Cursussen Statistiek
Courses in Statistics

2007-2008

PERMANENTE VORMING IN DE WETENSCHAPPEN

CONTINUING EDUCATION IN SCIENCE

UNIVERSITEIT GENT

inhoud/contents

adres

Centrum voor Statistiek

ADRES Krijgslaan 281 - S9, 9000 Gent
URL http://www.cvstat.UGent.be

Instituut voor Permanente Vorming in de Wetenschappen

ADRES Krijgslaan 281 - S3, 9000 Gent
TELEFOON 09 264 44 26 FAX 09 264 85 90
E-MAIL Isabel.DeZutter@UGent.be
URL www.ipvw-ices.UGent.be

Center for Statistics
Centrum voor Statistiek

Institute for Continuing Education in Science

Instituut voor Permanente Vorming in de Wetenschappen

logo opleidingscheques

Wij aanvaarden de opleidingscheques van de Vlaamse Gemeenschap.

Deze reeks includeert cursussen die zijn opgenomen in de Doctoral Schools programma's.

voorwoord

Statistiek is de laatste decennia onmisbaar geworden in heel wat takken van de wetenschap. Denk maar aan sociologen die oorzaken van gedrag natrekken, artsen en biowetenschappers die DNA-chips analyseren of de effectiviteit van nieuwe interventies evalueren, bio-ingenieurs die de kwaliteit van het milieu bewaken, industriëlen die de productiekwaliteit bijsturen, economen die financiële tijdreeksen bestuderen,…
Met de komst van gebruiksvriendelijke software wordt het opslaan en manipuleren van data haast kinderspel. Veel minder evident is het om relevante en kwaliteitsvolle gegevens te verzamelen, om efficiënt informatie te onttrekken en niet misleid te worden door naïeve conclusies. Een techniek en de interpretatie van haar resultaten hangen immers fundamenteel samen met het design en de implementatie van de studie, vaak ook met bijkomende onderstellingen over een complexe datastructuur.

Het Centrum voor Statistiek van de Universiteit Gent, in samenwerking met het Instituut voor Permanente Vorming van de Faculteit Wetenschappen (IPVW), organiseert daarom elk jaar cursussen die inspelen op de noden van gebruikers van statistische methoden. Het aanbod richt zich vooral op het verwerven van inzicht in de basis van het statistisch onderzoek. Practica's op pc stellen de kandidaat in staat om ook al doende te leren. Het doelpubliek bestaat uit professionelen en onderzoekers met een academische vorming. Of u nu kennis wil opfrissen, op de hoogte wil blijven van recente ontwikkelingen, of belangstelling heeft voor een nieuw onderzoeksdomein, deze formule wil u in staat stellen om gericht kennis en vaardigheden op te doen. De nieuwe inzichten zullen uw bedrijf en uw onderzoek de extra voorsprong geven die het verdient.
Ook de Vlaamse regering ziet zo'n training als een troef voor haar economisch beleid. Professionelen kunnen genieten van financiële steun onder de vorm van opleidingscheques. Meer informatie over dit 'stimulerend middel' vindt u op de websites www.vdab.be/opleidingscheques en www.opleidingscheques.be en via de link op de IPVW-website: www.ipvw-ices.UGent.be.

Tussen november 2007 en april 2008 wordt het pakket basismodules aangeboden, waarin statistische kennis gradueel wordt opgebouwd. Vooraf, in oktober 2007, geven twee modules een introductie tot het gebruik van de statistische software SPSS en SAS, waarvoor de Universiteit Gent een licentie heeft. Naar jaarlijkse gewoonte wordt dit programma aangevuld met een aantal gespecialiseerde cursussen. In november 2007 wordt de zeer succesvolle "Short Course on Survival Analysis" hernomen. In januari 2008 wordt voor het eerst een driedaagse over "Missing Data" georganiseerd. Het programma wordt afgesloten in april 2008 met de module "Data Mining".

Verschillende cursussen worden in het Engels gedoceerd. Ons doel blijft om iedereen maximaal te laten genieten van deze kans tot bijkomende opleiding. Het geheel verloopt in een gemoedelijke sfeer met ruime mogelijkheid tot interactie met de docenten.
De IPVW-activiteiten waaraan een examen is verbonden worden ook opgenomen in het programma van de 'Doctoral Schools' die, met ingang van academiejaar 2007-2008, de UGent-doctoraatstudent zullen ondersteunen bij zijn/haar onderzoek en opleiding.

Verder brengen we graag onder uw aandacht dat aan de Universiteit Gent ook een meer doorgedreven Master opleiding in Statistische Data-Analyse loopt, die wetenschappers uit diverse disciplines een grondige vorming aanbiedt in de methodes van de toegepaste statistiek. Meer info hierover vindt u op de website van het Centrum voor Statistiek: www.cvstat.ugent.be.

We hopen dat u het nieuwe aanbod kan smaken en wensen u alvast een leerrijk en productief jaar toe!

introduction

Statistics has become indispensable in many branches of science. Sociologists and psychologists searching to explain behaviour, biologists analysing DNA-chips, clinicians evaluating new interventions, bio-engineers monitoring the environment, managers performing quality control, economists studying time series, … they all rely on statistical methods. Today's user-friendly software allows anyone to store and manipulate data quite easily. It remains a challenge however, to gather relevant, high-quality data and retrieve information efficiently to draw accurate inference. Without training and due professionalism one runs a high risk of arriving at misleading conclusions. One must recognise how the appropriate statistical technique and justified interpretation depend fundamentally on the design and implementation of a study in combination with any assumptions about the data structure.

To meet the needs of users of statistical methods, the Centre for Statistics of Ghent University in co-operation with the Institute for Continuing Education of the Faculty of Science (ICES) organises yearly a series of courses. The goal is to train users of statistical software, providing them with insight in the basics of statistical research. Practical sessions on PC's allow participants to obtain this through hands-on experience. Our courses are aimed at professionals and participants with an academic training, who wish to refresh their knowledge, keep it up to date or discover new areas of research. The program is designed to offer specific knowledge and skills through separate modules. The new insights will give your company or research the extra edge it needs. The Flemish Community regards continuing training as an important element of its economic policy. Professionals are granted financial support through the government's introduction of training vouchers. More information about this stimulating initiative can be found on the websites: www.opleidingscheques.be and www.vdab.be/opleidingscheques (both in Dutch) or via the link on the ICES-website: www.ipvw-ices.UGent.be (English).

From November 2007 to April 2008, the program offers a classic series of basic modules which gradually build up statistical knowledge and techniques. Leading into this, two courses in October 2007 introduce the statistical software SPSS and SAS, for which Ghent University has a campus licence. Every year, the program offers an additional number of more specialised courses. In November 2007 the highly successful course on "Survival Analysis" will be taken up again. It deals with the planning, execution and interpretation of time-to-event data. In January 2008 a three-day course is offered on "Missing Data", i.e. data that were intended to be collected, but were not. Finally, in

April 2008 the program finishes with a course on "Data Mining", which deals with large, high-dimensional data. Several modules are taught in English to give international candidates the opportunity to participate. Classes take place in a pleasant atmosphere with ample opportunity to interact with the lecturers.
ICES-activities that include an exam will also be incorporated in the program of the Doctoral Schools, which will support UGent doctoral students, as from academic year 2007-2008, with their research and training.

The University of Ghent also presents a Masters program in Statistical Data-Analysis. This program offers a more thorough training in the methods of practical statistics to scientists in diverse areas. For more information visit our website www.cvstat.ugent.be.

Hoping the new program meets your expectations, we wish you an enjoyable and productive learning experience!

MODULE 1 – Inleiding tot SPSS

Dhr. Kris Erauw

Beschrijving

We leven in een kennismaatschappij. Heel veel mensen verzamelen gegevens of willen bepaalde ideeën met onderzoeksbevindingen ondersteunen. Denk aan de jongeren in een stedelijke jeugdraad die in hun jeugdwerkbeleidsplan de geformuleerde beleidslijnen moeten ondersteunen met onderzoeksbevindingen; of aan de voorzitter van een oudercomité die de standpunten van ouders op een wetenschappelijk verantwoorde manier wil bevragen. Het inzamelen en opslaan van al die gegevens is vaak niet zo evident als het lijkt. Zeker niet als het de bedoeling is de gegevens later op een professionele manier te verwerken.
Deze lessenreeks is er op gericht data in een bruikbare vorm te verzamelen, de ingezamelde data in SPSS op te slaan en met SPSS de eerste beschrijvende statistieken te produceren.
De verschillende lessen in de reeks zijn ervaringsgericht opgevat. De deelnemers worden met een aantal problemen geconfronteerd waarna mogelijke oplossingen besproken en gedemonstreerd worden.

Les 1
Data en dataverzameling: data in SPSS invoeren en definiëren, data uit andere programma's importeren en gebruiken.

Les 2
Elementaire bewerkingen: samenvattende statistieken en voorstellingen genereren, variabelen herschrijven en combineren, databestanden bewerken en combineren.

Les 3
Gemiddelden vergelijken: grafische voorstelling van gemiddelden, t-toetsen en one-way variantie-analyse.

Les 4
Meer uit je databestand halen: de eerste stappen in lineaire regressie.

Data

1, 2, 4, en 5 oktober 2007 telkens van 17u tot 20u.

Plaats

PC-klas 1 van de Faculteit Psychologische en Pedagogische Wetenschappen, Henri Dunantlaan 1, Gent.

Doelpubliek

Deze practica zijn bedoeld voor alle personen die gegevens inzamelen en/of opslaan, met de bedoeling deze statistisch te analyseren en te interpreteren.

Toelatingsvoorwaarden

Geen

Lesmateriaal

Documentatie- en oefeningenbundel.

Prijs

De deelnameprijs bedraagt 325 EUR voor deelnemers uit de private sector, 175 EUR voor UGent-personeelsleden en personeel uit de non-profit, social-profit en overheidssector. Een gereduceerde prijs van 125 EUR geldt voor doctoraatsstudenten. In deze prijs is o.a. het lesmateriaal ingesloten.

MODULE 2 – Introduction to SAS

Dr. Anita Prinzie

Course description

The amount of data stored in data warehouses grows exponentially every day. These high-dimensional and often noisy data are stored with the aim to eventually extract useful patterns and information on which strategies can be formed. However, the data as stored in a data warehouse are typically not suitable for immediate analysis. The data might be noisy, data might be stored across different tables or more rich information could be extracted by creating your own variables. Hence, before any information can be extracted and modeled, a first time-consuming task involves the manipulation of the data (missing-value imputation, creation of new variables) stored in multiple tables to obtain an analysis table appropriate for subsequent statistical analyses like regression analysis, analysis of variance or survival analysis. This course aims to empower you to manipulate huge data warehouses by learning the SAS programming language. Whereas other software packages like SPSS are offering limited capabilities to handle and manipulate high-dimensional data, with SAS language you are in total control of data manipulation. In addition to data manipulation, we will illustrate by hands-on practice how you can use SAS statistical procedures to subsequently analyse/model your data. Unlike the interactive SAS Enterprise Miner, the SAS language provides more flexibility to define model parameters and facilitates easy replication of analyses. After finishing this course, you should be able to perform data manipulation and basic statistical analyses using the SAS programming language.

Dates and venue

October 8, 9, 10 and 11, 2007 from 5.30 pm to 9.00 pm, at the Faculty of Science, Building S9, Campus Sterre, Krijgslaan 281, Ghent.

Target audience

This course will benefit investigators/data analysts from diverse areas in charge of analysing potentially high-dimensional data stored in a data warehouse. It is most suited to anyone working with high-dimensional data (either in terms of number of instances and/or number of variables). However, if you are working with low-dimensional data, you too might benefit from this course as mastering the SAS programming language delivers you more control to manipulate your data in order to create new variables and to replicate your work in a convenient way.

Course prerequisites

Participants should have experience in basic statistics at a level equivalent to the "Introductory Statistics" course of this program.

Course material

Copies of lecture notes.
Optional but highly recommended programming manual:
"The Little SAS Book: A Primer", Lora D. Delwiche,
Susan J. Slaughter (2003), 3rd ed.

Fees

The registration fee amounts to 325 EUR for participants of the private sector, 175 EUR for employees of Ghent University, the government, the non-profit and social-profit sector. A special rate of 125 EUR applies to doctoral students. The book is optional but recommended and can be bought at the additional cost of 50 EUR. Please indicate this clearly on the registration form.

MODULE 3 – Introductory Statistics. Basics of Statistical Inference

Prof. dr. Maria Ysebaert

Course description

This course aims to provide insight into basic statistical concepts with emphasis on practical applications. Mathematical formulae will be kept to a minimum. The theory and the methods of analysis will be extensively illustrated with examples relating to a wide variety of different fields.

We start with concise graphical and numerical descriptions of data obtained from observational or experimental studies. The most common and frequently used probability distributions of discrete and continuous variables will be presented. Statistical inference draws conclusions about a population based on sampled data. Chance variations are taken into account such that a level of confidence is attached to these conclusions. We present the reasoning behind significance tests for the comparison of observed data with a hypothesis, the validity of which we want to assess. We apply this procedure to data obtained either from one or from two populations. The correct use of the t-test will be discussed. Nonparametric methods are considered as a possible alternative in case the requirements of the t-test are not met. We cover the basic concepts of hypothesis testing for categorical data, including the chi-square test. Quite often the relationship between two variables, where the outcome of one variable is seen as depending on the value of the other, is the focus of scientific interest. We will give an introduction to linear regression analysis, where a regression line based on observations obtained in a sample describes this relation.

Dates and venue

November 6, 13, 20 and 27, December 4, 11 and 18, 2007 from 5.30 pm till 9.30 pm (each lecture is followed by a hands-on practical session) at the Faculty of Science, Building S9, Campus Sterre, Krijgslaan 281, Ghent.

Target audience

This course will benefit investigators from diverse areas, research scientists, clinical research associates, and, in general, anyone who comes in contact with data handling and who wants to acquire insight into basic statistical methods or who feels that his/her knowledge and practice of statistics needs refreshing. No extensive background in mathematics is required.

Exam

Participants can, if they wish, take part in an exam at a date which will be specified later. A certificate from Ghent University will be issued to participants with a university degree at the bachelors level or an equivalent degree upon succeeding in this test. As such this course can be incorporated in a doctoral training program.

Course prerequisites

The course is open to all interested persons.

Course material

Copies of lecture notes.

Recommended handbooks are:

Book 1: "Fundamentals of Biostatistics", Bernard Rosner, 6th ed. (2005), Thomson Learning.

Book 2: "Introduction to the Practice of Statistics", David S. Moore and George P. McCabe, 5th ed. (2005), W.H. Freeman and Co.

Fees

The registration fee amounts to 700 EUR for participants of the private sector, 300 EUR for employees of Ghent University, the government, the non-profit and social-profit sector. A special rate of 200 EUR applies to doctoral students. The book is optional and can be bought at the additional cost of 70 EUR. Please indicate this clearly on the registration form.

MODULE 4 – Survival Analysis

Prof. dr. David Harrington

Course description

This course will examine modern methods in the analysis of event time data, or survival analysis, with an emphasis on practical applications in medical research.

The course will begin with basic ideas in the non-parametric analysis of right censored data, including estimates of the hazard function, the Kaplan Meier estimate of a survival distribution and its standard error, the logrank test for comparing survival outcomes in two or more groups, and sample size calculations.

The course will then cover regression models for censored data, focusing primarily on the Cox proportional hazards model. Examples will be used to illustrate the methods for estimating models in SAS and S-Plus and for interpreting model coefficients. The regression modeling will cover the use of stepwise regression to choose parsimonious models, numerical and graphical methods to examine possible interactions, the use of time-dependent covariates, the analysis of data with left truncation, and model diagnostics to check the appropriateness of an estimated model.

The third unit of the course will cover special topics, such as the analysis of data with dependent competing risks, the analysis of clustered data with correlated cases, and the use of sequential designs in medical trials.

All methods will be illustrated with annotated computer output from SAS and S-Plus. A hands-on computer session will help you to practice the principles exposed in this course.

Dates and venue

November 21, 22 and 23, 2007 from 5.30 pm to 9.00 pm at the Faculty of Science, Building S9, Campus Sterre, Krijgslaan 281, Gent. The course includes a hands-on practical session.

Target audience

(Biomedical) researchers involved in the planning, execution and/or interpretation of time-to-event data. The latter arise in clinical trials of chronic diseases and the study of prospective cohorts in public health and epidemiology.

Exam

Participants can, if they wish, take part in an exam at a date which will be specified later. A certificate from Ghent University will be issued to participants with a university degree at the bachelors level or an equivalent degree upon succeeding in this test. As such this course can be incorporated in a doctoral training program.

Course prerequisites

Participants are expected to be familiar with linear regression.

Course material

Copies of lecture notes.

Fees

The registration fee amounts to 500 EUR for participants of the private sector and 200 EUR for employees of Ghent University and the non-profit sector. A special rate of 150 EUR applies to doctorandi.

MODULE 5 – Analysis of Variance

Prof. dr. Maria Ysebaert

Course description

Analysis of variance (ANOVA) is a statistical tool used in the comparison of means of a random variable in populations that differ in a characteristic (factor), e.g. treatment, age, sex, subject, etc. First, we cover one-way ANOVA, where only one factor is of concern. Depending on the type of the factor, the conclusions pertain to just those factor levels included in the study (fixed factor model), or the conclusions extend to a population of factor levels of which the levels in the study are a sample (random effects model). In two-way and multi-way ANOVA (populations differ in more than one characteristic), the effects of factors are studied simultaneously to obtain information about the main effects of each of the factors as well as about any special joint effects (factorial design). In nested designs, where each level of a second factor (mostly a random factor) occurs in conjunction with only one level of the first factor, analysis of variance enables us to extract the variability induced by the nested factor from the effects of the main factor. For correct analysis of the data in multi-way ANOVA, not only the linear model and the type of factor have to be considered but, also, the assumptions that must be satisfied.
In this course we will focus on correct execution of data analysis and understanding the results of this analysis. We will provide insight into the conclusions and pay attention to expressing these conclusions in a correct and understandable way. The different methods will be extensively illustrated with examples from scientific studies in a variety of fields.

Dates and venue

January 8, 15, 22 and 29, February 5, 12 and 19, 2008 from 5.30 pm to 9.30 pm (each lecture is followed by a hands-on practical session) at the Faculty of Science, Building S9, Campus Sterre, Krijgslaan 281, Ghent.

Target audience

This course will benefit investigators from a diversity of areas, who need to use statistical methods in the collection and handling of data in their research, in particular for assessing the effect of e.g. different treatments.

Exam

Participants can, if they wish, take part in an exam. A certificate from Ghent University will be issued to participants with a university degree at the bachelor level or an equivalent degree upon succeeding in this test. As such this course can be incorporated in the doctoral training program.

Course prerequisites

Participants are expected to have an active knowledge of the basic principles underlying statistical strategies, at a level equivalent to the 'Introductory Statistics' course of this program. In the first session, on January 8, these principles will be briefly reviewed. Participants who have recently followed the introductory course are exempt from that first session.

Course material

Copies of lecture notes.
Recommended handbook: "Applied Linear Statistical Models", Michael H. Kutner, Christopher J. Nachtsheim, John Neter and William Li, 5th ed. (2004), McGraw-Hill.

Fees

The registration fee amounts to 700 EUR for participants of the private sector, 300 EUR for employees of Ghent University, the government, the non-profit and social-profit sector. A special rate of 200 EUR applies to doctoral students. The book is optional and can be bought at the additional cost of 75 EUR. Please indicate this clearly on the registration form.

MODULE 6 – Missing Data

Prof. dr. Stijn Vansteelandt

Course description

Missing data (i.e. data that were intended to be collected, but were not) form a pervasive problem in many statistical data analyses, for the following two reasons. First, most standard statistical software routines only include subjects without missing data in the analysis. They thus make inefficient use of the observed data by discarding information from subjects whose data was only partially missing. Second, in many cases, subjects without missing data form a selective subgroup. Statistical results obtained for that group may not generalise to the intended study population.

The goal of this course is to develop an understanding of these problems caused by missing data. In addition, the goal is to learn how simple methods of correction for missing data (such as single imputation and last-value-carried-forward) may fail, and to provide methods for valid analysis in the presence of missing data (likelihood-based model estimation, weighting and multiple imputation). Particular emphasis will be placed on the assumptions underlying missing data analyses (e.g. whether the data are missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR)) and on the implications of these assumptions in standard analyses. In addition, considerable attention will be given to the relative advantages and limitations of the different missing data approaches. All methods will be illustrated with annotated computer output from SAS. Hands-on computer sessions will help practice the principles exposed in this course.

Dates and venue

January 16, 17 and 18, 2008 from 9 am to 12.15 am and from 1.15 pm to 4.30 pm at the Faculty of Science, Building S9, Campus Sterre, Krijgslaan 281, Ghent.

Target audience

The course targets researchers who need to analyse incomplete data sets and are seeking practical tools to handle missing data in their own analyses.

Exam

Participants can, if they wish, take part in an exam. A certificate from Ghent University will be issued to participants with a university degree at the bachelor level or an equivalent degree upon succeeding in this test. As such this course can be incorporated in a doctoral training program.

Course prerequisites

Participants are expected to be familiar with basic statistical data analysis and linear regression analysis.

Course material

Copies of lecture notes.
Recommended handbook: “Missing Data in Clinical Studies”, Geert Molenberghs and Mike Kenward, 1st ed. (2007), John Wiley.

Fees

The registration fee amounts to 700 EUR for participants of the private sector, 300 EUR for employees of Ghent University, the government, the non-profit and social-profit sector. A special rate of 200 EUR applies to doctoral students. The fee includes i.a. the lecture notes. The book is optional and can be bought at the additional cost of 95 EUR. Please indicate this clearly on the registration form.

MODULE 7 – Lineaire Regressie

Prof. dr. Huguette Reynaerts

Beschrijving

Vertrekkend van een eenvoudig lineair model om één uitkomst te voorspellen op basis van één predictor, wordt progressief een arsenaal van technieken opgebouwd om op flexibele wijze te modelleren en te voorspellen. Het zoeken naar significante (causale?) associaties voor de variatie in uitkomsten, het zo goed mogelijk inschatten van de verwachte uitkomst op basis van een gegeven reeks variabelen en het voorspellen van de verwachte uitkomst (betrouwbaarheidsinterval) of de uitkomsten zelf (predictie-interval), vormen de hoofdbrok van deze module. Bijzondere aandacht gaat naar de interpretatie van parameters in het regressiemodel. Deze hangt af van het model voor de verwachte uitkomst, van de residuele variatie rond die verwachte uitkomst en, last but not least, van de proefopzet. In functie van concrete vragen worden, consistent met een gegeven dataverzameling, zo efficiënt mogelijk modellen geselecteerd. Hoe meer a priori onderstellingen, des te nauwkeuriger het antwoord. De statistische basisonderstellingen worden geverifieerd en onderzocht op tegenspraak met het feitenmateriaal. Tevens worden uitschieters in de waarnemingen en hun invloed op de conclusies bekeken. Meervoudige lineaire regressiemodellen zijn zeer flexibel. Zij incorporeren de t-test, variantieanalyse en covariantieanalyse. Men kan dus tegelijkertijd de invloed onderzoeken van discrete en van continue variabelen, corrigeren voor confounding en effectmodificatie, en zekere niet-lineaire verbanden inbouwen. Er zijn echter ook grenzen aan de directe toepasbaarheid, die geleid hebben tot een waaier van uitbreidingen. In de cursus wordt naar oplossingen voor meer complexe problemen verwezen, zonder deze uit te diepen.

Data en Plaats

21 en 28 februari, 6, 13, 20 en 27 maart 2008, telkens van 11u tot 14u30 (theorieles gevolgd door een hands-on practicum) op de Faculteit Wetenschappen, Gebouw S9, Campus Sterre, Krijgslaan 281, Gent.

Doelpubliek

Deze cursus richt zich tot allen die geregeld in aanraking komen met data en die inzicht willen verwerven in statistische dataverwerking.

Examen

Voor deelnemers die dit wensen wordt de module afgesloten met een examen. Deelnemers die slagen voor dit examen en houder zijn van een universitair diploma op het niveau van bachelor of een gelijkwaardig diploma, bekomen een universitair getuigschrift. De module kan als dusdanig worden opgenomen in een doctoraatsopleiding.

Vereiste voorkennis

Basiskennis over statistische besluitvorming is vereist om duidelijk inzicht te krijgen in de verschillende modellen.

Lesmateriaal

Basistekst van de lesgever.
Aanbevolen handboek: “Applied Linear Statistical Models”, Michael H. Kutner, Christopher J. Nachtsheim, John Neter and William Li, 5th ed. (2004), McGraw-Hill.

Prijs

De deelnameprijs bedraagt 700 EUR voor deelnemers uit de private sector, 300 EUR voor UGent-personeelsleden en personeel uit de non-profit, social-profit en overheidssector. Een gereduceerde prijs van 200 EUR geldt voor doctoraatsstudenten. Het bovenvermelde boek is optioneel en kan besteld worden tegen een bijkomende kost van 75 EUR. Gelieve dit duidelijk op het inschrijvingsformulier aan te duiden.

MODULE 8 – Data Mining

Prof. dr. Stefan Van Aelst

Course description

The increased availability of computer storage space has lead to the collection of large databases in high dimensions. Also modern techniques such as micro-array analysis lead to high-dimensional data sets. This course focuses on modern statistical learning techniques to construct prediction rules based on such large, high-dimensional data. The outcome can be continuous (regression) or categorical (classification). The relation between bias, variance and model complexity will be explained. The course will discuss penalty methods, such as ridge regression and Lasso, local methods, such as nearest neighbors and kernels, splines, mixture models, classification and regression trees, support vector machines. Techniques such as cross-validation, bagging and boosting will be introduced as well. The curse of dimensionality will be explained and illustrated, and dimension reduction techniques such as principal component analysis will be discussed.

Dates and venue

March 25, 26 and 28, 2008 from 9.30 am till 12.30 pm and from 1.30 pm to 4.30 pm.

Target audience

This course targets researchers from all areas that are involved in prediction problems based on large and/or high-dimensional databases.

Exam

Participants can, if they wish, take part in an exam. A certificate from Ghent University will be issued to participants with a university degree at the bachelor level or an equivalent degree upon succeeding in this test. As such this course can be incorporated in a doctoral training program.

Course prerequisites

Participants are expected to be familiar with linear regression.

Course material

Copies of lecture notes.
The course is based on "The Elements of Statistical Learning: Data Mining , Inference and Prediction," by Hastie, T., Tibshirani, R., and Friedman, J. (2001) Springer-Verlag.

Fees

The registration fee amounts to 700 EUR for participants of the private sector, 300 EUR for employees of Ghent University, the government, the non-profit and social-profit sector. A special rate of 200 EUR applies to doctoral students. The fee includes i.a. the lecture notes. The book is optional and can be bought at the additional cost of 75 EUR. Please indicate this clearly on the registration form.

De lesgevers

Dhr. Kris Erauw is stafmedewerker bij de dienst onderwijsondersteuning van de Faculteit Psychologie en Pedagogische Wetenschappen aan de Universiteit Gent. Hij stond jarenlang mee in voor de begeleiding van studenten bij de vakken statistiek en methodologie, en bij het schrijven van hun scriptie. Daarnaast ondersteunt hij onderzoekers bij het ontwikkelen van een gepast onderzoeksopzet en bij het verwerken van hun data.

Prof. dr. David Harrington is Professor of Biostatistics in the Department of Biostatistics at the Harvard School of Public Health and Head of the Department of Biostatistics and Computational Biology at the Dana-Farber Cancer Institute. He conducts statistical research in survival analysis and collaborative research in cancer. He served as the Group Statistician for the Eastern Cooperative Oncology Group from 1990 to 2000, an organisation of approximately 300 treatment sites conducting clinical and basic research in all adult malignancies. He is currently the Director of the Biostatistics Core Facility for the Dana-Farber/Harvard Cancer Center, a consortium of Harvard Medical School teaching affiliates, academic departments and laboratories with more than 750 investigators directly involved in cancer research. He is also the principal statistician for the Cancer Care Outcomes Research and Surveillance Consortium, a network of cancer registries and cancer centers organised to study patterns of cancer care among US physicians and access to care for subpopulations.

Dr. Anita Prinzie is postdoctoraal onderzoeker van het Fonds Wetenschappelijk Onderzoek – Vlaanderen aan gastuniversiteit Universiteit Gent, Vakgroep Marketing. Zij behaalde haar diploma Master in Marketing Analysis and Planning en haar doctoraat aan de Universiteit Gent. In haar doctoraat bestudeerde zij het gebruik van sequentiële analyse methodes voor Customer Relationship Management. Na haar doctoraat werkte ze als gastonderzoeker aan Monash University, Australië. Momenteel doceert zij binnen de Master of Marketing Analysis (UGent) in de Marketing Modeling and Engineering cursus SAS macrotaal. Haar huidige onderzoeksinteresses omvatten het testen van de externe validiteit van haar nieuw ontwikkelde Random Multinomial Logit algoritme voor keuze-analyses en het begrijpen en analyseren van klantenbeslissingsprocessen vanuit een marketingactie perspectief.

Prof. dr. Huguette Reynaerts is hoofddocent aan de Universiteit Gent, Vakgroep Toegepaste Wiskunde en Informatica. Ze is verantwoordelijk voor de cursussen statistiek in de Faculteit Economische en Toegepaste Economische Wetenschappen. Naast haar vorming als wiskundige aan de UGent, specialiseerde zij zich in de statistiek, de econometrie en het operationeel onderzoek aan de VUB. Haar wetenschappelijk onderzoek situeert zich op het gebied van de financiële stochastiek.

Prof. dr. Stefan Van Aelst is docent aan de Universiteit Gent, Vakgroep Toegepaste Wiskunde en Informatica, Faculteit Wetenschappen, waar hij verantwoordelijk is voor cursussen statistiek in de opleidingen wiskunde, informatica, geologie en tandheelkunde. Zijn onderzoeksgebied is het ontwikkelen en bestuderen van robuuste methoden voor statistische modellen.

Prof. dr. Stijn Vansteelandt is docent aan de Universiteit Gent, Vakgroep Toegepaste Wiskunde en Informatica. Hij doceert er statistiek aan de Faculteiten Wetenschappen en Farmaceutische Wetenschappen en binnen de Master in Statistische Data-Analyse. Hij verrichtte post-doctoraal onderzoek aan de Harvard School of Public Health en aan de Universiteit Gent. Zijn huidige onderzoek spitst zich voornamelijk toe op missing data en causale besluitvorming in geclusterde sampling designs en longitudinale studies.

Prof. dr. Maria Ysebaert is eredocent van de Universiteit Gent. Zij ontving haar vorming als biochemicus aan de UGent, de University of Oregon Medical School en het Nobel Instituut te Stockholm. Naast biochemie, onderwees zij biostatistiek op kandidatuur- en postgraduaatniveau aan de Faculteit Diergeneeskunde van de UGent. Haar huidige wetenschappelijke interesse betreft statistische analyse in het onderzoek van moleculaire structuren van proteïnen.

Practical information

Registration

Please use the registration form in this brochure or on our website: **www.ipvw-ices.UGent.be**.
Your registration is valid only upon receiving a confirmation mail from ICES. If you have not received this mail within a week, please contact ICES to double check.
The registration fee covers tuition, some or all of course materials, use of auditoria and PCs, coffee and refreshments.

Payment

The registration fee is due within 30 days following receipt of the invoice. Payment is possible through bank transfer **with clear statement of the structured message on the invoice**.
All mentioned amounts are free from VAT.

Additional reduction

When 3 or more participants from the same company or institute enrol simultaneously for the same module(s), an additional overall reduction of 20% is granted. Please check before enrolling if anyone else at your institute or company might be interested or is already planning to participate. This reduction does not apply to (doctoral) students.

Cancellation

Cancelling is possible in writing until 5 days before the start of the concerning module, in which case 25% of the registration fee will be retained. In case of cancellation within 5 days before the start of the module, the full registration fee is due.

Training vouchers (Opleidingscheques)

The training vouchers are an initiative of the Flemish community to stimulate continuing education. There are 2 types of training vouchers:

For employers:
order them at **www.opleidingscheques.be**

For employees:
order them at **www.vdab.be/opleidingscheques**

On the mentioned sites you can read all necessary information about this initiative.

Doctoral schools (Information for UGent doctorandi)

As of academic year 2007-2008 five Doctoral Schools will be founded at the Ghent University. They are concentrated around the following domains of research:

- Arts, Humanities and Law
  (Director: prof. Dominique Willems)
- Social and Behavioural Sciences
  (Director: prof. André Vandierendonck)
- Natural Sciences
  (Director: n.n.)
- (Bioscience) Engineering
  (Director: n.n.)
- Life Sciences and Medicine
  (Director: prof. Jozef Vercruysse)

These constellations, in close consultation with the faculties, will play a supporting role to doctorandi: on the one hand they will organise specialised trainings and workshops in research skills and "transferable skills", on the other hand they will set up guest lectures, information sessions, promotional tasks and contacts with the industry.
At press time, the practical procedures were not yet known. Please keep an eye out for more information on our website and check the UGent-website (Onderzoek > Doctoreren > De oprichting van 'Doctoral Schools') for more information.

Registration form Statistics 2007-2008

This form can also be found on our website: **www.ipvw-ices.UGent.be**

Please send, fax or e-mail this form to ICES no later than 4 working days before the start of the first selected module.
Address: ICES – Krijgslaan 281, S3 – 9000 GHENT **Fax:** +32 (0)9 264 85 90 **E-mail:** Isabel.DeZutter@UGent.be

Your registration is only valid after you receive a confirming email from our service.

Last Name: First name:

Function:

Company or institute:

Address:

Phone: Fax: E-mail:

Date of birth*: / / 1 9 Place of birth*: Gender*: ☐ M ☐ F

* General information about age and gender of our participants is sent without names to Sodexho or Accor within the framework of the training vouchers and is used on the certificates

Required for UGent-participants: SAP internal order number°: **4 3** . . . . . .

° intern bestelbonnr, verplicht voor deelnemers van UGent, gelieve ook de bestelbon zelf mee op te sturen

Yes, I enrol for the following module(s) of the course in Statistics 2007-2008 organised by the Center for Statistics in co-operation with the Institute for Continuing Education in Science:

- ☐ **M1: Inleiding tot SPSS****
- ☐ **M2: Introduction to SAS** *I order:* ◯ *No book* ◯ *Book (Delwiche & Slaughter)*
- ☐ **M3: Introductory Statistics: Basics of Statistical Inference**
  I order: ◯ *No book* ◯ *Book 1 (Rosner)* ◯ *Book 2 (Moore & McCabe)*
- ☐ **M4: Survival Analysis**
- ☐ **M5: Analysis of Variance** *I order:* ◯ *No book* ◯ *Book (Kutner et al.)*
- ☐ **M6: Missing Data** *I order:* ◯ *No book* ◯ *Book (Molenberghs et al.)*
- ☐ **M7: Lineaire Regressie**** *I order:* ◯ *No book* ◯ *Book (Kutner et al.)*
- ☐ **M8: Data Mining** *I order:* ◯ *No book* ◯ *Book (Hastie et al.)*

** These courses are taught in Dutch.

The total amount for my registration adds up to ..... euro, of which ..... euro will be paid through use of : ◯ training vouchers for employees / ◯ training vouchers for employers.
I took note of the payment and cancellation procedure. All mentioned amounts are free from VAT.

I am:
- ☐ employed in the industry
- ☐ employed in the non-profit, social profit, government sector, schools, …
- ☐ (doctoral) student (Student card n° .....)

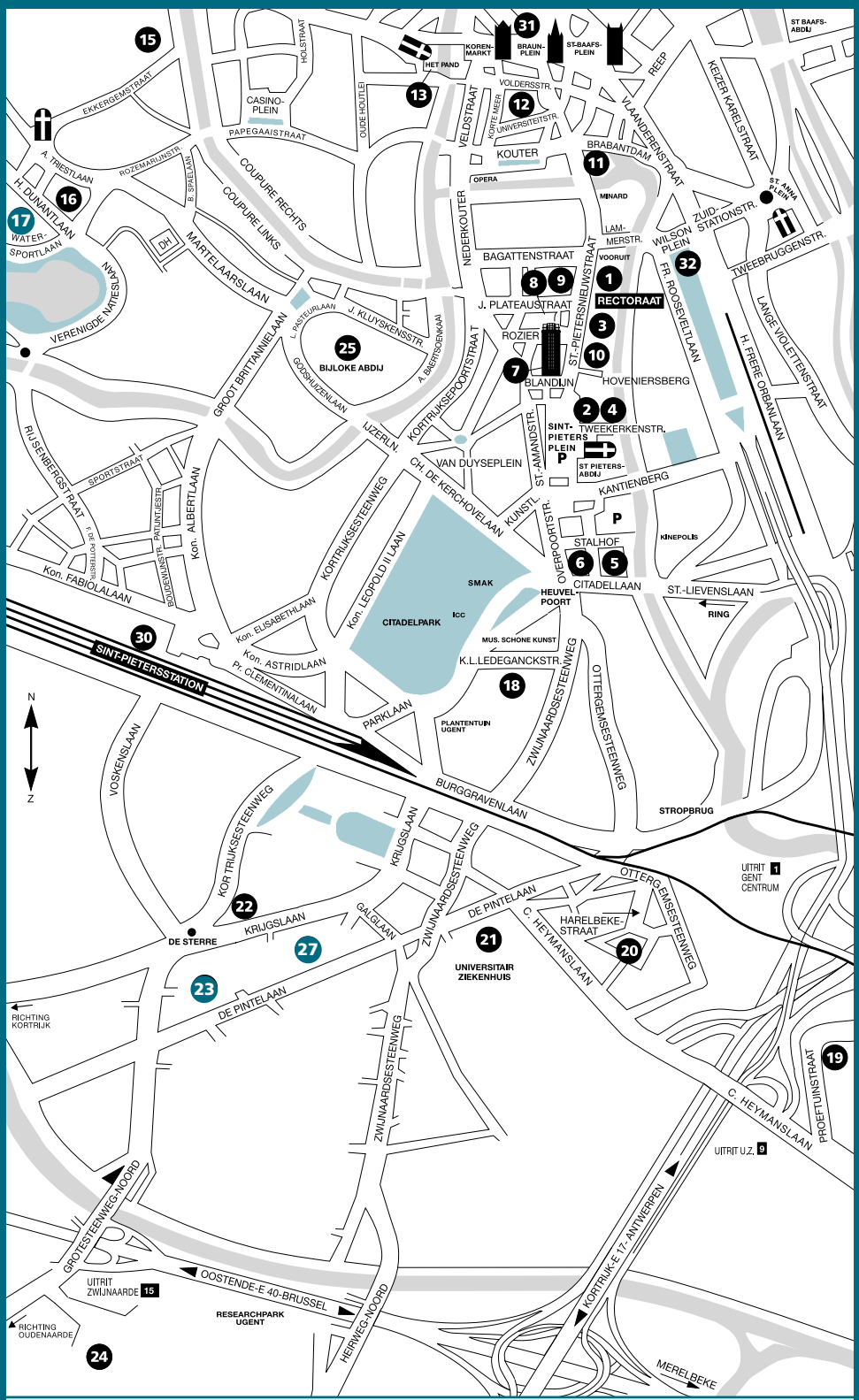
INVOICE ADDRESS Name: .....

Address: .....

Date: ..... Signature:

.....

Course locations



17	Fac. of Psychology and Educational Sciences (PP), H. Dunantlaan 1		
23	Fac. of Sciences (WE), Campus Sterre, Krijgslaan 281, building S9		
27	Fac. of Sciences (WE), IPVW-ICES, Campus Sterre, Krijgslaan 281, building S3		
2	Adviescentrum voor studenten	**30**	Station Gent Sint-Pieters

Voor alle verdere inlichtingen kan u zich wenden tot

UNIVERSITEIT GENT **Faculteit Wetenschappen**

IPVW, Krijgslaan 281, S3, 9000 Gent
Telefoon 09 264 44 26 – www.ipvw-ices.UGent.be