

SKOLEEKSAMEN MET4



Høst 2024

Dato: 27. november 2024

Tidsrom: 09:00 - 12:00

Antall timer: 3

Foreleser/emneansvarlig kan kontaktes av eksamensvakt på telefon: 92254759

TILLATTE HJELPEMIDLER:

Kalkulator Ja ☒ Nei ☐

Ordbok: én tospråklig ordbok tillatt.

Alle trykte/egenskrevne hjelpemidler.

Antall sider, inkludert forside og vedlegg: 9

OBS: Vedlegg 1 inneholder R-utregninger som kan være relevante i besvarelsen av noen av oppgavene.

Oppgave 1

Gjør sykehus oss friskere? Dette er et spørsmål innen helseøkonomi og i en amerikansk kontekst kan man se for seg at en eldre, fattig populasjon med begrenset tilgang på helsetjenester bruker legevakt og akuttten som primærhelsetjeneste. Når man oppsøker disse stedene, vil noen bli innlagt på sykehus. Denne typen behandling er dyr, fyller opp sykehusene, og potensielt er den ikke veldig effektiv for å forbedre helseutfallene til disse menneskene. Dette kan blant annet være fordi man blir eksponert for andre syke mennesker når man er på sykehus. Samtidig tilbyr sykehusene viktig helsehjelp til disse pasientene, og svaret på vårt initielle spørsmål kan fortsatt være “ja”.

The National Health Interview Survey inneholder spørsmål som kan hjelpe oss med å undersøke dette. Spørsmålene er som følger: “During the past 12 months, was the respondent a patient in a hospital over night?” og “Would you say your health in general is excellent, very good, good, fair, poor?”.¹ Svarene blir operasjonalisert ved å tilordne tallverdier fra 1 til 5, der 1 tilsvarer ‘poor’ og 5 tilsvarer ‘excellent’.

Nedenfor er en tabell over helseutfall fordelt på om de var innlagt på sykehus eller ikke.

Klassifisering av gruppe	N	Gj.snitt helsestatus	Std.error
Sykehus	7,774	3.21	0.014
Ikke sykehus	90,049	3.94	0.003

NB!

$$\text{Standard error} = \frac{\text{Standardavvik}}{\sqrt{N}}$$

- (a) Test om variansen er lik mellom gruppene.
- (b) Test om forventet helsestatus er lik for de som var innlagt på sykehus og de som ikke var innlagt på sykehus.
- (c) Diskuter hvorvidt resultatene fra testen i 1b) kan brukes til å konkludere om sykehus bidrar til å forbedre eller forverre helsen. Svar kort. Maks ett avsnitt.

¹Eksempelet er lånt fra Angrist, J. D., & Pischke, J. S. (2009, s. 13). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Oppgave 2

Skatteunndragelse er et samfunnsproblem som både reduserer inntektene staten kan bruke på velferdsgoder som sykehus, skole og politi m.m., samtidig som det gjør at noen er gratispassasjer på andres bidrag og dermed er skattebyrden ujevnt fordelt mellom folk.

En effektiv metode for å forhindre skatteunndragelse er tredjepartsrapportering. Dette betyr at en tredjepart, som en arbeidsgiver, rapporterer arbeidstakerens lønn og andre økonomiske forhold direkte til Skatteetaten. Dette gjør det vanskelig å unndra skatt.

Selvstendig næringsdrivende er ikke underlagt slik rapportering. For å måle skatteunndragelse blant dem, kan vi se på deres forbruk. Hvis selvstendig næringsdrivende i gjennomsnitt konsumerer mer enn forventet gitt inntekten deres, kan det indikere at de unndrar skatt.²

For å måle unndragelse for selvstendig næringsdrivende setter Bjørkheim og Nygård (2024) opp følgende regresjonslikning:³

$$\log(Y_{it}) = \beta_0 + \beta_1 \log(\text{Wage}_{it}) + \beta_2 \text{SE}_{it} + \beta_3 (\text{SE} \times \text{Male})_{it} + \beta \mathbf{X}_{it} + \epsilon_{it}$$

der

- Y_{it} er husholdning i sine donasjoner til veldedighet i år t , og er et mål på husholdningsforbruk.
- Wage_{it} er husholdning i sin inntekt i år t .
- SE_{it} er en dummy variabel som tar verdien 1 om husholdning i er definert som selvstendig næringsdrivende i år t .
- $(\text{SE} \times \text{Male})_{it}$ er en dummy variabel som tar verdien 1 om husholdning i er definert som selvstendig næringsdrivende i år t hvor den selvstendig næringsdrivende er mann.
- \mathbf{X}_{it} er en vektor for en rekke kontrollvariabler.

Utskrift av regresjonen er presentert i Vedlegg 2, kolonne (1).

Datasettet inneholder informasjon på husholdningsnivå om både selvstendig næringsdrivende og arbeidstakere, lønnen deres og hvor mye de konsumerer hvert år i tillegg til en rekke andre variabler fra 2012 til 2016.

- Bruk utskriften i Vedlegg 2 til å gi en fortolkning av koeffisienten β_1 .**
- Bruk utskriften til å gi en fortolkning av koeffisienten β_2 . Kommenter påstanden om at selvstendig næringsdrivende unndrar skatt.**
- Lag et 95%-konfidensintervall for β_1 i modellen over.**

Forskerne bekymrer seg for at selvstendig næringsdrivende og lønnstakere har ulike preferanser, og dermed har ulikt konsum. Om det er sant vil ikke regresjonsmodellen deres plukke opp skatteunndragelse, men kun det faktum at selvstendig næringsdrivende konsumerer mer eller

²Denne intuisjonen ble først brukt i en artikkel fra Storbritania for å måle skatteunndragelse der, for mer informasjon se Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of public economics*, 39(1), 17-32.

³Bjørkheim, J. B., & Nygård, O. E. (2024). Gender Differences in Tax Evasion: Evidence from Norwegian Administrative Data.

mindre enn lønnstakere. Ettersom de har 5 år med data over de samme husholdningene, kjører de en ny regresjonsmodell.

$$\log(Y_{it}) = \beta_0 + \beta_1 \log(\text{Wage}_{it}) + \beta_2 \text{SE}_{it} + \beta_3 (\text{SE} \times \text{Male})_{it} + \beta \mathbf{X}_{it} + \eta_i + \epsilon_{it} \quad (1)$$

- (d) Hva slags modell er dette, og hvorfor mener forskerne at denne modellen er bedre egnet til å fange opp skatteunndragelse?

Under visse forutsetninger kan vi estimere β_1 , β_2 , β_3 og β i modell (1) ved hjelp av minste kvadraters metode for modellen:

$$\widetilde{\log(Y_{it})} = \beta_1 \widetilde{\log(\text{Wage}_{it})} + \beta_2 \widetilde{\text{SE}_{it}} + \beta_3 \widetilde{(\text{SE} \times \text{Male})_i} + \beta \widetilde{\mathbf{X}_{it}} + \widetilde{\epsilon_{it}} \quad (2)$$

der notasjonen \widetilde{X}_{it} for en generell variabel X_{it} i datasettet betyr følgende transformasjon:

$$\widetilde{X}_{it} = X_{it} - \bar{X}_i = X_{it} - \frac{1}{5} \sum_{t=1}^5 X_{it}$$

- (e) Vis at modell (2) følger fra modell (1). Hvilke typer husholdninger er vi avhengige av å ha i datasettet for å kunne estimere β_2 på denne måten?

Oppgave 3

Hvert år går noen fra å være i arbeid til å bli arbeidsledige. Vi har tilgang på et survey datasett fra USA som inneholder data om 5220 personer som var i jobb da de ble spurt i april 2006.⁴ Dataene inneholder også deres ansettelsesstatus ett år senere, i april 2007, i tillegg til noen flere variabler. Disse variablene er beskrevet i Vedlegg 3 der vi også har tilpasset en logistisk regresjonsmodell til disse dataene (kolonne 1 i regresjonsutskriften).

- (a) Skriv opp modellen i Vedlegg 3 og gi en fortolkning av koeffisienten knyttet til variabelen `race2`.
- (b) Hva er sannsynligheten for at svart, gift kvinne på 27 år, som har en bachelorgrad og tjente \$400 i uken i 2006 er arbeidsledig i 2007?

I 2008 blir det finanskrisen i USA og mange mister jobben. Vi har tilgang på en tilsvarende survey fra 2009 med data på 5412 arbeidstakere. Dette er modellen i kolonne 2 av regresjonsutskriften i Vedlegg 3.

- (c) Sammenlign modellene i Vedlegg 3, hvor modell 1 er fra 2007 og modell 2 er fra 2009. Er det forskjeller i hvilke karakteristikk som er assosiert med arbeidsledighet i normale tider sammenlignet med under en finanskrisen? Svar kort. Maks 1 avsnitt.

⁴Surveyen heter Current Population Survey og kilden er Stock, J. H., & Watson, M. W. (2015). Introduction to econometrics (3rd updated edition)

Vedlegg 1: Diverse R-utregninger

```
qnorm(0.975)
```

```
## [1] 1.959964
```

```
qnorm(0.95)
```

```
## [1] 1.644854
```

```
qnorm(0.90)
```

```
## [1] 1.281552
```

```
qf(0.90, df1 = 7773, df2 = 90048)
```

```
## [1] 1.021512
```

```
qf(0.95, df1 = 7773, df2 = 90048)
```

```
## [1] 1.027693
```

```
qf(0.975, df1 = 7773, df2 = 90048)
```

```
## [1] 1.033076
```

Vedlegg 2: Regresjoner for skatteunndragelse

Table 2: Association between Self-Employment and Charitable Gifts, Pooled Cross-Section

	(1)	(2)	(3)	(4)	(5)	(6)
	All			Single		
	Annual inc.		Positive gifts	Permanent inc.		Positive gifts
	No gift conditioning	Tobit		Positive gifts	Positive gifts	
Log household income	0.777*** (0.004)	4.68*** (0.017)	0.170*** (0.002)	0.176*** (0.003)	0.193*** (0.004)	0.149*** (0.004)
Self-employed	0.427*** (0.010)	1.45*** (0.032)	0.027*** (0.005)	0.027*** (0.005)	-0.026*** (0.010)	-0.032*** (0.010)
Male self-employed	-0.358*** (0.011)	-1.24*** (0.038)	-0.020*** (0.006)	-0.021*** (0.006)	-0.012 (0.014)	-0.015 (0.014)
Female					0.094*** (0.004)	0.091*** (0.004)
k female	1.73 (0.022)	1.36 (0.009)	1.17 (0.033)	1.17 (0.032)	0.872 (0.047)	0.809 (0.056)
k male	1.09 (0.008)	1.04 (0.005)	1.04 (0.021)	1.04 (0.020)	0.821 (0.043)	0.733 (0.051)
pval test $k_f = k_m$	0.000	0.000	0.001	0.000	0.420	0.313
k all	1.24 (0.008)	1.13 (0.004)	1.08 (0.018)	1.08 (0.018)	0.844 (0.032)	0.766 (0.038)
pval test $k = 1$	0.000	0.000	0.000	0.000	0.000	0.000
Controls	YES	YES	YES	YES	YES	YES
N of female self-employed	142,822	142,822	48,117	48,117	9,551	9,551
N of male self-employed	384,148	384,148	101,021	101,021	12,269	12,269
N	6,562,565	6,562,565	1,642,645	1,643,596	355,411	355,651
R^2	0.102	0.000	0.0478	0.0477	0.0317	0.0288

Vedlegg 3: Logistisk regresjonsmodell

Ekstra variabel er definert $age2 = age * age$, og er alder ganget med seg selv.

Documentation for Employment_06_07

These data file contains data on 5220 workers who were survey in the April 2006 Current Population Survey and reported that they were employed. The data file contains their employment status in April 2007, one year later, along with some additional variables.

Variable Name	Description
<i>Variables from the 2007 Survey</i>	
<i>employed</i>	indicator =1 if employed in 2007
<i>unemployed</i>	indicator =1 if unemployed in 2007
<i>Variables from the 2006 Survey</i>	
<i>age</i>	age
<i>female</i>	indicator =1 if female
<i>married</i>	indicator =1 if
<i>race</i>	= 1 if self-identified race = white (only) = 2 if self-identified race = black (only) = 3 if self-identified race was not white (only) or black (only)
<i>union</i>	indicator =1 if a member of a union
<i>ne_states</i>	indicator =1 if from a northeastern state
<i>so_states</i>	indicator =1 if from a southern state
<i>ce_states</i>	indicator =1 if from a central state
<i>we_states</i>	indicator =1 if from a western state
<i>private</i>	indicator =1 if employed in a private firm
<i>government</i>	indicator =1 if employed by the government
<i>self</i>	indicator =1 if self-employed
<i>educ_lths</i>	indicator =1 if highest level of education is less than a high school graduate
<i>educ_hs</i>	indicator =1 if highest level of education is a high school graduate
<i>educ_somcol</i>	indicator =1 if highest level of education is some college
<i>educ_aa</i>	indicator =1 if highest level of education is AA degree
<i>educ_ba</i>	indicator =1 if highest level of education is BA or BS degree
<i>educ_adv</i>	indicator =1 if highest level of education is advanced degree
<i>earnwke</i>	average weekly earnings

```

##
## =====
##                               Dependent variable:
##                               -----
##                               unemployed
##                               (1)         (2)
## -----
## race2                        0.346      0.318
##                               (0.340)    (0.212)
##
## race3                        -0.332     0.107
##                               (0.525)    (0.262)
##
## age                         -0.037     -0.090**
##                               (0.063)    (0.040)
##
## age2                        0.0004     0.001**
##                               (0.001)    (0.0005)
##
## earnwke                     -0.001*    -0.0001
##                               (0.0003)   (0.0002)
##
## married                     -1.377***   -0.305**
##                               (0.269)    (0.149)
##
## female                      -0.642***   -0.456***
##                               (0.246)    (0.145)
##
## educ_hs                     -0.324     -0.473**
##                               (0.341)    (0.217)
##
## educ_somocol                -0.652*    -0.910***
##                               (0.379)    (0.246)
##
## educ_aa                     -0.176     -1.029***
##                               (0.440)    (0.308)
##
## educ_bac                    -1.072**    -0.918***
##                               (0.495)    (0.262)
##
## educ_adv                     -0.700     -1.370***
##                               (0.624)    (0.351)
##
## Constant                    -1.522     -0.147
##                               (1.108)    (0.739)
##

```



```
## -----
## Observations      4,569      4,773
## Log Likelihood    -367.535    -889.857
## Akaike Inf. Crit.  761.070     1,805.713
## =====
## Note:              *p<0.1; **p<0.05; ***p<0.01
```