

## Løsningsforslag

Oppgavesettet og løsningsforslaget er utarbeidet av Jarle Møen. Det anbefales at sensorene gir inntil 10 poeng per delspørsmål.

### Oppgave 1

- a) Vi bruker T-test for to uavhengige utvalg. Vi ser at standardavviket til matvarekonsumet for de to gruppene med selvstendig næringsdrivende er relativt likt, og bruker derfor testen som forutsetter lik varians.

$$S_p^2 = \frac{643 * 26172^2 + 2889 * 26986^2}{3532} = 720366788$$

$$T = \frac{37370 - 40769}{\sqrt{720366788 * \left( \frac{1}{644} + \frac{1}{2890} \right)}} = -2,91$$

Kritisk grense for tosidig T-test med 3532 frihetsgrader på 95% nivå er  $\pm 1,96$  og vi forkaster hypotesen om likt matvarekonsum.

(Evt. kan man bruke T-test for ulik varians om man tester for lik varians først.

Konklusjonene blir den samme.)

- b) Dersom logaritmen til matvarekonsumet er normalfordelt er ikke matvarekonsumet selv normalfordelt. I prinsippet er det et problem fordi T-testen forutsetter normalfordelte variabler. Testen vil imidlertid holde som en god tilnærming i store utvalg slik vi har her.
- c) Vi kan bruke enveis variansanalyse – som bygger på normalfordelte variabler – eller Kruskal-Wallis-testen som er et fordelingsfritt alternativ. Et tredje alternativ er å gjøre multiple sammenligninger med Tukeys HSD. Et fjerde alternativ er å bruke regresjon med do dummier.
- d) 
$$\begin{aligned} \text{Var}[\ln(\text{Mat}/\text{Inntekt})] &= \text{Var}[\ln(\text{Mat}) - \ln(\text{Inntekt})] \\ &= \text{Var}[\ln(\text{Mat})] + \text{Var}[\ln(\text{Inntekt})] - 2 * \text{Cov}(\text{Mat}, \text{Inntekt}) \\ &= \text{Var}[\ln(\text{Mat})] + \text{Var}[\ln(\text{Inntekt})] - \\ &\quad 2 * \text{Corr}[\ln(\text{Mat}), \ln(\text{Inntekt})] * \text{SD}(\ln(\text{Mat})) * \text{SD}(\ln(\text{Inntekt})) \\ &= 0,771^2 + 0,807^2 - 2 * 0,6601 * 0,771 * 0,807 \\ &= \underline{0,42} \end{aligned}$$

## Oppgave 2

- a) Kolonne (1) i tabell 3 viser at forskjellen i matvarekonsum mellom lønsmottagere og selvstendig næringsdrivende ikke er signifikant. At forskjellen er liten stemmer med det vi kan lese ut av tabell 1. Tabell 1 viser samtidig at selvstendig næringsdrivende har klart lavere log inntekt. Når vi i tabell 3, kolonne (2) kontrollerer for inntekt ved bruk av regresjon får vi fram at for gitt inntektsnivå har selvstendig næringsdrivende signifikant høyere matvarekonsum enn selvstendig næringsdrivende.
- b) Når inntekten øker med 1 prosent, øker matvarekonsumet med 0,63 prosent, og for gitt inntektsnivå har selvstendig næringsdrivende 12 prosent høyere matvarekonsum enn lønsmottakere. Bokstavelig tolket gir konstantleddet estimert konsum for en husholdning der den med høyest inntekt ikke er selvstendig næringsdrivende. Dersom det ikke finnes observasjoner med tilnærmet null inntekt er konstantleddet ganske enkelt skjæringspunktet med Y-aksen for regresjonslinjen.
- c) Forskjellen i logaritmisk matvarekonsum er  $0,0538 + 0,0492 = 0,103$ . Selvstendige næringsdrivende innenfor bygg og anlegg har altså 10,3 prosent høyere matvareforbruk for gitt inntekt enn andre selvstendige næringsdrivende.
- d) Ved å sammenligne kolonne (4) med kolonne (3) ser vi at variablene BA øker og blir signifikant når vi fjerner variabelen ant\_pers. Det må bety at variablene BA er positivt korrelert med antall personer i husholdningen. Husholdninger der den som har høyest inntekt jobber innenfor bygg og anlegg tenderer altså til å være større enn andre husholdninger. (Antall barn for selvstendige innenfor bygg og anlegg i tabell 1 er konsistent med dette.)
- e) Et ekstra barn øker både variabelen ant\_pers med én og variabelen ant\_barn med én.  $0,153 + 0,0301 = 0,183$ . Følgelig er matvarekonsumet i en tobarnsfamilie i gjennomsnitt 18,3 % høyere enn matvarekonsumet i en ettbarnsfamilie.
- f) 99 % konfidensintervall for  $b^{BA*SE} = 0,0623 \pm t_{0,005}^* 0,0256 = 0,0623 \pm 2,576 * 0,0256 = 0,0623 \pm 0,0659$ , dvs.  $[-0,004, 0,1282]$ . (Koeffisienten er altså ikke signifikant på 1 %-nivå.)
- g) Analysen indikerer at selvstendig næringsdrivende underreporterer sine inntekter, og at problemet er størst blant selvstendig næringsdrivende innenfor bygg og anlegg.
- h) Målefeil i en forklaringsvariabel vil trekke koeffisienten mot null. Følgelig må vi anta at koeffisienten til log inntekt ville vært større om vi hadde brukt permanentinntekten.
- i) Husholdningsspesifikke, uobserverte effekter knyttet til preferanser vil ligge i feilleddet. Gjentatte observasjoner av samme husholdning vil nå ha en felles komponent i feilleddet som skaper en form for autokorrelasjon. Problemet kan løses ved å estimere robuste feilledd eller bruke tilfeldig effekt-estimatoren. (Så lenge den uobserverte effekten ikke er korrelert med inkluderte variabler får vi ikke noe endogenitetsproblem.)

## Oppgave 3

- a)  $S^{2012}=23616$   
 $S^{2013}=0,5*23616+0,5*28016=25816$   
 $S^{2014}=0,5*25816+0,5*25311=25563,5$

Prediksjon for 2015 er 25563,5

PS.: Det er dessverre en trykkfeil på den relevante plansje i kompendiet.

De som har basert seg på den generelle formelen som er oppgitt vil få

$$S_t=0,5*25311+0,5^2*28016+0,5^3*23616=22611,5$$

w i det siste leddet skulle ikke vært der, så rett formel er

$$S_t=0,5*25311+0,5^2*28016+0,5^2*23616=25563,5$$

- b) ARIMA(0,0,0) er hvit støy. Det vil passe dårlig da vi ser at det er sterk autokorrelasjon med høyt strømforbruk i vintermånedene og lavt strømforbruk i sommermånedene.

ARIMA(1,0,0) er en AR(1)-prosess, dvs en stasjonær, førsteordens autoregressiv prosess,  $Y_t=\mu+\phi Y_{t-1}+u_t$ . Det vil kunne fange opp autokorrelasjonsmønsteret som vi ser i grafen og vil være vår foretrukne modell.

ARIMA(0,1,0) er en tilfeldig gang. Det vil ikke passe da det er opplyst at strømforbruket i en husholdning antas å være stasjonært. (Det vil ikke kunne vandre fritt, men må variere rundt en nokså stabil forventningsverdi.)