

# Inlämningsuppgift, Statistisk översiktscurs, VT25

## Introduktion, inlämning, format

I denna inlämningsuppgift ska ni självständigt i grupper analysera ett datamaterial i programmeringsspråket R. Till skillnad från datorlaborationerna finns det få kodexempel.

Datorlaborationerna går igenom de flesta momenten som behandlas i inlämningsuppgiften. Inlämningsuppgiften har fyra uppgifter (1, 2, 3, 4), med deluppgifter därunder. De fyra uppgifterna tar upp frågor som liknar det som görs igenom på laborationerna 1-4. När du är klar med laboration 1 ska du kunna göra uppgifterna under 1 i inlämningsuppgiften, osv.

Efter att ni lämnat in uppgiften kan ni antingen få godkänt eller komplettering (i Athenasystemet registreras kompletteringsbehov som underkänt). Därefter finns en möjlighet att komplettera, se datum nedan. Efter inlämningsdatum 2 finns ingen ytterligare chans att komplettera, inlämningsuppgiften behöver då göras om vid ett framtida kurstillfälle. Kursbeskrivningen innehåller ytterligare information om bedömningsgrunder, etc.

**1. Inlämning, senast**                      **Fredag 25 april kl. 16.59. Återlämnas 2 maj.**

**2. Komplettering, senast**            **Fredag 9 maj kl. 16.59.**

När ni lämnar in är det viktigt att svara på alla frågor och delfrågor. Om svar inte finns på vissa frågor vid första inlämningen kan ni inte få feedback på den uppgiften. Vi rekommenderar att ni går igenom hela uppgiften och svarar på alla frågor redan vid inlämning 1.

Inlämningen ska innehålla svar i form av siffrsvar, grafer, etc., beroende på ställd fråga. Varje uppgift ska också innehålla ett textsvar. Grafer ska ha rubrik, namn på axlarna, etc.

Ni kan arbeta i exv. Google Docs, Microsoft Word, eller annat dokument. Koden ni använder ska också finnas med i inlämningen, ni kan antingen klippa in den sist i ert dokument (börja på en ny sida och markera tydligt, i koden, exv med radbrytningar och förklarande rubriker, var varje deluppgift finns i koden). Om ni tycker det är bättre kan ni istället klistra in koden vid respektive uppgift, i huvuddokumentet, eller lämna in en separat kodfil (som ska gå att följa, uppgift för uppgift).

Skriv era namn, gruppnummer och inlämningsdatum längst upp i de dokument ni lämnar in.

På alla moment i kursen, och speciellt på jour och datorlabbar, finns möjlighet att ställa frågor och lärarna finns tillgängliga för att hjälpa till. Kontakta lärarna i god tid vid eventuella problem.

Alla gruppmedlemmar ska vara delaktiga och bidra till alla delar av rapporten och arbetet som leder upp till rapporten, dvs skriva kod, analysera data, tolka resultat, dra slutsatser och skriva rapporten.

1. I labb 1 jobbade vi med bland annat med data på kapitalavkastning över tid.

1.1 Definiera en vektor med tio värden (0.05, 0.07, 0.054, 0.07, 0.074, 0.087, 0.074, 0.094, 0.091, 0.089), där värdena är andelen av de tillfrågade som säger att de skulle rösta på XYZ-partiet, från opinionsundersökningar i januari för åren 1-10.

1.2 Gör en lämplig graf som visar opinionsutvecklingen för XYZ-partiet, med lämpliga värden etc. på axlarna och en tydlig rubrik.

1.3 Modifiera vektorn ovan på lämpligt sätt och gör en graf som visar opinionsutvecklingen, i **procent**, för XYZ-partiet, med lämpliga värden på axlarna och en tydlig rubrik.

1.4 I uppgift 1.1 ovan var data från åren 2010-2019. Istället för att skapa en vektor, använd följande kommando för att skapa en data frame med två kolumner:

```
data.frame(Year=c(2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019), Vote_intention = c(0.05, 0.07, 0.054, 0.07, 0.074, 0.087, 0.074, 0.094, 0.091, 0.089))
```

För att inte bara skriva ut ovanstående på skärmen, välj ett lämpligt namn på objektet (er data frame) och lagra ovanstående i objektet.

1.5 Skapa en graf som i 1.2 men använd istället kolumnerna i er data frame för att göra grafen.

1.6 Lägg till en tredje kolumn med riksdagsspärren i er data frame (vi antar här att er data frame heter AAA), med följande kommando:

```
AAA$threshold <- rep(0.04, 10)
```

(Kommandot repeterar värdet 0.04 tio gånger och lägger in som en ny variabel i AAA)

Titta på de första raderna i er data frame för att se att den nya kolumnen lagts till

```
head(AAA)
```

1.7 Plotta riksdagsspärren i samma graf som ni gjort i 1.5. Direkt efter ert plottkommando i uppgift 1.5, kör följande kommando (byt till ert namn, istället för AAA):

```
lines(AAA$threshold~AAA$Year)
```

lägg också till någon färg i kommandot så att linjen skiljer sig från opinionsdata

Uppdatera rubrik etc, i grafen, på lämpligt sätt.

1.8 Använd ett lämpligt R-kommando för att ta fram femte värdet i första kolumnen av er data frame.

**2.** I laboration 2 jobbade vi med bland annat med deskriptiv statistik, tabeller och grafer, fördelningar, och central- och spridningsmått.

**2.1** Ta fram medelvärde, median, varians och standardavvikelse, med ett eller flera R-kommandon, för den vektor ni skapade i uppgift 1.1 ovan

**Läs in datasetet titanic. Svara på följande frågor med hjälp av lämpliga R-kommandon:**

**2.2** Hur många män respektive kvinnor reste med Titanic?

**2.3** Är alla passagerare (inkl. besättning) registrerade som antingen man eller kvinna? (Tips – finns saknade värden?)

**2.4** Ta fram en lämplig tabell som visar överlevnadsfrekvensen (andelen överlevande) för män respektive kvinnor.

**2.5** Ta fram en lämplig figur som visar överlevnadsfrekvensen (andelen överlevande) för män respektive kvinnor.

**2.6** Ta fram en tabell som visar antal överlevande och antal icke-överlevande, uppdelat på kön. Vilken grupp är störst: antalet män som överlevde eller antalet kvinnor som överlevde?

**2.7** Gör ett histogram över variabeln "Paid" och beskriv hur fördelningen ser ut (modalitet, ev. skevhet, jämför median vs. medel). Testa att variera hur många staplar histogrammet har med argumentet breaks (labb2).