



LUNDS
UNIVERSITET

Ekonomihögskolan

**STAN48, Statistik: Programmering för data science, 7,5
högskolepoäng**
Statistics: Programming for Data Science, 7.5 credits
Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Institutionsstyrelsen vid Statistiska institutionen 2021-11-29 att gälla från och med 2022-08-29, höstterminen 2022.

Allmänna uppgifter

Kurs på avancerad nivå som ingår som rekommenderad kurs i en magister- eller masterexamen i statistik. Kursen kan även läsas som fristående eller inom andra magister- och masterprogram vid Lunds universitet.

Undervisningsspråk: Engelska

Huvudområde

Statistik

Fördjupning

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

Kursens mål

Kursen är en modern version av utbildningen i statistiska beräkningsmetoder. Den flyttar fokus från huvudsakligen metodologiska grunder till programmeringskunskaper som möjliggör ett effektivt genomförande av metodiken på ett av de ledande programmeringsspråken inom data science: såsom R eller Python.

Kunskap och förståelse

För godkänd kurs skall studenten

- kunna förklara samspelet mellan teori och dess implementering i en algoritm,
- kunna klargöra konceptualiseringen av data i form av en matematisk modell med stokastiska komponenter,
- kunna formalisera begreppet algoritmisk effektivitet, och
- kunna identifiera fördelar och nackdelar med olika datapaket, såsom R och Python, för ett konkret statistiskt problem.

Färdighet och förmåga

För godkänd kurs skall studenten

- självständigt kunna skriva en algoritm i valda programmeringsspråk, som R och/eller Python, som implementerar en given statistisk beräkningsmetod,
- kunna ladda ner och hantera massiva data från olika källor och ladda ner dem till olika datormiljöer,
- kunna utföra en fullständig analys av data inom ramen för en given metod och rapportera resultatet skriftligt,
- kunna förbättra ineffektivt skrivna algoritmer och visa fördelar med föreslagna lösningar,
- kunna välja ett programmeringspaket som är mest lämpligt för det aktuella problemet, och
- kunna kommunicera betydelsen av resultat från statistisk dataanalys med en icke-professionell och tolka hur den statistiska programvaran representerar dem.

Värderingsförmåga och förhållningssätt

För godkänd kurs skall studenten

- kunna bedöma sitt tillvägagångssätt och kritiskt diskutera och försvara utvalda lösningar,
- kunna dokumentera sitt arbete och framsteg med hjälp av standarder för kommentering och dokumentation, så att andra kan använda och fortsätta arbetet, och
- kunna fördela uppgifter för ett generellt dataanalysproblem mellan medlemmarna i sin grupp.

Kursens innehåll

Kursen presenterar modern statistisk datoranvändning inom data science genom implementeringar i populära programspråk såsom R och Python. Den behandlar följande ämnen:

1. Programmeringsmiljön: bibliotek för statistik
2. Arbeta med *data frames*, *arrays* och matriser
3. Metoder för slumpvalsgenerering
4. Monte Carlo-integration, inferens och variansreducering
5. Bootstrap och återsamlingsmetoder
6. Bayesiansk databehandling
7. Markov Chain Monte Carlo-metoder
8. Optimering och andra numeriska metoder

Kursens genomförande

Kursen är utformad som en serie föreläsningar, övningar och laborationer. Den avslutas med projektpresentation som utgör del av examinationen. Laborationsprojekten och quizen som är kopplade till laborationerna respektive övningarna bidrar till slutbetyget.

Kursens examination

Examinationen utgörs av slutprojektet och dess presentation som står för 50% av

slutbetyget, laborationerna 25% och quizen 25%. Både quiz och laborationer måste göras för att få godkänt betyg på var och en av dessa två komponenter.

Lunds universitet ser mycket allvarligt på fusk och kommer att vidta disciplinåtgärder mot alla slags försök till fusk i samband med tentamina och andra examinationsformer. Plagiering betraktas som ett mycket allvarligt akademiskt brott. Det straff som universitetets disciplinnämnd 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

90 hp i statistik, eller sammanlagt 90 hp i matematik, programmering, matematisk statistik och statistik varav minst 45 hp i statistik eller matematisk statistik, eller motsvarande.

Prov/moment för kursen STAN48, Statistik: Programmering för data science

Gäller från H22

- 2201 Projekt, 3,5 hp
Betygsskala: Underkänd, Godkänd
- 2202 Laborationer, 2,0 hp
Betygsskala: Underkänd, Godkänd
- 2203 Quiz, 2,0 hp
Betygsskala: Underkänd, Godkänd