

Pensum SOSGEO2920 Dataanalyse med kunstig intelligens

Emnet vil være rettet mot anvendt analyse av kvantitative data av flere forskjellige typer, men som i dag ikke dekkes i nåværende metodekurs SOSGEO1120. Det vil dessuten legges vekt på rutiner for dokumentasjon, og aktiv bruk av kunstig intelligens.

I tillegg til pensumet vil det utarbeides praktiske oppgaver med datasett der substansielle sosiologiske temaer undersøkes. Pensumet vil suppleres med undervisningsmateriale for kode og teknikker.

Emnet vil presentere grunnleggende analyse av ordinære data (f.eks. surveydata), og i tillegg gi en innføring i analyser som krever litt mer spesielle teknikker: 1) mål på dødelighet, fertilitet og fremskrivinger, 2) geografiske data, 3) sosiale nettverk, og 4) håndtering av «store data».

I pensumlisten er det oppgitt en video fra **Posit PBC**. Denne forklarer oppsett og bruk av den integrerte KI-assistenten for Rstudio. Dette er en ganske ny funksjonalitet og jeg har ikke funnet annet enn dokumentasjonsfilene hos Posit, men de er lite egnet som pensumtekst. Denne videoen er laget av produsenten og gir en god oversikt over funksjonaliteten. Posit Assistant kan settes opp med API-tilkobling mot GPT UiO, og det er som vil benyttes i hele kurset slik at all KI-bruk skjer innenfor UiO sine godkjente verktøy.

Pensumlisten er relativt kort, men det er ganske krevende saker med veldig mye hands-on. Så det er mye mer jobb enn det kanskje ser ut som ved første blick.

Pensumliste

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021–. <https://doi.org/10.1038/s41562-016-0021>

Kommentar: gir en innføring i viktigheten av dokumentasjon og etterprøvbarehet, som vil ligge som et premiss for all analysen på kurset, inkludert bruk av kunstig intelligens

Wickham, H., Çetinkaya-Rundel, M., & Golemund, G. (2023). *R for data science : import, tidy, transform, visualize, and model data* (2nd edition). O' Reilly. Kapittel 22

Kommentar: Teksten finnes åpent tilgjengelig online. Veldig store data håndteres med arrow-pakken og spesielle filformater, som er kompatible med tidyverse-logikken.

Kosuke Imai and Nora Webb Williams, (2022). *Quantitative Social Science: An Introduction in tidyverse*. Princeton University Press (pp. 238–278).

Kommentar: I *kompedium*. Kapittel 5.2 gir innføring i nettverksdata og sentralitetsmål, mens kapittel 5.3 gir enkle innføring i geografiske data og bruk av kart.

Lyngstad, Torkild (2026) *Liv og død. En introduksjon til demografi* (5, 6 og 11). <https://torkildl.github.io/assets/Liv-og-dod.pdf>

Kommentar: Bok under produksjon. Kapittel 5 og 6 gir analyse av dødelighet og fertilitet, mens kapittel 11 bruker dette til fremskrivninger.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning: with Applications in R (2nd ed. 2021.). Springer Nature. <https://doi.org/10.1007/978-1-0716-1418-1>

Kommentar: I *kompedium*. Kapittelet dekker principal components og k-means clustering.

Posit PBC (2026). Introducing Posit AI for Positron and RStudio, by Posit PBC (p. 29:29) [Video]. Youtube. Retrieved <https://www.youtube.com/watch?v=4er6aua7zyI>

Kommentar: gir en innføring i Posit Assistant integrert i Rstudio

Data Carpentry. (2026). R for Social Scientists. <https://datacarpentry.github.io/r-socialsci/>

Kommentar: Åpent tilgjengelig online. Dekker grunnleggende datahåndtering og grafikk med Tidverse for R.