

# GEAVANCEERDE STATISTISCHE METHODEN

**POSTACADEMISCHE  
OPLEIDING**

30 januari 2018 – 23 oktober 2018



**UNIVERSITEIT  
GENT**

# INLEIDING

Deze opleiding wil een praktisch inzicht geven in een aantal geavanceerde en veelgebruikte statistische methoden voor de analyse van uitgebreide datasets. Een keuze maken uit dit ruime aanbod van methoden berust op een grondige kennis van de voorwaarden waaronder de analysemethode ontwikkeld is. Eens deze selectie werd gemaakt, kan de uitvoering van de analyse worden aangevat met behulp van publiektoegankelijke software pakketten zoals R.

De opleiding start met een aantal lessen rond het pakket R, dat tevens het software pakket is dat in de ganse lessenreeks wordt gebruikt. Daarna komen de niet-parametrische en de meest gebruikte multivariate statistische methoden aan bod. Tijdens iedere les komen er tal van oefeningen aan bod (50% theorie en 50% oefeningen). In de laatste les van de modules 1 en 2 worden er enkel oefeningen gegeven.

Iedere module kan apart gevolgd worden.

## DOELPUBLIEK

De cursus richt zich tot personen in de bedrijfsweld, de dienstensector, de overheid, de sector gezondheidszorg en de farmaceutische wereld maar ook tot mensen uit het onderwijs en alle ondersteunende disciplines die gebruik maken van statistische methoden en modellen.

De cursus is toegankelijk voor iedereen die door zijn basisvorming of door zijn ervaring voldoende met het onderwerp vertrouwd is om met interesse en motivatie de opleiding te volgen.

Door het modulaire karakter van de opleiding kan de cursus de modulekeuze aanpassen aan zijn interesseprofiel en specifieke voorkennis.

## WETENSCHAPPELIJKE COÖRDINATIE

Em. Prof. Jean-Pierre Ottoy, Vakgroep Wiskundige Modelling, Statistiek en Bio-Informatica, Universiteit Gent

## LESGEVERS

- Prof. Tim De Meyer, Vakgroep Wiskundige Modelling, Statistiek en Bio-Informatica, Universiteit Gent
- Prof. Jan De Neve, Vakgroep Data-analyse, Universiteit Gent
- Dhr. Joris Meys, Vakgroep Wiskundige Modelling, Statistiek en Bio-Informatica, Universiteit Gent
- Mevr. Annelies Tourny, Statter (spin-off UGent)
- Dr. Karel Vermeulen, Vakgroep Wiskundige Modelling, Statistiek en Bio-Informatica, Universiteit Gent

## GETUIGSCHRIFT

U ontvangt een getuigschrift, indien u deelneemt aan minstens 2 van de 3 modules en slaagt in het bijbehorende examen (per module wordt een afzonderlijk examen voorzien).

# PROGRAMMA

## 1. CURSUS R VOOR GEVORDERDEN

R is een krachtige en vrij beschikbare programmeertaal gericht op dataverwerking en analyse. De grote kracht van R ligt in de uitgebreide mogelijkheden voor automatisering van het analyseproces en de talloze uitbreidingen in de vorm van R packages. Om de kracht van R goed te kunnen toepassen, moet men echter een goed begrip hebben van hoe R functioneert als programmeertaal. In deze cursus gaan we daar dieper op in.

In een eerste les overlopen we RStudio en de mogelijkheden die het biedt om je werk te organiseren. We kijken naar het gebruik van projecten, de verschillende opties en introduceren RMarkdown documenten.

In een tweede les bekijken we object types, hun eigenschappen en toepassingen. De focus ligt op een optimaal begrip van principes zoals vectorisatie en recycling.

In een derde les bekijken we functies. We gaan dieper in op de verwerking van argumenten en het toevoegen van foutmeldingen en waarschuwingen. Dit vormt de basis om in een vierde les dieper in te gaan op geavanceerde manipulatie van datasets. Hier komen zowel transformaties van variabelen als van volledige datasets aan bod.

De mogelijkheden voor grafische weergave van de data aan de hand van R en het pakket `ggplot2`, worden tijdens de vijfde les besproken.

Al deze onderwerpen worden gecombineerd in de oefeningentles op het einde van de cursus. Hierin werken we een voorbeeldanalyse uit inclusief automatische rapportering met RMarkdown.

Deze cursus bouwt voort op de introductie tot R uit het basispakket statistiek. Een goede voorkennis van R is vereist. Basisfuncties voor inlezen van data, dataverwerking en grafische weergave worden verondersteld gekend en enige vertrouwdheid met de RStudio omgeving is wenselijk.

**Lesgevers:** Joris Meys en Annelies Tourny

**Data:** 30 januari, 6, 20 en 27 februari, 6 en 13 maart 2018

## 2. NIET-PARAMETRISCHE METHODEN

Niet-parametrische methoden worden gekenmerkt door hun algemene geldigheid, zonder dat distributieve veronderstellingen (zoals normaliteit) over de data moeten gemaakt worden.

In de eerste les worden enkele klassieke niet-parametrische testen besproken. De Wilcoxon en de Mann-Whitney rank testen zijn de niet-parametrische alternatieven voor de t-testen, terwijl de Kruskal-Wallis test de niet-parametrische tegenhanger is van de F-test in een ANOVA. Deze testen maken gebruik van de asymptotische approximatie of van de exacte permutatiedistributie.

In de tweede les bespreken we hoe we volledige distributies met elkaar kunnen vergelijken. Op die manier kunnen varianties en scheefheden vergeleken worden. In een niet-parametrische statistische analyse is deze stap van gemiddelden naar volledige distributies eenvoudig te zetten.

In de derde les worden een aantal niet-parametrische testen besproken voor meer complexe studies (factoriële proeven met meerdere factoren en blokkenproeven), terwijl in de vierde les, de niet-parametrische tegenhangers van de lineaire regressieanalyse worden toegelicht, zoals smoothers (bv. LOESS en splines) en GAMS (generalized additive models). Vermits deze minder restrictief zijn, zijn ze breder toepasbaar.

In de vijfde les komt een algemene niet-parametrische methode aan bod voor het bekomen van approximatieve varianties en betrouwbaarheidsintervallen: de bootstrap. Deze methode wordt geïllustreerd aan de hand van enkele voorbeelden zoals regressie en correlatiecoëfficiënt.

Aangezien de oefeningen met het pakket R worden gemaakt is een grondige kennis van het gebruik van dit pakket vereist. Verder is er ook een grondige kennis van basisstatistiek vereist (t-toetsen, ANOVA en regressie).

**Lesgever:** Jan De Neve

**Data:** 17 en 24 april, 8, 15, 22 en 29 mei 2018

## 3. MULTIVARIATE METHODEN

Deze module handelt over enkele van de meest gebruikte multivariate statistische analysetechnieken, die dicht aanleunen bij datamining (clustering, classificatie, ...). Veelal is men niet in staat om een studie statistisch optimaal op te zetten en wordt men geconfronteerd met observationele studies. Men dient hierin een structuur te herkennen om tot een interpretatie en een besluit te komen.

Een eerste techniek waarmee de dimensionaliteit van een dataset gereduceerd kan worden is de principale componentenanalyse. Er wordt gezocht naar de belangrijkste (combinatie van) variabelen. Een stap verder is de factoranalyse. Dit soort analyses komt zowel voor in industrie als in marktgerichte diensten.

De canonische correlatieanalyse, die op dezelfde principes gebaseerd is, is een multivariate methode die gebruikt kan worden om verbanden op te sporen tussen twee multivariate subsets. Vervolgens wordt de clusteranalyse besproken. Deze heeft o.a. tot doel groepen te onderscheiden in een multivariate gegevensset.

Indien er reeds meer a-priori kennis is omtrent de groepen, kan een discriminantanalyse of een classificatieboom toegepast worden om een criterium te bepalen dat gebruikt kan worden om nieuwe observaties in één van de groepen onder te verdelen.

Aangezien de oefeningen met het pakket R worden gemaakt is een grondige kennis van het gebruik van dit pakket vereist.

**Lesgever:** Karel Vermeulen

**Data:** 18 en 25 september, 2, 9, 16 en 23 oktober 2018

**MEER INFO EN INSCHRIJVEN**

[www.ugain.ugent.be/stat](http://www.ugain.ugent.be/stat)

# PRAKTISCH

## PRIJS

Deze omvat lesgeld, hand-outs, frisdranken, koffie en broodjes. Betaling geschiedt na ontvangst van de factuur. Alle facturen zijn betaalbaar dertig dagen na dagtekening. Alle vermelde bedragen zijn vrij van BTW.

Module 1: Cursus R voor gevorderden	€ 900
Module 2: Niet-parametrische methoden	€ 900
Module 3: Multivariate methoden	€ 900
<b>Twee modules</b>	<b>5% korting op de modulaire prijzen</b>
<b>Drie modules</b>	<b>10% korting op de modulaire prijzen</b>

## KORTING

- Indien minstens één deelnemer van een bedrijf inschrijft voor de volledige opleiding wordt voor alle bijkomende gelijktijdige inschrijvingen van hetzelfde bedrijf een korting van 20% verleend. Facturatie geschiedt dan d.m.v. een gezamenlijke factuur.
- 10% korting op de in de tabel vermelde prijzen voor leden AIG en VBIG.
- Aangepaste prijzen voor personeel van UGent en geassocieerde hogescholen.
- Kortingen zijn niet cumuleerbaar.

## ANNULERING

Raadpleeg onze annulatievoorwaarden op [www.ugain.ugent.be/annulatievoorwaarden](http://www.ugain.ugent.be/annulatievoorwaarden)

## KMO-PORTEFEUILLE

Universiteit Gent aanvaardt betalingen via de KMO-portefeuille ([www.kmo-portefeuille.be](http://www.kmo-portefeuille.be); gebruik autorisatiecode DV.0103194).

## TIJDSTIP EN LOCATIE

- De lessen worden gegeven van **16u30 tot 20u**, in 2 delen, gescheiden door een broodjesmaaltijd en vinden plaats aan de **Universiteit Gent, UGain, Technologiepark 904, 9052 Zwijnaarde**.
- Data onder voorbehoud van wijzigingen om onvoorziene omstandigheden.

## HANDBOEKEN

De opleiding wordt ondersteund door de handboeken

- 'R for Dummies' van Andrie De Vries en Joris Meys: € 27,86 (incl. BTW) (optioneel bij module 1)
- 'Multivariate Statistical Methods' van Bryan F.J. Manly: € 44,96 (incl. BTW) (optioneel bij module 3)

Deze handboeken zijn optioneel. Deze zijn niet inbegrepen in de deelnameprijs en worden apart gefactureerd door de boekhandel.

## MEER INFO EN INSCHRIJVEN

[www.ugain.ugent.be/stat](http://www.ugain.ugent.be/stat)

## ORGANISATIE

Universiteit Gent  
UGain (UGent Academie voor Ingenieurs)  
Technologiepark 904, 9052 Zwijnaarde  
Tel: +32 9 264 55 82, Fax: +32 9 264 56 05  
E-mail: [ugain@ugent.be](mailto:ugain@ugent.be)



UNIVERSITEIT  
GENT

 FACULTEIT  
BIO-INGENIEURSWETENSCHAPPEN

 FACULTEIT INGENIEURSWETENSCHAPPEN  
EN ARCHITECTUUR