

CS130/230 Lecture 8

Linear Regression

Thursday, February 26, 2004

Regression Analysis

Regression analysis is a form of statistical analysis used for forecasting. Regression analysis estimates the relationship between variables, so that a particular variable can be predicted from one or more other variables. During regression analysis, we need to fit functions to data.

Trendlines are used to graphically display trends in data and to analyze problems of prediction. In other words we try to draw a line that best fits the data. By using regression analysis, you can extend a trendline in a chart beyond the actual data to predict future values.

This subject usually falls under statistics and mathematical modeling and can be applied to many different scientific and business applications. Understanding the various formulas for regression is beyond the scope of this class. However, you should understand that the line should be placed such that the distance or variation from each data point to the line is minimized.

Linear Regression

In linear regression we try to find a straight line that best fits our data. We first need to plot our data using Excel's XY or scatter chart. We then add the trendline to the chart and use the function to predict future values for our data.

The detailed steps are:

- Enter the data in an Excel worksheet and select the data you want to plot.
- Click on the chart wizard.
- Choose XY (scatter) plot.
- Check that the data range is correct.
- Enter the titles and labels.
- Click on the chart then select CHART from menu bar and ADD TRENDLINE from this menu.
- From the menu that appears, select the type of function that you would like to use for your model. In this example we will use the default, which is LINEAR REGRESSION.
- In order to have Excel display the equation of our regression line and the correlation coefficient, you need to click on the OPTIONS tab within the Add Trendline screen above, click on these two options, and then press OK.

You should be rewarded with a graph, equation and regression coefficient.

Problem 1

In the CS130 Pub folder is a file called Candy Bars.xls. Copy this file to your folder, open it and do the following.

Part I: Create a ScatterPlot of the data Carbohydrates and Sugars.

Part II: Add a trendline to your chart and display the function or equation.

Part III: What is the amount of sugars (in grams) that we can expect from a candy bar with 60 grams of carbohydrates?

Part IV: Add an empty column after name. In that column, place an asterisk for foods that have a carbohydrate count of 40grams or higher and a sugar count of 35 grams or higher.

Part V: Turn on the AutoFilter and find out the number of M&M/Mars candy that fits these criteria.

Regression Coefficient

The regression coefficient, also known as the R-squared value, is an indicator that ranges in value from 0 to 1 and reveals how closely the estimated values for the trendline correspond to your actual data. A trendline is most reliable when its R-squared value is at or near 1.

Problem 2

The following table gives data on the lean body mass (kilograms) and the resting metabolic rate for 12 women and 7 men who are subjects in a study of obesity. The researchers suspect that lean body mass (that is, the subject's weight leaving out all fat) is an important influence on metabolic rate.

Subject	Gender	Mass	Rate
1	M	62.0	1792
2	M	62.9	1666
3	F	36.1	995
4	F	54.6	1425
5	F	48.5	1396
6	F	42.0	1418
7	M	47.4	1362
8	F	50.6	1502
9	F	42.0	1256
10	M	48.7	1614
11	F	40.3	1189
12	F	33.1	913
13	M	51.9	1460

14	F	42.4	1124
15	F	34.5	1052
16	F	51.1	1347
17	F	41.2	1204
18	M	51.9	1867
19	M	46.9	1439

Part I: Import the above data into Excel.

Part II: Create a scatterplot of the mass and rate data.

Part III: Add a trendline to your plot. What can you say about the relationship between body mass and metabolic rate.

Part IV: Sort the data by gender.

Part V: Create a scatterplot for each gender, and add a linear trendline to your chart.

Part VI: What is the expected metabolic rate for a female who has a mass of 60?

Part VI: What is the expected mass of a male who has a metabolic rate of 1300?