

R-seminar 5: Oppgaver

STV1020 Vår 2021

Uke 16

Du skal bruke datasettet `wvs_us17.csv` og lineær regresjon for å undersøke relasjonen mellom (avhengig) AV og uavhengig (UV) variabel. Datasettet er en survey gjennomført i USA i 2017 for World Value Survey. Individer svarer på surveyen. Tilsvarende surveyer er gjort for mange andre land og år (se World Value survey sine nettsider for mer informasjon), men i dag bruker vi et datasett som bare inneholder observasjonen fra USA i 2017.

Datasettet inneholder følgende variabler:

- `country`: Land surveyen ble gjennomført i
- `year`: Året surveyen ble gjennomført
- `age`: Alder i år
- `gender`: 1 male, 2 female
- `corruption`: How would you place your views on corruption in your country on a 10-point scale, where 1 means "there is no corruption in my country" and 10 means "there is abundant corruption in my country".
- `imp_democracy`: How important is it for you to live in a country that is governed democratically? On this scale where 1 means it is "not at all important" and 10 means "absolutely important", what position would you choose?
- `income_group`: On this card is an income scale on which 1 indicates the lowest income group and 10 the highest income group in your country. We would like to know in what group your household is. Please, specify the appropriate number, counting all wages, salaries, pensions and other incomes that come in.

Oppgaver

1. Last inn datasettet `wvs_us17.csv` fra github og oppgi antall enheter og variabler i datasettet.
2. Finn navn på variablene i datasettet.

3. Opprett en nytt datasett med kun variablene `imp_democracy`, `age` og `income_group`. Pass på at klassen til variablene er `numeric`.
4. Vis hvordan du fjerner enheter som mangler opplysninger fra datasettet. Oppgi antall enheter i datasettet etter at du har fjernet enhetene.
5. Lag et spredningsplott av sammenhengen mellom `imp_democracy` (AV) og `age` (UV) Selv om `imp_democracy` egentlig har ordinalt målenivå kan du tenke på den som tilnærmet kontinuerlig i de følgende oppgavene.
6. Kjør en lineær regresjonsmodell med `imp_democracy` som AV og `age` som UV.
7. Tolk koeffisienten til `age`.
8. Print resultatet av regresjonen i en tabell ved hjelp av `stargazer()`. Lagre tabellen lokalt på pc-en din og åpne den i f.eks. `word` eller en nettleser.
9. Lag et plott med observerte verdier av `imp_democracy` på y-aksen og observerte verdier av `age` på x-aksen. Legg til en regresjonslinje, endre aksetitlene og legg til en tittel. Lagre plottet lokalt på pc-en din.
10. Lag en ny variabel `gender2` som tar verdien "Female" når observasjonen har verdien "2" på `gender` og "Male" når observasjonen har verdien "1" på `gender`. Sjekk at det ble riktig.
11. Estimer en ny multivariat regresjonsmodell med `imp_democracy` som AV og `age` og `gender 2` som UVs (dette er tema for seminar 6)