Regression Techniques

Probability and Statistics for Engineers and Scientists

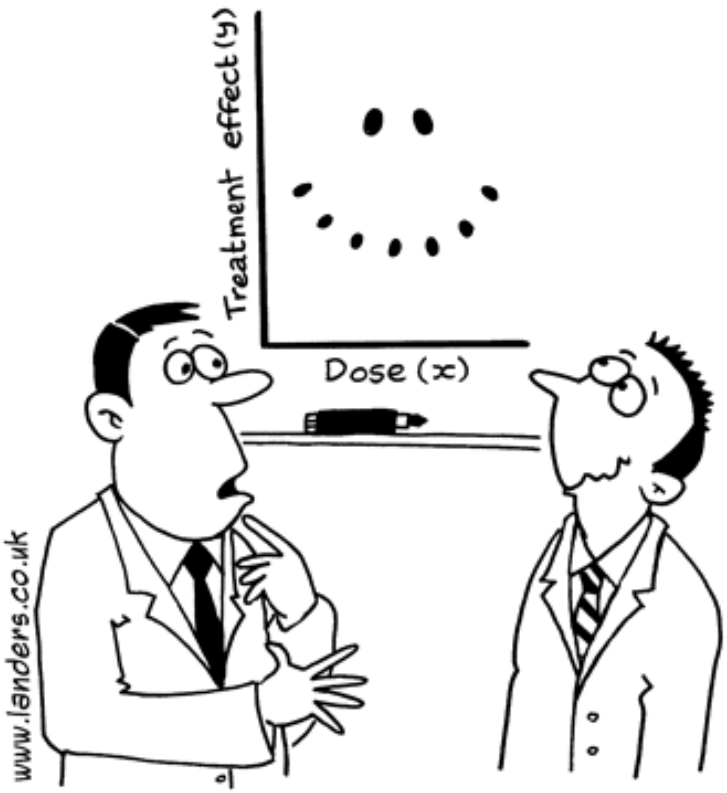
Steve Mazza

Naval Postgraduate School
Monterey, CA

SE3030, Winter/2014
Quantitative Methods of Systems Engineering

Introduction

Regression uses selected values of x and observed values of y in order to predict the most probable value of y for any value of x.



"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

What Is Regression?

Regression is a statistical process for estimating the relationships among variables. Usually regression is used to estimate the expected (or average) value of the dependent variable when the independent variables are fixed. The dependent variable is a function of the independent variables.

Techniques

- simple linear regression
- multiple linear regression
- non-linear regression

Simple Linear Regression

Simple linear regression is a method of a linear regression with a single explanatory variable. It utilizes a least squares approach to finding the best fit within the data. It fits a *straight line* through the set of n points in such a way that minimizes the sum of squared residuals of the model.

Simple Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \ i = 1 \ldots n$$

If you were to plot each of the data points on a graph, the line determined by simple linear regression would make the vertical distance from the fit line to each of the data points as small as possible.

Multiple Linear Regression

In multiple linear regression, the response variable y is modeled as a linear function of more than one input variable x_i.

Multiple Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, \ i = 1 \ldots n$$

While the equation is quadratic with respect to x, it is linear with respect to β.

Due to the complexity of the calculations, multiple linear regression often utilizes linear (matrix) algebra.

Non-linear Regression

We specify a function that relates the values of k input variables and parameters θ to the response variable y.

Non-linear Regression Model

$$y_i = f(x_1 \dots x_k; \theta_1 \dots \theta_p)$$

The unknown parameters are estimated by fitting the model to the data set. For example,

$$(y_1, x_{11}, \dots, x_{k1})$$

$$\vdots$$

$$(y_n, x_{1n}, \dots, x_{kn})$$

we chose parameters to minimize the sum of squares

$$\sum_{i=1}^{n} (y_i - f(x_{1i} \dots x_{ki}; \theta_1 \dots \theta_p))^2$$

Questions?