

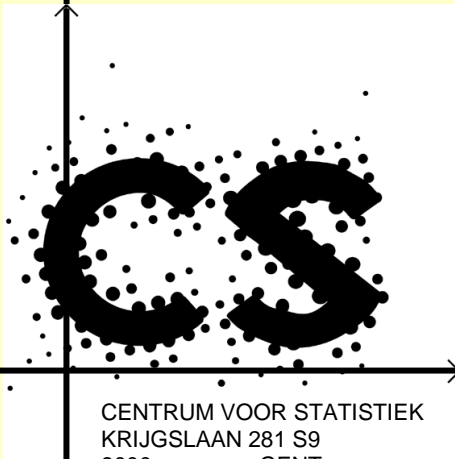
CURSUSSEN STATISTIEK

2001-2002

P



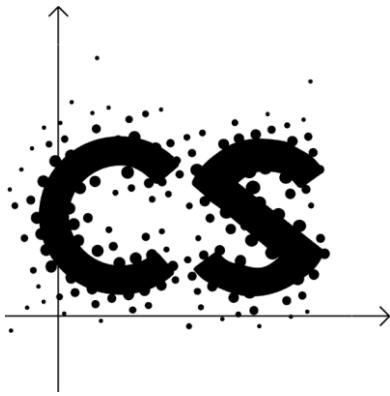
UNIVERSITEIT
GENT



CENTRUM VOOR STATISTIEK
KRIJGSLAAN 281 S9
9000 GENT



INSTITUTE FOR CONTINUING
EDUCATION IN SCIENCE
KRIJGSLAAN 281 S8
9000 GENT



**CENTRUM VOOR STATISTIEK
UNIVERSITEIT GENT**

STATISTIEK



**INSTITUTE FOR
CONTINUING EDUCATION
IN SCIENCE**

VOORWOORD

"Statistiek" is de laatste decennia onmisbaar geworden in heel wat takken van de wetenschap. Denk maar aan wetsgeleerden die bewijsmateriaal wegen, 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 organiseert daarom sinds enkele jaren 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. PC-practica 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.

Het programma voor het academiejaar 2001-2002 heeft een upgrade ondergaan. In de eerste plaats is er de samenwerking met het Instituut voor Permanente Vorming van de Faculteit Wetenschappen (IPVW). Het aanbod kadert daardoor binnen een groter geheel en kan rekenen op een ruimere infrastructuur.

In de tweede plaats worden verschillende cursussen in het Engels gedoceerd. Daardoor kunnen we ook aan buitenlandse onderzoekers een bijzondere kans tot nascholing bieden.

Last but not least worden er op inhoudelijk vlak 2 nieuwe modules geïntroduceerd die bouwen op de basiscursussen om te komen tot de analyse van 'categorische data' en tot 'survival analysis'.

We hopen dat u het nieuwe aanbod kan smaken!

MODULE 1: Inleiding tot SPSS

Prof. Dr. G. Schuyten en Drs. K. Erauw

Beschrijving: Je hoort het overal: we leven in een informatiemaatschappij. Heel veel mensen houden zich bezig met het verzamelen van gegevens of willen bepaalde ideeën met onderzoeksbevindingen ondersteunen. Denk maar aan de jongeren in een stedelijke jeugdraad die in hun jeugdwerkbeleidsplan de geformuleerde beleidslijnen moeten ondersteunen met onderzoeksbevindingen. Of de voorzitter van een oudercomité die de tevredenheid van ouders over de school op een wetenschappelijk verantwoorde manier wil bevragen. Het verzamelen en opslaan van al die gegevens is vaak niet zo evident als het lijkt. Zeker niet als het later de bedoeling is deze gegevens op een professionele manier te verwerken.

Deze lessenreeks is erop gericht data in een bruikbare vorm te verzamelen, die ingezamelde data in SPSS op te slaan en in SPSS de eerste beschrijvende statistieken te produceren.

De verschillende lessen in de reeks zijn vooral ervaringsgericht opgevat. De deelnemers worden met een aantal problemen geconfronteerd waarna mogelijke oplossingen besproken en gedemonstreerd worden.

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

Les 2: Samenvattende statistieken. Data-cleaning. Grafische voorstellingen van in SPSS ingevoerde data. SPSS-output in andere programma's gebruiken.

Les 3: Nieuwe variabelen aanmaken op basis van bestaande variabelen. Variabelen hercoderen. Werken met een gedeelte van een databestand. Verschillende SPSS-databestanden samen gebruiken.

Les 4: Verbanden tussen twee variabelen. Lineaire regressie, t-toets en variantieanalyse.

Data: 8, 9, 11 en 12 oktober 2001 telkens van 17u tot 20u.

Plaats: Faculteit Psychologische en Pedagogische Wetenschappen, Henri Dunantlaan 1, Gent.

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

Toelatingsvoorwaarden: Geen

Lesmateriaal: De syllabus die bij de practica beschikbaar is, is opgevat als zelfstudiepakket waarmee cursisten na of naast de practica op eigen houtje verder kunnen oefenen.

Prijs: De deelnameprijs bedraagt €100 (RUG-doctorandi), €150 (personeel RUG of non-profit sector), €300 (private sector) en omvat practica met SPSS.

MODULE 2: Algemene inleiding tot S-PLUS

verzorgd door de commerciële verdeler van het pakket: CANDiensten, Nederland

Beschrijving: S-PLUS is een modulaair gebouwd statistisch pakket met bijzondere grafische mogelijkheden. Het biedt de gebruiker heel wat flexibiliteit zowel bij het analyseren van een dataset met behulp van standaardtechnieken als het implementeren van eigen methoden. In deze initiatiecursus maakt U kennis met S-PLUS voor Windows. De cursus legt de nadruk op het gebruik van S-PLUS met behulp van de Grafische User Interface. Ook de S-PLUS-programmeertaal wordt kort behandeld. Na de cursus bent U in staat om via de Grafische User Interface:

- data in te lezen en te bewerken
- grafieken te maken en aan te passen,
- eenvoudige statistische analyses uit te voeren.

Datum: Zaterdag 27 oktober 2001 van 9u30 tot 16u30

Plaats: Faculteit Wetenschappen, Gebouw S9, PC-zaal K. Zuse (3^e verdieping), Campus Sterre, Krijgslaan 281, Gent.

Deelname: Deelname is gratis, maar u moet zich wel registreren. Dit kan door het sturen van een e-mail naar cursus@candienst.nl, met vermelding van uw naam, departement, instelling, telefoonnummer en de cursus die u wilt volgen. U ontvangt een bevestiging van deelname. Het aantal deelnemers is beperkt tot 20.

MODULE 3: Introductory Statistics. Basics of Statistical Inference

Prof. Dr. Maria Ysebaert

Course description: This course aims to provide insight into 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 mainly from biological and medical research.

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 and 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.

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. A regression line based on observations obtained in a sample describes this relation and enables us to predict one variable from the other. We will check whether the proposed model conforms to the data. Finally, we extend this technique to a more complex setting in which more than one variable explains the response.

Language: English will be used in all lectures and course notes.

Schedule: October 23 and 30, November 6, 13 and 27, December 4 and 11, 2001, from 17:00 to 21:00

Venue: Faculty of Science, Building S9, Campus Sterre, Krijgslaan 281, Gent

Target audience: This course will benefit medical investigators, research scientists, clinical research associates, and, in general, all researchers who come in contact with data handling and who want to acquire insight into statistical methods or who feel that their knowledge and practice of statistics needs refreshing. No extensive background in mathematics is required.

Course prerequisites: No specific requirements. A prior knowledge of some statistical principles would be helpful.

Required literature: "Introduction to the Practice of Statistics", David S. Moore and George P. McCabe, Third edition (1999), W.H. Freeman and Company, New York.

Fees: The registration fee amounts to €800 for participants of the private sector and €350 for employees of Ghent University and the non-profit sector. A special rate of €200 applies to doctorandi at Ghent University. The fee includes the textbook and copies of the transparencies used in the lectures.

MODULE 4: Short course on 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.

Data: Monday November 19th (9:30-17:30) and Tuesday November 20th (9:30-12:00). All methods will be illustrated with annotated computer output from SAS and S-PLUS but there will be no hands-on computer sessions. Registration starts at 9:00 on Monday.

Venue: Het Pand, Onderbergen 1 (for pictures and a map see website: <http://www.gent.be>)

Intended 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.

Course material: Copies of lecture notes.

Price: The registration fee amounts to €500 for participants of the private sector and €300 for employees of Ghent University and the non-profit sector.

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, concentration, subject, sex, age, etc. First, we cover one-way analysis of variance, 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 analysis of variance (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 two 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 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 analysis of variance, it is essential to state the linear model and to take the type of factor into account. Also, the conditions that must be fulfilled in the use of ANOVA, should be kept in mind. 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.

Language: English will be used in all lectures and course notes.

Schedule: January 8, 15, 22 and 29, February 5, 12 and 19, 2002, from 17:00 to 21:00.

Venue: Faculty of Science, Building S9, Campus Sterre, Krijgslaan 281, Gent

Target audience: This course will benefit medical investigators, research scientists, clinical research associates, who need to use statistical methods in the collection and handling of data in their research, in particular for assessing the effect of treatments, in order to evaluate in an exact way the biological and experimental variability. Participants are expected to have an active knowledge of the basic principles underlying statistical strategies, at a level equivalent to the introductory statistics course (module 3) of this program. Some experience with statistical software, such as SPSS, would be helpful.

Course prerequisites: Holder of a university degree at least to the bachelors' level or an equivalent degree.

Course material: "Applied Linear Statistical Models", John Neter, Michael H. Kutner, Christopher J. Nachtsheim and William Wasserman, Fourth edition (1996), Irwin Press, Chicago.

Fees: The registration fee amounts to €800 for participants of the private sector and €300 for employees of Ghent University and the non-profit sector. A special rate of € 200 applies to doctorandi at Ghent University. The fee includes the textbook and copies of the transparencies used in the lectures.

MODULE 6: Workshop on Resampling Methods

Prof. Dr. Philip Good

Course description: This course on hypothesis testing and estimation using only the data at hand is intended for industrial statisticians, research workers in other fields and others interested in applying distribution-free statistical methods based on resampling. Resampling uses brute computer power and it is not based on assumptions such as normal distributions but rather based on repeated samples from the researcher's own data. Real-world examples with both small and large samples are used to illustrate the many practical applications of the bootstrap, the permutation test, and cross-validation.

INTRODUCTION TO RESAMPLING METHODS

(The "what" requires only the ability to count;

The "why" requires familiarity with the concepts of Type I and Type II error)

Four Methods: one common basis

- 1.1. Bootstrap
- 1.2. Cross-Validation
- 1.3. Density-Estimation
- 1.4. Permutation Tests

When to use Resampling Methods

Four Easy Tests

- Two-sample comparison
- Correlation
- One-Way Analysis
 - Unordered categories
 - Ordered categories
- Randomized Blocks

Five Steps to a Permutation Test

Permutation or Parametric Distribution

HARDER PROBLEMS

(Assumes some familiarity with experimental design)

Comparing Variances

One-Sample

Bootstrap Sampling

Behrens-Fisher Problem

Assumptions

Weak and Strong Exchangeability

Permutation, Parametric, and Bootstrap

Contingency Tables

Multivariate Analysis

Date: Friday April 5, 2002, from 14:00 to 17:00.

Venue: Faculty of Science, Building S9, Campus Sterre, Krijgslaan 281, Gent

Fees: The registration fee amounts to €120 for participants of the private sector and €50 for employees of Ghent University and the non-profit sector.

MODULE 7: Categorische Data-analyse

Prof. dr. G. Schuyten en Dr. Yves Rosseel

Beschrijving: De laatste jaren zien we een toename in publikaties van het gebruik van multivariate technieken voor categorische data van nominaal en ordinaal meetniveau. De welgekende klassieke analyses van contingentietabellen met chi-kwadraat toetsen en associatiematen voldoen niet meer. De concepten die gepaard gaan met deze nieuwe technieken zoals odds ratios, marginale en conditionele odds, het algemeen log-lineaire model, effect parameters en dergelijke zijn nog onvoldoende gekend bij een breed publiek. De module staart met een behandeling van deze nieuwe concepten. Vervolgens worden de belangrijkste methodes besproken die in de praktijk gebruikt worden voor het analyseren van categorische data. Zowel deze waarbij de respons (afhankelijke) variabele categorisch is, als deze waarbij alle variabelen waarvan men de samenhang wenst te onderzoeken categorische variabelen zijn.

Vervolgens komen de statistisch modellen aan bod waarop deze technieken gebaseerd zijn. Het meest populaire model voor een categorische binaire respons variabele (bvb. ‘geslaagd’ of ‘niet-geslaagd’) is de logistische regressie die vele gelijkenissen vertoont met het gewone, welbekende regressiemodel. Dit model wordt vervolgens gegeneraliseerd naar respons variabelen met meer dan twee categorieën. Het log-lineaire model ligt aan de basis van analyse van verbanden en interactie-patronen tussen categoriale variabelen. Beiden modellen zijn bijzondere gevallen van het veralgemeende lineaire model.

Tot slot behandelen we enkele gespecialiseerde (logit) modellen voor (1) categorische variabelen van ordinaal meetniveau, (2) herhaalde metingen, en (3) kleine steekproeven.

Data: 14, 21, 28 maart, 18, 25 april, 2 en 16 mei 2002 van 17u tot 20.30u.

Plaats: Faculteit Psychologische en Pedagogische Wetenschappen, PC-klas, Henri Dunantlaan 1, Gent.

Doelpubliek: Deze module richt zich tot alle personen in de sector gezondheidszorg, onderwijs, dienstensector en in de bedrijfswereld die geconfronteerd worden met de analyse van categorische data. Voorkennis over categorische data-analyse is niet noodzakelijk. Een basiskennis over gewone regressie en variantieanalyse is aangeraden om duidelijk inzicht te krijgen in de diverse modellen.

Toelatingsvoorwaarden: Houder van een universitair diploma of van niet-universitair hoger onderwijs.

Lesmateriaal: Handboek: “An Introduction to Categorical Data Analysis” van Alan Agresti, en kopies van transparanten van de lesgevers.

Prijs: Het cursusgeld bedraagt €900 voor deelnemers uit de private sector, €350 voor RUG-personeel of personeel uit non-profit sector. Een gereduceerde prijs van €250 geldt voor doctorandi van RUG.

DE LESGEVERS

Prof. Dr. Phillip Good, a frustrated novelist, is the author of “Resampling Methods” (Birkhauser, 2nd Ed, 2001), “Permutation Tests” (Springer-Verlag, 2nd ed 2000), “Applying Statistics in the Courtroom” (Chapman-Hall, 2001) and “Managers’ Guide to Design and Conduct of Clinical Trials” (CRC, 2002). Nine other books, 600 plus popular articles, 30 scholarly articles, and eight of his short stories are also in print. Holder of the Ph.D. and A.B. in Mathematical Statistics from the University of California at Berkeley, he is a former Calloway Professor of Computer Science at the University of Georgia, Division Head and Professor of Biology and Physics at West Coast University, and Associate Professor of Applied Mathematics at Claremont College

Prof. Dr. David Harrington received his doctoral degree in statistics from the University of Maryland in 1976, and has held academic appointments at the University of Virginia and the Harvard School of Public Health. He is currently Professor of Biostatistics in the Department of Biostatistical Science at the Dana-Farber Cancer Institute. Professor Harrington conducts statistical research in survival analysis and in longitudinal data and collaborative research in cancer. He served as the Group Statistician for the Eastern Cooperative Oncology Group from 1990 to 2000, an organization 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 Consortium for Outcomes Research, a recently funded network of cancer registries and cancer centers organized to study patterns of cancer care among US physicians and access to care for subpopulations.

Prof. Dr. Gilberte Schuyten is hoogleraar aan de Vakgroep Data-analyse van de RUG en verantwoordelijk voor het onderwijs in statistiek, empirisch onderzoek en kwantitatieve methoden aan psychologen en pedagogen. Haar huidig onderzoek spitst zich toe op het inzetten van elektronische leeromgevingen in het statistiekonderwijs.

Prof. dr. Maria Ysebaert is eredocent aan de Universiteit Gent. Zij ontving haar vorming als biochemicus aan de RUG, 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 RUG. Haar huidige wetenschappelijke interesse betreft statistische analyse in het onderzoek van moleculaire structuren en proteïnen.

Drs. Kris Erauw is assistent aan de Vakgroep Data Analyse, waar hij betrokken is bij de ondersteuning van de vakken statistiek en methodologie. Bovendien begeleidt hij studenten bij het maken van hun scriptie voor de statistische en methodologische verwerking.

Dr. Yves Rosseel is doctor-assistent aan de Vakgroep Data-analyse van de RUG. Hij behaalde in deze instelling zijn doctoraat als experimenteel psycholoog, verrichtte post-doctoraal onderzoek aan de University of Warwick en de KUL. Zijn huidig onderzoek spitst zich toe op categorisatie, mixture models en data-mining.

DEELNEMINGSFORMULIER

Terug te sturen of te faxen (09 264 49 83) bij voorkeur 4 dagen voor de aanvang van de door u geselecteerde module.

Naam:

Voornaam:

Functie:

Onderneming:

Adres:

.....

Telefoon: Fax:

E-mail:

Ik schrijf mij in voor de volgende modules van de Cursussen Statistiek 2001-2002 ingericht door het Centrum voor Statistiek van de Universiteit Gent in samenwerking met het IPVW:

- Module 1: Inleiding tot SPSS
- Module 2: Algemene inleiding tot S-PLUS
- Module 3: Introductory Statistics. Basics of Statistical Inference
- Module 4: Short course on Survival Analysis
- Module 5: Analysis of Variance
- Module 6: Workshop on Resampling Methods
- Module 7: Categorische Data-analyse

.....
FAKTURATIE-ADRES

Naam:

Adres:

.....

Datum: Handtekening:

PRAKTISCHE INLICHTINGEN

Het cursusgeld dient gestort te worden na ontvangst van de factuur. De facturen zijn contant betaalbaar dertig dagen na dagtekening. Alle vermelde bedragen zijn vrij van BTW.

De antwoordstrook dient terugstuurd te worden aan:
Heidi De Dobbelaere
Krijgslaan 281 S3
9000 Gent
Fax +32 (0)9 264 49 83

U kan ook inschrijven door het sturen van een e-mail naar Heidi.DeDobbelaere@rug.ac.be
Uw inschrijving is slechts geldig na ontvangst van een bevestiging per e-mail.

Vanaf de derde inschrijving van (niet-RUG) personeelsleden van eenzelfde bedrijf of instelling wordt een korting van 20% toegestaan.

In de deelnameprijs zijn begrepen: het lesgeld, het lesmateriaal, het gebruik van de lokalen, koffie en broodjes.

Annulatie is mogelijk:

Indien men dit schriftelijk bevestigt 10 dagen of meer voor het begin van de betreffende module; er is evenwel een vergoeding van 25% van de deelnameprijs verschuldigd;

Indien men minder dan 10 dagen voor de aanvang van de betreffende module annuleert, dan is de volledige deelnameprijs verschuldigd.

Bijkomende inlichtingen over deze cursussen en over het Centrum voor Statistiek zijn te bekomen op de website: <http://cvstat.rug.ac.be>