

Project 1

Academiejaar 2013-2014

Voor deze opgave analyseren jullie gegevens afkomstig van de Vereniging van Vlaamse Steden en Gemeenten (<http://www.vvsg.be>). Deze vereniging brengt gegevens uit verschillende databanken samen en levert uitgebreide informatie over demografische en economische indicatoren per gemeente, regio, provincie, etc. Binnen het Vlaamse gewest worden 13 centrumsteden herkend, nl. Aalst, Antwerpen, Brugge, Genk, Gent, Hasselt, Kortrijk, Leuven, Mechelen, Oostende, Roeselare, Sint-Niklaas en Turnhout. Elke groep analyseert de gegevens van een regio bestaande uit een centrumstad en zijn omliggende gemeenten.

De datasets bevatten de meest recente gegevens (afkomstig van 2011 of 2012) voor de volgende variabelen:

Demografisch	Totaal aantal mannen
	Totaal aantal vrouwen
	Aantal geboorten
	Aantal overlijdens
	Immigratie saldo (Inwijkelingen-uitwijkelingen)
	80+/60+ verhouding (in%) (0-19)/totaal verhouding (in %)
Economisch	Aantal belastingaangiften < 10.000 euro
	Aantal belastingaangiften > 50.000 euro
	Gemiddeld inkomen per aangifte
	werkzaamheidsgraad (% werkenden van (18-64+) bevolking)
	werkloosheidsgraad (% werkzoekenden van (18-64+) bevolking)
	Gemiddelde verkoopprijs van woonhuizen

De groepen worden samengesteld tijdens het practicum. De datasets worden als volgt verdeeld:

Groep	Dataset
1	Aalst
2	Gent
3	Hasselt
4	Kortrijk
5	Leuven
6	Mechelen
7	Sint-Niklaas
8	Turnhout

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 **4 april 2014** via Toledo. Dit project telt mee voor 2 punten van het eindresultaat.

Opgave 1: PCA

1. Voer een PCA analyse uit op jullie gegevens. Argumenteer waarom je de analyse baseert op de correlatie- of de covariantiematrix van de gegevens. Bepaal het aantal componenten dat je verkiest te behouden en verklaar je keuze.
2. Bekijk de loadings van de weerhouden componenten en tracht de PCs te interpreteren aan de hand van de loadings. Ga na welke originele variabelen het belangrijkste zijn voor elk van de weerhouden PCs.
3. Onderzoek de impact van de centrumstad op het resultaat van de PCA analyse.

Opgave 2: Testen multivariate normaliteit

1. Onderzoek of de veronderstelling van multivariate normaliteit aannemelijk zou zijn voor jullie gegevens. Gebruik hiervoor ook de scores van de PCA analyse.
2. Als de multivariate normaliteit niet aannemelijk zou blijken te zijn, onderzoek dan of je de data kan aanpassen om beter aan deze veronderstelling te voldoen.

Opgave 3: Clustering

1. Voer een cluster analyse uit op de gegevens. Kies op basis van de resultaten volgens verschillende clustermethodes een “optimaal” aantal clusters.
2. Stel de clusterresultaten grafisch voor en bespreek de kwaliteit van de gevonden clustering. Bespreek gelijkenissen en/of verschillen tussen de methoden.
3. Je kan ook een clustering uitvoeren vertrekkende van de scores van de PCA analyse. Zou dit zinvol kunnen zijn of heeft een voorafgaande PCA analyse weinig zin? Levert dit een ander resultaat voor jullie gegevens?

Veel succes!