

## Project 2

Academiejaar 2013-2014

Voor deze opgave analyseren jullie opnieuw gegevens afkomstig van de Vereniging van Vlaamse Steden en Gemeenten (<http://www.vvsg.be>). Elke groep analyseert de gegevens voor dezelfde regio als in het eerste project. Let op: de datasets verschillen van deze voor de eerste opgave. Alle gegevens zijn nu afkomstig van het jaar 2010 (het meest recente jaar waarvoor alle gegevens al beschikbaar zijn). De datasets bevatten gegevens voor de volgende variabelen:

Demografisch	Relatief migratie saldo (% van totale bevolking) Grijze 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)
Armoede	% leefloners t.o.v. totale bevolking % geboorten in kansarme gezinnen % bejaarden met inkomensgarantie (i.e. uitkering)
Respons	Gemiddelde verkoopprijs van woonhuizen

Je beantwoordt de onderstaande vragen door 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 12 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 **26 mei 2014** via Toledo. Dit project telt mee voor 2 punten van het eindresultaat.

### Opgaven:

1. Voer een regressieanalyse uit op jullie gegevens. Modeleer de respons (gemiddelde verkoopprijs van woonhuizen) aan de hand van de overige variabelen. Evalueer dit model en geef duidelijk aan welke problemen er eventueel met dit model zijn.
2. We veronderstellen nu dat het bekomen model een goed regressiemodel is. Geef dan een interpretatie voor de geschatte regressiecoëfficiënt bij de volgende variabelen: natuurlijke loop, gemiddeld inkomen per aangifte, en % geboorten in kansarme gezinnen.

3. Onderzoek of het model nog verbeterd kan worden door transformaties van respons en/of predictorvariabelen. Argumenteer je keuze.
4. Is er multicollineariteit aanwezig in je dataset? Zo ja, welke oplossing stel je voor.
5. Onderzoek of er invloedrijke uitschieters in je dataset aanwezig zijn.
6. Bouw een goed regressiemodel als je dit model hoofdzakelijk wil gebruiken om de gemiddelde verkoopprijs van woonhuizen in andere regio's te voorspellen.

Veel succes!