

## Project 3

Academiejaar 2013-2014

---

De dataset `regios.txt` bevat voor de steden en gemeenten uit 8 Vlaamse regio's gegevens voor de volgende variabelen in het jaar 2010:

Demografisch	Relatief migratie saldo (% van totale bevolking) Grijs druk (% 65+ t.o.v. 20-64) Groene druk (% -20 t.o.v. 20-64) natuurlijke loop (+1 bij aangroei, -1 bij daling)
Economisch	% belastingsaangiften > 50.000 euro Gemiddeld inkomen per aangifte werkzaamheidsgraad (% werkenden van (18-64) bevolking) werkloosheidsgraad (% werkzoekenden van (18-64) bevolking) Gemiddelde verkoopprijs van woonhuizen
Armoede	% leefloners t.o.v. totale bevolking % geboorten in kansarme gezinnen % bejaarden met inkomensgarantie (i.e. uitkering)
Geografisch	Regio Stad

De variabele 'Regio' is een indicator die aangeeft welke steden en gemeenten tot eenzelfde regio behoren. De variabele 'Stad' geeft weer of de observatie een stad (1) of gemeente (0) is.

Je beantwoordt de onderstaande vragen door individueel gepaste analyses uit te voeren met R. De bespreking van de resultaten en de nodige figuren verwerk je in een schriftelijk rapport dat maximaal uit 6 bladzijden mag bestaan (12pt lettergrootte). In een appendix voeg je de gebruikte R-code toe. Rapporteer enkel resultaten en interpretaties, herhaal geen theorie uit de cursus! Het rapport dien je in pdf vorm in ten laatste op **12 juni 2014** via Toledo. Dit project telt mee voor 5 punten van het eindresultaat.

*Opgaven:*

Het doel is om te onderzoeken of steden van gemeenten kunnen onderscheiden worden op basis van de beschikbare variabelen.

1. Gebruik een regressiemodel om de respons 'Stad' te verklaren aan de hand van de andere beschikbare variabelen.
2. Geeft dit regressiemodel een aanvaardbare fit?
3. Is de variabele 'Regio' nuttig in het regressiemodel of kan deze variabele beter weggelaten worden?
4. Zijn er invloedrijke observaties of andere afwijkingen die best opgelost worden om een beter model te bekomen?
5. Hoe goed kan je met dit model steden van gemeenten onderscheiden?
6. Via discriminantanalyse kan je ook steden van gemeenten proberen te onderscheiden. Levert dit een beter resultaat op dan met voorgaand model?

Veel succes!