



LUNDS  
UNIVERSITET

Ekonomihögskolan

## STAN53, Statistik: Högdimensionell dataanalys, 7,5 högskolepoäng

*Statistics: High-dimensional Data Analysis, 7.5 credits*

Avancerad nivå / Second Cycle

---

### Fastställande

Kursplanen är fastställd av Institutionsstyrelsen vid Statistiska institutionen 2022-08-29 (U 2022/536) och senast reviderad 2024-03-07 (U 2024/138). Den reviderade kursplanen träder i kraft 2024-03-08 och gäller från och med höstterminen 2024.

### 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 Fördjupning*

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

### Kursens mål

På ett allmänt plan ska studenterna kunna förstå konceptet att analysera multivariata och högdimensionella data. De ska ha en grundläggande minimumnivå av matriskompetens, känna till allmänna aspekter av att hantera multivariata data och förstå utmaningen i arbeta med data vars dimension är av samma storlek som eller större än antalet observationer. De studenter som klarar kursen, kommer att ha grundläggande förståelse och färdighet för utbudet av multivariata metoder, dimensionsreduktion och regulariseringstekniker, samt kommer att kunna sammanfatta och tolka multivariata och *high-throughput* experimentella data, tillämpa principalkomponentanalys och faktoranalys och visa hur dessa koncept utnyttjas vid visualisering. Studenterna kommer också att kunna använda maskininlärningsmetoder för att analysera *high-throughput* data och dra relevanta slutsatser.

## Kunskap och förståelse

För godkänd kurs skall studenten

- visa kunskap om utbudet av tillgängliga multivariata tekniker med fokus på metoder för högdimensionella och *high-throughput* data,
- visa förståelse för de utmaningar som finns med att använda multivariata tekniker på moderna högdimensionella data, och
- visa fördjupad kunskap om regulariseringsmetoder, klusteranalys och algoritmer för prediktion som *k-nearest neighbours*, tillsammans med koncepten träningsdata, testdata, felrisiker och korsvalidering.

## Färdighet och förmåga

För godkänd kurs skall studenten

- kunna tillämpa regulariseringsmetoder, klusteranalys och algoritmer för prediktion som *k-nearest neighbours*, tillsammans med koncepten träningsdata, testdata, felrisiker och korsvalidering,
- kunna sammanfatta resultaten av analyser, inklusive visualiseringsmetoder, och
- kunna förklara resultaten för en forskare som inte är van vid data.

## Värderingsförmåga och förhållningssätt

För godkänd kurs skall studenten

- identifiera rätt tekniker och beräkningstekniker för att genomföra statistiska analys av multivariata och högdimensionella empiriska data.

## Kursens innehåll

Kursen inleds med en introduktion till matriser och multivariat normalfördelning. Detta följs av singularvärdesuppdelning och dess geometriska tolkning. Sedan behandlas principalkomponentanalys inklusive dess funktionalformulering. Prediktionssteori inklusive prediktion med högdimensionella prediktorer presenteras härnäst med fokus på *penalised* regression och prediktion. Detta följs av singularvärdesuppdelning av glesa matriser och linjär diskriminantanalys. Kursen avslutas med storskalig inferens.

## Kursens genomförande

Kursen utgörs av föreläsningar, övningstillfällen och datorövningar samt ett avslutande seminarium.

## Kursens examination

Examinationen utgörs av quiz, laborationsrapporter och ett avslutande projekt. Projektet presenteras skriftligt och muntligt vid ett avslutande seminarium. Vid seminariet får studenterna redogöra för sin förståelse av de tillämpade metoderna.

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 universitets 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.

## Betyg

Betygsskalan omfattar betygsgraderna: U=Otillräckligt, E=Tillräckligt, D=Tillfredsställande, C=Bra, B=Mycket bra, A=Utmärkt

**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.

Betyg på hel kurs bestäms som en sammanvägning av resultaten på quizen (40 %), laborationsrapporterna (40 %) och det avslutande projektet (20 %).

## Förkunskapskrav

90 hp i statistik, eller motsvarande.

## Övrigt

Kursen ersätter STAN41 Statistik: Multivariat analys. De två kurserna kan inte ingå i en examen tillsammans.