ARTICLE IN PRESS

Early Human Development xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Early Human Development

journal homepage: www.elsevier.com/locate/earlhumdev

Best practice guidelines

WASP (Write a Scientific Paper): Multivariate analysis

Victor Grech[a,*], Neville Calleja[b]

[a] Academic Department of Paediatrics, Mater Dei Hospital, Malta
[b] Department for Policy in Health - Health Information and Research, Malta

ARTICLE INFO

Keywords:
Software
Computers, statistics
Biostatistics

ABSTRACT

Linear regression is the equation which provides of straight line that best describes the association between two continuous variables, x and y. However, it is often the case that the dependent variable y is influenced by more than one variable and not just a single x variable. Multivariate analysis is a statistical modeling technique wherein multiple x variables are analysed simultaneously for their effect on y, resulting in an additive model (via an equation) that explains the observation/s and corrects for confounding association/s using one dependent and several independent variables, assigning a gradient to each of these independent variables, and with all product terms of gradient and magnitude of the independent variables adding up to estimate 'y'. This paper outlines these various techniques and applications.

1. Introduction

The previous paper in this series in *Early Human Development* introduced linear regression as the equation which provides of straight line that best describes the association between two continuous variables, x and y (Grech linear regressionxx). This relationship is modelled by providing a line of least squares fit for the available data which is described by the equation $y = a + bx$, where a is the intercept and b is the slope, i.e. a line defined by the lowest sum of perpendicular distances from the points being fitted (Figs. 1 and 2).

2. Multivariate analysis

However, it is often the case that the dependent variable y is influenced by more than one variable and not just x. These may be labelled x_1, x_2, x_3 and so on. Multivariate analysis is a statistical modeling technique wherein multiple x variables are analysed simultaneously for their effect on y, resulting in an additive model (via an equation) that explains the observation/s and corrects for confounding association/s using one dependent and several independent variables, assigning a gradient to each of these independent variables, and with all product terms of gradient and magnitude of the independent variables adding up to estimate 'y' (Figs. 5 and 6).

Multivariate methods comprise a range of analysis techniques depending on the data at hand, yielding an equation that resembles:

$$Y = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_nx_n$$

3. Logistic regression

Logistic regression is a similar technique wherein the outcome (dependent variable) is binary e.g. alive/deceased. So, rather than a gradient, one gets an odds ratio [1]. The model is ultimately still additive, but on the log scale. However, the independent variables may be binary or continuous.

For example, a paper entitled "Epidemiology and prognostic factors in meningococcal disease in a small island population: Malta 1994–1998" analysed "factors at presentation [with meningococcal disease] and in the management of meningococcal disease which may influence mortality" in the cohort studied [2]. Mortality is a binary variable: alive/deceased and logistic regression was therefore employed.

4. Approaches

Several regression techniques are available:

4.1. Forward selection

In this technique, independent variables are introduced into a working model one at a time and the order is juggled by the strength of the correlation of each added independent variable. The effect is assessed as each new variable is added. Variables that do not significantly add to the success of the model are excluded. The best sequence with which to add independent variables would be to rank them in order of increasing significant association with y when analysed separately, and

* Corresponding author at: Department of Paediatrics, Mater Dei Hospital, Malta.
E-mail addresses: victor.e.grech@gov.mt (V. Grech), neville.calleja@gov.mt (N. Calleja).

https://doi.org/10.1016/j.earlhumdev.2018.04.012

Please cite this article as: Grech, V., Early Human Development (2018), https://doi.org/10.1016/j.earlhumdev.2018.04.012

	A	B	C	D	E	F	G
1	Year	GDP growth/yr (%)	M/F		n	49	
2	1961	12.04	0.5144		r^2	0.435	=CORREL(B2:B50,C2:C50)
3	1962	8.91	0.5148				
4	1963	8.47	0.5137		df	47	=F1-2
5	1964	11.68	0.5143		t*	3.31	=F2*SQRT(F4/(1-F2^2))
6	1965	5.82	0.5129		2-tailed p	0.00180	=2*TDIST(ABS(F5),F4,1)
7	1966	10.64	0.5184				
8	1967	11.08	0.5129				
9	1968	12.88	0.5171				
10	1969	12.48	0.5173				
11	1970	-1.02	0.5172				
12	1971	4.70	0.5162				
13	1972	8.41	0.5157				
14	1973	8.03	0.5151				
15	1974	-1.23	0.5155				
16	1975	3.09	0.5149				
17	1976	3.97	0.5150				
18	1977	4.39	0.5147				
19	1978	5.27	0.5145				
20	1979	5.48	0.5150				
21	1980	2.82	0.5146				
22	1981	4.18	0.5143				
23	1982	3.38	0.5133				
24	1983	3.06	0.5138				
25	1984	4.46	0.5132				
26	1985	6.33	0.5136				
27	1986	2.83	0.5143				
28	1987	4.11	0.5141				
29	1988	7.15	0.5136				
30	1989	5.37	0.5137				
31	1990	5.57	0.5132				
32	1991	3.32	0.5139				
33	1992	0.82	0.5146				
34	1993	0.17	0.5136				
35	1994	0.86	0.5135				
36	1995	1.94	0.5126				
37	1996	2.61	0.5137				
38	1997	1.60	0.5126				
39	1998	-2.00	0.5132				
40	1999	-0.20	0.5135				
41	2000	2.26	0.5142				
42	2001	0.36	0.5133				
43	2002	0.29	0.5138				
44	2003	1.69	0.5133				
45	2004	2.36	0.5128				
46	2005	1.30	0.5130				
47	2006	1.69	0.5129				
48	2007	2.19	0.5137				
49	2008	-1.04	0.5128				
50	2009	-5.53	0.5131				

Fig. 1. Correlation calculation.

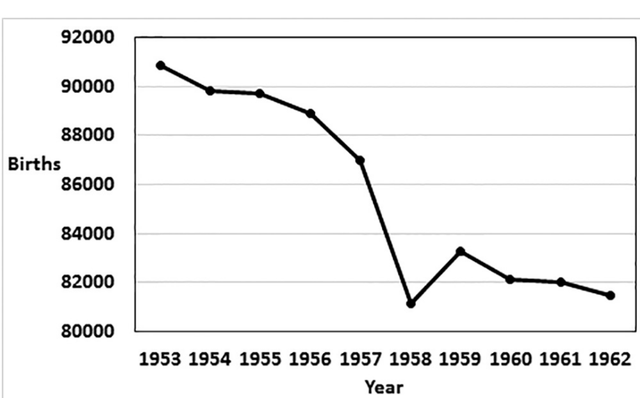

Fig. 2. Annual live births in Finland, 1953–1962 as per Table 1.

Table 1
Annual live births in Finland, 1953–1962.

Year	Births
1953	90,866
1954	89,845
1955	89,740
1956	88,896
1957	86,985
1958	81,148
1959	83,253
1960	82,129
1961	81,996
1962	81,454