



## INSTITUTIONEN FÖR DATA- OCH INFORMATIONSTEKNIK

### **DIT862 Statistiska metoder för Data Science, 7,5 högskolepoäng**

Statistical Methods for Data Science, 7.5 credits

*Avancerad nivå / Second Cycle*

---

#### **Fastställande**

Kursplanen är fastställd av Institutionen för data- och informationsteknik 2018-02-02 och senast reviderad 2018-09-27. Den reviderade kursplanen gäller från och med 2018-11-05, höstterminen 2018.

*Utbildningsområde:* Naturvetenskapligt 100 %

*Ansvarig institution:* Institutionen för data- och informationsteknik

#### **Inplacering**

Kursen är obligatorisk inom Applied Data Science masterprogram. Den ges även som fristående kurs vid Göteborgs universitet.

Kursen kan ingå i följande program: 1) Applied Data Science masterprogram (N2ADS)

#### *Huvudområde*

Data Science

#### *Fördjupning*

A1F, Avancerad nivå, har kurs/er på avancerad nivå som förkunskapskrav

#### **Förkunskapskrav**

För tillträde till kursen krävs att studenten har en examen på kandidatnivå inom något ämne, eller har minst 90 hp i datavetenskap, software engineering eller motsvarande. Specifikt krävs följande:

- minst 15 hp avklarade kurser i programmering,
- en av kurserna DIT851 Introduktion till Data Science, 7,5 hp, eller DIT856 Tillämpat matematisk tänkande, masterkurs, 7,5 hp, alternativt minst 7,5 hp i matematik.

Följande kunskapsnivå i Engelska krävs; Engelska 6/Engelska B eller motsvarande från ett erkänt internationellt test, t.ex. TOEFL, IELTS.

## Lärandemål

Efter godkänd kurs ska studenten kunna:

### *Kunskap och förståelse*

- redogöra för grundläggande begrepp inom sannolikhets- och statistikteori, med fokus på tillämpningar och experiment inom data science
- ge exempel på hur probabilistiska modeller används för olika tillämpningar inom data science

### *Färdigheter och förmåga*

- med hjälp av statistiska datorverktyg utföra dataanalyser såsom deskriptiv statistik och datavisualisering
- implementera ett antal probabilistiska modeller och tillämpa dem i användningsområden inom data science
- tillämpa statistiska testmetoder för att utvärdera analysystem inom data science

### *Värderingsförmåga och förhållningssätt*

- motivera vilken typ av statistisk metod som är lämplig för de vanligaste typerna av experiment inom data science-tillämpningar
- diskutera fördelar och nackdelar med olika typer av probabilistiska modeller som kan vara lämpliga för en given tillämpning inom data science
- diskutera hur felaktigt använda statistiska tekniker kan leda till tveksamma slutsatser

## Innehåll

Kursen ger en introduktion till sannolikhets- och statistikteori, dataanalys med hjälp av deskriptiv statistik, datavisualisering, samt tillämpningar av probabilistiska modeller inom data science-området.

Inom kursen täcks följande områden:

- dataanalys inklusive deskriptiv statistik och datavisualisering
- sannolikhetssteori inklusive grundläggande sannolikhetskalkyl, stokastiska variabler, fördelningar
- statistiska metoder inklusive punkt- och intervallskattning, hypotesprövning, regression
- probabilistiska modeller i tillämpningar inom data science-området, exempelvis Naive Bayes-klassificering och topic modeling för text, eller Hidden Markov Models för sekvenser

### *Delkurser*

1. **Hemtentamen** (*Take-home examination*), 4 hp  
Betygsskala: Väl godkänd (VG), Godkänd (G) och Underkänd (U)
2. **Inlämningsuppgifter** (*Assignments*), 3,5 hp  
Betygsskala: Väl godkänd (VG), Godkänd (G) och Underkänd (U)

### **Former för undervisning**

Föreläsningar, övningar, datorlaborationer.

*Undervisningsspråk:* engelska

### **Former för bedömning**

Kursen examineras genom en individuell skriftlig hemtentamen, samt obligatoriska skriftliga inlämningsuppgifter som redovisas genom rapportinlämning, varav vissa kommer att genomföras individuellt och andra i grupper om normalt 2-4 studenter.

Det kommer att ges frivilliga individuella uppgifter som ger bonuspoäng på tentamen. Dessa bonuspoäng är giltiga under ett helt läsår.

Försenad inlämning av hemtentamen innebär att lösningen får betyget Underkänd (U), om inte särskilda skäl föreligger. Underkänd hemtentamen omexamineras genom en ny hemtentamen.

Om student som underkänts två gånger på samma examinerande moment önskar byte av examinator inför nästa examinationstillfälle, ska sådan begäran inlämnas skriftligt till kursansvarig institution och bifallas om det inte finns särskilda skäl däremot (HF 6 kap § 22).

I det fall en kurs har upphört eller genomgått större förändringar ska studenten i normalfallet garanteras tillgång till minst tre provtillfällen (inklusive ordinarie provtillfälle) under en tid av åtminstone ett år med utgångspunkt i kursens tidigare uppläggning.

### **Betyg**

På kursen ges något av betygen Väl godkänd (VG), Godkänd (G) och Underkänd (U). För godkänt betyg på hel kurs krävs godkänt betyg på samtliga delkurser.

För att få betyget Väl Godkänd (VG) på hela kursen krävs betyget VG på båda delkurserna.

**Kursvärdering**

Kursen utvärderas genom möten både under och efter kursen mellan lärare och studentrepresentanter. Därutöver används en anonym enkät för att få skriftlig information. Resultatet av utvärderingen används för att förbättra kursen genom att visa på delar som kan läggas till, förbättras, ändras eller tas bort.

**Övrigt**

Kurslitteratur kommer att publiceras senast 8 veckor innan kursstart.

Kursen ersätter kursen DIT861, 7,5 hp. Den här kursen kan inte ingå i en examen som innehåller DIT861. Den kan inte heller ingå i en examen som bygger på en annan examen där DIT861 ingår.