

# TNIU66: Statistik och sannolikhetslära

## Kursinformation 2020

Kunskaper i statistik och sannolikhetslära är väsentliga för nästan alla människor i ett modernt samhälle. Vi översköljs av påståenden baserade på statistik av mer eller mindre vederhäftig karaktär, och behöver förmågan att kunna skilja det värdefulla från det rena tramset. Som universitetsutbildad inom ett kvantitativt ämne (vare sig det är logistik i någon form eller som ingenjör) har du ett särskilt ansvar att se till att den statistik som används främjar ett gott samhälle och klargör samband mer än förvirrar. Den här kursen bär inte hela den vägen, men den är en god start för ett kreativt användande av starka matematiska verktyg i resten av din utbildning, och förhoppningsvis även i livet efter studierna.

## Kursens mål, innehåll och förväntade läranderesultat

Kursens mål är att ge en introduktion till matematisk modellering av slumpmässiga försök och till statistiska metoder och begrepp. Den ska även visa på tillämpningar relevanta för de program för vilken den ges. Innehållet och förväntade läranderesultat är tagna ur kursplanen, och återges här för fullständighetens skull. Före kursens början är dessa svåra att läsa, varför det rekommenderas att du återvänder till dem under kursens gång.

**Sannolikhetslära:** Sannolikhetsbegreppen. Utfallsrum och händelse. Mängdlära och kombinatorik. Betingade sannolikheter, med Bayes sats, och begreppet oberoende händelser. Diskreta och kontinuerliga slumpvariabler med fördelningsmått såsom väntevärde, varians/standardavvikelse, kovarians och korrelation. De vanligaste sannolikhetsmodellerna, inkluderande likformig-, binomial-, poisson-, exponential- och normalfördelningsmodellen. Centrala gränsvärdessatsen.

**Statistik:** Deskriptiv statistik med begreppen medelvärde, median och standardavvikelse samt visualiseringar. Punkt- och intervallskattningar av väntevärde (med och utan känd standardavvikelse), sannolikhet och intensitet. Punktskattning av varians och standardavvikelse. Hypotesprövning, inklusive  $p$ -värde och styrkefunktion. Korrelation. Linjär regression för två variabler med kontroll av förutsättningar samt konfidens- och prediktionsintervall.

Efter genomförd kurs ska du kunna:

1. analysera fördelningen hos en datamängd avseende centralvärde och spridning, såsom medelvärde och median samt standardavvikelse, samt visualisera detta.
2. redogöra för olika synsätt på begreppet sannolikhet.
3. beräkna sannolikheter för händelser, med användning av begrepp och verktyg som oberoende, betingning, oförenlighet, komplementhändelse, union, snitt, kombinatorik, trädidiagram.
4. formulera en sannolikhetsmodell med hjälp av slumpvariabler, även med centrala gränsvärdessatsen, och använda den för att bestämma egenskaper hos dess fördelning samt beräkna sannolikheter.
5. beräkna punktskattningar av väntevärde, varians, standardavvikelse, sannolikhet och intensitet samt bedöma deras lämplighet.

- beräkna konfidensintervall för väntevärde (med och utan känd standardavvikelse), sannolikhet och intensitet samt tolka resultatet.
- formulera och genomföra en hypotesprövning, och däri kunna tolka begreppen styrkefunktion och  $p$ -värde.
- genomföra en korrelationsanalys och tolka resultatet.
- ställa upp och tolka en linjär regressionsmodell med två variabler, avgöra om en linjär modell är tillämpbar, samt bedöma tillförlitligheten hos skattningar av såväl väntevärden som enskilda observationer.
- använda datorstöd för alla beräkningar där det är relevant.

## Förkunskaper

En högskolekurs inom matematik, exempelvis matematisk grundkurs för BI eller matematik för SL och FTL. Gymnasiematematik, kurserna Ma1 – Ma3 (eller motsvarande). I den mån repetition är nödvändig är det *ditt* ansvar att se till detta (givetvis kan du fortfarande fråga om detaljer).

## Litteratur, webbstöd, programvara och miniräknare

Kurslitteratur är boken Tillämpad statistik – en grundkurs, andra upplagan, av Karl Wahlin, utgiven av Sanoma utbildning, Stockholm, 2015. Det finns en läsanvisning på Lisam, då inte hela boken ingår i kursen. Boken behandlar heller inte riktigt allt som kursen tar upp, och därför finns även kompletterande materiel på Lisam.

På webben finns mycket stöd för den som önskar ta del av sådant. Dessvärre finns även sådant som är av mer tvivelaktig karaktär (läs exempelvis ”statistik” på svenska wikipedia), varför det krävs viss baskunskap för att kunna bedöma vad som är värt att studera. Likaså kan beteckningar och konventioner skilja sig mellan det som hittas på webben och de som vi använder i den här kursen. En god källa som tar upp det mesta i kursen på ett vederhäftigt sätt är Khan Academy.

Kursens webbsida ligger under Lisam. **Du får full tillgång till kurssidans först när du registrerar dig på kursen.** Där kan du hitta det som hör till kursen, bl.a. denna kursinformation, alla bilder som visas på föreläsningarna, exempelsamling, laborationsuppgifter samt schema. Inlämning av laborationsuppgifter sker enbart via Lisam.

Den programvara som används vid laborationerna är Excel, orsaken är att du ska ha lätt att göra statistiska analyser även efter avslutad kurs. Excel är ett program som du med stor sannolikhet kommer att ha tillgång till även efter avslutad utbildning. På Lisam finns länkar till korta videotutorials som introduktion till Excel för den student som inte kommit i kontakt med programmet tidigare.

I kursen behövs en miniräknare. En grafritande sådan är att föredra, då den oftast innehåller även de vanligaste fördelningsfunktionerna (framför allt normalfördelningen och  $t$ -fördelningen), samt stöd för vissa statistiska beräkningar. Du *kan* klara dig med en mycket enkel räknare med i princip enbart de fyra räknesätten, men måste då vara beredd på en hel del skriv- och bokföringsarbete. *Du förväntas själv ta reda på hur just din miniräknare fungerar och vilka funktioner den har.* Den miniräknare som används vid tentamen får inte ge möjlighet till annan kontakt med yttervärlden än via tangentbord och skärm. Exempelvis får den inte ha möjlighet till wifi-uppkoppling.

## Lärare och kontaktytor

Lärare och examinator för kursen är Michael Hörnquist, som träffas säkrast i samband med undervisningen. Enklare frågor av administrativ eller allmän karaktär kan även ställas via e-post till michael.hornquist@liu.se.

## Organisation av undervisningen

Kursen är uppdelad i fyra moduler, A – D, var och en syftande till några av lärandemålen. Avsikten är att det ska bli enklare för dig att se vad du ska kunna efteråt. I slutet av den här kursinformationen finns en undervisningsplan, där det anges vad som tas upp på respektive föreläsning och lektion. Inramande uppgifter är bonusgrundande förberedelseuppgifter, se mer i nästa avsnitt under ”Lektioner”.

## Om undervisningen, inklusive examinationen

**Föreläsningar:** Föreläsningarna genomförs på konventionellt sätt. De *introducerar* begrepp och ideer, men *förutsätter* att du själv studerar de relevanta avsnitten i kurslitteraturen. Alla bilder som visas på en föreläsning kan laddas ned från Lisam senast vid lunch dagen före. Till varje föreläsning finns ett antal övningsuppgifter markerade som ”Efter”. Dessa är lämpliga att göra snarast efter föreläsningen för att börja träna på att använda de begrepp föreläsningen har tagit upp.

**Eget arbete:** De schemalagda tomma passen efter föreläsningarna, märkta ”Eget arbete” är avsedda att ge tid både till efterbearbetning, exempelvis till angivna efteruppgifter och till att förbereda kommande lektion. De är inlagda så ofta övriga schemat har tillåtit det, men får inte tolkas som att de erbjuder tillräcklig självstudietid. För 6 hp är rimlig arbetsinsats motsvarande fyra veckors heltidsarbete, det vill säga 160 timmar.

**Laborationer, LAB1:** Det finns tre laborationer i kursen, vilka genomförs i grupper om tre studenter (undantagsvis två). Grupperna formerar ni själva och det går bra att variera grupper mellan laborationerna. Det är obligatorisk närvaro vid laborationspassen, och om du av någon orsak inte kan närvara vid ditt pass uppmanas du i första hand att bevista den andra gruppen.<sup>1</sup> Till varje laboration finns på Lisam de uppgifter du ska göra. **Några av dessa är förberedelseuppgifter, dessa ska vara gjorda innan du påbörjar laborationen.**<sup>2</sup> Redovisning av laborationen sker genom att gruppen lämnar in lösningar på laborationsuppgifterna (anvisningar ges i samband med laborationsinstruktionerna) samt att ni innan ni lämnar datorsalen tillsammans i gruppen kortfattat berättar vad ni har gjort och vad som eventuellt återstår innan ni är klara. **Det är viktigt att du också skriver upp ditt namn på den lista som tillhandahålls.** Alla laborationsuppgifter lämnas in via Lisam. Om du av någon orsak inte skulle lämna in i tid, eller inte får redovisningen godkänd, erbjuds ytterligare fyra (4) möjligheter till inlämning under året (motsvarande omtentamina).<sup>3</sup> Det är kl. 07.00 den 30 mars, 14 april, 12 juni, samt 31 augusti, 2020, som är sista tidpunkt. Notera att inlämnade rapporter *inte* rättas kontinuerligt, utan först efter respektive sista inlämningsdag. Den student som inte är godkänd på alla tre laborationer efter sista inlämningstillfället, 31 augusti 2020, behöver normalt göra om hela momentet LAB1 nästa år. Samarbete mellan grupperna är tillåtet, men inte att skriva av

<sup>1</sup>Det finns även två uppsamlingstillfällen, se schemat, vilka kan användas av den som har missat ordinarie tillfälle, exempelvis på grund av sjukdom.

<sup>2</sup>Detta kommer att kontrolleras, och den som inte kan visa upp egna lösningar, eller försök till lösningar, kommer att avvisas från salen.

<sup>3</sup>Fem (5) möjligheter för första laborationen, se undervisningsplanen.

varandra. Varje gruppledmedlem ansvarar fullt ut för den lösning gruppen lämnar in, och läraren kan begära muntlig förklaring av vad som har gjorts för att säkerställa enskilda students insats.<sup>4</sup>

**Lektioner:** Till varje lektion finns ett antal uppgifter föreskrivna. Dessa kommer att behandlas vid tavlan för att belysa den teori som föreläsningarna presenterat. Här förväntas du som student ha förberett dig för att kunna ställa frågor kring det som är oklart vid genomgången. Vid sex av lektionpassen finns på undervisningsplanen en inramad uppgift, dessa är studentredovisningsuppgifter som kan ge bonuspoäng till tentamen (syns i kursplanen som UPG1). Till dessa lektioner ska du, om du önskar bonuspoängen, ha med dig en handskriven lösning till den inramade uppgiften, att lämna in. Med hjälp av tillämpad sannolikhetslära väljs bland de inlämnade lösningarna vem som ska få redovisa vid tavlan. Den inlämnade lösningen får givetvis användas som stöd vid tavelredovisningen. Alla som anger att de är beredda att redovisa, samt lämnar in en lösning, får *en* (1) bonuspoäng, även den som inte redovisar vid tavlan. Bonusen gäller till och med omtentamen i augusti 2020. Totalt kan alltså upp till sex (6) bonuspoäng erhållas. Lösningen som redovisas behöver inte vara perfekt, men den ska vara ett ärligt försök, och du som redovisar måste visa att du har förstått uppgiften. Om det är uppenbart att du enbart har skrivit av någon annan, utan att själv sätta dig in i uppgiften, tappas samtliga bonuspoäng som du hitintills har samlat ihop under kursen. Uppgifterna angivna som ”efter” är lämpliga som repetition.

**Resurs:** Avsikten med resurstillfällena är att dina frågor ska stå i centrum. Det finns ingen planering på vad som ska gås genom; läraren finns i angiven sal för att svara på frågor. Om flera studenter frågar om samma moment kan kortare genomgångar förekomma.

**Tentamen, TEN1:** Tentamen består av sex uppgifter à sex poäng. Till det kommer upp till sex bonuspoäng du kan ha med dig från lektionsdelen av kursen. För betyget  $n$  räcker  $6n - 1$  poäng. Uppgifterna på tentamen kopplar direkt till de fyra modulerna (modulen Sannolikhetslära och modulen Statistik kopplar till två uppgifter vardera) och därmed även till lärandemålen. Notera dock att ämnet är kumulativt i sin uppbyggnad, delarna bygger på varandra, varför senare avsnitt ofta bygger på sådant du har tillgodogjort dig i tidigare avsnitt.

Tillåtna hjälpmedel på tentamen är kurslitteraturen, dvs Wahlin,<sup>5</sup> samt (nästan) valfri miniräknare med tömda minnen (miniräknaren får inte ge möjlighet till kontakt med omvärlden annat än via tangentbord och skärm, exempelvis får den inte ha wifi). Det får finnas anteckningar i boken, sådana man ofta gör i en lärobok som studeras.<sup>6</sup> Likaså är det ok med små (ett par centimeter) påklitrade sidmarkeringar, ”pagemarkeringar”. Tidigare års (2018 och 2019) tentamina finns på Studieinfo för kursen det året, se under ”Övriga dokument”. Nytt för i år är att svaren som lämnas in ska anges på en bifogad svarsblankett, och poängsättningen kommer att utgå från vad som står på svarsblanketten. För att säkerställa den egna prestationen ska dock den kalkyl som lett fram till givna svar bifogas svarsblanketten.

Välkommen till kursen!  
*Michael Hörnquist*

---

<sup>4</sup>Att ha sitt namn med på en lösning där man inte kan förklara det som står är att försöka vilseleda examinator att prestationen är ens egen, utan att den är det. Det är inte acceptabelt, och sådant beteende kommer att anmälas till Disciplinnämnden som grundad misstanke om fusk.

<sup>5</sup>Som *alternativ* kan valfri annan bok i ämnet användas, dock enbart *en* (1) bok totalt.

<sup>6</sup>Ingen kommer att anmälas till Disciplinnämnden för att det finns för mycket skrivet i boken.

## Undervisningsplan

Uppgifterna är hämtade från exempelsamlingen på Lisam (anges som x.y), ur kurslitteraturen Wahlin (anges som Wx.y) samt från tidigare års tentamina (anges som TeÅR.x.y; x är tentans nummer det året och y är uppgift).

### A Deskriptiv statistik och datorstöd

Denna modul svarar mot lärandemålen 1 och 10, och ger grunderna i deskriptiv (=beskrivande) statistik. Laborationen syftar primärt till att komma igång med Excel. För den som är ovan vid Excel rekommenderas de video-tutorials som är länkade till på Lisam.

Fö A. Wahlin: kap. 2.1 – 4. Efter: W2.1abcdef, W2.3

Le A. 1.1, 1.2, 2.1, 2.4, 2.5, 2.6. Efter: 1.3, 1.4, 2.2, 2.3, 2.8, Te19.1.1

Lab A. Finns på Lisam (inlämnas senast fredagen 31 januari 2020 kl. 07.00, alternativt/komplettering måndagen 17 februari 2020 kl. 07.00)

### B Sannolikhetslära

Denna modul svarar mot lärandemålen 2 – 4. Det här är den mest teoretiska delen av kursen, och ger många av de verktyg som behövs för de mer tillämpade statistikdelarna.

Fö B1. Wahlin: kap. 3. Efter: W3.1, W3.3, W3.7, W3.12

Le B1. 3.1a, 3.1b, 3.2, 3.4, 3.5. Efter: Te19.1.2, W3.14

Fö B2. Wahlin: kap. 4.1 – 2 (ej 4.2.2, 4.2.4). Efter: W4.1, W4.3, W4.8

Le B2. 4.1, 4.2, 4.3, 4.4. Efter: W4.4, W4.5, Te19.2.2

Fö B3. Wahlin: kap. 4.3 (ej 4.3.4), 5 (ffa medelvärden), kompletterande materiel.

Efter: W4.11, W4.12, W4.13, W4.14, W4.15

Le B3. 4.5abce, 4.8, 4.9ace, 4.11, 5.1. Efter: 4.5df, 4.6, 4.7, 4.9bd, 4.10, W4.17, Te19.1.3

Extra. Gästföreläsning av Valentin Polishchuk, främst för FTL men alla är välkomna

### C Statistik

Denna modul svarar mot lärandemålen 5 – 7 och 10. Normalfördelningen och Centrala gränsvärdessatsen är väsentliga *förkunskaper* för den här modulen.

Fö C1. Wahlin: kap. 5, 6 (konfidensintervall). Efter: W6.1, W6.2

Le C1. 6.1, 6.2, 6.3, 6.4a, 6.4bc. Efter: W6.6, W6.7, W6.9a, Te19.1.4

Lab C. Finns på Lisam (inlämnas senast lördagen 29 februari 2020 kl. 07.00)

Fö C2. Wahlin: kap. 6 (hypotesprövning och  $p$ -värden). Efter: W6.3, W6.4, W6.14

Le C2. 6.5, 6.6, 6.7, 6.8a, 6.8b. Efter: W6.9b, W6.12, Te19.2.5

### D Sambandsanalys

Denna modul svarar mot lärandemålen 8 – 10. Största delen tas upp av så kallad enkel linjär regression, men även korrelation ingår. Endast samband mellan kvantitativa variabler ingår.

Fö D. Wahlin: kap. 10.1 – 4. Efter: W10.2abcd

Le D. 7.1, 7.2, 7.4, 7.6. Efter: 7.3, W10.2fghi, 7.5

Lab D. Finns på Lisam (inlämnas senast fredagen 13 mars 2020 kl. 07.00)