Stockholms universitet

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Höstterminen 2024

Michael Carlson (examinator)
2024-10-01

Kursbeskrivning för
Statistik och dataanalys II, 15 högskolepoäng, ST1202

KURSENS INNEHÅLL

Kursen är en fortsättning på kursen Statistik och dataanalys I (SDA I) och ger fördjupade och utvidgade kunskaper om klassiska och moderna statistiska metoder för dataanalys samt undersökningsmetodik. Kursens lärandemål och förväntade studieresultat specificeras i kursplanen. Kursen består av två delar:

Del 1: Regressions- och tidserieanalys, totalt 10 hp

Kursdelen bygger vidare på regressionsanalysen som introducerades på kursen SDA I. Inledningsvis ligger fokus på linjär regression, men med ett större fokus på regressionsmodeller som sannolikhetsmodeller där vi använder oss av teorin för punktskattning, intervallskattning, och hypotesprövning från andra halvan av SDA I. Stor vikt läggs på att kontrollera modellantaganden och att undersöka konsekvenserna av avvikelser från dessa. Vidare expanderas klassen av regressionsmodeller till att inkludera logistisk regression och Poissonregression. Ett antal metoder och tekniker från maskininlärning – exempelvis regularisering, kNN, och icke-linjär regression – introduceras. En naturlig konsekvens av att introducera en omfattande statistisk verktygslåda är att vi behöver lära oss sätt att välja mellan modeller; för detta ändamål introduceras F-test men även andra metoder såsom korsvalidering ingår. Kursdelen behandlar även analys och modellering av tidsseriedata. Tidsseriedelen täcker grundläggande koncept såsom säsongsrensning och trender, men även exponentiell utjämning samt en kortfattad introduktion till ARIMA-modeller ingår. Såväl regressions- som tidsseriemomentet betonar prediktion och klassificiering. Genomgående läggs stor vikt vid ett kritiskt förhållningssätt till modellering och inferens samt vid tillämpad analys och förmåga att kunna implementera modeller i programspråket R.

Del 2: Undersökningsmetodik, totalt 5 hp

För att kunna analysera egenskaper och attityder hos individer, företag och andra typer av objekt behövs data. Undersökningsmetodik handlar om just detta, hur man samlar in användbara data på bästa sätt i syfte att kunna göra relevanta analyser utifrån givna frågeställningar. Grundläggande kunskaper i undersökningsmetodik utgör en viktig del av den statistiska kunskapsbanken. I kursdelen ingår bl.a.: hur man planerar en statistisk undersökning; operationalisering av frågeställningar; olika insamlingsmetoder; användning av officiell statistik och registerdata; konstruktion av frågeformulär; kodning; datahantering och editering; olika feltyper, och särskilt mät- och bortfallsfel; kvalitetskriterier och utvärdering; experiment och kvasi-experimentella undersökningar; skyddet av konfidentiella uppgifter. En kort introduktion till s.k. sannolikhetsurval dragna från ändliga populationer och parameterskattningar ges. Övningar med R-programmering ingår.

LÄRARE OCH ALLMÄN INFORMATION

Statistiska institutionen finns på Campus Albano i Hus 4, plan 6. Allmän information om institutionen (expeditionstider, studie- och karriärvägledning, telefonnummer, schema etc.) finns utlagd på institutionens hemsida, www.statistics.su.se. Kursspecifik information förmedlas huvudsakligen via Athena samt i samband med undervisningen.

Lärare	Rum	E-post
Michael Carlson, kursansvarig & examinator - föreläsare & tentamen Del 2	A4631	michael.carlson@stat.su.se
Karl Wahlin - föreläsare & tentamen Del 1	---	karl.wahlin@stat.su.se
Anders Fredriksson - datorlabbar, inlämningsuppgift	A4629	anders.fredriksson@stat.su.se
Fredrik Stenkvist - datorlabbar, inlämningsuppgift	A4629	fredrik.stenkvist@stat.su.se

Samtliga lärare har mottagning efter överenskommelse.

KURSLITTERATUR OCH ÖVRIGA LÄROMEDEL

Del 1: Regressions- och tidsserieanalys

- Chatterjee, S och, A.S. (2012). *Regressions Analysis by Example*. 5:e upplagan. Wiley.
- Dahmström, K. (2011). *Från datainsamling till rapport – att göra en statistisk undersökning*. 5:e upplagan. Studentlitteratur, Lund. Endast avsnitt 10.5 - 10.6.
- Hyndman, R. J. och Athanasopoulos, G. (2021) *Forecasting: principles and practice*. Otexts. 3:e upplagan. Elektronisk resurs: länk.

Del 2: Undersökningsmetodik

Huvudlitteratur:

- Dahmström, K. (2011). *Från datainsamling till rapport – att göra en statistisk undersökning*. 5:e upplagan. Studentlitteratur, Lund. Kapitel 1-2, (3), 4-9, (10), 12-13, 16, (17)

Kompletterande litteratur med länkar (finns även på Athena):

- Statistiska centralbyrån 2001. *Fråga rätt! Utveckla, testa, utvärdera och förbättra blanketter*. Elektronisk resurs: länk. Endast kap 1-5.
- Svenska statistikfrämjandet (2023). *Svenska statistikfrämjandets etiska kod för statistiker och statistisk verksamhet*. Elektronisk resurs: länk.
- Surveyföreningen 2014. *Kvalitet i webbpanelundersökningar*. Elektronisk resurs: länk. Endast kap 2-4.

Övrigt kursmaterial såsom föreläsningsanteckningar, instruktioner, övningstentor, hjälpmedel vid tentamen såsom formel- och tabellsamling, mm. läggs löpande upp på Athena.

UNDERVISNING

Schema finns på länk. Läsanvisningar ges i en särskild **Läsanvisning** som är tillgänglig på Athena senast vid kursstart. Undervisningen består av:

Del 1: Regressions- och tidsserieanalys

- 16 föreläsningar (F)
- 8 övningar (Ö)
- 5 datorlaborationer (D)
- 5 jourpass

Del 2: Undersökningsmetodik

- 8 föreläsningar (F)
- 3 kombinerade seminarier/föreläsningar (S1-3)
- 1 seminarium med muntlig redovisning av inlämningsuppgiften (S4)
- 1 datorlaboration (D)
- 3 jourpass

Jourpassen är schemalagda 2-timmarspass då ni kan träffa övningslärarna för att ställa frågor och få hjälp med studierna och inlämningsuppgifterna.

OBLIGATORISK NÄRVARO

För all undervisning gäller att närvaro är frivillig. Detta betyder att du som student själv bestämmer vilka undervisningstillfällen som du vill närvara vid.

Observera dock att **seminarium S4 i kursdel 2 är obligatorisk** då den utgör en del av examinationen.

Närvaro rekommenderas dock vid följande undervisningstillfällen (se schema för tid och plats):

Del 1: Regressionsanalys och tidsserieanalys

- Föreläsning F1 - information om kursen och kursdelen, upplägg, mm.
- Övning Ö1 - indelning i arbetsgrupper och information om inlämningsuppgiften.
- Datorlabbar D1-D5 - lärarledd instruktion och hjälp med inlämningsuppgiften.

Del 2: Undersökningsmetodik

- Föreläsning F1 - information om kursdelen, upplägg, mm.
- Datorlabb D1 - lärarledd instruktion och hjälp med inlämningsuppgiften.,
- Seminarium S4 - **Obligatorisk!** Muntlig presentation (examination) av inlämningsuppgift.

SÄRSKILT PEDAGOGISKT STÖD

Särskilt pedagogiskt stöd i samband med undervisningen och/eller examinationen kan vid behov tillåtas efter begäran hos institutionens studie- och karriärvägledare och efter medgivande av examinator. Stöd kan t.ex. vara förlängd skrivtid, att få sitta i enskild sal, språklexikon mm. Kontakta institutionens studie- och karriärvägledare (studievagledare@stat.su.se) i god tid innan tentamen äger rum, minst tre veckor innan tentamen äger rum rekommenderas.

KUNSKAPSKONTROLL OCH BETYGSÄTTNING

Kursen examineras genom totalt fyra prov:

Prov 1. SDA II, tentamen 1: Regressions- och tidserieanalys, 7.5hp
Prov 2. SDA II, inlämningsuppgift 1: Regressions- och tidserieanalys, 2.5hp
Prov 3. SDA II, tentamen 2: Undersökningsmetodik, 3.5hp
Prov 4. SDA II, inlämningsuppgift 2: Undersökningsmetodik, 1.5hp

Det två kursdelarna examineras var för sig genom en skriftlig individuell salstentamen och en inlämningsuppgift i form av ett grupparbete med tre-fyra studenter per arbetsgrupp. Övningstentor tillhandahålls som exempel på hur tentamen kan vara utformad. Särskilda anvisningar för inlämningsuppgifterna delas ut i samband med undervisningen. Inlämningsuppgiften i undersökningsmetodik ska redovisas både skriftligt och muntligt. Antiplagieringsverktyg kan komma att användas vid rättningen.

Tentorna (prov 1 och 3) bedöms var för sig enligt en sjugradig målrelaterad betygsskala där A = Utmärkt, B = Mycket bra, C = Bra, D = Tillfredsställande, E = Tillräckligt, Fx = Otillräckligt och F = Helt otillräckligt.

Inlämningsuppgifterna (prov 2 och 4) bedöms var för sig enligt en två-gradig målrelaterad betygsskala, där G = godkänd och U = underkänd. Den enskildes prestation bedöms genom en dokumenterad beskrivning av arbetsfördelningen inom gruppen samt genom individuella deluppgifter.

Se även avsnittet *Beskrivning av proven och betygskriterier* nedan.

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 i enlighet med kursplanen genom ett viktat genomsnitt av resultaten på prov 1 och prov 3 enligt följande tablå:

		Betyg prov 3				
		E	**D**	**C**	**B**	**A**
Betyg prov 1	**E**	E	E	D	D	D
	D	D	D	D	C	C
	C	D	C	C	C	B
	B	C	C	B	B	B
	A	B	B	B	A	A

Övriga villkor

- Såväl Fx som F är underkända betyg och kräver omexamination. Möjlighet till komplettering av betyget Fx upp till godkänt betyg medges inte på denna kurs.
- Studerande som har fått ett underkänt betyg på något av proven har rätt att genomgå ytterligare prov så länge kursen ges för att uppnå ett godkänt betyg.
- Studerande som fått betyget E-B på prov 1 eller 3 får inte genomgå förnyat prov för högre betyg.
- Examinationsuppgifter som inte lämnas in i tid bedöms ej.
- Studerande som fått ett underkänt betyg på ett prov två gånger i rad av en och samma examinator har rätt att få en annan examinator utsedd vid nästkommande examinationstillfälle för det aktuella provet, om inte särskilda skäl talar emot det. Framställan om detta ska göras skriftligt till prefekten vid Statistiska institutionen.

EXAMINATIONSTILLFÄLLEN OCH VILLKOR

För varje prov på kursen ges två examinationstillfällen i anslutning till innevarande termin.

Prov 2 Inlämningsuppgift: - Regressions- och tids-serieanalys	Inlämning av skriftlig rapport: *fastställs inom kort* Återkoppling: *fastställs inom kort* Andra chans/komplettering: *fastställs inom kort*
Prov 4 Inlämningsuppgift: - Undersökningsmetodik	Inlämning av skriftlig rapport: *fastställs inom kort* Muntlig redovisning: **fredag 10 januari (S4)**, tider se schema Återkoppling: *fastställs inom kort* Andra chans/komplettering: *fastställs inom kort*

- Missar du första inlämningsdatumet för en inlämningsuppgift har du en chans till på dig att lämna in (andra chansen) vilket likställs med ett andra examinationstillfälle.
- Om inlämningsuppgiften underkänns vid första inlämningstillfället har du möjlighet att komplettera och rätta till fel. En reviderad version av rapporten lämnas in senast den tidpunkt som anges ovan (komplettering).

  Obs! Om du missar första inlämningsdagen och får underkänt, kan du inte komplettera innevarande termin utan får återkomma nästa gång kursen ges eller senare.
- Missar du den muntliga redovisningen av Prov 4 ges möjlighet att särredovisa efter överenskommelse med seminarieläraren eller examinatorn. Detta tillfälle jämställs med ett andra examinationstillfälle. Särredovisning måste i så fall ske senast fredagen den **14 juni 2023**.
- 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.

Prov 1 Tentamen – Regressions och tidsserieanalys	måndag 9 december, 2024, kl. 10-15
Sista anmälningsdag	onsdag 29 november, 2024
Omtentamen – Regressions och tidsserieanalys	*ej fastställt*
Prov 3 Tentamen – Undersökningsmetodik	torsdag 16 januari, 2024, kl. 8-12
Sista anmälningsdag	torsdag 19 december, 2025
Omtentamen – Undersökningsmetodik	*ej fastställt*

- Anmälan till tentamen är obligatorisk och krävs för att få delta vid tentamen samt omtentamen. Det betyder att du måsta anmäla dig i förväg, annars får du inte delta. Se även *Tentamensinformation* på länken Under utbildningen - Statistiska institutionen (su.se).
- Anmälan ska göras via studenttjänsten i Ladok (https://student.ladok.se). Anmälan till tentamen öppnar ca 30 dagar före tentamensdatum.
- Anmäl dig i god tid och kontrollera att du är anmäld. Om det uppstår problem vid anmälan måste du mejla expedition@stat.su.se **senast sista anmälningsdag** (se ovan).
- Om du är omregistrerad på en gammal kurskod kan du endast anmäla dig via mejl till expedition@stat.su.se. Kontakta studentexpeditionen i god tid innan sista anmälningsdag.
- För att få delta vid tentamen måste du kunna visa upp giltig ID-handling vid tentamenstillfället.

BESKRIVNING AV PROVEN OCH BETYGSKRITERIER

Inlämningsuppgifterna

Villkor för proven, datum för inlämning, betygsättning mm. beskrivs i avsnitten *Kunskapskontroll och betygsättning* och *Examinationstillfällen* ovan.

Prov 2: Regressionsanalys och tidsserieanalys, inlämningsuppgift, 2.5 högskolepoäng

Provet utförs i form av ett grupparbete med 3-4 studenter per grupp och går ut på att utifrån en given frågeställning och givna data finna lämpliga modeller. Redovisningen sker genom en skriftlig rapport. Deluppgifter som ska lösas individuellt ingår för att kunna bedöma den individuella prestationen. Betygskriterierna är följande:

G: Tydlig frågeformulering där data beskrivs deskriptivt med passande diagram och grafer. En tydlig redogörelse för val av ursprunglig modell samt slutgiltig modell ska ingå. Det ska tydligt framgå vilka kriterier som används i valet av slutgiltig modell. Även en modellkontroll ska ingå, där det tydligt framgår vilka antagandena är. Svar på teoretiska frågor ska vara väl motiverade. En beskrivning av hur arbetet har utförts och arbetsfördelningen inom arbetsgruppen ska finnas. De individuella uppgifterna ska vara utförda. Den skriftliga redovisningen har lämnats in i tid.

U: Något av följande: Alla uppgifter är inte lösta. Delar av kraven för betyg G är ej uppfyllda. Beskrivning av arbetsfördelningen saknas. Redovisningen har inte lämnats in i tid.

Prov 4: Undersökningsmetodik, inlämningsuppgift, 1.5 högskolepoäng

Provet utförs i form av ett grupparbete med 3-4 studenter per grupp och går ut på att utifrån en given frågeställning planera en statistisk undersökning. Redovisningen sker genom en skriftlig rapport samt en muntlig presentation. Deluppgifter som ska lösas individuellt ingår för att kunna bedöma den individuella prestationen. Betygskriterierna är följande:

G: Syftet med undersökningen och en adekvat undersökningsplan ska beskrivas. Operationaliseringen av frågeställningen ska förklaras och motiveras. En tänkt målpopulationen, en lämplig ram för urvalsdragningen samt insamlingsmetod ska vara definierade och motiverade. Ett frågeformulär med relevanta och passande frågor har konstruerats med utgångspunkt från frågeställningen och insamlingsmetod. En relevant diskussion om undersökningens design med metodkritik ska finnas. En beskrivning av hur arbetet har utförts och arbetsfördelningen inom arbetsgruppen ska finnas. De individuella uppgifterna ska vara utförda. Den skriftliga redovisningen har lämnats in i tid samt presenterats muntligen.

U: Något av följande: Alla uppgifter är inte lösta. Delar av kraven för betyg G är ej uppfyllda. Beskrivning av arbetsfördelningen saknas. Redovisningen har inte lämnats in i tid. Ej genomfört den muntliga redovisningen.

Salstentorna

Villkoren för proven, betygsättning mm. beskrivs i avsnitten *Kunskapskontroll och betygsättning* och *Examinationstillfällen* ovan.

Prov 1: Regressionsanalys och tidsserieanalys, tentamen, 7.5 högskolepoäng

Provet består av en individuell skriftlig tentamen. Följande betygskriterier gäller (totalpoäng = 100):

A: Utmärkt. Studenten skall på ett kritiskt och välstrukturerat sätt kunna tillämpa regressions- och tidsserieanalys med tillhörande statistisk inferens och modellutvärdering som inte nödvändigtvis direkt behandlas i kursmaterialet. Studenten skall vidare klart och tydligt kunna redogöra för begrepp, metoder och teori som används vid genomförande av regressions- och tidsserieanalys. Kräver minst 90 poäng på den skriftliga tentamen.

B: Mycket bra. Studenten skall på ett kritiskt och välstrukturerat sätt kunna tillämpa regressions- och tidsserieanalys med tillhörande statistisk inferens och modellutvärdering som direkt behandlas i kursmaterialet. Studenten skall vidare klart och tydligt kunna redogöra för begrepp, metoder och teori som används vid genomförande av regressions- och tidsserieanalys. Ges för 80 - 89 poäng på den skriftliga tentamen.

C: Bra. Studenten skall på ett välstrukturerat sätt kunna tillämpa regressions- och tidsserieanalys med tillhörande statistisk inferens och modellutvärdering som direkt behandlas i kursmaterialet. Studenten skall vidare på ett bra sätt kunna redogöra för begrepp, metoder och teori som används vid genomförande av regressions- och tidsserieanalys. Ges för 70 – 79 poäng på den skriftliga tentamen.

D: Tillfredsställande. Studenten skall på ett korrekt sätt kunna tillämpa regressions- och tidsserieanalys med tillhörande statistisk inferens och modellutvärdering som direkt behandlas i kursmaterialet. Studenten skall vidare på ett nöjaktigt sätt kunna redogöra för begrepp, metoder och teori som används vid genomförande av regressions- och tidsserieanalys. Ges för 60 – 69 poäng på den skriftliga tentamen.

E: Tillräcklig. Studenten skall på ett huvudsakligen korrekt sätt kunna tillämpa regressions- och tidsserieanalys som direkt behandlas i kursmaterialet. Studenten skall vidare på ett huvudsakligen korrekt sätt kunna redogöra för begrepp, metoder och teori som används vid genomförande av regressions- och tidsserieanalys. Ges för 50 – 59 poäng på den skriftliga tentamen.

Fx: Otillräcklig. Studenten uppfyller endast delvis kraven för betyg E. Ges för 40-49 poäng på den skriftliga tentamen. OBS! Omtentamen krävs.

F: Helt otillräcklig. Studenten uppfyller endast i mindre omfattning eller inte alls kraven för betyg E. Ges för 0 – 39 poäng på tentamen. OBS! Omtentamen krävs.

Prov 3: Undersökningsmetodik, tentamen, 3.5 högskolepoäng

Provet består av en individuell skriftlig tentamen. Följande betygskriterier gäller (totalpoäng = 100):

A: Utmärkt. Studenten skall på ett kritiskt och välstrukturerat sätt lösa problem inom undersökningsmetodik som ligger nära men som inte direkt behandlats i kursmaterialet. Studenten skall kunna jämföra och välja lämpliga ansatser för att lösa olika problem och på ett klart och tydligt sätt argumentera för dessa val. Studenten skall vidare på ett klart och uttömmande sätt kunna redogöra för begrepp, metoder och teori som används vid genomförandet av statistiska urvalsundersökningar. Kräver minst 90 poäng på den skriftliga tentamen.

B: Mycket bra. Studenten skall på ett kritiskt och välstrukturerat sätt kunna lösa problem inom undersökningsmetodik som behandlats direkt i kursmaterialet. Studenten skall kunna jämföra och välja lämpliga ansatser för att lösa olika problem och argumentera för dessa val. Studenten skall vidare i stort sett uttömmande kunna redogöra för och förklara begrepp, metoder och teori som används vid genomförandet av statistiska urvalsundersökningar. Ges för 80 - 89 poäng på den skriftliga tentamen.

C: Bra. Studenten skall på ett korrekt sätt kunna lösa problem inom undersökningsmetodik som direkt behandlats i kursmaterialet. Studenten skall kunna jämföra och välja lämpliga ansatser för att lösa olika problem. Studenten skall vidare kunna redogöra för begrepp, metoder och teori som används vid genomförandet av statistiska urvalsundersökningar. Ges för 70 – 79 poäng på den skriftliga tentamen.

D: Tillfredsställande. Studenten skall på ett i huvudsak korrekt sätt kunna lösa problem inom undersökningsmetodik som direkt behandlas i kursmaterialet. Studenten skall kunna jämföra och välja lämpliga ansatser för att lösa olika problem och förklara dessa val. Studenten skall vidare nöjaktigt kunna redogöra för begrepp, metoder och teori som används vid genomförandet av statistiska urvalsundersökningar. Ges för 60 – 69 poäng på den skriftliga tentamen.

E: Tillräcklig. Studenten skall på ett till största delen tillräckligt sätt kunna lösa elementära problem inom undersökningsmetodik som direkt behandlats i kursmaterialet. Studenten skall kunna välja lämpliga ansatser för att lösa olika problem och förklara dessa val. Studenten skall vidare nöjaktigt kunna redogöra för begrepp, metoder och teori som används vid genomförandet av statistiska urvalsundersökningar. Ges för 50 – 59 poäng på den skriftliga tentamen.

Fx: Otillräcklig. Studenten uppfyller endast delvis kraven för betyg E. Ges för 40-49 poäng på den skriftliga tentamen. OBS! Omtentamen krävs.

F: Helt otillräcklig. Studenten uppfyller endast i mindre omfattning eller inte alls kraven för betyg E. Ges för 0 – 39 poäng på tentamen. OBS! Omtentamen krävs.

TILLÅTNA HJÄLPMEDEL OCH FUSK VID EXAMINATION

De två inlämningsuppgifterna utförs huvudsakligen i grupp där samarbete inom gruppen självklart är tillåtet. Notera dock att betygsättningen är individuell och kan variera inom en arbetsgrupp då det är den enskildes prestation inom arbetsgruppen som bedöms. Samarbete *mellan* grupper är också tillåtet men varje grupp ska lämna in sin unika rapport. All typ av plagiering är självklart otillåtet och programvara för textmatchning kan vid behov komma att användas.

Salstentor är individuella och alla former av samarbete är otillåtet liksom andra hjälpmedel än de som medgetts av examinator. Miniräknare utan lagrad text och data är tillåtet; mobiltelefoner med räknarfunktion är inte ett tillåtet hjälpmedel. Formel- och tabellblad för kursdel 1 Regression- och tidsserieanalys delas ut vid tentamen.

Användning av AI-Verktyg

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.

Se även information om fusk och plagiering på universitetets webbsida (länk) samt Regler och handläggningsordning för disciplinärenden. Notera att institutionen alltid måste anmäla misstanke om fusk och plagiering till SU:s disciplinnämnd.