

Course description: Sampling and estimation

7,5 ECTS-credits

Swedish title: Urval och estimation

Advanced and masters level

The course can also be a moment in some other courses at the advanced level

Description of the course

The course consists of one moment:

1. Sampling and estimation. 7.5 ECTS credits

The course covers representative sampling and how best to exploit the prior information available.

The course covers the most common sampling methods like simple random sampling, stratified sampling, sampling with varying inclusion probabilities (known as π ps), cluster sampling, multistage sampling and systematic sampling. The course also deals with sampling methods where the time aspect is important, for example longitudinal studies, rotating or panel sampling and methods which provide a good geographical or spatial distribution. It also discusses special designs whose primary purpose is to measure changes over time at both micro and macro level. It also discusses how to choose between the methods and how the plans are implemented. When a survey has been finished, one must estimate the various parameters such as population totals, underlying entities, relationships and special dimensions. The course discusses different estimation procedures, particularly when there are different types of auxiliary information in the frame.

Learning goals

After completing the course students should be able to

- explain the advantages and disadvantages of standard sampling plans
- select appropriate sampling plans for different selection problems.
- carry out estimation and precision estimation on data from different sampling plans, with and without auxiliary information, with the help of computers and for small samples also without a computer
- analyze both model- and design-based sampling
- choose suitable estimators depending on the problem and the access to auxiliary information
- describe and use many common estimation methods for non-response problems, including multiple and single imputation, calibration
- on his own solve most problems in typical standard textbooks at this level within a limited time.
- describe the literature and account for the major journals in the field.

Forms of examination and grading criteria

The course is examined through both individual assignments and a written test. Textbooks and written notes are allowed at the test. All assignments are graded as approved or not approved. If they are not approved the assignment can be reworked and handed in once more. One of the individual assignments is reported in the form of an oral presentation where attendance is mandatory. In addition all assignments are graded with credit points. The assignment grades are added together and rescaled so that the maximal credit sum is 100 credit points. The written test can also give a maximum of 100 points. To pass the moment (grades A – E) it is required that all assignments are approved, the written test is given at least 40 points and the total score exceeds 100 points. Everything needs to be achieved at the same "semester", which means during the course including the first reexamination opportunity.

The final grade on the moment is based on a total count of the credit points from the test and the assignments.

The maximal total credit is thus 200 points. Grades are given on a seven-point rating scale

- A. Excellent - 180-200 points
- B. Very good - 160-179 points
- C. Good - 140-159 points
- D. Satisfactory - 120-139 points

E. Enough - 100-119 points

Fx. Insufficient - 60-99 points or more than 100 points but all assignments have not been approved or the exam gave less than 40 points

F. Completely insufficient - equivalent to less than 59 points

To pass the whole course a minimum grade of E on the moment is required. The grade for the course is the same as for the moment.

Students who have passed an examination may not retake the test for higher grades. Exceptions are made for those students who passed all moments but not reached a total count of 100 points. They may rewrite the written exam. However, assignments and writing results are not valid after the first reexamination opportunity. Next semester the students who did not pass the moment must make all assignments of that semester and rewrite the written test.

Teaching

There will be lecture sessions at ten occasions, most of them consisting of four hours. The time schedule can be found at the department's website (<http://gauss.stat.su.se/site/scheman/vt12/uoe.pdf>). To get good results on the course it is important to attend all training sessions. Of these 38 hours about 20 hours will be like ordinary lectures, about 2 hours will be guest lectures, about 12 hours will be devoted to problem solving and 4 hours to presentation of assignments. If there will be many students at the course, additional presentation opportunities may be scheduled. Before the problem solving exercises student are assumed to have read through and tried to answer all of the exercises given in advance. The course is self-contained and will formally not require any previous knowledge of sampling but in practice one introductory course to sampling is strongly recommended.

The course is equivalent to 7.5 credits, which in turn corresponds to 10 half-time weeks or 200 hours workload. A typical workload of a student may be allocated as follows: 36 hours lectures, 39 hours of home assignments, 5 hours test, 60 hours trying to solve the tutorials (by hand or computer) and 60 hour reading the literature.

Required literature

Material distributed during the course. (Lecture notes and other material which are posted on the department's web page are considered as being distributed in during the course).

Sharon Lohr, (2010), Sampling: Design and Analysis, 2nd edition, chapters 1-9, 12, 15 Duxbury

Reference literature

Urval – från teori till praktik, Handbok 2008:1, SCB – finns att hämta på http://www.scb.se/statistik/publikationer/OV9999_2007A01_BR_X99BR0801.pdf.

Teachers

Daniel Thorburn, 08-16 29 56, Daniel.Thorburn@stat.su.se (lectures)

Nicklas Pettersson 08-16 26 96 Nicklas.Pettersson@stat.su.se (exercises)

Preliminary course content

Standard sampling methods. What is sampling? Differences between model- and design-based inference. Auxiliary information. Probability sampling. Some common sampling methods. Regression estimators, GREG. Stratified sampling, optimization, quota sampling. Sampling with varying inclusion probabilities, HT-estimation, SYG-variance estimation, various sampling techniques π ps, simultaneous choice of estimation and sampling techniques. Cluster sampling, indirect sampling, sampling when the frame is unknown *a priori*. Variance estimation, resampling, Taylor linearisation. Longitudinal studies, rotating samples, panel sampling, change and level estimates. Estimation in the presence of non response, MAR, MCAR, and various forms of imputation and reweighed, propensity score, NMAR methods. Other parameters, such as quantiles, Gini coefficients and domain

estimates. Normality, asymptotic results. Embedded experiments, subsampling of the loss, quality studies.

Preliminary schedule

(13-17 if nothing else is said)

16/1	Introduction, paradigms, SRS model-based, SRS design-based auxiliary variables, some simple sampling designs
24/1	Regression-type estimators, Difference, ratio, regression, GREG- and prediction estimators Exercises (NP)
30/1 (14-18)	Stratified sampling Optimal allocation, number of strata, post stratification; quota sampling Exercises (NP)
7/2	Sampling with varying inclusion probabilities HT and SYG estimators, combination of design and estimation techniques, cluster sampling, indirect sampling, sampling from an unknown frame Exercises (NP)
13/2	Estimation with non-response, Missing data, MAR, MCAR, post stratification, some reweighting and imputation techniques Exercises (NP)
20/2	Large sample properties Central limit theorem for independent random variables and sampling from finite populations, Berry-Esseen type central limit theorems Exercises (NP)
27/2 (14-18)	Coordinated samples Longitudinal studies, negative-positive coordination, permanent random numbers, rotating panels, Estimates of level and/or change. Exercises (NP)
5/3	Miscellaneous topics Gauss-approximation, resampling, quantiles, other functions, quality and evaluation studies, embedded experiments Exercises (NP)
9/3	Bayesian probability and sampling, Paper presentations Exercises (NP)
15/3 (10-12)	Paper presentations

Kursbeskrivning: Urval och estimation,

AN 7,5 högskolepoäng

Engelsk titel. Sampling and Estimation, AL, 7,5 ECTS-credits

Avancerad nivå och masternivå

Kursen kan också ingå som moment i vissa kurser på avancerad nivå

Beskrivning av kursen

Kursen består av ett moment:

1. Urval och estimation, 7,5 högskolepoäng

Kursen behandlar hur man gör representativa urval och hur man på bästa sätt kan utnyttja i förväg tillgänglig information.

Kursen behandlar de vanliga urvalsmetoderna för en urvalsundersökning: obundet slumpmässigt urval, stratifierat urval, olika former av urval med varierande urvalssannolikheter (sk pps) klusterurval, flerstegsurval och systematiska urval. Kursen behandlar även urvalsmetoder som utnyttjar tidsaspekten t ex roterande urval eller panelurval och metoder som ger en bra geografisk eller spatial spridning. Den diskuterar också speciellt undersökningar vars primära syfte är att studera förändringar över tiden både på mikro- och makronivå. Vidare behandlas hur man väljer mellan metoderna och hur planerna implementeras. När sedan undersökningen genomförts skall man sedan skatta olika parametrar såsom populationstotaler, bakomliggande storheter, samband och speciella mått. Kursen diskuterar olika skattningsförfaranden speciellt vid tillgång till olika former av hjälpinformation i ramen.

Lärandemål

Efter att ha genomgått kursen förväntas studenten kunna

- redogöra för fördelar och nackdelar med vanliga urvalsplaner
- självständigt välja lämpliga urvalsplaner för olika urvalsproblem.
- genomföra skattningar och osäkerhetsberäkningar på data från olika urvalsplaner, med och utan hjälpinformation, med och vid mindre urval utan dator
- analysera både modell- och designbaserade urval
- välja skattningsmetod beroende på problemet och tillgången på hjälpinformation
- redogöra för och använda många vanliga skattningsmetoder vid problem som bortfall t ex multipel och enkel imputering, kalibrering
- på egen hand lösa problemen i vanliga läroböcker på denna nivå inom begränsad tid.
- redogöra för litteraturen och de huvudsakliga tidskrifterna inom området.

Kunskapskontroll och examination och betygskriterier

Momentet examineras genom dels individuella inlämningsuppgifter, dels ett skriftligt prov. Läroböcker, anteckningar och annan litteratur är tillåtna vid det skriftliga provet. En av inlämningsuppgifterna redovisas i form av en muntlig presentation den 14 februari (ev behövs fler tider), då närvaro alltså är obligatorisk. Inlämningsuppgifterna poängbedöms och kan tillsammans (efter normering) ge maximalt 100 poäng. Skrivningen kan också ge maximalt 100 poäng. Förutom poäng åsätts inlämningsuppgifterna även betyget ej godkänd eller godkänd. För godkänt resultat på momentet (betyg A-E) krävs att samtliga inlämningsuppgifter åsatts betyget godkänd, att tentamen gett minst 40 poäng samt att totalpoängen överstiger 100 poäng. Allt skall vara uppnått vid samma "kurstillfälle" dvs under kursen fram till och med första omtentamenstillfälle. Betyget på momentet baseras på en sammanräkning av poängen som alltså maximalt kan vara 200 poäng.

Betygssättning sker efter en sjugradig betygsskala

- A. Utmärkt – motsvarar 180 – 200 poäng
- B. Mycket bra – motsvarar 160 – 179 poäng
- C. Bra – motsvarar 140 – 159 poäng
- D. Tillfredsställande – motsvarar 120 – 139 poäng

- E Tillräckligt – motsvarar 100 – 119 poäng
Fx Otillräckligt – motsvarar 60 – 99 poäng eller
mer än 100 poäng utan att alla inlämningsuppgifter är godkända eller skrivningen har givit mer än 40 poäng
F Helt Otillräckligt – motsvarar mindre än 59 poäng

För att få godkänt på hela kursen krävs lägst betyget E på momentet. Betyget på kursen blir detsamma som på momentet.

Studerande som godkänts på prov får ej genomgå förnyat prov för högre betyg. Undantag görs för de studenter som godkänts på samtliga moment men ej uppnått 100 poäng totalt. Dessa får göra om den skriftliga tentamen. Normalt förfaller dock inlämningsuppgifter och skrivningsresultat efter första omtentamenstillfället. Efter detta skall student som ännu ej godkänts på hela momentet göra aktuell termins uppgifter och ny skriftlig tentamen.

Undervisning

Undervisningen sker vid 9 tillfällen om fyra timmar. Schema framgår av institutionens hemsida (<http://gauss.stat.su.se/site/scheman/vt12/uo.pdf>). För att få gott resultat på kursen är det viktigt att närvara vid samtliga undervisningstillfällen. Av dessa 36 timmar kommer preliminärt cirka 18 timmar att vara av föreläsningstyp, 2 timmar gästföreläsare, 12 timmar räkneövningstyp och 4 timmar presentation. Vid stort antal elever kan ytterligare presentationstillfällen tillkomma. Vid räkneövningarna förutsätts de studerande ha läst igenom och försökt lösa samtliga de övningsexempel som meddelats i förväg. Kursen är fristående och kommer formellt inte att kräva några förkunskaper för provtagning men i praktiken rekommenderas en introduktionskurs i sampling starkt.

Kursen motsvarar 7,5 högskolepoäng vilket i sin tur motsvarar 10 halvtidsveckor eller 200 arbetstimmar. En normal arbetsinsats för en student skulle kunna fördelas enligt följande: 36 timmar undervisningstillfällen, 39 timmar arbete med inlämningsuppgifter, 5 timmar skrivning, 60 timmar självständigt genomräkning av övningsexempel (för hand eller med dator) samt 60 timmar läsning av litteraturen.

Kurslitteratur

Material som utdelas i samband med undervisningen. Föreläsningssanteckningar och annat som läggs ut på nätet räknas som utdelade i samband med undervisningen).
Sharon Lohr, (2010), *Sampling: Design and Analysis*, 2nd edition, kapitel 1-9, 12, 15 Duxbury

Referenslitteratur

Urval – från teori till praktik, Handbok 2008:1, SCB – finns att hämta på
http://www.scb.se/statistik/publikationer/OV9999_2007A01_BR_X99BR0801.pdf.

Lärare

Daniel Thorburn, 08 – 16 29 56, Daniel.thorburn@stat.su.se
Nicklas Pettersson 08 – 16 26 96 Nicklas.pettersson@stat.su.se

Preliminärt kursinnehåll

Allmänt om urvalsmetoder. Vad är sampling? Skillnader mellan modell- och designbaserad inferens. Hjälpinformation. Sannolikhetsurval. Några vanliga urvalsmetoder. Regressionsestimatorer, GREG. Stratifierade urval, optimering, kvoturval. Urval med varierande urvalssannolikheter, HT-skattning, SYG-variansskattning,

olika urvalsmetoder för π ps, kombination val av skattning och urvalsmetod. Klusterurval, indirekta urval, urval vid i förväg okänd ram. Variansskattning, resampling, Taylorlinjarisering. Longitudinella undersökningar, roterande urval, panelurval, förändrings- och nivåskattningar. Estimation vid bortfall, MAR, MCAR, , olika former av imputering och omvägning, propensity score, NMAR-modeller. Andra målvariabler, t ex kvantiler, GINI-koefficienter och domänskattningar. Normalitet, asymptotiska resultat. Inbäddade experiment, subsampling i bortfallet, kvalitetsstudier.