

Statistisk dataanalyse med Jamovi

Daniel Hammarström

2024-01-02

Table of contents

Forord	3
1 Deskriptiv dataanalyse	4
1.1 Data i regnearkprogrammer	4
1.2 Dataanalyse i praksis	5
2 Summary	7
Referanser	11

Forord

Målet med denne boken er at du skal bli kjent med verktøy og metoder for statistisk dataanalyse, hvordan man gjennomfører slike, og rapporterer resultater fra dem. I denne boken vil vi introdusere Jamovi, en gratis programvare for statistisk dataanalyse med tilstrekkelig kraft fleksibilitet for analyser av forskningsdata. I tillegg vil vi snakke om bruk av regnearkprogrammer så som Microsoft Excel eller Google Sheets og andre programvarer som du vil ha nytte av når du forfatter rapporter som er baserte på data. Målet med boken er å gi en introduksjon til de verktøy du trenger for å forfatte en bachelor eller masteroppgave.

Hva trenger jeg? Denne boken bygger videre på Statistisk Dataanalyse forfattet av Christer Thrane. Vi rekommanderer denne som en introduserende tekst.

For å følge med i denne boken trenger du Jamovi, som du kan installere til din Mac eller PC ved å laste ned programvaren [her](#). Det er også mulig å bruke [Jamovi i skyen](#).

Det finnes flere alternativer til regnearkprogrammer. Det mest kjente er Microsofts Excel, men du kan også bruke Open Office eller Google sheets. For å forfatte rapporter kan du bruke et ordbehandlingsprogram så som Microsofts Word eller lignende (Google docs, open office, libre office, etc). Til sist bruker vi en presentasjonsprogramvare (Microsoft PowerPoint) for å redigere figurer. I denne boken prøver vi å gå litt mer i dybden på noen konsepter som presenteres i Thranes Statistisk Dataanalyse. For å få mer dybde i pensum rekommanderes Innføring i statistikk og dataanalyse for studenter i Idretts- og helsefag.

[Jamovi her](#).

1 Deskriptiv dataanalyse

I denne modul vil vi introdusere de programmer som vi vil bruke i emnet. Vi vil snakke om hvordan vi kan lagre data og om hvordan vi kan gjennomføre analyser på en ryddig måte.

I pensum snakker man om deskriptiv dataanalyse og det typiske i data. Vi prøver å sette dette inn i en kontekst for vitenskapelig dataanalyse.

Modulen avsluttes med et arbeidskrav hvor du forventes å gi korrekt svar på alle spørsmål. Arbeidskravet kan besvares så mange ganger som kreves for å få alle besvarelser korrekt. Her anbefaler vi en strategi hvor du noterer spørsmål i et dokument som en utgangspunkt for å identifisere spørsmål som du finner vanskelige.

1.1 Data i regnearkprogrammer

Excel er et av flere programmer som du kan bruke for lagre data, lage figurer, gjennomføre analyse og beregninger ved hjelp av data.

Det finnes flere alternative til excel, de fleste har lignende funksjonalitet hvor data i form av tall eller tekst kan mates inn i celler. Celler kan ha funksjoner som beregner eksempelvis tall basert på andre celler. Excel er per i dag, hva vi vet, det program med flest celler! Over 17 milliarder celler. I tillegg har excel over 400 funksjoner som kan brukes for å beregne eller manipulere data. Alt dette gjør excel til en versting hva gjelder å skape trøbbel!

Et kjent problem i excel er konvertering av tekst til datoer. Dette problemet er av betydelse bland annet når man studerer gener hvor noen navn på gener ligner på datoer (eks. SEPT1).

Det finnes måter å bruke excel (og lignende programmer) som sikkerstiller at dette, og andre problemer ikke påvirker analyser og resultater. Den enkleste måten er å være konservativ: Excel (og lignede programmer) fungerer utmerket til å registrere og lagre data. Ønsker du å leve livet mer risikofylt bruker du excel for å lage figurer, gjennomføre beregninger, beregne statistikk eller innhente data fra brukere.

Når vi lagrer data i excel ønsker vi å skape lykkelige datasett. Her finnes en analogi til en kjent roman, Anna Karenina, hvor forfatteren, Leo Tolstoy åpner med å si at **“Alle lykkelige familier er like; hver ulykkelig familie er ulykkelig på sin egen måte.”** Dette kan sies å være sant også for data. Lykkelig data er konsekvent, “tidy” hvilket betyr en rad per

observasjon og en kolonne per variabel. Enkle, men beskrivende navn på variabler på første raden, ingen tomme celler, og ingen beregninger.

Ulykkelig data kan være ulykkelig på så mange forskjellige måter. For eksempel, vi kan ha flere variabler per kolonne, spesialtegn i variabelnavn eller celler, miks av datatyper osv.

Når vi bruker excel (eller lignende programmer) får å mate inn data kan vi bruke datavalidering for å sikre at den data vi før inn er hva vi forventer.

Hvert datasett bør ha en kodebok. En kodebok inneholder en beskrivelse av variablene som er representert i datasettet. I excel kan vi bruke en ny fane for å legge in denne informasjonen. Kodeboken hjelper deg å de du samarbeider med å forstå hvilke data som du har.

Excel ønsker å lagre dine filer i formatet `.xlsx`, dette er ikke alltid hva du ønsker. `xlsx` formatet inneholder formatteringer som ikke er mulige å føre over til ander programmer. Et bedre alternativ er lagre data som `csv`-filer. Dette formatet er åpent og kan leses av flere programvarer, også i fremtiden.

1.2 Dataanalyse i praksis

I dette emnet så foreslår vi at dere bruker excel (eller lignende programmer) for å mate inn og lagre data, og Jamovi for å gjennomføre dataanalyse og lage enkle figurer. Denne kombinasjonen er mer enn nok for å på en god måte klare å ferdigstille et vitenskapelig arbeid hvor du forventes å presentere og tolke data fra et eksperiment eller observasjoner.

Vi foreslår dette fordi kombinasjonen er fritt tilgjengelig for deg i et fremtidig arbeidsliv hvor du ikke vil ha adgang til mer kostbare programvarer, og det finnes muligheter for å gå videre til mer avanserte programvarer (eks. R) med utgangspunkt i Jamovi.

En dataanalyse i praksis er ikke begrenset til hvilke programvarer du bruker. Det å lære å være systematisk og strukturert kommer å spare deg og dine medarbeidere mye hodebry og tid.

En systematisk og organisert dataanalyse er også reproducerbar. Når vi snakker om reproducerbar dataanalyse i denne sammenhengen mener vi at du kan gi din data og analysene til en tredje person som i sin tur kan spore de valg du gjort i dataanalysen.

Reproducerbar dataanalyse er noe som mange snakker om nå da mange mener vi står mitt i en replikasjonskrise. Vitenskapelige funn er ikke alltid mulige å bekrefte i nye studier og det viser seg at andelen funn som ikke kan bekreftes er urovekkende mange! Forskjellige problemer og løsninger er foreslått for lage bedre vitenskap og en stor del av dette går ut på å gjøre vitenskapelig analysearbeid mer transparent.

En **reproducerbar dataanalyse** inneholder all informasjon som kreves for å gjenskape analyseresultatene.

En **transparent dataanalyse** inneholder også beskrivninger av hvorfor og hvordan man valgt å lage analysen på en gitt måte

For å gjøre dette mulig så kreves ytterligere struktur til et prosjekt.

Et enkelt oppsett kan være å tenke på dataanalysen som en isolert mappe på din PC. I mappen finner man alt som kreves for å gjenskape eller forstå din analyse. Her finnes:

- Rådata: Data som er urørt etter det at man matet inn eller innhentet den i forskjellige programmer osv.
- Delvis bearbeidet data: Data som er organisert for data analyse
- Analysefiler: Filer som er kan lese til eks. Jamovi og inneholder analyser av din data
- Rapporter og figurer: Disse er sluttprodukter av deres arbeid, disse kan settes sammen til eks. en bachelor-oppgave.

For å beskrive alle disse delene bør du også ha en fil som beskriver de forskjellige filene i analysen og den overordnede hensikten med hele prosjektet. Denne informasjonen kan beskrives i en fil som vi navngir **README**.

En **README**-fil bør skrives i et format som ikke krever spesielle programvarer for å lese (eks. `.txt`). I presentasjonen finner dere et eksempel på en **README**-fil for et prosjekt *in progress*. **README**-filen er et levende dokument og bør gjenspeile forandringer i prosjektet, en overskrift med **oppdateringer** kan hjelpe å holde styr på fremgangen i prosjektet.

Til sist bør vi vurdere hvordan vi navngir prosjekter. Da disse bør være isolerte (self-contained), og inneholde all informasjon så bør også mappen/prosjektet ha et navn som beskriver innhold. Unngå eks. **Prosjekt1**, **Prosjekt2** osv. Prøv istedenfor å lage beskrivende navn, noe som gir en hint om hva prosjektet ønsker å gjøre eller besvare.

2 Summary

In summary, this book has no content whatsoever.

3

4

5

Referanser