Stockholms universitet

Statistiska institutionen, VT 2025

Statistik och dataanalys I, 15 hp, ST1101

Kursens innehåll

Kursen ger kunskaper om klassiska och moderna statistiska metoder för dataanalys samt dess teoretiska grunder. Centralt är förståelse för hela analysprocessen: från datakällor, datainsamling och datahantering via estimering till inferens, prediktion och andra tillämpningar. Stor vikt läggs vid praktisk datahantering, visualisering och analys genom programmering i R. Genomgående läggs vikt vid ett kritiskt förhållningssätt vid användning av statistiska metoder och tolkning av resultat.

Kursen behandlar:

- Insamlingsmetoder och datakällor. Olika datatyper såsom numeriska och kategoriska men även text, bild och spatiala data. Grafiska och numeriska beskrivningar av data.
- Regressionsanalys; modeller med en och flera förklaringsvariabler, antaganden, skattning, inferens, prediktion, modellutvärdering. Kopplingen till moderna dataanalysmetoder så som maskininlärning tas upp.
- Sannolikhetsteori; grundläggande begrepp, sannolikhetsmodeller, diskreta och kontinuerliga slumpvariabler, sannolikhetsfördelningar, väntevärde och varians, kovarians och korrelation, några olika standardfördelningar, linjärkombinationer av flera slumpvariabler, samplingfördelningar och centrala gränsvärdessatsen.
- Statistisk inferens; punkt- och intervallskattning, hypotestest, p-värden och prediktion, introduktion till likelihood och Bayesiansk inferens, chi2-test.

Kursen examineras genom 4 prov:

Prov 1. SDA I, tentamen 1. Individuell salstentamen, 4.5 hp, provkod 11ST.

Prov 2. SDA I, inlämningsuppgift 1. Inlämningsuppgift i grupp, 3 hp, provkod 12SI.

Prov 3. SDA I, tentamen 2. Individuell salstentamen, 6 hp, provkod 13ST.

Prov 4. SDA I, inlämningsuppgift 2. Inlämningsuppgift i grupp, 1.5 hp, provkod 14SI.

Förväntade studieresultat

För godkänt resultat på kursen ska studenten kunna:

- Bearbeta data, beskriva data numeriskt och grafiskt samt utföra enklare analyser i R.
- Redogöra för olika insamlingsmetoder och datakällor.
- Formulera statistiska modeller och lösa grundläggande problem inom sannolikhetsteori och inferens.
- Utföra regressionsanalys och tolka resultaten.
- Göra lämpliga metodval och kritiskt granska resultat.

Lärare och allmän information

Lärare	Roll
Oskar Gustafsson	Kursansvarig och examinator, föreläsningar del 2
Karl Sigfrid	Föreläsningar del 1
Fredrik Stenkvist	Datorövningar, jour och inlämningsuppgifter
Anders Fredriksson	Datorövningar och inlämningsuppgifter
Ester Lagerlöf	Plenara övningar och jour
Maria Bennis	Plenara övningar och jour

Alla lärare har mottagningstid efter överenskommelse, kontakta oss via e-post för att boka tid för möte (antingen på campus eller via Zoom).

Statistiska institutionen ligger i nybyggda Campus Albano, Albanovägen 12, Hus 4, plan 6. Allmän information som rör institutionen finns på institutionshemsidan. Generell information om kursen samt länk till schema hittas via kurshemsidan.

Kursutvärdering

Efter att kursen är avslutad görs en utvärdering av kursen. Kursutvärderingen används som underlag för kvalitetsarbetet med kursen och som ett led i studentinflytandet. Utvärderingen utförs genom att en enkät skickas via mejl till samtliga registrerade kursdeltagare. Kursdeltagarnas svar på enkäten sammanställs och läggs tillsammans med den kursansvariga lärarens slutrapport/kursutvärdering upp på Athena.

Undervisning och närvaro

Undervisningen består av 24 föreläsningar, 15 plenara räkneövningar och 8 datorövningar enligt schema. Dessutom finns 7 jourtillfällen schemalagda, där det finns möjlighet att ställa frågor och få hjälp med räkneuppgifter. En beskrivning av innehållet samt läsanvisningar/uppgifter från kurslitteraturen ges för respektive undervisningstillfälle på kursens github-sida https://statisticssu.github.io/SDA1/.

- På de två undervisningstillfällena **D1** Del I och **D5** Del II ges information om vad som gäller för inlämningsuppgift I respektive inlämningsuppgift II samt görs indelning i arbetsgrupper. Dessa tillfällen är kurskrav för att få godkänt betyg på Prov 2 (SDA I, inlämningsuppgift 1, 3 hp, provkod 12SI) respektive Prov 4 (SDA I, inlämningsuppgift 2, 1.5 hp, provkod 14SI). Kontakta din seminarielärare före övningstillfället vid förhinder. **Obs!** Detta gäller även för omregistrerade studenter.

Närvaro vid kursens övriga undervisningstillfällen är frivillig men starkt rekommenderad.

Kunskapskontroll och examination

Kursen examineras genom kunskapskontroll av de förväntade studieresultaten. Kunskapskontrollen sker genom de fyra proven listade ovan i avsnittet Kursens innehåll.

Prov 1 och **Prov 3** är individuella skriftlig salstentamina som betygsätts enligt en sjugradig målrelaterad betygsskala:

Sjugradig betygsskala	
A	Utmärkt
B	Mycket bra
C	Bra
D	Tillfredställande
E	Tillräckligt
Fx	Underkänd, något mer arbete krävs
F	Underkänd, mycket mer arbete krävs

Prov 2 och **Prov 4** är inlämningsuppgifter som utförs i grupp och betygsätts enligt en tvågradig betygsskala med betygen godkänd (G) och underkänd (U).

- För godkänt slutbetyg på kursen krävs lägst betyget E på Prov 1 och 3 samt betyget G på Prov 2 och 4.
- Sammanvägt betyg på hela kursen sätts genom ett viktat genomsnitt av resultaten på Prov 1 och Prov 3. Bokstavsbetygen A–E omvandlas till siffrorna 5–1 och sammanräknas till ett medelbetyg där delkurserna viktas med det antal högskolepoäng som respektive prov utgör av poängantalet för Prov 1 och 3 tillsammans. Avrundning sker till närmaste heltal. Se betygsmatrisen på sista sidan i detta dokument.
- Studerande som fått lägst betyget E på prov får inte genomgå förnyat prov för högre betyg.
- Såväl Fx som F på tentamen är underkända betyg och kräver omexamination. Studenter som får betyget Fx kan alltså inte komplettera för högre betyg.
- Studerande som fått betyget Fx eller F på ett prov har rätt att genomgå minst fyra ytterligare prov så länge kursen ges för att uppnå lägst betyget E.
- Studerande som fått betyget Fx eller F på prov två gånger av en examinator har rätt att begära att en annan examinator utses för att bestämma betyg vid nästa provtillfälle. Framställan härom ska göras skriftligt till prefekten.
- För varje prov på kursen ges två examinationstillfällen innevarande termin.

Regler för Prov 1 – SDA I tentamen 1 (4.5 hp) och Prov 3 - SDA I tentamen 2 (6 hp)

Prov 1 och Prov 3 är individuella skriftliga tentamina.. Följande gäller för båda tentamina:

- Från och med det datum som tentamensresultatet meddelas finns tentorna att hämta ut på studentexpeditionen under deras ordinarie öppettider (se hemsidan) eller skickas elektroniskt. Lösningsförslag läggs upp på Athena senast detta datum.
- Salstentor är individuella och under tentorna är samarbete otillåtet liksom andra hjälpmedel än de som medgetts av examinator. För denna kurs är endast miniräknare utan lagrad text och data tillåtet. Mobiltelefoner med räknarfunktion är inte ett tillåtet hjälpmedel. Häftet ”Formelblad och tabellsamling” bifogas tentamen och återlämnas efter skrivningen.
- Särskilt stöd kan vid behov tillåtas efter begäran hos institutionens studie- och karriärvägledare och efter medgivande av examinator. Kontakta studie- och karriärvägledare i god tid innan tentamen äger rum, helst inte senare än tre veckor innan tentamen äger rum.
- Regler som gäller för tentamensskrivningar vid Stockholms universitet finns på följande länk: Regler för salstentamen.

Obs!

Du måste själv anmäla dig till tentamen. Anmälan ska normalt göras senast 10 dagar innan tentamen, men i vissa fall är deadline tidigare än så. Kontrollera därför alltid i Ladok vilken deadline som gäller för varje enskild tentamen. Anmälan öppnar 30 dagar innan tentamen. Har du anmält dig korrekt genereras en anonymkod direkt på skärmen. Om du får problem med din anmälan, kontakta studentexpeditionen expedition@stat.su.se före sista anmälningsdag. **Om du missar att anmäla dig kan du inte tentera!**

Tentamenstillfällen

- **Tentamen del 1**: 13 februari, kl. 14–18
- **Tentamen del 2**: 19 mars, kl. 14–19
- **Omtentamen del 1**: 26 mars, kl. 8–12
- **Omtentamen del 2**: 13 maj, kl. 8–13

Regler för Prov 2 - SDA I inlämningsuppgift 1 (3 hp) och Prov 4 – SDA I inlämningsuppgift (1.5 hp)

Inlämningsuppgifterna görs som grupparbeten med 3 personer per grupp. Indelning i grupper gör ni själva i Athena i dokumentet "Skapa arbetsgrupp" som ligger i mappen Inlämningsuppgift 1 respektive 2. Instruktioner för uppgifterna kommer att finnas tillgängliga på Athena vid kursstart.

Samarbete inom gruppen är självklart tillåtet, dock kan individuell prövning och betygsättning inom gruppen förekomma. Observera att samtliga gruppmedlemmar står för och ska kunna svara för alla delar av arbetet som redovisas i rapporten. Alla gruppmedlemmar ska deltaga i alla delar av arbetet: problemförståelse, programmering, analys och rapportskrivande. Samarbete mellan grupperna är tillåtet men varje grupp ska lämna in sin unika rapport.

Prov 2 betygsätts med Godkänd (alla deluppgifter i båda projekt godkända) eller Underkänd (minst en deluppgift underkänd). Om en eller flera deluppgifter får underkänt betyg ges en chans till komplettering innevarande termin.

Obs! Det är inte möjligt att komplettera om inlämningen sker först vid andra tillfället. Detta innebär att om man missar en inlämning och istället lämnar in vid den senare tidpunkten och då får underkänt, kan man inte komplettera uppgiften.

Obs! Samtliga deluppgifter ska vara behandlade och godkända under innevarande termin för att hela inlämningsuppgiften ska kunna godkännas. Resultat från deluppgifter sparas inte och kan inte överföras till kommande terminer.

AI-verktyg och plagiering

- Användning av AI-verktyg är tillåtet som hjälp när det gäller kunskapsinhämtning och instudering men inte för att producera material till någon form av examination.
- All typ av plagiering är otillåten. Detta innefattar även text genererad av AI-verktyg. Användning av AI-verktyg för förbättring av en ursprungligen egenskriven text är otillåten.
- Programvara för textmatchning och detektorer för AI-genererad text används av institutionen. Läs Riktlinjer för diciplinärenden vid Stockholms universitet.

Inlämningstillfällen

- **Deadline för inlämning del 1**: 11 februari 18.00. Återlämnas 18 februari. Deadline för eventuell komplettering: 25 februari 18.00.
- **Deadline för inlämning del 2**: 17 mars 18.00. Återlämnas 24 mars. Deadline för eventuell komplettering: 31 mars 18.00.

Betygskriterier

Prov 1 – SDA I tentamen 1 (4.5 hp)

Prov 1 är en individuell skriftlig datortentamen. Den skriftliga tentamen behandlar material i enlighet med kursinnehållet. Följande betygskriterier gäller:

Betygskriterier Prov 1, provkod 11ST

A (Utmärkt): Studenten kan på ett utmärkt sätt använda begrepp och satser inom dataanalys och regression som inte nödvändigtvis direkt behandlas i kursdelen. Studenten kan på ett korrekt och välstrukturerat sätt lösa och tolka problem av fördjupad karaktär inom dataanalys och regression. Kräver minst 90 poäng på den skriftliga tentamen.

B (Mycket bra): Studenten kan på ett mycket bra sätt använda begrepp och satser inom dataanalys och regression som behandlas i kursdelen. Studenten kan på ett korrekt och välstrukturerat sätt lösa och tolka problem av fördjupad karaktär inom dataanalys och regression. Kräver 80-89 poäng på den skriftliga tentamen.

C (Bra): Studenten kan på ett bra sätt använda begrepp och satser inom dataanalys och regression som behandlas i kursdelen. Studenten kan på ett korrekt och välstrukturerat sätt lösa och tolka problem av fördjupad karaktär om fördelningar, ställa upp enkla statistiska modeller för några konkreta situationer. Kräver 70-79 poäng på den skriftliga tentamen.

D (Tillfredsställande): Studenten kan på ett tillfredsställande sätt använda begrepp och satser inom dataanalys och regression som behandlas i kursdelen. Studenten kan på ett korrekt sätt lösa och tolka problem av fördjupad karaktär inom dataanalys och regression. Kräver 60-69 poäng på den skriftliga tentamen.

E (Tillräckligt): Studenten kan på ett tillräckligt sätt använda begrepp och satser inom dataanalys och regression som behandlas i kursdelen. Studenten kan på ett mestadels korrekt sätt lösa och tolka problem av fördjupad karaktär inom dataanalys och regression. Kräver 50-59 poäng på den skriftliga tentamen.

Fx (Underkänd, något mer arbete krävs): Motsvarar 40-49 poäng på den skriftliga tentamen. Omtentamen krävs.

F (Underkänd, mycket mer arbete krävs): Motsvarar 0-39 poäng på den skriftliga tentamen. Omtentamen krävs.

Prov 2 – SDA III inlämningsuppgift 1 (1.5 hp), provkod 12SI

Betygssättning på inlämningsuppgiften sker med betygen Godkänd (G) eller Underkänd (U). Följande betygskriterier gäller:

Betygskriterier Prov 2

G (Godkänd): Studenten har ställt upp lämpliga statistiska modeller för givna situationer, har visat tillräcklig förmåga att använda grundläggande terminologi inom dataanalys, har använt den statistiska programvaran R på ett korrekt sätt, samt har presenterat resultaten i en skriftlig rapport i enlighet med instruktionerna. Studenten skall ha varit närvarande vid alla obligatoriska undervisningstillfällen.

U (Underkänd): Studentens prestationer är otillräckliga med avseende på minst ett av kriterierna för G.

Prov 3 – SDA I tentamen 2 (6 hp), provkod 13ST

Prov 3 är en individuell skriftlig datortentamen. Den skriftliga tentamen behandlar material i enlighet med kursinnehållet. Följande betygskriterier gäller:

Betygskriterier Prov 3

A (Utmärkt): Studenten kan på ett utmärkt sätt använda begrepp och satser inom den sannolikhetslära och statistisk inferensteori som inte nödvändigtvis direkt behandlas i kursdelen. Studenten kan på ett korrekt och välstrukturerat sätt beräkna och tolka punkt- och intervallskattningar av modellparametrar samt utföra hypotesprövning av modellparametrar. Kräver minst 90 poäng på den skriftliga tentamen.

B (Mycket bra): Studenten kan på ett mycket bra sätt använda begrepp och satser inom den sannolikhetslära och statistisk inferensteori som behandlas i kursdelen. Studenten kan på ett korrekt och välstrukturerat sätt beräkna och tolka punkt- och intervallskattningar av modellparametrar samt utföra hypotesprövning av modellparametrar. Kräver 80–89 poäng på den skriftliga tentamen.

C (Bra): Studenten kan på ett bra sätt använda begrepp och satser inom den sannolikhetslära och statistisk inferensteori som behandlas i kursdelen. Studenten kan på ett korrekt och välstrukturerat sätt beräkna och tolka punkt- och intervallskattningar av modellparametrar. Kräver 70–79 poäng på den skriftliga tentamen.

D (Tillfredsställande): Studenten kan på ett tillfredsställande sätt använda begrepp och satser inom den sannolikhetslära och statistisk inferensteori som behandlas i kursdelen. Studenten kan på ett korrekt sätt beräkna och tolka punkt- och intervallskattningar av modellparametrar. Kräver 60–69 poäng på den skriftliga tentamen.

E (Tillräckligt): Studenten kan på ett tillräckligt sätt använda begrepp och satser inom den sannolikhetslära och statistisk inferensteori som behandlas i kursdelen. Studenten kan på ett mestadels korrekt sätt beräkna och tolka punkt- och intervallskattningar av modellparametrar. Kräver 50–59 poäng på den skriftliga tentamen.

Fx (Underkänd, något mer arbete krävs): Motsvarar 40–49 poäng på den skriftliga tentamen. Omtentamen krävs.

F (Underkänd, mycket mer arbete krävs): Motsvarar 0–39 poäng på den skriftliga tentamen. Omtentamen krävs.

Prov 4 – SDA I inlämningsuppgift 2 (1.5 hp), provkod 14SI

Betygssättning av inlämningsuppgiften sker med betygen Godkänd (G) eller Underkänd (U). Följande betygskriterier gäller:

Betygskriterier Prov 4

G (Godkänd): Studenten har visat tillräcklig förmåga att beräkna och tolka punkt- och intervallskattningar och utföra hypotesprövning av modellparametrar, har visat tillräcklig förmåga att använda grundläggande statistisk terminologi, har använt den statistiska programvaran R på ett korrekt sätt, samt har presenterat resultaten i en skriftlig rapport i enlighet med instruktionerna. Studenten skall ha varit närvarande vid alla obligatoriska undervisningstillfällen.

U (Underkänd): Studentens prestationer är otillräckliga med avseende på minst ett av kriterierna för G.

Kurslitteratur

- De Veaux, R., Velleman, P. och Bock, D. *Stats: Data and Models*. Pearson Global Edition, femte upplagan.

Övrigt kursmaterial såsom extramaterial, föreläsningsanteckningar, övningsuppgifter, instruktioner till inlämningsuppgifter med mera läggs upp på kursens sida i Athena och på kurshemsidan. På kurshemsidan finns även ett preliminärt schema över innehåll respektive litteratur alt. övningsuppgifter för varje undervisningstillfälle.

Betygsmatris

Betyg tentamen 2 \ Betyg tentamen 1	E	D	C	B	A
E	E	E	D	D	C
D	D	D	D	C	C
C	D	C	C	C	B
B	C	C	B	B	B
A	C	B	B	A	A