

ag comments

hent-ssb-data.Rmd

- Dere har brukt samme navn på chunks flere ganger. Dette gjør at dokumentet ikke kjører. Jeg har endret dem til mutate, mutate1, mutate2 etc.. Ikke særlig informativt, men dokumentet kjører.
- Vær forsiktig med bruk av opsjonene knitr::opts_chunk\$set(echo = TRUE, include = TRUE, eval = TRUE). Settes eval til FALSE vil koden ikke bli kjørt, man vil bruke tidligere genererte verdier og vil ikke få beskjed om eventuelle feil.
- Bruk heller tidyverse funksjonen rename() i en pipe, istedenfor konstruksjoner som `names(pop_08_17_ya)[[1]] <- "knr"`

```
names(pop_08_17_ya)[[2]] <- "aar"
```

```
names(pop_08_17_ya)[[3]] <- "ya_Menn"
```

```
names(pop_08_17_ya)[[4]] <- "ya_Kvinner"
```

– Kan skrives som

```
# langt enklere å se hva som blir gjort
pop_08_17_ya <- pop_08_17_ya %>%
  rename(
    knr = Region,
    aar = Tid,
    ya_Menn = sex1,
    ya_Kvinner = sex2
  )
```

- Tidyverse funksjonen write_csv() er også mer moderne og bedre enn klassisk Rs write.csv() så bruk heller denne.
- Ellers stort sett fin og klar kode

model.Rmd

- Dokumentet kjører på første forsøk. Bra!
- Heteroskedastisitet ii). Dere skriver:
 - I denne testen har vi problemer med heteroskedastisitet, fordi p-verdien er mest sannsynlig større en 0,05.
 - Dette er feil! Poenget er at sannsynligheten for at vi skal observere en slik testverdi hvis H0 (homoskedastisitet) er sann er SVÆRT SVÆRT liten.
 $p\text{-value} < 2.2e-16$
- Diskusjon i) (etter facet plot) . Dere skriver:

Kvaliteten på modellen er ikke helt bra, ettersom det er stor variasjon i grafene.

- Det viktigste her er ikke variasjonen, men systematikken. Vi ser fra plot-ene at enkelte kommuner har positive (eller negative) residualer for *hele* perioden. Modellen over/under-estimerer altså systematisk prisen i disse kommunene. Dette (bias antakelig pga. en eller flere utelatte variabler) er et langt større problem enn tilfeldig variasjon (residualer fordelt rundt 0 linjen) som vi kan betrakte som tilfeldig støy.
- Plot for kommunene “1119”, “1120”, “1127”, “1121”, “1130”, “1135”, “1106”, “1149” punkt ii). Som kommentar til figuren skriver dere:
 - Kommuner som overvurderes på pris per kvadratmeter er de som ligger i nærheten av Stavanger. De som ligger nærmere Haugesund undervurderes.
 - Dette er feil. Husk at det er *residualene* vi her plotter, dvs. kommunene rundt Stavanger har systematisk positive residualer, mens kommunene rundt Haugesund har systematisk negative residualer. Dette innebærer at vår enkle modell systematisk estimerer for *lav* pris rundt Stavnager og for *høy* pris rundt Haugesund. Dette tyder på at det er en viktig underliggende variabel som vi ikke har med i vår enkle modell.