



Verskillende logistieke regressiemetodes waarby volledige of kwasivolledige skeiding in die steekproef teenwoordig is

Author:M. Botes¹**Affiliation:**¹Department of Statistics, University of Pretoria, South Africa**Correspondence to:**

M. Botes

Email:

michelleb@lightstone.co.za

Postal address:

Private Bag X20, Hatfield 0028, South Africa

How to cite this abstract:

Botes, M., 2014, 'Verskillende logistieke regressiemetodes waarby volledige of kwasivolledige skeiding in die steekproef teenwoordig is', *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* 33(1), Art. #1193, 1 page. <http://dx.doi.org/10.4102/satnt.v33i1.1193>

Note:

A selection of conference proceedings: Student Symposium in Science, 07 and 08 November 2013, University of Pretoria, South Africa. Organising committee: Mr Rudi W. Pretorius (Department of Geography, University of South Africa) and Ms Andrea Lombard (Department of Geography, University of South Africa), Dr Hertzog Bisset (South African Nuclear Energy Corporation [NECSA]) and Prof. Philip Crouse (Department of Chemical Engineering, University of Pretoria).

Copyright:

© 2014. The Authors. Licensee: AOSIS OpenJournals. This work is licensed under the Creative Commons Attribution License.

Read online:

Scan this QR code with your smart phone or mobile device to read online.

Different logistic regression methods containing complete or quasi-complete separation in the sample. Complete or quasi-complete separation can occur when applying a logistic regression model. Exact logistic regression, Firth's method and hidden logistic regression are three of numerous methods used to deal with separation. These methods are compared to each other for different sample sizes and types of covariates.

Die inligting wat van die grootste belang in eksperimentele ontledings is, is hoe die veranderinge in een veranderlike (die onafhanklike veranderlike) 'n ander veranderlike (die afhanklike veranderlike) beïnvloed. Die kenmerkendste verskil tussen 'n logistieke regressiemodel en 'n liniêre regressiemodel is die aard van die afhanklike veranderlike. 'n Liniêre regressiemodel het 'n skalaar- afhanklike veranderlike, terwyl die uitkoms vir die logistieke regressiemodel binêr of digotoom is.

Die logistieke regressiefunksie is die logaritme van die waarskynlikheid van 'n sukses, voorgestel as 'n liniêre kombinasie van al die onafhanklike veranderlikes wat by die model ingesluit is. Vir 'n eenvoudige binêre model waarby elke waarneming net een van twee moontlike uitkomstes kan aanneem, word een afsnywaarde vir die waarskynlikheid van 'n sukses van die afhanklike veranderlike geïmplementeer. Indien die waarskynlikheid van 'n sukses vir 'n spesifieke waarneming bo die afsnypunt val, word die waarneming in 'n spesifieke groep, sê groep 1, geklassifiseer. Indien die waarskynlikheid van 'n sukses vir 'n spesifieke waarneming onder die afsnypunt val, word die waarneming in 'n ander groep geklassifiseer, sê groep 2.

Wanneer een van die onafhanklike veranderlikes die afhanklike veranderlike perfek in die onderskeie groepe kan klassifiseer, het die waarskynlikheidsfunksie geen maksimum nie, en kan geen eindige waarde dus vir die beraamers van die regressiekoëffisiënte gevind word nie. Daar is drie verskillende onderling uitsluitende en allesomvattende klasse waarin die data van 'n logistieke regressie geklassifiseer kan word: volledige skeiding, kwasivolledige skeiding en oorvleuelende data. Volledige en kwasivolledige skeiding impliseer dat slegs 'n oneindige of 'n nulmaksimum-waarskynlikheidsberaming vir die regressiekoëffisiënte verkry kan word.

Verskeie metodes waarvolgens volledige skeiding en kwasivolledige skeiding hanteer kan word, is mettertyd ontwikkel. Eksakte logistieke regressie, Firth se metode en verborge logistieke regressie word hier bespreek en aan die hand van praktiese voorbeelde illustreer. Hierdie metodes word met mekaar vergelyk in verskillende situasies waarin twee onafhanklike veranderlikes oorweeg word. 'n Klein steekproef met 20 waarnemings word ondersoek en vergelyk met 'n datastel met 100 waarnemings waarby onderskeidelik volledige skeiding en kwasivolledige skeiding teenwoordig is.

Indien die data nie oorvleuel nie, word eksakte logistieke regressie, Firth se metode en verborge logistieke regressie op die datastel toegepas. Vir elk van hierdie drie modelle word die regressiekoëffisiënte, die pasgehalte van die model (Pearson se chi-kwadraatstatistiek, afwyking en die Hosmer-Lemeshow-toetsgrootte) sowel as die klassifikasietabel geëvalueer. Wanneer daar met 'n groot datastel gewerk word wat in kategorieë verdeel is, word die beste resultate in die algemeen volgens eksakte logistieke regressie verkry. Firth se metode gee beduidende regressiekoëffisiëntberamings vir alle gevalle en verander die data op 'n logistiese kromme wat geleidelik verhoog tot 'n beraamde waarskynlikheid van tussen 0 en 1. Die verborge logistieke regressiemodel verskaf perfekte klassifikasie vir alle gevalle, maar verteenwoordig 'n model onder volledige of kwasivolledige skeiding.