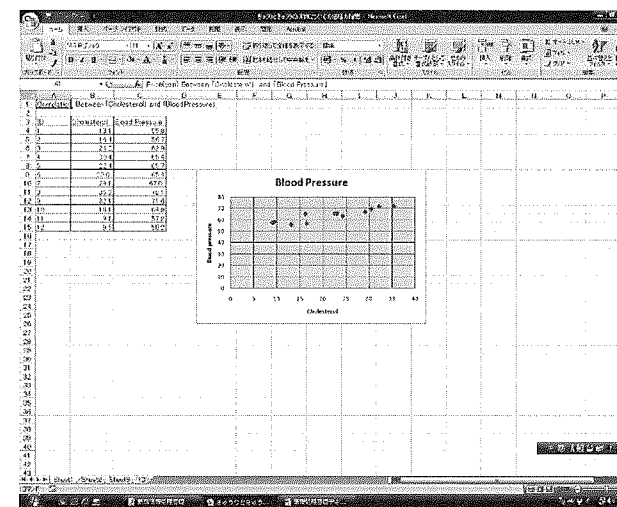X/numerical-value axis: correlation of cholesterol

Y / numerical axis: correlation of the pickles cholesterol

Click the [next].

[Introductory notes] Click a tab and it unchecks the check box of "displaying introductory notes."
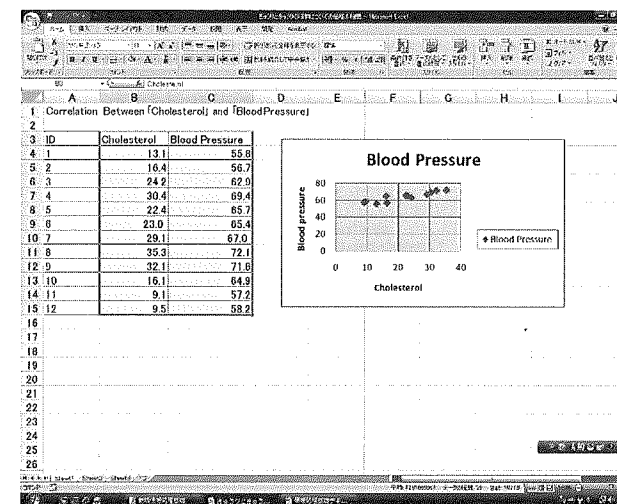
Click the [next], choose an "object" from a bull down menu and click [OK].

The scatter diagram was made now.

Furthermore, a scale is adjusted so that "correlation" may become legible, or appearance is prepared.

Since a setup of the Y-axis is changed, it right-clicks on the Y-axis.

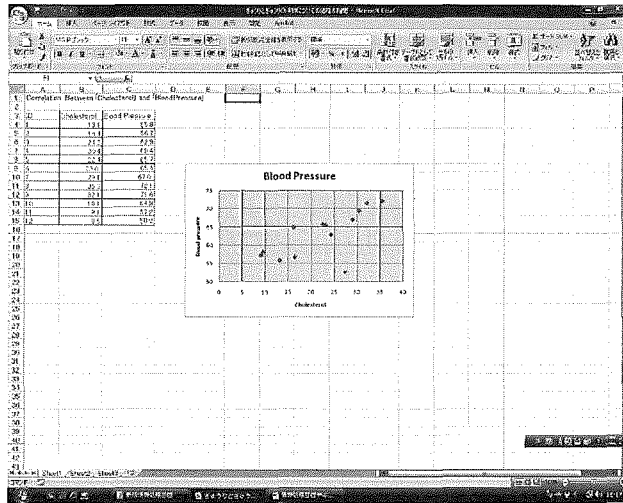If it does so, since the formatting dialog of an axis will open, a [scale] tab is clicked.

Minimum: 50

Maximum: 75

Scale resolution: Five

It changes as mentioned above, respectively.

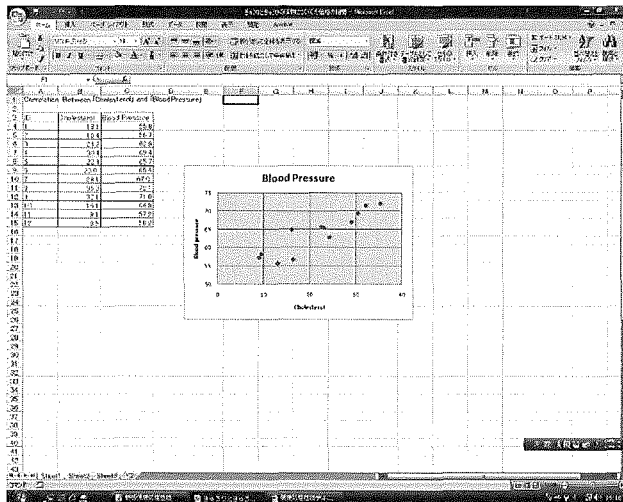Character arrangement of the Y-axis is made into length.

The X-axis is also changed so that a scale may be [every / 10] over in a similar manner.

minimum: Zero

maximum: 40

Peak interval: Ten

It is completion now.