

STAN46, Statistik: Analys av funktionell data, 7,5 högskolepoäng

Statistics: Functional Data Analysis, 7.5 credits

Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Institutionsstyrelsen vid Statistiska institutionen 2018-03-12 att gälla från och med 2018-03-13, vårterminen 2018.

Allmänna uppgifter

Undervisningsspråk: Engelska

Huvudområde

Statistik

Fördjupning

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

Kursens mål

Kunskap och förståelse

För godkänd kurs skall studenten

- visa förståelse för både tekniska och konceptuella aspekter av hantering av dataobservationer som funktioner,
- visa kunskap om och förståelse för metoder som utvidgar multivariata standardmetoder genom att införa en dimensionsfri metod som tar hänsyn till ordningen i observerade variabler,
- visa förståelse för den funktionella dataansats som generaliserar statistisk metoder, såsom principalkomponentanalys och multivariat regression,
- visa förståelse för hur man representerar data med hjälp av funktioner, när observationernas dimension är för hög för att effektivt kunna behandlas som en vektor, och
- visa förståelse derivatfunktionernas roll vid lösande av praktiska problem.

Färdighet och förmåga

För godkänd kurs skall studenten

- visa förmåga att effektivt hantera data med en implicit ordning och/eller är högdimensionella genom att använda funktionell dekomposition, och
- visa färdighet i ett antal algoritmiska tillvägagångssätt, med hjälp av mjukvarupaket, både för att analysera och visualisera funktionella data.

Värderingsförmåga och förhållningssätt

För godkänd kurs skall studenten

- visa insikt om tillgängliga val för visualisering och utforskning av data i ett funktionellt rum,
- visa förmåga att göra bedömningar av vilken typ av tillvägagångssätt som är mest lämpad för det aktuella problemet, och
- kritiskt diskutera resultaten av analys som erhållits med en viss metod.

Kursens innehåll

I många praktiska problem som härrör från områden såsom geofysiska vetenskaper, astronomi, medicin etc. är observationerna ofta mätningar av en kontinuerlig process, vilken är en funktion av tid och/eller rum. Med snabba teknologiska framsteg blir sådana mätningar allt mer vanlig förekommande, vilket kräver en utveckling av högdimensionell dataanalys. Inom dataforskning löser man ofta utmaning med högdimensionalitet av observerade data genom att använda en dimensionsfri analys där man betraktar en observation som en funktion. Detta är ett sätt att i analys kringgå de tekniska svårigheter med högdimensionalitet, som ibland går under benämningen "the curse of dimensionality".

Kursen introducerar deltagarna till olika aspekter av dimensionsfri analys, vilka undanröjer de tekniska och beräkningshinder som hör samman med högdimensionella data. Detta sätt att hantera högdimensionella data är ett framväxande och snabbt utvecklande fält som kräver förståelse för både etablerade metoder och nyutvecklade tekniker. Huvudsyftet med kursen är att fokusera på tillämpning av funktionella dataanalystekniker på verkliga problem, varför praktisk problemlösning prioriteras före matematisk teori.

Kursen börjar med en genomgång av grunderna i dataanalys av data som kan anses vara "funktioner". Vi kommer att diskutera olika visualiserings- och explorativa dataanalystekniker. Statistiska metoder som vi kommer att använda består av icke parametrisk spline-utjämning, funktionella linjära modeller, funktionell PCA, regulariseringsmetoder, analys av derivat, registrering och icke-linjär utjämning.

Studenterna skall arbeta med projekt där de skall tillämpa kursens analystekniker på verkliga problem. Kursen programvara kommer att vara R och/eller Python, men eleverna får använda en annan mjukvara som möjliggör arbete med analys av kursens uppgifter, t.ex. Matlab. Under genomförande av sina projekt kan studenterna diskutera och jämföra idéer med varandra och få särskild vägledning från instruktörerna. Insatser kommer att göras för att hjälpa studenter att formulera integrera verkliga problem i matematiska modeller så att lämpliga algoritmer kan tillämpas med hänsyn till beräkningsbegränsningar. Genom att olika projekt kommer att kräva olika lösningsmetoder kommer studenterna att bli introduceras till ett brett utbud av nya metoder.

Kursens genomförande

Kursen är utformad som en serie föreläsningar, redovisningar och laborationer med rapporter.

Kursens examination

Examinationen utgörs av inlämningsuppgifter enskilt och i grupp, vilka redovisas skriftligt och muntligen.

Lunds universitet ser mycket allvarligt på fusk och kommer att vidta disciplinåtgärder mot alla slags försök till fusk i samband med tentamina eller andra examinationsformer. Plagiering betraktas som ett mycket allvarligt akademiskt brott. Påföljder som universitetet kan utdela för detta, och för andra slags fusk i samband med olika former av examination, inkluderar avstängning från universitetet under en viss tidsperiod.

Om så krävs för att en student med varaktig funktionsnedsättning ska ges ett likvärdigt examinationsalternativ jämfört med en student utan funktionsnedsättning, så kan examinator efter samråd med universitetets avdelning för pedagogiskt stöd fatta beslut om alternativ examinationsform för berörd student.

Prov/moment för denna kurs finns i en bilaga i slutet av dokumentet.

Betyg

Betygsskalan omfattar betygsgraderna Underkänt, E, D, C, B, A.

A (Utmärkt) 85-100 poäng/procent. Ett framstående resultat som är utmärkt vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet.

B (Mycket bra) 75-84 poäng/procent. Ett mycket bra resultat som karakteriseras av mycket bra teoretiskt djup, praktisk relevans, analytisk förmåga samt självständighet.

C (Bra) 65-74 poäng/procent. Ett bra resultat som karakteriseras av bra teoretiskt djup, praktisk relevans, analytisk förmåga samt självständighet.

D (Tillfredsställande) 55-64 poäng/procent. Ett resultat som är tillfredsställande vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet.

E (Tillräckligt) 50-54 poäng/procent. Ett resultat som möter minimikraven vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet, men inte mer.

U (Otillräckligt/Underkänt) 0-49 poäng/procent. Ett resultat som är otillräckligt vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet.

För att få godkänt på en kurs måste studenten få betyg E eller högre.

Förkunskapskrav

Allmänna förkunskaper för masterprogrammet i statistik, samt STAN41 eller en kurs i linjär algebra med matristeori eller motsvarande.

Övrigt

I kursen kommer paketet fda i R eller pyFDA i Python att användas. Paketet Functional Data Analysis finns tillgängligt för Matlab.

Prov/moment för kursen STAN46, Statistik: Analys av funktionell data

Gäller från V19

- 1901 Laboration, 2,0 hp
Betygsskala: Underkänd, Godkänd
- 1902 Inlämningsuppgift, 2,0 hp
Betygsskala: Underkänd, Godkänd
- 1903 Muntlig presentation, 3,5 hp
Betygsskala: Underkänd, Godkänd