Explain the difference between multiple regression and multivariate regression, with minimal use of symbols/math

Are multiple and multivariate regression really different? What *is* a variate anyways?

regression | multiple-regression | terminology | multivariate-regression

edited Oct 28 '15 at 9:32
amoeba
25.4k 5 89 150

asked Sep 3 '10 at 18:54
Neil McGuigan
3,337 7 38 54

3 Answers

Very quickly, I would say: 'multiple' applies to the number of predictors that enter the model (or equivalently the design matrix) with a single outcome (Y response), while 'multivariate' refers to a matrix of response vectors. Cannot remember the author who starts its introductory section on multivariate modeling with that consideration, but I think it is Brian Everitt in his textbook An R and S-Plus Companion to Multivariate Analysis. For a thorough discussion about this, I would suggest to look at his latest book, Multivariable Modeling and Multivariate Analysis for the Behavioral Sciences.

For 'variate', I would say this is a common way to refer to any random variable that follows a known or hypothesized distribution, e.g. we speak of gaussian variates X_i as a series of observations drawn from a normal distribution (with parameters μ and σ^2). In probabilistic terms, we said that these are some random *realizations* of X, with mathematical expectation μ, and about 95% of them are expected to lie on the range $[\mu - 2\sigma; \mu + 2\sigma]$.

edited Sep 19 '10 at 9:32

answered Sep 3 '10 at 19:03

chl ♦
36.9k 6 122 241

Even coursera.org/learn/machine-learning/home/week/2 uses the term multivariate regression instead of multiple regression… – Franck Dernoncourt Oct 28 '15 at 2:32

I think the same confusion arises with people using the term GLM for General Linear Model (e.g., in neuroimaging studies) vs. Generalised Linear Model. I have seen many instances of "multivariate logistic regression" where there's only one outcome, and I don't think this matters so much as long as the term is clearly defined by the author. – chl ♦ Oct 28 '15 at 10:38

Here are two closely related examples which illustrate the ideas. The examples are somewhat US centric but the ideas can be extrapolated to other countries.

Example 1

Suppose that a university wishes to refine its admission criteria so that they admit 'better' students. Also, suppose that a student's grade Point Average (GPA) is what the university wishes to use as a performance metric for students. They have several criteria in mind such as high school GPA (HSGPA), SAT scores (SAT), Gender etc and would like to know which one of these criteria matter as far as GPA is concerned.

Solution: Multiple Regression

In the above context, there is one dependent variable (GPA) and you have multiple independent variables (HSGPA, SAT, Gender etc). You want to find out which one of the independent variables are good predictors for your dependent variable. You would use multiple regression to make this assessment.

Example 2

Instead of the above situation, suppose the admissions office wants to track student performance across time and wishes to determine which one of their criteria drives student performance across time. In other words, they have GPA scores for the four years that a student stays in school (say, GPA1, GPA2, GPA3, GPA4) and they want to know which one of the independent variables predict GPA scores better on a year-by-year basis. The admissions office hopes to find that the *same* independent variables predict performance across all four years so that their choice of admissions criteria ensures that student performance is consistently high across all four years.

Solution: Multivariate Regression

In example 2, we have multiple dependent variables (i.e., GPA1, GPA2, GPA3, GPA4) and multiple independent variables. In such a situation, you would use multivariate regression.

answered Sep 3 '10 at 19:27

user28

Simple regression pertains to **one** dependent variable (y) and **one** independent variable (x): $y = f(x)$

Multiple regression (aka multivariable regression) pertains to **one** dependent variable and **multiple** independent variables: $y = f(x_1, x_2, \ldots, x_n)$

Multivariate regression pertains to **multiple** dependent variables and **multiple** independent variables: $y_1, y_2, \ldots, y_m = f(x_1, x_2, \ldots, x_n)$. You may encounter problems where both the dependent and independent variables are arranged as matrices of variables (e.g. $y_{11}, y_{12}, \ldots$ and $x_{11}, x_{12}, \ldots$), so the expression may be written as $Y = f(X)$, where capital letters indicate matrices.

Further reading:

- "R Cookbook" by P. Teetor, O'Reilly publisher, 2011, Chapter 11 on "Linear Regression and ANOVA".
- Quora question "What is the difference between a multiple linear regression and a multivariate regression?"
- Mathworks (Matlab) tutorial on linear regression.

answered Jul 18 at 3:29

stackoverflowuser2010
370 2 12

I understand the definition. But what is the effect of treating a multi-variate regression as a system of uni-variate regressions? – LKS Aug 3 at 19:30