Survey Methodology, June 1997
Vol.23, No. 1, pp. 77-78
Statistics Canada

A Simple Derivation of the Linearization of the Regression Estimator

KEES ZEELENBERG[1]

ABSTRACT

We show how the use of matrix calculus can simplify the derivation of the linearization of the regression coefficient estimator and the regression estimator.

KEYWORDS: matrix calculus, regression estimator, Taylor expansion

1. INTRODUCTION

Design-based sampling variances of non-linear statistics are often calculated by means of a linear approximation obtained by a Taylor expansion; examples are the variances of the general regression coefficient estimator and the regression estimator. The linearizations usually need some complicated differentiations. The purpose of this paper is to show how matrix calculus can simplify these derivations, to the extent that even the Taylor expansion of the regression coefficient estimator can be derived in one line, which should be compared with the nearly one page that Särndal et al (1992, pp. 205-6) need. To be honest, the use of matrix calculus requires some more machinery to be set up, which is not needed for traditional methods. However this set-up can be regarded as an investment: once it has been learned, it can be used fruitfully in many other applications. After this paper had been written, Binder (1996) appeared, in which similar techniques are used to derive variances by means of linearization. The present paper can be seen as a pedagogical note, in which the use of differentials is exposed.

2. MATRIX DIFFERENTIALS

2.1. Introduction

We will use the matrix calculus by means of differentials, as set out by Magnus and Neudecker (1988); this calculus differs somewhat from the usual methods, which focus on derivatives instead of differentials. Therefore in this section we will briefly describe the definitions and properties of differentials (see Zeelenberg, 1993, for a more extensive survey). We first define differentials for vector functions, and then generalize to matrix functions.

2.2. Vector Functions

Let f be a function from an open set $S \subset R^m$ to R^n; let x_0 be a point in S. The function f is *differentiable* at x_0 if there exists a real $n \times m$-matrix A, depending on x_0, such that for any $u \in R^m$ for which $x_0 + u \in S$, there holds

$$f(x_0 + u) = f(x_0) + A_{x_0} u + o(u), \tag{1}$$

where $o(u)$ is a function such that $\lim_{|u|\to 0} |o(u)|/|u| = 0$; the matrix A is called the *first derivative* of f at x_0; it is denoted as $\mathsf{D}f(x_0)$ or $\partial f/\partial(x')\rfloor_{x=x_0}$. The derivative $\mathsf{D}f$ is equal to the matrix of partial derivatives, i.e. $\mathsf{D}f(x)_{ij} = \partial f_i/\partial x_j$. The linear function $\mathrm{d}\, f_{x_0} : R^m \mapsto R^n$ defined by $\mathrm{d}\, f_{x_0} : u \mapsto A_{x_0} u$ is called the *differential* of f at x_0. Usually we write $\mathrm{d}\, x$ instead of u so that $\mathrm{d}\, f_{x_0}(\mathrm{d}\, x) = A_{x_0}\, \mathrm{d}\, x$. From (1) we see that the differential corresponds to the linear part of the function, which can also be written as

$$y - y_0 = A_{x_0}(x - x_0),$$

where $y_0 = f(x_0)$. Therefore the differential of a function is the linearization of the function: it is the equation of the hyperplane through the origin that is parallel to the hyperplane tangent to the graph of f at x_0; so the linearized function can be written as

$$f(x) \doteq f(x_0) + A_{x_0}(x - x_0). \tag{2}$$

Alternatively, if B is a matrix such that $\mathrm{d}\, f_{x_0}(\mathrm{d}\, x) = B\, \mathrm{d}\, x$, then B is the derivative of f at x_0 and contains the partial derivatives of f at x_0. This one-to-one relationship between differentials and derivatives is very useful, since differentials are easy to manipulate.

Finally, we usually omit the subscript 0 in x_0, so that we write $\mathrm{d}\, f = A_x\, \mathrm{d}\, x$.

2.3. Matrix Functions

A matrix function F from an open set $S \subset R^{m\times n}$ to $R^{p\times q}$ is differentiable if $\operatorname{vec} F$ is differentiable. The derivative $\mathsf{D}F$ is the derivative of $\operatorname{vec} F$ with respect to $\operatorname{vec} X$, and is also denoted by $\partial \operatorname{vec} F/\partial(\operatorname{vec} X)'$. The differential $\mathrm{d}\, F$ is the matrix function defined by $\operatorname{vec} \mathrm{d}\, F_{X_0}(U) = A_{X_0} \operatorname{vec} U$.

[1]Kees Zeelenberg, Department of Statistical Methods, Statistics Netherlands, P.O. Box 4000, 2270 JM Voorburg, The Netherlands.
The views expressed in this article are those of the author and do not necessarily reflect the policy of Statistics Netherlands.

2.4. Properties of Differentials

Let A be a matrix of constants, F and G differentiable matrix functions, and α a real scalar. Then the following properties are easily proved:

$$\begin{aligned} \mathrm{d}\,A &= 0, & (3)\\ \mathrm{d}(\alpha F) &= \alpha\,\mathrm{d}\,F, & (4)\\ \mathrm{d}(F+G) &= \mathrm{d}\,F + \mathrm{d}\,G, & (5)\\ \mathrm{d}(FG) &= (\mathrm{d}\,F)G + F(\mathrm{d}\,G), & (6)\\ \mathrm{d}\,F^{-1} &= -F^{-1}(\mathrm{d}\,F)F^{-1}. & (7)\end{aligned}$$

The last property can be proved by taking the differential of $FF^{-1} = I$ and rearranging.

3. LINEARIZATION OF THE REGRESSION COEFFICIENT ESTIMATOR

The π-estimator (Horvitz-Thompson estimator) of the finite population regression coefficient (cf. Särndal et al, 1992, section 5.10) is

$$\hat{B} = \hat{T}^{-1}\hat{t}, \tag{8}$$

where

$$\begin{aligned} \hat{T} &= \sum_{k\in s} \frac{x_k x_k'}{\pi_k},\\ \hat{t} &= \sum_{k\in s} \frac{x_k y_k}{\pi_k}, \end{aligned}$$

y_k is the variable of interest for individual k, x_k is the vector with the auxiliary variables for individual k, π_k is the inclusion probability for individual k, and s denotes the sample. Taking the total differential of (8), using properties (6) and (7), and evaluating at the point where $\hat{T} = T$, $\hat{t} = t$, we get

$$\mathrm{d}\,\hat{B} = -T^{-1}(\mathrm{d}\,\hat{T})T^{-1}t + T^{-1}(\mathrm{d}\,\hat{t}). \tag{9}$$

Because of the connection between differentials and linear approximation, as given in equation (2), it immediately follows that (9) corresponds to the linearization of the regression coefficient estimator:

$$\hat{B} \doteq B - T^{-1}(\hat{T}-T)T^{-1}t + T^{-1}(\hat{t}-t) = B + T^{-1}(\hat{t}-\hat{T}B),$$

where $B = T^{-1}t$.

4. LINEARIZATION OF THE REGRESSION ESTIMATOR

The regression estimator of a population total is (cf. Särndal et al, 1992, section 6.6)

$$\hat{t}_{yr} = \hat{t}_{y\pi} + (t_x - \hat{t}_{x\pi})'\hat{B}, \tag{10}$$

where $\hat{t}_{y\pi}$ is the π-estimator of the variable of interest, t_x is the vector with the population totals of the auxiliary variables, $\hat{t}_{x\pi}$ is the vector with the π-estimators of the auxiliary variables, and $\hat{B}$ is the estimator of the regression coefficient of the auxiliary variables on the variable of interest. Taking the total differential of (10), using properties (3) and (6), and evaluating at the point where $\hat{t}_{y\pi} = t_y$, $\hat{t}_{x\pi} = t_x$, and $\hat{B} = B$, we get the linear approximation of the regression estimator

$$\mathrm{d}\,\hat{t}_{yr} = \mathrm{d}\,\hat{t}_{y\pi} - (\mathrm{d}\,\hat{t}_{x\pi})'B,$$

so that

$$\begin{aligned} \hat{t}_{yr} &\doteq t_y + \hat{t}_{y\pi} - t_y + (t_x - \hat{t}_{x\pi})'B\\ &= \hat{t}_{y\pi} + (t_x - \hat{t}_{x\pi})'B. \end{aligned}$$

Note that for the linearization of the regression estimator we do not need that of the regression coefficient estimator B.

ACKNOWLEDGEMENTS

I wish to thank Jeroen Pannekoek, Jos de Ree, Robbert Renssen, two referees, and an Associate Editor for their comments.

REFERENCES

BINDER, D.A. (1996). Linearization methods for single phase and two-phase samples: a cookbook approach. *Survey Methodology*, 22, 17-22.

MAGNUS, J.R. and NEUDECKER, H. (1988). *Matrix Differential Calculus.* New York: Wiley.

SÄRNDAL, C.-E., SWENSSON, B., and WRETMAN, J. (1992). *Model Assisted Survey Sampling.* New York: Springer.

ZEELENBERG, C. (1993). *A Survey of Matrix Differentiation.* Research Paper, Department of Statistical Methods. Voorburg: Statistics Netherlands.