

Bakgrunn

ISS har bidratt til 40-gruppen i datavitenskap i samarbeid med ISV med emnene SGO1910 og SOS2901: <https://www.uio.no/studier/emnegrupper/sv/40-DATAVIT/>

SOS2901 har hatt for få studenter og etter nylig gjennomgang ISS sin emneportefølje er emnet stoppet. ISS er forpliktet til å levere emne til emnegruppen 40-DATAVIT, men må ha emner som tiltrekker seg flere studenter enn tilfellet har vært med SOS2901. Dette forslaget til et nytt emne er bredere og er dermed også mindre personavhengig.

Det foreslås derfor et nytt emne som erstatning: «Deskriptiv og eksplorerende dataanalyse» som er innrettet langt bredere mot praktisk anvendelse for samfunnsvitenskap. All analyse vil i praksis innebære deskriptive analyser og kreve håndtering av ulike datastrukturer før man kan gjøre mer avansert modellering. Dette emnet fokuserer på nettopp dette første leddet med dataanalyse. Emnet styrer dermed helt unna regresjonsanalyser og statistisk inferens.

Emnet vektlegger gjennomgående prinsipper for «Open Science» i kombinasjon med kunstig intelligens der transparens og dokumenterbarhet er sentralt. KI skaper helt nye utfordringer for nettopp dokumentasjon og etterprøvbarhet. Det legges opp til å bruke R og Rstudio gjennomgående, der den innebygde assistenten kan settes opp mot API'et til GPT UiO slik at de lærer assistert koding fra første dag, innenfor UiO sine godkjente systemer.

Det foreslås en muntlig eksamen der deler av arbeidet gjort underveis i emnet presenteres. Det kan være et prosjekt med innlevering underveis som gjennomføres alene eller i gruppe. Ved gruppearbeid blir det likevel levert inn individuelt, og inkluderer en redegjørelse for eget bidrag i arbeidet.

Bruk av kunstig intelligens, dokumentasjon og etterrettelighet er læringsmål, og portfolioen bør derfor dokumentere arbeidet. Det kan bl.a. gjøres ved fortløpende innleveringer via [github.uio.no](https://github.com/uio-no) med sporbar historikk. Selve dokumentasjonen inngår i vurderingen som sådan.

SOS1XXX – Deskriptiv og eksplorerende dataanalyse

Dette emnet gir en innføring i deskriptiv og eksplorerende dataanalyse for samfunnsvitenskapelige formål. Emnet dekker både grunnleggende deskriptiv statistikk, herunder befolkningsmodeller, nettverk og visualisering på kart. Emnet gir også en innføring i teknikker for håndtering av høydimensjonale data gjennom reduksjonsteknikker som klusteranalyse og principal components.

Alt arbeidet gjøres med R med integrert tekst og kode i Quarto. Et gjennomgående tema er dokumenterbarhet og reproduserbar kode. I tillegg vektlegges praktisk bruk av koding assistert med kunstig intelligens gjennom API-tilgang for GPT UiO. Kurset vil gjennomgående bruke de KI-verktøyene som til enhver tid er godkjente for bruk ved UiO.

Hva lærer du?

Etter endt kurs skal studentene kunne følgende:

Kunnskap

- Kjenne sentrale teknikker for deskriptive og visuell fremstillinger av kvantitative data
- Kjenne viktigste prinsipper og grunnlag for open science
- Kjenne til kontekststyring av LLM

Ferdigheter

- Kunne bruke R til å gjøre deskriptive og eksplorerende analyser
- Gjøre analyser med integrert kode og tekst i Quarto / Markdown
- Kunne konfigurere LLM til egnet bruk med kontekststyring
- Kunne gjennomføre reproduserbare analyser

Generell kompetanse

- Gjennomføre selvstendige analyser med datasett i R for samfunnsvitenskapelige formål

Forkunnskaper

Ingen

Undervisning

Forelesninger og praktiske oppgaver.

Studentene må ha tilgang til egen datamaskin med internettilgang med R og Rstudio installert.

Eksamen

Muntlig eksamen basert på en presentasjon av eget arbeid gjennom emnet. Det kan gjennomføres alene eller i gruppe.

Karakterskala

Emnet bruker karakterskala fra A til F, der A er beste karakter og F er stryk. Les mer om [karakterskalaen](#).

Aktuelt pensum og undervisning (foreløpig)

Aktuelt pensum kan være bl.a. utdrag fra følgende bøker. I tillegg blir noen enkelte artikler og egenutviklet undervisningsmateriale. Dette er eksempler på aktuelt pensum, og det vil jobbes videre med dette fremover.

Yu, B. and Barter R. L. (2024) *Veridical Data Science. The Practice of Responsible Data Analysis and Decision Making*. The MIT Press

<https://mitpress.mit.edu/9780262049191/veridical-data-science/>

Christensen, G., Freese J., and Miguel E. (2019) *Transparent and Reproducible Social Science Research. How to Do Open Science*. UC Press.

<https://www.ucpress.edu/books/transparent-and-reproducible-social-science-research/paper>

Wickham et al (2023) *R for Data Science*, O'Reilly Media, <https://r4ds.hadley.nz/>

Omtrentlige temaer per uke:

1. Open science, eksplorerende og konfirmerende analysestrategier
2. Introduksjon til R: oppsett med KI-assistent i Rstudio, kontekststyring
3. Systematisk KI-bruk: system, oppsett og etterrettelighet
4. Reproduserbar arbeidsflyt med Quarto
5. Deskriptive teknikker: tabeller og grafikk
6. Multivariat datareduksjon: klusterteknikker og principal components
7. Befolkningsmodeller
8. Fremskrivninger
9. Grafisk fremstilling på kart
10. Deskriptiv nettverksanalyse