



Stockholms
universitet

STOCKHOLMS
UNIVERSITET
Statistiska institutionen
VT 2024

Oscar Oelrich
2023-12-16

Kursbeskrivning för Statistik och dataanalys I, 15 högskolepoäng, 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 och datainsamling, datahantering, estimering, inferens, prediktion och praktiska 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 dataanalytiska metoder såsom maskininlärning tas upp.
- Sannolighetsteori; 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, χ^2 -test.

Kursen består av två delar som examineras genom fyra prov:

1. SDA I, tentamen 1, tentamen, 4.5 hp, provkod 11ST
2. SDA I, inlämningsuppgift 1, 3 hp, provkod 12SI
3. SDA I, tentamen 2, tentamen, 6 hp, provkod 13ST
4. SDA I, inlämningsuppgift 2, 1.5 hp, provkod 14SI

KURSENS LÄRANDEMÅL OCH FÖRVÄNTADE STUDIERESULTAT

Efter att ha genomgått kursen förväntas 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 sannolikhets teori och inferens
- utföra regressionsanalys och tolka resultaten
- göra lämpliga metodval och kritiskt granska resultat

KURSLITTERATUR OCH ÖVRIGA LÄROMEDEL

De Veaux, R., Velleman, P. och Bock, D. (2021). Stats: Data and Models, Pearson Global Edition. 5:e upplagan.

Övrigt kursmaterial såsom extra övningsuppgifter, instruktioner till inlämningsuppgifter, tidigare tentor samt formel- och tabellsamling läggs upp på Athena och på kurshemsidan <https://statisticssu.github.io/SDA1/>.

EXAMINATOR, KURSANSVARIG OCH ALLMÄN INFORMATION

Kursansvarig och examinator: Oscar Oelrich

E-post: oscar.oelrich@stat.su.se

Mottagningstid: Efter överenskommelse.

Föreläsare, del 1: Karl Sigfrid.

E-post: karl.sigfrid@stat.su.se

Mottagningstid: Efter överenskommelse.

Föreläsare, del 2: Oscar Oelrich (se ovan),

Plenara övningar och Jour: Mona Sfaxi

E-post: mona.sfaxi@stat.su.se

Mottagningstid: Efter överenskommelse.

Datorövningar: Sofia Edlund

E-post: sofia.edlund@stat.su.se

Mottagningstid: Efter överenskommelse.

Datorövningar: Sebastian Hedberg

E-post: sebastian.hedberg@stat.su.se

Mottagningstid: Efter överenskommelse.

Statistiska institutionen ligger i nybyggda Campus Albano, Albanovägen 12, Hus 4, plan 6. Allmän information om institutionen finns på www.statistics.su.se Information om kursen hittas via kurshemsidan: <https://www.su.se/sok-kurser-och-program/st1101-1.621586>

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

Undervisningen består av 24 **föreläsningar** (F1-F10 på Del I, F11-F24 på Del II), 15 **plenara statistikövningar** (Ö1-Ö6 på del I och Ö7-Ö15 på del II), och 8 **datorövningar** (D1, D2, D3, D4a och D4b på Del I och D5-D7 på Del II) enligt schema, se länk till Time edit på kurshemsidan, <https://statisticssu.github.io/SDA1/>.

En beskrivning av innehållet samt läsanvisningar till kurslitteraturen ges i avsnittet Undervisningsplan nedan.

Under kursens gång finns **jour** schemalagd en gång per vecka, det är frivilligt att delta. Under jouden finns möjlighet att få hjälp med räkneuppgifter eller fråga om kursinnehåll man inte förstår.

OBLIGATORISK NÄRVARO

- 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.

KUNSKAPSKONTROLL OCH EXAMINATION

Kursen examineras genom kunskapskontroll av de förväntade studieresultaten. Kunskapskontrollen sker genom skriftliga prov i form av individuella salstentor och skriftliga redovisningar av gruppuppgifter. Betygssättning av Prov 2 och Prov 4 (inlämningsuppgift 1 och 2) sker enligt en tvågradig betygsskala med betygen godkänd (G) och underkänd (U). Betygssättning av Prov 1 och Prov 3 (tentamen 1 och 2) sker enligt en sjugradig målrelaterad betygsskala:

- A = Utmärkt
- B = Mycket bra
- C = Bra
- D = Tillfredsställande
- E = Tillräckligt
- Fx = Underkänd, något mer arbete krävs
- F = Underkänd, mycket mer arbete krävs

Sammanvägt betyg på hela kursen

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.

		Betyg tentamen 1				
		E	D	C	B	A
Betyg tentamen 2	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

Övrigt

- 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 skriftligt göras till prefekten.
- För varje prov på kursen ges två examinationstillfällen innevarande termin.

Datum och regler för tentamina (Prov 1 och Prov 3)

Tentamen 1	Fredag den 9 februari kl. 14.00–18.00
Omtentamen 1	Onsdag den 27 mars kl. 08.00–12.00
Tentamen 2	Söndag den 17 mars kl. 09.00–14.00
Omtentamen 2	Lördag den 27 april kl. 14.00–19.00

- Från och med det datum som tentamensresultatet meddelas finns tentorna att hämta ut på studentexpeditionen under deras ordinarie öppettider (se hemsidan) eller skickas elektroniskt. Lösningförslag 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:
<https://www.su.se/medarbetare/organisation-styrning/styrdokument-regelboken/utbildning/regler-för-salstentamen-1.513161>

Obs! Kom ihåg att du måste **anmäla dig senast 10 dagar innan tentamen i Ladok**. Anmälan till tentamen öppnar 30 dagar innan tentamensdatum. Om du missar att anmäla dig kan du inte tentera!

Datum och regler för inlämningsuppgifterna (Prov 2 och Prov 4)

Inlämningsuppgift 1 Inlämning 1: Mån 12/2 kl 12:00
Återlämning: Mån 19/2 kl 18:00
Inlämning 2/Komplettering: Mån 26/2 kl 18:00

Inlämningsuppgift 2 Inlämning 1: Tis 19/3 kl 17:00
Återlämning: Tis 26/3 kl 18:00
Inlämning 2/Komplettering: Tis 4/4 kl 18:00

- 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.
- Uppgiften skall redovisas i form av en skriftlig rapport. Instruktioner för uppgiften finns på Athena.
- 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 delta i alla delar av arbetet: problemförståelse, programmering, analys och rapportskrivande. Samarbete mellan grupperna är tillåtet men varje grupp ska lämna in sin unika rapport. Alla typer av plagiering är otillåtna (detta inefattar även text genererad av AI, eller förbättrad med hjälp av AI) och programvara för textmatchning kan vid behov komma att användas. Läs ”Riktlinjer för dciplinärenden vid Stockholms universitet”
<https://www.su.se/medarbetare/organisation-styrning/styrdokument-regelboken/utbildning/regler-och-handläggningsordning-för-disciplinärenden-1.605869>
- Om en inlämningsuppgift underkänns efter inlämning 1 får man möjlighet att komplettera. Kompletteringar ska lämnas in senast vid tidpunkten för inlämning 2 som anges ovan och

beräknas vara rättade senast fem arbetsdagar senare. Återlämning sker efter överenskommelse med seminarieläraren eller kursansvarig.

- 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.
- **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 inom resp. inlämningsuppgift 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.
- **Obs!** Fullgjord obligatorisk närvaro krävs för godkänt betyg, se betygskriterier nedan.

BETYGSKRITERIER

Prov 1. Tentamen 1 4.5 hp, provkod 11ST

Prov 1 är en individuell skriftlig tentamen. Den skriftliga tentamen behandlar materialet enligt kursinnehållet. Skrivningen ger maximalt 100 poäng. Skrivtiden är fyra timmar. Betygssättning på tentamen sker enligt en sjugradig målrelaterad betygsskala. För godkänt resultat finns betygen A, B, C, D och E, där A är högst och E är lägst. För underkänt resultat finns Fx och F. Ingen komplettering är möjlig vid underkänt betyg. Följande betygskriterier gäller:

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. Inlämningsuppgift 1, 3 hp, provkod 12SI

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

G: 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: Studenten har felaktigt utfört analyser och/eller har feltolkat resultaten och/eller har misslyckats med att använda grundläggande terminologi inom dataanalys i samband med lösandet av inlämningsuppgiften. Om studenten inte har varit närvarande vid alla obligatoriska undervisningstillfällen meddelas betyget underkänd.

Prov 3. Tentamen 2, 6 hp, provkod 13ST

Prov 3 är en individuell skriftlig tentamen. Den skriftliga tentamen behandlar materialet enligt kursinnehållet. Skrivningen ger maximalt 100 poäng. Skrivtiden är fem timmar. Betygsättning på tentamen sker enligt en sjugradig målrelaterad betygsskala. För godkänt resultat finns betygen A, B, C, D och E, där A är högst och E är lägst. För underkänt resultat finns Fx och F. Ingen komplettering är möjlig vid underkänt betyg. Följande betygskriterier gäller:

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. Inlämningsuppgift 2, 1.5 hp, provkod 14SI

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

G: 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: Betyget underkänt meddelas om studenten har visat otillräcklig förmåga att beräkna och tolka punkt- och intervallskattningar och utföra hypotesprövning av modellparametrar och/eller har feltolkat resultaten och/eller har misslyckats med att använda grundläggande statistisk terminologi/begrepp i samband med lösandet av inlämningsuppgiften. Om studenten inte har varit närvarande vid alla obligatoriska undervisningstillfällen meddelas betyget underkänd.

UNDERVISNINGSPLAN

Ett preliminärt schema över innehåll respektive litteratur alt. övningsuppgifter vid varje undervisningstillfälle finns på kurshemsidan: <https://statisticssu.github.io/SDA1/>