

# Praktijkgerichte Statistiek

Een praktische opleiding voor alle gebruikers van statistische methodes

## Postacademische opleiding

### Module 0: Inleiding S-plus

3 oktober 2002

### Module 1: Basiscursus Statistiek

8, 15 en 22 oktober, 5, 12 en 19 november 2002

### Module 2: Regressie-analyse

14, 21 en 28 januari, 4, 11 en 18 februari 2003

### Module 3: Multivariate Methodes

25 februari, 6, 11, 18 en 25 maart en 2 april 2003

## Wetenschappelijke coördinatie:

Prof. dr. Jean-Pierre Ottoy

Vakgroep Toegepaste Wiskunde, Biometrie en Procesregeling

Wij aanvaarden de  
opleidingscheques van  
het Vlaams Gewest

Vernieuwde uitgave



INSTITUUT VOOR  
PERMANENTE  
VORMING

50% theorie en 50% oefeningen  
met **S-PLUS<sup>®</sup> home** in elke module



Dit programma laat toe een getuigschrift van de Universiteit Gent te verwerven.

## **Het Instituut voor Permanente Vorming**

Het Instituut voor Permanente Vorming (IVPV) van de Universiteit Gent berust op een continu samenwerkingsverband tussen de universiteit en de bedrijfswereld. De hoeksteen van de vormingsprogramma's is de overdracht van basiskennis en bedrijfspraktische knowhow, teneinde de deelnemers in staat te stellen technologische vernieuwingen in hun omgeving toe te passen.

*Prof. dr. ir. Luc Boullart*  
Directeur

*Dr. ir. Erik Dejonghe*  
Voorzitter Stuurgroep

## **Getuigschrift van Postacademische Opleiding van de Universiteit Gent**

Dit programma is een onderdeel van de post-academische opleidingen van de Universiteit Gent.

De aanwezigheid tijdens de sessies en de evaluatie aan het einde van de opleiding bepaalt de facto of de deelnemer slaagt. Er is, per module, een afzonderlijk examen voorzien. Na afloop zal aan de geslaagde deelnemers die minstens twee van de drie modules (1, 2 en 3) doorlopen hebben en met succes examen aflegden, een Getuigschrift van Postacademische Opleiding van de Universiteit Gent uitgereikt worden.

## **Doelpubliek**

De opleiding richt zich tot alle personen in de bedrijfswereld, de dienstensector, de overheid, de sector gezondheidszorg en het onderwijs die, in ondersteunende disciplines, gebruik maken van statistische methodes en modellen. De opleiding is toegankelijk voor eenieder 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 cursist de modulekeuze aanpassen aan zijn interesseprofiel en specifieke voorkennis.

## **Vormende waarde**

De doelstelling van deze opleiding is een praktisch inzicht te geven in veelgebruikte en nuttige statistische methodes voor de bedrijfswereld. Een probleem bij het analyseren van data is de keuze die men dient te maken tussen het ruime aanbod aan statistische methodes. Deze keuze berust immers op een grondige kennis van de voorwaarden waaronder de statistische methode ontwikkeld is. Eens een verantwoorde methode geselecteerd is, kan de uitvoering van de analyse worden aangevat. Dit stadium is de jongste jaren sterk vereenvoudigd wegens de ruime verspreiding van statistische software pakketten (S-PLUS, SAS, SPSS, Statistica, ...). Tenslotte dienen de resultaten van de analyse geïnterpreteerd te worden. Ook hierbij is een grondige kennis van de kracht en de tekortkomingen van de gebruikte statistische techniek onontbeerlijk.

In deze lessenreeks zal op deze drie fasen worden ingegaan en tijdens een aantal begeleide oefeningen zullen de cursisten zelf gegevens kunnen analyseren en interpreteren. Er zal ook nadruk worden gelegd op een fase waaraan maar al te vaak te weinig aandacht wordt geschonken: de opzet van een studie. Deze is meestal onontbeerlijk om tot een geldige statistische analyse en een verantwoord besluit te komen.



De opleiding bestaat uit drie modules: een basiscursus statistiek, een cursus regressie-analyse en een cursus over multivariate methodes. Een onderscheid tussen de modules kan worden gemaakt op basis van enerzijds de methode van data-acquisitie en anderzijds de probleemstelling. De technieken uit de eerste module zijn geschikt voor de analyse van experimentele data, deze uit de tweede en derde module worden vooral gebruikt voor observationele gegevens, respectievelijk met en zonder apriori status van predictor of respons.

De modules zijn zodanig opgebouwd dat cursisten zich voor elke module afzonderlijk (uitgezonderd voor module 0) of voor de volledige cursus kunnen inschrijven.

## Module 0: Inleiding S-PLUS

Deze stoomcursus is bedoeld om beginnende S-PLUS gebruikers op weg te helpen in het S-PLUS data-analyse en visualisatie systeem. Men behandelt onder andere het bewerken van gegevens en het maken van grafieken.

**Duur:** 1 lesavond: 100% oefeningen

**Lesgevers:** Prof. Dr. J.P. Ottoy, Dr. Ir. O. Thas en lic. E. Deschepper

**Lesmateriaal:**

- Een persoonlijk exemplaar van het handboek 'An introduction to S-PLUS for Windows' van Longhow Lam (Uitg. Candiensten)
- Softwarelicentie S-plus

**Data module 0:** 3 oktober 2002

## Module 1: Basiscursus Statistiek

In deze module worden vooreerst de basisbegrippen, de concepten en de algemene gedachtegang aangebracht. Dit vormt een belangrijke basis om tot een zinvolle en eenduidige interpretatie van een statistische analyse te kunnen komen. Dit houdt in dat er gestart wordt met een herhaling van de meest voorkomende distributies (normale, t, F, ...), de principes van een statistische toets, de constructie van hypothesen, steekproeven, enz. Eens deze begrippen gekend zijn, komen de klassieke statistische technieken aan bod. Voor het vergelijken van gemiddelden zijn de t- en de F-toets de meest gekende. De variantie-analyse met 1 of met meerdere factoren vormen een directe uitbreiding hierop. Deze zijn gebaseerd op een statistisch model, wat een belangrijke bouwsteen is in de toegepaste statistiek. Deze modellen zijn eveneens onontbeerlijk in de ontwerp-fase van een experiment of studie: door het doordacht opzetten van een experiment kan met een minimale kost maximale informatie bekomen worden. In de lessen zal de nadruk gelegd worden op het probleemoplossend vermogen van de statistische technieken, de voorwaarden voor toepassing, de interpretatie en de opzet van experimenten. Reële voorbeelden zullen uitgewerkt worden.

**Duur:** 6 lesavonden: 50% theorie en 50% oefeningen

**Lesgevers:** Prof. Dr. J.P. Ottoy, lic. E. Deschepper en lic. H. Wouters

**Lesmateriaal:**

- Een persoonlijk exemplaar van het handboek 'Applied Linear Statistical Models' van Neter, Wasserman en Kutner (MacGraw Hill)
- Handleiding S-PLUS van J.P. Ottoy, O. Thas, E. Deschepper en H. Wouters
- Kopies van de transparanten van de lesgevers
- Softwarelicentie S-plus (ook op eigen PC) gedurende de volledige cursus

**Data module 1:** 8, 15 en 22 oktober, 5, 12 en 19 november 2002

**Doelpubliek:** R&D, marketing en productie: allen die geregeld in aanraking komen met data en inzicht in statistische dataverwerking en proefopzet willen verkrijgen.

## Module 2: Regressie-analyse

De regressie-analyse is één van de bekendste en meest gebruikte statistische technieken, zowel in onderzoekinstellingen als in de industrie. Niettegenstaande de techniek redelijk eenvoudig is in gebruik, worden er in de praktijk toch veel problemen ondervonden: aan welke veronderstellingen moeten de gegevens voldoen? Wat als er niet aan deze veronderstellingen voldaan is? Hoe worden betrouwbaarheids- en predictieintervallen geïnterpreteerd? Indien er meerdere predictoren zijn, hoe wordt het beste regressiemodel dan gevonden? Kan een discrete variabele als predictor gebruikt worden? Deze en vele andere problemen worden in deze lessenreeks uitvoerig behandeld, steeds met de nadruk op de toepassingen. In de eerste twee lessen wordt de enkelvoudige lineaire regressie behandeld. In dit meest eenvoudige geval is er slechts 1 predictor. Aan de hand van voorbeelden wordt er aangeleerd hoe de voorwaarden eenvoudig nagegaan kunnen worden en hoe hieraan eventueel verholpen kan worden (bv. transformaties). Hiertoe wordt de nadruk gelegd op grafische voorstellingen van de data en van het regressiemodel, zodat de bevindingen eenvoudig gecommuniceerd kunnen worden. Tevens wordt aangebracht hoe de statistische significantie van de predictor nagegaan wordt (t-toets), hoe betrouwbaarheidsintervallen berekend en geïnterpreteerd dienen te worden, en wat het verschil is met predictieintervallen indien het model voor predictiedoeleinden gebruikt wordt. In de volgende twee lessen komt de meervoudige lineaire regressie aan bod. Dit zijn regressiemodellen met meerdere predictoren. Dezelfde topics als bij de enkelvoudige lineaire regressie worden hier behandeld, maar het zal snel duidelijk zijn dat er hier extra problemen bij komen kijken. We vermelden hier bijvoorbeeld het probleem van modelbouw: hoe kies je uit een reeks van predictoren de beste uit? En wat zijn de gevolgen van te veel of te weinig predictoren geselecteerd te hebben? Het begrip correlatie tussen de predictoren zal hier een belangrijke rol spelen.

Vervolgens wordt er een les besteed aan niet-lineaire regressiemodellen. Ook gaat er aandacht uit naar enkele meer recente regressietechnieken zoals regressiebomen en niet-parametrische regressie.

Alle concepten en technieken zullen met duidelijke voorbeelden geïllustreerd worden. In de oefeningenlessen krijgen de cursisten de kans om onder begeleiding zelf regressiemodellen te bouwen en te analyseren.

**Duur:** 6 lesavonden: 50% theorie en 50% oefeningen

**Lesgevers:** Prof. dr. J.P. Ottoy, Dr. ir. O. Thas en lic. H. Wouters

**Lesmateriaal:**

- Een persoonlijk exemplaar van het handboek 'Applied Linear Statistical Models' van Neter, Wasserman en Kutner (MacGraw Hill)
- Handleiding S-PLUS van J.P. Ottoy, O. Thas, E. Deschepper en H. Wouters
- Kopies van de transparanten van de lesgevers
- Softwarelicentie S-plus (ook op eigen PC) gedurende de volledige cursus

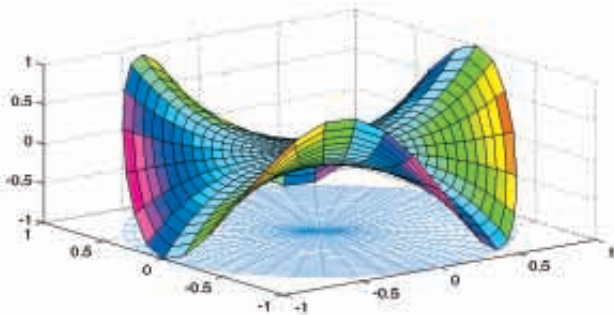
**Data module 2:** 14, 21 en 28 januari, 4, 11 en 18 februari 2003

**Doelpubliek:** R&D, marketing en productie: allen die geregeld in aanraking komen met data en die inzicht in statistische dataverwerking en proefopzet willen verkrijgen door middel van modelbouw.

## Module 3: Multivariate Methodes

In deze lessenreeks worden meer exploratieve statistische technieken uiteengezet. Veelal is men niet in staat een studie statistisch optimaal op te zetten en wordt men geconfronteerd met observationele studies. Men dient hierin vervolgens 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 voor in zowel industriële afdelingen als in marktgerichte diensten. De canonische correlatie-analyse, 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 tot doel groepen te onderscheiden in een multivariate gegevensset en kan eventueel een tweede stap vormen in de analyse, na eerst een principale componentenanalyse of factoranalyse uitgevoerd te hebben om de dimensionaliteit te reduceren.

Indien er reeds meer apriori kennis is omtrent de groepen, kan een discriminantanalyse toegepast worden om een criterium te bepalen dat gebruikt kan worden om nieuwe observaties in één der groepen onder te verdelen. Geen van deze tools is rechtstreeks gebaseerd op de technieken die in de eerste module besproken werden, maar een basiskennis ervan is aangeraden om een duidelijk onderscheid te kunnen maken tussen de doelstellingen van de methodes. Dit onderscheid zit voornamelijk in het verschil tussen analyse van experimentele of observationele studies. De eerste soort leent zich tot het toepassen van bevestigende statistische methodes: hypothese toetsen. Op de tweede soort daarentegen worden eerder exploratieve en hypothese genererende technieken toegepast.



**Duur:** 6 lesavonden: 50% theorie en 50% oefeningen

**Lesgevers:** Dr. Ir. O. Thas en lic. E. Deschepper

- Lesmateriaal:**
- Een persoonlijk exemplaar van het handboek 'Multivariate Statistical Methods' van Bryan F.J. Manly, Chapman & Hall.
  - Handleiding S-PLUS van J.P. Ottoy, O. Thas, E. Deschepper en H. Wouters
  - Kopies van de transparanten van de lesgevers
  - Softwarelicentie S-plus (ook op eigen PC) gedurende de volledige cursus

**Data module 3:** 25 februari, 6, 11, 18 en 25 maart en 2 april 2003

**Doelpubliek:** R&D, marketing en productie: allen die geconfronteerd worden met grote multivariate datasets waarin een structuur dient gezocht te worden om tot het vormen van conclusies te komen.

## DEELNEMINGS- EN INLICHTINGSFORMULIER

Terug te sturen bij voorkeur minstens 1 week vóór aanvang van de eerste les die u wenst te volgen.

Naam: \_\_\_\_\_

Voornaam: \_\_\_\_\_

Functie: \_\_\_\_\_

Onderneming: \_\_\_\_\_

Adres: \_\_\_\_\_

Telefoon: \_\_\_\_\_

Fax: \_\_\_\_\_

E-mail: \_\_\_\_\_

Bedrijvigheid: \_\_\_\_\_

Privé-adres: \_\_\_\_\_

**Ik schrijf in voor het vormingsprogramma:**

### 'PRAKTIJKGERICHTE STATISTIEK'

- Module 0:** Inleiding S-PLUS (kan enkel in combinatie met minstens één andere module gevolgd worden)
- Module 1:** Basiscursus Statistiek
- Module 2:** Regressie-analyse
- Module 3:** Multivariate Methodes
- Modules 0 t.e.m. 3**
- Modules 1 t.e.m. 3**
- Ik wens het Getuigschrift van de Universiteit Gent te behalen
- Ik betaal . . . . . Euro d.m.v. opleidingscheques
- Ik wens blijvend geïnformeerd te worden over de vormingsprogramma's van het IVPV

**Facturatie-adres**

Naam: \_\_\_\_\_

Adres: \_\_\_\_\_

BTW nr. \_\_\_\_\_

Datum: \_\_\_\_\_

Handtekening: \_\_\_\_\_



Deze gegevens blijven strikt binnen het IVPV en zullen worden gebruikt om u op de hoogte te houden van latere programma's van permanente en postacademische vorming.

## De Lesgevers

**Prof. dr. Jean-Pierre Ottoy**, wetenschappelijk coördinator, is hoogleraar aan de vakgroep Toegepaste Wiskunde, Biometrie en Procesregeling van de RUG en verantwoordelijk voor het onderwijs in de wiskunde, informatica en Statistische Dataverwerking aan de Bio-ingenieurs. Het wetenschappelijk onderzoek is gesitueerd in het domein van de biomathematica, meer in het bijzonder de toegepaste statistiek, waarbij de modelvorming en de kwantitatieve analyse een vooraanstaande rol spelen.



**Dr. Ir. Olivier Thas** is doctor-assistent aan de vakgroep Toegepaste Wiskunde, Biometrie en Procesregeling. Hij verzorgt o.a. de basiscursus statistiek voor verscheidene masteropleidingen en de cursussen Multivariate Statistiek en Proefopzet voor de studenten Bio-ingenieur. Tevens is hij actief in de statistische dienstverlening, zowel binnen als buiten de universiteit. Zijn onderzoek kadert voornamelijk binnen de niet-parametrische en multivariate statistiek.

**Lic. Ellen Deschepper** en **Lic. Heidi Wouters** zijn assistent aan de vakgroep Toegepaste Wiskunde, Biometrie en Procesregeling, waar zij betrokken zijn bij de ondersteuning van de vakken Statistische Dataverwerking, proefopzet en multivariate statistiek.

Beiden verlenen ondersteuning aan studenten en onderzoekers voor de statistische verwerking van experimenten in het domein van de toegepaste biologische wetenschappen. Ze zijn tevens betrokken bij het wetenschappelijk onderzoek, geleid door Prof. Ottoy. Hun onderzoek kadert in de milieugerichte en niet-parametrische statistiek.

Frankeren  
als brief

Universiteit Gent

Instituut voor Permanente Vorming

T.a.v. Els Van Lierde

Technologiepark 913

9052 Gent - Zwijnaarde



## Praktische inlichtingen

Het programma is modulair opgebouwd. Elke module kan apart gevolgd worden, behalve module 0 die moet gecombineerd worden met één of meerdere andere modules.

Behalve op 3 oktober, 6 maart en 2 april worden alle sessies op dinsdagavond gegeven, telkens van 18u tot 21u30, in twee lessen, gescheiden door een koffiepauze/broodjes-lunch.

Alle lessen worden gegeven aan de Universiteit Gent, Instituut voor Permanente Vorming, Gebouw Magnel, Technologiepark 904, 9052 Zwijnaarde

Bijkomende inlichtingen krijgt u op het secretariaat:

**Universiteit Gent, Instituut voor Permanente Vorming**

Els Van Lierde

Technologiepark 913

9052 Zwijnaarde

Tel: +32 9 264 55 82

Fax: +32 9 264 56 05

E-mail: [ivpv@rug.ac.be](mailto:ivpv@rug.ac.be)

<http://www.ivpv.rug.ac.be/statistiek>

**Daar de RUG erkend is als opleidingsverstrekker in het kader van de opleidingscheques van het Vlaams Gewest, kan uw bedrijf tot de helft besparen op de deelnemingsprijs van deze opleiding.**

**Voor meer informatie en bestelling van uw opleidingscheques zie [www.vlaanderen.be/opleidingscheques](http://www.vlaanderen.be/opleidingscheques)**

Indien u deze folder meerdere malen mocht ontvangen, verzoeken wij u vriendelijk deze aan uw collega's te bezorgen en ons dit te melden via e-mail.

## Annulering

Annulering is mogelijk onder de volgende voorwaarden:

- gelieve steeds schriftelijk te bevestigen (per brief of fax)
- bij annulering van de inschrijving 10 dagen of meer vóór de aanvang van het programma is een vergoeding verschuldigd van 25% van de deelnemingsprijs
- bij annulering minder dan 10 dagen vóór de aanvang van het programma is de volledige deelnemingsprijs verschuldigd.

## Deelnemingsprijs

De deelnemingsprijs omvat het lesgeld, een handboek, een handleiding S-PLUS voor de oefeningen, kopies van de transparanten, gebruik van de leslokalen, oefeningen op PC met S-PLUS, S-PLUS licentie om te oefenen op eigen PC, frisdranken, koffie en broodjes.

Betaling geschiedt na ontvangst van de factuur.

Alle facturen zijn contant betaalbaar dertig dagen na dagtekening. Inschrijving is mogelijk per module of voor de volledige cursus. Alleen module 0 kan niet afzonderlijk gevolgd worden. Alle vermelde bedragen zijn vrij van BTW.

Indien minstens één deelnemer van een bedrijf inschrijft voor de volledige opleiding (modules 0 t.e.m. 3 of 1 t.e.m. 3), wordt voor alle bijkomende gelijktijdige inschrijvingen van hetzelfde bedrijf, per module of volledig pakket, een korting van 20% verleend. Facturatie geschiedt dan d.m.v. een gezamenlijke factuur.

Inschrijving gebeurt door terugzending van het aangehecht deelnemingsformulier of via de website.

Module 0: 1 avond € 75

Module 1: 6 avonden € 1.000

Module 2: 6 avonden € 1.000

Module 3: 6 avonden € 1.000

Modules 0 t.e.m. 3: € 2.525

Modules 1 t.e.m. 3: € 2.500